Journal of Econometrics 170 (2012) 519–537

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Assessing misspecified asset pricing models with empirical likelihood estimators

Caio Almeida [a], René Garcia [b,*]

[a] Graduate School of Economics, Getulio Vargas Foundation, Rio de Janeiro, Brazil
[b] EDHEC Business School, 393, Promenade des Anglais, BP 3116, 06202 Nice Cedex 3, France

ARTICLE INFO

Article history:
Available online 2 June 2012

JEL classification:
C1
C5
G1

Keywords:
Stochastic discount factor
Euler equations
Generalized minimum contrast estimators
Model misspecification
Cressie–Read discrepancies

ABSTRACT

Hansen and Jagannathan (1997) compare misspecified asset pricing models based on least-square projections on a family of admissible stochastic discount factors. We extend their fundamental contribution by considering Minimum Discrepancy projections where misspecification is measured by a family of convex functions that take into account higher moments of asset returns. The Minimum Discrepancy problems are solved on dual spaces producing a family of estimators that captures the least-square problem as a particular case. We derive the asymptotic distributions of the estimators for the Cressie–Read family of discrepancies, and illustrate their use with an assessment of the Consumption Asset Pricing Model.

1. Introduction

Asset pricing models can be seen as useful approximations of reality to explain empirical stylized facts. Hansen and Jagannathan (1997), hereafter HJ, suggested that an appropriate way to compare their performance consisted in evaluating functions of their implied pricing errors on corresponding Euler equations. They proposed a useful test for comparisons of possibly misspecified asset pricing models, based on a least-square projection of a proxy model on a family of admissible stochastic discount factors (SDFs). This test has been used in a number of empirical papers as a tool for model diagnostics as well as model selection (see for instance, Jagannathan and Wang, 1996, Hodrick and Zhang, 2001, Wang and Zhang, 2005, Chen and Ludvigson, 2009, Chen et al., 2008, and Kan and Robotti, 2009 among many others).

The idea of adopting least-square theory to determine admissible discount factors that are close to asset pricing proxies is intuitive. First, it provides an easy interpretation of the degree of misspecification of a model as a maximum pricing error measure in the space of payoffs. It is also easy to implement by using duality theory in convex optimization problems (see Luenberger, 1969). However, the quadratic metric has one important limitation. It implies that misspecification is provided by a quadratic form on the pricing errors of the primitive securities that fails to take into account moments of the payoffs (returns) distributions other than mean and variance.

There is a large body of research indicating the importance of considering skewness and kurtosis when pricing assets.[1] In econometrics, there is also a considerable literature proposing increasingly more sophisticated Empirical Likelihood-type estimators that are robust against distributional assumptions and that possess good properties analogous to those of parametric likelihood procedures (see Kitamura, 2006b).[2] In particular, Stutzer (1995) and Kitamura and Stutzer (2002) have suggested the use of relative entropy to develop a research program that parallels that of HJ (1991,1997).

In this paper, we propose alternative methods to measure the degree of misspecification of asset pricing models that make use of

* Corresponding author.
E-mail addresses: calmeida@fgv.br (C. Almeida), rene.garcia@edhec.edu (R. Garcia).

[1] See for instance Kraus and Litzenberger (1976), Rubinstein (1973), Baroni-Adesi (1985), Harvey and Siddique (2000), Dittmar (2002), and Vanden (2006), among many others.

[2] For instance, Owen (1988) proposed the Empirical Likelihood estimator, and Kitamura and Stutzer (1997) the Exponential Tilting Estimator (see also Imbens et al., 1998). Smith (1997) proposed a large class of Generalized Empirical Likelihood (GEL) estimators, later shown to be equivalent to the subset of Minimum Discrepancy estimators with Cressie–Read discrepancies (Newey and Smith, 2004). Recently, Smith (2007) developed tests for conditional moment restrictions models based on a kernel-weighted version of the Cressie–Read power divergence family of discrepancies.

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.05.020

the theory of Minimum Discrepancy (MD) estimators (Corcoran, 1998). The idea is to consider general convex functions ϕ to calculate the distance between a certain asset pricing proxy model y and the family M of admissible SDFs (that prices a set of underlying payoffs x on primitive securities). [3] We formulate this problem within an MD framework where the goal is to obtain an SDF m^* that is admissible (i.e., satisfies the moment conditions by pricing primitive securities) and that is the closest possible, in the spirit of Csiszar's I-divergencies (1975,1991), to the asset pricing proxy, by minimizing $\phi(1+m-y)$. We make use of duality theory (see Kitamura, 2006b, and Borwein and Lewis, 1991) to estimate this MD probability measure and its distance to the proxy y, by solving simpler finite-dimensional problems.

When the MD problems are specialized to the class of Cressie and Read (1984) discrepancies, we show that, under our formulation, the dual optimization problems reduce to a class of GEL estimators where the proxy model y appears only in the discrepancy function ϕ and not in the moment conditions as it is generally the case. This formulation makes clear that we are interested, like HJ (1997), in measuring the degree of misspecification of a model y with respect to a family of admissible SDFs M that is invariant to changes in the model vector of parameters θ. Moreover, this family will also be invariant with respect to changes of models and should depend only on the primitive payoffs x and prices q.

By looking at the first-order conditions of the dual quadratic problem, HJ (1997) found a nice interpretation for their least-square solutions as corrections to the asset pricing proxy model. They showed that the solutions are given by the proxy y from which we subtract the optimal linear combination of primitive asset payoffs ($\lambda'_{\rm HJ}x$) that is the smallest linear correction (in the least squares sense) for y to become an admissible SDF. It happens that in our MD problems we have similar interpretations of solutions as proxy corrections to become admissible SDFs. The solutions to our MD problems give additive correction terms to the proxy y that are nonlinear functions of the optimal linear combination of primitive asset's payoffs ($\lambda'_{\rm MD}x$), which are the smallest correction (in the ϕ divergence additive sense) for y to become an admissible SDF.

Regarding model assessment and parameter estimation, we extend Hansen et al. (HHL, 1995) who derive consistent estimators of the specification-error and volatility bounds with a least-square criterion and an asymptotic distribution theory. We propose consistent estimators of information bounds and specification-error bounds that are based on the MD measures of the Cressie–Read family and derive their corresponding asymptotic distributions to make statistical inferences. In particular, we will be able to test whether a model-based specific SDF is within the information bound or not and whether the degree of misspecification of a parametric asset pricing model is significantly different from zero or not. To develop our estimators and the asymptotic distribution theory we follow Kitamura (2000) and Kitamura and Stutzer (1997) who develop nonparametric maximum likelihood estimation procedures based on the Kullback–Leibler Information Criterion. We extend these estimation procedures to the family of Cressie–Read discrepancy measures.

Regarding model comparison, HJ (1997) suggest as a first step the estimation of model parameters by minimizing the HJ distance between the model and the family of admissible SDFs. Then, assuming the existence of a set of model candidates whose parameters were previously estimated based on the HJ distance, they suggest selecting the model with the smallest distance.[4] Similarly in our MD problems, for any fixed discrepancy function ϕ, our theory suggests the estimation of model parameters by minimizing the discrepancy chosen, and the use of the MD distance to rank candidate asset pricing models.[5]

Our empirical illustration consists in applying the estimators for several members of the Cressie–Read family to the Consumption CAPM (Breeden, 1979). We first test its ability to price a set of primitive securities (bond and S&P 500) at different regions of the parametric space. The discrepancy between asset pricing proxies and admissible SDFs is measured by the following Cressie–Read functions: Pearson's Chi-Square, Empirical Likelihood (EL, Owen, 1988), Hellinger's distance, Exponential Tilting (ET, Kitamura and Stutzer, 1997), Euclidean Likelihood or Continuous Updating Estimator (CUE) (Hansen et al., 1996). We then conduct estimation and testing for these various measures of misspecification.

The rest of the paper is organized as follows. Section 2 introduces the market structure, defines admissible SDFs, and presents the original HJ framework for model estimation and selection. Section 3 formulates our generalization to the HJ methodology that considers MD optimization problems. It presents the main theorem that provides a family of metrics that contains HJ (1997) as a particular case. It also defines implied probabilities, presents their relation to admissible SDFs, and provides some particular model selection procedures based on known discrepancy functions belonging to the Cressie–Read family. Section 4 shows the consistency of the estimators and derives their asymptotic distributions. Section 5 provides an illustration of the estimation and testing procedures with the Consumption CAPM. It describes the model and the data and provides estimation and tests results. Section 6 presents a discussion on the results, analyzing the relation between implied admissible SDFs, pricing errors, and discrepancies adopted. Section 7 concludes.

2. Stochastic discount factors and asset pricing proxies

Following the lead of Harrison and Kreps (1979), Chamberlain and Rothschild (1983), and HJ (1997) we model portfolio payoffs as elements of a Hilbert space. We assume that assets are purchased at a certain time t and that the payoffs are received at a time $T > t$. Let G_T represent the sigma-algebra that represents the conditioning information at date T, and L^2 denote the space of all square-integrable (i.e., finite second moments) random variables that are measurable with respect to G_T. Assume there exists a set of N primitive securities whose payoffs are represented by a vector $x \in \Re^N$, with $x \in L^2$ and in addition having a nonsingular second moment matrix Exx'. A payoff p will be any square integrable variable which is obtained as a linear combination of the payoffs of the N primitive securities:

$$P \equiv \left\{xc : c \in \Re^N\right\}. \tag{2.1}$$

3 Similar to HJ (1997), we formulate two types of problems: one based on the family M of admissible SDFs, and the other based on the family M^+ of admissible strictly positive SDFs.

4 Hodrick and Zhang (2001) use the HJ (1997) distance to compare ten different asset pricing models based on the 25 Fama and French (1993) test assets. Recently, Kan and Robotti (2009) suggest a more formal selection by developing a test to compare the HJ distance between models, and obtain the asymptotic distribution of this test, for any combination of correctly specified, misspecified, and nested or non-nested model candidates.

5 Formal statistical tests for model comparisons are not developed in this paper. We refer the reader to a recent literature on model choice with empirical likelihood estimators (see Kitamura, 2000, Ramalho and Smith, 2002, Hong et al., 2003, Kitamura, 2006a, and Chen et al., 2007, among others). In particular, Kitamura (2000) proposes nonparametric likelihood ratio tests based on the Exponential Tilting estimator to compare possibly misspecified moment-based econometric models. Hong et al. (2003) suggests a GEL-based model selection criteria alternative to the non-nested model selection test of Kitamura (2000). The framework in any of these two papers can be adapted to our above-mentioned GEL problems to derive formal statistical tests for model comparison and selection.

We further assume that the payoffs in P satisfy the Law of One Price and that the pricing functional π is continuous and linear on P.[6]

An admissible SDF will be any square-integrable random variable m that correctly prices all asset payoffs $p \in P$

$$\pi(p) = E(mp). \tag{2.2}$$

An asset pricing model y will be an approximation for an admissible SDF, and will possibly price some payoffs in P with error:

$$\pi_y(p) = E(yp), \tag{2.3}$$

where the error is measured by the difference $\pi(p) - \pi_y(p)$. We assume that it will be a function of a set of parameters θ and also that it may depend on a subset of elements of vector $z = [x\ u]$, with x representing the payoffs of the primitive securities, and u a set of economic factors like, for instance, consumption growth or any other macroeconomic variable.[7]

2.1. Hansen and Jagannathan's (1997) least-squares distance for proxy models

Given a proxy asset pricing model $y(\theta)$, parameterized by a vector of parameters $\theta \in \Re^k$, HJ (1997) suggest to measure its degree of misspecification by obtaining the least-squares projection of this proxy into the space of admissible SDFs M:

$$\delta_{\mathrm{HJ}}(\theta)^2 = \min_{m \in M} \|m - y(\theta)\|^2 = \min_{m \in M} E\{(m - y(\theta))^2\}. \tag{2.4}$$

This problem can be rewritten by noticing that $m \in M$ can be re-expressed as $m \in L^2$ satisfying the moment condition (2.2) for the particular set of primitive securities:

$$\delta_{\mathrm{HJ}}(\theta)^2 = \min_{m \in L^2} E\{(m - y(\theta))^2\}$$
$$\text{subject to } E(mx) = \pi(x) = q. \tag{2.5}$$

Making use of Lagrange Multipliers the problem becomes:

$$\delta_{\mathrm{HJ}}(\theta)^2 = \min_{m \in L^2} \sup_{\lambda \in \Re^N} E\{(m - y(\theta))^2 - 2\lambda'(mx - q)\}. \tag{2.6}$$

By fixing the Lagrange Multipliers and solving the minimization on the variable m, HJ (1997) obtained the following dual optimization problem:

$$\delta_{\mathrm{HJ}}(\theta)^2 = \max_{\lambda \in \Re^N} E\{y(\theta)^2 - (y(\theta) - \lambda' x)^2 - 2\lambda' q\}, \tag{2.7}$$

which admits as a solution the (square of) Hansen and Jagannathan's distance:

$$\delta_{\mathrm{HJ}}(\theta)^2 = (Exy(\theta) - q)' Exx'^{-1} (Exy(\theta) - q). \tag{2.8}$$

HJ (1997) also solved a variation of problem (2.4) where they restricted the optimization set to admissible SDFs that are strictly positive random variables:

$$\delta^+_{\mathrm{HJ}}(\theta)^2 = \min_{m \in M+} \|m - y(\theta)\|^2 = \min_{m \in M+} E\{(m - y(\theta))^2\}, \tag{2.9}$$

where $M+ = \{\tilde{m} \in M, \tilde{m} \gg 0\}$, with $\tilde{m} \gg 0$ indicating that $\tilde{m}$ is a strictly positive random variable.

The same techniques of dual optimization described above can be applied to solve this problem:

$$\delta^+_{\mathrm{HJ}}(\theta)^2 = \max_{\lambda \in \Re^N} E\{y(\theta)^2 - [(y(\theta) - \lambda' x)^+]^2 - 2\lambda' q\}, \tag{2.10}$$

where $(y(\theta) - \lambda' x)^+ = \max(y(\theta) - \lambda' x, 0)$ (see HJ (1997) for more details).

The main advantage of using the set $M+$ instead of M is that any SDF in $M+$ allows one to obtain arbitrage-free prices of any derivatives of the basis assets, while this is not true for elements in M that achieve negative values in at least one state.[8]

2.1.1. Interpreting the primal and dual problems

The dual optimization problem (2.7) is nicely interpreted by HJ (1997) as an optimal portfolio problem with a quadratic utility function. The Lagrange Multipliers represent the portfolio weights on the different primitive securities payoffs. Stutzer (1995) explores this portfolio interpretation in a nonparametric setting based on ET obtaining a CARA (exponential) utility function, and Almeida and Garcia (2008) generalize Stutzer's interpretation in a strictly nonparametric setting with general Cressie–Read discrepancy functions providing a portfolio interpretation with HARA (Hyperbolic Absolute Risk Aversion) utility functions.

The first-order conditions from problem (2.7) also give an interesting interpretation, this time for the solution of the primal problem:

$$q = E\{(y(\theta) - \lambda'_{\mathrm{HJ}} x)x\}. \tag{2.11}$$

Eq. (2.11) shows that the optimal Lagrange Multipliers λ_{HJ} that solve this problem find the smallest correction in the mean square sense to the proxy $y(\theta)$ such that it becomes an admissible SDF. With an extension to MD estimators, we will obtain a similar interpretation but the correction to the proxy will be non-linear in the primitive payoffs x.

