Stochastic Discount Factor Bounds and Rare Events:
A Review
Maurício da Silva Medeiros Júnior

Stochastic Discount Factor Bounds and Rare Events:
A Review

Dissertação submetida a Escola de Pós-Graduação em Economia como requisito parcial para a obtenção do grau de Mestre em Economia.

Orientador: Caio Almeida

Rio de Janeiro
9 de Março de 2016
Medeiros Júnior, Maurício da Silva
Stochastic discount factor bounds and rare events: a review / Maurício da Silva Medeiros Júnior. – 2016.
48 f.

Dissertação (mestrado) - Fundação Getulio Vargas, Escola de Pós-Graduação em Economia.
Orientador: Caio Almeida.
Inclui bibliografia.


CDD – 332
MAURÍCIO DA SILVA MEDEIROS JÚNIOR

Stochastic Discount Factor Bounds and Rare Events: A Review.

Dissertação apresentada ao Curso de Mestrado em Economia da Escola de Pós-Graduação em Economia para obtenção do grau de Mestre em Economia.

Data da defesa: 22/03/2016

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

[Assinaturas]

Camille Ibsen Rodrigues de Almeida
Orientador (a)

Carlos Eugenio Ellery Lustosa Da Costa

José Valentim Machado Vicente
Abstract

We aim to provide a review of the stochastic discount factor bounds usually applied to diagnose asset pricing models. In particular, we mainly discuss the bounds used to analyze the disaster model of Barro (2006). Our attention is focused in this disaster model since the stochastic discount factor bounds that are applied to study the performance of disaster models usually consider the approach of Barro (2006). We first present the entropy bounds that provide a diagnosis of the analyzed disaster model which are the methods of Almeida and Garcia (2012, 2016); Ghosh et al. (2016). Then, we discuss how their results according to the disaster model are related to each other and also present the findings of other methodologies that are similar to these bounds but provide different evidence about the performance of the framework developed by Barro (2006).

KEYWORDS: Stochastic Discount Factors; Information-Theoretic Bounds; Implicit Utility Maximizing Weights; Rare Events; Disaster Models.
Contents

1 Introduction ................................................................. 5

2 Stochastic Discount Factor Bounds .................................... 7
   a Standard Assumptions .............................................. 14
   b Hansen and Jagannathan (1997) Distance ....................... 15
   c Entropy Bounds .................................................... 17
      c.1 Almeida and Garcia (2012) .................................. 17
      c.2 Ghosh et al. (2016) ......................................... 22
   d Comparison Between the Methodologies ......................... 26
      d.1 Comparing the Entropy Bounds and the Hansen and Jagannathan
          (1997) distance ............................................. 26
      d.2 Comparing the Entropy Bounds ............................. 30

3 Rare Events Hypothesis ................................................. 32
   a Disaster Model .................................................... 36
   b Application of the Stochastic Discount Factor Bounds to Analyze the
      Disaster Model .................................................. 37

4 Final Considerations .................................................... 42
1 Introduction

Asset pricing models capture characteristics of reality to describe how assets can be priced in a certain economy. However, these models only provide approximations of the real world which in general possess misspecification issues. Hansen and Jagannathan (1997) proposed one of the first methods to test for comparisons between misspecified asset pricing models based on a proxy of admissible stochastic discount factors provided by their asset pricing model derived from a least-square criterion. Nonetheless, this methodology has some limitations. As Almeida and Garcia (2012); Ghosh et al. (2016) note, the relevance of skewness and kurtosis on the pricing of assets is not taking into account in the Hansen and Jagannathan (1997) distance. Almeida and Garcia (2012); Ghosh et al. (2016) address this issue providing techniques that consider higher moments to test the misspecification of asset pricing models.

In this work, our purpose is providing a literature review of these stochastic discount bounds usually applied to diagnose asset pricing models. As these models have limitations, there is a wide range of bounds discussed in the literature to analyze the ability of certain models to explain financial instruments. We focus the analysis on the bounds used to diagnose disaster models (Barro, 2006). In particular, we mainly discuss the framework presented by Almeida and Garcia (2012); Ghosh et al. (2016) which are based on entropy functions to analyze dispersion of returns in order to correct the examined asset pricing model. There other bounds that represent a generalization of the Hansen and Jagannathan (1997) method (see De Araujo (2010) for a literature review on other discount factor bounds), however we aim to describe only those that are applied to analyze disaster models.

Almeida and Garcia (2012) provide a different approach based on Minimum Discrepancy projections that aim to correct asset pricing models using information about assets returns. Different from the method of Hansen and Jagannathan (1997), Almeida and Garcia (2012) methodology corrects the proxy of the stochastic discount factor to identify the admissible ones by adding a hyperbolic function of a certain linear
combination of basis assets returns. In a subsequent paper, Almeida and Garcia (2016) present evidence that their methodology is specially useful in environments that pricing kernel dispersion are caused by skewness and kurtosis. Their model allow them to distinguish these two types of models, while methods as Hansen and Jagannathan (1997) characterize them as equivalents.

The approach presented by Ghosh et al. (2016) is based on different assumptions from the one proposed by Almeida and Garcia (2012), but these methodologies have a similar spirit. In the work of Ghosh et al. (2016), they consider asset pricing models correctly specified which can be factorized into and observable component and a potentially unobservable one. As in the method proposed by Almeida and Garcia (2012), the model of Ghosh et al. (2016) does not perform the minimum adjustment using a least square sense. They proposed an approach that provides the most likely stochastic discount factor given the data using only the information about asset returns and an entropy criterion of adjustment.

There are other bounds based on entropy functions as the ones presented by Backus et al. (2011); Liu (2015). While Backus et al. (2011) use the methodology developed by Bansal and Lehmann (1997); Alvarez and Jermann (2005) to build their bounds, Liu (2015) generalizes the bound of Hansen and Jagannathan (1991) through the Holder's inequality. Both provide the diagnose of certain asset pricing models aiming to provide a better understanding of the analyzed models. Backus et al. (2011); Liu (2015) are directly related to the studies performed by Almeida and Garcia (2012, 2016) and, consequently, Ghosh et al. (2016).

The discussion presented in this paper is restricted to the rare disaster models and the application of the aforementioned bounds to diagnose this framework. Initially, we present the entropy bounds that are mainly discussed which are the approaches of Almeida and Garcia (2012) and Ghosh et al. (2016). Their similarities are also discussed and compared with the bound of Hansen and Jagannathan (1991) and the distance presented by Hansen and Jagannathan (1997). Then, we show some applications of these bounds and others that are related to them to analyze disaster
models (Barro, 2006) aiming to observe how well these models perform.

Pioneered by Rietz (1988) and revived by a growing literature (Barro, 2006; Gabaix, 2012), the rare events hypothesis emerged to rationalize the Equity Premium Puzzle (EPP) (Mehra and Prescott, 1985). Barro (2006) study if the rare events hypothesis can be used to explain the discrepancies observed between the U.S. stock market data (and also other countries). The results obtained by the representative agent model applied by him indicate that this is a possible solution to the EPP. Barro (2006) shows that this model assigns higher probabilities to bad states as the rare events hypothesis suggests to explain the EPP. In light of this methodology, our analysis is based on the application of the previously mentioned bounds to study the asset pricing model of Barro (2006) despite that fact that are some extensions of this framework (see Barro (2009); Barro and Jin (2011); Nakamura et al. (2013) for other versions of this framework).

The paper proceed as follows. Section (2) discusses the entropy bounds proposed by Almeida and Garcia (2012) and Ghosh et al. (2016) and compares the two bounds previously mentioned to observe their similarities and possible differences. Section (3) presents the rare events hypothesis and the disaster model built by Barro (2006) in order to study if the EPP can be explained by this hypothesis. In this section we also show how the entropy bounds are applied to verify the performance of asset pricing models. In particular, we discuss applications related to the disaster model of Barro (2006). Finally, Section (4) concludes this work.

2 Stochastic Discount Factor Bounds

In this chapter we present a literature review on the development of stochastic discount factor bounds that are applied to diagnose the disaster model of Barro (2006). We first present the methodology that inspired them (Hansen and Jagannathan, 1991, 1997) and the standard assumptions of the stochastic discount factor bounds that are analyzed in this paper. Then, we describe the methodology of
Almeida and Garcia (2012) and Ghosh et al. (2016). We also compare these two approaches since they have some similarities that are useful in the analysis that we do related to the disaster model of Barro (2006).

The nonexistence of arbitrage opportunities implies the existence of a pricing kernel commonly defined as the stochastic discount factor, such that equilibrium price of a traded security can be represented as the conditional expectation of the future pay-off discounted by the pricing kernel. The standard consumption-based asset pricing model, which is built in the representative agent and time-separable power utility framework, characterizes the pricing kernel as a simple parametric function of consumption growth.

Nonetheless, as Ghosh et al. (2016) present, there is an extensive literature showing that pricing kernels based on consumption growth alone cannot explain either the historically observed levels of returns, i.e., the EPP and Risk Free Rate Puzzle (RFRP) (Mehra and Prescott, 1985; Weil, 1989), or the cross-sectional dispersion of returns between different classes of financial assets (Hansen and Singleton, 1983; Mankiw and Shapiro, 1986; Breeden et al., 1989; Campbell, 1996). It is important to note, nevertheless, that a considerable number of papers present evidence that consumption risk does matter for explaining asset returns (Lettau and Ludvigson, 2001a,b; Parker and Julliard, 2005; Hansen et al., 2008; Savov, 2011).

