What is the Consumption-CAPM Missing?
An Information-Theoretic Framework for the Analysis of Asset Pricing Models*

Anisha Ghosh† Christian Julliard‡ Alex P. Taylor§

June 18, 2015

Abstract

We consider asset pricing models in which the SDF can be factorized into an observable component and a potentially unobservable one. Using a relative entropy minimization approach, we non-parametrically extract the most likely SDF and its unobservable component. Empirically, we find it to have a business cycle pattern, and significant correlations with market crashes and the Fama–French factors. Moreover, we derive novel entropy bounds for the SDF that are tighter, and have higher information content, than the existing ones. We show that commonly used consumption-based SDFs correlate poorly with the most likely one, require high risk aversion to satisfy the bounds, and understate market crash risk.

Keywords: Pricing Kernel, Stochastic Discount Factor, Consumption Based Asset Pricing, Entropy Bounds.
JEL Classification Codes: G11, G12, G13, C52.

*We benefited from helpful comments from Mike Chernov, George Constantinides, Darrell Duffie, Bernard Dumas, Burton Hollifield, Ravi Jagannathan, Nobu Kiyotaki, Albert Marcet, Bryan Routledge, Michael Stutzer, and seminar and conference participants at Carnegie Mellon University, the London School of Economics, INSEAD, Johns Hopkins University, 2011 Adam Smith Asset Pricing Conference, 2011 NBER Summer Institute, 2011 Society for Financial Econometrics Conference, 2011 CEPR ESSFM at Gerzensee, 2012 Annual Meeting of the American Finance Association.

†Tepper School of Business, Carnegie Mellon University; anishagh@andrew.cmu.edu.
‡Department of Finance and FMG, London School of Economics, and CEPR; c.julliard@lse.ac.uk.
§Department of Finance, Manchester Business School; alex.taylor@mbs.ac.uk.

I Introduction

The absence of arbitrage opportunities implies the existence of a pricing kernel, also known as the stochastic discount factor (SDF), such that the equilibrium price of a traded security can be represented as the conditional expectation of the future payoff discounted by the pricing kernel. The standard consumption-based asset pricing model, within the representative agent and time-separable power utility framework, identifies the pricing kernel as a simple parametric function of consumption growth. However, pricing kernels based on consumption growth alone cannot explain either the historically observed levels of returns, giving rise to the Equity Premium and Risk Free Rate Puzzles (e.g., Mehra and Prescott (1985), Weil (1989)), or the cross-sectional dispersion of returns between different classes of financial assets (e.g., Hansen and Singleton (1983), Mankiw and Shapiro (1986), Breeden, Gibbons, and Litzenberger (1989), Campbell (1996)).[1]

Nevertheless, there is considerable empirical evidence that consumption risk does matter for explaining asset returns (e.g., Lettau and Ludvigson (2001a, 2001b), Parker and Julliard (2005), Hansen, Heaton, and Li (2008), Savov (2011)). Therefore, a burgeoning literature has developed based on modifying the preferences of investors and/or the structure of the economy. In such models the resulting pricing kernel can be factorized into an observable component consisting of a parametric function of consumption growth, and a potentially unobservable, model-specific, component. Prominent examples in this class include: the external habit model, where the additional component consists of a function of the habit level (Campbell and Cochrane (1999); Menzly, Santos, and Veronesi (2004)); the long run risks model based on recursive preferences, where the additional component consists of the return on total wealth (Bansal and Yaron (2004)); and models with housing risk, where the additional component consists of the growth in the expenditure share on non-housing consumption (Piazzesi, Schneider, and Tuzel (2007)). The additional, and potentially unobserved, component may also capture deviations from rational expectations (e.g., Brunnermeier and Julliard (2007)), models with robust control (e.g., Hansen and Sargent (2010)), heterogeneous agents (e.g., Constantinides and Duffie (1996)), ambiguity aversion (e.g., Ulrich (2010)), as well as a liquidity factor arizing from solvency constraints

[1] Recently, Julliard and Ghosh (2012) show that pricing kernels based on consumption growth alone cannot explain either the equity premium puzzle, or the cross-section of asset returns, even after taking into account the possibility of rare disasters.

(e.g., Lustig and Nieuwerburgh (2005)).

In this paper, we propose a new methodology to analyze dynamic asset pricing models, such as those described above, for which the SDF can be factorized into an observable component and a potentially unobservable one. Our no-arbitrage approach allows us to: *a*) filter from the data the *most likely* estimate of the time series of the unobserved pricing kernel; *b*) construct entropy bounds to assess the empirical plausibility of candidate SDFs; *c*) estimate, given a fully observable pricing kernel, the minimum (in the information sense) adjustment of the SDF needed to correctly price asset returns. This methodology provides useful diagnostic tools for studying the ways in which various models might fail empirically, and allows us to characterize some properties that a successful model must satisfy.

First, we show that, given a set of asset returns and consumption data, a relative entropy minimization approach can be used to extract, non-parametrically, the time series of both the SDF and its unobservable component (if any). This methodology identifies the most likely, in an information-theoretic sense, time series of the SDF and its unobservable component: we show that the entropy minimization approach is equivalent to maximizing the expected risk neutral likelihood under a set of no arbitrage restrictions. Moreover, given a fully observable pricing kernel, this procedure identifies the most likely modification of the SDF that enables it to price asset returns correctly, while adding to it a *minimum* amount of extra information. Along this dimension our paper is close in spirit to, and innovates upon, the long tradition of using asset prices (mostly options) to estimate the risk neutral probability measure (see, e.g., Jackwerth and Rubinstein (1996), and Ait-Sahalia and Lo (1998)) and use this information to extract an implied pricing kernel (see, e.g., Ait-Sahalia and Lo (2000), Rosenberg and Engle (2002), and Ross (2011)).

Empirically, our estimated time series for the unobservable pricing kernel is substantially (but far from perfectly) correlated with the Fama and French (1993) factors, for a variety of sample frequencies and assets used in the estimation (even using only assets, like the industry and momentum portfolios, that are not well priced by the Fama–French factors).[2] This suggests that our approach successfully identifies the pricing kernel, and provides a rationalization of the empirical success of the Fama–French factors. The estimated most

[2]This correlation ranges from .45 to .81 when Fama–French portfolios are used in the estimation of the most likely SDF, while it is reduced to the .43–.70 range when considering only industry or momentum portfolios.

likely SDF has a clear business cycle pattern but also shows significant and sharp reactions to stock market crashes (even if these crashes do not result in economy-wide contractions). Moreover, we show that while the SDFs of most of the equilibrium models tend to adequately account for business cycle risk, they nevertheless fail to show significant reactions to market crashes, and this hampers their ability to price asset returns—that is, all models seem to be missing a market crash risk component.

Second, we construct entropy bounds that restrict the admissible regions for the SDF and its unobservable component. Our results complement and improve upon the seminal work by Hansen and Jagannathan (1991), which provides minimum variance bounds for the SDF, and Hansen and Jagannathan (1997) (the so called *second* Hansen–Jagannathan distance), which identifies the minimum variance (linear) modification of a candidate pricing kernel needed for it to be consistent with asset returns. The use of an entropy metric is also closely related to Stutzer (1995, 1996), which first suggested constructing entropy bounds based on asset pricing restrictions, and Alvarez and Jermann (2005), who derive a lower bound for the volatility of the permanent component of investors' marginal utility of wealth (see also Backus, Chernov, and Zin (2011), Bakshi and Chabi-Yo (2011) and Kitamura and Stutzer (2002)). We show that a second order approximation of the risk neutral entropy bounds (Q bounds) have the canonical Hansen–Jagannathan (HJ) bounds as a special case, but are generally tighter since they naturally impose a non-negativity restriction on the pricing kernel. Using the multiplicative structure of the pricing kernel, we are able to provide novel bounds (M bounds) that have higher information content, and are tighter, than the risk neutral entropy bounds and those of Hansen and Jagannathan (1991). Moreover, our approach improves upon Alvarez and Jermann (2005) in that a decomposition of the pricing kernel into permanent and transitory components is not required (but is still possible), and we can accommodate an asset space of arbitrary dimension.

Our methodology can also be used to construct bounds (Ψ bounds) for the potentially unobserved component of the pricing kernel. We show that for models in which the pricing kernel is a function only of observable variables, the Ψ bounds are the tightest ones, and can be satisfied if and only if the model is actually able to correctly price assets. Moreover, when the pricing kernel is fully observable, our Ψ bounds are closely related to the second Hansen–Jagannathan distance: HJ identify the minimum variance *linear* adjustment, while our approach identifies the minimum entropy *multiplicative* (or log-linear) adjustment that would make a candidate pricing kernel consistent with the observed asset returns. We show that the key difference between the two approaches is that the entropy one focuses not only on the second moment deviations, but also on all other higher moments. In an empirical example using stock return data we find that these higher moments play an important role: driving about 22%–26% of the entropy of the most likely pricing kernel.

Third, we demonstrate how our methodology provides useful diagnostic tools to assess the plausibility of some of the most well known consumption-based asset pricing models, and lends new insights into their empirical performance. For the standard time separable power utility model, we show that the pricing kernel satisfies the Hansen and Jagannathan (1991) bound for large values of the risk aversion coefficient, and the Q and M bounds for even higher levels of risk aversion. However, the Ψ bound is tighter and is not satisfied for *any* level of risk aversion. We show that these findings are robust to the use of the long run consumption risk measure of Parker and Julliard (2005), despite the fact that this measure of consumption risk is able to explain a substantial share of the cross-sectional variation in asset returns with a small risk aversion coefficient. Considering more general models of dynamic economies, such as models with habit formation, long run risks in consumption growth, and complementarities in consumption, we find that the SDFs implied by all of them *a*) correlate poorly with the filtered most likely SDF, *b*) require implausibly high levels of risk aversion to satisfy the entropy bounds, and *c*) tend to understate market crash risk, in particular the risk associated with market crashes that do not result in recessions. Moreover, the empirical application illustrates that inferences based on the entropy bounds deliver results that are much more stable, in evaluating the plausibility of a given model across different sets of assets and data frequencies, than the cross-sectional R^2 (which, instead, tends to vary wildly for the same model).

Compared to the previous literature, our nonparametric approach offers five main advantages: *i*) it can be used to extract information not only from options, as is common in the literature, but also from any type of financial asset; *ii*) instead of relying exclusively on the information contained in financial data, it allows us to also exploit the information about the pricing kernel contained in the time series of aggregate consumption, thereby connecting our results to macro-finance modeling; *iii*) the relative entropy extraction of the

SDF is akin to a nonparametric maximum likelihood procedure and thereby provides the most likely estimate of its time series; *iv*) the methodology has considerable generality, and may be applied to any model that delivers well-defined Euler equations and for which the SDF can be factorized into an observable component and an unobservable one (these include investment-based asset pricing models, and models with heterogenous agents, limited stock market participation, and fragile beliefs); *v*) it relies not only on the second moment of the pricing kernel, but also on all higher moments.

The remainder of the paper is organized as follows. Section II presents the information-theoretic methodology, the entropy bounds developed, and their properties. Section III uses the Consumption-CAPM with power utility as an illustrative example of the application of our methodology. Section IV applies the diagnostic tools developed in this paper to the analysis of more general models of dynamic economies. Section V concludes and discusses extensions. The Appendix contains proofs, additional empirical results and theoretical details, and a thorough data description.

II Entropy and the Pricing Kernel

In the absence of arbitrage opportunities, there exists a strictly positive pricing kernel, M_{t+1}, or stochastic discount factor (SDF), such that the equilibrium price, P_{it}, of any asset i delivering a future payoff, X_{it+1}, is given by

$$P_{it} = \mathbb{E}_t \left[M_{t+1} X_{it+1} \right]. \tag{1}$$

where $\mathbb{E}_t$ is the rational expectation operator conditional on the information available at time t. For a broad class of models, the SDF can be factorized as follows

$$M_t = m(\theta, t) \times \psi_t \tag{2}$$

where $m(\theta, t)$ denotes the time t value of a known, strictly positive, function of observable data and the parameter vector $\theta \in \Theta \subseteq \mathbb{R}^k$ with true value θ_0, and ψ_t is a potentially unobservable component. In the most common case, $m(\theta, t)$ is simply a function of consumption growth, i.e., $m(\theta, t) = m(\theta, \Delta c_t)$, where $\Delta c_t \equiv \log \frac{C_t}{C_{t-1}}$ and C_t denotes the consumption flow at time t.

Equations (1) and (2) imply that, for any set of tradable assets, the following vector of Euler equations must hold in equilibrium

$$\mathbf{0} = \mathbb{E}\left[m\left(\theta, t\right)\psi_t \mathbf{R}_t^e\right] \equiv \int m\left(\theta, t\right)\psi_t \mathbf{R}_t^e dP \tag{3}$$

where $\mathbb{E}$ is the unconditional rational expectation operator,[3] $\mathbf{R}_t^e \in \mathbb{R}^N$ is a vector of excess returns on different tradable assets, and P is the unconditional physical probability measure. Under weak regularity conditions, the above pricing restrictions for the SDF can be rewritten as

$$\mathbf{0} = \int m\left(\theta, t\right)\frac{\psi_t}{\bar{\psi}}\mathbf{R}_t^e\, dP = \int m\left(\theta, t\right)\mathbf{R}_t^e\, d\Psi \equiv \mathbb{E}^{\Psi}\left[m\left(\theta, t\right)\mathbf{R}_t^e\right]$$

where $\bar{x} \equiv \mathbb{E}\left[x_t\right]$, and $\frac{\psi_t}{\bar{\psi}} = \frac{d\Psi}{dP}$ is the Radon–Nikodym derivative of Ψ with respect to P. For the above change of measure to be legitimate, we need absolute continuity of the measures Ψ and P.

Therefore, given a set of consumption and asset returns data, for any θ, one can obtain a non-parametric maximum likelihood estimate (as formally shown in Appendix A.1) of the Ψ probability measure as follows:

$$\Psi^*(\theta) \equiv \arg\min_{\Psi} D\left(\Psi || P\right) \equiv \arg\min_{\Psi} \int \frac{d\Psi}{dP}\ln\frac{d\Psi}{dP}dP \;\; \text{s.t. } \mathbf{0} = \mathbb{E}^{\Psi}\left[m\left(\theta, t\right)\mathbf{R}_t^e\right], \tag{4}$$

where, for any two absolutely continuous probability measures A and B, $D\left(A||B\right) := \int \ln\frac{dA}{dB}dA \equiv \int \frac{dA}{dB}\ln\frac{dA}{dB}dB$ denotes the relative entropy of A with respect to B, i.e., the Kullback–Leibler Information Criterion (KLIC) divergence between the measures A and B (White (1982)). Note that $D\left(A||B\right)$ is always non-negative, and has a minimum at zero that is reached when A is identical to B. This divergence measures the additional information content of A relative to B and, as pointed out by Robinson (1991), is very sensitive to any deviation of one probability measure from another. Therefore, the above equation is a relative entropy minimization under the asset pricing restrictions coming from the Euler equations. That is, the minimization in Equation (4) estimates the unknown measure Ψ as the one that adds the minimum amount of additional information needed for the pricing

[3] Our setting can accomodate departures from rational expectations as long as the objective and subjective probability measures are absolutely continuos (i.e., as long as the two measures have the same zero probability sets). If agents had subjective beliefs of this type, Equation (3) would still hold, with $\mathbb{E}$ denoting rational expectations, but ψ_t would contain a change of measure element capturing the discrepancy between subjective beliefs and the rational expectations (see, e.g., Hansen (2014, footnote 35) and Basak and Yan (2010)).

kernel to price assets.

To understand the information-theoretic interpretation of the estimator of Ψ, let F be the set of all probability measures on $\mathbb{R}^{N+N'}$, where N' denotes the dimensionality of the observables in $m(\theta, t)$, and for each parameter vector $\theta \in \Theta$, define the following set of probability measures $\Psi(\theta) \equiv \left\{\psi \in F : \mathbb{E}^{\psi}\left[m(\theta, t)\, \mathbf{R}_t^e\right] = \mathbf{0}\right\}$ which are also absolutely continuous with respect to the physical measure P in Equation (3). If the observable component of the SDF, $m(\theta, t)$, correctly prices assets at the given value of θ, we have that $P \in \Psi(\theta)$, and P solves Equation (4) delivering a KLIC value of 0. On the other hand, if $m(\theta, t)$ is not sufficient to price assets, P is not an element of $\Psi(\theta)$ and there is a positive KLIC distance $D(\Psi||P) > 0$ attained by the solution $\Psi^*(\theta)$. Thus, the estimation approach searches for a $\Psi^*(\theta)$ that adds the minimum amount of additional information needed for the pricing kernel to price asset returns.

The above approach can also be used, as first suggested by Stutzer (1995), to recover the risk neutral probability measure (Q) from the data as

$$Q^* \equiv \arg\min_Q D(Q||P) \equiv \arg\min_Q \int \frac{dQ}{dP} \ln \frac{dQ}{dP} dP \;\; \text{s.t. } \mathbf{0} = \int \mathbf{R}_t^e dQ \equiv \mathbb{E}^Q\left[\mathbf{R}_t^e\right] \tag{5}$$

under the restriction that Q and P are absolutely continuous.

The definition of relative entropy, or KLIC, implies that this discrepancy metric is not symmetric, that is, generally $D(A||B) \neq D(B||A)$ unless A and B are identical (hence their divergence is always zero).[4] This implies that for measuring the information divergence between Ψ and P, as well as between Q and P, we can also invert the roles of Ψ and P in Equation (4), and the roles of Q and P in Equation (5), to recover Ψ and Q as

$$\Psi^*(\theta) \equiv \arg\min_\Psi D(P||\Psi) \equiv \arg\min_\Psi \int \ln \frac{dP}{d\Psi} dP \;\; \text{s.t. } \mathbf{0} = \mathbb{E}^\Psi\left[m(\theta, t)\, \mathbf{R}_t^e\right], \tag{6}$$

$$Q^* \equiv \arg\min_Q D(P||Q) \equiv \arg\min_Q \int \ln \frac{dP}{dQ} dP \;\; \text{s.t. } \mathbf{0} = \mathbb{E}^Q\left[\mathbf{R}_t^e\right]. \tag{7}$$

The divergence $D(P||\Psi)$ can be thought of as the information loss from measure Ψ to measure P (and similarly for $D(P||Q)$). This alternative approach, once again, chooses Ψ

[4] Information theory provides an intuitive way of understanding the asymmetry of the KLIC: $D(A||B)$ can be thought of as the expected minimum amount of extra information bits necessary to encode samples generated from A when using a code based on B (rather than using a code based on A). Hence generally $D(A||B) \neq D(B||A)$ since the latter, by the same logic, is the expected information gain necessary to encode a sample generated from B using a code based on A.

and Q such that assets are priced correctly and such that the estimated probability measures are as close as possible (i.e., minimizing the information loss of moving from one measure to the other) to the physical probability measure P.

Note that the approaches in Equations (4) and (6) identify $\{\psi_t\}_{t=1}^{T}$ only up to a positive scale constant. Nevertheless, this scaling constant can be recovered from the Euler equation for the risk free asset (if one is willing to assume that such an asset is observable).

But why should relative entropy minimization be an appropriate criterion for recovering the unknown measures Ψ and Q? There are several reasons for this choice.

First, as formally shown in Appendix A.1, the approaches in Equations (4)–(7) deliver *non-parametric maximum likelihood* estimates of the Q and Ψ measures—that is, the *most likely* estimate given the data at hand, without having to make parametric assumptions about the functional form of the SDF in the case of Equations (5) and (7) or the ψ component of the SDF in the case of Equations (4) and (6). That is, the above KLIC minimization is equivalent to maximizing the likelihood in an unbiased procedure for finding the pricing kernel or the ψ_t component of the pricing kernel. Note that this is also the rationale behind the *principle of maximum entropy* (see, e.g., Jaynes (1957a, 1957b)) in the physical sciences and Bayesian probability, which states that subject to known testable constraints (in our case, the asset pricing Euler restrictions), the probability distribution that best represents our knowledge is the one with maximum entropy, or minimum relative entropy in our notation.

Second, the use of relative entropy, due to the presence of the logarithm in the objective functions in Equations (4)–(7), naturally forces the the pricing kernel to be non-negative. This, for example, is not imposed in the identification of the minimum variance pricing kernel of Hansen and Jagannathan (1991).[5]

Third, our approach to uncover the ψ_t component of the pricing kernel satisfies the criterion of Occam's razor (the law of parsimony), since it adds the *minimum amount of information* needed for the pricing kernel to price assets. This is due to the fact that the relative entropy is measured in units of information.

Fourth, it is straightforward to add conditioning information to construct a conditional

[5]Hansen and Jagannathan (1991) offer an alternative bound that imposes this restriction, but it is computationally cumbersome (the minimum variance portfolio is basically an option in this case). See also Hansen, Heaton, and Luttmer (1995).

version of the entropy bounds presented in the next section: given a vector of conditioning variables $\mathbf{Z}_{t-1}$, one simply has to multiply (element by element) the argument of the integral constraints in Equations (4), (5), (6) and (7) by the conditioning variables in $\mathbf{Z}_{t-1}$.

Fifth, there is no ex-ante restriction on the number of assets that can be used in constructing ψ_t, and the approach can naturally handle assets with negative expected rates of return (cf. Alvarez and Jermann (2005)).

Sixth, as implied by Brown and Smith (1990), the use of entropy is desirable if we think that tail events are an important component of the risk measure.[6]

Finally, this approach is numerically simple when implemented via duality (see, e.g., Csiszar (1975)). That is, when implementing the entropy minimization in Equation (4), each element of the series $\{\psi_t\}_{t=1}^{T}$ can be estimated, up to a positive constant scale factor, by

$$\psi_t^*(\theta) = \frac{e^{\lambda(\theta)' m(\theta,t)\mathbf{R}_t^e}}{\sum_{t=1}^{T} e^{\lambda(\theta)' m(\theta,t)\mathbf{R}_t^e}}, \quad \forall t \tag{8}$$

where $\lambda(\theta) \in \mathbb{R}^N$ is the solution to the following unconstrained convex problem

$$\lambda(\theta) \equiv \arg\min_{\lambda} \frac{1}{T}\sum_{t=1}^{T} e^{\lambda' m(\theta,t)\mathbf{R}_t^e}, \tag{9}$$

and this last expression is the dual formulation of the entropy minimization problem in Equation (4).

Similarly, the entropy minimization in Equation (6) is solved by

$$\psi_t^*(\theta) = \frac{1}{T(1 + \lambda(\theta)' m\,(\theta, t)\,\mathbf{R}_t^e)}, \quad \forall t \tag{10}$$

where $\lambda(\theta) \in \mathbb{R}^N$ is the solution to

$$\lambda(\theta) \equiv \arg\min_{\lambda} -\sum_{t=1}^{T} \log(1 + \lambda' m\,(\theta, t)\,\mathbf{R}_t^e), \tag{11}$$

and this last expression is the dual formulation of the entropy minimization problem in Equation (6).

[6] Brown and Smith (1990) develop what they call "a Weak Law of Large Numbers for rare events,"that is, they show that the empirical distribution that would be observed in a very large sample converges to the distribution that minimizes the relative entropy.

Note also that the above duality results imply that the number of free parameters available in estimating $\{\psi\}_{t=1}^{T}$ is equal to the dimension of (the Lagrange multiplier) λ—that is, it is simply equal to the number of assets considered in the Euler equation.

Moreover, since the $\lambda(\theta)$s in Equations (9) and (11) are akin to Extremum Estimators (see, e.g., Hayashi (2000, Ch. 7)), under standard regularity conditions (see, e.g., Amemiya (1985, Theorem 4.1.3)), one can construct asymptotic confidence intervals for both $\{\psi_t\}_{t=1}^{T}$ and the entropy bounds presented in the next section.

To summarize, we estimate the ψ_t component of the SDF non-parametrically, using the relative entropy minimizing procedures in Equations (4) and (6). The estimate $\{\psi_t^*(\theta)\}_{t=1}^{T}$ is then multiplied by the observable component $m(\theta, t)$ to obtain the overall SDF, $M_t^* = m(\theta, t)\,\psi_t^*(\theta)$. Since we have proposed two different relative entropy minimization approaches, we get two different estimates of the most likely SDF given the data. Asymptotically, the two should be equal because of the MLE property of these procedures. Nevertheless, in any finite sample they could potentially be very different. As shown in our empirical analysis, the two estimates are very close to each other, suggesting that their asymptotic behavior is well approximated in our sample.

II.1 Entropy Bounds

Based on the relative entropy estimation of the pricing kernel and its component ψ outlined in the previous section, we now turn our attention to the derivation of a set of entropy bounds for the SDF, M, and its components.

Dynamic equilibrium asset pricing models identify the SDF as a parametric function of variables determined by the consumers' preferences and the state variables driving the economy. A substantial research effort has been devoted to developing diagnostic methods to assess the empirical plausibility of candidate SDFs, as well as to provide guidance for the construction and testing of other—more realistic—asset pricing theories.

The seminal work of Hansen and Jagannathan (1991) identifies, in a model-free no-arbitrage setting, a variance minimizing benchmark SDF, $M_t^*\left(\bar{M}\right)$, whose variance places a lower bound on the variances of other admissible SDFs:

Definition 1 (Canonical HJ bound) *For each* $\mathbb{E}[M_t] = \bar{M}$*, the Hansen and Jagan-*

nathan (1991) minimum variance SDF is

$$M_t^*\left(\bar{M}\right) \equiv \underset{\left\{M_t\left(\bar{M}\right)\right\}_{t=1}^{T}}{\arg\min} \sqrt{Var\left(M_t\left(\bar{M}\right)\right)} \ s.t.\ \mathbf{0} = \mathbb{E}\left[\mathbf{R}_t^e M_t\left(\bar{M}\right)\right]. \qquad (12)$$

The solution to the above minimization is $M_t^*\left(\bar{M}\right) = \bar{M} + \left(\mathbf{R}_t^e - \mathbb{E}\left[\mathbf{R}_t^e\right]\right)' \beta_{\bar{M}}$, *where* $\beta_{\bar{M}} = Cov\left(\mathbf{R}_t^e\right)^{-1}\left(-\bar{M}\mathbb{E}\left[\mathbf{R}_t^e\right]\right)$, *and any candidate stochastic discount factor* M_t *must satisfy* $Var\left(M_t\left(\bar{M}\right)\right) \geq Var\left(M_t^*\left(\bar{M}\right)\right)$.