HJ (1997) also interpret the primal problem (2.4) as a maximum pricing error problem per unit norm. A linear functional $\pi_a = \pi - \pi_y$ representing the approximate pricing errors is defined. They show that δ_{HJ} is the norm of this functional, and moreover that this norm is achieved by a special payoff $\tilde{p}$ obtained with the application of the Riesz representation theorem to the functional π_a.

2.1.2. Model estimation based on the HJ distance

HJ (1997) suggest estimating the parameter vector θ by minimizing either the HJ distance or the HJ distance with positivity constraint:

$$\operatorname*{argmin}_{\theta \in \Re^k} \delta_{\mathrm{HJ}}(\theta), \quad \text{or} \tag{2.12}$$

$$\operatorname*{argmin}_{\theta \in \Re^k} \delta^+_{\mathrm{HJ}}(\theta), \tag{2.13}$$

as alternative estimators to the GMM (Hansen, 1982):

$$\operatorname*{argmin}_{\theta \in \Re^k} g(\theta)' W g(\theta) \tag{2.14}$$

where $g(\theta) = E(y(\theta)x)$ represents the moment conditions, and W is an $n \times n$ symmetric positive definite matrix. Note that the HJ estimators in Eqs. (2.12) and (2.13), as well as the GMM estimator in

6 We assume the existence of the second moments to be able to work in a Hilbert space. For a treatment of the case with inexistent moments (not performed here), the payoffs should be in a Banach (L^j) space (see Royden, 1988), and, in principle, the existence of a linear pricing functional could be questioned. However, we refer to Araujo and Monteiro (1989) who provide a proof of existence of equilibrium in L^j spaces, therefore guaranteeing the existence of a linear pricing functional in such spaces.

7 For simplicity of notation, up to Section 4 we represent the asset pricing proxy by $y(\theta)$, where θ is a vector of parameters. In Section 4, which provides asymptotic properties of our estimators, we need to consider an explicit dependence of y on the vector z.

8 In addition, any SDF in $M+$ can be taken as the solution of an equilibrium model. Again, this is not true for SDFs that assume negative values.

(2.14), are special cases of the minimum distance estimators with a quadratic norm.

In an asset pricing context, HJ (1997) showed that the main difference between the HJ estimator in Eq. (2.12) and GMM is that in general the optimal matrix W in Eq. (2.14) (see Hansen and Singleton, 1982) will depend on the particular proxy model $y(\theta)$ adopted, while the normalizing matrix is fixed at $(Exx')^{-1}$ in the case of the HJ estimator in Eq. (2.12). The HJ distance has to be preferred since it gives weights to the pricing errors that are invariant to the asset pricing proxy $y(\theta)$.

3. Minimum Discrepancy distance for proxy models

3.1. The Minimum Discrepancy problem

Given a proxy asset pricing model $y(\theta)$ and a convex discrepancy function ϕ, the MD problem is to find an admissible SDF which is as close as possible to $y(\theta)$ in the ϕ discrepancy sense[9]:

$$\delta_{\mathrm{MD}}(\theta) = \min_{m \in L^2(y)} E\{\phi(1 + m - y(\theta))\} \quad \text{subject to } E(mx) = q \tag{3.1}$$

where $L^2(y) = \{m \in L^2, m \gg y(\theta) - 1\}$.

We also solve the constrained case where we search the strictly positive admissible SDF that is closest to $y(\theta)$ in the ϕ sense:

$$\delta^{+}_{\mathrm{MD}}(\theta) = \min_{m \in L^2(y)} E\{\phi(1 + m - y(\theta))\} \quad \text{subject to } E(mx) = q,\ m \gg 0. \tag{3.2}$$

These problems should be of interest when either the asset pricing proxy model $y(\theta)$ can depend nonlinearly on the underlying primitive securities or when the underlying primitive securities themselves include assets with non-Gaussian returns.[10] In the first case, it is not clear that corrections to the asset pricing proxy should be linear combinations of basis assets payoffs like in HJ (1997). In the second case, it is not clear that the penalty for a proxy asset pricing model $y(\theta)$ should only depend on the second moments of the pricing errors. Therefore, adopting more general discrepancies will probably be more appropriate when dealing with assets with nonlinear or asymmetric payoffs such as options, mortgages, credit derivatives, other exotic but liquid instruments, and also equities with skewed and fat tailed returns.

We make use of arguments found in Borwein and Lewis (1991) to solve our discrepancy problems based on simpler optimization problems on dual spaces. The corresponding dual optimization problem is given by:

$$v_{\mathrm{MD}}(\theta) = \max_{\lambda \in \Re^N} \lambda' q - E\{\phi^*(\lambda' x, y(\theta))\}, \tag{3.3}$$

where ϕ^* denotes the convex conjugate of ϕ (see Luenberger, 1969) appropriately translated by the proxy model $y(\theta)$:

$$\phi^*(z, y(\theta)) = \sup_{w > y(\theta) - 1} zw - \phi(1 + w - y(\theta)), \tag{3.4}$$

for the optimization problem in Eq. (3.1), and:

$$\phi^{*+}(z, y(\theta)) = \sup_{w > \max(0, y(\theta) - 1)} zw - \phi(1 + w - y(\theta)), \tag{3.5}$$

for the optimization problem in Eq. (3.2).[11] We further restrict the domain of ϕ^* and ϕ^{*+} to z's such that $\partial\phi^{-1}(z) > 0$, where $\partial\phi^{-1}$ denotes the inverse of the first derivative of ϕ.[12]

Newey and Smith (2004) show that when the discrepancy function is chosen within the Cressie and Read (1984) family, the dual problem belongs to the class of GEL estimators. In a recent paper, Almeida and Garcia (2008) generalize Hansen and Jagannathan (HJ, 1991) nonparametric bounds considering an arbitrary number of moments of returns[13] by specializing the Cressie–Read discrepancy problem to a nonparametric setting. The results in this paper extend those in Almeida and Garcia (2008) to explicitly consider the existence of a parametric model $y(\theta)$ in order to generalize HJ (1997). The optimization problem now consists in obtaining admissible SDFs that combine parametric aspects coming from $y(\theta)$ with nonparametric aspects coming from an optimal linear combination of primitive assets' payoffs (optimal in the divergence sense).

Aiming at incorporating the parametric model $y(\theta)$ but keeping the moment conditions as in HJ (1997), we formulate MD problems slightly different from Newey and Smith (2004) and Almeida and Garcia (2008). We introduce a translation of the function ϕ by $1 - y(\theta)$. This translation guarantees that the MD problem in Eq. (3.1) (or (3.2)) has solution $m = y(\theta)$ whenever the asset pricing proxy $y(\theta)$ is an admissible SDF (or a strictly positive admissible SDF). In addition, the inclusion of $y(\theta)$ in the divergence function (and not in the moment condition explicitly) allows us to interpret our MD problem as a genuine generalization of HJ (1997). Indeed, as in HJ (1997) we measure the distance of the proxy model $y(\theta)$ to a fixed family M (or $M+$) of admissible SDFs.

The next theorem provides the type of optimization problems that will have to be solved to find the discrepancy of $y(\theta)$ with respect to the family M, and the corresponding admissible SDF closest to $y(\theta)$, when the discrepancy belongs to the Cressie and Read (1984) family of discrepancies, most adopted in the current econometric literature.

Theorem 1. *Let $y(\theta)$ represent the asset pricing proxy, parameterized by a vector of parameters $\theta \in \Theta$. Let the discrepancy function belong to the class of Cressie–Read functions: $\phi(\pi) = \frac{\pi^{\gamma+1}-1}{\gamma(\gamma+1)}$ with $\gamma \in \Re$. In this case for a fixed vector of parameters θ, the optimization problem (3.1) specializes to:*

$$\delta_{\mathrm{CR}}(\theta) = \min_{m \in L^2(y)} E\left\{\frac{(1 + m - y(\theta))^{\gamma+1} - 1}{\gamma(\gamma+1)}\right\} \quad \text{subject to } E(mx) = q. \tag{3.6}$$

Then the GEL problem dual to the MD problem is given by:

$$v_{\mathrm{CR}}(\theta) = \max_{\lambda \in \Re^N} \lambda' q - E\left\{\frac{(\gamma \lambda' x)^{\frac{\gamma+1}{\gamma}}}{\gamma + 1}\right.$$

9 We are particularly interested in a strictly positive argument for the discrepancy function ϕ because this argument should be a normalized version of a Radon–Nikodym derivative (see Section 3.2), and all Radon–Nikodym derivatives are strictly positive by construction. This restriction on the argument implies that we should search for SDFs $m \in L^2$ satisfying the restriction $1 + m - y(\theta) \gg 0$. Note that, as long as the discrepancy is defined for negative arguments, we could relax this positivity restriction with the cost of not having a Radon–Nikodym derivative anymore.

10 In Almeida and Garcia (2008), we illustrate graphically how different discrepancy measures (that is different γs in the CR family put implicitly different weights on moments higher than two, in particular skewness and kurtosis).

11 It is interesting to note that ϕ^{*+} is a version of ϕ^*_+ in Borwein and Lewis (1991) that accounts for the translation given by the proxy model $y(\theta)$.

12 This restriction guarantees that the critical point of $H(w) = zw - \phi(1 + w - y(\theta))$ lies in the interval $[y(\theta) - 1, +\infty)$. It is obtained by calculating the first-order condition of $H(w)$, inverting for z and imposing that $1 + w - y(\theta) > 0$. Kitamura (2006b) suggests as an alternative control the modification of ϕ to achieve $+\infty$ whenever its argument is negative, implying that the optimizer searches for solutions only at positive elements of the primal problem domain.

13 See also Snow (1991) who generalizes HJ (1991) bounds to account for higher moments of returns.

$$+ (y(\theta)-1)\lambda' x + \frac{1}{\gamma(\gamma+1)} \Bigg\}, \tag{3.7}$$

and the admissible SDF which is closest to the asset pricing proxy y is given by:

$$m_{CR}(\theta) = y(\theta) - 1 + (\gamma\lambda_*' x)^{\frac{1}{\gamma}}, \tag{3.8}$$

where λ_* is the solution of the optimization problem (3.7).

Proof of Theorem 1. The goal is to find the convex conjugate ϕ^*, to use it in Eq. (3.3) to obtain the dual optimization problem with a Cressie and Read divergence. Letting $H(x) = zx - \frac{(1+x-y(\theta))^{\gamma+1}-1}{\gamma(\gamma+1)}$ defined in $[y(\theta)-1, \infty)$, and differentiating in x, we obtain its supremum at $x^{\sup} = y(\theta) - 1 + (\gamma z)^{\frac{1}{\gamma}}$. Note that since $x^{\sup} > y(\theta) - 1$ for any z in the domain of ϕ^*, $x^{\sup}$ belongs to the domain of H. By substituting $x^{\sup}$ in ϕ^* we obtain $\phi^*(z, y(\theta)) = (y(\theta)-1)z + \frac{(\gamma z)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + \frac{1}{\gamma(\gamma+1)}$. Applying ϕ^* in Eq. (3.3) gives the optimization problem (3.7). The first-order conditions of this optimization problem with respect to λ are:

$$\frac{\partial v_{CR}}{\partial \lambda} = q - E\left\{\left(y(\theta) - 1 + (\gamma\lambda_*' x)^{\frac{1}{\gamma}}\right) x\right\} = 0, \tag{3.9}$$

showing that m_{CR} in Eq. (3.8) is an admissible SDF that minimizes the MD problem (3.1) when the divergence is a member of the Cressie–Read family. □

We assume that there exists a risk-free asset on the set of primitive securities paying interest rate equal to r_f. The existence of such an asset is also assumed by HJ (1997, Assumption 1.2) and it is important to guarantee that our discrepancy problems are well posed in the sense that the mean of any admissible SDF will be equal to $\frac{1}{r_f}$. Of course, if in practice such an asset does not exist, we can augment the primitive securities payoff space by a synthetic risk-free asset. We provide a corollary to Theorem 1 that simplifies the dual optimization problem by taking into account the existence of this risk-free asset.

Corollary 1. *Assuming that there is a risk-free asset among the primitive securities then the dual optimization problem in Eq. (3.7) can be simplified to (by also eliminating the constant term):*

$$v_{CR}^{co}(\theta) = \max_{\tilde\lambda \in \Re^{N-1}} \tilde\lambda' q^{co} - E\left\{\frac{(1+\gamma\tilde\lambda' x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(\theta)-1)\tilde\lambda' x\right\}, \tag{3.10}$$

where q^{co} is the vector of prices of the $N-1$ remaining primitive securities other than the risk-free asset. The corresponding admissible SDF that solves this problem is given by:

$$m_{CR}^{co}(\theta) = y(\theta) - 1 + (1 + \gamma\tilde\lambda_*' x)^{\frac{1}{\gamma}}, \tag{3.11}$$

where $\tilde\lambda_*$ is the solution of the optimization problem (3.10).

Proof of Corollary 1. To prove this corollary just observe that the risk-free asset has a constant payoff equal to 1, which allows the separation of the maximization in two parts:

$$v_{CR}^{co}(\theta) = \max_{\tilde\lambda\in\Re^{N-1}, \alpha\in\Re} \frac{1}{r_f}\cdot\alpha + \tilde\lambda' q^{co} - E\left\{\frac{(\gamma\alpha + \gamma\tilde\lambda' x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(\theta)-1)(\tilde\lambda' x + \alpha)\right\}. \tag{3.12}$$

By taking the derivative of Eq. (3.12) with respect to α, eliminating $\tilde\lambda$ and equating to zero, we obtain the concentrated value $\alpha^* = \frac{(\frac{1}{r_f} - E_\mu[y(\theta)-1])^\gamma}{\gamma}$. For the particular case where the proxy model y prices the risk-free asset, α^* becomes $\frac{1}{\gamma}$ and by substituting α^* in (3.12) and by eliminating constant terms (not depending on $\tilde\lambda$) the result follows. □

Corollary 2. *If in Theorem 1 we substitute the optimization problem (3.1) by (3.2), then:*

Problem (3.2) specializes to:

$$\delta_{CR}^+(\theta) = \min_{m\in L^2(y)} E\left\{\frac{(1+m-y(\theta))^{\gamma+1}-1}{\gamma(\gamma+1)}\right\}$$
$$\text{subject to } E(mx) = q,\ m \gg 0. \tag{3.13}$$

The corresponding GEL problem dual to the MD problem is given by:

$$v_{CR}^+(\theta) = \max_{\lambda\in\Re^N} \lambda' q - E\left\{\left(\frac{(\gamma\lambda' x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(\theta)-1)\lambda' x\right) \times I_{\left\{(\gamma\lambda' x)^{\frac{1}{\gamma}} + (y(\theta)-1) > 0\right\}}\right\} \tag{3.14}$$

and the admissible SDF which is closest to the asset pricing proxy y is given by:

$$m_{CR}^+(\theta) = \left(y(\theta) - 1 + (\gamma\lambda_*' x)^{\frac{1}{\gamma}}\right)^+, \tag{3.15}$$

where λ_* is the solution of the optimization problem (3.14), and $I\{\cdot\}$ represents a set indicator function.