Therefore, there is a burgeoning literature that focuses in analyzing asset pricing models based on consumption growth and observe what kind of information it is possible to be extracted from these methodologies. The first tentative to perform this analysis is Hansen and Jagannathan (1991) who describe a region of admissible stochastic discount factors of a certain asset pricing models. This framework develop a admissible bound for stochastic discount factors based only on the second moment of the stochastic discount factor distribution.

Hansen and Jagannathan (1997) pursue a better understanding of this analysis of asset pricing models developing a method to test for comparisons of possible mis-
specified asset pricing models. This approach is based on a least-square projection of an asset pricing proxy on a family of admissible stochastic discount factors to provide a diagnosis of asset pricing models, analyzing problems of model selection and specification in the examined frameworks. There are several empirical works that use this methodology to analyze their asset pricing models such as Hodrick and Zhang (2001), Kan and Robotti (2009) and Wang and Zhang (2012).

The Hansen and Jagannathan (1997) methodology is frequently applied in the literature, however there are some important limitations. The role that skewness and kurtosis play on the pricing of assets is crucial as observed by Rubinstein (1973), Alan Kraus (1976), Barone-Adesi (1985), Harvey and Siddique (2000), Dittmar (2002) and Vanden (2006), among others. Nonetheless, the framework of Hansen and Jagannathan (1997) does not account for it (Almeida and Garcia, 2012, 2016). Another issue regarded to this approach is that it can not differentiate among asset pricing models well when the set of basis assets is composed by the 25 Fama and French (1993) portfolios (Lewellen et al., 2010; Li et al., 2010).

Due to the importance of skewness and kurtosis on the pricing of assets, there is a wide class of asset pricing models that account for higher order moments and/or nonlinearities. This brings a need for tools that are able to analyze and to evaluate those models. As the Hansen and Jagannathan (1997) approach only considers two moments of returns, the absence of higher moments in this methodology is an usual criticism to this framework. Chabi-Yo (1989) is a work that indicates a concern about this issue related to the Hansen and Jagannathan (1997) method.

Lewellen et al. (2010) and Li et al. (2010) observe that the Hansen and Jagannathan (1997) model has a bad performance when it has to compare asset pricing models which the set of basis assets is composed by the 25 Fama and French (1993) portfolios. In this context, Gospodinov et al. (2016) also observe this pattern applying other statistical tests with a positive constraint that are not considered by Li et al. (2010).
Give the necessity for metrics that generalize the Hansen and Jagannathan (1997) methodology, Almeida and Garcia (2012) provide a metrics that measures the degree of misspecification of asset pricing models taking into account moments of higher order. They consider a family of convex functions to compute the distance between an asset pricing proxy $y$ and the family $M$ of admissible stochastic discount factors that prices a set of predetermined primitive securities.

They build this problem using a Minimum Discrepancy framework where the goal is to obtain an admissible stochastic discount factor that prices the basis assets, i.e., in the discrepancy sense, is the closest possible to $y$. They use the duality theory (Kitamura, 2006) to estimate the stochastic discount factor and its distance to $y$ through a finite-dimensional problem. Almeida and Garcia (2012) specialize the convex functions to the Cressie-Read family of discrepancy functions whose the dual problems have the interpretation of HARA portfolio problems. It provides a framework that can accounts for an infinite number of moments returns and captures the Hansen and Jagannathan (1997) and Chabi-Yo (1989) methodologies.

The first implication of Almeida and Garcia (2012) methodology is allowing the researcher to correct the asset pricing model proxy to become an admissible stochastic discount factor by adding a nonlinear function of a linear combination of basis asset returns. They show that these nonlinear corrections are a excellent tool to detect a class of model misspecification not identified by Hansen and Jagannathan (1997) method. Another advantage of this technology is providing a manner to analyze economies where nonlinear risk is priced. In this scenario, the Hansen and Jagannathan (1997) approach is worse than Almeida and Garcia (2012) method.

This methodology is also related to a literature that started to pursue methods that estimate nonlinear stochastic discount factors. Bansal and Viswanathan (1993), for instance, use a neural network methodology to build a nonlinear stochastic discount factor that attends the specifications by Alan Kraus (1976) and Glosten and Jagannathan (1994). Dittmar (2002) also studies nonlinear pricing kernels and observes that, under decreasing absolute risk version, a cubic pricing kernel is able to
best describe a cross-section of industry portfolios. As stated by Almeida and Garcia (2016), the methodology of Almeida and Garcia (2012) also embeds cubic nonlinearities implicitly which is another argument showing the generalization provided by Almeida and Garcia (2012) approach.

Almeida and Garcia (2016) apply the methodology proposed by Almeida and Garcia (2012) which generalizes variance (Hansen and Jagannathan, 1997), entropy (Backus et al., 2011) and higher-moment (Snow, 1991) bounds. In this work they show how valuable their methodology is to diagnose models with nonlinearities in the pricing kernels and non-Gaussian features in the returns. Almeida and Garcia (2016) verify the admissibility of disaster and long-run risk models. In Disaster models (Barro, 2006) the price kernel dispersion is caused mainly from asymmetric negative jumps on consumption growth, while Long Run Risk models have kurtosis as an important of dispersion. Applying the methodology of Almeida and Garcia (2012), Almeida and Garcia (2016) obtain great insights about the pricing kernel of these models.

The framework of Almeida and Garcia (2012) is also applied by Almeida and Garcia (2016) to compute information bounds for different sets of portfolios. They aim to analyze the suggestion proposed by Lewellen et al. (2010) which is adding alternative portfolios like industry ones to Fama and French (1993) portfolios in order to break its strong three-factor structure in tests of asset pricing models. Thus, Almeida and Garcia (2016) verify that for different discrepancies in their family, industry portfolios bring non-redundant information with respect to Fama and French (1993) portfolios under different ranges of stochastic discount factors means.¹

Almeida and Garcia (2016) also use their methodology to show the application of their nonlinear stochastic discount factors to evaluate the performance of some simple option-based strategies. They observe that the return distributions of these strategies are usually skewed and fat-tailed, what brings the need for a discounting that goes beyond the first two moments. Applying their discrepancy functions, they

¹Given a fixed minimum distance bound and a pair of sets of primitive securities, Almeida and Garcia (2016) define non-redundant information of one set in relation other as the first set produces a higher bound that the second.
capture the sensitivity of performance of these higher moments. Almeida and Garcia (2016) also compare these results to the evidence obtained from linear stochastic discount factors.

The building of nonparametric stochastic discount factors bounds is an issue addressed by other works as well. Stutzer (1995), for instance, suggests a nonparametric bound to diagnose asset pricing models based on estimators that attend the Kullback-Leibler Information Criterion (KLIC). Bansal and Lehmann (1997) also propose an entropic bound based on a maximization of growth portfolio. As discussed by Almeida and Garcia (2016), Backus et al. (2011) apply the method of Bansal and Lehmann (1997) to measure entropy in tests of disaster-based models. Recently, Ghosh et al. (2016) propose a nonparametric entropy bound of stochastic discount factors that can be factorized into an observable component and an unobservable.

Even though pricing kernels based only on consumption risk have given rise the EPP and the RFRP, there is considerable empirical evidence that corroborates the importance of consumption risk to explain asset returns. To provide a better understanding of this relation, there is a literature mainly based on modifying the preferences of investors and/or the structure of the economy. These models have pricing kernels that can be factorized into an observable and unobservable, model-specific, component as Ghosh et al. (2016) observe.

Some examples that illustrate these models are the external habit model (Campbell and Cochrane, 1999; Menzly et al., 2004), the long run risk models (Bansal and Yaron, 2004) and models with housing risk (Piazzesi et al., 2007). In the external habit model, the the model-specific component is a function of the habit level, the long run risk models have as possibly unobservable component the return of total wealth and the housing risk models have as additional component the growth in the expenditure share on non-housing consumption (Ghosh et al., 2016).

In these frameworks, Ghosh et al. (2016) provide a methodology to analyze dynamic asset pricing models. Through their approach it is possible to estimate non-
parametrically the time series of the unobserved pricing kernel under a set pricing restrictions from the data, Ghosh et al. (2016) also build entropy bounds to assess the empirical plausibility of the stochastic discount factors which are candidates and they estimate the minimum (in the information sense) adjustment of the stochastic discount factor required to correctly price asset returns. This method provides a useful way to diagnose certain asset pricing models through an analysis of models that might fail empirically and also allows them to characterize some properties that a successful model must satisfy.

Ghosh et al. (2016) show that it is possible to extract the time series of both the stochastic discount factor and the unobservable component through an approach that is equivalent to maximising the expected risk neutral likelihood under a set of no arbitrage restrictions. Their time series for the unobservable pricing kernel is substantially correlated to the Fama and French (1993) factors for a variety of sample frequencies and assets used in the estimation procedure. This result suggests that their model successfully identifies the pricing kernel.