The HJ bound offers a natural benchmark for evaluating the potential of an equilibrium asset pricing model since, by construction, any SDF that is consistent with the observed data should have a variance that is not smaller than that of $M_t^*\left(\bar{M}\right)$. However, the identified minimum variance SDF does not impose a non-negativity constraint on the pricing kernel. In fact, since $M_t^*\left(\bar{M}\right)$ is a linear function of the returns, the restrction is not generally natisfied.[7]

As noticed in Stutzer (1995), using the Kullback–Leibler Information Criterion minimization in Equation (5), one can construct an entropy bound for the risk neutral probability measure that naturally imposes the non-negativity constraint on the pricing kernel. We generalize the idea of using an entropy minimization approach to construct risk neutral bounds—Q bounds—for the pricing kernel.

In what follows, for a given risk neutral probability measure Q with Radon–Nikodym derivative $\frac{dQ}{dP} = \frac{M_t}{\bar{M}}$, we use $D\left(P||Q\right)$ and $D\left(P||\frac{M_t}{\bar{M}}\right)$ interchangeably, i.e., $D\left(P||\frac{M_t}{\bar{M}}\right) \equiv D\left(P||Q\right) \equiv \int \ln\left(\frac{dP}{dQ}\right) dP \equiv -\int \ln\left(\frac{M_t}{\bar{M}}\right) dP$. Similarly, $D\left(\frac{M_t}{\bar{M}}||P\right) \equiv D\left(Q||P\right) \equiv \int \ln\left(\frac{dQ}{dP}\right) dQ \equiv \int \frac{dQ}{dP} \ln\left(\frac{dQ}{dP}\right) dP \equiv \int \frac{M_t}{\bar{M}} \ln\left(\frac{M_t}{\bar{M}}\right) dP$.

Definition 2 (Q bounds) *We define the following risk neutral probability bounds for any candidate stochastic discount factor* M_t*:*

1. *The Q1 bound:*
$$D\left(P||\frac{M_t}{\bar{M}}\right) \equiv \int -\ln\frac{M_t}{\bar{M}} dP \geqslant D\left(P||Q^*\right)$$
where Q^* *solves Equation (7).*

[7] We call the bound in Definition 1 the "canonical" HJ bound since Hansen and Jagannathan (1991, 1997) also provide an alternative bound, which forces the non-negativity of the pricing kernel, but that is computationally more complex.

2. *The Q2 bound (Stutzer (1995)):*

$$D\left(\frac{M_t}{\bar{M}}||P\right) \equiv \int \frac{M_t}{\bar{M}} \ln \frac{M_t}{\bar{M}} dP \geqslant D\left(Q^*||P\right)$$

where Q^* solves Equation (5).

These bounds, like the HJ bound, use only the information contained in the asset returns but, unlike the latter, they impose the restriction that the pricing kernel must be positive. Moreover, under mild regularity conditions, we show that (see Remark 2 in Appendix A.2), to a second order approximation, the problem of constructing canonical HJ bounds and Q bounds are equivalent, in the sense that approximate Q bounds identify the minimum variance bound for the SDF.[8] The intuition behind this result is simple: a) a second order approximation of (the log of) a smooth pdf results in an approximately Gaussian distribution (see, e.g., Schervish (1995)); b) the relative entropy of a Gaussian distribution is proportional to its variance; c) the *diffusion invariance principle* (see, e.g., Duffie (2005, Appendix D)) implies that in the continuous time limit the (equivalent) change of measure does not change the volatility.

Both the HJ and Q bounds use information about asset returns but do not utilize information about consumption growth or the structure of the pricing kernel. Instead, we propose a novel approach that, while also forcing the non-negativity of the pricing kernel, a) takes into account more information about the form of the kernel, therefore yielding sharper bounds, and b) allows us to construct bounds for both the pricing kernel as a whole and for its individual components.

Consider an SDF that, as in Equation (2), can be factorized into two components, i.e., $M_t = m(\theta, t) \times \psi_t$ where $m(\theta, t)$ is a known non-negative function of observable variables (generally consumption growth) and the parameter vector θ, and ψ_t is a potentially unobservable component. A large class of equilibrium asset pricing models, including ones with time separable power utility with a constant coefficient of relative risk aversion, external habit formation, recursive preferences, durable consumption goods, housing, and disappointment aversion, fall into this framework. Based on the above factorization of the SDF we

[8]The (sufficient, but not necessary) regularity conditions required for this approximation result are typically satisfied in consumption-based asset pricing models.

can define the following bounds.

Definition 3 (M bounds) *For any candidate stochastic discount factor of the form of Equation (2), and given any choice of the parameter vector θ, we define the following bounds:*

1. *The $M1$ bound:*

$$D\left(P||\frac{M_t}{\bar{M}}\right) \equiv \int -\ln\frac{M_t}{\bar{M}}dP \geqslant D\left(P||\frac{m(\theta,t)\,\psi_t^*}{\overline{m(\theta,t)\,\psi_t^*}}\right) \equiv \int -\ln\frac{m(\theta,t)\,\psi_t^*}{\overline{m(\theta,t)\,\psi_t^*}}dP$$

where ψ_t^* solves Equation (6) and $\overline{m(\theta,t)\,\psi_t^*} \equiv \mathbb{E}\left[m(\theta,t)\,\psi_t^*\right].$

2. *The $M2$ bound:*

$$D\left(\frac{M_t}{\bar{M}}||P\right) \equiv \int \frac{M_t}{\bar{M}}\ln\frac{M_t}{\bar{M}}dP \geqslant D\left(\frac{m(\theta,t)\,\psi_t^*}{\overline{m(\theta,t)\,\psi_t^*}}||P\right) \equiv \int \frac{m(\theta,t)\,\psi_t^*}{\overline{m(\theta,t)\,\psi_t^*}}\ln\frac{m(\theta,t)\,\psi_t^*}{\overline{m(\theta,t)\,\psi_t^*}}dP$$

where ψ_t^* solves Equation (4).

The above bounds for the SDF are tighter than the Q bounds since, denoting by Q^* the minimum entropy risk neutral probability measure,

$$D\left(P||\frac{m(\theta,t)\,\psi_t^*}{\overline{m(\theta,t)\,\psi_t^*}}\right) \geq D\left(P||Q^*\right) \text{ and } D\left(\frac{m(\theta,t)\,\psi_t^*}{\overline{m(\theta,t)\,\psi_t^*}}||P\right) \geq D\left(Q^*||P\right) \qquad (13)$$

by construction, and are also more informative since not only is the information contained in the asset returns used in their construction, but also a) the structure of the pricing kernel in Equation (2) and b) the information contained in $m(\theta,t)$.

Information about the SDF can also be elicited by constructing bounds for the ψ_t component itself. Given the $m(\theta,t)$ component, these bounds identify the minimum amount of information that should be added to ψ_t for the pricing kernel M_t to be able to price asset returns.[9]

Definition 4 (Ψ bounds) *For any candidate stochastic discount factor of the form of Equation (2), and given any choice of the parameter vector θ, we define two lower bounds for the relative entropy of ψ_t:*

[9] As with the Q and M bounds, we use interchangeably $D(P||\Psi)$ and $D\left(P||\frac{\psi_t}{\bar{\psi}}\right)$, as well as $D(\Psi||P)$ and $D\left(\frac{\psi_t}{\bar{\psi}}||P\right)$.

1. *The $\Psi 1$ bound:*

$$D\left(P||\frac{\psi_t}{\bar{\psi}}\right) \equiv -\int \ln \frac{\psi_t}{\bar{\psi}} dP \geqslant D\left(P||\frac{\psi_t^*}{\bar{\psi}^*}\right)$$

where ψ_t^* solves Equation (6);

2. *The $\Psi 2$ bound*

$$D\left(\frac{\psi_t}{\bar{\psi}}||P\right) \equiv \int \frac{\psi_t}{\bar{\psi}} \ln \frac{\psi_t}{\bar{\psi}} dP \geqslant D\left(\frac{\psi_t^*}{\bar{\psi}^*}||P\right)$$

where ψ_t^* solves Equation (4).

Besides providing an additional check for any candidate SDF, the Ψ bounds are useful in that a simple comparison of $D\left(\frac{\psi_t^*}{\bar{\psi}^*}||P\right)$, $D\left(\frac{m(\theta,t)}{\overline{m(\theta,t)}}||P\right)$ and $D\left(Q^*||P\right)$ can provide a very informative decomposition in terms of the entropy contribution to the pricing kernel, which is logically similar to the widely used variance decomposition analysis. For example, if $D\left(\frac{\psi_t^*}{\bar{\psi}^*}||P\right)$ happens to be close to $D\left(Q^*||P\right)$, while $D\left(\frac{m(\theta,t)}{\overline{m(\theta,t)}}||P\right)$ is substantially smaller, the decomposition implies that most of the ability of the candidate SDF to price assets comes from the ψ_t component.

We note that in principle a volatility bound, similar to the Hansen and Jagannathan (1991) bound for the pricing kernel, can be constructed for the ψ_t component. Such a bound, presented in Definition 5 of Appendix A.2, identifies a minimum variance $\psi_t^*\left(\bar{\psi}^*\right)$ component with standard deviation given by

$$\sigma_{\psi^*} = \bar{\psi}^* \sqrt{\mathbb{E}\left[\mathbf{R}_t^e m\left(\theta,t\right)\right]' Var\left(\mathbf{R}_t^e m\left(\theta,t\right)\right)^{-1} \mathbb{E}\left[\mathbf{R}_t^e m\left(\theta,t\right)\right]}. \tag{14}$$

This bound, like the entropy based Ψ bounds in Definition 4, uses information about the structure of the SDF but, unlike the latter, does not constrain ψ_t and M_t to be non-negative as implied by economic theory. Moreover, using the same approach employed in Remark 2, this last bound can be obtained as a second order approximation of the entropy based Ψ bounds in Definition 4.

Equation (14), viewed as a second order approximation to the entropy Ψ bounds, also makes clear why bounds based on the decomposition of the pricing kernel as $M_t = m\left(\theta,t\right)\psi_t$ offer sharper inference than bounds based on only M_t. Consider for example the case in which the candidate SDF is of the form $M_t = m\left(\theta,t\right)$, that is, $\psi_t = 1$ for any t. In this case, it can easily happen that there exists a $\tilde{\theta}$ such that

$$Var\left(M_t\left(\tilde{\theta}\right)\right) \equiv Var\left(m\left(\tilde{\theta},t\right)\right) \geq Var\left(M_t^*\left(\bar{M}\right)\right)$$

where $Var\left(M_t^*\left(\bar{M}\right)\right)$ is the HJ bound in Definition 1, that is there exists a $\tilde{\theta}$ such that the HJ bound is satisfied. Nevertheless, the existence of such a $\tilde{\theta}$ does not imply that the candidate SDF is able to price asset returns. This would be the case if and only if the volatility bound for ψ_t is also satisfied since, from Equation (14), we have that under the assumption of constant ψ_t the bound can be satisfied only if $\mathbb{E}\left[\mathbf{R}_t^e m\left(\theta_0,t\right)\right] \equiv \mathbb{E}\left[\mathbf{R}_t^e M_t\left(\theta_0\right)\right] = \mathbf{0}$, that is, only if the candidate SDF is able to price asset returns.

II.1.1 Residual ψ and the Second Hansen–Jagannathan Distance

If we want to evaluate a model of the form $M_t = m\left(\theta,t\right)$, i.e., a model without an unobservable component, the Ψ bounds will offer a tight selection criterion since, under the null of the model's being true, we should have $D\left(\frac{\psi_t^*}{\psi^*}||P\right) = D\left(P||\frac{\psi_t^*}{\psi^*}\right) = 0$ and this is a tighter bound than the HJ, Q, and M bounds defined above. The intuition for this is simple: Q bounds (and HJ bounds) require the candidate model to deliver at least as much relative entropy (variance) as the minimum relative entropy (variance) SDF, but they do not require that the $m\left(\theta,t\right)$ under scrutiny should also be able to price the assets. That is, it might be the case—as in practice we will show is the case—that for some values of θ both the Q bounds and the HJ bounds will be satisfied, but nevertheless the SDF grossly violates the pricing restrictions in the Euler equation (3).

Note that when considering a model of the form $M_t = m\left(\theta,t\right)$, the estimated ψ^* component is a residual one, i.e., it captures what is missed, for pricing assets correctly, by the pricing kernel under scrutiny. The residual ψ^* and the entropy bounds are also closely related to the *second* Hansen and Jagannathan bound. Given a model that identifies a SDF M, Hansen and Jagannathan (1997) assume that portfolio payoffs are elements of a Hilbert space and consider the minimum squared deviation between M and a pricing kernel $q \in \mathcal{M}$ (or $\mathcal{M}^+$ if non-negativity is imposed), where $\mathcal{M}$ denotes the set of all admissible SDFs.

That is, the second HJ distance is defined as

$$d_{HJ}^2 := \min_{q\in\mathcal{M}} \mathbb{E}\left[(M_t - q_t)^2\right].$$

Note that $q \in \mathcal{M}$ can be rewritten as $q \in L^2$ satisfying the pricing restriction (1), i.e.

$$d_{HJ}^2 \equiv \min_{q\in L^2} \mathbb{E}\left[(M_t - q_t)^2\right] \quad \text{s.t. } \mathbf{0} = \mathbb{E}\left[q_t \mathbf{R}_t^e\right] \equiv \mathbb{E}^Q\left[\mathbf{R}_t^e\right],$$

where the constraint in the above formulation is the same one that we impose for constructing our entropy bounds.

In practice, the second HJ bound looks for the minimum—in a least squares sense—linear adjustment that makes $M_t - \lambda'\mathbf{R}_t^e$ an admissible SDF (where λ arises from the linear projection of M on the space of returns). This idea of the minimum adjustment of the second HJ distance is strongly connected to our M and Ψ bounds and residual ψ.

Consider the decomposition $M_t = m(\theta, t)\,\psi_t$ in its extreme form: $M_t \equiv m(\theta, t)$, i.e., the case in which the candidate SDF is fully observable and, under the null of the model under scrutiny, ψ^m (the model-implied ψ) should simply be a constant. In this case, we can estimate a *residual* $\{\psi_t^*\}_{t=1}^T$ that should be constant if the model is correct. In this case, the $M1$ bound defines the distance

$$d_{M1} = \min_{\{\psi_t\}_{t=1}^T} D\left(P||M_t\psi_t\right) - D\left(P||M_t\right) \equiv \min_{\{\psi_t\}_{t=1}^T} D\left(P||\psi_t\right) \quad \text{s.t. } \mathbf{0} = \mathbb{E}\left[q_t \mathbf{R}_t^e\right]$$

where $q_t := M_t\psi_t$ and we have normalized ψ_t to have unit mean to simplify exposition, and note that the second equality is nothing but the $\Psi1$ bound. Note that in this case we have $\log\psi_t \equiv \log q_t - \log M_t$. That is, while the second HJ distance focuses on the deviation between q and M, our entropy approach focuses on the log deviations. By construction, $M_t\psi_t^* \in \mathcal{M}$ (or $\mathcal{M}^+$ if M is nonnegative), that is, once again the relative entropy minimization identifies an admissible SDF in the Hansen and Jagannathan (1997) sense. To illustrate the link between the second HJ distance and the d_{M1} distance above, we follow the cumulant expansion approach of Backus, Chernov, and Zin (2011). Recall that the cumulant generating function (i.e., the log of the moment generating function) of a random variable $\ln x_t$ is $k^x(s) = \ln\mathbb{E}\left[e^{s\ln x_t}\right]$, and, with appropriate regularity conditions, it admits the power series expansion

$$k^x(s) = \sum_{j=1}^{\infty} \kappa_j^x \frac{s^j}{j!},$$

where the jth cumulant, κ_j^x, is the jth derivative of $k^x(s)$ evaluated at $s = 0$. That is, κ_j^x captures the jth moment of the variable $\ln x_t$, i.e., κ_1^x reflects the mean of the variable, κ_2^x the variance, κ_3^x the skewness, κ_4^x the kurtosis, and so on.[10]

Using the cumulant expansion, the d_{M1} distance above can be rewritten as

$$d_{M1} = \frac{\kappa_2^{\psi^*}}{2!} + \frac{\kappa_3^{\psi^*}}{3!} + \frac{\kappa_4^{\psi^*}}{4!} + ... \tag{15}$$

where $\kappa_j^{\psi^*}$ denotes the jth cumulant of (log) ψ^*, and ψ^* solves

$$\arg\min_{\{\psi_t\}_{t=1}^T} \left(\frac{\kappa_2^{\psi}}{2!} + \frac{\kappa_3^{\psi}}{3!} + \frac{\kappa_4^{\psi}}{4!} + ... \right) \quad \text{s.t. } \mathbf{0} = \mathbb{E}^{\Psi}\left[m(\theta, t)\,\mathbf{R}_t^e\right]. \tag{16}$$

The above implies that the ψ^* component identified by our $M1$ (and $\Psi1$) bound has a very similar interpretation to the second HJ distance: it provides the minimum—in the entropy sense—multiplicative (or log linear) adjustment that would make $m(\theta,t)\,\psi_t^*$ an admissible SDF. The key difference between the second HJ bound and our $M1$ bound is that the former focuses only on the minimum second moment deviation, i.e., on the variance of $q_t - M_t$, while our bound takes into consideration not only the second moment (captured by the κ_2^{ψ} cumulant in Equation (15)), but also all other moments (captured by the $\kappa_{j>2}^{\psi}$ cumulants) of the log deviation $\log q_t - \log M_t \equiv \log \psi_t$. This implies that if skewness, kurtosis, tail probabilities, etc. are relevant for asset pricing, our approach would be more likely to capture these higher moments more effectively than the least squares one. Moreover, note that the cumulant generating function cannot be a finite-order polynomial of degree greater than two (see Theorem 7.3.5 of Lukacs (1970)). That is, if the mean and variance are not sufficient statistics for the distribution of the true SDF, then *all* the other higher moments become relevant for characterizing the SDF, and their relevance for asset pricing is captured by our entropy approach given the one to one mapping between relative entropy and cumulants. In Table A1 of Appendix A.4, we compute the minimum adjustment to the CCAPM SDF required to make it an admissible pricing kernel using both of the above

[10]For instance, if $\ln x_t \sim N\left(\mu_x; \sigma_x^2\right)$, we have $\kappa_1^x = \mu_x$, $\kappa_2^x = \sigma_x^2$, $\kappa_{j>2}^x = 0$.

approaches. The results show that, for a wide variety of test assets, the HJ adjustment leads to an SDF that has a close to Gaussian distribution. The relative entropy adjustment, on the other hand, results in an SDF having substantial skewness and kurtosis.

The cumulant decomposition also allows us to assess the relevance of higher moments for pricing asset returns. In particular, with the estimated $\{\ln \psi_t^*\}_{t=1}^T$ at hand, we can estimate its moments using sample analogs, use these moments to compute the cumulants, and finally compute the contribution of the jth cumulant to the total entropy of ψ^* as

$$\frac{\kappa_j^{\psi^*}/j!}{\sum_{s=2}^{\infty} \kappa_s^{\psi^*}/s!} \equiv \frac{\kappa_j^{\psi^*}/j!}{D\left(P||\Psi^*\right)} \tag{17}$$

as well as the total contribution of cumulants of order larger than j as

$$\frac{\sum_{s=j+1}^{\infty} \kappa_s^{\psi^*}/s!}{\sum_{s=2}^{\infty} \kappa_s^{\psi^*}/s!} \equiv \frac{D\left(P||\Psi^*\right) - \sum_{s=2}^{j} \kappa_s^{\psi^*}/s!}{D\left(P||\Psi^*\right)}. \tag{18}$$

These statistics are important for comparing the informativeness of our bounds to that of the second HJ distance since, if the minimum variance deviation had all the relevant information for pricing asset returns, we would expect

$$\frac{D\left(P||\Psi^*\right) - \kappa_2^{\psi^*}/2!}{D\left(P||\Psi^*\right)} \cong 0 \text{ and } \frac{\kappa_j^{\psi^*}/j!}{D\left(P||\Psi^*\right)} \cong 0 \qquad \forall j > 2.$$

As we will show in the empirical section below, this is not the case.

III An Illustrative Example: the C-CAPM with Power Utility

We first illustrate our methodology for the Consumption-CAPM (C-CAPM) of Breeden (1979), Lucas (1978) and Rubinstein (1976), when the utility function is time and state separable with a constant coefficient of relative risk aversion. For this specification of preferences, the SDF takes the form

$$M_{t+1} = \delta \left(C_{t+1}/C_t\right)^{-\gamma}, \tag{19}$$

where δ denotes the subjective time discount factor, γ is the coefficient of relative risk aversion, and C_{t+1}/C_t denotes the real per capita aggregate consumption growth. Empirically, the above pricing kernel fails to explain i) the historically observed levels of returns, giving rise to the Equity Premium and Risk Free Rate Puzzles (e.g., Mehra and Prescott (1985), Weil (1989)), and ii) the cross-sectional dispersion of returns between different classes of financial assets (e.g., Mankiw and Shapiro (1986), Breeden, Gibbons, and Litzenberger (1989), Campbell (1996), Cochrane (1996)).

Parker and Julliard (2005) argue that the covariance between contemporaneous consumption growth and asset returns understates the true consumption risk of the stock market if consumption is slow to respond to return innovations. They propose measuring the risk of an asset by its ultimate risk to consumption, defined as the covariance of its return and consumption growth over the period of the return and many following periods. They show that while the ultimate consumption risk would correctly measure the risk of an asset if the C-CAPM were true, it may be a better measure of the true risk if consumption responds with a lag to changes in wealth. The ultimate consumption risk model implies the following SDF:

$$M_{t+1}^{S} = \delta^{1+S} \left(C_{t+1+S}/C_{t}\right)^{-\gamma} R_{t+1,t+1+S}^{f}, \tag{20}$$

where S denotes the number of periods over which the consumption risk is measured and $R_{t+1,t+1+S}^{f}$ is the risk free rate between periods $t+1$ and $t+1+S$. Note that the standard C-CAPM obtains when $S = 0$. Parker and Julliard (2005) show that, empirically, the specification of the SDF in Equation (20), unlike the one in Equation (19), explains a large fraction of the variation in expected returns across assets for low levels of the risk aversion coefficient.

The functional forms of the above two SDFs fit into our framework in Equation (2). For the contemporaneous consumption risk model, $\theta = \gamma$, $m(\theta, t) = (C_t/C_{t-1})^{-\gamma}$, and $\psi_t^m = \delta$, a constant, for all t. For the ultimate consumption risk model, $\theta = \gamma$, $m(\theta, t) = (C_{t+S}/C_{t-1})^{-\gamma}$, and $\psi_t^m = \delta^{1+S} R_{t,t+S}^{f}$. Therefore, for each model, we construct entropy bounds for the SDF and its components using quarterly data[11] on per capita real personal consumption expenditures on nondurable goods and returns on the 25 Fama–French portfolios for 1947:Q1–2009:Q4 and compare them with the HJ bound.[12] We also obtain the non-parametrically extracted (called "filtered" hereafter) SDF and its components for

[11] See Appendix A.3 for a thorough data description.

[12] We use the 25 Fama–French portfolios as test assets because they have been used extensively in the literature to test the C-CAPM and also constituted the set of base assets in Parker and Julliard (2005).

$\gamma = 10$. For the ultimate consumption risk model, we set $S = 11$ quarters because the fit of the model is the greatest at this value, as shown in Parker and Julliard (2005).

Figure 1, Panel A plots the relative entropy (or KLIC) of the filtered and model-implied SDFs and their ψ components as functions of the risk aversion coefficient γ and the $Q1$, $M1$, and $\Psi1$ bounds for the contemporaneous consumption risk model in Equation (19). The black curve with circles shows the relative entropy of the model-implied SDF as a function of the risk aversion coefficient. For this model, the missing component of the SDF, ψ_t, is a constant, hence it has zero relative entropy for all values of γ, as shown by the grey horizontal line with triangles. The grey dashed curve and the dark grey dotted curve show, respectively, the relative entropy (as a function of the risk aversion coefficient) of the filtered SDF and its missing component. The model satisfies the HJ bound (not reported in the figure) for high values of $\gamma \geqslant 64$. It satisfies the $Q1$ bound for even higher values of $\gamma \geqslant 72$, as shown by the intersection of the black curve with circles (i.e. the entropy of the model implied SDF) and the black horizontal dashed-dotted line (the risk neutral bound). The minimum value of γ at which the $M1$ bound is satisfied is given by the value corresponding to the intersection of the black curve with circles and the light grey dashed curve, i.e., it is the minimum value of γ for which the relative entropy of the model-implied SDF exceeds that of the filtered SDF. The figure shows that this corresponds to $\gamma = 107$. Finally, the $\Psi1$ bound identifies the minimum value of γ for which the missing component of the model-implied SDF has a higher relative entropy than the missing component of the filtered SDF. Since the former has zero relative entropy while the latter has a strictly positive value for all values of γ, the model fails to satisfy the $\Psi1$ bound for any value of γ.

Panel B shows that very similar results are obtained for the $Q2$, $M2$, and $\Psi2$ bounds. The $Q2$ and $M2$ bounds are satisfied for values of γ at least as large as 73 and 99, respectively, while the $\Psi2$ bound is not satisfied for any value of γ. Overall, as suggested by the theoretical predictions, the Q bounds are tighter than the HJ bound, the M bounds are tighter than the Q bounds, and the Ψ bounds are tighter than the M bounds.

We also construct confidence bands for the above relative entropy bounds using $1,000$ bootstrapped samples. The 95% confidence bands for the $Q1$- and $Q2$ bounds extend over the intervals $[70.0, 109.0]$ and $[69.5, 109.0]$, respectively, and those for the $M1$- and $M2$ bounds cover the intervals $[94.5, 157.5]$ and $[86.0, 150.0]$, respectively. Finally, the $\Psi1$- and

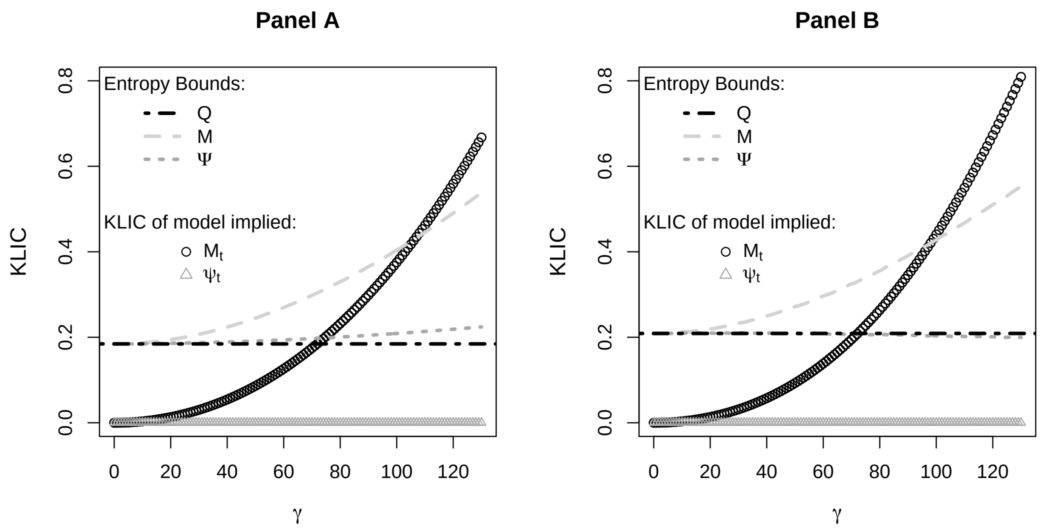


Figure 1: The KLIC of the model SDF, $M_t = \delta\,(C_t/C_{t-1})^{-\gamma}$, and the model ψ (equal to zero in this case), as well as the Q, M and Ψ bounds as functions of the risk aversion coefficient. The Q (M) bound is satisfied when the KLIC of M_t is above it, while the Ψ bound is satisfied when the KLIC of ψ_t is above it. Panels A and B show, respectively, results when ψ_t^* is estimated using Equations (6) and (4), quarterly data 1947:Q1–2009:Q4, and the 25 Fama–French portfolios.