Proof of Corollary 2. We follow the proof of Theorem 1 noticing that due to the positivity constraint of the admissible SDFs, the supremum of $H(x) = zx - \phi(1 + x - y(\theta))$ should be taken in the interval $[\max(0, y(\theta)-1), \infty)$. The critical point obtained from differentiating H is $x^* = y(\theta) - 1 + (\gamma z)^{\frac{1}{\gamma}}$ (note that $x^* > y(\theta) - 1$ for any z in the domain of ϕ^*). Now, as $H'(x)$ changes sign at $x = x^*$, if $x^* > 0$, this point is a maximum and the solution is identical to the one in Theorem 1. However, as $H'(x) < 0$ for all x such that $x > x^*$, if $x^* < 0$ then $y(\theta) - 1 < 0$, $H(\cdot)$ is decreasing in the whole domain $[0, \infty)$, and zero is a maximum of the function. Thus $x^{\sup} = (y(\theta) - 1 + (\gamma z)^{\frac{1}{\gamma}})^+$, and consequently $\phi^*(z, y(\theta)) = \left(\frac{(\gamma z)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(\theta)-1)z + \frac{(1-y(\theta))^{\gamma+1}}{\gamma(\gamma+1)}\right) I_{\{(\gamma z)^{\frac{1}{\gamma}} + (y(\theta)-1) > 0\}} + \frac{1}{\gamma(\gamma+1)} - \frac{(1-y(\theta))^{\gamma+1}}{\gamma(\gamma+1)}$. By eliminating terms that do not depend on z in ϕ^*, and applying it in Eq. (3.3) we obtain the optimization problem (3.14). The first-order conditions of this problem with respect to λ are:

$$\frac{\partial v_{CR}}{\partial \lambda} = q - E\left\{\left(y(\theta) - 1 + (\gamma\lambda_*' x)^{\frac{1}{\gamma}}\right)^+ x\right\} = 0 \tag{3.16}$$

showing that m_{CR} in Eq. (3.15) is an admissible SDF that minimizes the MD problem (3.2) when the divergence is a member of the Cressie–Read family. □

At this point, it can be immediately stated without proof:

Corollary 3. *If there is a risk-free asset in the economy and we are searching for the closest SDF in the MD ϕ-sense within the set* *of nonnegative SDFs M+, the problem can be solved via the dual problem:*

$$v_{\mathrm{CR}}^{co+}(\theta) = \max_{\lambda\in\Re^{N-1}} \lambda' q^{co} - E\left\{\left(\frac{(1+\gamma\lambda' x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(\theta)-1)\lambda' x\right) \times I_{\{(1+\gamma\lambda' x)^{\frac{1}{\gamma}}+(y(\theta)-1)>0\}}\right\}, \qquad (3.17)$$

where we have suppressed the variables constant in λ for simplification. In addition, the admissible SDF which is closest to the asset pricing proxy $y(\theta)$ is given by:

$$m_{\mathrm{CR}}^{co,+}(\theta) = \left(y(\theta) - 1 + (1+\gamma\lambda_*' x)^{\frac{1}{\gamma}}\right)^+, \qquad (3.18)$$

where λ_* is the solution of the optimization problem (3.17).

3.2. Generalized Minimum Contrast estimators and MD SDF problems

Suppose that the econometrician observes IID realizations of a $\Re^N$ random variable z with probability law μ, and that he is interested in the model defined by the following set of moment conditions[14]:

$$E[g(z,\theta)] = \int g(z,\theta)d\mu = 0, \quad \theta\in\Re^k. \qquad (3.19)$$

Given a general convex function ϕ, define a function D that will measure the discrepancy between two probability measures P and Q, with P absolutely continuous with respect to Q, by:

$$D(P,Q) = \int \phi\left(\frac{dP}{dQ}\right) dQ \qquad (3.20)$$

where $\frac{dP}{dQ}$ represents the Radon–Nikodym derivative of P with respect to Q.

A GMC estimator will search for the parameter vector θ giving a probability measure $\pi \in P(\theta) = \{\pi$ probability in $\Re^N$ such that $\int g(z,\theta)d\pi = 0\}$ that minimizes the discrepancy with respect to the true unknown probability measure μ:

$$V = \inf_{\theta\in\Theta} \inf_{\pi\in P(\theta)} D(\pi,\mu). \qquad (3.21)$$

To interpret our MD problem in the context of GMC estimators, assume that μ is the unknown probability law generating the observable payoffs x of the primitive securities. Given the family M of admissible SDFs m pricing x, our MD problem finds the admissible SDF that is as close as possible (in the ϕ-sense) to the asset pricing proxy y. This solution is achieved by constructing a probability measure $\tilde{\pi}$ with Radon–Nikodym derivative $\frac{d\tilde{\pi}}{d\mu} = \frac{1+m-y(\theta)}{E_\mu(1+m-y(\theta))}$, and searching among the admissible SDFs m the one that generates the $\tilde{\pi}$ closest to the unknown probability μ.[15] Defining $\tilde{P}(\theta) = \{\tilde{\pi}$ prob. in $\Re^N$ such that : $(\frac{d\tilde{\pi}}{d\mu} + y(\theta) - 1) \in M\}$, our MD problem is equivalent to the following GMC problem:

$$V = \inf_{\theta\in\Theta} \inf_{\tilde{\pi}\in\tilde{P}(\theta)} D(\tilde{\pi},\mu). \qquad (3.22)$$

The solution to this problem according to Corollary 1 is given by:

$$\frac{d\tilde{\pi}_{\mathrm{CR}}}{d\mu} = \frac{1+m_{\mathrm{CR}}-y(\theta)}{E_\mu(1+m_{\mathrm{CR}}-y(\theta))} = \frac{(1+\gamma\lambda_*' x)^{\frac{1}{\gamma}}}{E_\mu(1+\gamma\lambda_*' x)^{\frac{1}{\gamma}}}. \qquad (3.23)$$

A normalized version of the Radon–Nikodym derivative $\frac{d\tilde{\pi}_{\mathrm{CR}}}{d\mu}$, on its sample version, generates strictly positive $\pi_{\mathrm{CR},i}$'s that are denominated implied probabilities.[16]

$$\pi_{\mathrm{CR},i} = \frac{(1+\gamma\tilde{\lambda}_*' x_i)^{\frac{1}{\gamma}}}{\sum_{j=1}^{T}(1+\gamma\tilde{\lambda}_*' x_j)^{\frac{1}{\gamma}}}, \quad i=1,\ldots,T. \qquad (3.24)$$

Implied probabilities are useful in a variety of applications. For instance, Brown and Newey (1998) obtain an efficient estimation of moment conditions based on implied probabilities. Brown and Newey (2002) suggest their use on the estimation of probability distribution functions via bootstrapping schemes. Smith (2004) shows how these probabilities can be used to obtain efficient moment estimation for GEL estimators. Antoine et al. (2007) show, in a context of Euclidean Likelihood, how implied probabilities contain important information coming from overidentifying restrictions that can be used to decrease the variance of the estimator. In an asset pricing context, Almeida and Garcia (2008) show that implied probabilities can be used to derive nonparametric bounds for SDFs that by construction take into account an arbitrary combination of moments of returns from primitive securities.

3.3. Interpreting the dual optimization problem

In a seminal work, Stutzer (1995) proposed a portfolio interpretation for the ET estimator based on a standard two-period model of optimal portfolio choices (see Huang and Litzenberger, 1988). He showed that the ET entropy minimization problem corresponds to an optimal portfolio problem with a CARA utility function. Based on the same two-period model, Almeida and Garcia (2008) extended his interpretation to the whole Cressie–Read family in a nonparametric setting. Here we generalize Almeida and Garcia (2008) interpretation to dual optimization problems in a semi-parametric setting.

The dual MD optimization problem appearing in Corollary 1 can be interpreted as an allocation problem among the N basis assets with the objective to maximize the utility function U defined by:

$$u(W, y(\theta)) = -\frac{1}{\gamma+1}(1-\gamma W)^{\left(\frac{\gamma+1}{\gamma}\right)} + (y(\theta)-1)W. \qquad (3.25)$$

Suppose an investor distributes his/her initial wealth W_0 putting λ_j units of wealth on the risky asset R_j and the remaining $W_0 - \sum_{j=1}^{N}\lambda_j$ in a risk-free asset paying $r_f = \frac{1}{a}$. Terminal wealth is then $W = W_0 * r_f + \sum_{j=1}^{N}\lambda_j * (R_j - r_f)$. Assume in addition that this investor maximizes the utility function provided above in Eq. (3.25), solving the following optimal portfolio problem:

$$\Omega = \sup_{\lambda\in\Lambda} E\,(u(W)) \qquad (3.26)$$

where $\Lambda = \{\lambda : u(W(\lambda))$ is strictly increasing and concave$\}$, and expectation is taken with respect to W and $y(\theta)$.

[14] We follow Kitamura (2006b, Section 3) to formalize the concept of a Generalized Minimum Contrast (GMC) estimator. For a sample version of GMC estimators, see Corcoran (1998).

[15] Note that although we do not restrict m to be strictly positive, due to the restriction $1+m-y(\theta) \gg 0$ (see Section 3.1), we guarantee that $\frac{1+m-y(\theta)}{E_\mu(1+m-y(\theta))}$ is indeed a Radon–Nikodym derivative.

[16] Our implied probabilities are strictly positive due to the restrictions in our optimization problems that guarantee that $1+m-y(\theta) \gg 0$. Back and Brown (1993) provide GMM implied probabilities, Owen (1988) for EL, Kitamura and Stutzer (1997) for ET, and Imbens et al. (1998) for Cressie–Read discrepancy estimators. Brown and Newey (2002) provide implied probabilities for GEL estimators.

Note that the utility is composed of two terms, one linear in wealth and the other given by a HARA utility. The linear term corresponds to a risk-neutral economy with SDF $y(\theta) - 1$. If the term $y(\theta) - 1$ did not exist, by scaling the original vector λ to be $\tilde{\lambda} = \frac{\lambda}{(1+\gamma W_0.rf)}$, we would be able to decompose the utility function in $u(W) = u(W_0 * r_f) * \left(1 + \gamma \tilde{\lambda}\left(R - \frac{1}{a}\right)\right)^{\left(\frac{\gamma+1}{\gamma}\right)}$. This decomposition would essentially show that solving a GEL optimal problem for excess returns would measure the gain when switching from a total allocation of wealth at the risk-free asset paying r_f to an optimal diversified allocation (in the utility u sense) that includes both risky assets and the risk-free asset. $y(\theta) - 1$ works as a penalty factor to avoid the resulting marginal utility from the HARA part to be too far from the translated proxy $y(\theta) - 1$. This penalty factor is compatible with the fact that in the original MD problem we are searching among all admissible SDFs the one that satisfies a convexity criterion of proximity to the asset pricing model y.[17]

3.4. Some special Cressie–Read discrepancies

In this section we specialize the results in Theorem 1 to provide the dual optimization problems and corresponding admissible SDFs solutions for some special discrepancies in the Cressie–Read family frequently adopted in the econometric literature. We begin by investigating the relation between Euclidean Likelihood and the HJ (1997) distance. In a sequence, we provide the optimization problems and solutions under EL (CR with $\gamma = -1$) and ET (CR with $\gamma = 0$) discrepancies. The other two discrepancies adopted in the empirical section, namely Pearson's Chi-Square (Cressie–Read with $\gamma = -2$) and Hellinger's distance (CR with $\gamma = -\frac{1}{2}$), can be obtained directly by application of Theorem 1 with their specific γ's.

3.4.1. Hansen and Jagannathan distance derived from Euclidean Likelihood

Euclidean likelihood or CUE is obtained by fixing $\gamma = 1$ on the Cressie–Read discrepancy. By using this value of gamma in Corollary 1 above and dropping the constant terms, we obtain the following optimization problem[18]:

$$\begin{aligned} v^{co}_{CUE}(\theta) &= \max_{\lambda \in \Re^{N-1}} \lambda' q - E\left\{\frac{1}{2}(1+\lambda' x)^2 + (y(\theta) - 1)\lambda' x\right\} \\ &= \max_{\lambda \in \Re^{N-1}} \lambda' q - E\left\{\frac{1}{2}(\lambda' x)^2 + y\lambda' x\right\}. \end{aligned} \tag{3.27}$$

From the first-order conditions, we obtain $m_{CUE} = y(\theta) + \tilde{\lambda}'_* x$, which is exactly the linear correction term found by HJ (1997). By comparing Eq. (3.27) to Eq. (2.7) we note that the two problems are equivalent. Thus, with a slight modification of our proposed MD problem, the HJ distance becomes one element within the particular Cressie–Read family.

[17] The term $y(\theta) - 1$ appears instead of $y(\theta)$ because our MD problems are based on original GMC problems were the elements in search are probability measures. More precisely, while GMC problems search for the absolutely continuous probability measure that is closest to the unknown true probability measure μ, we convert our search for an SDF m that is closest to the asset pricing model $y(\theta)$, to a search of an absolutely continuous probability measure $\tilde{\pi}$ with Radon–Nikodym derivative $\frac{d\tilde{\pi}}{d\mu} = 1 + m - y(\theta)$, that is closest to the unknown true probability μ (see Section 3.2). In this sense, -1 works simply as a scaling factor. Note that when $y(\theta) = 1$, there is no more a proxy model and we go back to the nonparametric case studied in Almeida and Garcia (2008).

[18] To obtain a precise equivalence with respect to HJ (1997), in this particular case we do not restrict $1 + m - y(\theta)$ to be strictly positive, meaning that it will not necessarily be a Radon–Nikodym derivative.

3.4.2. Empirical Likelihood ($\gamma = -1$)

In this limiting interesting case, the Cressie–Read discrepancy converges to $\phi(\pi) = -\ln(\pi)$. Our MD problem becomes:

$$\begin{aligned} \delta_{EL}(\theta) &= \min_{m \in L^2} E\{-\ln(1 + m - y(\theta))\} \\ &\quad \text{subject to } E(mx) = q. \end{aligned} \tag{3.28}$$

Noting that the expression $\frac{(1+\gamma x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + \frac{1}{\gamma(\gamma+1)}$ converges to $-1 - \ln(1-x)$ when $\gamma \to -1$, and applying the results in Corollary 1, the dual optimization problem becomes:

$$\begin{aligned} v^{co}_{EL}(\theta) &= \max_{\tilde{\lambda} \in \Re^{N-1}} \tilde{\lambda}' q^{co} \\ &\quad - E\left\{-\ln(1 - \tilde{\lambda}' x) + (y(\theta) - 1)\tilde{\lambda}' x\right\}. \end{aligned} \tag{3.29}$$

The corresponding admissible SDF that solves this problem is given by:

$$m^{co}_{EL}(\theta) = y(\theta) - 1 + \frac{1}{(1 - \tilde{\lambda}'_* x)}. \tag{3.30}$$

3.4.3. Exponential Tilting ($\gamma = 0$)

The ET discrepancy is also a limiting case on the Cressie–Read family studied by Kitamura and Stutzer (1997). The Cressie–Read discrepancy converges in this case to $\phi(\pi) = \pi \ln \pi$, with convex conjugate e^{z-1}.