Another useful application of their model is to use the entropy bounds to construct bounds for the potentially unobserved component of the pricing kernel. An interesting finding that they obtain is that their bounds is the tightest ones when the pricing kernel is a function only of observable variables, and it is satisfied if and only if the model is actually able to correctly price assets. In this case their bounds are closely related to the Hansen and Jagannathan (1997) distance which identify the minimum variance linear adjustment.

Ghosh et al. (2016) also describe the relevance of their model to analyze and to diagnose the most well known consumption-based asset pricing models. Studying the standard time separable power utility model, Ghosh et al. (2016) show that the pricing kernel satisfies the Hansen and Jagannathan (1991) bound for large values of the risk aversion coefficient applying their methodology. They consider more general models of dynamic economies such as models of habit formation, long run risks and complementarities in consumption. In these applications they observe that the
estimated stochastic discount factor implied by the models require implausibly high levels of risk aversion to satisfy their entropy bounds, usually all understate the market crash risk and all are poorly correlated with the filtered stochastic discount factor.²

The model proposed by Ghosh et al. (2016) have five main advantages. First, it is a model that can be used to extract information not only from options, but also from any type of financial asset. Second, the estimated pricing kernel from their model explore information of financial data and aggregate consumption. Third, the relative entropy extraction of the stochastic discount factor is akin to a non-parametric maximum likelihood procedure and provides an estimate of its time series. Fourth, this methodology does not consider only the second moment, but also all higher moments. Fifth, the authors argue that it is a method with considerable generality despite the restriction imposed in the stochastic discount factor, which must have a observable and a possibly unobservable components.

a Standard Assumptions

We consider an economy similar to the one presented by Almeida and Garcia (2012); Ghosh et al. (2016). Thus, suppose that assets are purchased at a certain period $t$ and the payoffs are realized at period $T > t$. Define $\Gamma_T$ as the sigma-algebra that represents the information at time $T$ and $L^2$ the space of all integrable random variable that are measurable with respect to $\Gamma_T$. Consider a set of $n$ primitive securities whose payoffs are represented by $x \in \mathbb{R}^n$. Define as a payoff any function $p \in L^2$ obtained as a element of:

$$P = \{x \cdot c : c \in \mathbb{R}^n\},$$

such that $x \in L^2$ and $E[xx']$ a nonsingular second moment matrix. Then, we can guarantee that the Law of One Price is satisfied for every $p \in P$ and the pricing func-

²The stochastic discount factors obtained non-parametrically are called the filtered stochastic discount factor by Ghosh et al. (2016). They also obtain its components for the risk aversion coefficient equalized to 10.
tional \( \pi \) is continuous and linear on \( P \) which can be written as:

\[
\pi(p) = \pi(c \cdot x) = c \cdot q,
\]

where \( c \in \mathbb{R}^n \) is the only portfolio that replicates the payoff \( p \).

An admissible stochastic discount factor will be a random variable \( m \in L^2 \) that correctly prices asset payoffs \( p \in P \) such that:

\[
\pi(p) = \mathbb{E}[m \cdot p]. \tag{1}
\]

The Riesz representation theorem guarantees the existence of at least one stochastic discount factor. Under no arbitrage hypothesis, to guarantee that the variable \( m \in L^2 \) satisfies Equation (1) it is sufficient to have that:

\[
q = \mathbb{E}[m \cdot x]. \tag{2}
\]

For the empirical application, we consider that Equation (2) is valid to each period \( t \) for the sequence \( \{(m_t, x_t, q_t)\} \) (Hansen and Jagannathan, 1997).

b  Hansen and Jagannathan (1997) Distance

An asset pricing model \( y \) will approximate an admissible stochastic discount factor described as \( \pi_y(p) = \mathbb{E}[y \cdot p] \) with a pricing error given by \( \pi(p) - \pi_y(p) \). Therefore, we characterize an asset pricing model as misspecified if there is not a parameter \( \theta \in \Theta \) such that:

\[
q = \mathbb{E}[y(\theta) \cdot x]. \tag{3}
\]

Given a proxy asset pricing model \( y(\theta) \), parameterized by a vector of parameters \( \theta \in \mathbb{R}^k \), Hansen and Jagannathan (1997) measure the degree of misspecification considering a least-squares criterion. They use a least-squares projection of \( y(\theta) \) into the space of all admissible stochastic discount factors defined as \( M \). Thus, their mis-
specification criterion is given by:

$$\delta_{HJ}(\theta)^2 = \min_{m \in M} \|m - y(\theta)\| = \min_{m \in M} \mathbb{E}\{(m - y(\theta))^2\}. \quad (4)$$

As $m \in M$ can be interpreted as $m \in L^2$ satisfying Equation (1) for the specific set of primitive securities, Equation (4) can be described as:

$$\delta_{HJ}(\theta)^2 = \min_{m \in L^2} \mathbb{E}\{(m - y(\theta))^2\} \text{ s.t. } q = \pi(x) = \mathbb{E}[m \cdot x]. \quad (5)$$

Using the Lagrange multipliers the problem becomes:

$$\delta_{HJ}(\theta)^2 = \min_{m \in L^2} \sup_{\lambda \in \mathbb{R}^n} \mathbb{E}\{(m - y(\theta))^2 - 2\lambda'(m \pi - q)\}. \quad (6)$$

Fixing the Lagrange multiplier and solving the minimization problem on $m$, Hansen and Jagannathan (1997) obtained the following dual optimization problem:

$$\delta_{HJ}(\theta)^2 = \max_{\lambda \in \mathbb{R}^n} \mathbb{E}\{y^2 - (y - \lambda'x)^2 - 2\lambda'q\}. \quad (7)$$

The solution of the optimization problem described by Equation (8) is given by:

$$\delta_{HJ}(\theta)^2 = \mathbb{E}\{(\mathbb{E}\{xy\} - q)'\mathbb{E}\{xx'\}^{-1}(\mathbb{E}\{xy\} - q)\}^{1/2}. \quad (8)$$

To estimate the parameter vector $\theta$, Hansen and Jagannathan (1997) suggest to estimate it by solving the following problem:

$$\arg \min_{\theta \in \mathbb{R}^k} \delta_{HJ}(\theta) \quad (9)$$

as an alternative to the GMM estimator (Hansen, 1982) which would be obtained from the following equation:
arg \min_{\theta \in \mathbb{R}^k} g(\theta)' \cdot W \cdot g(\theta),

(10)

where \( g(\theta) = \mathbb{E}\{y(\theta) \cdot x\} \) represents the moment conditions, and \( W \) is an \( n \) symmetric positive definite matrix that might depend on the sample observations.

It is important to observe that the GMM estimator and the Hansen and Jagannathan (1997) estimator are special cases of minimum distance estimators with a quadratic norm. The difference between them is that while the weight matrix \( W \) presented in the GMM estimator will depend on the particular asset pricing model \( y \) adopted as proxy, the Hansen and Jagannathan (1997) estimator has a weight matrix invariant to the asset pricing proxy. To observe the importance of this difference, consider the case which the GMM criterion is used to select among possibly misspecified models. Since \( g(\theta) \) is the pricing error, the GMM metric will weight pricing errors differently across models. If the Hansen and Jagannathan (1997) estimator is applied, the pricing errors weights will not change with the asset pricing model chosen. As Hansen and Jagannathan (1997) show, this is a crucial advantage of their methodology in relation to GMM estimator. It allows the researcher compare different possible misspecified asset pricing models.

c Entropy Bounds
c.1 Almeida and Garcia (2012)

Hansen and Jagannathan (1997) propose a measure of misspecification which is mainly related to the second moment of payoffs that compose the basis asset returns of the economy analyzed. Aiming to build a measure that assigns different set of weights on higher moments of stochastic discount factors, Almeida and Garcia (2012) propose a metric of misspecification that is robust to the choice of the payoffs moments. Almeida and Garcia (2012) consider the Cressie-Read family of discrepancy functions to generalize the entropy estimators used to access how a model is misspecified, i.e., to define a criterion of distance between \( y(\theta) \) and the set \( M \) of
stochastic discount factors. An element of the Cressie-Read family is described as:

\[ \phi(m) = \frac{m^{\gamma+1} - 1}{\gamma(\gamma + 1)} \]  

(11)

where, for \( \gamma = -1 \) or \( \gamma = 0 \), \( \phi(m) = -\ln(m) \) and \( \phi(m) = n \ln(m) \) respectively (representing, respectively, the Empirical Likelihood estimator (EL) and the Exponential Tilting estimator (ET)), since these are the limits of the function when \( \phi(\cdot) \) converges to \(-1\) and \(0\).

Fixing the parameter \( \gamma \), the misspecification of the model \( y(\theta) \) will be defined as:

\[ \delta_\gamma(\theta) = \min_{m \in M} E\{\phi(1 + m - y(\theta))\} \]  

(12)

Observing Equation (11) it is possible to observe that the population moment determined by \( \gamma \) will be considered in the misspecification measure of Almeida and Garcia (2012). Defining \( \gamma = 1 \), the measure based on Equation (11) is a generalization of the Hansen and Jagannathan (1997) distance. Almeida and Garcia (2012) show that accounting for different moments different from the second might be useful. Cases where \( y(\theta) \) depends on the basis assets of the economy non-linearly or when the asset have non-gaussian payoffs represent situations that considering different population moments can bring relevant information to identify the misspecification of a model.