$\Psi 2$ bounds are not satisfied for any finite value of the risk aversion coefficient in any of the bootstrapped samples. The bootstrap results reveal two points. First, it demonstrates the robustness of our approach: the two different definitions of relative entropy produce very similar results. Second, the confidence bands are quite tight in contrast with the large values of the standard error typically obtained when using GMM type approaches to estimate the risk aversion parameter.

Figure 2 presents analogous results to Figure 1 for the ultimate consumption risk model in Equation (20). Panel A shows that the $Q1$, and $M1$ bounds are satisfied for $\gamma \geqslant 23$, and 46, respectively (the HJ bound, not reported, is satisfied for values above 22). These are almost three times, more than three times, and more than two times smaller, respectively, than the corresponding values in Figure 1, Panel A, for the contemporaneous consumption risk model. As for the latter model, the $\Psi 1$ bound is not satisfied for any value of γ. Panel B shows that the $Q2$ and $M2$ bounds are satisfied for $\gamma \geqslant 24$ and 47, respectively, while the $\Psi 2$ bound is not satisfied for any value of γ. The bootstrapped 95% confidence bands for the $Q1$ and $Q2$ bounds extend over the intervals $[23.0, 35.0]$ and $[24.0, 37.0]$, respectively, and those for the $M1$ and $M2$ bounds cover the intervals $[36.0, 60.0]$ and $[40.0, 74.0]$, respectively. Also, similar to the contemporaneous consumption risk model, the $\Psi1$ and $\Psi2$ bounds are not satisfied for any finite value of the risk aversion coefficient in any of the bootstrapped samples.

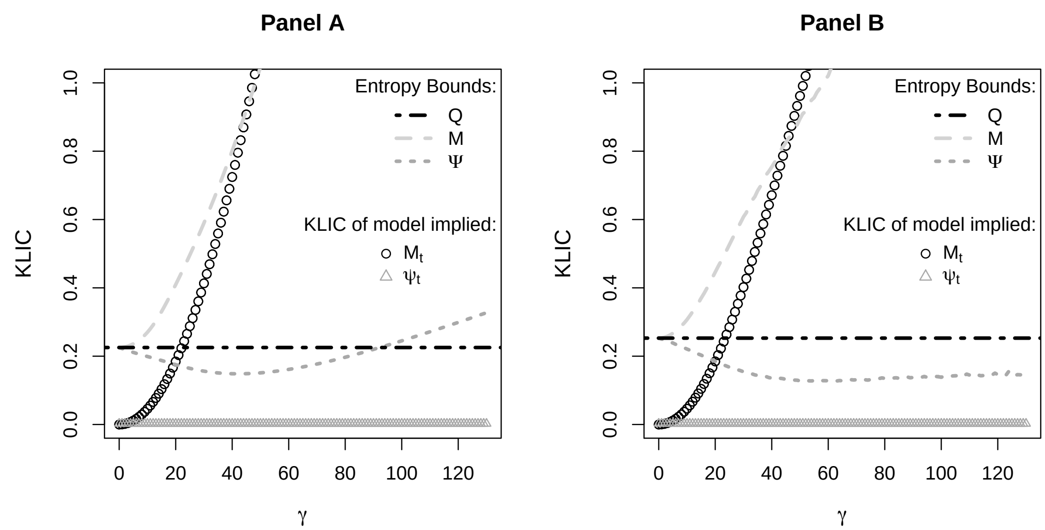


Figure 2: KLIC of the model SDF, $M_t = \delta^{1+S}(C_{t+S}/C_{t-1})^{-\gamma} R^f_{t,t+S}$, and the model ψ (equal to zero in this case), as well as the Q, M and Ψ bounds as functions of the risk aversion coefficient. The Q (M) bound is satisfied when the KLIC of M_t is above it, while the Ψ bound is satisfied when the KLIC of ψ_t is above it. Panels A and B show, respectively, results when ψ_t^* is estimated using Equations (6) and (4), quarterly data 1947:Q1–2009:Q4, and the 25 Fama–French portfolios.

It is important to note that even though the best fitting level for the RRA coefficient for the ultimate consumption risk model is smaller than 10 ($\hat{\gamma} = 1.5$), and at this value of the coefficient the model is able to explain about 60% of the cross-sectional variation in returns across the 25 Fama–French portfolios, all the bounds reject the model for low RRA, and the Ψ bounds are not satisfied for any level of RRA. This stresses the power of the proposed approach.

The above results indicate that our entropy bounds are not only theoretically, but also empirically, tighter than the HJ variance bounds. Using the cumulants decomposition introduced in the previous section, we can identify the information content added by taking into account higher moments of the SDF and its components. In particular, the statistics in Equations (17) (black dashed-dotted line) and (18) (grey dashed line) are plotted in the left panels of Figure 3 (for $S = 0$) and Figure 4 (for $S = 11$).

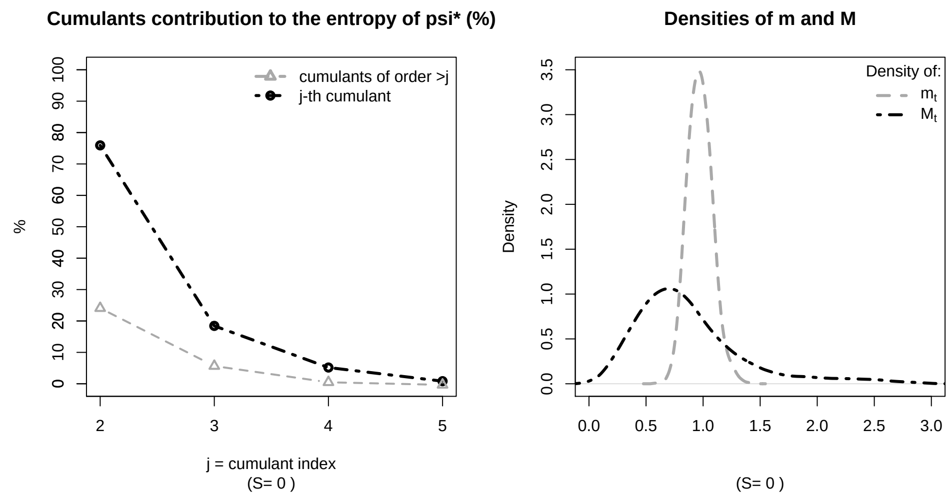


Figure 3: Left panel: Relative contribution of the cumulants of ψ_t^* to $D(P||\Psi^*)$. Right panel: Densities of $m_t := (C_t/C_{t-1})^{-\gamma}$ and $M_t^* := (C_t/C_{t-1})^{-\gamma} \psi_t^*$. ψ_t^* is estimated using Equation (6), quarterly data 1947:Q1–2009:Q4, and the 25 Fama–French portfolios, with $\gamma = 10$.

The figures show that the contribution of the second moment to $D(P||\Psi^*)$ is large—being in the 74%–78%—but that higher moments also play a very important role, with their cumulated contribution being in the 22%–26% range. Among these higher moments, the lion's share goes to the skewness, with its individual contribution being about 18% for both $S = 0$ and $S = 11$.

The relevance of skewness is also outlined in the right panels of Figure 3 (for $S = 0$) and Figure 4 (for $S = 11$) where the (Epanechnikov kernel estimates of the) densities of $m_t := \left(\frac{C_{t+S}}{C_{t-1}}\right)^{-10} R_{t,t+S}^f$ and $M_t^* := \left(\frac{C_{t+S}}{C_{t-1}}\right)^{-10} R_{t,t+S}^f \psi_t^*$ are presented. These figures show that besides the increase in variance generated by ψ^*, there is also a substantial increase in the skewness of our estimated most likely pricing kernel. This point is also shown in Figures 5 (for $S = 0$) and 6 (for $S = 11$), where the left panels present the cumulant decomposition of the entropy of $m_t := \left(\frac{C_{t+s}}{C_{t-1}}\right)^{-10} R_{t,t+S}^f$ while the right panel presents the cumulant decomposition for $M_t^* := m_t \psi_t^*$. The figures show that the sources of entropy of our filtered most likely pricing kernel ($m_t \psi_t^*$) are very different than those of the consumption growth component alone (m_t): almost all (99%) the entropy of m_t is generated by its second moment, while higher cumulants have basically no role; instead, about a quarter (24%–25%) of the entropy of $m_t \psi_t^*$ is generated by the third and higher cumulants.

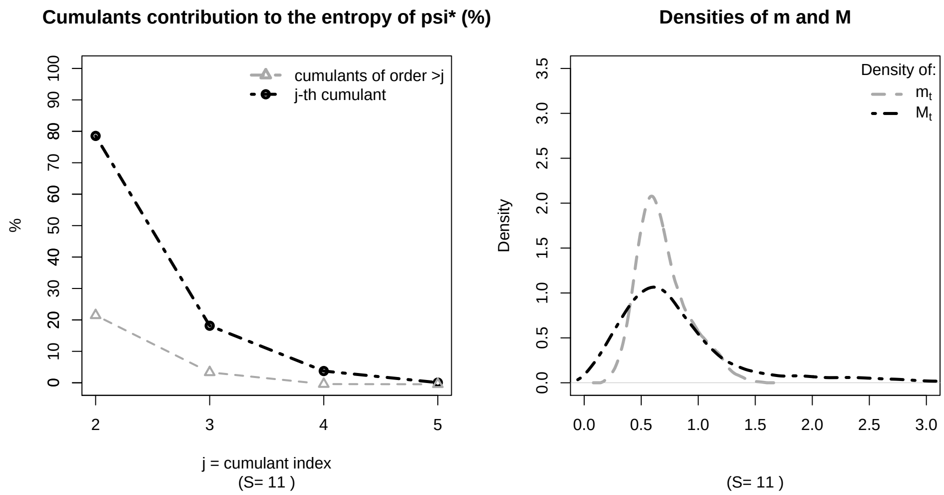


Figure 4: Left panel: Relative contribution of the cumulants of ψ_t^* to $D(P||\Psi^*)$. Right panel: Densities of $m_t := (C_{t+S}/C_{t-1})^{-\gamma} R^f_{t,t+S}$ and $M_t^* := (C_{t+S}/C_{t-1})^{-\gamma} R^f_{t,t+S}\psi_t^*$. ψ_t^* is estimated using Equation (6), quarterly data 1947:Q1–2009:Q4, and the 25 Fama–French portfolios, with $S = 11$ and $\gamma = 10$.

We now turn to the analysis of the time series properties of the candidate SDFs considered. Figure 7, Panel A plots the time series of the filtered SDF and its components estimated using Equation (6) for $\gamma = 10$ for the contemporaneous consumption risk model ($S = 0$). The black dashed line plots the component of the SDF that is a parametric function of consumption growth, $m(\theta, t) = (C_t/C_{t-1})^{-\gamma}$. The dotted line with circles plots the filtered unobservable component of the SDF, ψ_t^*, estimated using Equation (6). The black solid line plots the filtered SDF, $M_t^* = (C_t/C_{t-1})^{-\gamma}\psi_t^*$. The grey shaded areas represent NBER-dated recessions while dark grey dashed-dotted vertical lines correspond to the major stock market crashes identified in Mishkin and White (2002).[13] The figure reveals two main points. First, the estimated SDF has a clear business cycle pattern, but also shows significant and sharp reactions to financial market crashes that do not result in economy-wide contractions. Second, the time series of the SDF almost coincides with that of the unobservable component. In fact, the correlation between the two time series is .996. The observable consumption growth component of the SDF, on the other hand, has a correlation

[13] Mishkin and White (2002) identify a stock market crash as a period in which either the Dow Jones Industrial, the S&P500, or the NASDAQ index drops by at least 20% in a time window of either one day, five days, one month, three months, or one year. Consequently, in yearly figures, we classify a given year as having a stock market crash if any such event was recorded in that year. Similarly, in quarterly figures, we identify a given quarter as being a crash period if either a crash was registered in that quarter or if the entire year (containing the quarter) was identified by Mishkin and White as a stock market crash year.

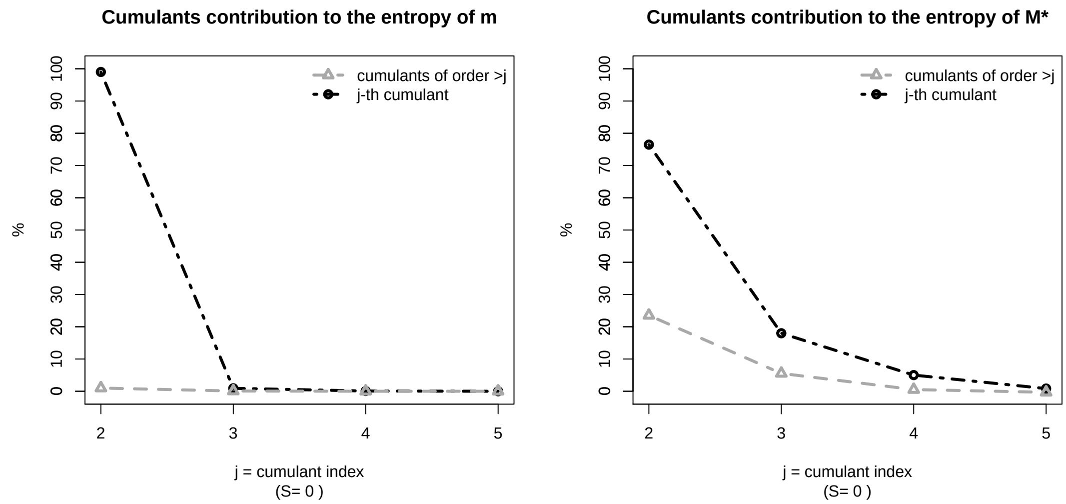


Figure 5: Left panel: Contribution of the cumulants of $(C_t/C_{t-1})^{-\gamma}$ to $D\left(P||\,(C_t/C_{t-1})^{-\gamma}\right)$. Right panel: Contribution of the cumulants of $(C_t/C_{t-1})^{-\gamma}\,\psi_t^*$ to $D\left(P||\,(C_t/C_{t-1})^{-\gamma}\,\psi_t^*\right)$. ψ_t^* is estimated using Equation (6), quarterly data 1947:Q1–2009:Q4, and the 25 Fama–French portfolios, with $\gamma = 10$.

of only .06 with the SDF. Therefore, most of the variation in the SDF comes from variation in the unobservable component, ψ, and not from the consumption growth component. In fact, the volatility of the SDF and its unobservable component are very similar, with the latter accounting for about 99% of the volatility of the former, while the volatility of the consumption growth component accounts for only about 1% of the volatility of the filtered SDF. Similar results are shown in Panel B, which plots the time series of the filtered SDF and its components estimated using Equation (4) for $\gamma = 10$.

Finally, Figure 8, Panel A plots the time series of the filtered SDF and its components estimated using Equation (6) for $\gamma = 10$ for the ultimate consumption risk model ($S = 11$). The figure shows that, as in the contemporaneous consumption risk model, the estimated SDF has a clear business cycle pattern, but also shows significant and sharp reactions to financial market crashes that do not result in economy-wide contractions. However, unlike the latter model, the time series of the consumption growth component is much more volatile and more highly correlated with the SDF. The volatility of the consumption growth component is 21.7%, more than 2.5 times higher than that for the standard model. The correlation between the filtered SDF and its consumption growth component is .37, an order of magnitude bigger than the correlation of .06 in the contemporaneous consumption risk model. This explains the ability of the model to account for a much larger fraction of the

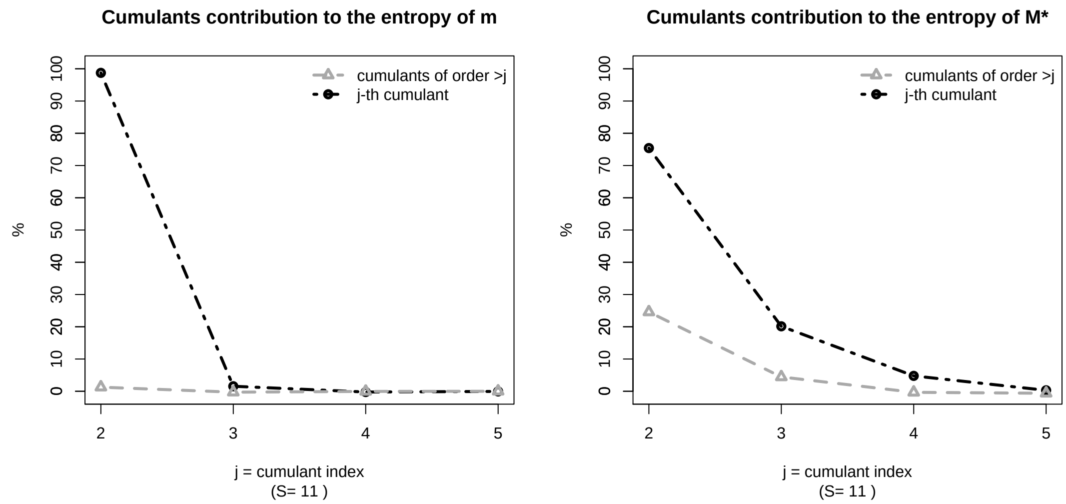


Figure 6: Left panel: Contribution of the cumulants of $(C_{t+S}/C_{t-1})^{-\gamma} R^f_{t,t+S}$ to $D\left(P||\left(C_{t+S}/C_{t-1}\right)^{-\gamma} R^f_{t,t+S}\right)$. Right panel: Contribution of the cumulants of $(C_{t+S}/C_{t-1})^{-\gamma} R^f_{t,t+S}\psi^*_t$ to $D\left(P||\left(C_{t+S}/C_{t-1}\right)^{-\gamma} R^f_{t,t+S}\psi^*_t\right)$. ψ^*_t is estimated using Equation (6), quarterly data 1947:Q1–2009:Q4, and the 25 Fama–French portfolios, with $S = 11$ and $\gamma = 10$.

variation in expected returns across the 25 Fama–French portfolios for low levels of the risk aversion coefficient. In fact, the cross-sectional R^2 of the model is 54.1% (for $\gamma = 10$), an order of magnitude higher than the value of 5.2% for the standard model. However, the correlation between the ultimate consumption risk SDF and its unobservable component is still very high at .92, showing that the model is missing important elements that would further improve its ability to explain the cross-section of returns. Similar results are shown in Panel B, which plots the time series of the filtered SDF and its components estimated using Equation (4) for $\gamma = 10$.

Overall, the results show that our methodology provides useful diagnostics for dynamic asset pricing models. Moreover, the very similar results obtained using the two different types of relative entropy minimization in Equations (4) and (6) suggest the robustness of our approach.

IV Application to More General Models of Dynamic Economies

Our methodology provides useful diagnostics to assess the empirical plausibility of a large class of consumption-based asset pricing models where the SDF, M_t, can be factorized into an observable component consisting of a parametric function of consumption, C_t, as

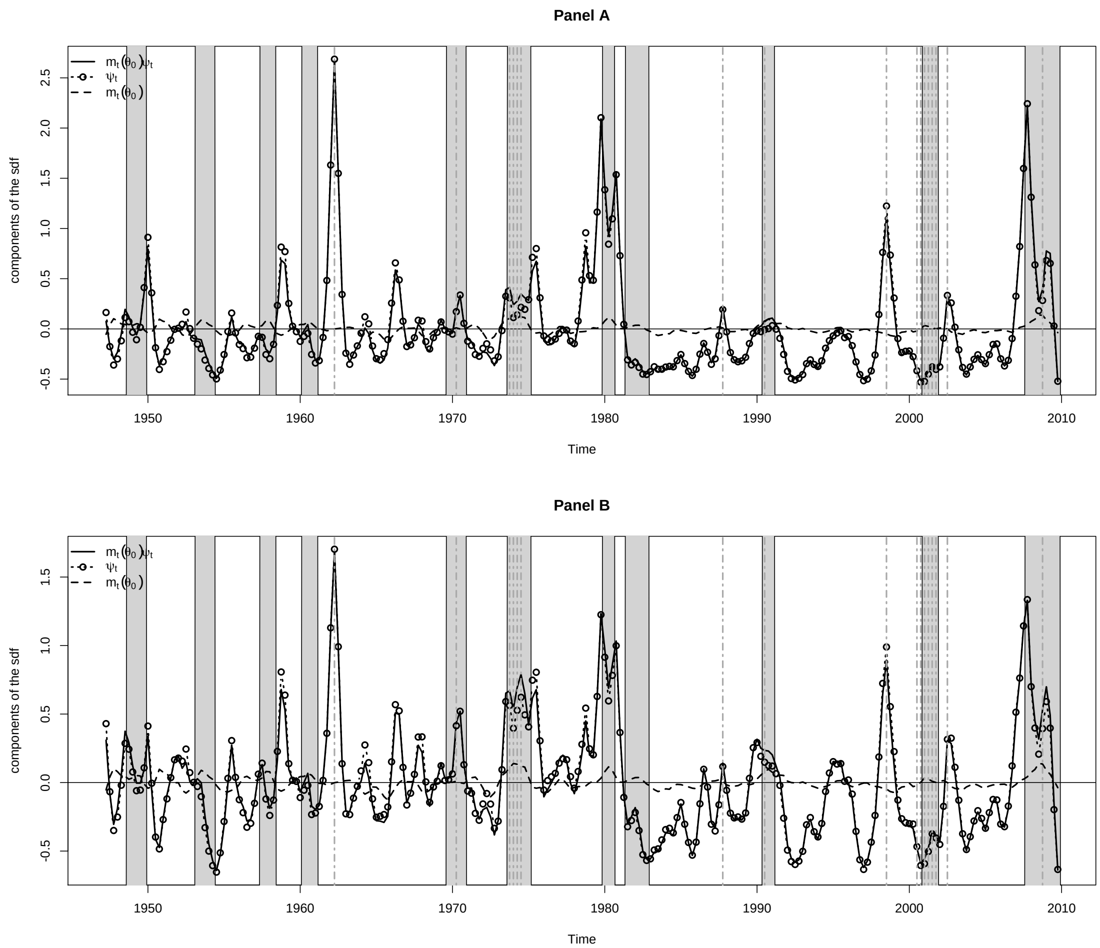


Figure 7: The (de-meaned) time series of the filtered SDF, $M_t^* = m(\theta; t)\,\psi_t^*$, and its components for the standard CCAPM for γ =10. Panels A and B show, respectively, results when ψ_t^* is estimated using Equations (6) and (4), quarterly data 1947:Q1–2009:Q4, and the 25 Fama–French portfolios. Shaded areas are NBER recession periods. Vertical dot-dashed lines are the stock market crashes identified by Mishkin and White (2002).

in the standard time-separable power utility model, and a potentially unobservable one, ψ_t, that is model-specific. In this section, we apply it to a set of "winner" asset pricing models, i.e., frameworks that can successfully explain the Equity Premium and the Risk Free Rate Puzzles with "reasonable" calibrations. In particular, we consider the external habit formation models of Campbell and Cochrane (1999) and Menzly, Santos, and Veronesi (2004), the long-run risks model of Bansal and Yaron (2004), and the housing model of Piazzesi, Schneider, and Tuzel (2007). We apply our methodology to assess the empirical plausibility of these models in two ways. First, since our methodology identifies the most likely time series of the SDF, we compare this time-series with the model-implied time series

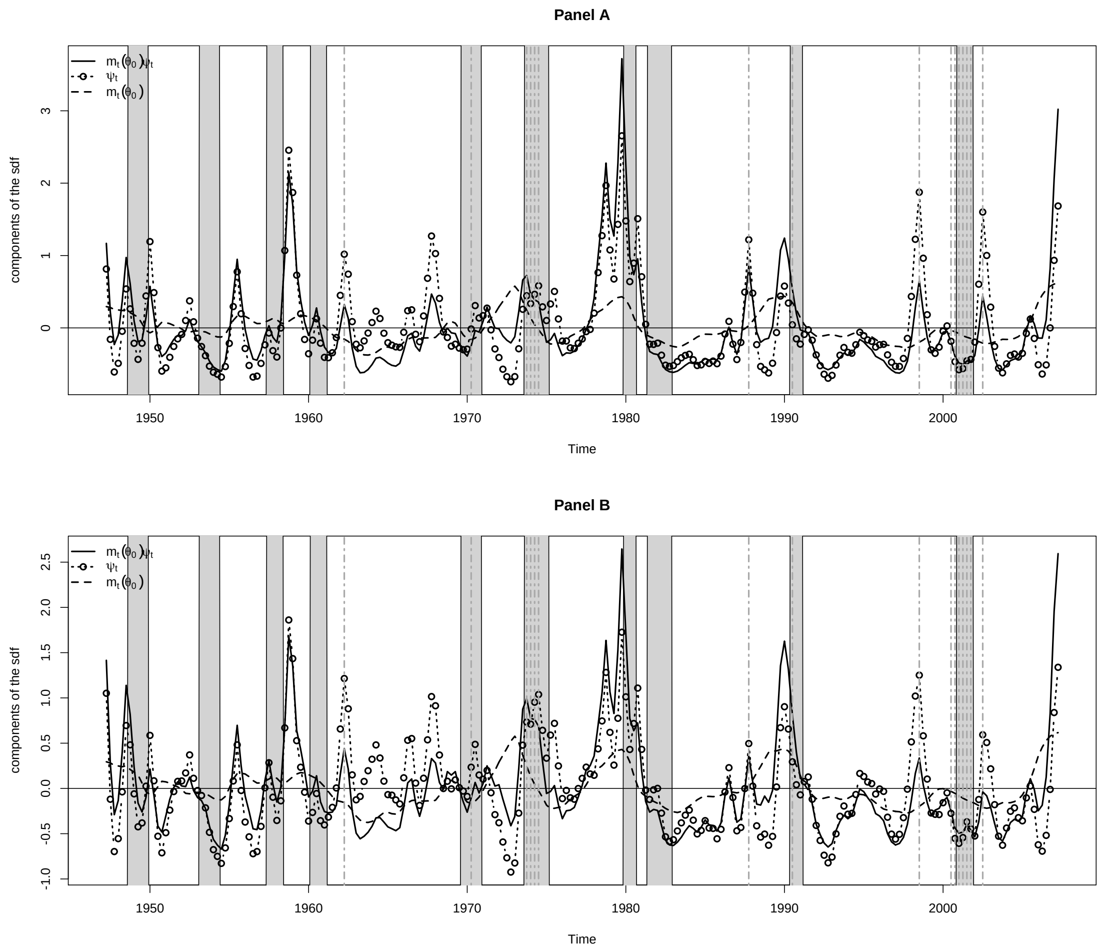


Figure 8: The (demeaned) time series of the filtered SDF, $M_t^* = m(\theta; t)\,\psi_t^*$, and its components for the ultimate consumption risk CCAPM of Parker and Julliard (2005) for γ =10. Panels A and B show, respectively, results when ψ_t^* is estimated using Equations (6) and (4), quarterly data 1947:Q1–2009:Q4, and the 25 Fama–French portfolios. Shaded areas are NBER recession periods. Vertical dot-dashed lines are the stock market crashes identified by Mishkin and White (2002).

of the SDF for each model. Second, for each model we compute the values of the power coefficient, γ, at which the model-implied SDF satisfies the HJ, Q, M, and Ψ bounds.