Then the MD problem becomes:

$$\begin{aligned} \delta_{ET}(\theta) &= \min_{m \in L^2} E\{(1 + m - y(\theta)) \ln(1 + m - y(\theta))\} \\ &\quad \text{subject to } E(mx) = q. \end{aligned} \tag{3.31}$$

Noting that the expression $\frac{(1+\gamma x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + \frac{1}{\gamma(\gamma+1)}$ converges to e^x when $\gamma \to 0$ and applying the results in Corollary 1, the dual optimization problem becomes:

$$v^{co}_{ET}(\theta) = \max_{\tilde{\lambda} \in \Re^{N-1}} \tilde{\lambda}' q^{co} - E\left\{e^{\tilde{\lambda}' x} + (y(\theta) - 1)\tilde{\lambda}' x\right\}. \tag{3.32}$$

The corresponding admissible SDF that solves this problem is given by:

$$m^{co}_{ET}(\theta) = y(\theta) - 1 + e^{\tilde{\lambda}'_* x}. \tag{3.33}$$

3.5. Model estimation based on Minimum Discrepancy bounds

Researchers have been using the HJ (1997) distance to estimate asset pricing models by finding the parameter vector θ^* that minimizes this distance. Following Kitamura (2006b) and the whole literature on MD estimators, we propose estimating the above asset pricing models by finding the parameter vector θ_{MD} that minimizes any specific discrepancy function either described by Eq. (3.1) or (3.2):

$$\theta_{MD} = \operatorname*{argmin}_{\theta \in \Re^k} \delta_{MD}(\theta), \tag{3.34}$$

or

$$\theta^+_{MD} = \operatorname*{argmin}_{\theta \in \Re^k} \delta^+_{MD}(\theta). \tag{3.35}$$

Note that, according to the results presented in Section 3.1, these problems can be expressed as min–max optimization problems:

$$\theta_{\text{MD}} = \underset{\theta\in\Re^k}{\text{argmin}}\, \upsilon_{\text{MD}}(\theta) = \underset{\theta\in\Re^k}{\text{argmin}}\, \underset{\lambda\in\Re^N}{\max}\, \lambda' q - E\{\phi^*(\lambda' x, y(\theta))\}, \tag{3.36}$$

$$\theta^+_{\text{MD}} = \underset{\theta\in\Re^k}{\text{argmin}}\, \upsilon^+_{\text{MD}}(\theta) = \underset{\theta\in\Re^k}{\text{argmin}}\, \underset{\lambda\in\Re^N}{\max}\, \lambda' q - E\{\phi^{*,+}(\lambda' x, y(\theta))\}. \tag{3.37}$$

For any fixed θ_0 in the parameter space Θ, the inner problem will deliver an admissible SDF closest (in the ϕ sense) to $y(\theta_0)$, obtained via a set of Lagrange Multipliers λ_0 representing portfolio weights from a HARA utility function (see Section 3.3).

When the discrepancy is a member of the Cressie–Read family, the equivalent problems are:

$$\theta_{\text{CR}} = \underset{\theta\in\Re^k}{\text{argmin}}\, \delta_{\text{CR}}(\theta), \tag{3.38}$$

$$\theta^+_{\text{CR}} = \underset{\theta\in\Re^k}{\text{argmin}}\, \delta^+_{\text{CR}}(\theta), \tag{3.39}$$

$$\theta_{\text{CR}} = \underset{\theta\in\Re^k}{\text{argmin}}\, \upsilon_{\text{CR}}(\theta), \tag{3.40}$$

$$\theta^+_{\text{CR}} = \underset{\theta\in\Re^k}{\text{argmin}}\, \upsilon^+_{\text{CR}}(\theta), \tag{3.41}$$

where $\delta_{\text{CR}}(\theta)$, $\upsilon_{\text{CR}}(\theta)$, $\delta^+_{\text{CR}}(\theta)$, $\upsilon^+_{\text{CR}}(\theta)$ are respectively defined in Eqs. (3.6), (3.7), (3.13) and (3.14).

When a risk-free rate is assumed to exist in the economy, we provide below the dual problems that should be solved to estimate the parameters θ when the discrepancy belongs to the Cressie–Read family[19]:

$$\theta_{\text{CR}} = \underset{\theta\in\Re^k}{\text{argmin}}\ \underset{\lambda\in\Re^{N-1}}{\max}\ \lambda' q - E\left\{\frac{(1+\gamma\lambda' x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(\theta)-1)\lambda' x + \frac{1}{\gamma(\gamma+1)}\right\}, \tag{3.42}$$

$$\theta^+_{\text{CR}} = \underset{\theta\in\Re^k}{\text{argmin}}\ \underset{\lambda\in\Re^{N-1}}{\max}\ \lambda' q - E\left\{\left(\frac{(1+\gamma\lambda' x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(\theta)-1)\lambda' x\right) I_{\left\{(1+\gamma\lambda' x)^{\frac{1}{\gamma}}+y(\theta)-1>0\right\}}\right\}. \tag{3.43}$$

4. Properties of the estimators

In order to be able to perform hypothesis tests with the new proposed discrepancy measures, we must provide the statistical properties of our estimators. The asymptotic properties of the estimators proposed by HJ (1997) have been studied by HJ (1997) and by HHL (1995). More recently, Kan and Robotti (2009) offered a detailed analysis of linear asset pricing models, extending previous results to consider formal model comparisons in terms of their HJ distances, for either correctly specified or misspecified models.

For MD estimators that do no rely on a quadratic criterion, Kitamura and Stutzer (1997) derived asymptotic properties for the ET estimator, while Kitamura (2000) extended the results in Kitamura and Stutzer (1997) to consider possibly misspecified models. Newey and Smith (2004) obtained the corresponding properties of consistency and asymptotic normality for members of the Cressie–Read family, under the assumption of i.i.d. data, and of correct model specification.

In this section, we analyze the asymptotic properties of our MD estimators considering that the asset pricing models analyzed are misspecified. We derive properties for the particular case where the estimators belong to the family of Cressie–Read discrepancies.[20] Under the same set of assumptions provided by Kitamura and Stutzer (1997) and Kitamura (2000), we prove consistency and asymptotic normality of our estimators.

4.1. Consistency of estimators

We derive the asymptotic properties of the estimator that is obtained from Eq. (3.42).[21] To accommodate cases where the asset pricing model depends on factors that are not payoffs (for example consumption), we assume a more general dependence of y on $z = [x, u]$ where x represent basis assets as before and u represent factors that are not on the payoff space. For each fixed γ in the family of Cressie–Read discrepancies, define the following functions:

$$f^\gamma(\theta, \lambda, z) = -\lambda' q + \frac{(1+\gamma\lambda' x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(z,\theta)-1)\lambda' x, \tag{4.1}$$

$$\mathfrak{M}^\gamma(\theta,\lambda,z) = -\upsilon^{\text{co}}_{\text{CR}}(\theta) = E\{f^\gamma(\theta,\lambda,z)\} = -\lambda' q + E\left\{\frac{(1+\gamma\lambda' x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(z,\theta)-1)\lambda' x\right\}, \tag{4.2}$$

and its corresponding version for a sample of size n:

$$\mathfrak{M}^\gamma_n(\theta,\lambda,z) = -\lambda' q + \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{(1+\gamma\lambda' x_t)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(z_t,\theta)-1)\lambda' x_t\right\}. \tag{4.3}$$

For any $\theta \in \Theta$, the dual parameters (Lagrange Multipliers) λ from Eq. (3.42) can be obtained by:

$$\lambda(\theta) = \underset{\lambda\in\Re^{N-1}}{\text{argmin}}\, \mathfrak{M}^\gamma(\theta,\lambda,z), \tag{4.4}$$

and its corresponding estimator for a sample of size n is:

$$\hat{\lambda}_n(\theta) = \underset{\lambda\in\Re^{N-1}}{\text{argmin}}\, \mathfrak{M}^\gamma_n(\theta,\lambda,z). \tag{4.5}$$

Also note by rewriting Eq. (3.42) as a function of $\mathfrak{M}^\gamma(\theta,\lambda,z)$ that the pseudo-true value for the parameter θ can be obtained by:

$$\theta_* = \underset{\theta\in\Theta}{\text{argmax}}\, \mathfrak{M}^\gamma(\theta,\lambda(\theta),z), \tag{4.6}$$

with $\lambda(\theta)$ defined by Eq. (4.4).

[19] Due to its importance in empirical work, we also choose to analyze its statistical properties in the following section.

[20] The derivation of these properties for more general MD estimators should follow precisely the same principles.

[21] For asymptotic properties of the estimator obtained from Eq. (3.43), that restricts the SDF space to nonnegative SDFs, we refer the reader to HHL (1995) for the quadratic case of $\gamma = 1$. Apparently for a generalization of their results on nonnegative SDFs for an arbitrary member of the Cressie–Read family, concavity of the dual problems is the key condition, and such condition is satisfied by the whole family.

The corresponding MD estimator for θ_* is given by:

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\operatorname{argmax}}\, \mathfrak{M}_n^{\gamma}(\theta, \hat{\lambda}_n(\theta), z), \quad (4.7)$$

with $\hat{\lambda}_n(\theta)$ defined by Eq. (4.5).

Let λ_* denote the pseudo-true Lagrange multipliers, the value of the vector λ that is obtained in Eq. (4.4) when $\theta = \theta_*$:

$$\lambda_* = \lambda(\theta_*). \quad (4.8)$$

Letting $\Gamma(\theta, \Delta)$ denote an open sphere with center θ and radius Δ, building on Kitamura and Stutzer (1997) we make the following set of assumptions:

Assumption 1. (a.1) The process z_t is stationary and ergodic.

(a.2) The process z_t is strong mixing with mixing coefficients α_n satisfying $\sum_{n=1}^{\infty} \alpha_n^{1-\frac{1}{b}} \langle\infty, b\rangle 1$;

(b) The set of parameters Θ is a compact k-dimensional set.

(c) For sufficiently small $\Delta > 0$, $E\left[\sup_{\theta\in\Gamma(\bar{\theta},\Delta)} d'q - \left\{\frac{(1+\gamma d'x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(z,\theta)-1)d'x\right\}\right] < \infty$, for all vectors d in a neighborhood of λ_*, for any $\bar{\theta} \in \Theta$.

(d) $E(xx')$ is nonsingular.

(e) If a sequence θ_j, $j = 1, 2, \ldots$ converges to some $\theta \in \Theta$ as $j \to \infty$, then $y(z, \theta_j)$ converges to $y(z, \theta)$, for all z except for a null set that may vary with θ.

(f) There is a unique θ_* solving Eq. (4.6).

Before stating the main theorem on the consistency of the estimators for the parameters θ, we prove an auxiliary lemma that shows the consistency of the Lagrange Multipliers λ appearing in Eq. (4.5).

Lemma 1. *Under Assumption 1, $\hat{\lambda}_n(\theta_*)$ converges to the pseudo-true value λ_* in probability.*

Proof of Lemma 1. We follow Kitamura (2000) and make use of Theorem 2.7 of Newey and McFadden (1994), which for completeness, we present within our proof:

(Theorem 2.7 Newey and McFadden, 1994). Suppose that there is a function $Q_0(\lambda)$ such that: (i) $Q_0(\lambda)$ is uniquely maximized at λ_0; (ii) λ_0 is an element of the interior of a convex set Λ, and $\hat{Q}_n(\lambda)$ is concave, where $\hat{Q}_n(\lambda)$ represents the corresponding sample estimator for $Q_0(\lambda)$ on a sample of size n; and (iii) $\hat{Q}_n(\lambda) \xrightarrow{p} Q_0(\lambda)$ for all $\lambda \in \Lambda$. Then, there exists, with probability approaching one, a sequence $\hat{\lambda}_n$ of parameter estimates satisfying $\hat{\lambda}_n \xrightarrow{p} \lambda_0$.

Now, if we set $Q_0(\lambda) = -\mathfrak{M}^{\gamma}(\theta, \lambda(\theta), z)$ and $\hat{Q}_n(\lambda) = -\mathfrak{M}_n^{\gamma}(\theta, \lambda(\theta), z)$, let us show that they satisfy conditions (i)–(iii) of Newey and McFadden's theorem.

Assumption 1(d) implies condition (i). Indeed, for any fixed $\theta \in \Theta$, uniqueness of $\lambda(\theta)$ in Eq. (4.4) is guaranteed since the Jacobian of the first-order condition for $-\mathfrak{M}^{\gamma}$ with respect to λ given by $-E(xx'(1 + \gamma\lambda(\theta)'x)^{\frac{1}{\gamma}-1})$ is nonsingular and negative definite, implying that $-\mathfrak{M}^{\gamma}(\theta, \lambda(\theta), z)$ is strictly concave in λ. In addition, the strict concavity of $-\mathfrak{M}^{\gamma}(\theta, \lambda(\theta), z)$ and the fact that $\Lambda = \Re^{N-1}$ (see Eq. (4.4)) also imply that $\lambda(\theta)$ is an element on the interior of a convex set. As for the second part of condition (ii), we note that $-\mathfrak{M}_n^{\gamma}(\theta, \lambda(\theta), z)$ is concave since the Jacobian of its first-order condition with respect to λ is given by $-\sum_{t=1}^{n}(x_t x_t'(1+\gamma\lambda(\theta)'x_t)^{\frac{1}{\gamma}-1})$ which is negative definite.

Lemma 1 in Hong et al. (2003) show that condition (iii) is satisfied if the following four (sufficient) conditions are valid:

(1) $\lambda(\theta)$ in Eq. (4.4) is continuous in θ;

(2) $\frac{\partial f(\theta,\lambda)}{\partial\lambda} = -q + (1 + \gamma\lambda'x)^{\frac{1}{\gamma}}x + (y(z, \theta) - 1)x$ is uniformly continuous in θ, $f(.,.)$ given by (4.1);

(3) $E\left[\sup_{\theta\in\Theta,\lambda\in\Re^{N-1}} \lambda q - \left\{\frac{(1+\gamma\lambda'x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(z,\theta)-1)\lambda'x\right\}\right] < \infty$;

(4) $\{z_t\}_{t=1}^{n}$ are i.i.d. random variables.