As Almeida and Garcia (2012) observe, using the Cressie-Read family is useful since it allows a dual formulation that provides a simple solution to the problem described by Equation (12). Due to this dual formulation, as in the Hansen and Jagannathan (1997) distance solution, the variable \( m \in L^2 \) that solves Equation (12) can obtained from a n-dimensional vector of Lagrange multipliers (Almeida and Garcia, 2012). Therefore, consider the convex conjugate of \( \phi(\cdot) \) called \( \phi^*(\cdot) \) such that:

\[ \phi^*(b) = \sup_{m \in L^2} m' b - \phi(1 - m - y(\theta)). \]
As Borwein and Lewis (1991) show, the dual program of Equation (12) is given by \[
\max_{\lambda'q} \lambda'q - E[\phi^*(\lambda'x)].
\]
Through this finding of Borwein and Lewis (1991), Almeida and Garcia (2012) compute \(\delta\) as:

\[
\delta_i(\theta) = \max_{\lambda'q} \lambda'q - \mathbb{E} \left\{ \frac{(\gamma \lambda'x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(\theta) - 1)\lambda'x + \frac{1}{\gamma(\gamma+1)} \right\}.
\] (13)

Applying misspecified models to price assets have its difficulties since these frameworks are associated with pricing errors. Therefore, aiming to use a misspecified asset pricing model to this purpose, Almeida and Garcia (2012) offer an interesting alternative. First fix a measure of misspecification and then find the parameter \(\theta\) that minimize it. Almeida and Garcia (2012), therefore, obtain \(\theta\) using the following framework:

\[
\theta^* = \arg\min_{\theta} \delta(\theta)
\]

It is important to observe that in correctly specified models \(\theta\) is an element of the parametric space \((\Theta)\) that correctly price the asset which is invariant with the choice of the measure used to define if the model is misspecified or not. Nonetheless, if the model is misspecified, then \(\theta^*\) will depend on the choice of the measure.

To estimate \(\theta\), Almeida and Garcia (2012) suppose the existence of a time series of returns \(\{R_t\}_{t=1}^T\), which by definition the price is 1, and a time series of explicative variables of \(\{y_i(\theta)\}_{t=1}^T\). Thus, Almeida and Garcia (2012) estimate \(\theta^*\) associated to a measure of misspecification using the following equation:

\[
\hat{\theta} = \arg\min_{\theta \in \Theta} \lambda'1_n - \sum_{t=1}^T \frac{1}{T} \{\phi^*(\lambda'R_t)\}.
\] (14)

The asymptotic properties of this estimator are obtained by Almeida and Garcia (2012) considering that the models are misspecified. This methodology provides an adjustment summing nonlinear functions of the basis asset returns in the objective function when compared to the previously methods build to measure misspecifici-
cation of asset pricing models. Almeida and Garcia (2016) use this methodology to build bounds that are used to analyze the performance of asset pricing models. Their stochastic discount factor bounds is defined as:

$$\arg \min_{m} \frac{1}{T} \sum_{t=1}^{T} \phi(m_t)$$

$$\frac{1}{T} \sum_{t=1}^{T} m_t (R_t - R_f) = 0$$

$$\frac{1}{T} \sum_{t=1}^{T} m_t = 1/R_f$$

$$m_t > 0 \ \forall t$$

Almeida and Garcia (2012) show an equivalence between this framework and other that corrects the model through a multiplicative adjustment. In this case the adjustment over an asset pricing model \(y(\theta)\) can be characterized by a random variable \(m\) such that \(\mathbb{E}\{m\} = 1\) and:

$$\mathbb{E}\{my(\theta)x\} = q.$$

From this equation, considering that \(m\) can be interpreted as a Radon-Nikodym derivative if the adjust is absolutely continuous, Equation (16) can be written as (Ghosh et al. (2016) also do it)

$$\mathbb{E}^m[y(\theta)x] = q.$$ (16)

Once again Almeida and Garcia (2012) apply the Cressie-Read family of discrepancy functions to measure misspecification. They define the size of the multiplicative adjustment as \(\mathbb{E}\{\phi(m)\}\), where \(\phi(.)\) is a multiplicative function and convex (then, \(\mathbb{E}\{\phi(m)\} \geq \phi(\mathbb{E}\{m\}) = 0\)), and \(\delta(\theta)\) is the measure of misspecification which is defined as:
\[ \delta(\theta) = \min_{m \geq 0} \mathbb{E}[\phi(m)] \] (17)
\[ \mathbb{E}[m y(\theta) x] = q \]
\[ \mathbb{E}[m] = 1 \] (18)

Therefore, \( \delta(\theta) \) is the least adjustment required to correct the model. If the model is correctly specified, then \( m = 1 \) is feasible and the distance of \( y(\theta) \) is zero in the true value of \( \theta \). If the model is misspecified, the distance will be positive and will consider the population moment defined by \( \gamma \) in the Cressie-Read function optimized. This optimization problem is solved as:

\[ \delta(\theta) = \min_{m \geq 0} \frac{\sum_{t=1}^{T} [\phi(m_t)]}{\sum_{t=1}^{T} [m_t y(t, \theta) x_t]} = q \]
\[ \sum_{t=1}^{T} m_t = 1 \] (19)

In this model, the purpose is to identify the estimative \( \theta^* \) which is obtained as \( \theta^* = \arg \min_{\theta \in \Theta} \delta(\theta) \) and from this optimization problem it is found that:\(^3\)

\[ m^* = \frac{\left( \gamma \left( \frac{\alpha^*}{m^*} \right)^\gamma (y(\theta^*) x - q) + 1 \right)^{1/\gamma}}{\mathbb{E} \left[ \left( \gamma \left( \frac{\alpha^*}{m^*} \right)^\gamma (y(\theta^*) x - q) + 1 \right)^{1/\gamma} \right]} \] (20)

where \( \alpha^* \in \mathbb{R}^n \) (which is a function of \( \theta^* \)) is the solution to the dual problem of the entropy minimization problem in Equation (19) as Kitamura (2006) show. This problem is given by:

\(^3\)Almeida and Garcia (2012) also derive the asymptotic properties of the estimator of \( \theta \) considering the model misspecified in the multiplicative adjustment case.
\[
\alpha^* = \arg \min_{\alpha \in \mathbb{A}} - \sum_{t=1}^{n} \log(1 + \alpha'(m(t, \theta)x_t))
\]  

Almeida and Garcia (2016) use this methodology to build entropy bounds aiming to diagnose and to analyze asset pricing models. They use this methodology to generalize entropy bounds and to show how parameters affect pricing kernel dispersion in asset pricing models. These methodology is useful to distinguish between models where dispersion comes mainly from skewness and kurtosis in the distribution of the stochastic discount factor. Almeida and Garcia (2016) provide bounds that impose a data-driven balance between the amount of skewness and kurtosis that any admissible pricing kernel should satisfy.

c.2 Ghosh et al. (2016)

Ghosh et al. (2016) present a similar methodology in comparison to the one proposed by Almeida and Garcia (2012). To build their methodology, it is necessary to suppose that the model is correctly specified. Ghosh et al. (2016) also consider in their method only the models that the stochastic discount factor can be factorized as an observable part and other part that can be unobservable. In this scenario they will be able to describe their methodology which imposes more restrictions than Hansen and Jagannathan (1997) bound, however allows them to obtain bounds for the stochastic discount factors with higher information content than the commonly used ones.

Define \( M \) as the stochastic discount factor such as \( M = z(\theta) \cdot \psi \), where \( z(\theta) \) follows the same properties of \( y(\theta) \) presented in the methodology of Almeida and Garcia (2012) (in other words, this variable it is the observable part) and \( \psi \) it is the potentially unobservable component.\(^4\) Thus, for any set of tradable asset, the following restriction based on Euler equations must hold in equilibrium:

\(^4\)Ghosh et al. (2016) consider \( z(\theta) \) strictly positive.
\[ 0 = \mathbb{E}[z(\theta)\psi R^e] = \mathbb{E}^\psi[z(\theta)R^e] \]  

where \( R^e \in \mathbb{R}^n \) is a vector of excess returns on different tradable assets.

Then, given a set of consumption and asset returns data, for any \( \theta \), it is possible to estimate the probability measure \( \psi \) as follows:

\[ \psi^* \equiv \arg \min_{\psi} D(\psi \| P) \equiv \arg \min_{\psi} \int \frac{dP}{d\psi} d\psi \text{ s.t. } 0 = \mathbb{E}^\psi[z(\theta)R^e] \]  

or

\[ \psi^* \equiv \arg \min_{\psi} D(P \| \psi) \equiv \arg \min_{\psi} \int \frac{d\psi}{dP} \ln \frac{d\psi}{dP} d\psi \text{ s.t. } 0 = \mathbb{E}^\psi[z(\theta)R^e] \]  

where the above equations are the relative entropy (or Kullback-Leibler Information Criterion (KLIC)) minimization under the restrictions derived from the Euler equations.