In the next sub-section we present the models considered. The reader familiar with these models can go directly to Section IV.2, which presents the empirical results, without loss of continuity. A detailed description of the data is presented in Appendix A.3.

IV.1 The Models Considered

IV.1.1 External Habit Formation Model: Campbell and Cochrane (1999)

In this model, identical agents maximize power utility defined over the difference between consumption and a slow-moving habit or time-varying subsistence level. The SDF is given by

$$M_t^m = \underbrace{(C_t/C_{t-1})^{-\gamma}}_{m(\theta,t)}\underbrace{\delta\,(S_t/S_{t-1})^{-\gamma}}_{\psi_t^m}, \tag{21}$$

where δ is the subjective time discount factor, γ is a curvature parameter that provides a lower bound on the time varying coefficient of relative risk aversion, $S_t = \frac{C_t - X_t}{C_t}$ denotes the surplus consumption ratio, and X_t is the habit component. Note that the ψ^m component depends on the surplus consumption ratio, S, that is not directly observed. To obtain the time series of ψ^m, we extract the surplus consumption ratio from the observed data using two different procedures.

First, we extract the time series of the surplus consumption ratio from the consumption data. In this model, the aggregate consumption growth is assumed to follow an i.i.d. process:

$$\Delta c_t = g + \upsilon_t, \quad \upsilon_t \sim \text{i.i.d.} N\left(0, \sigma^2\right).$$

The log surplus consumption ratio evolves as a heteroskedastic AR(1) process:

$$s_t = (1-\phi)\,\overline{s} + \phi s_{t-1} + \lambda\,(s_{t-1})\,\upsilon_t, \tag{22}$$

where $s_t := \ln S_t$ and $\overline{s}$ is the steady state log surplus consumption ratio, and

$$\lambda\,(s_t) = \begin{cases} \frac{1}{\overline{S}}\sqrt{1-2\,(s_t-\overline{s})}-1, & \text{if } s_t \leq s_{max} \\ 0, & \text{if } s_t > s_{max} \end{cases},$$

$$s_{max} = \overline{s} + \frac{1}{2}\left(1-\overline{S}^2\right), \qquad \overline{S} = \sigma\sqrt{\frac{\gamma}{1-\phi}}.$$

For each value of γ, we use the calibrated values of the model preference parameters (δ, ϕ) in Campbell and Cochrane (1999), the sample mean (g) and volatility (σ) of the consumption growth process, and the innovations in real consumption growth, $\widehat{\upsilon}_t = \Delta c_t - g$, to extract the time series of the surplus consumption ratio using Equation (22) and, thereby, obtain the time series of the model-implied SDF and its ψ^m component.

Second, in this model, the equilibrium market-wide price–dividend ratio is a function of the surplus consumption ratio alone, although the form of the function is not available in closed form. Using numerical methods, we invert this function to extract the time series of the surplus consumption ratio from the historical time series of the price–dividend ratio and, thereby, obtain the time series of the model-implied SDF and its ψ^m component from Equation (21).

IV.1.2 External Habit Formation Model: Menzly, Santos, and Veronesi (2004)

In this model, the SDF is analogous to that of Campbell and Cochrane (1999) disussed above. The aggregate consumption growth is also assumed to follow an i.i.d. process:

$$dc_t = \mu_c dt + \sigma_c dB_t,$$

where μ_c is the mean consumption growth, $\sigma_c > 0$ is a scalar, and B_t is a Brownian motion. The point of departure from the Campbell and Cochrane (1999) framework is that Menzly, Santos, and Veronesi (2004) assume that the *inverse* surplus consumption ratio, $Y_t := \frac{1}{S_t}$, follows a mean reverting process that is perfectly negatively correlated with innovations in consumption growth:

$$dY_t = k\left(\overline{Y} - Y_t\right)dt - \alpha\left(Y_t - \lambda\right)\left[dc_t - E\left(dc_t\right)\right], \tag{23}$$

where $\overline{Y}$ is the long run mean of the inverse surplus consumption ratio and k controls the speed of the mean reversion. To obtain the time series of ψ^m (the model implied ψ component), we extract the surplus consumption ratio from the observed data using two different procedures.

First, for each value of γ,[14] we use the calibrated values of the model parameters $\left(\delta,\, k,\, \overline{Y},\, \alpha,\, \lambda\right)$ in Menzly, Santos, and Veronesi (2004), the sample values of μ_c and σ_c, and the innovations in real consumption growth, $\widehat{dB_t} = \frac{[dc_t - E(dc_t)]}{\sigma_c}$, to extract the time series of the surplus consumption ratio, and this allows us to compute the time series of the model-implied SDF.

Second, in this model, the equilibrium price–consumption ratio of the total wealth portfolio is a function of the surplus consumption ratio alone. However, this function is not available in closed form except when $\gamma = 1$. Therefore, we rely on log-linear approximations

[14] Note that the Menzly, Santos, and Veronesi (2004) model assumes that the representative agent has log utility, i.e., γ is set equal to 1, in order to derive a closed form solution for the price–consumption ratio. For other values of γ, the model does not admit a closed form solution. Nevertheless, the pricing kernel is well defined even if γ is different from one, hence we will be considering this more general case.

to the return on the total wealth portfolio to express the equilibrium log price–consumption ratio as an affine function of the log surplus consumption ratio for all values of γ. Details of this procedure are described in Appendix A.5. We, then, invert this affine function to extract the time series of the surplus consumption ratio from the historical time series of the market-wide price–dividend ratio and, thereby, obtain the time series of the model-implied SDF and its ψ^m component from Equation (21). Note that approximating the total wealth price–consumption ratio by the market-wide price–dividend ratio is the approach used by Menzly, Santos, and Veronesi (2004).

IV.1.3 Long-Run Risks Model: Bansal and Yaron (2004)

The Bansal and Yaron (2004) long-run risks model assumes that the representative consumer has the version of Kreps and Porteus (1978) preferences developed by Epstein and Zin (1989) and Weil (1989) for which the SDF is given by

$$M_{t+1}^m = \delta^\theta \left(\frac{C_{t+1}}{C_t}\right)^{-\frac{\theta}{\rho}} R_{c,t+1}^{\theta-1},$$

where $R_{c,t+1}$ is the unobservable gross return on an asset that delivers aggregate consumption as its dividend each period, δ is the subjective time discount factor, ρ is the elasticity of intertemporal substitution, $\theta := \frac{1-\gamma}{1-1/\rho}$, and γ is the relative risk aversion coefficient.

The aggregate consumption and dividend growth rates, Δc_{t+1} and Δd_{t+1}, respectively, are modeled as containing a small persistent expected growth rate component, x_t, that follows an AR(1) process with stochastic volatility, and fluctuating variance, σ_t^2, that evolves according to a homoscedastic linear mean reverting process.

Appendix A.6 shows that, for the log-linearized model, the log of the SDF and its ψ^m component are given by

$$\ln M_{t+1}^m = \underbrace{c_2 \Delta c_{t+1}}_{\ln m(\theta,t+1)} + \underbrace{c_1 + c_3 x_{t+1} + c_4 \sigma_{t+1}^2 + c_5 x_t + c_6 \sigma_t^2}_{\ln \psi_{t+1}^m} \tag{24}$$

where the parameters $(c_1, c_2, c_3, c_4, c_5, c_6)$ are known functions of the underlying time series and preference parameters of the model.

To obtain the time series of the SDF and ψ^m, we extract the state variables, x_t and σ_t^2, from the observed data using two different procedures. First, we extract them from the consumption data. Second, we extract them from the asset market data, in particular, from the market-wide price–dividend ratio and the risk free rate. The extraction of the state variables using these two procedures is described in Appendix A.6. Finally, for each value of γ, we use the calibrated parameter values from Bansal and Yaron (2004) and the time series of the state variables to obtain the time series of the SDF and its ψ^m component from Equation (24).

IV.1.4 Housing: Piazzesi, Schneider, and Tuzel (2007)

In this model, the pricing kernel is given by

$$M_t^m = \delta \left(C_t/C_{t-1}\right)^{-\gamma} \left(A_t/A_{t-1}\right)^{\frac{\gamma\rho-1}{\rho-1}},$$

where A_t is the expenditure share on non-housing consumption, γ^{-1} is the intertemporal elasticity of substitution, and ρ is the intratemporal elasticity of substitution between housing services and non-housing consumption.

Taking logs, we obtain

$$\ln M_t^m = \underbrace{-\gamma \Delta c_t}_{\ln m(\theta,t)} + \underbrace{\ln \delta + \frac{\gamma\rho - 1}{\rho - 1}\Delta a_t}_{\ln \psi_t^m}. \tag{25}$$

Note that, in this model, ψ^m depends only on the observable variables and therefore does not need to be extracted from consumption or asset market data. For each value of γ, we use the calibrated values of the model parameters (δ, ρ) in Piazzesi, Schneider, and Tuzel (2007) to obtain the time series of the model-implied SDF and its ψ^m component from Equation (25).

IV.2 Empirical Results

For our empirical analysis, we focus on two data samples: an annual data sample starting at the onset of the Great Depression (1929–2009), and a quarterly data sample starting in the post World War II period (1947:Q1–2009:Q4). A detailed description of the data is presented in Appendix A.3. Note that, in any finite sample, the extracted time series of the SDF, as well as the information bounds on the SDF and its unobservable component, depend on the set of test assets used for their construction. Since the Euler equation holds for any traded asset as well as any adapted portfolio of assets, this gives infinitely many moment restrictions. Nevertheless, econometric considerations necessitate the choice of only a subset of assets to be used. As a consequence, in our empirical analysis, we compute the bounds, and filter the time series of the SDF and its components, using a broad cross-section of test assets. In particular, at the quarterly frequency, the test assets include the 6 size and book-to-market-equity sorted portfolios of Fama–French, 10 industry-sorted portfolios, and 10 momentum-sorted portfolios. Due to the smaller available time series at the annual frequency, we restrict the cross-section of test assets to include the 6 size and book-to-market-equity sorted portfolios, 5 industry-sorted portfolios, and the first and last deciles of the 10 momentum-sorted portfolios.

IV.2.1 The Time Series of the Most Likely SDF

Our first approach to assessing the empirical plausibility of these models is based on the observation that our methodology identifies the most likely time-series of the SDF, which we call the *filtered SDF*. That is, given a candidate SDF with observable component $m(\theta, t)$, we use the relative entropy minimizing procedures in Equations (4) and (6) to estimate a time series for the unobservable (or residual, if the SDF is fully observable) component $\{\psi_t^*(\theta)\}_{t=1}^T$, and obtain the filtered SDF as $m(\theta, t)\,\psi_t^*$.

Note that the filtered SDF and its missing component depend on the local curvature of the utility function γ, since changing γ modifies the constraints in Equations (4) and (6). Therefore, for each model, we fix γ at the authors' calibrated value, and extract the time series of the filtered SDF and its components. We compare the filtered SDF $(m(\theta, t)\,\psi_t^*)$ with the model-implied SDF $(m(\theta, t)\,\psi_t^m)$ for each model.

Table I presents the results at quarterly frequency. Panel A presents the results when the model-implied SDF and its components are obtained by extracting the state variable(s) from the consumption data. Panel B presents the results when extracting the state variable(s) from the asset market data. The first column presents the correlation between the filtered time series $\{\ln \psi_t^*\}_{t=1}^T$ of the missing component of the SDF and the corresponding model-implied time series, $\{\ln \psi_t^m\}_{t=1}^T$. The second column shows the correlation between the filtered SDF, $\{\ln M_t^* = \ln(m(\theta, t)\,\psi_t^*)\}_{t=1}^T$, where $m(\theta, t) = (C_t/C_{t-1})^{-\gamma}$, and the model-implied SDF, $\{\ln M_t^m = \ln(m(\theta, t)\,\psi_t^m)\}_{t=1}^T$. The 95% confidence intervals for these correlations, reported in square brackets, are obtained by bootstrapping with replacement from the data.

Consider first the results for the CC external habit model that are presented in the first row of each panel. For this model, the utility curvature parameter is set to the calibrated value of $\gamma = 2$. Panel A, column 1 shows that when the model-implied state variable is extracted from the consumption data, the correlation between the filtered and model-implied ψ is only .10 when ψ^* is estimated using Equation (6). Column 2 shows that the correlation between the filtered and model-implied SDFs is marginally higher at .13. When ψ^* is estimated using Equation (4), the correlations are very similar: .07 and .09. Panel B shows that the correlations between the filtered and model-implied SDFs and ψs remain small when the model state variable is extracted from the market-wide price–dividend ratio.

Table I: Correlation of Filtered and Model SDFs, 1947:Q1–2009:Q4

	Correlation of filtered and model SDF		Cross-sectional R^2	
	$\rho\left(\ln\psi_t^*, \ln\psi_t^m\right)$	$\rho\left(\ln M_t^*, \ln M_t^m\right)$	no intercept	free intercept
Panel A: State Variables Extracted From Consumption Data				
CC	.10 / .07 [−.09,.18] [−.11,.18]	.13 / .09 [−.07,.20] [−.09,.19]	−1.19 [−3.14,.02]	.002 [.00,.38]
MSV	−.01 / .003 [−.07,.18] [−.09,.18]	.05 / .04 [−.07,.20] [−.09,.19]	−.79 [−2.72,.06]	.002 [.000,.37]
BY	−.02 / .03 [−.14,.12] [−.11,.18]	.16 / .09 [−.03,.25] [−.12,.18]	−.71 [−2.83,.02]	.005 [.00,.35]
PST	−.12 / −.14 [−.24,.02] [−.24,.03]	−.03 / −.04 [−.16,.09] [−.21,.09]	−.91 [−3.21,.14]	.03 [.00,.36]
Panel B: State Variables Extracted From Asset Prices				
CC	.17 / .16 [−.10,.18] [−.10,.18]	.18 / .17 [−.10,.18] [−.10,.19]	−.77 [−3.13,.08]	.31 [.00,.39]
MSV	.18 / .23 [−.10,.19] [−.10,.22]	.19 / .24 [−.10,.20] [−.10,.22]	−.46 [−3.78,.00]	.04 [.00,.48]
BY	.03 / .06 [−.11,.17] [−.11,.21]	.04 / .07 [−.11,.17] [−.10,.21]	−1.26 [−3.23,−.39]	.24 [.00,.52]

Correlation between the filtered and the model-implied ψ-components of the SDFs (column 1), the correlation between the filtered and the model-implied SDFs (column 2), the cross-sectional R^2 implied by the model-specific SDFs when no intercept is allowed in the cross-sectional regression (column 3), and the cross-sectional R^2 when an intercept is allowed in the regression (column 4), using quarterly data 1947:Q1–2009:Q4. The bootstrapped 95% confidence intervals are given in square brackets below. Each cell in columns 1 and 2 has two entries, indicating whether the filtered ψ^*-component and, therefore, the filtered SDF is estimated using Equation (6), shown on the left, or Equation (4), shown on the right. Panel A presents results when the models' state variables and, therefore, the model-implied SDFs, are extracted from consumption data. Panel B presents the results when the state variables are extracted from asset prices. The acronyms CC, MSV, BY and PST, denote, respectively, the models of Campbell and Cochrane (1999), Menzly, Santos, and Veronesi (2004), Bansal and Yaron (2004) and Piazzesi, Schneider, and Tuzel (2007).

The second row in each panel presents the results for the MSV external habit model. In this case, γ is set equal to 1, which is the calibrated value in the model. Row 2 in each panel shows that the results for the MSV model are similar to those for the CC model. When ψ^* is estimated using Equation (6), the correlations between the filtered and model-implied ψ components of the SDFs are small, varying from $-.01$ when the surplus consumption ratio is extracted from the consumption data, to .18 when the state variable is extracted using the price–dividend ratio. The correlations between the filtered and model-implied SDFs are marginally higher, varying from .05 when the surplus consumption ratio is extracted from the consumption data, to .19 when it is extracted using the price–dividend ratio. Similar results are obtained when ψ^* is estimated using Equation (4).

Table II: Correlation of Filtered and Model SDFs, 1929–2009

	Correlation of filtered and model SDF		Cross-sectional R^2	
	$\rho\left(\ln\psi_t^*, \ln\psi_t^m\right)$	$\rho\left(\ln M_t^*, \ln M_t^m\right)$	no intercept	free intercept
Panel A: State Variables Extracted From Consumption Data				
CC	.35 / .31	.39 / .34	.082	.504
	[−.04,.44] [−.04,.41]	[−.00,.48] [−.02,.92]	[−2.19,.74]	[.00,.81]
MSV	.33 / .22	.41 / .34	.76	.82
	[−.02,.41] [−.04,.37]	[.06,.46] [−.02,.96]	[−2.05,.76]	[.00,.80]
BY	−.17 / −.028	.27 / .20	.45	.47
	[−.31,.22] [−.44,.21]	[−.03,.48] [−.16,.77]	[−2.25,.75]	[.00,.80]
PST	−.09 / −.001	−.004 / −.013	−.73	.09
	[−.23,.24] [−.25,.25]	[−.20,.21] [−.26,.26]	[−2.39,.09]	[.00,.40]
Panel B: State Variables Extracted From Asset Prices				
CC	.19 / .14	.24 / .17	−.20	.60
	[−.12,.35] [−.10,.28]	[−.11,.37] [−.08,.29]	[−2.86,.53]	[.00,.63]
MSV	−.04 / .13	.01 / .18	−.16	.001
	[−.10,.33] [−.10,.27]	[−.08,.35] [−.09,.28]	[−2.69,.27]	[.00,.52]
BY	−.01 / .10	−.02 / .09	−.15	.005
	[−.21,.34] [−.23,.31]	[−.21,.29] [−.29,.32]	[−.77,.25]	[.00,.27]

Correlation between the filtered and the model-implied ψ-components of the SDFs (column 1), the correlation between the filtered and the model-implied SDFs (column 2), the cross-sectional R^2 implied by the model-specific SDFs when no intercept is allowed in the cross-sectional regression (column 3), and the cross-sectional R^2 when an intercept is allowed in the regression (column 4), using annual data 1929–2009. The bootstrapped 95% confidence intervals are shown in square brackets below. Each cell in columns 1 and 2 has two entries, indicating whether the filtered ψ^*-component and, therefore, the filtered SDF is estimated using Equation (6), shown on the left, or Equation (4), shown on the right. Panel A presents the results when the models' state variables and, therefore, the model-implied SDFs, are extracted from consumption data. Panel B presents the same when the state variables are extracted from asset prices. Other notation as in Table I.

The third row in each panel presents the results for the BY long run risks model. The parameter γ is set equal to the BY calibrated value of 10. Row 3, Panel A, column 1 shows that when the state variables are extracted from the consumption data, the correlation between the filtered and model-implied ψ components is $-.02$ (.03) when ψ^* is estimated using Equation (6) (Equation (4)). Column 2 shows that the correlation between the filtered and model-implied SDFs is .16 (.09). Similar results are shown in Panel B, where the state variables are extracted from the market-wide price–dividend ratio.

Table II presents results analogous to those in Table I, but at an annual frequency. The results are largely similar to those in Table I. Notable exceptions are the two habit models when the state variable is extracted from the consumption data. In this case the correlations between filtered and model implied SDFs and ψ components are much higher than at the quarterly frequency, being in the .31–.39 range for CC and .22–.41 for MSV.

The last two columns of Tables I and II show the cross-sectional R^2s, along with 95% confidence bands, in square brackets below, implied by the model-specific SDFs. The cross-sectional R^2 is obtained by performing a cross-sectional regression of the historical average returns on the model-implied expected returns. Column 3 presents the cross-sectional R^2 when there is no intercept in the regression while column 4 presents the results when an intercept is included. The results reveal that the cross-sectional R^2s vary wildly for the same model, and often have large negative values when an intercept is not allowed in the cross-sectional regression, or when the model-implied state variables are extracted using either the consumption or the asset market data. Moreover, they have very wide confidence intervals. As we show in the next sub-section, this is in stark contrast with the results based on entropy bounds in Tables VI and VII, that tend instead to give consistent results and tighter confidence bands for each model across different samples and procedures used to extract the model state variables.

Overall, Tables I and II make two main points. First, they demonstrate the robustness of our estimation methodology—very similar results are obtained using either Equation (6) or (4) to filter $\psi^*$and M^*. Second, they show that, regardless of the data frequency and the procedure used to extract the model-implied SDFs, all the asset pricing models considered yield SDFs that tend to have low correlations with the filtered SDF—the *most likely* SDF given the data. While the results in Tables I and II are obtained using the combined set of size and book-to-market-equity sorted, momentum-sorted, and industry-sorted portfolios, very similar results are obtained using the 25 Fama–French portfolios as test assets.[15]

[15]The results are available from the authors upon request.

The correlations between the model specific SDFs and the filtered SDFs discussed above would have little significance if the filtered discount factors had no clear economic interpretation. In order to address this concern, we show below that our filtered pricing kernel has clear economic content since *a*) it is always highly correlated with the Fama–French factors (that can be interpreted as proxies for the true unknown sources of systematic risk), *b*) it implies that the most likely SDF should have a strong business cycle pattern, and *c*) it reacts significantly to financial market crashes.

Tables III and IV show the correlations between the filtered and model-implied log SDFs and the three Fama–French (FF) factors at quarterly and annual frequencies, respectively. Column 1 presents the correlation between the model-implied SDF when the state variables are extracted from the consumption data, and the three FF factors. This is computed by performing a linear regression of the model-implied time series of the SDF, $\{\ln(M_t^m)\}_{t=1}^T$, on the three FF factors and computing the correlation between $\ln(M^m)$ and the fitted value from the regression. Column 2 presents the correlation when the model-implied state variables are extracted from asset market data. Columns 3 and 4 present, respectively, the correlations of the filtered SDF and its missing component with the three FF factors.

Consider first Table III. Panel A, column 3 shows that the log of the filtered SDF, $M_t^* \equiv m(\theta, t)\, \psi_t^*$, correlates strongly with the FF factors, having correlation coefficients ranging from .49 to .59 when the set of test assets consists of the 25 size and book-to-market-equity sorted portfolios of Fama–French. Column 4 reveals that this high correlation is due almost entirely to the ψ^* component, and not $m(\theta, t)$, since the correlation between the filtered SDF and the FF factors is the same as that between the filtered missing component of the SDF and the FF factors.

The above results are perhaps not surprising because the FF factors are known to be quite successful at explaining a large fraction of the cross-sectional variation in returns of the 25 size and book-to-market-equity sorted portfolios. However, Panels B and C show that the filtered SDF correlates strongly with the FF factors independently of the set of test assets used to extract the filtered SDF. When the set of test assets consists of the 10 momentum-sorted portfolios, the correlations vary from .51 to .55. For the 10 industry-sorted portfolios, the correlations vary from .53 to .69. Column 4 of Panels B and C reveals that this high correlation is almost entirely driven by the missing component of the SDF

Table III: Correlations with FF3, 1947:Q1–2009:Q4

	$(\ln M_t^m)_{cons}$	$(\ln M_t^m)_{prices}$	$\ln M_t^*$	$\ln \psi_t^*$
Panel A: 25 Fama–French				
CC	.18	.20	.54/.59	.54/.59
MSV	.21	.95	.54/.59	.54/.59
BY	.25	.45	.54/.58	.52/.57
PST	.07	–	.49/.52	.45/.50
Panel B: 10 Momentum				
CC	.18	.20	.52/.52	.51/.51
MSV	.21	.95	.52/.52	.51/.51
BY	.25	.45	.55/.53	.50/.50
PST	.07	-	.53/.51	.43/.43
Panel C: 10 Industry				
CC	.18	.20	.65/.69	.64/.68
MSV	.21	.95	.65/.69	.65/.68
BY	.25	.45	.66/.69	.62/.65
PST	.07	-	.53/.55	.47/.51

Correlations between the 3 Fama–French factors and *i)* the model-implied SDF with state variables extracted from consumption (column 1) and stock market (column 2) data, *ii)* the filtered SDF (column 3), and *iii)* the filtered ψ^* component of the SDF (column 4), using quarterly data 1947:Q1–2009:Q4 and a different set of portfolios in each panel. Each cell in columns 3 and 4 has two entries, indicating whether the filtered ψ^*-component and, therefore, the filtered SDF is estimated using Equation (6), shown on the left, or Equation (4), shown on the right. Other notation as in Table I.

and not the consumption growth component.

Row 1, column 1 of each panel shows that, for the CC model, while the filtered SDF correlates strongly with the FF factors, the model-implied SDF has a small correlation coefficient of .18 when the surplus consumption ratio is extracted from the consumption data. Row 1, column 2 shows that the correlation rises only marginally to .20 when the state variable is extracted from the market-wide price–dividend ratio.

For the MSV model, the correlation between the model-implied SDF and the FF factors is small, at .21, when the surplus consumption ratio is extracted from the consumption data. However, when the state variable is extracted from the price–dividend ratio, the correlation between the model-implied SDF and the FF factors is very high: .95—much higher than the correlation between the filtered SDF and the FF factors for each set of test assets.

Row 3 in each panel shows that for the BY model, the correlation between the model-implied SDF and the FF factors is .25 when the state variables are extracted from the consumption data. The correlation increases to .45 when the asset price data are used in the extraction of the model-implied state variables.