Condition (1), that is, the continuity of λ as a function of θ, is implied by Assumption 1(d) since we have seen that it guarantees that the Jacobian of the f.o.c. for $-\mathfrak{M}^{\gamma}$ is nonsingular. Condition (2) is implied by Assumption 1(b) and (e). While Assumption 1(e) gives the continuity of function $y(z, \theta)$, Assumption 1(b) imposes that Θ is a compact set, directly implying that $y(z, \theta)$ is uniformly continuous in Θ and consequently that $\frac{\partial f(\theta,\lambda)}{\partial\lambda}$ is uniformly continuous in Θ. Condition (3) is guaranteed by Assumption 1(b) and (c) since a compact set can be covered by a finite number of open spheres, and for each sphere in the cover, moments of the supremum are finite by Assumption 1(c). Finally, Kitamura (2000) shows that condition (4) of i.i.d. random variables can be substituted by Assumption 1(a.2) of a strongly mixing condition for process z_t. Therefore, $-\mathfrak{M}^{\gamma}$ satisfies the three conditions of Newey and McFadden's theorem, that proves the consistency of $\hat{\lambda}_n(\theta_*)$. □

Theorem 2. *Under Assumption 1, $\hat{\theta}_n$ converges to the pseudo-true value θ_* in probability.*

Proof of Theorem 2. From Assumption 1(f) we have that[22]:

$$\mathfrak{M}^{\gamma}(\theta, \lambda(\theta), z) < \mathfrak{M}^{\gamma}(\theta_*, \lambda(\theta_*), z) \le \frac{1}{1+\gamma} I_{\{\gamma\in\Re,\gamma\neq-1\}}, \quad \text{for all } \theta \in \Theta,\ \theta \neq \theta_*, \quad (4.9)$$

where $I_{\{A\}}$ is the indicator function for set A. For each pair $(\bar{\theta}, \Delta)$, define the following random variable $M(\bar{\theta}, \Delta) = \sup_{\theta\in\Gamma(\bar{\theta},\Delta)} -\lambda(\theta)'q + \{\frac{(1+\gamma\lambda(\theta)'x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(z, \theta) - 1)\lambda(\theta)'x\}$. Assumption 1(c) guarantees that $E\left\{M(\bar{\theta}, \Delta)\right\} < \infty$. Assumption 1(d) and (e) guarantee continuity of $\lambda(\theta)$ and $y(z, \theta)$, which coupled with the compactness imposed by Assumption 1(b), guarantees that $\lim_{\Delta\to 0} M(\bar{\theta}, \Delta) = -\lambda(\bar{\theta})'q + \{\frac{(1+\gamma\lambda(\bar{\theta})'x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(z, \bar{\theta}) - 1)\lambda(\bar{\theta})'x\}$. Applying the Dominated Convergence Theorem we obtain that:

$$\lim_{\Delta\to 0} E\left\{M(\bar{\theta}, \Delta)\right\} = \mathfrak{M}^{\gamma}(\bar{\theta}, \lambda(\bar{\theta}), z). \quad (4.10)$$

By Eqs. (4.9), (4.10), by the stationarity and ergodicity of z from Assumption 1(a.1), and the fact that $\Theta - \Gamma(\theta_*, \Delta)$ is a compact set that can be covered by a finite number of open spheres, we obtain the following inequality (see Kitamura and Stutzer, 1997 for a detailed derivation): For each $\epsilon > 0$ there exist $\Delta > 0$ and $h > 0$ such that:

$$\text{Prob}\left\{\sup_{\theta\in\Theta-\Gamma(\theta_*,\Delta)} \mathfrak{M}_n^{\gamma}(\theta, \hat{\lambda}(\theta), z) > \mathfrak{M}^{\gamma}(\theta_*, \lambda_*, z) - h\right\} < \frac{\epsilon}{2}, \quad \text{for sufficiently large } n. \quad (4.11)$$

Now, Assumption 1(a.2) and (c), and the consistency of $\hat{\lambda}_n(\theta_*)$ imply that:

$$\text{Prob}\left\{\mathfrak{M}_n^{\gamma}(\theta_*, \hat{\lambda}(\theta_*), z) < \mathfrak{M}^{\gamma}(\theta_*, \lambda_*, z) - \frac{h}{2}\right\} < \frac{\epsilon}{2} \quad \text{for sufficiently large } n. \quad (4.12)$$

[22] If a model y is correctly specified then $\lambda_* = 0$, and $M(\theta_*, 0) = \frac{1}{1+\gamma}$, if $\gamma \neq -1$ and $M(\theta_*, 0) = 0$ for $\gamma = -1$.

The probabilities in Eqs. (4.11) and (4.12) imply that $\hat{\theta}_n \xrightarrow{p} \theta_*$.[23] □

Now we turn our attention to the estimator of the MD Cressie–Read distance δ_{CR}, which can be obtained as a function of $\mathfrak{M}^\gamma$:

$$\delta_{CR} = \max_{\theta \in \Theta} \mathfrak{M}^\gamma(\theta, \lambda(\theta), z). \tag{4.13}$$

Its corresponding estimator for a sample of size n is:

$$\hat{\delta}_{n\,CR} = \max_{\theta \in \Theta} \mathfrak{M}_n^\gamma(\theta, \hat{\lambda}_n(\theta), z). \tag{4.14}$$

Knowledge of the asymptotic distribution of the estimator for the MD distance δ_{CR} will be important to assess the statistical significance of the degree of model misspecification. In what follows we state a result of consistency for the estimator $\hat{\delta}_{n\,CR}$:

Theorem 3. *Under Assumption 1, $\hat{\delta}_{n\,CR}$ converges to δ_{CR} in probability.*

Proof of Theorem 3. In Lemma 1, we proved: (i) The continuity and uniqueness of $\lambda(\theta)$ as a function of θ, (ii) That $E\left[\sup_{\theta\in\Theta,\lambda\in\Re^{N-1}} \lambda q - \{\frac{(1+\gamma\lambda' x)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(z,\theta)-1)\lambda' x\}\right] < \infty$, and (iii) The uniform continuity of $\frac{\partial f(\theta,\lambda)}{\partial \lambda}$ with respect to θ. Now, according to Lemma 1 of Hong et al. (2003), these three properties coupled with Assumption 1(a.2), and the compactness of the space of parameters Θ given by Assumption 1(b), are sufficient conditions to guarantee that:

$$\max_{\theta\in\Theta} \mathfrak{M}_n^\gamma(\theta, \hat{\lambda}(\theta), z) \xrightarrow{p} \mathfrak{M}^\gamma(\theta_*, \lambda_*, z), \tag{4.15}$$

which is the desired result. □

4.2. Asymptotic normality

In this section, we first want to show that the consistent estimators that we described in the previous section are also asymptotically normal. As it is usual, we start by the first-order condition $g(\widehat{\Phi}) = \frac{\partial}{\partial \Phi'} \mathfrak{M}_n^\gamma(\widehat{\Phi}, z)$, where $\hat{\Phi} = [\hat{\theta}', \hat{\lambda}']'$, and perform a short Taylor expansion about the true value of the parameter vector Φ_*. Given some invertibility conditions on the Hessian, and relying on a law of large numbers and a central limit theorem, we show in Appendix that the estimator $\hat{\Phi}$ is indeed asymptotically normal.

Theorem 4. *Under Assumptions 1 and 3 (see Appendix),*

$$\sqrt{n}\left(\hat{\Phi} - \Phi_*\right) \to_d N(0, V_\Phi). \tag{4.16}$$

The V_Φ matrix is given by: $V_\Phi = H_{\Phi\Phi}^{-1} S_\Phi H_{\Phi\Phi}^{-1}$, where $H_{\Phi\Phi}$ is the Hessian matrix and S_Φ the variance–covariance matrix of the score vector. More explicit expressions for the elements of these matrices are given in Appendix.

We next provide the asymptotic distribution for the Cressie—Read distance δ_{CR}. The asymptotic distribution of δ_{CR} will be useful to verify if the distance of a certain asset pricing model $y(z, \theta)$ is or not statistically different from zero. Since our extension to the Cressie–Read family maintains the important properties of concavity of the conjugate problem and convexity of the constraints, we are able to mimic the asymptotic results provided by HHL (1995) when proving our theorem.[24]

Assumption 2. (a) $\sqrt{n}\left(\mathfrak{M}_n^\gamma(\theta_*, \lambda(\theta_*), z) - \mathfrak{M}^\gamma(\theta_*, \lambda(\theta_*), z)\right) \to_d N(0, \sigma^2)$.
(b) $\sqrt{n}(\frac{1}{n}\sum_n m_{CR}^{co}(\theta_*)_n x_n - q - E(m_{CR}^{co}(\theta_*)x - q)) \to_d N(0, \sigma_m^2)$, where m_{CR}^{co} is defined in Eq. (3.11).
(c) The distance $\delta_{CR} \neq 0$ (the model is misspecified).[25]

Theorem 5. *Under Assumptions 1–3,*

$$\sqrt{n}\left(\hat{\delta}_{n\,CR} - \delta_{CR}\right) \to_d N(0, \sigma^2) \tag{4.17}$$

where σ^2 is the asymptotic variance on the central limit approximation of $\mathfrak{M}_n^\gamma(\theta_*, \lambda(\theta_*), z)$ in Assumption 2(a).

Proof of Theorem 5. Following the proof of Proposition 2.2 in HHL (1995), and noticing that $\delta_{CR} = -\mathfrak{M}^\gamma(\theta_*, \lambda(\theta_*), z)$, we decompose the estimator into two terms:

$$\begin{aligned}\sqrt{n}\left(\hat{\delta}_{n\,CR} - \delta_{CR}\right) = {} & \sqrt{n}(-\mathfrak{M}_n^\gamma(\hat{\theta}_n, \lambda(\hat{\theta}_n), z) \\ & + \mathfrak{M}_n^\gamma(\theta_*, \lambda(\theta_*), z)) \\ & + \sqrt{n}(-\mathfrak{M}_n^\gamma(\theta_*, \lambda(\theta_*), z) \\ & + \mathfrak{M}^\gamma(\theta_*, \lambda(\theta_*), z)).\end{aligned} \tag{4.18}$$

We center our analysis at the first term of this decomposition since the second term, according to Assumption 2(a) converges in distribution to $N(0, \sigma^2)$.

Letting $X_n = \sqrt{n}(q - \frac{1}{n}\sum_n m_{CR}^{co}(\hat{\theta}_n)_n x_n - E(q - m_{CR}^{co}(\hat{\theta}_n)x))$, and $Y_n = (\hat{\lambda}_n(\hat{\theta}_n) - \lambda_*)$, the concavity of function δ_{CR} in λ, and the first-order conditions for the population conjugate problem imply the following gradient inequality[26]:

$$0 \le \sqrt{n}\left(-\mathfrak{M}_n^\gamma(\hat{\theta}_n, \lambda(\hat{\theta}_n), z) + \mathfrak{M}_n^\gamma(\theta_*, \lambda(\theta_*), z)\right) \le X_n Y_n. \tag{4.19}$$

Using the consistency of θ ($\hat{\theta}_n \to_p \theta_*$) and the continuity of $m_{CR}^{co}(\theta)$ by Assumption 1(e), we apply the Slutsky Theorem, to show that $\frac{1}{n}\sum_n m_{CR}^{co}(\hat{\theta}_n)_n x_n - \frac{1}{n}\sum_n m_{CR}^{co}(\theta_*)_n x_n \to_p 0$. This convergence and the central limit theorem for the pricing errors in Assumption 2(b) imply that $X_n \to_d N(0, \sigma_m^2)$. The consistency of θ and of the Lagrange Multipliers λ ($\hat{\lambda}_n(\theta_*) \to_p \lambda_*$), imply that $Y_n \to_p 0$. By another application of the Slutsky Theorem, the product $X_n Y_n \to_p 0$, which by Eq. (4.19) implies that $\sqrt{n}\left(-\mathfrak{M}_n^\gamma(\hat{\theta}_n, \lambda(\hat{\theta}_n), z) + \mathfrak{M}_n^\gamma(\theta_*, \lambda(\theta_*), z)\right) \to_p 0$. The convergence in probability of the first term on the right-hand side of Eq. (4.18) to zero and the central limit theorem in Assumption 2(a) for the second term on the right-hand side of Eq. (4.18) ($\mathfrak{M}_n^\gamma(\theta_*, \lambda(\theta_*), z)$) imply by another application of the Slutsky theorem that their sum converges in distribution to $N(0, \sigma^2)$, which is the desired result.[27] □

23 An alternative proof would be to use Lemma 1 in Hong et al. (2003) to show that $\mathfrak{M}_n^\gamma(\theta, \hat{\lambda}(\theta), z) \xrightarrow{p} \mathfrak{M}^\gamma(\theta, \lambda(\theta), z)$, for all, $\theta \in \Theta$, and then invoke the Theorem of Newey and McFadden (1994) again but this time for the parameter vector estimator $\hat{\theta}$.

24 For more primitive assumptions implying the central limit approximation of Theorem 5 see the references in HHL (1995).

25 We limit ourselves to the case of misspecified models. Other papers have treated the least-square case for both correctly and incorrectly specified models. See, for instance, Jagannathan and Wang (1996) and Kan and Robotti (2009).

26 We refer to the gradient inequality in Eq. (24), p. 248 in HHL (1995).

27 Note that similarly to HHL (1995), the first term in the decomposition of the estimator in Eq. (4.18) does not contribute to the asymptotic variance of the discrepancy statistic.

To use this limiting distribution in practice we need to obtain an estimate of the scalar asymptotic variance σ^2. Given a sample with n observations, we start by obtaining the estimates for the vector of parameters $\hat{\theta}_n$ and the corresponding estimates for the Lagrange Multipliers $\hat{\lambda}_n(\hat{\theta}_n)$. We form a time series sequence $f_t^{\gamma}(\hat{\theta}_n, \hat{\lambda}_n(\hat{\theta}_n), z) = -\hat{\lambda}_n(\hat{\theta}_n)'q + \frac{(1+\gamma\hat{\lambda}_n(\hat{\theta}_n)'x_t)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + (y(z_t, \hat{\theta}_n) - 1)\hat{\lambda}_n(\hat{\theta}_n)'x_t$, for $t = 1, \ldots, n$. Noticing that the sample mean of $f_t^{\gamma}(\hat{\theta}_n, \hat{\lambda}_n(\hat{\theta}_n), z)$ is $\mathfrak{M}_n^{\gamma}(\hat{\theta}_n, \hat{\lambda}_n(\hat{\theta}_n), z)$ (see Eq. (4.3)), σ^2 can be estimated by using a Newey and West (1987) type of frequency zero spectral density estimator applied to the time series sequence $u_t = f_t^{\gamma}(\hat{\theta}_n, \hat{\lambda}_n(\hat{\theta}_n), z) - \mathfrak{M}_n^{\gamma}(\hat{\theta}_n, \hat{\lambda}_n(\hat{\theta}_n), z)$, $t = 1, \ldots, n$, as suggested by HJ (1997).

5. Empirical application

HJ (1997) illustrated the usefulness of their least-square projection by analyzing the degree of misspecification of the canonical consumption-based asset pricing model of Breeden (1979) and Lucas (1978) for various values of the preference parameters. We perform a similar analysis considering several Cressie–Read discrepancy functions: Pearson's, EL, Hellinger's, ET, and CUE. For illustration purposes, we limit our analysis to solutions of problem (3.1), that is solutions to the MD problems for the whole space M of admissible SDFs and not for the restricted strictly positive SDFs.

The Consumption CAPM (CCAPM) SDF is given by:

$$m_t^{\text{ccapm}} = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\theta} \tag{5.1}$$

where C_t denotes the time t aggregate consumption in the economy considered.

We use the annual (1890–1985) time-series data on stocks and bonds of Campbell and Shiller (1989) updated to 2004 and the corresponding aggregate consumption annual series.[28] Similarly to HJ (1997), we first consider a small grid of values for the risk aversion coefficient θ in order to analyze the sensitivity of different discrepancy functions to changes in the parametric space. Then we estimate the model by following the procedure outlined in Section 3.5.

We concentrate on the risk-aversion parameter as it is the most important parameter in the CCAPM and since it is the one generating nonlinearities in the model. For this reason, for each value of θ, β is fixed to a value that guarantees that the mean of the CCAPM SDF proxy is always equal to 0.98, the averaged value of the historical 1-month Treasury Bill.[29]

For fixed values of the parameters β and θ, and given a time series of consumption growth rates we can compute the SDF m^{ccapm}. Once we know the SDF proxy, it is possible to compute pricing errors, to estimate the discrepancy distance δ_{CR} and Lagrange Multipliers for any discrepancy in the Cressie–Read family for a fixed value of γ.