It is important to note that the above equations are representations of both the EL estimator and the ET estimator respectively. Solving the problem of Equations (24) and (25), Ghosh et al. (2016) characterize an element of the sequence \( \{\psi_t\}_{t=1}^T \) as:

a) From Equation (24):

\[ \psi_t^* = \frac{1}{T (1 + \lambda(\theta)'z(\theta, t)R^e_t)} \]  

where \( \lambda(\theta) \in \mathbb{R}^n \) is the solution to the dual problem of the entropy minimization problem in Equation (24) which is analogous to Equation (22).

b) From Equation (25):

\[ \psi_t^* = \frac{\exp \{\lambda(\theta)z(\theta, t)R^e_t\}}{\sum_{t=1}^{T} \exp \{\lambda(\theta)z(\theta, t)R^e_t\}} \]  

where \( \lambda(\theta) \in \mathbb{R}^n \) is the solution to the dual problem of the entropy minimization problem in Equation (24) which is analogous to Equation (22).
where $\lambda(\theta) \in \mathbb{R}^n$ is the solution to the dual problem of the entropy minimization problem in Equation (25) which is analogous to Equation (22).

This approach also allows to recover the risk neutral probability measure ($Q$) from the data as (Stutzer, 1995):

$$Q^* \equiv \arg \min_Q D(P\|Q) \equiv \arg \min_Q \int \ln \frac{dP}{dQ} dP \text{ s.t. } 0 = \int R_i^t dQ \equiv \mathbb{E}^Q [R_i^t] \quad (28)$$

and

$$Q^* \equiv \arg \min_Q D(Q\|P) \equiv \arg \min_Q \int \frac{dQ}{dP} \ln \frac{dQ}{dP} dP \text{ s.t. } 0 = \mathbb{E}^Q [R_i^t] \quad (29)$$

where $Q$ and $P$ are absolutely continuous.5

Based on the relative entropy estimation of the stochastic discount factor and its component $\phi$ described previously, Ghosh et al. (2016) derive a set of entropy bounds for the stochastic discount factor, $M$ and its components. Their bounds are defined as:

**Definition 1 (Q-bounds)** Ghosh et al. (2016) define the following risk neutral probability bounds for any candidate stochastic discount factor $M_t$ as:

1. **Q1-bound:**

   $$D \left( P \left\| \frac{M_t}{M} \right. \right) \equiv \int - \ln \frac{M_t}{M} dP \geq D \left( P \left\| Q^* \right. \right) \quad (30)$$

   where $Q^*$ solves Equation (28).

2. **Q2-bound:** (Stutzer, 1995)

---

5Observe that Equation (28) is the EL estimation procedure to obtain the risk neutral probability measure and the Equation (29) represents the ET estimation framework.
where $Q^*$ solves Equation (29).

In this definition Ghosh et al. (2016) consider that $M_t = z(\theta, t) \cdot \phi_t$ where $z(\theta, t)$ is a known non negative function of observable variables and the parameter vector $\theta$, and $\phi_t$ is a potentially unobservable component.

**Definition 2 (M-bounds)** For any candidate stochastic discount factor of the form $M_t = z(\theta, t) \cdot \phi_t$, and given any choice of the parameter vector $\theta$, Ghosh et al. (2016) define the following bounds:

1. **$M_1$-bound:**

$$ D \left( P \| \frac{M}{M} \right) = \int - \ln \frac{M}{M_t} dP \geq D \left( P \| z(\theta, t) \phi_t^* \right) = \int - \ln \frac{z(\theta, t) \phi_t^*}{z(\theta, t) \phi_t^*} dP $$

where $\phi_t^*$ solves Equation (32) and $z(\theta, t) \phi_t^* \equiv \mathbb{E}[z(\theta, t) \phi_t^*]$.

2. **$M_2$-bound:**

$$ D \left( \frac{M_t}{M} \| P \right) = \int \frac{M_t}{M} \ln \frac{M_t}{M} dP \geq D \left( \frac{z(\theta, t) \phi_t^*}{z(\theta, t) \phi_t^*} \| P \right) = \int \frac{z(\theta, t) \phi_t^*}{z(\theta, t) \phi_t^*} \ln \frac{z(\theta, t) \phi_t^*}{z(\theta, t) \phi_t^*} dP $$

where $\phi_t^*$ solves Equation (33).
d Comparison Between the Methodologies

d.1 Comparing the Entropy Bounds and the Hansen and Jagannathan (1997) distance

As mentioned in the Subsection (b), the methodology of Hansen and Jagannathan (1997) to assess misspecification in an asset pricing model is based on a least-square criterion observing only the second moment between all the elements of the stochastic discount factor set and a proxy of asset pricing model \(y(\theta)\) as described by Equation (34). Therefore, in this framework occurs:

\[
d_{HJ}(\theta)^2 = \min_{m \in \mathcal{L}} \mathbb{E}\{(m - y(\theta))^2\} \quad s.t. \quad q(x) = \mathbb{E}[m \cdot x].
\]

Observing this equation and comparing to the results of Almeida and Garcia (2012) we can observe that indeed the Hansen and Jagannathan (1997) distance is obtained from the Cressie-Read family. Fixing \(\gamma = 1\), the measure based on Equation (11) (which is the function that described the Cressie-Read family of discrepancy functions) is a generalization of the Hansen and Jagannathan (1997) distance. The difference between Almeida and Garcia (2012) and Hansen and Jagannathan (1997) is that the first accounts for higher moments changing the value of \(\gamma\). That is why Almeida and Garcia (2012) provide a general version of Hansen and Jagannathan (1997) distance.

Expanding the method provided by Almeida and Garcia (2012), Almeida and Garcia (2016) measure the importance of each moment of the pricing kernel to the overall model discrepancy. Each member of the Cressie Read family will give different weights to the moments of a pricing kernel, thus they argue that it might help to distinguish between different asset pricing models. Almeida and Garcia (2016) use a Taylor expansion to observe it based on the work of Backus et al. (2011) which verify the relevance of analyzing the entropy of a pricing kernel through the cumulant-generating function.

Through the Taylor expansion applied in the expected value of \(\phi(m) = \frac{m^{\gamma+1} - a^{\gamma+1}}{(\gamma\gamma+1)}\)
around the stochastic discount factor mean $a$, Almeida and Garcia (2016) identify the role played by each moment. They consider $\phi(a) = 0$, $\phi'(m) = \frac{m^\gamma}{\gamma}$, $\phi''(m) = m^{\gamma - 1}$, $\phi'''(m) = (\gamma - 1)m^{\gamma - 2}$, ..., and Taylor expanding $\phi$ and taking expectations on both sides they obtain:

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

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

Therefore, the weights that Almeida and Garcia (2016) obtain to skewness and kurtosis are respectively $\frac{(\gamma - 1)a^{\gamma-2}}{3!}$ and $\frac{(\gamma - 1)(\gamma - 2)a^{\gamma-3}}{4!}$. These results imply that for $\gamma$ close to one, the discrepancy functions do not capture much of the higher moment activity of pricing kernels. Once the value of $\gamma$ become more negative the skewness and kurtosis receive considerable weights in the expansion.

Another important finding related to Equation (34) is the relative weights that are assigned to skewness and kurtosis by different Cressie Read functions. Note that for values of $\gamma$ such that $-2 < \gamma < 1$ the weight given to kurtosis is smaller than the corresponding weight given to skewness. Now, observing values of $\gamma$ that are less than $-2$, the discrepancy function assigns more weight to kurtosis rather than to skewness. As Almeida and Garcia (2016) observe, even higher-moments receive more absolute weight than their corresponding odd higher-moments in this region of $\gamma$.

Ghosh et al. (2016) also have some similarities to the work of Hansen and Jagannathan (1997). Comparing the constraint of Equation (34) it is possible to observe that this is the same constraint used to build Ghosh et al. (2016) risk neutral entropy bounds (defined by $D(P\|Q)$). However, the nature of the minimization is different since in the methodology of Ghosh et al. (2016) is obtained the most likely pricing kernel, while the Hansen and Jagannathan (1997) approach is found an $m$ that minimizes the second moment of the deviation from the candidate stochastic discount factor.
Therefore, there is an interesting conceptual difference between the Hansen and Jagannathan (1997) distance and the risk neutral $D^2(P||Q)$ bound. While the Hansen and Jagannathan (1997) obtain the minimum adjustment that makes $\gamma(\theta) - X^R\gamma$ an admissible stochastic discount factor based on a least square metric, Ghosh et al. (2016) identifies the most likely stochastic discount factor given the data and offers it as a benchmark to which a candidate stochastic discount factor $\gamma(\theta)$ can be compared.

As Ghosh et al. (2016) observe, their entropy bounds focus on the space of distribution functions. Due to the one to one mapping relation between distributions and moments generating functions, their proposed bounds carry information that is related not only about second moments, but also about all the other moments of the stochastic discount factors. As skewness, kurtosis, tail probabilities etc. are relevant for asset pricing, Ghosh et al. (2016) argue that this a crucial advantage of their models.

Ghosh et al. (2016) consider the risk neutral entropy Q1-bound which, given a candidate stochastic discount factor $M$, this defines the entropy distance:

$$d_{Q1} = D\left(P||\hat{Q}\right) - D\left(P||M\right)$$

where $\hat{Q}$ solves Equation and Ghosh et al. (2016) have normalized $M$ to have unit mean to simplify the exposition.