Table IV: Correlations with FF3, 1929–2009

	$(\ln M_t^m)_{cons}$	$(\ln M_t^m)_{prices}$	$\ln M_t^*$	$\ln \psi_t^*$
Panel A: 6 Fama–French				
CC	.19	.12	.73/.78	.72/.77
MSV	.26	.87	.73/.78	.72/.77
BY	.38	.73	.77/.77	.68/.72
PST	.35	–	.81/.76	.65/.67
Panel B: 10 Momentum				
CC	.19	.12	.55/.63	.58/.61
MSV	.26	.87	.55/.62	.57/.61
BY	.38	.73	.69/.69	.51/.57
PST	.35	-	.73/.70	.50/.55
Panel C: 10 Industry				
CC	.19	.12	.49/.53	.49/.53
MSV	.26	.87	.50/.54	.50/.55
BY	.38	.73	.42/.39	.38/.42
PST	.35	-	.41/.27	.34/.37

Correlations between the 3 Fama–French factors and *i)* the model-implied SDF with state variables extracted from consumption (column 1) and stock market (column 2) data, *ii)* the filtered SDF (column 3), and *iii)* the filtered ψ^* component of the SDF (column 4), using annual data 1929–2009 and a different set of portfolios in each panel. Each cell in columns 3 and 4 has two entries, indicating whether the filtered ψ^*-component and, therefore, the filtered SDF is estimated using Equation (6), shown on the left, or Equation (4), shown on the right. Other notation as in Table I.

Finally, row 4 in each panel shows that for the PST model, the correlation between the model-implied SDF and the FF factors is very small: .07.

Table IV shows that very similar results are obtained at an annual frequency. Tables III and IV demonstrate the soundness of our estimation methodology: the filtered time series of the SDF and its ψ^* component are quite robust, in terms of their correlations with the FF factors, to the choice of the utility curvature parameter γ, the set of assets, and the data frequency considered. Moreover, our filtered SDF and ψ^* are consistently highly correlated with the FF factors independently of the sample frequency and the cross-section of assets used for the estimation (even assets, such as the industry and momentum portfolios, that are not well priced by the FF factors). This finding has several important implications. First, it suggests that our estimation approach successfully identifies the unobserved pricing kernel, since there is substantial empirical evidence that the FF factors do proxy for asset risk sources. Second, our finding provides a rationalization of the empirical success of the FF factors in pricing asset returns. Finally, although the filtered SDF is highly correlated with the FF factors, the correlation coefficient is substantially smaller than unity, particularly for the industry and momentum portfolios (see, e.g., Table IV), suggesting that the FF factors cannot fully capture all the underlying sources of systematic risk that are important in pricing these assets.

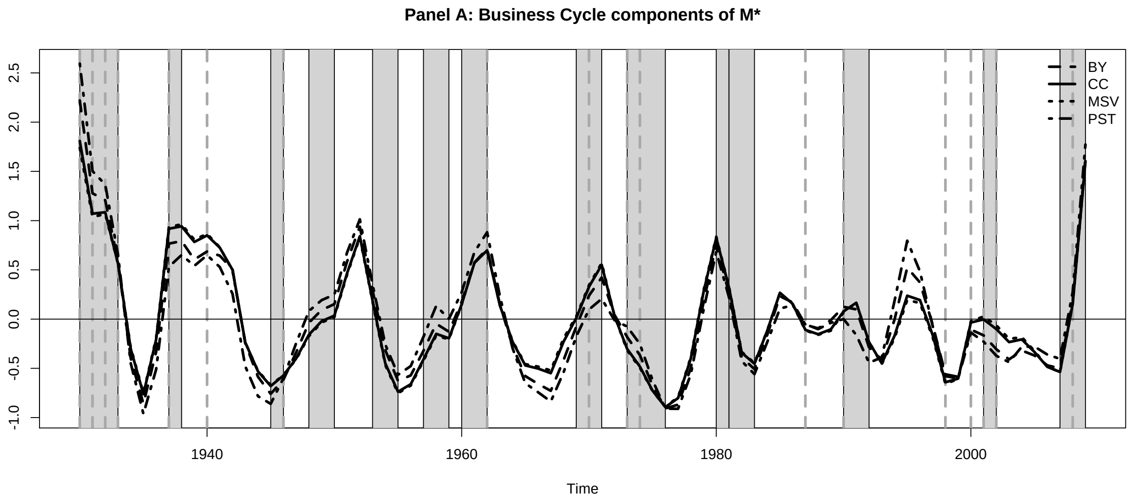
The reason behind the stable correlation results between our filtered SDFs and the three Fama French factors seems to be the fact that, independently of the set of assets used for the filtering, the most likely SDF tends to have a very similar time series behavior. In particular, it shows a clear business cycle pattern, and significant and sharp reactions to stock market crashes (even if these crashes do not necessarily result in economy-wide contractions). This feature of the filtered SDFs is illustrated in Figure 9 (annual frequency) and Figure 10 (quarterly frequency). In each figure we show the business cycle component (Panel A) and the residual component (Panel B) of the filtered M^* for the different models.[16] At both data frequencies, independently of the model considered, both the business cycle and residual components are extremely similar across the models.

In Table V we compare the business cycle and market crash properties of the filtered SDFs with the model-implied ones. For each model considered, and for both the filtered (M^*) and model-implied (M^m) pricing kernels, the table presents the risk neutral probabilities of recessions (column 1), and stock market crashes non-concomitant with recessions (column 2) as well as, in the first row of each panel, the sample frequency of these events.[17] For the model-implied pricing kernels, we present the probabilities when the state variables are extracted using the consumption data as well as using the asset price data (in brackets below).

Focusing on quarterly data (Panel A), column 1 shows that the filtered SDFs (M^*) imply a risk neutral probability of a recession in the 25%–26% range. Comparing this with the model-implied probabilities shows that whether the state variables are extracted using the consumption or the asset market data, all the model-implied pricing kernels deliver a similar risk neutral probability of recessions, one that is similar to that of our filtered SDFs

[16]The decomposition into a business cycle and a residual component is obtained by applying the Hodrick and Prescott (1997) filter to the estimated M^*.

[17]To compute the risk neutral probabilities, note that for any quantity A_t and function $f(.)$, we have that $\mathbb{E}^Q[f(A_t)] = \int f(A_t)\frac{dQ}{dP}dP = \int f(A_t)\frac{M_t}{M}dP$. Hence, given an SDF M_t (either filtered or model-implied) the risk neutral expectation can be estimated (assuming ergodicity) using the sample analog $\widehat{\mathbb{E}^Q[f(A_t)]} = \frac{1}{T}\sum_t^T f(A_t)\frac{M_t}{M}$. For instance, to estimate the probability of a recession, we replace $f(A_t)$ with an index function that takes the value 1 if the economy was in an NBER-designated recession at time t and zero otherwise. See also Remark 1 in Appendix A.1.



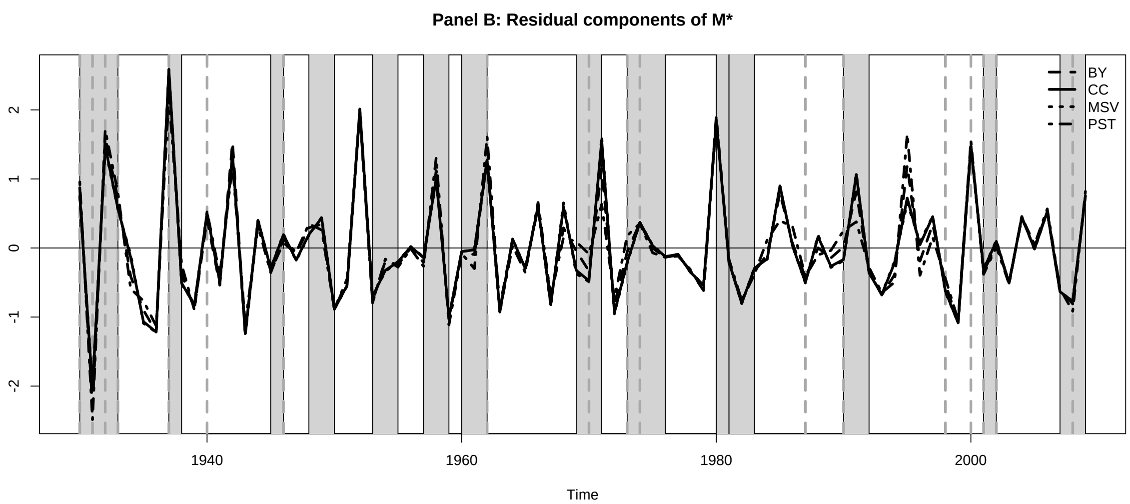


Figure 9: Business cycle (Panel A) and residual (Panel B) components of the most likely (log) SDF ($M_t^* = (C_t/C_{t-1})^{-\gamma}\,\psi_t^*$) filtered using Equation (6), annual data 1929–2009, 6 size and book-to-market-equity portfolios, 10 industry portfolios, and the first and last decile of 10 momentum portfolios, for the different models considered: Bansal and Yaron (2004) (BY), Campbell and Cochrane (1999) (CC), Menzly, Santos, and Veronesi (2004) (MSV), and Piazzesi, Schneider, and Tuzel (2007) (PST). The difference between the models is driven by the value of the utility curvature parameter γ that is set to the authors' original calibrations. Decomposition into business cycle and residual component obtained applying the Hodrick and Prescott (1997) filter to M^*. Shaded areas denote NBER recession years, and vertical dashed lines indicate the major stock market crashes identified by Mishkin and White (2002).

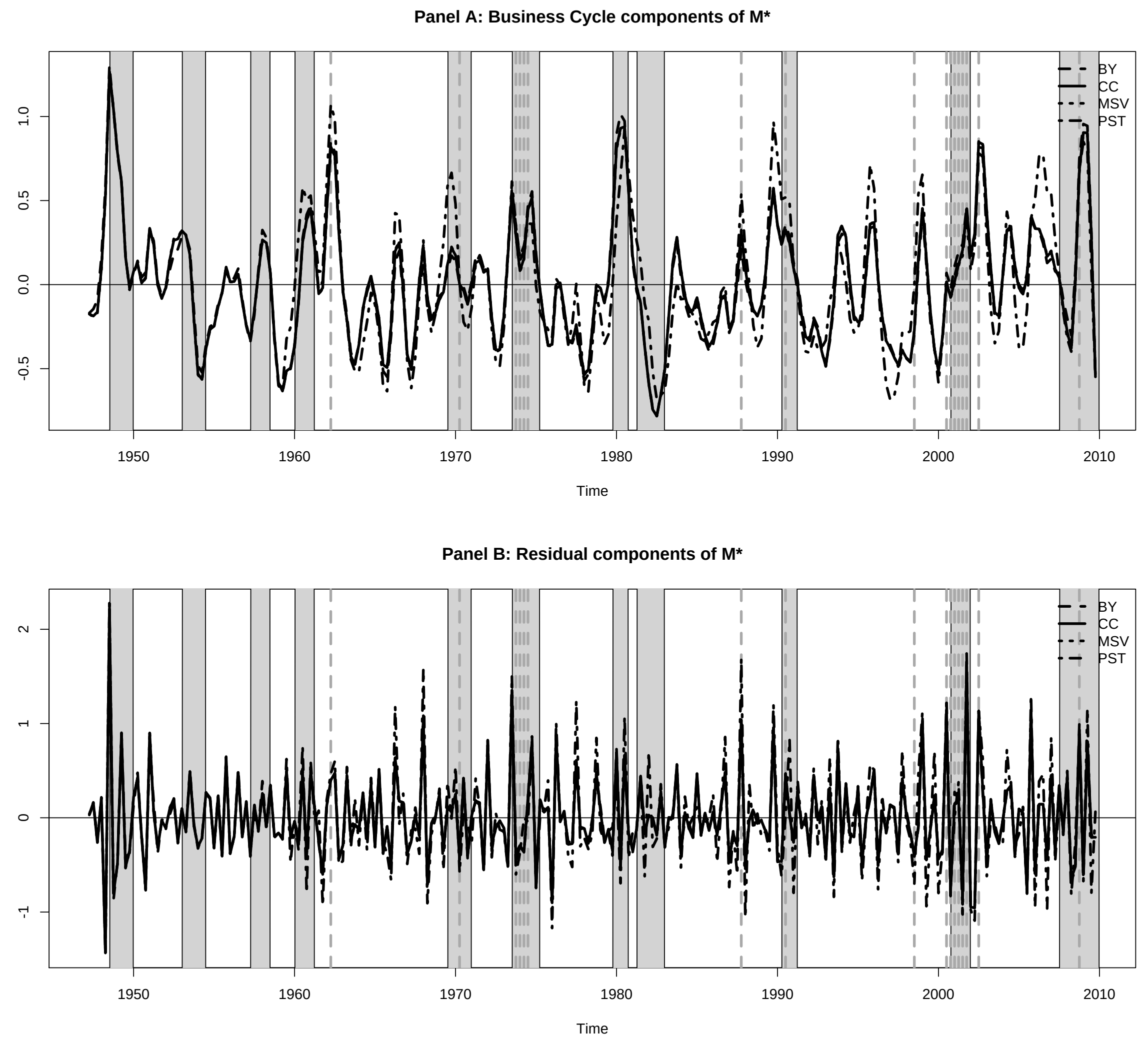


Figure 10: Business cycle (Panel A) and residual (Panel B) components of the most likely (log) SDF ($M^*_t = (C_t/C_{t-1})^{-\gamma}\,\psi^*_t$) filtered using Equation (6), quarterly data 1947:Q1–2009:Q4, 6 size and book-to-market-equity portfolios, 10 industry portfolios, and 10 momentum portfolios, for the same models as in Figure 9. The difference between the models is driven by the value of the utility curvature parameter γ that is set to the authors' original calibrations. Decomposition into business cycle and residual component obtained applying the Hodrick and Prescott (1997) filter to M^*. Shaded areas denote NBER recession years, and vertical dashed lines indicate the major stock market crashes identified by Mishkin and White (2002).

Table V: Recession and Market Crash Probabilities of M^m and M^*

		Recession Probability	Market Crash without Recession Probability
Panel A: Quarterly Data, 1947:Q1–2009:Q4			
Sample		.22	.024
CC	M^m	.25 [.24]	.024 [.025]
	M^*	.25/.25	.054/.059
BY	M^m	.22 [.55]	.024 [.034]
	M^*	.26/.26	.049/.057
MSV	M^m	.22 [.22]	.024 [.028]
	M^*	.25/.25	.055/.059
PST	M^m	.21	.028
	M^*	.25/.25	.086/.065
Panel B: Annual Data, 1929–2009			
Sample		.375	.088
CC	M^m	.61 [.49]	.055 [.068]
	M^*	.58/.56	.092/.119
BY	M^m	.41 [.59]	.083 [.227]
	M^*	.59/.59	.085/.097
MSV	M^m	.38 [.39]	.086 [.098]
	M^*	.57/.56	.094/.122
PST	M^m	.37	.067
	M^*	.60/.60	.103/.093

Risk-neutral probability of recessions (column 1) and stock market crashes non-concomitant with recessions (column 2) implied by the model (M^m) and filtered (M^*) SDFs at quarterly (Panel A) and annual (Panel B) frequencies. Each cell in the rows corresponding to the model SDF has two entries, indicating whether the models' state variables are extracted from the consumption data, shown on the left, or from the asset market data, shown on the right. Each cell in the rows corresponding to the filtered SDF has two entries, indicating whether the filtered ψ^*-component and, therefore, the filtered SDF is estimated using Equation (6), shown on the left, or Equation (4), shown on the right. Other notation as in Table I.

(with the notable exception of the BY pricing kernel that, extracting the state variables using asset market data, implies a risk neutral probability of recession of about 55%). More interestingly, column 2 shows that the model-implied kernels fail to show the significant and sharp reaction to stock market crashes exhibited by the filtered SDFs: the probabilities of stock market crashes non-concomitant with recessions implied by the filtered SDFs are between 104% and 207% higher than those implied by the model specific kernels when the model-implied state variables are extracted from the consumption data, and between 44% and 207% higher when the state variables are extracted from the asset price data. Panel B presents similar findings for the annual data, but also shows that MSV and PST imply too low probablities of recessions and BY—only when extracted from the asset prices—implies a very high probability of a market crash[18]

Overall, the above results suggest that the explanatory power of these models for asset pricing would be improved by augmenting the pricing kernels with a component that exhibits sharp reactions to market crashes that are not perfectly correlated with the business cycle.

IV.2.2 Entropy Bounds Analysis

Our second approach to assessing the empirical plausibility of the asset pricing models considered relies on the entropy bounds derived in Section II.1. For each model we compute the minimum values of the power coefficient, γ, at which the model-implied SDF satisfies the HJ, Q, M, and Ψ bounds. We also compute the 95% confidence bands via bootstrapping. Table VI presents the results for the quarterly data. Panels A and B show the results when the state variables needed to construct the time series of the model-implied SDF and its components are extracted from the consumption (Panel A) and the asset market data (Panel B). Consider first the results for the HJ, $Q1$, $M1$, and $\Psi1$ bounds. The first row in each panel presents the bounds for the CC model. Panel A shows that when the surplus consumption ratio is extracted from the consumption data, the minimum values of γ at which the pricing kernel satisfies the HJ, $Q1$, $M1$, and $\Psi1$ bounds are 10.2, 16.1, 16.4, and 23.2, respectively. Therefore, as suggested by the theoretical predictions, the Q bound is tighter than the HJ bound, and the M bound is tighter than the Q bound. Note that in this model, the curvature of the utility function is $\frac{\gamma}{S_t}$, where S_t is the surplus consumption ratio, and this ratio is almost identical to the coefficient of relative risk aversion (see, e.g., the discussion in Campbell and Cochrane (1999)). For $\gamma = 2$, the calibrated value in CC, the curvature varies over $[19.7, \infty)$. Panel A shows that the Q bound is satisfied for $\gamma \geqslant 16.1$, implying that the curvature varies over $[56.6, \infty)$, the M bound is satisfied for $\gamma \geqslant 16.4$, implying that the curvature varies over $[57.2, \infty)$, and the Ψ bound is satisfied for $\gamma \geqslant 23.2$, implying that the curvature varies over $[68.5, \infty)$. A similar ordering of the bounds is obtained when the surplus consumption ratio is extracted from the market-wide price–dividend ratio in Panel B except that, in this case, even higher values of risk aversion are needed in order to satisfy the bounds. Also, very similar results are obtained for the $Q2$, $M2$, and $\Psi2$ bounds, stressing the robustness of our approach.

[18]Note that, at an annual frequency, a year is designated as a recession year if at least one of its quarters is in an NBER recession period.

Table VI: Bounds for RRA, Quarterly Data, 1947:Q1–2009:Q4

	HJ Bound	$Q1/Q2$ Bounds	$M1/M2$ Bounds	$\Psi1/\Psi2$ Bounds
Panel A: State Variables Extracted From Consumption Data				
CC	10.2	16.1 / 15.7 [16.0,38.0] [14.4,34.8]	16.4 / 16.0 [16.0,38.0] [14.6,36.8]	23.2 / 23.9 [23.0,>100] [21.2,>100]
MSV	32.6	40.8 / 40.4 [38.0,62.0] [38.0,59.0]	43.4 / 43.5 [40.0,64.0] [40.0,64.0]	61.3 / 62.8 [59,113] [59.0,>100]
BY	> 100	> 100 / > 100 [>100,>100] [>100,>100]	> 100 / > 100 [>100,>100] [>100,>100]	> 100 / > 100 [>100,>100] [>100,>100]
PST	73.8	99.0 / 92.6 [96.0,172.0] [88.0,161.0]	111.1 / 102.2 [102.0,183.0] [93.0,172,1]	96.2 / 90.5 [94.0,187.0] [86.0,176.0]
Panel B: State Variables Extracted From Asset Prices				
CC	19	43 / 46 [43.0,50.0] [46.0,49.0]	46 / 46 [46.0,50.0] [46.0,49.0]	47 / 48 [47.0,51.0] [48.0,50.0]
MSV	73.3	90.3 / 90.0 [92.0,>100] [89.5,>100]	> 100 / > 100 [>100,>100] [>100,>100]	> 100 / > 100 [>100,>100] [>100,>100]
BY	4.0	5 / 5 [5.0,6.0] [5.0,6.0]	5 / 5 [5.0,6.0] [5.0,6.0]	5 / 5 [5.0,6.0] [5.0,6.0]

Minimum values of the utility curvature parameter γ at which the model-implied SDF satisfies the HJ (column 1), Q (column 2), M (column 3), and Ψ (column 4) bounds using quarterly data 1947:Q1–2009:Q4. The bootstrapped 95% confidence intervals are shown in square brackets below. Columns 2–4 have two entries in each cell, which indicate whether the filtered ψ^*-component of the SDF and, therefore, the filtered SDF are estimated using Equation (6), shown on the left, or Equation (4), shown on the right. Panels A and B present the results when the models' state variables are extracted from the consumption data and the asset market data, respectively. Other notation as in Table I.

The second row in each panel presents the bounds for the MSV model. When the surplus consumption ratio is extracted from the consumption data, the HJ, $Q1$, $M1$, and $\Psi1$ bounds are satisfied for a minimum value of $\gamma = 32.6$, 40.8, 43.4, and 61.3, respectively. Very similar results are obtained for the $Q2$, $M2$, and $\Psi2$ bounds. Therefore, this model requires much higher values of risk aversion than CC to be consistent with the observed asset returns. Note, however, that for both models and both procedures used to extract the model-implied SDFs, the risk aversion coefficients at which the models satisfy the bounds are very high.

The third row in each panel presents the bounds for the BY model. Panel A shows that when the model-implied state variables are extracted from the consumption data, the model-implied pricing kernel fails to satisfy the HJ, Q, M, and Ψ bounds for any value of the risk aversion parameter smaller than 100. On the other hand, when the model-implied state variables are extracted from the asset market data (Panel B), the HJ bound is satisfied for a minimum value of $\gamma = 4.0$ while the $Q1$, $M1$, and $\Psi1$ bounds are all satisfied by a relative risk aversion as small as 5. Similar results are obtained for the $Q2$, $M2$, and $\Psi2$ bounds. Therefore, the results show that the empirical performance of the BY framework crucially depends on how the latent state variables are extracted from the data.

Finally, the fourth row of Panel A presents the bounds for the PST model. Note that, in this model, the SDF is a function of the observable data alone, hence there is no need to extract any state variable from the asset market data. Therefore, we do not have a fourth row in Panel B. The model satisfies the HJ, $Q1\,(Q2)$, $M1\,(M2)$, and $\Psi1\,(\Psi2)$ bounds for minimum values of $\gamma = 73.8,\ 99.0\,(92.6),\ 111.1\,(102.2)$, and $96.2\,(90.5)$, respectively. Therefore, this model requires very high levels of risk aversion to be consistent with observed asset returns.

Overall, Table VI demonstrates that, in line with the theoretical underpinnings of the various bounds, the Q bound is tighter than the HJ bound because it naturally exploits the restriction that the SDF is a strictly positive random variable. The M bound is tighter than the Q bound because it formally takes into account the ability of the SDF to price assets and the dependency of the pricing kernel on consumption. Furthermore, the results suggest that all the models considered require very high levels of risk aversion to satisfy the bounds, with the only exception being the long run risks model of BY (but only when the model state variables are extracted from the asset price data).

Table VII: Bounds for RRA, Annual Data, 1929–2009

	HJ Bound	$Q1/Q2$ Bounds	$M1/M2$ Bounds	$\Psi1/\Psi2$ Bounds
Panel A: State Variables Extracted From Consumption Data				
CC	.7	5.1 / 2.7	5.2 / 2.7	7.6 / 3.6
		[4.0,41.0] [3.0,8.0]	[4.0,41.0] [3.0,8.0]	[5.0,>100] [4.0,23.2]
MSV	17	28.7 / 24.4	30.3 / 26.6	> 100 / 76.5
		[19.0,53.3] [23.7,35.0]	[20.0,53.3] [24.7,35.4]	[>100,>100] [81.0,>100]
BY	50	53 / 71	60 / > 80	55 / > 80
		[22.0,71.0] [69.7,>80]	[24.0,72.0] [>80,>80]	[49.0,>80] [2.0,>80]
PST	17.1	28.6 / 24.1	31.4 / 27.0	22.0 / 18.6
		[19.0,51.7] [23.0,35.4]	[20.0,51.3] [24.0,35.4]	[14.0,42.7] [19.7,29.0]
Panel B: State Variables Extracted From Asset Prices				
CC	4	7 / 6	7 / 6	8 / 7
		[4.0,12.0] [6.0,9.0]	[4.0,12.0] [6.0,9.0]	[4.0,14.0] [7.0,11.0]
MSV	23.7	39.1 / 33.4	42.2 / 37.0	> 100 / > 100
		[22.0,69.5] [29.5,45.0]	[26.0,69.5] [30.5,45.0]	[>100,>100] [>100,>100]
BY	5	6 / 6	6 / 6	6 / 6
		[5.0,6.0] [2.0,7.0]	[5.0,6.0] [2.0,6.0]	[5.0,6.0] [2.0,6.0]

Minimum values of the utility curvature parameter γ at which the model-implied SDF satisfies the HJ (column 1), Q (column 2), M (column 3), and Ψ (column 4) bounds using annual data 1929–2009. The bootstrapped 95% confidence intervals are shown in square brackets below. Columns 2–4 have two entries in each cell, which indicate whether the filtered ψ^*-component of the SDF and, therefore, the filtered SDF are estimated using Equation (6), shown on the left, or Equation (4), shown on the right. Panels A and B present results when the models' state variables are extracted from consumption data and asset market data, respectively. Other notation as in Table I.

Table VII presents analogous bounds to those in Table VI for the annual data. The table shows that, at this frequency, all the bounds tend to be satisfied with smaller values of the utility curvature parameter, suggesting that the models considered can more easily rationalize asset pricing dynamics at the annual, rather than quarterly, level. However, once again in line with the theoretical predictions, the Q bound is tighter than the HJ bound, and the M bound is tighter than the Q bound.

Note that the above results on the bounds have tight confidence bands and are much more consistent in evaluating the plausibility of a given model across the different procedures used to extract the model-implied SDF and its components, than the cross-sectional R^2 measures shown in Tables I and II, which vary wildly for the same model and have very wide confidence intervals.