We chose values for the risk aversion parameter θ from two very distinct regions of the parametric space. Small values of θ ($\theta = 1, 5$) will correspond to small volatility CCAPM SDFs that will have more difficulty in pricing the stock returns (S&P 500). On the other hand, high values of θ ($\theta = 20, 50$) will generate more volatile CCAPM SDFs that will have variation compatible with the extreme variation of equity returns. This behavior description for the CCAPM SDF is compatible with the equity premium puzzle document by Mehra and Prescott (1985) and re-expressed in terms of SDF's variance bounds by HJ (1991).

5.1. Lagrange Multipliers (portfolio weights) and implied SDFs

We start by analyzing results obtained with the Pearson, EL, Hellinger, ET, and CUE discrepancies. For each fixed discrepancy and parameter θ value, we solve the MD optimization problem proposed in Eq. (3.7) to obtain the Lagrange Multipliers (LM), the corresponding implied admissible SDF, and the value that minimizes the discrepancy.

Table 1 presents the LM estimated with each CR discrepancy. As noticed in Section 3.3, those LM estimates correspond to optimal portfolio weights from the maximization of a HARA utility function (plus a linear term) when the agent can invest in a short-term bond and/or the S&P 500. We observe that for all values of the parameter θ within the grid, all discrepancies agree on the signs of the weights attributed to the bond and the S&P: they all sell the bond and buy the S&P. According to the results appearing in Corollary 1, the admissible SDFs that solve the concentrated MD problem should be negatively correlated to the S&P returns whenever the nonparametric term given by $(1 + \gamma\tilde{\lambda}_{\text{ad}}'x)^{\frac{1}{\gamma}}$ dominates the parametric term coming from the CCAPM. This is exactly what can be observed in Table 2 and Figs. 1 and 2.

Table 2 presents the correlation of admissible SDFs with the S&P 500 returns. In the last column it presents the correlation of the parametric CCAPM model with the S&P 500 returns. For all discrepancies, the correlation is decreasing in absolute value with the value of θ. This is well illustrated in Figs. 1 and 2. They present for two very different values of θ (5 and 50) and each discrepancy function, the CCAPM SDF (dashed line) and the corresponding admissible SDF (solid line) that is closest to the CCAPM SDF in that region of the parametric space. Note that for the smaller value of θ in Fig. 1, the nonparametric part of the admissible SDF generates more variability than the CCAPM (compare the solid and dashed lines). For the larger values of θ (see Fig. 2), the CCAPM term shows high variability and the correlation between admissible SDF and S&P returns goes from high negative values to small negative values, which is precisely the correlation of the CCAPM SDF with the S&P returns (see last column of Table 2).

Still observing the portfolio weights (or LMs) in Table 1, we can see that the weights for the S&P are not very sensitive to changes in the parameter value θ while the weights in the short-term bond clearly decrease with θ. This has an intuitive interpretation: since when increasing the value of θ we increase the variability of the CCAPM SDF, any admissible SDF that will solve the MD problem should present the nonparametric term $(1 + \gamma\tilde{\lambda}_{\text{ad}}'x)^{\frac{1}{\gamma}}$ with volatility of the magnitude of the parametric term (the CCAPM SDF). The way to achieve this high volatility is to keep higher weights on the S&P and lower weights in the bond.

5.2. Discrepancy measures and implied probabilities

We next move to the analysis of the MD values obtained by solving the dual maximization problems that will capture the degree of misspecifation of the CCAPM model in each region of the parametric space. Table 3 presents the minimizing values for the discrepancy functions adopted. Not surprisingly, all Cressie–Read discrepancies achieve their smallest value (considering the parameter grid) when $\theta = 50$. In principle, if we had to choose a parameter value based on any of these discrepancy problems we would choose the same as HJ (1997), which corresponds to our CUE quadratic problem. However, the behavior of the implied admissible SDFs for each discrepancy function varies a lot, especially for the smaller value of the parameter $\theta = 5$ (see again Figs. 1 and 2). For instance, while Cressie–Read estimators with non-positive γ (Pearson, EL, Hellinger, and ET) produce SDFs that are positively

[28] The dataset is available on Shiller's website, http://www.econ.yale.edu/~shiller/data.htm.

[29] Note that there will still exist pricing errors when pricing the risk-free rate since it varies across time.

Table 1
Lagrange Multipliers for the CCAPM under different CR discrepancies. Risk factors are composed by annual returns over the period 1891–2004. The Bond and Stock risk factors are represented respectively by a short-term bond and S&P 500 returns as in Campbell and Shiller (1989). Cressie–Read estimators solve HARA utility maximization problems whose portfolios are linear combinations of the listed risk factors with an extra linear term including the asset pricing proxy model *y* (the CCAPM), and where the Lagrange Multipliers are the portfolio weights. A fixed SDF mean equal to 0.98 is adopted.

CCAPM parameter (θ)	CR discrepancies				
	Pearson's	*EL*	*Hellinger's*	*ET*	*CUE/HJ*
$\theta = 1$					
Bond	−0.5050	−1.0665	−1.1658	−1.1968	−1.0917
S&P	1.1105	1.4252	1.4991	1.477	1.3828
$\theta = 5$					
Bond	−0.8363	−1.2279	−1.3527	−1.4005	−1.3070
S&P	1.0330	1.4109	1.4846	1.468	1.3703
$\theta = 20$					
Bond	−0.8363	−1.4414	−1.5735	−1.6290	−1.5544
S&P	1.0330	1.3165	1.3741	1.365	1.2690
$\theta = 50$					
Bond	−0.8027	−0.1691	−0.1105	−0.0373	0.1245
S&P	1.0295	1.2847	1.3274	1.3312	1.2265

Table 2
Correlation between implied SDFs and the S&P 500 returns under different CR discrepancies. Risk factors are composed by annual returns over the period 1891–2004. The Bond and Stock risk factors are represented respectively by a short-term bond and S&P 500 returns as in Campbell and Shiller (1989). Cressie–Read SDFs are obtained from the first-order condition of HARA utility maximization problems whose portfolios are linear combinations of the listed risk factors with an extra linear term including the asset pricing proxy model *y* (the CCAPM). A fixed SDF mean is set equal to 0.98.

CCAPM (θ)	CR discrepancies					
	Pearson's	*EL*	*Hellinger's*	*ET*	*CUE/HJ*	*CCAPM*
1	−0.8045	−0.8804	−0.9197	−0.9443	−0.9656	−0.0318
5	−0.6617	−0.7593	−0.7858	−0.8013	−0.8133	−0.0385
20	−0.3104	−0.3335	−0.3312	−0.3280	−0.3213	−0.0413
50	−0.0978	−0.1111	−0.1095	−0.1079	−0.1050	−0.0071

Table 3
Measuring misspecification of the CCAPM via different CR discrepancies. Risk factors are composed by annual returns over the period 1891–2004. The Bond and Stock risk factors are represented respectively by a short-term bond and S&P 500 returns as in Campbell and Shiller (1989). Cressie–Read estimators solve HARA utility maximization problems whose portfolios are linear combinations of the listed risk factors with an extra linear term including the asset pricing proxy model *y* (the CCAPM). A fixed SDF mean is set equal to 0.98.

CCAPM parameter (θ)	CR discrepancies				
	Pearson's	*EL*	*Hellinger's*	*ET*	*CUE/HJ*
1	0.0343	0.0377	0.0382	0.0380	0.0356
5	0.0333	0.0371	0.0376	0.0374	0.0351
20	0.0295	0.0323	0.0327	0.0325	0.0306
50	**0.0277**	**0.0313**	**0.0316**	**0.0315**	**0.0299**

skewed with respect to the constant 1 (have more extreme positive values) the estimator with a positive γ (CUE) produces a SDF that is negatively skewed with respect to the constant 1.

Based on Eq. (3.23), we also computed the implied probabilities corresponding to these admissible SDFs. Here we have a total of 115 annual observations what generates constant empirical probabilities equal to $\pi^{emp} = \frac{1}{115} = 0.0087$. Figs. 3 and 4 show these implied probabilities. Note that for any fixed value of the CCAPM parameter θ, the variability of those probabilities around the empirical probabilities (dashed line) appears to be a decreasing function of the Cressie–Read parameter γ.

5.3. Model estimation

We follow the estimation procedure described in Section 3.5 and solve the optimization problems for all the CR discrepancies previously analyzed. Fig. 5 presents estimation results. In each panel corresponding to a particular value of γ, the estimated risk aversion coefficient θ is reported in the legend, while the graph plots the admissible SDF that is the closest to the estimated parametric model (the CCAPM). Table 4 presents the parameter estimates with the corresponding standard errors obtained based on the asymptotic results described in Section 4. The standard errors where estimated based on analytical formulas obtained for the asymptotic information and Hessian matrices as functions of the returns, consumption growth, Lagrange Multipliers for S&P and risk-free rate, and model parameter (risk aversion coefficient).

Table 4
Estimation of the CCAPM via different CR discrepancies. This table presents the values and standard errors of the CCAPM risk aversion parameter estimated under different CR discrepancies. Standard errors appear within parentheses below the parameter values. Risk factors are composed by annual returns over the period 1891–2004. The Bond and Stock risk factors are represented respectively by a short-term bond and S&P 500 returns as in Campbell and Shiller (1989). Cressie–Read estimators solve HARA utility maximization problems whose portfolios are linear combinations of the listed risk factors with an extra linear term including the asset pricing proxy model *y* (the CCAPM). A fixed SDF mean is set equal to 0.98.

CR discrepancies				
Pearson's	*EL*	*Hellinger's*	*ET*	*CUE/HJ*
35.3	34.7	34.4	33.9	33.0
(8.7)	(7.0)	(6.4)	(6.1)	(5.7)

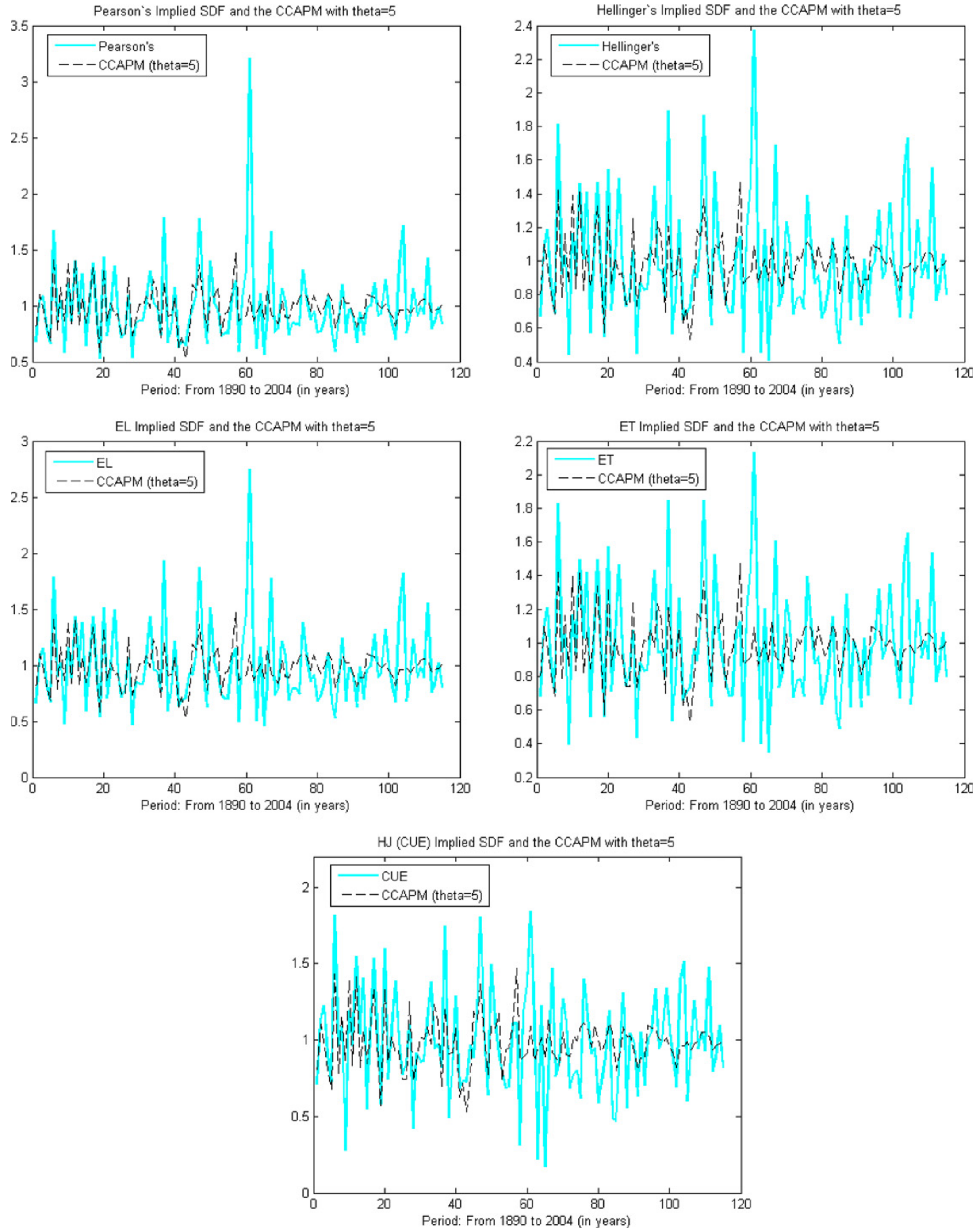

Fig. 1. Admissible SDFs and the CCAPM with $\theta = 5$. This picture presents Admissible SDFs under different CR discrepancies and the CCAPM SDF with a risk aversion coefficient (θ) of equal to 5. The MD problems are solved based on annual returns on S&P 500 and a short-term bond over the period 1890–2004. The CCAPM SDF is based on consumption growth data over the period 1890–2004. All SDF means are fixed at 0.98.

We see that all CR discrepancy estimators produce very high values for the risk aversion coefficient, which confirms the failure of the CCAPM model to explain the equity premium with a reasonable risk aversion coefficient. The estimated value for the risk aversion is in the order of 34.0, ranging from a minimum of 33.0 under the CUE ($\gamma = 1$) estimator to a maximum of 35.3 under the Pearson estimator. Standard errors are a decreasing function of the Cressie–Read parameter gamma, going from 8.7 under the Pearson estimator down to 5.7 under the CUE estimator.