As Ghosh et al. (2016) consider, this distance must be non-positive for the candidate stochastic discount factor $M$ to satisfy the Q1-bound. Defining $\hat{q}$ the minimum entropy stochastic discount factor that the bound identifies, the above distance can be rewritten as:

$$d_{Q1} = \ln E\left[\exp\{\ln \hat{Q}\}\right] - \ln E\left[\exp\{\ln M\}\right] + \int \ln MdP - \int \ln \hat{q}dP.$$
where they consider by construction $E[M_l] = E[q] = 1$. Ghosh et al. (2016) also observe that the relative entropy minimization identifies an admissible stochastic discount factor in the Hansen and Jagannathan (1997) sense.

Applying the cumulant expansion approach of Backus et al. (2011), Ghosh et al. (2016) show the connection between their bounds and the second Hansen and Jagannathan (1997) distance. To present it, consider the cumulant generating function of a random variable $\ln x_t$ as:

$$k^x(s) = \ln E\{\exp\{s \ln x_t\}\}$$

and, with appropriate regularity conditions, it admits the power series expansion:

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

where the $j$-th cumulant, $\kappa_j$, is the $j$-th derivative of $\kappa_j(s)$ evaluated at $s = 0$ as Ghosh et al. (2016) describe. For instance, if $\ln x_t \sim N(\mu_x, \sigma_x^2)$, it is possible to define $\kappa_1^x = \mu_x$, $\kappa_2^x = \sigma_x^2$, $\kappa_{j>2}^x = 0$. Thus, $d_{Q_1}$ and $D\left(P\|\hat{Q}_1\right)$ can be written as:

$$d_{Q_1} = \kappa_1^x(1) - \kappa_1^x - (k^M(1) - \kappa_1^M) = \frac{\kappa_2^x - \kappa_2^M}{2!} + \frac{\kappa_3^x - \kappa_3^M}{3!} + \frac{\kappa_4^x - \kappa_4^M}{4!} + \ldots \quad (35)$$

and

$$D\left(P\|Q\right) = \frac{\kappa_2^x}{2!} + \frac{\kappa_3^x}{3!} + \frac{\kappa_4^x}{4!} + \ldots \quad (36)$$

Thus, Ghosh et al. (2016) observe that the first term in the infinite sums presented in Equation (36) captures the Gaussian terms of the distributions, while the other terms represent the nonnormal elements. This means that while the Hansen and Jagannathan (1997) distance only taking into account the second moments deviations, the entropy bounds developed by Ghosh et al. (2016) looks at all the moments.
of the candidate stochastic discount factor as the methodology of Almeida and Garcia (2016).

The idea of minimum adjustment of Hansen and Jagannathan (1997) is also strongly connected to Ghosh et al. (2016) bounds. Defining $M_t = z(\theta, t)\psi_t$, they observe that the $\psi^*$ identified by their $M1$ bound is similar to the Hansen and Jagannathan (1997) distance since in the entropy sense it provides the minimum adjustment that would make $z(\theta, t)\psi_t$ and admissible stochastic discount factor. Ghosh et al. (2016) also verify that while their bounds take into account all possible moments, the method of Hansen and Jagannathan (1997) only focuses in the second moment. Ghosh et al. (2016) perform the same calculation presented in the analysis of the Q1 bound to verify how the $M1$ bound considers higher moments of the candidate stochastic discount factor.

### d.2 Comparing the Entropy Bounds

Initially, comparing both methodologies, we must observe the difference between them in the assumptions considered. While Almeida and Garcia (2012) provide a manner to correct misspecified problems, Ghosh et al. (2016) consider the model correctly specified. Another important difference is that Ghosh et al. (2016) restrict their method to analyze models which the stochastic discount factor can be factorized as an observable part and other part that can be unobservable. Another important difference is the number of estimators accounted for in both methodologies. Almeida and Garcia (2012) use different estimators of the Cressie-Read family while Ghosh et al. (2016) apply the EL and ET as estimators for their bounds.

To compare both approaches, we focus our attention only in the EL estimator of both models since the analysis performed to observe the equivalence of both models is similar for the ET estimator used by Almeida and Garcia (2012) and Ghosh et al. (2016). It is important to observe that the estimators used in Ghosh et al. (2016) are the limiting cases of the distance criterion used by Almeida and Garcia (2012), since this fact allows us to discuss how general it is the Almeida and Garcia (2012)
Aiming to observe the equivalence between the two methodologies, initially observe that the moment condition considered by Almeida and Garcia (2012) in the multiplicative adjustment is \( E[my(\theta)x] = q \). Define \( R^e \) as the vector of excess return. Here we also use the fact that \( E[m] = 1 \) and the result follows from \( E[my(\theta)R^f] = 1 \), where \( R^f \) is the risk-free rate. Thus, it is possible to rewrite this moment condition as:

\[
E[my(\theta)R^e] = 0 \quad (37)
\]

From the result of Kitamura (2006) and applying the definitions of the model presented by Almeida and Garcia (2012), the EL estimator of \( m \) (which here we define as \( m^* \)), considering Equation (37) as the moment condition, is given by:

\[
m^*_t(\theta) = \frac{1}{T(1 + \alpha(\theta)'y(\theta,t)R^e_t)} \quad (38)
\]

where the variables follow the definitions presented in Subsection (b).

Analyzing Equation (38), it is possible to observe that this equation is equivalent to Equation (26) which is the EL estimator obtained by Ghosh et al. (2016). Using a similar argument, it is straightforward to show that the same equivalence is valid to the ET estimators found by both methodologies. As the Ghosh et al. (2016) only accounts for the EL and ET estimators, this framework is equivalent to the limiting cases of the framework provided by Almeida and Garcia (2012).

Finally, we observe that the method of Almeida and Garcia (2012) is a generalization of the framework provided by Ghosh et al. (2016). Almeida and Garcia (2012) present a manner to estimate the admissible stochastic discount factor generalizing the entropy estimators used to identify the misspecification of the asset pricing models. This methodology includes the EL and ET estimators since it uses the Cressie-Read family to define its criterion of misspecification.
Almeida and Garcia (2016) use the methodology developed by Almeida and Garcia (2012) to build bounds and to provide diagnoses of certain dynamic asset pricing models. They apply this methodology to analyze the long-run risk model of Bansal and Yaron (2004). Calibrating the model as Bansal and Yaron (2004), Almeida and Garcia (2016) first observe if the model attends their frontiers using a set of value and size portfolios. They verify that the model passes the market frontier for all values of $\rho$, however it stays outside of the six-portfolio frontiers for all the values of $\rho$. Ghosh et al. (2016) also apply their methodology to estimate bounds analyzing the long-run risk framework developed by Bansal and Yaron (2004). The equivalence between their frameworks is presented by their similar results related to Bansal and Yaron (2004) model.

3 Rare Events Hypothesis

In this chapter we initially present how the rare events hypothesis are considered in certain frameworks to provide an explanation of the EPP. We also discuss the limitations of the models based on the rare events hypothesis. The next step that we adopt is to present the disaster model of Barro (2006) in order to study the application of the stochastic discount factor bounds presented to analyze this methodology. We focus in the Barro (2006) approach since the bounds that are applied to study disaster models focus mainly in this approach. Thus, we present a detailed literature review on the application of the bounds in diagnosing the disaster model of Barro (2006).

The EPP is the contradictory difference observed between the average excess of return on the U.S. stock market relative to the one-month Treasury bill (i.e., the risk premium), which has been about 7% per year over the last century, and the results obtained from the representative agent model with time-separable CRRA utility when this model is calibrated to match micro evidence on households' risk attitude, time-series properties of consumption growth and asset returns. As in the representative agent model the results indicate a risk premium less than 1%, the discrepancy
between this finding and the evidence obtained from the U.S. stock market data generates the EPP (Mehra and Prescott, 1985).

Pioneered by Rietz (1988) and revived by a growing literature (Barro, 2006; Gabaix, 2012), the rare events hypothesis emerged to rationalize the EPP. As Julliard and Ghosh (2012) explain this hypothesis, consider that in every state there is an ex ante small probability of an extreme stock market crash and economic downturn. In this case risk-averse investors demand a high equity premium to compensate the extreme losses that can occur in these unlikely, but severe, states. In finite sample, if these bad states occur with a frequency lower than their true probability, ex post realized risk premia will be high; in other words, equity owners will be compensated by crashes and disasters that do not occur in the sample period. Therefore, for the economists, investors will appear irrational since their risk aversion is overestimated and the consumption risk is underestimated.

Thus, the EPP can be explained by assigning higher probabilities to bad states. There are some works that pursue a response to the EPP using this theoretical background. Barro (2006) review the rare events hypothesis and use it to explain the discrepancies observed between the U.S. stock market data (and also other countries) and the results obtained by the representative agent model usually applied in the literature. As Barro (2006) describes, they calibrate the representative agent framework with disaster probabilities from the twentieth century global history, especially the contractions of the World War I, the Great Depression and the World War II.