The results in Tables VI and VII are obtained by allowing only the utility curvature parameter, γ, to vary while holding constant all the other model parameters at the authors' calibrated values. Note that most consumption based asset pricing models, including the ones considered in this paper, are highly parametrized. Since the state variables are not directly observed in many of the models, the parameters governing their dynamics are typically chosen to match some of the moments of the data. Consequently, the properties of the SDF are quite sensitive to not only γ but also the values of all the other parameters. Therefore, we also compute the minimum values of the power coefficient, γ, at which the model-implied SDFs satisfy the HJ, Q, M, and Ψ bounds while allowing the remaining model parameters to simultaneously vary over intervals two standard errors around their calibrated values. The results, shown in Table A2 of Appendix A.7.1, remain qualitatively unchanged. In particular, for each model, the HJ, Q, M, and Ψ bounds are satisfied for smaller values of γ when the other parameters are allowed to vary simultaneously compared to Tables VI and VII where the other parameters are held fixed. However, as in the latter tables, the CC, MSV, and PST models still require much larger values of risk aversion to satisfy the bounds compared to the authors' calibrated values.

Also note that we have used *excess* returns (in excess of the risk free rate) on a broad cross-section of risky assets to extract the most likely SDF and obtain entropy bounds on the SDF and its components. However, it is well known that the level of the risk free asset constrains models quite dramatically. Therefore, in order to check the robustness of our results, we repeated the empirical exercise, using as test assets the gross returns (instead of excess returns) on the same assets considered so far plus the risk free asset. The methodology needs to be slightly modified in this case and is described in Appendix A.7.2. The results, shown in Table A3 of Appendix A.7.2, show that the inclusion of the risk free rate as an additional asset leaves the HJ, Q, M, and Ψ bounds on the SDF and its components very similar to those obtained in Tables VI and VII without the risk free rate, for all the models considered.

IV.2.3 What Are The Consumption-Based Models Missing?

As shown in Section II.1.1, modelling the SDF as being fully observable, i.e., setting $m(\theta, t) = M_t^m$ where M_t^m is the *entire* pricing kernel of the model under consideration (given in Equations (21), (24) and (25)), we can extract a residual ψ^{resid} component such that $M_t^* := M_t^m \times \psi_t^{resid}$ prices assets correctly. The ψ^{resid} component can once again be estimated using the relative entropy minimization procedures in Equations (6) and (4) replacing m with M^m. The ψ^{resid} multiplicative adjustment of the pricing kernel: *a*) still has a maximum likelihood interpretation; *b*) adds the minimum amount of information needed for M^* to be able to price assets correctly; and *c*) most importantly, as the second Hansen–Jagannathan distance, it provides a useful diagnostic for detecting what the pricing kernels are missing in order to be consistent with the observed asset returns.

We first examine the relative importance of the two components of M^*, M^m and ψ^{resid} in pricing a broad cross-section of assets. We do this by computing the contribution of each component to the overall entropy of the pricing kernel. The results are shown in Table VIII. Columns 1 and 2 present the relative entropy, or KLIC, of the model-implied SDF, M_t^m, and the residual component, ψ_t^{resid}, respectively. Column 3 presents the KLIC of ψ_t^{resid} as a fraction of the KLIC of the overall filtered kernel $M_t^m \times \psi_t^{resid}$.

Each row of column 1 presents the KLIC, or relative entropy, of M_t^m. There are four numbers for this quantity since there are two possible ways of computing the KLIC (as $D(P||M^m)$, shown on the left, and $D(M^m||P)$, shown on the right), and two possible ways of extracting the models' state variables (from the consumption data, top numbers, and from the asset market data, bottom numbers in square brackets). Similarly, four numbers with the same ordering are shown in the remaining two columns. Consider first Panel A,

Table VIII: Relative Entropy of SDF and Its Components

	$KLIC\left(M_t^m\right)$	$KLIC\left(\psi_t^{resid}\right)$	$\frac{KLIC\left(\psi_t^{resid}\right)}{KLIC\left(M_t^m \psi_t^{resid}\right)}$
Panel A: Quarterly, 1947:Q1–2009:Q4			
CC	.035/.037	.26/.32	.772/.786
	[.018] [.019]	[.30] [.33]	[.909] [.859]
MSV	.0002/.0002	.31/.36	.992/.996
	[.004] [.004]	[.30] [.35]	[.950] [.953]
BY	.003/.003	.30/.35	.957/.971
	[1.69] [1.70]	[.59] [.39]	[.647] [.448]
PST	.008/.008	.39/.39	1.01/.989
Panel B: Annual, 1929–2009			
CC	.379/.660	.66/.69	.688/.676
	[.164] [.169]	[.76] [.73]	[.815] [.767]
MSV	.001/.001	.85/.85	.972/.974
	[.023] [.023]	[.85] [.81]	[.973] [.906]
BY	.023/.022	.82 / .84	.932/.959
	[2.66] [1.75]	[2.33] [1.02]	[1.44] [.712]
PST	.19/.27	.96/.91	1.06/.996

KLIC of the model-implied SDF (column 1), the KLIC of the residual psi (column 2), and the ratio of the KLIC of the residual psi and the KLIC of the product of the model-implied SDF and the residual psi (column 3) at the quarterly (Panel A) and annual (Panel B) levels. Each cell has four entries, which indicate whether the models' state variables are extracted from consumption data, shown at the top, or from asset market data, shown at the bottom, and whether the KLIC between measure A and the physical measure P is computed as $D(P||A)$, shown on the left, or as $D(A||P)$, shown on the right. Other notation as in Table I.

which presents the results obtained from the quarterly data. Columns 1 and 2 show that for the CC model, the relative entropy of ψ^{resid} is an order of magnitude bigger than that of M^m, regardless of whether ψ_t^{resid} is estimated using Equation (6) or (4), or whether M_t^m is obtained by extracting the state variable from the consumption or the asset market data. This point is further highlighted in column 3, which shows that the KLIC of ψ_t^{resid} accounts for the lion's share of the KLIC of the overall kernel: 77.2%–78.6% when the model-implied state variable is extracted from the consumption data and 85.9%–90.9% when it is extracted from the asset price data. Very similar results are obtained for the MSV, BY, and PST models in rows 2–4, and also for the annual data in Panel B. Overall, the results suggest that for each model considered, most of the ability of the kernel to price assets comes from the residual component and very little from the model-implied component, i.e., all the pricing kernels under consideration seem to miss a substantial share of the information needed to price correctly the observed asset returns.

In order to assess whether these models are missing similar features of the data, Table IX presents the correlations between the ψ^{resid} of different models at the quarterly (Panel A) and annual (Panel B) frequencies. As in the previous table, for all the entries, we have four numbers given by the two ways of computing the relative entropy (the left and right numbers corresponding to Equations (6) and (4)) and the two ways of extracting the models' state variables (from the consumption data in the top numbers and from the asset prices for the numbers below in square brackets). Panel A shows that when the models' state variables are extracted from the consumption data, the correlations between the residual ψs are extremely high, varying from .85 (between CC and PST) to (almost) 1.0 (between MSV and BY) when the ψ^{resid} component is estimated using Equation (6). When the ψ^{resid} component is estimated using Equation (4), the correlations are very similar, varying from .93 to (almost) 1.0. When the models' state variables are extracted from the asset prices, the correlations among the various ψ^{resid} are almost unchanged, with one important exception: in this case, the correlation between the residual component of the BY model and all other models becomes much smaller, ranging from .1 to .41. This implies that the BY pricing kernel changes a lot, depending on whether its state variables are extracted from the market data or the consumption data. Similar results were obtained for the annual data, shown in Panel B, although the correlations are generally smaller at this frequency.[19]

Table IX: Correlation of Residual ψs

	MSV	BY	PST
Panel A: Quarterly, 1947:Q1–2009:Q4			
CC	.96/.93 [.96] [.95]	.97/.96 [.32] [.10]	.85/.93 [.93] [.94]
MSV		1.0/1.0 [.41] [.20]	.91/.97 [.89] [.94]
BY			.91/.97 [.26] [.10]
Panel B: Annual, 1929:2009			
CC	.87/.66 [.91] [.78]	.88/.77 [.40] [.22]	.80/.51 [.83] [.53]
MSV		.99/.95 [.52] [.27]	.92/.71 [.89] [.62]
BY			.88/ .62 [.38] [−.03]

Correlations between the residual ψs of the different asset pricing models using quarterly data 1947:Q1–2009:Q4 (Panel A) and annual data 1929–2009 (Panel B). Each cell has four entries, which indicate whether the models' state variables are extracted from the consumption data, shown at the top, or from the asset market data, shown at the bottom, and whether the residual psi is estimated using Equation (6), shown on the left, or using Equation (4), shown on the right. Other notation as in Table I.

Figure 11 plots the time series of the residual ψs for the four models at the quarterly (Panel A) and annual (Panel B) frequencies, with state variables extracted from the con-

[19] Note that the estimates from the annual data are inherently more imprecise, due to the small sample size available, than those from the quarterly data.

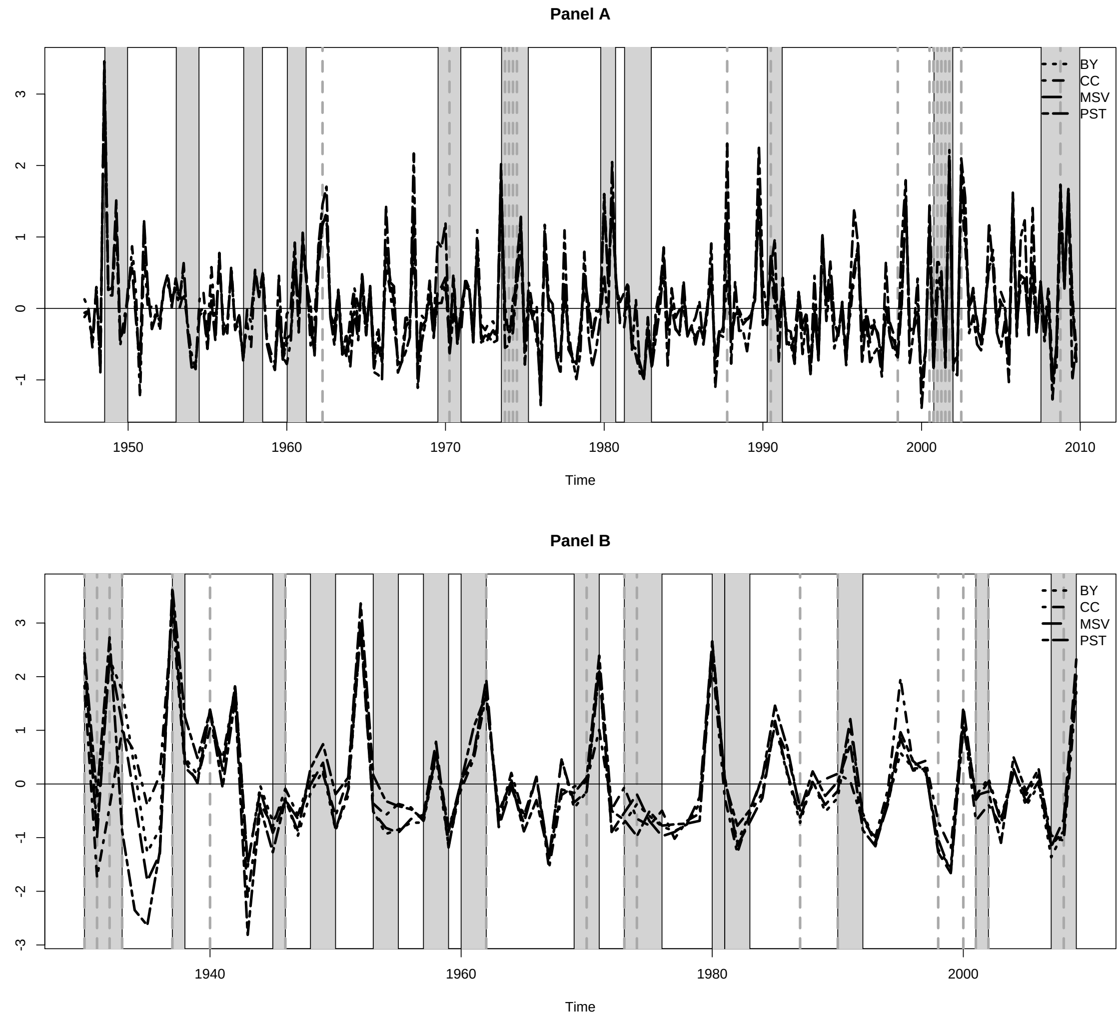


Figure 11: The (log) residual ψ components, $\ln(\psi_t^{resid})$, of the SDFs ($M_t^* = M_t^m \psi_t^{resid}$) filtered using Equation (6). Quarterly data, 1947:Q1–2009:Q4, in Panel A, and annual data, 1929–2009, in Panel B, for the same models as in Figure 9. Shaded areas denote NBER recession years, and vertical dashed lines indicate the major stock market crashes identified by Mishkin and White (2002).

sumption data and ψ^{resid} estimated using Equation (6). The results suggest that these models are all missing a very similar component that would improve their ability to explain the asset return dynamics. In particular, all the ψ^{resid} have a clear business cycle pattern, but also show significant and sharp reactions to financial market crashes that do not result in economy-wide contractions.

To further illustrate this point, Table X presents the changes in the model-implied risk neutral probabilities needed to rationalize the stock returns according to ψ^{resid}, that is, the percentage change caused by replacing M^m with $M^m \times \psi^{resid}$. As before, we have four entries per model since we compute the probabilities when the state variables are extracted

Table X: Change in Risk Neutral Probabilities due to Residual ψs

	Recession Probability	Market Crash Probability	Market Crash without Recession Probability
Panel A: Quarterly Data, 1947:Q1–2009:Q4			
CC	10/11	60/59	72/105
	[9/14]	[78/78]	[133/144]
BY	14/15	69/68	107/136
	[−65/−67]	[−31/−32]	[84/144]
MSV	15/15	78/74	124/144
	[12/11]	[53/53]	[93/126]
PST	17/20	98/75	232/148
Panel B: Annual Data, 1929–2009			
CC	−1/ − 1	−2/1	10/36
	[21/17]	[73/85]	[11/81]
BY	42/37	84/86	−2/37
	[−22/−8]	[−45/43]	[5/−24]
MSV	50/46	92/92	7/39
	[43/39]	[61/63]	[3/33]
PST	58/57	64/71	−3/69

Percentage changes in risk neutral probabilities generated by the the residual ψ component. Recession probabilities. Columns 1 to 3 focus on the probabilities of, respectively, recession, market crash, and market crash without recession. Each cell has four entries, which indicate whether the models' state variables are extracted from consumption data, shown at the top, or from asset market data, shown at the bottom, and whether the residual psi is estimated using Equation (6), shown on the left, or using Equation (4), shown on the right. Other notation as in Table I.

using the consumption data as well as using the asset price data (in brackets below), and using two minimum entropy methods (left and right numbers). Focusing on the quarterly data in Panel A, three patterns emerge. First (column 1), ψ^{resid} implies a relatively small increase in the risk neutral probability of recessions, suggesting that the models considered tend to adequately capture business cycle risk at this frequency (with the exception of BY when the state variables are extracted from the asset prices, which seems to imply too much recession risk). Second (column 2), all the models seem to imply a too low risk neutral probability of a market crash, i.e., ψ^{resid} increases this quantity by about 53%–98% (with again the exception of BY, which seems to imply too much crash risk). Third (column 3), all the models imply a much too low probability of market crashes not concomitant with recessions: ψ^{resid} increases the risk neutral likelihood of these events by about 72%–232%. Panel B shows a similar pattern, albeit the probability of market crashes without recessions are harder to identify at this frequency. Overall, Table X suggests that the models do not seem to price correctly the market crash risk, especially for market crashes that do not lead to large real economic contractions.

To summarize, the results in this section suggest that the consumption based asset pricing models we have considered would benefit from being augmented with a component that exhibits significant reactions to financial market crashes, in particular crashes that do not result in macroeconomic contractions. Moreover, not only the standard C-CAPM with power utility, but also most of the more recent models that have been proposed in the literature, seem to be missing this component.

V Conclusion

In this paper, we propose an information-theoretic approach as a diagnostic tool for dynamic asset pricing models. The models we consider are characterized by having a pricing kernel that can be factorized into an observable component, consisting of a parametric function of observable variables, and a potentially unobservable one that is model-specific.

Based on this decomposition of the pricing kernel, we provide three major contributions. First, using a relative entropy minimization approach, we show how to non-parametrically extract the time series of both the SDF and its unobservable component. Given the data, this methodology identifies the most likely—in the information theoretic sense—time series of the SDF and its unobservable component. Moreover, given a fully observable pricing kernel, this procedure delivers the most likely modification of the SDF that would enable it to price asset returns correctly. Applying this methodology to the data, we find that the estimated SDF has a clear business cycle pattern, but also shows significant and sharp reactions to financial market crashes that do not result in economy-wide contractions. Moreover, we find that the non-parametrically extracted SDF, independently of the set of assets used for its construction, is substantially (yet not perfectly) correlated with the risk factors proposed in Fama and French (1993). This provides a rationalization of the empirical success of the Fama French factors in pricing asset returns, and suggests that our filtering procedure does successfully identify theSDF.

Second, we construct a new set of entropy bounds that build upon and improve the ones suggested in the previous literature in that *a*) they naturally impose the non-negativity of the pricing kernel, *b*) they are generally tighter and have higher information content, and *c*) allow using jointly the information contained in consumption data and a large cross-section of asset returns.

Third, applying the methodology developed in this paper to a large class of dynamic asset pricing models, we find that the SDFs implied by all of these models correlate poorly with our filtered most likely SDF, require implausibly high levels of risk aversion to satisfy our entropy bounds, and are all missing a similar component that exhibits significant reactions to financial market crashes that do not result in economy-wide macroeconomic contractions These results are robust to the choice of test assets used as well as the frequency of the data.

The methodology developed in this paper is quite general, and may be applied to any model that delivers well-defined Euler equations, such as models with heterogenous agents, limited stock market participation, and biased beliefs.

References

Ait-Sahalia, Y., and A. W. Lo (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53(2), 499–547.

——— (2000): "Nonparametric risk management and implied risk aversion," *Journal of Econometrics*, 94(1-2), 9–51.

Alvarez, F., and U. J. Jermann (2005): "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73(6), 1977–2016.

Amemiya, T. (1985): *Advanced econometrics*. Harvard University Press, Cambridge: Mass.

Backus, D., M. Chernov, and S. E. Zin (2011): "Sources of Entropy in Representative Agent Models," Working Paper 17219, National Bureau of Economic Research.

Bakshi, G. S., and F. Chabi-Yo (2011): "Variance Bounds on the Permanent and Transitory Components of Stochastic Discount Factors," *Journal of Financial Economics*, Forthcoming.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

Basak, S., and H. Yan (2010): "Equilibrium Asset Prices and Investor Behaviour in the Presence of Money Illusion," *The Review of Economic Studies*, 77(3), pp. 914–936.

Breeden, D. T. (1979): "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–96.

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger (1989): "Empirical Test of the Consumption-Oriented CAPM," *The Journal of Finance*, 44(2), 231–262.

Brown, D. E., and R. L. Smith (1990): "A Correspondence Principle for Relative Entropy Minimization," *Naval Research Logistics*, 37(2).

Brunnermeier, M. K., and C. Julliard (2007): "Money Illusion and Housing Frenzies," *Review of Financial Studies*, 20(5).

Campbell, J. Y. (1996): "Understanding Risk and Return," *Journal of Political Economy*, 104(2), 298–345.

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–51.

Cochrane, J. H. (1996): "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy*, 104(3), 572–621.

Constantinides, G. M., and D. Duffie (1996): "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 104(2), 219–40.

Constantinides, G. M., and A. Ghosh (2011): "Asset Pricing Tests with Long Run Risks in Consumption Growth," *The Review of Asset Pricing Studies*, 1, 96–136.

Csiszar, I. (1975): "I-Divergence Geometry of Probability Distributions and Minimization Problems," *Annals of Probability*, 3, 146–158.

Duffie, D. (2005): *Dynamic Asset Pricing Theory*. New Age International (P) Ltd.

Epstein, L. G., and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–968.

Fama, E. F., and K. R. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *The Journal of Financial Economics*, 33, 3–56.

Hamilton, J. D., and J. C. Wu (2012): "Identification and estimation of Gaussian affine term structure models," *Journal of Econometrics*, 168(2), 315–331.

Hansen, L. P. (2014): "Nobel Lecture: Uncertainty Outside and Inside Economic Models," *Journal of Political Economy*, 122(5), 945 – 987.

Hansen, L. P., J. Heaton, J. Lee, and N. Roussanov (2007): "Intertemporal Substitution and Risk Aversion," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6 of *Handbook of Econometrics*, chap. 61. Elsevier.

Hansen, L. P., J. Heaton, and E. G. J. Luttmer (1995): "Econometric Evaluation of Asset Pricing Models," *The Review of Financial Studies*, 8(2), pp. 237–274.

Hansen, L. P., J. C. Heaton, and N. Li (2008): "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 116(2), 260–302.

Hansen, L. P., and R. Jagannathan (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99(2), 225–62.

Hansen, L. P., and R. Jagannathan (1997): "Asssessing Specification Errors in Stochastic Discount Factor Models," *The Journal of Finance*, 52, 557–590.

Hansen, L. P., and T. J. Sargent (2010): "Fragile Beliefs and the Price of Uncertainty," *Quantitative Economics*, 1(1), 129–162.

Hansen, L. P., and K. J. Singleton (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91, 249–68.

Hayashi, F. (2000): *Econometrics*. Princeton University Press: Princeton.

Hodrick, R. J., and E. C. Prescott (1997): "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit, and Banking*, 29(1), 1–16.

Jackwerth, J. C., and M. Rubinstein (1996): "Recovering Probability Distributions from Option Prices," *Journal of Finance*, 51(5), 1611–32.

Jaynes, E. T. (1957a): "Information Theory and Statistical Mechanics," *The Physics Review*, 106(4), 620–630.

——— (1957b): "Information theory and statistical mechanics II," *The Physical Review*, 108(2), 171–190.

Julliard, C., and A. Ghosh (2012): "Can Rare Events Explain the Equity Premium Puzzle?," *Review of Financial Studies*, 25(10), 3037–3076.

Kitamura, Y. (2006): "Empirical Likelihood Methods in Econometrics: Theory and Practice," Cowles Foundation Discussion Papers 1569, Cowles Foundation, Yale University.

Kitamura, Y., and M. Stutzer (2002): "Connections between entropic and linear projections in asset pricing estimation," *Journal of Econometrics*, 107(1-2), 159–174.

Kreps, D. M., and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46(1), 185–200.

Lettau, M., and S. Ludvigson (2001a): "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56(3), 815–49.

——— (2001b): "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, 109, 1238–1286.

Lucas, Jr., R. E. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1426–14.

Lukacs, E. (1970): *Characteristic Functions (2nd Edition)*, Griffin books. Hafner Publishing, London.

Lustig, H. N., and S. G. V. Nieuwerburgh (2005): "Housing Collateral, Consumption Insurance, and Risk Premia: An Empirical Perspective," *The Journal of Finance*, 60(3), pp. 1167–1219.

Mankiw, N. G., and M. D. Shapiro (1986): "Risk and Return: Consumption Beta Versus Market Beta," *Review of Economics and Statistics*, 68, 452–59.

Mehra, R., and E. C. Prescott (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15(2), 145–61.

Menzly, L., T. Santos, and P. Veronesi (2004): "Understanding Predictability," *Journal of Political Economy*, 112(1), 1–47.

Mishkin, F. S., and E. N. White (2002): "U.S. Stock Market Crashes and Their Aftermath: Implications for Monetary Policy," NBER Working Papers 8992, National Bureau of Economic Research, Inc.

Owen, A. B. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75(2), 237–249.

——— (1991): "Empirical Likelihood for Linear Models," *The Annals of Statistics*, 19(4), 1725–1747.

——— (2001): *Empirical Likelihood*. Chapman and Hall.

Parker, J. A., and C. Julliard (2005): "Consumption Risk and the Cross-Section of Expected Returns," *Journal of Political Economy*, 113(1).

Piazzesi, M., M. Schneider, and S. Tuzel (2007): "Housing, consumption and asset pricing," *Journal of Financial Economics*, 83(3), 531–569.

Robinson, P. M. (1991): "Consistent Nonparametric Entropy-based Testing," *Review of Economic Studies*, 58(3), 437–453.

Rosenberg, J. V., and R. F. Engle (2002): "Empirical pricing kernels," *Journal of Financial Economics*, 64(3), 341–372.

Ross, S. A. (2011): "The Recovery Theorem," Working Paper 17323, National Bureau of Economic Research.

Rubinstein, M. (1976): "The Valuation of Uncertain Income Streams and the Pricing of Options," *Bell Journal of Economics*, 7, 407–25.

Savov, A. (2011): "Asset Pricing with Garbage," *Journal of Finance*, 66(1), 177–201.

Schervish, M. J. (1995): *Theory of statistics*, Springer series in statistics. Springer-Verlag.

Stutzer, M. (1995): "A Bayesian approach to diagnosis of asset pricing models," *Journal of Econometrics*, 68(2), 367 – 397.

——— (1996): "A simple Nonparametric Approach to Derivative Security Valuation," *Journal of Finance*, LI(5), 1633–1652.

Ulrich, M. (2010): "Observable Long-Run Ambiguity and Long-Run Risk," Columbia University Manuscript.

Weil, P. (1989): "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421.

White, H. (1982): "Maximum Likelihood of Misspecified Models," *Econometrica*, 50, 1–25.

A Appendix

A.1 Maximum Likelihood Interpretation

To formally show the analogy between our estimation approach for the measures Ψ and Q and an MLE procedure, we have to consider the two definitions of relative entropy (and corresponding estimators) separately.

First, consider the entropy minimization problem of the type $D(P||x)$, with x being either the Q or the Ψ measure, used to contruct the estimators in Equations (7) and (6). Let the vector $\mathbf{z}_t$ be a sufficient statistic for the state of the economy at time t. That is, $\mathbf{z}_t$ can be thought of as an augmented state vector (e.g., containing the beginning of period state variables, as well as the time t realizations of the shocks and expectations about the future). Given $\mathbf{z}_t$, the equilibrium quantities, such as returns $\mathbf{R}^e$ and the SDF M, are just a mapping from $\mathbf{z}$ on to the real line, i.e.,

$$M(\mathbf{z}) : \mathbf{z} \rightarrow \mathbb{R}_+, \quad \mathbf{R}^e(\mathbf{z}) : \mathbf{z} \rightarrow \mathbb{R}^N, \quad M_t \equiv M(\mathbf{z}_t), \quad \mathbf{R}^e_t \equiv \mathbf{R}^e(\mathbf{z}_t)$$

where $\mathbf{z}_t$ is the time t realization of $\mathbf{z}$.