Note also that as we are solving problems that search admissible SDFs in the set M, we are not restricting the SDFs to be positive. For this reason we can observe that some admissible SDFs achieve negative values in some states. Alternatively, we could have solved the MD problem in $M+$ (set of nonnegative admissible SDFs) to guarantee the nonnegativity of the SDFs. Li et al. (forthcoming) indicate that for the HJ distance, when they search within the set of nonnegative SDFs $M+$ they are able to distinguish asset pricing models better than when they search within the set of admissible SDFs M. Gospodinov et al. (2010) analyze the case of linear asset pricing models. They provide a detailed analysis of the conditions under which searching in both M and $M+$ results in significant different results in terms of model choice. Given all this recent research on the topic of HJ distance with no-arbitrage constraint, we believe comparing results in the two sets of SDFs (M and $M+$)

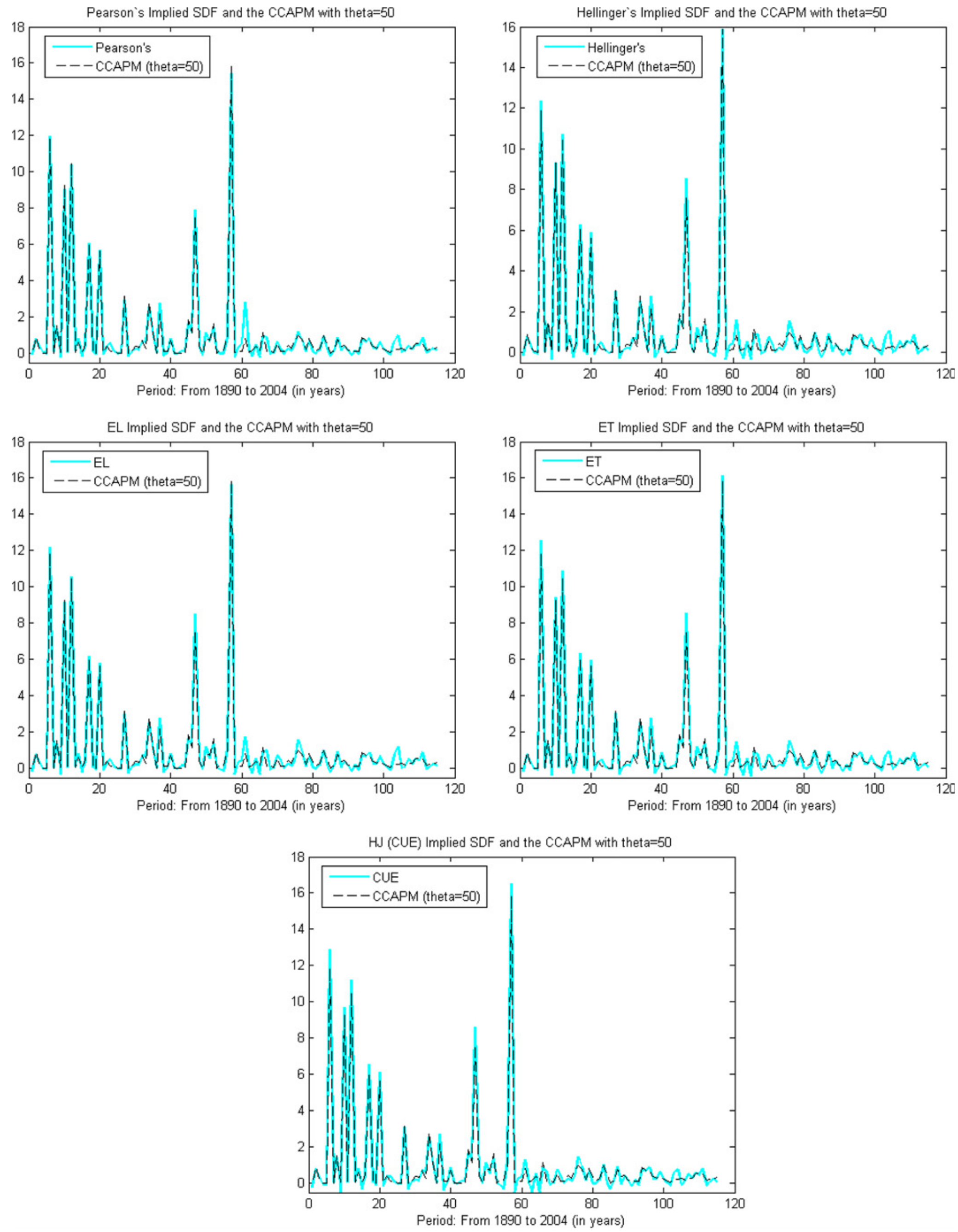

Fig. 2. Admissible SDFs and the CCAPM with $\theta = 50$. This picture presents Admissible SDFs under different CR discrepancies and the CCAPM SDF with a risk aversion coefficient (θ) equal to 50. The MD problems are solved based on annual returns on S&P 500 and a short-term bond over the period 1890–2004. The CCAPM SDF is based on consumption growth data over the period 1890–2004. All SDF means are fixed at 0.98.

under our general MD discrepancies is an interesting and natural avenue to pursue in future research, especially with a larger set of asset pricing models.

6. Discussion

A few lessons can be learned from the empirical application we just described. While applying our new discrepancy measures to the CCAPM may appear like beating a dead horse, the exercise is in fact very instructive in terms of misspecification. Indeed, the Cressie–Read discrepancy measures all concur in showing that the CCAPM model becomes compatible with return data only for high values of the risk-aversion. However, our approach allows us to recover the implied nonparametric SDF or equivalently how the MD estimators correct the pricing errors obtained by the CCAPM and this has information about the potential source of misspecification. Indeed, as usual in econometrics, a good misspecification test should provide some insight about the source of the specification problem and suggest some direction for improving the model. For instance, in a regression model, analyzing and testing some characteristics of the residuals will reveal some missing variables. Similarly, in our context, the functional form of basis assets with which the estimator corrects pricing errors can be suggestive of directions to improve the asset pricing model. For instance, while the HJ distance will correct the CCAPM with a linear combination of the returns on the short term interest rate and S&P, a estimator like EL will correct the CCAPM with the log of a linear combination of basis assets. Depending

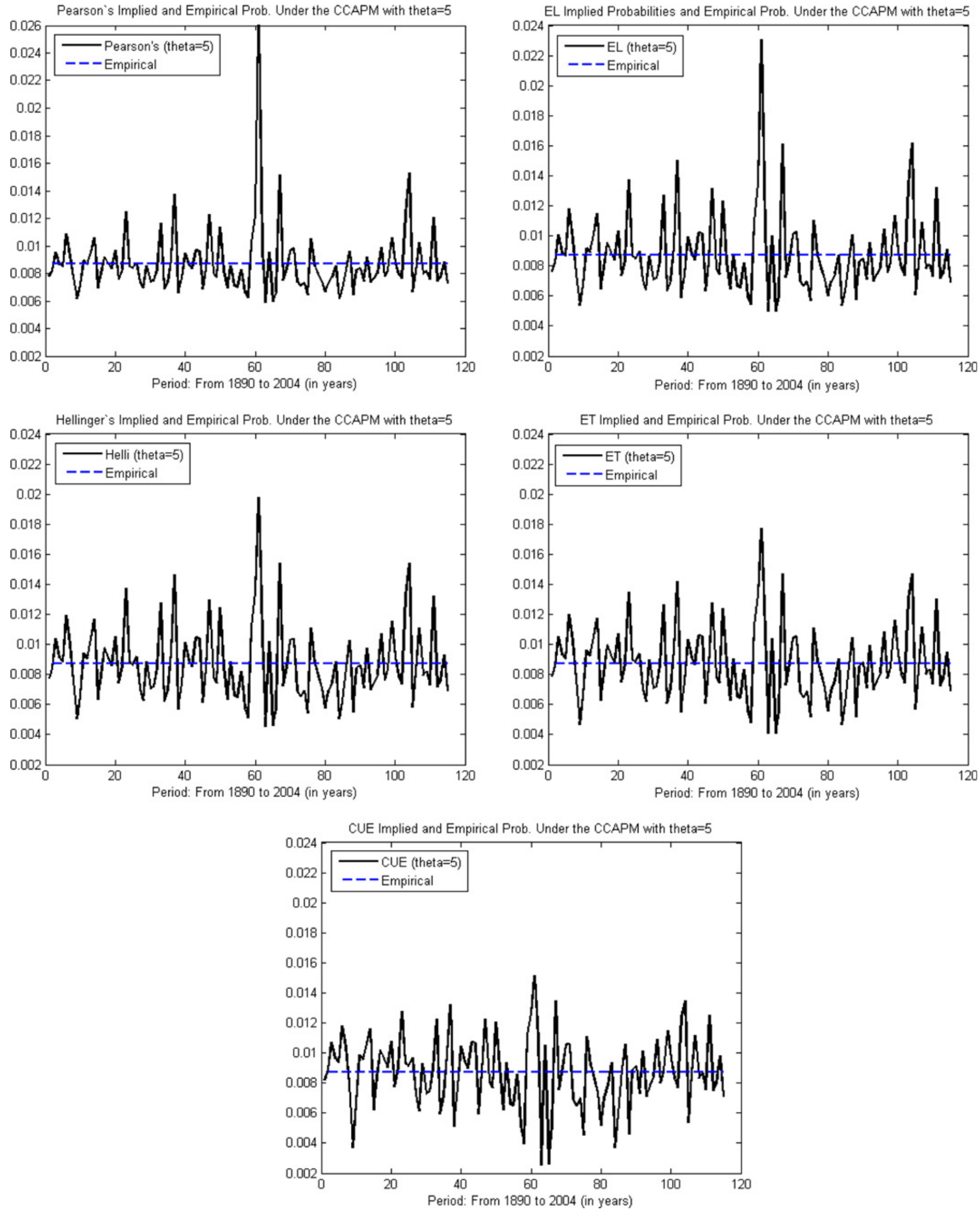

Fig. 3. Implied probabilities under the CCAPM with $\theta = 5$. This picture presents implied probabilities for different discrepancy measures from the Cressie–Read family, when measuring the degree of misspecification of the CCAPM asset pricing model for a fixed risk aversion coefficient $\theta = 5$. Results are based on annual returns from S&P 500 and a short-term bond over the period 1890–2004. The implied probabilities are obtained by solving the dual optimization problems that have a portfolio interpretation in terms of maximization of a HARA utility function. Implied probabilities are hyperbolic functions of the returns.

on how much the powers of consumption growth and the S&P returns covary, a nonlinear correction of the CCAPM might be more effective. An important question is what is the best way to correct the CCAPM model using S&P and short term rate returns, so as to price S&P returns? That is the potential kind of question that our methodology allows to answer. According to how much weight each estimator gives to the moments of returns, the correction will be different and will affect the value of the parameter estimated, as can be observed in Section 5.3. Statistical tests can be constructed via the distances obtained under different MD estimators, with the asymptotic properties described in Section 4.

The richness of our framework in terms of discrepancy measures may be construed as a hurdle since a criterion has to be chosen to pick a γ among possible values of the Cressie–Read family or even to choose between many families of discrepancies. In Almeida and Garcia (2008), we discuss robustness issues related to diagnosing asset pricing models and performance evaluation. In the case of our misspecification measure and the corresponding statistics that come with it, varying the γ will tell us to what extent the model assessment is dependent upon the discrepancy measure chosen. Allowing for this robustness analysis is in our view a good feature of our approach since it can tell us in which direction to improve the asset pricing models at hand.

7. Conclusion

We extend the least-square projection proposed by HJ (1997) to measure the degree of misspecification of asset pricing models by suggesting more general projections based on the minimization of discrepancy convex functions. Solutions to these MD problems

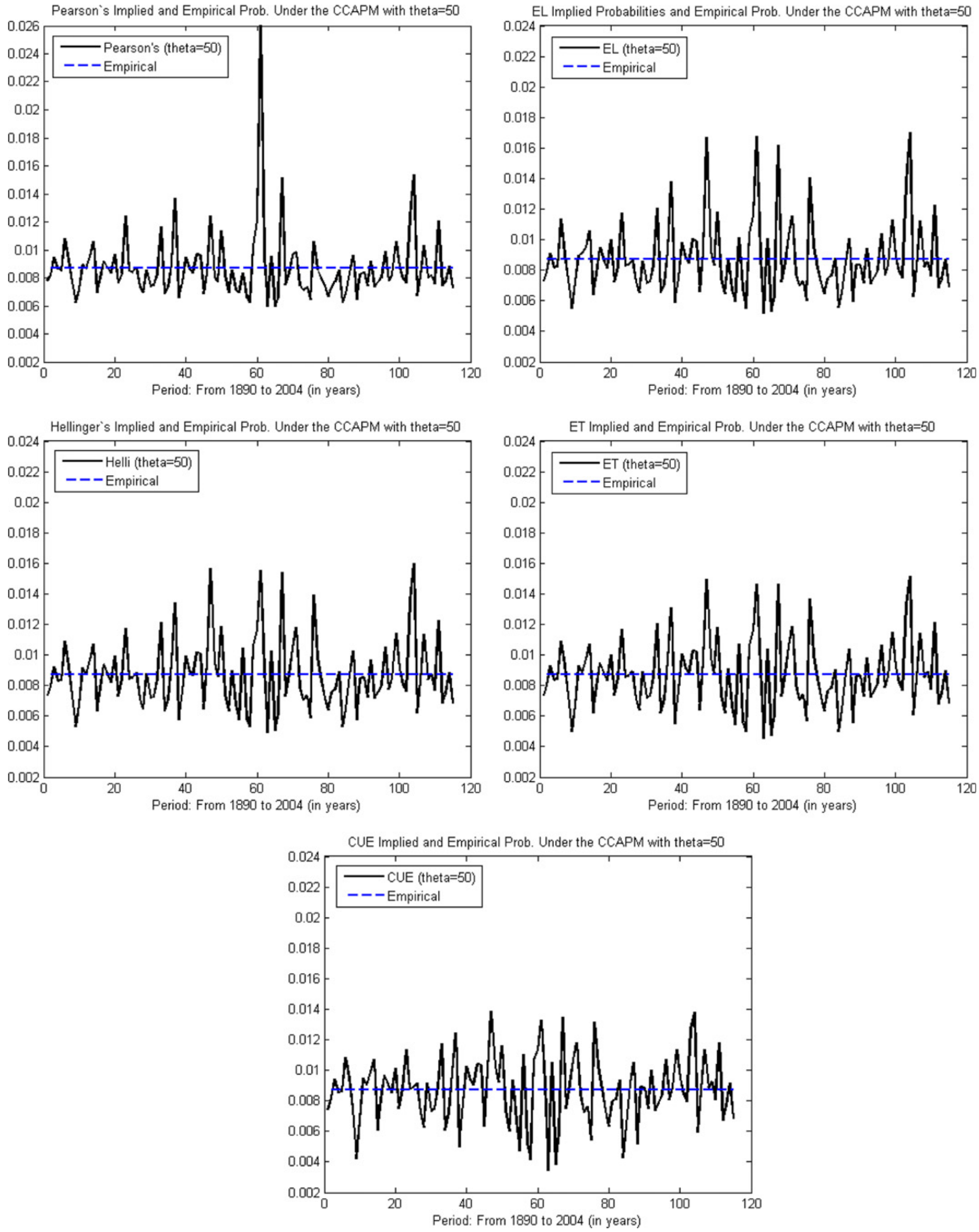

Fig. 4. Implied probabilities under the CCAPM with $\theta = 50$. This picture presents implied probabilities for different discrepancy measures from the Cressie–Read family, when measuring the degree of misspecification of the CCAPM asset pricing model for a fixed risk aversion coefficient $\theta = 50$. Results are based on annual returns from S&P 500 and a short-term bond over the period 1890–2004. The implied probabilities are obtained by solving the dual optimization problems that have a portfolio interpretation in terms of maximization of a HARA utility function. Implied probabilities are hyperbolic functions of the returns.

naturally imply semiparametric and nonlinear SDFs that take into account an arbitrary number of moments of the distributions of assets returns. We relate the problem of finding general MD projections of asset pricing models onto the family of admissible SDFs to that of solving an optimal portfolio problem. When specializing to the Cressie–Read family of discrepancies, our projections are obtained as solutions to optimal portfolio problems based on HARA utility functions added to a linear term on the asset pricing proxy considered as an imperfect SDF benchmark. We also relate the MD admissible SDFs to the implied probabilities from the econometric literature (see Newey and Smith, 2004), showing that in our context those probabilities are a normalized version of the admissible SDFs translated by an affine function of the asset pricing proxy model. Finally, we illustrate our methodology with an application to the CCAPM model, making use of a number of well-known Cressie–Read discrepancies, namely Pearson's, EL, Hellinger's, ET, and CUE. All estimators produce a high value for the risk-aversion parameter, reinforcing the equity premium puzzle.