Barro (2006) argue that the main criticism related to the work of Rietz (1988) is related to the belief that its theory is based on counterfactually high probabilities and size of economics disasters. Thus, an important analysis that they perform to justify their calibration is measure the frequency and sizes of the international economic disasters described previously. With this measures they can justify the magnitude of the calibrated disaster probabilities. Barro (2006) find evidence that this theory explains the EPP.6

---

6 Barro (2006) also conclude that the rare events hypothesis can explain other asset-pricing puzzles as the RFRP found by Weil (1989).
In line with the work of Barro (2006), Gabaix (2012) also shows evidence that the rare events hypothesis can be used to provide a better understanding about certain asset-pricing puzzles. Gabaix (2012) proposes a model with a variable-severity version of the rare disasters hypothesis. Usually in the literature the rare events hypothesis is formulated with constant severity of disasters as Barro (2006) do. However, as Gabaix (2012) explains, assuming constant severity of disasters implies that the framework does not account for key features of asset markets, such as volatile price-dividend ratios for stocks, volatile bond risk premia, and return predictability.

Adapting the model to variable-severity of disasters, Gabaix (2012) can study the effect of time-varying disaster severity on the prices of stocks and bonds as well as on the predictability of their returns. He observes that many asset-pricing puzzles can be qualitatively understood through his methodology. Through a calibration procedure, Gabaix (2012) also verifies that real and nominal variables are sufficiently sensitive to disasters.

As Barro (2006) and Gabaix (2012), Julliard and Ghosh (2012) analyze if the EPP can be explained by the rare events hypothesis. Since to solve the EPP the rare events hypothesis implies that the probability of the disaster has to be increased, Julliard and Ghosh (2012) apply an information-theoretic alternative to the Generalize Method of Moments (GMM) usually used to estimate the Consumption-CAPM. They use this methodology because it allows them to impose less restrictions in the probability of each state than the GMM.

The information-theoretic estimators that Julliard and Ghosh (2012) use are the Empirical Likelihood (EL) and Bayesian Exponentially Tilted Empirical Likelihood (BETEL) which are methods that are solved as a nonparametric maximum likelihood. Instead of assuming a parametric distribution for the data, the value of the probability mass at each sample is treated as a variable to be estimated as Julliard and Ghosh (2012) describe. They show that these statistical methodologies provide the necessary correction to the probabilities of each state to simulate the disasters hypothesis.
as the rare events hypothesis predicts.

Julliard and Ghosh (2012) argue that if a certain sample has a frequency of economic disasters lower than their true probability, the GMM estimators will rationalize the realized risk premium with a higher risk aversion coefficient. Observe that restricting the probability of each state equalizing it with the sample frequencies as the GMM does will imply in the EPP according to the rare events hypothesis. Therefore, Julliard and Ghosh (2012) use an estimator that relaxes the constraint imposed by the GMM and has results that corroborates the rare event hypothesis to verify if it can explain the EPP.

The first finding of Julliard and Ghosh (2012) is to estimate the consumption Euler equation for the equity risk premium allowing the probabilities related to different states differ from their sample properties. Estimating the Consumption-CAPM they still reject it by the data, since their results ask a very high level of risk aversion to rationalize the stock market risk premium. Using data from sizes of the economic contractions and stock market returns during disasters, they identify that the disasters present in the world data do not offer support for the rare events hypothesis. Julliard and Ghosh (2012) show that disasters should be happening every 6-10 years, while the usual calibration frequency is a disaster every fifty nine years.

Through an innovative approach, Julliard and Ghosh (2012) use their estimators to build nonparametrically the rare events distribution needed to rationalize the EPP with low risk aversion. They also generate counterfactual histories of the data that have the same length as the historical time series. Julliard and Ghosh (2012) find that the puzzle itself would be very unlikely to happen using their counterfactual time series. For them, this indicates that if one believes that the rare events hypothesis explain the EPP, one is willing to believe that the puzzle itself is a rare event.

Another result analyzed by Julliard and Ghosh (2012) is whether rare events can rationalize the poor performance of the Consumption-CAPM in pricing the cross-section of asset returns. They observe that using their framework to build the proba-
bility distribution that solves the EPP through the rare events hypothesis worsen the ability of the Consumption-CAPM to explain the cross-section of stock returns. Julliard and Ghosh (2012) justify this result explaining that to solve the EPP they have to assign higher probabilities to bad states. Doing it, they also assign more weight to states that present low consumption growths (since bad states are recessions usually) and all assets in the cross-section tend to yield low returns, reducing the dispersion of consumption risk across assets, jeopardizing the performance of the framework.

Therefore, through a methodology that is used to build the bounds of Ghosh et al. (2016), Julliard and Ghosh (2012) verify that the EPP is hardly explained by the rare events hypothesis. Their conclusion implies that the model of Barro (2006) do not solve the EPP with their distribution of consumption drop. To guarantee that the rare events hypothesis provide an explanation to the EPP, Julliard and Ghosh (2012) find that is necessary a large drop in the consumption of the economy. However, this drop is not observed in the data usually considered in the literature. Despite the evidence presented by Barro (2006); Gabaix (2012) that the rare events hypothesis can explain the EPP, Julliard and Ghosh (2012) argue in the opposite direction using a similar structure to the bounds of Ghosh et al. (2016). In the next step of this paper, we provide a wider discussion on the application of entropy bounds to diagnose the framework of Barro (2006).

a Disaster Model

Barro (2006) presents a model where a disaster-like drop in aggregate consumption growth implies in a large equity premium and describes other non-normal features of asset returns. The left tail of the probability distribution is calibrated using international evidence of large drops in consumption growth around the world. In this framework, the distribution of consumption growth has a gaussian component and a jump component translating in non-normal asset returns. Thus, the logarithm of consumption growth is described as:
where \( \eta_{t+1} \) is the normal component \( N(\mu, \sigma^2) \) and \( J_{t+1} \) is a Poisson mixture of normals. Define \( j \) as the number of jumps which takes integer values with probabilities \( \exp\{-\tau\}\frac{t^j}{j!} \), where \( \tau \) is the jump intensity. Also consider \( J_t \) normal conditionally on the number of jumps \( (J_t|j \sim N(j\alpha, j\lambda^2)) \).

In this framework (Barro, 2006; Almeida and Garcia, 2016), the logarithm of the stochastic discount factor with power utility is:

\[
\log m_{t+1} = \log \beta - \zeta \eta_{t+1}
\]

where \( \zeta \) is the coefficient of relative risk aversion. Therefore, the mean of the stochastic discount factor is:

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

b Application of the Stochastic Discount Factor Bounds to Analyze the Disaster Model

Almeida and Garcia (2016) use their methodology to compute entropy bounds for the disaster model as well. They compute the Cressie-Read bounds with the returns on the S&P 500 index and equity options strategies on this index. As the left tail of the option return should be affected by large drops in consumption, the disaster model proposed by Barro (2006) should be able to price these derivatives portfolios. Thus, Almeida and Garcia (2016) provide bounds based on the Cressie-Read discrepancy functions considering different values of \( \gamma \).

Varying the magnitude of the disaster, Almeida and Garcia (2016) observe that the disaster model is admissible for quadratic bound (\( \gamma = 1 \)). However, this is not the case for the entropy bounds where \( \gamma = 0 \) and \( \gamma = -1 \). This finding is justified by
analyzing the Taylor expansions of Subsection (d.1). Through it, Almeida and Garcia (2016) observe that the more negative the $\gamma$ the higher (in absolute value) are the negative weights given to skewness of the model implied pricing kernel. Having negative values of mean size of disaster jump risk ($\alpha$) generates more positive skewness on the pricing kernel making it harder for the model, for a fixed value of $\alpha$, to pass the frontiers for more negative values of $\gamma$ as Almeida and Garcia (2016) note.

Almeida and Garcia (2016) also vary the risk aversion parameters of the representative investor in the spirit of Hansen and Jagannathan (1991) to see how their methodology performs to diagnose the disaster model. They observe that the quadratic bound ($\gamma = 1$) is admissible for any value of the risk aversion parameter above 5. As they lower the value of $\gamma$, which makes the bound more restrictive, they increase the value of risk aversion parameter at which the model becomes admissible. Once again analyzing the Taylor expansion of Subsection (d.1), Almeida and Garcia (2016) verify that increasing the risk-aversion coefficients strongly contributes to raise the skewness of the pricing kernel. As skewness receive a negative weight for $\gamma = -1$ and $\gamma = 0$, it becomes harder to the model to pass their bounds for more negative values of $\gamma$ for a fixed value of risk-aversion coefficient.

This methodology that considers the disaster Poisson model has shown that a large drop in consumption growth will make a consumption-based asset pricing model easily admissible considering the bounds of Hansen and Jagannathan (1991). However, tighter bounds capturing higher moments of the basis asset returns with some non-normalities in returns impose more stringent conditions on the admissibility of the model as Almeida and Garcia (2016) verify. There are other papers that analyze the disaster model of Barro (2006), such as Backus et al. (2011) and Liu (2015), using generalized entropy bounds which observe similar findings as the ones presented by Almeida and Garcia (2016) considering index option returns.

Backus et al. (2011) also analyze the performance of disaster models. Aiming to diagnose models such as the one developed by Barro (2006), they use index option returns to obtain information about extreme events. Equity index options are a use-
ful source of information as Backus et al. (2011) observe, because their price reveal how market participants value extreme events whether they happen in the analyzed sample or not. As the issue is to find the shape of the left tail of the probability distribution of the consumption growth, Backus et al. (2011) use the information obtained through equity index options to observe if what the distribution that they estimate is compatible with the one estimated from macroeconomic data.