Equipped with the above definition, we can rewrite the Euler equation (3) as

$$0 = \mathbb{E}[\mathbf{R}^e_t M_t] \equiv \int \mathbf{R}^e_t M_t dP = \int \mathbf{R}^e(\mathbf{z}) M(\mathbf{z}) p(\mathbf{z}) d\mathbf{z} \tag{26}$$

where $p(\mathbf{z})$ is the pdf associated with the physical measure P. Moving to the risk neutral measure we have

$$0 = \mathbb{E}[\mathbf{R}^e_t M_t] = \mathbb{E}^Q[\mathbf{R}^e_t] = \int \mathbf{R}^e(\mathbf{z}) q(\mathbf{z}) d\mathbf{z} \tag{27}$$

where $q(\mathbf{z})$ is the pdf associated with the risk neutral measure Q and $M/\bar{M} = dQ/dP$. Note that

$$D(P||Q) = \int \ln \frac{dP}{dQ} dP = \int p(\mathbf{z}) \ln p(\mathbf{z}) d\mathbf{z} - \int p(\mathbf{z}) \ln q(\mathbf{z}) d\mathbf{z}.$$

Since the first term on the right hand side of the above expression does not involve q, $D(P||Q)$ is minimized, with respect to q, by choosing the distribution that maximizes the second term, i.e.,

$$Q^* \equiv \arg\min_Q D(P||Q) \equiv \arg\max_q \mathbb{E}[\ln q(\mathbf{z})] \text{ s.t. } \mathbb{E}^Q[\mathbf{R}^e_t] = 0.$$

That is, the minimum entropy estimator in Equation (7) maximizes the expected—risk neutral—log likelihood. Note that ML with a risk neutral likelihood is not uncommon, for instance in term structure modelling (see, e.g., Hamilton and Wu (2012)). Following Owen (1988, 1991, 2001), approximating the continuous distribution $q(\mathbf{z})$ with a multinomial distribution $\{q_t\}_{t=1}^T$ that assigns probability weight q_t to the time t realization of $\mathbf{z}$, a non-parametric maximum likelihood estimator (NPMLE) of Q can be obtained as

$$\{q_t^*\}_{t=1}^T = \arg\max \frac{1}{T}\sum_{t=1}^T \ln q_t \tag{28}$$

$$\text{s.t. } q_t \in \Delta^T \equiv \left\{(q_1, q_2, \ldots, q_T) : q_t \geqslant 0, \sum_{t=1}^T q_t = 1\right\} \text{ and (27) holds,}$$

provided that

$$\frac{1}{T}\sum_{t=1}^{T}\ln q_t \underset{T\to\infty}{\overset{p}{\longrightarrow}} \mathbb{E}\left[\ln q\left(\mathbf{z}\right)\right].$$

Note also that the NPMLE of $p(\mathbf{z})$ is simply $p_t = 1/T \ \forall t$ (see, e.g., Owen (1988, 1991, 2001)), i.e., the maximum entropy distribution. Therefore q^* contains all the necessary information to recover the state-price density from the Radon–Nikodym derivative dQ/dP.

Similarly, we have that

$$\begin{aligned}\Psi^* &\equiv \arg\min_{\Psi} D\left(P||\Psi\right) \equiv \arg\min_{\psi} \int p\left(\mathbf{z}\right)\ln p\left(\mathbf{z}\right)d\mathbf{z} - \int p\left(\mathbf{z}\right)\ln\psi\left(\mathbf{z}\right)d\mathbf{z} \\ &\equiv \arg\max_{\psi}\ \ \mathbb{E}\left[\ln\psi\left(\mathbf{z}\right)\right]\ \text{ s.t. }\ \mathbb{E}^{\Psi}\left[\mathbf{R}_t^e m_t\right] = 0\end{aligned}$$

where $\psi\left(\mathbf{z}\right)$ is the pdf associated with the measure Ψ. That is, the Ψ^* estimator in Equation (6) is also an MLE. Moreover, in a very similar fashion, one can show that $\psi^* m$ provides a MLE of q under the restriction that the pricing kernel has the multiplicative representation $M = m\psi$.

Hence, the estimates Q^* and Ψ^* maximize the log likelihoods of the data, but not the physical ones: the risk neutral log likelihood in the first case and an intermediate one in the second case (and Ψ^* can also be interpreted as maximizing the risk-neutral log likelihood under the constraint that $M_t = m_t\psi_t$).

Remark 1 *The above implies that, for any equilibrium quantity A_t, we have that $A_t \equiv A\left(\mathbf{z}_t\right)$. Hence, the risk neutral expectation of any function $f\left(.\right)$ of A, defined as*

$$\mathbb{E}^{Q}\left[f\left(A_t\right)\right] \equiv \int f\left(A\left(\mathbf{z}\right)\right)q\left(\mathbf{z}\right)d\mathbf{z},$$

can be estimated by (see, e.g., Kitamura (2006))

$$\widehat{\mathbb{E}^{Q}\left[f\left(A_t\right)\right]} = \sum_{t=1}^{T} f\left(A_t\right)q_t^*,$$

where q_t^* is the relative entropy minimizing risk neutral measure. For instance, the risk neutral probability of a recession in a given year, i.e., $\mathbb{E}^{Q}\left[\mathbf{1}_{\{rec.\ at\ t\}}\right]$, where $\mathbf{1}_{\{rec.\ at\ t\}}$ is an indicator function that takes the value one if year t was an NBER designated recession and zero otherwise, can be estimated by $\sum_{t=1}^{T}\mathbf{1}_{\{rec.\ at\ t\}}q_t^*$.

Second, consider the entropy minimization problem of the type $D\left(x||P\right)$ with x being either the Q or the Ψ measure. This alternative definition of relative entropy in Equations (5) and (4) also delivers non-parametric maximum likelihood estimates of the Q and Ψ measures, respectively. We establish this result for Ψ^* since for Q^* the same result can be shown by a simplified version of the same argument.

To see why the estimation problem in Equation (4) delivers an MLE of ψ_t, consider the following procedure for constructing (up to scale) the series $\{\psi_t\}_{t=1}^{T}$. First, given an integer $N >> 0$, distribute to the various points in time $t = 1, ..., T$, at random and with equal probabilities, the value $1/N$ in N independent draws. That is, draw a series of values (probability weights) $\left\{\tilde{\psi}\right\}_{t=1}^{T}$ given by

$$\tilde{\psi}_t \equiv \frac{n_t}{N}$$

where n_t measures the number of times that the value $1/N$ has been assigned to time t. Second, check whether the drawn series $\left\{\tilde{\psi}\right\}_{t=1}^{T}$ satisfies the pricing restriction $\sum_{t=1}^{T} m(\theta,t)\, R_t^e \tilde{\psi}_t = 0$. If it does, use this series as the estimator of $\{\psi_t\}_{t=1}^{T}$, and if it doesn't, draw another series. Obviously, a more efficient way of finding an estimate for ψ_t would be to choose the most likely outcome of the above procedure. Noting that the distribution of the $\tilde{\psi}_t$ is, by construction, a multinomial distribution with support given by the data sample, we have that the likelihood of any particular sequence $\left\{\tilde{\psi}_t\right\}_{t=1}^{T}$ is

$$L\left(\left\{\tilde{\psi}_t\right\}_{t=1}^{T}\right) = \frac{N!}{n_1! n_2! \ldots n_T!} \times T^{-N} = \frac{N!}{N\tilde{\psi}_1! N\tilde{\psi}_2! \ldots N\tilde{\psi}_T!} \times T^{-N}.$$

Therefore, the most likely value of $\left\{\tilde{\psi}_t\right\}_{t=1}^{T}$ maximizes the log likelihood

$$\ln L\left(\left\{\tilde{\psi}_t\right\}_{t=1}^{T}\right) \propto \frac{1}{N}\left(\ln N! - \sum_{t=1}^{T} \ln\left(N\tilde{\psi}_t!\right)\right).$$

Since the above procedure of assigning probability weights will become more and more accurate as N increases, we would ideally like to have $N \to \infty$. But in this case one can show[20] that

$$\lim_{N\to\infty} \ln L\left(\left\{\tilde{\psi}_t\right\}_{t=1}^{T}\right) = -\sum_{t=1}^{T} \tilde{\psi}_t \ln \tilde{\psi}_t.$$

Therefore, taking into account the constraint for the pricing kernel, the MLE of ψ_t would solve

$$\left\{\hat{\psi}_t\right\}_{t=1}^{T} \equiv \arg\max -\sum_{t=1}^{T} \tilde{\psi}_t \ln \tilde{\psi}_t, \quad \text{s.t.} \ \left\{\widetilde{\psi}_t\right\}_{t=1}^{T} \in \Delta^T, \ \sum_{t=1}^{T} m(\theta,t)\, \mathbf{R}_t^e \widetilde{\psi}_t = \mathbf{0}.$$

But the solution of the above MLE problem is also the solution of the relative entropy minimization problem in Equation (4) (see, e.g., Csiszar (1975)). That is, the KLIC minimization is equivalent to maximizing the likelihood in an unbiased procedure for finding the ψ_t component of the pricing kernel.

A.2 Additional Bounds and Derivations

Remark 2 *(HJ bounds as approximate Q bounds). Let p and q denote the densities of the state x associated, respectively, with the physical, P, and the risk neutral, Q, probability measures.*[21] *Assume that*

A.1 q and p are twice continuously differentiable;

and that there exists a $\mu_p < \infty$ and a $\mu_q < \infty$ such that

[20]Recall that from Stirling's formula, we have

$$\lim_{N\tilde{\psi}_t\to\infty} \frac{N\tilde{\psi}_t!}{\sqrt{2\pi N\tilde{\psi}_t}\left(\frac{N\tilde{\psi}_t}{e}\right)^{N\tilde{\psi}_t}} = 1.$$

[21]For expositional simplicity, we focus on a scalar state variable, but it is straightforward to extend the result to a vector state.

A.2 (Existence of maxima)

$$\left.\frac{\partial \ln p}{\partial x}\right|_{x=\mu_p} = 0, \quad \left.\frac{\partial \ln q}{\partial x}\right|_{x=\mu_q} = 0; \tag{29}$$

A.3 (Finite second moments)

$$-\left[\left.\frac{\partial^2 \ln p}{\partial x^2}\right|_{x=\mu_p}\right]^{-1} \equiv \sigma_p^2 < \infty, \; -\left[\left.\frac{\partial^2 \ln q}{\partial x^2}\right|_{x=\mu_q}\right]^{-1} \equiv \sigma_q^2 < \infty. \tag{30}$$

We have that in the limit of a small time interval, a second order approximation of the Q bounds yields[22]

$$D\left(P||\frac{M_t}{\bar{M}}\right) \propto Var\left(M_t\right), \tag{31}$$

$$D\left(\frac{M_t}{\bar{M}}||P\right) \propto Var\left(M_t\right). \tag{32}$$

Proof of Remark 2. We can then rewrite the $Q1$ and $Q2$ bounds as

$$D\left(P||\frac{M_t}{\bar{M}}\right) \equiv \int \ln\frac{dP}{dQ}dP = \int p \ln\frac{p}{q}dx \tag{33}$$

and

$$D\left(\frac{M_t}{\bar{M}}||P\right) \equiv \int \frac{dQ}{dP}\ln\frac{dQ}{dP}dP = \int \ln\frac{dQ}{dP}dQ = \int q\ln\frac{q}{p}dx. \tag{34}$$

Given conditions $A.1$–$A.3$, we have from a second order Taylor approximation that

$$\ln q \propto \frac{1}{2}\left.\frac{\partial^2 \ln q}{\partial x^2}\right|_{x=\mu_q} \left(x-\mu_q\right)^2 \equiv -\frac{1}{2}\frac{\left(x-\mu_q\right)^2}{\sigma_q^2}$$

$$\ln p \propto \frac{1}{2}\left.\frac{\partial^2 \ln p}{\partial x^2}\right|_{x=\mu_p} \left(x-\mu_p\right)^2 \equiv -\frac{1}{2}\frac{\left(x-\mu_p\right)^2}{\sigma_p^2}$$

That is, q and p are approximately (up to second order) Gaussian

$$q \approx N\left(\mu_q;\sigma_q^2\right), \quad p \approx N\left(\mu_p;\sigma_p^2\right).$$

Note also that in the limit of a small time interval, by the diffusion invariance principle, we have $\sigma_q^2 = \sigma_p^2 = \sigma^2$. Therefore, plugging the above approximation into Equation (33), we

[22] For the $Q2$ bound only, using the dual objective function of the entropy minimization problem, Stutzer (1995) provides a similar approximation result to the one in Equation (32) that is valid when the variance bound is sufficiently small. Moreover, for the case of Gaussian i.i.d. returns, Kitamura and Stutzer (2002) show that the approximation of the $Q2$ bound in Equation (32) is exact.

have that in the limit of a small time interval

$$\int p \ln \frac{p}{q} dx \approx \int \left[-\frac{1}{2} \frac{(x-\mu_p)^2}{\sigma^2} + \frac{1}{2} \frac{(x-\mu_q)^2}{\sigma^2} \right] pdx$$
$$= \frac{1}{2\sigma^2} \left[-\sigma^2 + \int (x-\mu_q)^2 \, pdx \right]$$
$$= \frac{1}{2\sigma^2} \{ -\sigma^2 + \int \left[(x-\mu_p)^2 + (\mu_p - \mu_q)^2 \right.$$
$$\left. +2 (\mu_p - \mu_q)(x-\mu_p) \right] pdx \}$$
$$= \frac{1}{2\sigma^2} (\mu_p - \mu_q)^2 = \frac{1}{2\sigma^2} \sigma^2 \sigma_\xi^2 = \frac{1}{2} \sigma_\xi^2$$

where the density ξ is a (strictly positive) martingale defined by $\xi \equiv \frac{dQ}{dP}$, and the one to the last equality comes from the change of drift implied by the Girsanov's Theorem (see, e.g., Duffie (2005, Appendix D)).
Similarly, from Equation (34) we have

$$\int q \ln \frac{q}{p} dx = \frac{1}{2} \sigma_\xi^2.$$

Since Q and P are equivalent measures, $M_t \propto \xi_t$. Therefore, in the limit of a small time interval, $Var(M_t) \propto \sigma_\xi^2$, implying

$$D\left(P || \frac{M_t}{\bar{M}}\right) \propto Var(M_t), \quad D\left(\frac{M_t}{\bar{M}} || P\right) \propto Var(M_t).$$

■

Definition 5 (Volatility bound for ψ_t) *For each $E[\psi_t] = \bar{\psi}$, the minimum variance ψ_t is*

$$\psi_t^*(\bar{\psi}) \equiv \arg\min_{\{\psi_t(\bar{\psi})\}_{t=1}^T} \sqrt{Var(\psi_t(\bar{\psi}))} \ \textit{s.t.} \ \mathbf{0} = \mathbb{E}\left[\mathbf{R}_t^e m(\theta,t) \psi_t(\bar{\psi})\right]$$

and any candidate SDF must satisfy the condition $Var(\psi_t) \geq Var\left(\psi_t^*(\bar{\psi})\right)$.

The solution of the above minimization for a given θ is

$$\psi_t^*(\bar{\psi}) = \bar{\psi} + (\mathbf{R}_t^e m(\theta,t) - \mathbb{E}[\mathbf{R}_t^e m(\theta,t)])' \beta_{\bar{\psi}}$$

where $\beta_{\bar{\psi}} = Var(\mathbf{R}_t^e m(\theta,t))^{-1} \left(-\bar{\psi}\mathbb{E}[\mathbf{R}_t^e m(\theta,t)]\right)$ and the lower volatility bound is given by

$$\sigma_{\psi^*} \equiv \sqrt{Var\left(\psi_t^*(\bar{\psi})\right)} = \bar{\psi}\sqrt{\mathbb{E}[\mathbf{R}_t^e m(\theta,t)]' Var(\mathbf{R}_t^e m(\theta,t))^{-1} \mathbb{E}[\mathbf{R}_t^e m(\theta,t)]}.$$

A.3 Data Description

For the quarterly data, we use 4 different sets of assets: *i*) the 25 Fama–French portfolios, *ii*) the 10 momentum-sorted portfolios, *iii*) the 10 industry-sorted portfolios, and *iv*) a combined set of 10 industry, 10 momentum and 6 size and book-to-market sorted portfolios. For the annual data, we use the same sets of assets except the 25 Fama–French portfolios, which are replaced by the 6 portfolios formed by sorting stocks on the basis of size and book-to-market-equity because of the small time series dimension available at the annual level.

Our proxy for the market return is the Center for Research in Security Prices (CRSP) value-weighted index of all stocks on the NYSE, AMEX, and NASDAQ. The proxy for the risk-free rate is the one-month Treasury Bill rate obtained from the CRSP files. The returns on all the portfolios are obtained from Kenneth French's data library. Quarterly (annual) returns for the above assets are computed by compounding monthly returns within each quarter (year), and converted to real returns using the personal consumption deflator. Excess returns on the assets are then computed by subtracting the risk free rate.

Finally, for each dynamic asset pricing model, the information bounds and the non-parametrically extracted and model-implied time series of the SDF depend on the consumption data. For the standard Consumption-CAPM of Breeden (1979) and Rubinstein (1976), the external habit models of Campbell and Cochrane (1999) and Menzly, Santos, and Veronesi (2004), and the long-run risks model of Bansal and Yaron (2004), we use per capita real personal consumption expenditures on nondurable goods from the National Income and Product Accounts (NIPA). We make the standard "end-of-period" timing assumption that consumption during quarter t takes place at the end of the quarter. For the housing model of Piazzesi, Schneider, and Tuzel (2007) aggregate consumption is measured as expenditures on non-durables and services excluding housing services.

A.4 HJ Kernel Versus Minimum Entropy Kernel

Table A1: Moments of SDF, 1947:Q1–2009:Q4

	$\sigma\left(M_t^*\right)$	$Sk\left(M_t^*\right)$	$Kurt\left(M_t^*\right)$
Panel A: HJ Kernel			
25 FF	.45	−.01	3.12
Market	.22	.61	3.91
10 Momentum	.41	.05	3.41
10 Industry	.32	.54	4.21
Panel B: Minimum Entropy Kernel			
25 FF	.91/.71	4.53/2.07	28.4/9.31
Market	.26/.24	3.14/1.84	19.1/8.59
10 Momentum	.69/.57	3.82/1.78	22.0/7.22
10 Industry	.45/.39	5.08/2.32	39.6/11.8

Moments of the SDF computed using the *(i)* the HJ minimum linear adjustment (Panel A) and *(ii)* the minimum relative entropy log-linear adjustment (Panel B). The test assets used in the estimation of the minimum adjustment consist of the 25 size and book-to-market-equity portfolios (row 1), the market portfolio (row 2), the 10 momentum portfolios (row 3), and the 10 industry portfolios (row 4). Quarterly data 1947:Q1–2009:Q4.

A.5 Extracting the Model-Implied SDF for the Menzly, Santos, and Veronesi (2004) Model

The SDF in this model is given by

$$M_t = \delta\left(C_t/C_{t-1}\right)^{-\gamma}\left(S_t/S_{t-1}\right)^{-\gamma}, \tag{35}$$

where δ is the subjective time discount factor, γ is the utility curvature parameter, $S_t = \frac{C_t - X_t}{C_t}$ denotes the surplus consumption ratio, and X_t is the habit component.

The inverse surplus, $Y_t = \frac{1}{S_t}$, follows a mean-reverting process:

$$dY_t = k\left(\overline{Y} - Y_t\right)dt - \alpha\left(Y_t - \lambda\right)\sigma_c dB_t.$$

Therefore, using Ito's Lemma, $s_t \equiv \ln\left(S_t\right) = -\ln\left(Y_t\right)$ follows the process

$$\begin{aligned} ds_t &= -\frac{1}{Y_t}dY_t + \frac{1}{2Y_t^2}\left(dY_t\right)^2 \\ &= -\frac{1}{Y_t}k\left(\overline{Y} - Y_t\right)dt + \frac{1}{Y_t}\alpha\left(Y_t - \lambda\right)\sigma_c dB_t + \frac{1}{2Y_t^2}\alpha^2\left(Y_t - \lambda\right)^2\sigma_c^2 dt \\ &= \left[k\left(1 - \overline{Y}S_t\right) + \frac{1}{2}\alpha^2\left(1 - \lambda S_t\right)^2\sigma_c^2\right]dt + \alpha\left(1 - \lambda S_t\right)\sigma_c dB_t. \end{aligned}$$

Therefore, discretizing the process, we have

$$\Delta s_{t+1} = k\left(1 - \overline{Y}S_t\right) + \frac{1}{2}\alpha^2\left(1 - \lambda S_t\right)^2\sigma_c^2 + \alpha\left(1 - \lambda S_t\right)\sigma_c\varepsilon_{t+1},$$

where $\varepsilon_{t+1} \sim$ i.i.d.$N\left(0, 1\right)$.

Now, the Euler equation for the return on the aggregate consumption claim is

$$E_t\left(e^{m_{t+1} + r_{c,tt+1}}\right) = 1, \tag{36}$$

where $r_{c,t+1}$ denotes the continuously compounded return on the consumption claim. We rely on log-linear approximations for $r_{c,t+1}$, as in Campbell and Shiller (1988):

$$r_{c,t+1} = \kappa_0 + \kappa_1 z_{t+1} - z_t + \Delta c_{t+1}, \tag{37}$$

where z_t is the log price–consumption ratio, $\kappa_1 = \frac{e^{\overline{z}}}{1+e^{\overline{z}}}$ and $\kappa_0 = \log(1 + e^{e^{\overline{z}}}) - \kappa_1 e^{\overline{z}}$, and $\overline{z}$ denotes the long-run mean of the log price–consumption ratio. We conjecture that z_t is affine in the single state variable s_t:

$$z_t = A_0 + A_1 s_t. \tag{38}$$

In order to verify the conjecture and also solve for A_0 and A_1, we substitute the expressions for $r_{c,t+1}$ and z_t from Equations (37) and (38), respectively, into the Euler equation (36):

$$E_t\left(\exp\left\{\ln\delta - \gamma\Delta c_{t+1} - \gamma\Delta s_{t+1} + \kappa_0 + \kappa_1 z_{c,t+1} - z_t + \Delta c_{t+1}\right\}\right) = 1,$$

$$\Rightarrow E_t\left(\exp\left\{\begin{array}{c} \ln\delta - \gamma\mu_c - \gamma\sigma_c\varepsilon_{t+1} - \gamma k\left(1 - \overline{Y}S_t\right) - \frac{1}{2}\gamma\alpha^2\left(1 - \lambda S_t\right)^2\sigma_c^2 - \gamma\alpha\left(1 - \lambda S_t\right)\sigma_c\varepsilon_{t+1} \\ +\kappa_0 + \kappa_1 A_0 + \kappa_1 A_1\left[k\left(1 - \overline{Y}S_t\right) + \frac{1}{2}\alpha^2\left(1 - \lambda S_t\right)^2\sigma_c^2 + \alpha\left(1 - \lambda S_t\right)\sigma_c\varepsilon_{t+1} + s_t\right] \\ -A_0 - A_1 s_t + \mu_c + \sigma_c\varepsilon_{t+1} \end{array}\right\}\right) = 1.$$

Using the properties of conditionally lognormal random variables, we have

$$\begin{aligned} 0 = {} & \ln\delta - \gamma\mu_c - \gamma k + \gamma k\overline{Y}S_t - \frac{1}{2}\gamma\alpha^2\sigma_c^2 - \frac{1}{2}\gamma\alpha^2\lambda^2\sigma_c^2 S_t^2 + \gamma\alpha^2\sigma_c^2\lambda S_t + \kappa_0 + \kappa_1 A_0 \\ & + \kappa_1 A_1 k - \kappa_1 A_1 k\overline{Y}S_t + \frac{1}{2}\kappa_1 A_1\alpha^2\sigma_c^2 + \frac{1}{2}\kappa_1 A_1\alpha^2\sigma_c^2\lambda^2 S_t^2 - \kappa_1 A_1\alpha^2\sigma_c^2\lambda S_t + \kappa_1 A_1 s_t \\ & - A_0 - A_1 s_t + \mu_c + \frac{1}{2}\left[-\gamma - \gamma\alpha\left(1 - \lambda S_t\right) + \kappa_1 A_1\alpha\left(1 - \lambda S_t\right) + 1\right]^2\sigma_c^2, \end{aligned}$$

which implies

$$0=\begin{pmatrix}\ln\delta-\gamma\mu_c-\gamma k-\frac{1}{2}\gamma\alpha^2\sigma_c^2+\kappa_0+\kappa_1A_0+\kappa_1A_1k+\frac{1}{2}\kappa_1A_1\alpha^2\sigma_c^2\\ -A_0+\mu_c+\frac{1}{2}\left[-\gamma-\gamma\alpha+\kappa_1A_1\alpha+1\right]^2\sigma_c^2\end{pmatrix}$$
$$+\begin{pmatrix}\gamma k\overline{Y}+\gamma\alpha^2\sigma_c^2\lambda-\kappa_1A_1k\overline{Y}-\kappa_1A_1\alpha^2\sigma_c^2\lambda\\ +\left[\gamma\alpha\lambda-\kappa_1A_1\alpha\lambda\right]\left[-\gamma-\gamma\alpha+\kappa_1A_1\alpha+1\right]\sigma_c^2\end{pmatrix}S_t$$
$$+\left(\kappa_1A_1-A_1\right)s_t$$
$$+\left(-\frac{1}{2}\gamma\alpha^2\lambda^2\sigma_c^2+\frac{1}{2}\kappa_1A_1\alpha^2\sigma_c^2\lambda^2+\frac{1}{2}\left(\gamma\alpha\lambda-\kappa_1A_1\alpha\lambda\right)^2\sigma_c^2\right)S_t^2.$$

Using the approximations $s_t\approx S_t-1$ and $S_t^2\approx-\overline{S}^2+2\overline{S}S_t$, we obtain

$$0=\begin{pmatrix}\ln\delta-\gamma\mu_c-\gamma k-\frac{1}{2}\gamma\alpha^2\sigma_c^2+\kappa_0+\kappa_1A_0+\kappa_1A_1k+\frac{1}{2}\kappa_1A_1\alpha^2\sigma_c^2\\ -A_0+\mu_c+\frac{1}{2}\left[-\gamma-\gamma\alpha+\kappa_1A_1\alpha+1\right]^2\sigma_c^2\end{pmatrix}$$
$$+\begin{pmatrix}\gamma k\overline{Y}+\gamma\alpha^2\sigma_c^2\lambda-\kappa_1A_1k\overline{Y}-\kappa_1A_1\alpha^2\sigma_c^2\lambda\\ +\left[\gamma\alpha\lambda-\kappa_1A_1\alpha\lambda\right]\left[-\gamma-\gamma\alpha+\kappa_1A_1\alpha+1\right]\sigma_c^2\end{pmatrix}S_t$$
$$+\left(\kappa_1A_1-A_1\right)\left(S_t-1\right)$$
$$+\left(-\frac{1}{2}\gamma\alpha^2\lambda^2\sigma_c^2+\frac{1}{2}\kappa_1A_1\alpha^2\sigma_c^2\lambda^2+\frac{1}{2}\left(\gamma\alpha\lambda-\kappa_1A_1\alpha\lambda\right)^2\sigma_c^2\right)\left(-\overline{S}^2+2\overline{S}S_t\right).$$

We use the method of undetermined coefficients and set to zero the constant term and the coefficient of S_t to obtain two equations in the two unknowns A_0 and A_1:

$$0=\begin{pmatrix}\ln\delta-\gamma\mu_c-\gamma k-\frac{1}{2}\gamma\alpha^2\sigma_c^2+\kappa_0+\kappa_1A_0+\kappa_1A_1k+\frac{1}{2}\kappa_1A_1\alpha^2\sigma_c^2\\ -A_0+\mu_c+\frac{1}{2}\left[-\gamma-\gamma\alpha+\kappa_1A_1\alpha+1\right]^2\sigma_c^2\end{pmatrix}$$
$$-\left(\kappa_1A_1-A_1\right)$$
$$-\left(-\frac{1}{2}\gamma\alpha^2\lambda^2\sigma_c^2+\frac{1}{2}\kappa_1A_1\alpha^2\sigma_c^2\lambda^2+\frac{1}{2}\left(\gamma\alpha\lambda-\kappa_1A_1\alpha\lambda\right)^2\sigma_c^2\right)\overline{S}^2. \tag{39}$$

and

$$0=\begin{pmatrix}\gamma k\overline{Y}+\gamma\alpha^2\sigma_c^2\lambda-\kappa_1A_1k\overline{Y}-\kappa_1A_1\alpha^2\sigma_c^2\lambda\\ +\left[\gamma\alpha\lambda-\kappa_1A_1\alpha\lambda\right]\left[-\gamma-\gamma\alpha+\kappa_1A_1\alpha+1\right]\sigma_c^2\end{pmatrix}$$
$$+\left(\kappa_1A_1-A_1\right)$$
$$+2\overline{S}\left(-\frac{1}{2}\gamma\alpha^2\lambda^2\sigma_c^2+\frac{1}{2}\kappa_1A_1\alpha^2\sigma_c^2\lambda^2+\frac{1}{2}\left(\gamma\alpha\lambda-\kappa_1A_1\alpha\lambda\right)^2\sigma_c^2\right). \tag{40}$$

Solving the equations for A_0 and A_1 gives the equilibrium solution for the log price–consumption ratio in Equation (38). Note that Equation (40) implies a quadratic equation for A_1:

$$0=\left(-\kappa_1^2\alpha^2\lambda\sigma_c^2+\overline{S}\kappa_1^2\alpha^2\lambda^2\sigma_c^2\right)A_1^2$$
$$+\begin{pmatrix}-\kappa_1k\overline{Y}-\kappa_1\alpha^2\sigma_c^2\lambda+\gamma\alpha\lambda\kappa_1\alpha\sigma_c^2-\kappa_1\alpha\lambda\left[-\gamma-\gamma\alpha+1\right]\sigma_c^2\\ +\kappa_1-1+\overline{S}\kappa_1\alpha^2\sigma_c^2\lambda^2-2\overline{S}\gamma\alpha^2\lambda\sigma_c^2\kappa_1\end{pmatrix}A_1$$
$$+\left(\gamma k\overline{Y}+\gamma\alpha^2\sigma_c^2\lambda+\gamma\alpha\lambda\left[-\gamma-\gamma\alpha+1\right]\sigma_c^2-\overline{S}\gamma\alpha^2\lambda^2\sigma_c^2+\overline{S}\gamma^2\alpha^2\lambda^2\sigma_c^2\right).$$

We choose the smaller root of the quadratic equation as the economically meaningful solution because it implies a positive relation between the log price–consumption ratio and the surplus consumption ratio, unlike the larger root, which implies a negative relation between the variables.