They study a framework of the macro-finance literature based on consumption growth and compare it with an approach of the option-pricing literature aiming to obtain a better understanding of how disasters can be modeled. The difficulty of Backus et al. (2011) work lies in linking the risk-neutral distribution of equity returns implied by options to the true distribution of consumption growth in order to compare models based on consumption and option data to provide a diagnose of them. They connect pricing kernel to a concept of entropy to relate the risk-neutral distribution to the true distribution of consumption growth. Their entropy bounds are derived by Bansal and Lehmann (1997); Alvarez and Jermann (2005).

As Backus et al. (2011) note, using entropy as a concept of dispersion in studies of disasters allow them to account for properties of returns that are difficult to explain in lognormal framework. In their setting, Backus et al. (2011) perform three analysis. In the first one they compare pricing kernels obtained from macro-finance and option-pricing models. Then, in the second they compare option prices derived from a macro-finance model to those that are observe. Finally, in the third analysis, Backus et al. (2011) compare the distribution of consumption growth derived from option prices through a macro-finance model based on macroeconomic data. All the examinations that Backus et al. (2011) present lead to the conclusion that options imply smaller probabilities of disasters than have been estimated from macroeconomic data.

Almeida and Garcia (2016) observe that their methodology has a direct link with the measure entropy of Backus et al. (2011). When $\gamma = -1$ their discrepancy function is equal to the entropy function obtained by Backus et al. (2011). They also note
that their framework has a connection with the Stutzer (1995) approach as Backus et al. (2011). However, Almeida and Garcia (2016) show that their methodology complement the work of Backus et al. (2011) since their bounds bring additional non-redundant information when analyzing asset pricing models. Another important finding of the Almeida and Garcia (2016) approach is that, in the dual space, it matches to the maximum excess log return of the growth-optimal portfolio of Bansal and Lehmann (1997).

Liu (2015) follows the Backus et al. (2011) work generalizing the bounds developed by Hansen and Jagannathan (1991) (through an application of the Holder's inequality) based on index option returns. This system of nonparametric bounds encodes all the distributional information of pricing kernel as Liu (2015) notes. Through moment expansions, as Backus et al. (2011), Liu (2015) shows how different moments of the pricing kernel affects the entropy function and how weighted asset return moments provide information on the entropy functions. In his empirical work, Liu (2015) focuses in obtain a better understanding of disaster models.

The approach presented by Liu (2015) differs from the one used by Backus et al. (2011) in certain aspects. First, Liu (2015) makes no assumption about the link between macro fundamentals and the asset market, thus his results are robust to misspecification. Second, Liu (2015) takes the realized option returns as given and studies their effects on the pricing kernel. Finally, Liu (2015) analyzes disaster models through the spectrum of new bounds. He observes that the ability of these bounds in providing incremental information on high order moments of pricing kernel and asset returns gives a sharper inference of disaster models.

Focusing the attention on the analysis of index option returns, Liu (2015) initially uses option implied bounds to confront standard rare disaster models. He marks up the parameter region that all nonparametric bounds are simultaneously satisfied. Liu (2015) shows that his bounds provide the sharpest restriction on the model compared to existing bounds. To take statistical uncertainty under consideration, Liu (2015) develops a formal testing framework that allows him to analyze multiple
assets and different types of bounds. Through this framework, he rejects the benchmark disaster model and alternative specifications.

Liu (2015) proposes a framework similar to the one developed by Almeida and Garcia (2012, 2016). Both methodologies are based on the approach of Hansen and Jagannathan (1991), the difference is how they derive their bounds. While Liu (2015) generalizes the Hansen and Jagannathan (1991) bound through the Holder's inequality, Almeida and Garcia (2012, 2016) generalize it using the Cressie-Read family of discrepancy functions which comprehends a wide class of other entropy estimators. They also show through a Taylor expansion how their bounds can bring new information analyzing high order moments of the pricing kernels.

Despite the similarities presented between the works of Almeida and Garcia (2016) and Liu (2015), there is an important difference between them. Consider the bound presented by Liu (2015) using the generalized entropy function (GEF) which is:


g_{\text{EF}}(s; M) = \log \mathbb{E}(M) - \frac{1}{s} \log \mathbb{E}(M^s) \tag{42}

where $M$ is the stochastic discount factor of the asset pricing model studies, $R$ is the asset returns and define $s$ as any real number. He shows that this is an extension of the entropy measure adopted by Backus et al. (2011). Thus, his bound is written as:


g_{\text{EF}}(s; M) \geq \frac{s-1}{s} \log \mathbb{E}(R^{s-1}) - \log(R_f), \ \forall s \in (-\infty, 1) \tag{43}

Through a Taylor expansion, as Backus et al. (2011); Almeida and Garcia (2012, 2016), Liu (2015) describes his bound as the following:


g_{\text{EF}}(s; M) \equiv \mathbb{E}(\log R) - \log R_f + \sum_{i=2}^{\infty} \frac{\kappa_i(\log R)}{i!} \left( \frac{s}{s-1} \right)^{i-1} \tag{44}

where the first term of the right-hand side is the risk premium, the second term is high order moments and $\kappa_i(\log R)$ is the $i$-th return cumulant.
Note that the relevance of the high order moments is directly related to the value of $s$. Comparing to the methodology of Almeida and Garcia (2012, 2016), the value of $s$ in the Liu (2015) framework plays the role that $\gamma$ plays in the approaches of Almeida and Garcia (2012, 2016). Nonetheless, there is an important difference between the bounds. Comparing Equation (34) and Equation (44), we note that the effect of high order moments in the stochastic discount factor bound of Almeida and Garcia (2012, 2016) is not connected to the asset returns while the methodology of Liu (2015) is directly related to it.

This is a relevant advantage of Almeida and Garcia (2016) approach due to the influence that the asset returns may incur when one varies the value of $\gamma$ or $s$ (in the Liu (2015) model) to differentiate the contribution from different high-order moments. While in the Almeida and Garcia (2016) model this influence does not occur with variations on $\gamma$, the model of Liu (2015) can have its findings affected by the asset returns when one varies the value of $s$.

4 Final Considerations

The methodologies developed by Almeida and Garcia (2012); Ghosh et al. (2016), which are influenced by Hansen and Jagannathan (1991) bound and Hansen and Jagannathan (1997) distance, are usually applied to diagnose asset pricing models. These frameworks have a wide range of applications to different models. In this paper, we focus our attention to understand the specific characteristics of these methods and their use in the analysis of disaster models (Barro, 2006). We discuss the relation between the entropy bounds built by Almeida and Garcia (2012), which is directly related to Almeida and Garcia (2016) and Ghosh et al. (2016) works, then we show how these methodologies and other bounds related to them can be used to examine disaster models (Barro, 2006).

Describing the Almeida and Garcia (2012) and Ghosh et al. (2016) frameworks,
initially, we observe that there is a equivalence between these methods. The approach of Almeida and Garcia (2012) is a generalization of the framework provided by Ghosh et al. (2016). Almeida and Garcia (2012) present a manner to estimate the admissible stochastic discount factor generalizing the entropy estimators used to identify the misspecification of the asset pricing models. This methodology includes the EL and ET estimators since it uses the Cressie-Read family to define its criterion of misspecification. As the Ghosh et al. (2016) only accounts for the EL and ET estimators, this framework is equivalent to the limiting cases of the framework provided by Almeida and Garcia (2012). The equivalence between these two bounds is corroborated by their similar results related to the diagnose of Bansal and Yaron (2004) model.

In this work, we discuss the application of these frameworks to analyze disaster models (Barro, 2006). Pioneered by Rietz (1988) and revived by a growing literature (Barro, 2006; Gabaix, 2012), the rare events hypothesis emerged to rationalize the EPP and motivates the work of Barro (2006); Gabaix (2012) to build their disaster models. The entropy bounds constructed by Almeida and Garcia (2012); Ghosh et al. (2016) are specifically design to analyze different asset pricing models and we present results about it. The literature has a detailed analysis of these methods according to performance, and we observe that the entropy bounds allow a study of the role that the higher moments of pricing kernels plays in the performance of asset pricing models. In particular, in the disaster models, higher moments have important implications as a variety of works present (Backus et al., 2011; Almeida and Garcia, 2016; Ghosh et al., 2016; Liu, 2015).

Therefore, we present the literature that discuss entropy bounds developed to the generalize the approach of Hansen and Jagannathan (1991) to evaluate asset pricing models and to take into account information of higher moments that are not considered in the initial methodology of Hansen and Jagannathan (1991). The methodologies of Backus et al. (2011); Almeida and Garcia (2016); Ghosh et al. (2016); Liu (2015) offer this advantage which clearly offers an important contribution in the understanding of asset pricing models. In particular, we focus our analysis in disaster
models. Since the stochastic discount factor bounds that are applied to analyze disaster models usually consider the approach of Barro (2006), we examine the findings related to the diagnose of this model through the lens of the aforementioned bounds. There are other asset pricing models which are analyzed in the scope of the entropy bounds but are not the purpose of this work (see Backus et al. (2011); Almeida and Garcia (2016); Ghosh et al. (2016); Liu (2015) for further details on it).

References