We proxy the log price–consumption ratio using the observable log price–dividend ratio and use Equation (38) to extract the time series of the state variable s_t. This extracted time series can then be used to obtain the time series of the model-implied SDF and its missing component.

Note that the model is calibrated quarterly. Since we evaluate the empirical plausibility of models at the quarterly as well as annual frequencies, we obtain the annual estimates of the model parameters as follows. First, we simulate a long sample (five million observations) of the state variable Y from

$$\Delta Y_{q,\tau+1} = k_q \left(\overline{Y}_q - Y_{q,\tau} \right) - \alpha_q \left(Y_{q,\tau} - \lambda_q \right) \sigma_{q,c} \varepsilon_{\tau+1}, \quad \varepsilon_{\tau+1} \sim \text{i.i.d.} N(0,1),$$

treating the calibrated quarterly parameter values as the truth. The subscript q in the above equation denotes quarterly. Second, we aggregate the simulated data into annual non-overlapping observations:

$$\begin{aligned} Y_{a,t} &= Y_{q,\tau} + Y_{q,\tau-1} + Y_{q,\tau-2} + Y_{q,\tau-3}, \text{ for } \tau = 1, 2, 3, \ldots \\ \Delta Y_{a,t+1} &= Y_{a,t+1} - Y_{a,t}, \end{aligned}$$

where τ denotes quarter τ and t denotes year t. Finally, we estimate the model parameters at the annual frequency from the equation

$$\Delta Y_{a,t+1} = k_a \left(\overline{Y}_a - Y_{a,t} \right) - \alpha_a \left(Y_{a,t} - \lambda_a \right) \sigma_{a,c} e_{t+1}, \quad \varepsilon_{t+1} \sim \text{i.i.d.} N(0,1),$$

treating the state variable $Y_{a,t}$ as observed and using the method of moments approach. This step produces the following annual estimates of the parameters: $\overline{Y}_a = 33.99531$, $k_a = .8689003$, $\alpha_a = 3.49499$, $\lambda_a = 29.843719$. The mean, $\mu_{a,c}$, and volatility, $\sigma_{a,c}$, of the aggregate consumption growth are set equal to their sample values.

A.6 Extracting the Model-Implied SDF for the Bansal and Yaron (2004) Model

The SDF in this model is given by

$$M_{t+1} = \delta^{\theta} \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{\theta}{\rho}} R_{c,t+1}^{\theta-1},$$

where $R_{c,t+1}$ is the unobservable gross return on an asset that delivers aggregate consumption as its dividend each period.

Using the Campbell–Shiller log-linearization for $r_{c,t+1} \equiv \ln \left(R_{c,t+1} \right)$,

$$r_{c,t+1} = \kappa_0 + \kappa_1 z_{t+1} - z_t + \Delta c_{t+1},$$

where z_t is the log price–consumption ratio, and noting that the model implies that the equilibrium $z_t = A_0 + A_1 x_t + A_2 \sigma_t^2$, we have

$$\begin{aligned} \ln M_t = {} & \left[\theta \ln \delta + (\theta - 1) \left(\kappa_0 + \kappa_1 A_0 - A_0 \right) \right] - \gamma \Delta c_{t+1} \\ & + (\theta - 1) \kappa_1 A_1 x_{t+1} + (\theta - 1) \kappa_1 A_2 \sigma_{t+1}^2 - (\theta - 1) A_1 x_t - (\theta - 1) A_2 \sigma_t^2, \end{aligned} \tag{41}$$

This is Equation (24) in the text. To obtain the time series of the SDF and its ψ component, we extract the state variables, x_t and σ_t^2, from the observed data using two different procedures.

First, we extract the state variables from the consumption data. In order to do so, we assume the same time series specification for the aggregate consumption growth process as in Bansal and Yaron (2004), with the only exception that we introduce a square-root process for the variance (as in Hansen, Heaton, Lee, and Roussanov (2007)):

$$\Delta c_{t+1} = \mu + x_t + \sigma_t \eta_{t+1} \tag{42}$$
$$x_{t+1} = \rho x_t + \phi_e \sigma_t e_{t+1} \tag{43}$$
$$\sigma_{t+1}^2 = \sigma^2(1-\nu_1) + \nu_1 \sigma_t^2 + \sigma_w \sigma_t w_{t+1}. \tag{44}$$

Note that the Bansal and Yaron (2004) model is calibrated at the monthly frequency with the *monthly* parameter values being: $\mu = .0015$, $\rho = .979$, $\phi_e = .044$, $\sigma = .0078$, $\nu_1 = .987$, $\sigma_w = .00029487$. We need to extract the *quarterly* state variables, $x_{t,q}$ and $\sigma_{t,q}^2$. First, we simulate a long sample (five million observations) from the above system, treating the given parameter values as the truth and retaining the simulated state variables. Second, we aggregate the simulated data into quarterly non-overlapping observations:

$$\begin{aligned} \Delta c_{t,q} &= \Delta c_t + \Delta c_{t-1} + \Delta c_{t-2}, \text{ for } t = 3, 6, 9, \ldots \\ x_{t,q} &= x_t + x_{t-1} + x_{t-2} \\ \sigma_{t,q}^2 &= \sigma_t^2 + \sigma_{t-1}^2 + \sigma_{t-2}^2 \end{aligned}$$

Third, we estimate the model parameters in Equations (42)–(44) using these quarterly observations and treating the state variables as observed, which produces the following quarterly estimates of the parameters:

$$\begin{aligned} \rho_q &= \rho_m^3 = .9383137 \\ v_{1,q} &= v_{1,m}^3 = .9615048 \\ \mu_q &= 3 \times \mu_m = .0045 \\ \sigma_q^2 &= Mean\left(\sigma_{t,q}^2\right) = .0001822490 \\ \phi_{e,q} &= \sqrt{\frac{Var\left(x_{t+1,q} - \rho_q x_{t,q}\right)}{\sigma_q^2}} = .1084845 \\ \sigma_{w,q} &= \sqrt{\frac{Var\left(\sigma_{t+1,q}^2 - \sigma_q^2\left(1 - v_{1,q}\right) - v_{1,q}\sigma_{t,q}^2\right)}{\sigma_q^2}} = 0.0007328592, \end{aligned}$$

where the variables with subscript m are the monthly calibrated values, and the means and variances are the ones obtained in the simulated sample. The fourth step is to run a Bayesian smoother through the historical quarterly consumption growth treating the quarterly parameters as being known with certainty. This smoother produces estimates of the quarterly state variables $\hat{x}_{t,q}$ and $\hat{\sigma}_{t,q}^2$.

The same steps can be applied to obtain the parameter estimates and, therefore, the time series of the state variables at the annual frequency. In this case, we have: $\rho_a = .7751617$; $v_{1,a} = .8546845$; $\mu_a = .018$; $\sigma_a^2 = .0007299038$; $\phi_{e,a} = .3853643$; $\sigma_{w,a} = .00270020$.

Using the point estimates of the parameters and the extracted time series of the state variables at the relevant frequency, the SDF and its missing ψ component are obtained from Equation (24).

Our second procedure for extracting the state variables relies on the asset market data. For the log-linearized version of the model, the observable log market-wide price–dividend ratio, $z_{m,t}$, and the log gross risk free rate, $r_{f,t}$, are affine functions of the state variables

x_t and σ_t^2. Therefore, Constantinides and Ghosh (2011) argue that these affine functions may be inverted to express the unobservable state variables, x_t and σ_t^2, in terms of the observables, $z_{m,t}$ and $r_{f,t}$. Following this approach, the pricing kernel in Equation (41) can be expressed as a function of the observable variables:

$$\ln M_t = c_1' - \gamma \Delta c_t + c_3' \left(r_{f,t} - \frac{1}{\kappa_1} r_{f,t-1} \right) + c_4' \left(z_{m,t} - \frac{1}{\kappa_1} z_{m,t-1} \right), \tag{45}$$

where the parameters (c_1', c_3', c_4') are functions of the underlying time-series and preference parameters.

Since the model is calibrated at the monthly frequency, we obtain the pricing kernels at the quarterly and annual frequencies by aggregating the monthly kernels. For instance, the quarterly pricing kernel, M^q, is obtained by

$$\ln M_t^q = -\gamma \Delta^q c_t + \ln \psi_t^q,$$

where $\Delta^q c_t$ denotes the quarterly log-consumption difference and $\ln \psi_t^q$ is given by

$$\ln \psi_t^q = 3c_1' + \sum_{i=0}^{2} \left[c_3' \left(r_{f,t-i} - \kappa_1 r_{f,t-i-1} \right) + c_4' \left(z_{m,t-i} - \kappa_1 z_{m,t-i-1} \right) \right].$$

Therefore, using the monthly calibrated parameter values from Bansal and Yaron (2004) and the historical monthly time series of the market-wide price–dividend ratio and risk free rate, we obtain the time series of the SDF and its missing component at the quarterly and annual frequencies from the above two equations.

A.7 Additional Robustness Checks

A.7.1 Entropy Bounds When All Model Parameters Are Simultaneously Allowed to Vary

In the empirical analysis on the entropy bounds, we focused on one-dimensional bounds as a function of the risk aversion parameter, γ, while fixing the other parameters at the authors' calibrated values. In other words, we computed the minimum values of γ at which the model-implied SDFs satisfy the HJ, Q, M, and Ψ bounds, while holding the remaining model parameters fixed at their calibrated values. As a robustness check, in this section, we compute the minimum values of γ at which the model-implied SDFs satisfy the bounds, while allowing the remaining model parameters to simultaneously vary over intervals that extend for two standard errors around their calibrated values.

For the external habit models of Campbell and Cochrane (1999) and Menzly, Santos, and Veronesi (2004), the model-implied SDFs are obtained by extracting the surplus consumption ratio from the aggregate consumption data. While the state variable may also be extracted from the price–dividend ratio, the Menzly, Santos, and Veronesi (2004) model admits a closed form solution for the price–dividend ratio only for $\gamma = 1$, and this motivates our choice for the extraction of the state variable in the external habit models. For the Bansal and Yaron (2004) long run risks model, on the other hand, we extract the two state variables by inverting the closed form solutions for the price–dividend ratio and risk free rate. While the state variables can also be extracted from the aggregate consumption data using Bayesian smoothing procedures, the computing time makes this prohibitively expensive if allowing all the parameters to vary simultaneously (since the Bayesian smoothing would have to be computed for each set of parameter values considered). Finally, for the Piazzesi, Schneider, and Tuzel (2007) model, the state variable is directly observable from the BEA tables and, therefore, does not need to be extracted from either the consumption or the asset market data.

The results are presented in Table A2, which shows that, for each model, the HJ, Q, M, and Ψ bounds are satisfied for smaller values of γ when the other parameters are allowed to vary simultaneously compared to Tables VI and VII where the other parameters are held fixed. However, as in the latter tables, the CC, MSV, and PST models still require much larger values of risk aversion to satisfy the bounds compared to the authors' calibrated values at the quarterly frequency.

Table A2: Bounds for RRA When All Parameters Are Allowed to Vary

	HJ Bound	$Q1/Q2$ Bounds	$M1/M2$ Bounds	$\Psi1/\Psi2$ Bounds
Panel A: Quarterly Data, 1947:Q1–2009:Q4				
CC	2.2	4.0/3.8	4.0/3.8	4.3/4.2
MSV	29.0	36.2/35.9	38.0/38.1	50.9/52.5
BY	3.0	4.0/4.0	4.0/4.0	4.0/4.0
PST	19.1	25.2/24.0	25.4/24.1	24.1/23.1
Panel B: Annual Data, 1929–2009				
CC	0.1	0.1/0.1	0.1/0.1	0.1/0.1
MSV	11.3	18.6/16.2	19.3/17.2	28.6/27.2
BY	4.0	4.0/4.0	4.0/4.0	4.0/4.0
PST	4.3	6.8/5.8	6.8/5.8	6.3/5.4

Minimum values of the utility curvature parameter γ at which the model-implied SDF satisfies the HJ (column 1), Q (column 2), M (column 3), and Ψ (column 4) bounds using quarterly data 1947:Q1–2009:Q4 (Panel A) and annual data 1929–2009 (Panel B). Columns 2–4 have two entries in each cell, which indicate whether the filtered ψ^*-component of the SDF and, therefore, the filtered SDF are estimated using Equation (6), shown on the left, or Equation (4), shown on the right. Other notation as in Table I.

A.7.2 Entropy Bounds When the Risk Free Rate is Included as an Additional Test Asset

In the empirical analysis, we have used the *excess* returns (in excess of the risk free rate) on a broad cross-section of risky assets to extract the most likely SDF and obtain entropy bounds for the SDF and its components. As a robustness check, we repeat the empirical exercise using as test assets the gross returns (instead of the excess returns) on the cross-section of size- and book-to-market-equity-sorted, momentum-sorted, and industry-sorted portfolios, and the return on the risk free asset.

In this case, the relevant Euler equation is

$$\mathbf{1}_N = \mathbb{E}\left[m\left(\theta, t\right)\psi_t \mathbf{R}_t\right]$$

where $\mathbf{R}_t \in \mathbb{R}^N$ is a vector of gross returns and $\mathbf{1}_N$ is an N-dimensional vector of ones. Under weak regularity conditions, the above pricing restrictions for the SDF can be rewritten as

$$\bar{\psi}^{-1}\mathbf{1}_N = \mathbb{E}^{\Psi}\left[m\left(\theta, t\right)\mathbf{R}_t\right]$$

or, as

$$\bar{M}^{-1}\mathbf{1}_N = \mathbb{E}^{Q}\left[\mathbf{R}_t\right]$$

where $\bar{x} \equiv \mathbb{E}\left[x_t\right]$, $\frac{\psi_t}{\bar{\psi}} = \frac{d\Psi}{dP}$, and $\frac{M_t}{\bar{M}} = \frac{dQ}{dP}$. Therefore, Equations (4)–(7) can be reformulated, respectively, as Equations (46)–(49) below:

$$\hat{\Psi} \equiv \arg\min_{\Psi} D\left(\Psi||P\right) \equiv \arg\min_{\Psi} \int \frac{d\Psi}{dP} \ln \frac{d\Psi}{dP} dP \ \ \text{s.t.} \ \ \bar{\psi}^{-1}\mathbf{1}_N = \mathbb{E}^{\Psi}\left[m\left(\theta,t\right)\mathbf{R}_t\right], \tag{46}$$

with its dual solution given (up to a positive scale constant) by

$$\hat{\psi}_t = \frac{e^{\lambda(\theta)'\left[m(\theta,t)\mathbf{R}_t - \bar{\psi}^{-1}\mathbf{1}_N\right]}}{\sum_{t=1}^{T} e^{\lambda(\theta)'\left[m(\theta,t)\mathbf{R}_t - \bar{\psi}^{-1}\mathbf{1}_N\right]}} = \frac{e^{\lambda(\theta)' m(\theta,t)\mathbf{R}_t}}{\sum_{t=1}^{T} e^{\lambda(\theta)' m(\theta,t)\mathbf{R}_t}}, \quad \forall t$$

where $\lambda(\theta) \in \mathbb{R}^N$ is the solution to the following unconstrained convex problem

$$\lambda(\theta) \equiv \arg\min_{\lambda} \frac{1}{T}\sum_{t=1}^{T} e^{\lambda'\left[m(\theta,t)\mathbf{R}_t - \bar{\psi}^{-1}\mathbf{1}_N\right]};$$

$$\hat{Q} \equiv \arg\min_{Q} D\left(Q||P\right) \equiv \arg\min_{Q} \int \frac{dQ}{dP} \ln \frac{dQ}{dP} dP \ \ \text{s.t.} \ \ \bar{M}^{-1}\mathbf{1}_N = \mathbb{E}^{Q}\left[\mathbf{R}_t\right], \tag{47}$$

with its dual solution given (up to a positive scale constant) by

$$\hat{M}_t = \frac{e^{\lambda'\mathbf{R}_t}}{\sum_{t=1}^{T} e^{\lambda'\mathbf{R}_t}}, \quad \forall t$$

where $\lambda \in \mathbb{R}^N$ is the solution to

$$\lambda(\theta) \equiv \arg\min_{\lambda} \frac{1}{T}\sum_{t=1}^{T} e^{\lambda'\left[\mathbf{R}_t - \bar{M}^{-1}\mathbf{1}_N\right]};$$

$$\hat{\Psi} \equiv \arg\min_{\Psi} D\left(P||\Psi\right) \equiv \arg\min_{\Psi} \int \ln \frac{dP}{d\Psi} dP \ \ \text{s.t.} \ \ \bar{\psi}^{-1}\mathbf{1}_N = \mathbb{E}^{\Psi}\left[m\left(\theta,t\right)\mathbf{R}_t\right], \tag{48}$$

with its dual solution given (up to a positive scale constant) by

$$\hat{\psi}_t = \frac{1}{T\left[1 + \lambda(\theta)'\left(m\left(\theta,t\right)\mathbf{R}_t - \bar{\psi}^{-1}\mathbf{1}_N\right)\right]}, \quad \forall t$$

where $\lambda(\theta) \in \mathbb{R}^N$ is the solution to

$$\lambda(\theta) \equiv \arg\min_{\lambda} -\sum_{t=1}^{T} \log(1 + \lambda'\left(m\left(\theta,t\right)\mathbf{R}_t - \bar{\psi}^{-1}\mathbf{1}_N\right));$$

$$\hat{Q} \equiv \arg\min_{Q} D\left(P||Q\right) \equiv \arg\min_{Q} \int \ln \frac{dP}{dQ} dP \ \ \text{s.t.} \ \ \bar{M}^{-1}\mathbf{1}_N = \mathbb{E}^{Q}\left[\mathbf{R}_t\right] \tag{49}$$

with its dual solution given (up to a positive scale constant) by

$$\hat{M}_t = \frac{1}{T\left[1 + \lambda(\theta)'\left(\mathbf{R}_t - \bar{M}^{-1}\mathbf{1}_N\right)\right]}, \quad \forall t$$

where $\lambda(\theta) \in \mathbb{R}^N$ is the solution to

$$\lambda(\theta) \equiv \arg\min_{\lambda} -\sum_{t=1}^{T} \log(1 + \lambda' \left(\mathbf{R}_t - \bar{M}^{-1}\mathbf{1}_N\right)).$$

Two observations are in order about the above results. First, looking at the dual optimizations, it is clear that different $\bar{M}$ and $\bar{\psi}$ will now matter in determining the solution, i.e., changes in the means will change the estimated SDF, and not simply as a scaling. Second, $\bar{M}$ can be calibrated easily, since from the Euler equation we have

$$\bar{M} \equiv \mathbb{E}\left[m(\theta,t)\,\psi_t\right] = \mathbb{E}\left[1/R_t^f\right],$$

and, therefore, can be estimated using a sample analogue. $\bar{\psi}$, on the other hand, can be recovered from

$$\bar{M} \equiv \mathbb{E}\left[m(\theta,t)\,\psi_t\right] = Cov\left(m(\theta,t)\,;\psi_t\right) + \bar{m}\bar{\psi},$$
$$\therefore \bar{\psi} = \frac{\bar{M} - Cov\left(m(\theta,t)\,;\psi_t\right)}{\bar{m}}.$$

Therefore, to calibrate $\bar{\psi}$, we can employ the following iterative procedure:

1. Set $\bar{\psi} = \frac{\bar{M}}{\bar{m}} = \frac{\frac{1}{T}\sum_{t=1}^{T} 1/R_t^f}{\frac{1}{T}\sum_{t}^{T} m(\theta,t)}$ as a starting guess.

2. Given $\bar{\psi}$, use the above entropy minimization procedures to estimate $\left\{\hat{\psi}_t\right\}_{t=1}^{T}$ (up to a positive constant κ).

3. Identify the scaling constant κ using the fact that the Euler equation for the risk free rate implies (as $T \to \infty$)

$$\kappa\frac{1}{T}\sum_{t=1}^{T} m(\theta,t)\,\hat{\psi}_t = \frac{1}{T}\sum_{t=1}^{T} 1/R_t^f \Rightarrow \kappa = \frac{\sum_{t=1}^{T} 1/R_t^f}{\sum_{t=1}^{T} m(\theta,t)\,\hat{\psi}_t}.$$

4. Compute an updated $\bar{\psi}$ using

$$\bar{\psi} = \frac{\bar{M} - \kappa\widehat{Cov}\left(m(\theta,t)\,;\hat{\psi}_t\right)}{\bar{m}} = \frac{\frac{1}{T}\sum_{t=1}^{T} 1/R_t^f - \kappa\widehat{Cov}\left(m(\theta,t)\,;\hat{\psi}_t\right)}{\frac{1}{T}\sum_{t}^{T} m(\theta,t)}$$

   where $\widehat{Cov}(.)$ is the sample analogue based covariance estimator.

5. With the new $\bar{\psi}$ in hand, go back to Step 2 and repeat until convergence of $\bar{\psi}$ is achieved. Once convergence is achieved, the exact estimate (no longer merely up to a constant) of ψ_t is given by $\kappa \times \hat{\psi}_t$.

Using the above approach, Table A3 repeats the analysis in Table VI when the set of assets consists of the gross returns (instead of excess returns) on the 6 size and book-to-market-equity sorted portfolios of Fama–French, 10 industry-sorted portfolios, 10 momentum-sorted portfolios, and the risk free asset. The table shows that the inclusion of the risk free rate as an additional asset in the estimation leaves the HJ, Q, M, and Ψ bounds on the SDF and its components virtually unchanged for all the asset pricing models considered.

Table A3: Bounds for RRA, Quarterly Data, 1947:Q1–2009:Q4

	HJ Bound	$Q1/Q2$ Bounds	$M1/M2$ Bounds	$\Psi1/\Psi2$ Bounds
Panel A: State Variables Extracted From Consumption Data				
CC	9	16/14	14/14	19/21
MSV	31	41/38	41/42	60/61
BY	> 100	$> 100/ > 100$	$> 100/ > 100$	$> 100/ > 100$
PST	69	93/86	112/106	86/85
Panel B: State Variables Extracted From Asset Prices				
CC	18	39/43	33/46	47/48
MSV	69	90/84	$> 100/ > 100$	$> 100/ > 100$
BY	4	5/5	5/5	5/5

Minimum values of the utility curvature parameter γ at which the model-implied SDF satisfies the HJ (column 1), Q (column 2), M (column 3), and Ψ (column 4) bounds using quarterly data 1947:Q1–2009:Q4. Columns 2–4 have two entries in each cell, which indicate whether the filtered ψ^*-component of the SDF and, therefore, the filtered SDF are estimated using Equation (6), shown on the left, or Equation (4), shown on the right. Panels A and B present results when the models' state variables are extracted from consumption data and asset market data, respectively. Other notation as in Table I.

The results in the other tables also remain largely similar upon inclusion of the risk free rate and are omitted for the sake of brevity.