Acknowledgments

We are grateful for comments and suggestions from Allan Timmermann, Luis Braido, Olesya Grishchenko, Raymond Kan, Marcelo Maia, Ross Valkanov, Andre Silva, two anonymous referees, the editors, and seminar and conference participants at the 2009 EFA in Bergen, the 2009 ESEM in Barcelona, 2009 Lubrafin in Rio, 2010 NFA in Winnipeg, Rady School at UCSD, and Getulio Vargas Foundation. The first author thanks CNPq-Brazil for financial support. The second author is a research Fellow of CIRANO and CIREQ.

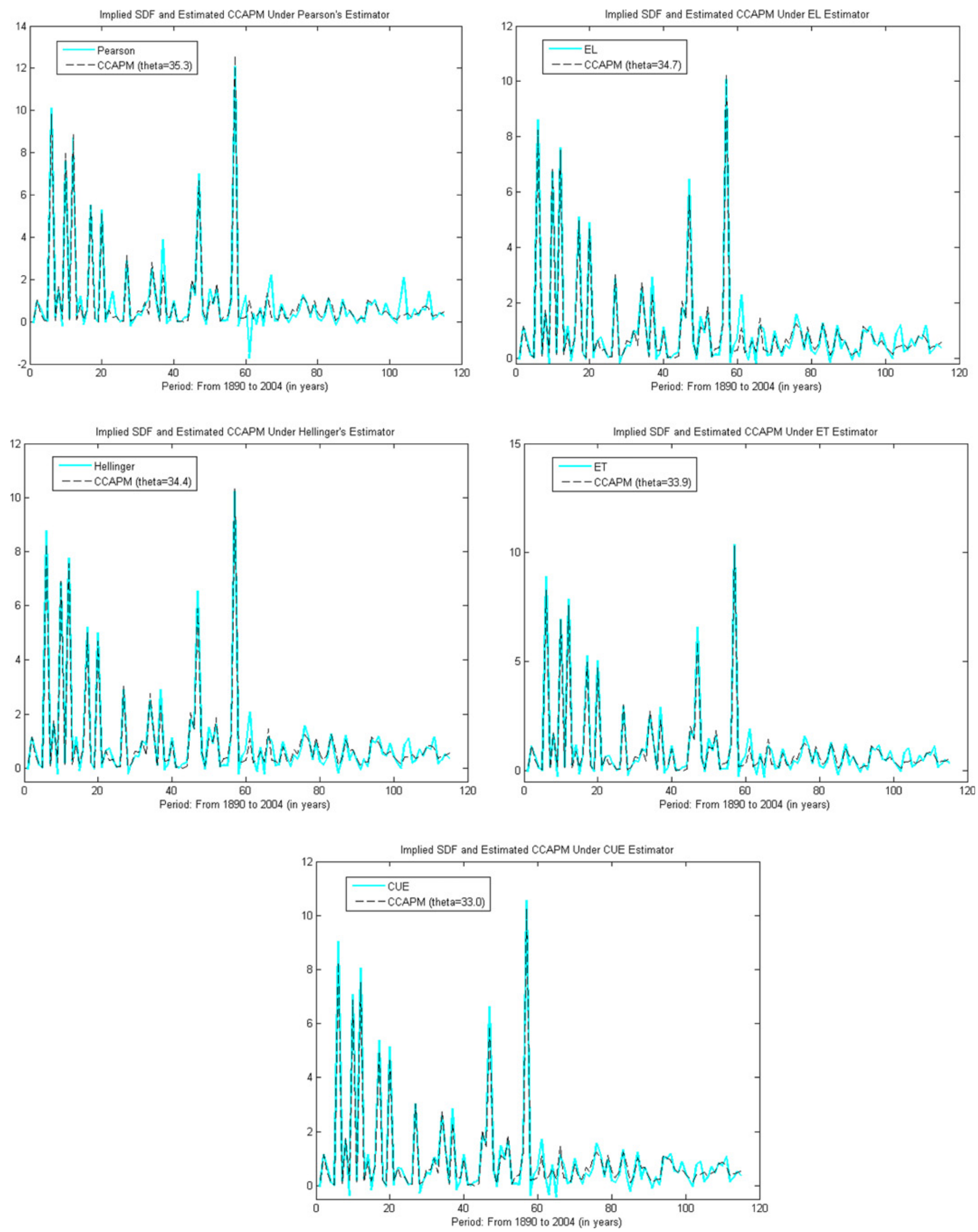

Fig. 5. Minimum Discrepancy estimators of the CCAPM model. This picture presents admissible SDFs and corresponding estimated CCAPM SDFs for different discrepancy measures from the Cressie–Read family. Results are based on annual returns from S&P 500 and a short-term bond over the period 1890–2004. The SDFs are obtained by solving a double optimization problem where we search for the parameter theta that minimizes the Cressie–Read discrepancy which contains a linear term in the CCAPM asset pricing proxy.

Appendix. Asymptotic normality of estimators

First, we provide explicit expressions for the various quantities that will enter our assumptions to be able to prove asymptotic normality of the estimators.

Given the function $f^{\gamma}(\theta,\lambda,z)$ from Eq. (4.1), for some assumptions in deriving the asymptotic normality results, we will need to know its derivative with respect to λ:

$$\begin{aligned} h^{\gamma}(\theta,\lambda,z) &= \frac{\partial}{\partial\lambda'}f^{\gamma}(\theta,\lambda,z) \\ &= -q+(1+\gamma\lambda'x)^{\frac{1}{\gamma}}x+(y(z,\theta)-1)x. \end{aligned} \tag{A.1}$$

We will also need first and second derivatives of this function $h^{\gamma}(\cdot)$:

$$\frac{\partial}{\partial\theta'}h^{\gamma}(\theta,\lambda,z)=y'(\theta,z), \tag{A.2}$$

$$\frac{\partial^2}{\partial\theta\partial\theta'}h^{\gamma}(\theta,\lambda,z)=y''(\theta,z). \tag{A.3}$$

The elements of the score function vector $g(\Phi)=\frac{\partial}{\partial\Phi'}\mathfrak{M}_n^{\gamma}(\Phi,z)$ are given by:

$$\frac{\partial\mathfrak{M}_n^{\gamma}(\Phi,z)}{\partial\theta'}=\frac{1}{n}\sum_{t=1}^{n}y'(\theta,z_t)\lambda'x_t \tag{A.4}$$

$$\begin{aligned} \frac{\partial\mathfrak{M}_n^{\gamma}(\Phi,z)}{\partial\lambda'} &= -q+\frac{1}{n}\sum_{t=1}^{n}(1+\gamma\lambda'x_t)^{\frac{1}{\gamma}}x_t \\ &\quad +(y(z_t,\theta)-1)x_t. \end{aligned} \tag{A.5}$$

The expectation of the cross-product of the score vector by its transpose will provide the matrix $S(\Phi)$.

The elements of the Hessian $H_{\Phi\Phi}$ are given by:

$$\frac{\partial^2 \mathfrak{M}_n^{\gamma}(\Phi, z)}{\partial\theta'\partial\lambda'} = \frac{1}{n}\sum_{t=1}^{n} y'(\theta, z_t)x_t \tag{A.6}$$

$$\frac{\partial^2 \mathfrak{M}_n^{\gamma}(\Phi, z)}{\partial\theta'\partial\theta'} = \frac{1}{n}\sum_{t=1}^{n} y''(\theta, z_t)\lambda' x_t \tag{A.7}$$

$$\frac{\partial^2 \mathfrak{M}_n^{\gamma}(\Phi, z)}{\partial\lambda'\partial\lambda'} = \frac{1}{n}\sum_{t=1}^{n} x_t x_t'(1+\gamma\lambda' x_t)^{\frac{1-\gamma}{\gamma}}. \tag{A.8}$$

To prove the asymptotic normality of the estimator $\hat{\Phi}$ we need the following assumptions:

Assumption 3. (a) The process z_t is strong mixing with mixing coefficients α_n satisfying $\sum_{n=1}^{\infty} \alpha_n^{1-\frac{1}{b}} < \infty$, $b > 1$;

(b) $\mathrm{Var}\left(n^{-\frac{1}{2}}\left[\left(\frac{\partial}{\partial\Phi'}\right)\mathfrak{M}_n^{\gamma}(\Phi_*, z)\right]\right) \to S_\Phi > 0 (n \to \infty)$;

(c) $H_{\Phi\Phi}$ is of full rank, where:

$$H_{\Phi\Phi} = E\left[\frac{\partial^2}{\partial\Phi\partial\Phi'} f^{\gamma}(\theta^*, \lambda^*, z)\right] = \begin{bmatrix} H_{\theta\theta} & H_{\theta\lambda} \\ H'_{\theta\lambda} & H_{\lambda\lambda}\end{bmatrix};$$

(d) $E\left[\|\frac{\partial}{\partial\theta'} h^{\gamma}(\theta_*, \lambda, z)\|^{2b}\right] < \infty$, for any $\lambda \in \Re$;

(e) θ^* is an interior point of Θ and $h^{\gamma}(\theta, \lambda, z)$ is twice continuously differentiable at $\theta = \theta_*$, for any $\lambda \in \Re$, z-almost surely;

(f) There exists a constant $\epsilon > 0$ such that:

$$E\left[\sup_{\bar\theta\in\Gamma(\theta^*,\Delta)} \|h^{\gamma}(\bar\theta, \lambda, z)\|^{2+\epsilon}\right] < \infty,$$

$$E\left[\sup_{\bar\theta\in\Gamma(\theta^*,\Delta)} \left\|\frac{\partial}{\partial\theta'} h^{\gamma}(\bar\theta, \lambda, z)\right\|^{2+\epsilon}\right] < \infty,$$

$$E\left[\sup_{\bar\theta\in\Gamma(\theta^*,\Delta)} \left\|\frac{\partial^2}{\partial\theta\partial\theta'} h^{\gamma}(\bar\theta, \lambda, z)\right\|^{2+\epsilon}\right] < \infty,$$

for a sufficiently small $\Delta > 0$, for any $\lambda \in \Re$.

Proof of Theorem 4. We start by expanding $g(\widehat{\Phi}) = \frac{\partial}{\partial\Phi'}\mathfrak{M}_n^{\gamma}(\widehat{\Phi}, z)$ around $\Phi_* = [\theta_*', \lambda_*']'$:

$$0 = g(\widehat{\Phi}) = g(\Phi_*) + H_{\Phi\Phi n}(\bar\Phi)(\widehat{\Phi} - \Phi_*), \tag{A.9}$$

where $H_{\Phi\Phi n} = \frac{\partial^2}{\partial\Phi\partial\Phi'}\mathfrak{M}_n^{\gamma}(\Phi, z)$ and $\bar\Phi$ is a convex combination of Φ_* and $\widehat{\Phi}$, which may be different for each parameter of the Φ vector. Solving (A.9) for $(\widehat{\Phi} - \Phi_*)$ and rewriting that every factor is $O(1)$ yields, assuming the invertibility of $H_{\Phi\Phi n}(\bar\Phi)$ in the neighborhood of Φ_*:

$$n^{\frac{1}{2}}(\widehat{\Phi} - \Phi_*) = -H_{\Phi\Phi n}(\bar\Phi)^{-1}\left(n^{\frac{1}{2}} g(\Phi_*)\right). \tag{A.10}$$

If the first term on the right-hand side turns out to be asymptotically nonstochastic, and the second term turns out to be asymptotically normal, it will follow that $n^{\frac{1}{2}}(\widehat{\Phi} - \Phi_*)$ must be asymptotically normal. We first show that $H_{\Phi\Phi n}(\bar\Phi)$ tends to a certain nonstochastic matrix as $n \to \infty$. Under Assumptions 1(c) and 2(a), (d), (e), (f), one can invoke the uniform weak law of large numbers with $sup_{\Phi\in\Gamma(\Phi_*,\Delta)}\|H_{\Phi\Phi n} - EH_{\Phi\Phi n}\| \to_p 0$. Therefore $H_{\Phi\Phi n}(\bar\Phi)$ must tend to the asymptotic Hessian $H_{\Phi\Phi}(\bar\Phi)$ as $n \to \infty$. But since $\widehat{\Phi}$ is consistent for Φ_* and $\bar\Phi$ lies between $\widehat{\Phi}$ and Φ_*, it is clear that $H_{\Phi\Phi n}(\bar\Phi)$ must also tend to $H_{\Phi\Phi}(\Phi_*)$. Since by (c) $H_{\Phi\Phi}(\Phi_*)$ is of full rank:

$$n^{\frac{1}{2}}(\widehat{\Phi} - \Phi_*) =_a H_{\Phi\Phi}^{-1}(\Phi_*)\left(n^{\frac{1}{2}} g(\Phi_*)\right). \tag{A.11}$$

The only stochastic element on the right-hand side is

$$n^{\frac{1}{2}} g(\Phi_*). \tag{A.12}$$

By Assumptions 1(c), 2(a)–(d), we can invoke a central limit theorem (see Kitamura and Stutzer, 1997) that guarantees that:

$$n^{\frac{1}{2}} g(\Phi_*) \to_d N(0, S_\Phi), \tag{A.13}$$

and the desired result in Theorem 4 follows, where the V_Φ matrix is given by: $V_\Phi = H_{\Phi\Phi}^{-1} S_\Phi H_{\Phi\Phi}^{-1}$. □

References

Almeida, C., Garcia, R., 2008. Empirical likelihood estimators for stochastic discount factors. Working Paper. Getulio Vargas Foundation and EDHEC Business School.

Antoine, B., Bonnal, H., Renault, E., 2007. On the efficient use of the informational content of estimating equations: implied probabilities and Euclidean empirical likelihood. Journal of Econometrics 138, 461–487.

Araujo, A., Monteiro, P.K., 1989. Equilibrium without uniform conditions. Journal of Economic Theory 48 (2), 416–427.

Back, K., Brown, D.P., 1993. Implied probabilities in GMM estimators. Econometrica 61, 971–975.

Baroni-Adesi, G., 1985. Arbitrage equilibrium with skewed asset returns. Journal of Financial and Quantitative Analysis 20 (3), 299–313.

Borwein, J.M., Lewis, A.S., 1991. Duality relationships for entropy-like minimization problems. SIAM Journal on Control and Optimization 29 (2), 325–338.

Breeden, D., 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. Journal of Financial Economics 7 (3), 265–296.

Brown, B.W., Newey, W.K., 1998. Efficient semiparametric estimation of expectations. Econometrica 66, 453–464.

Brown, B.W., Newey, W.K., 2002. Generalized method of moments, efficient bootstrapping, and improved inference. Journal of Business and Economic Statistics 20, 507–517.

Campbell, J., Shiller, R., 1989. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1 (3), 195–228.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure, and mean–variance analysis on large asset markets. Econometrica 51 (5), 1281–1304.

Chen, X., Favilukis, J., Ludvigson, S., 2008. An estimation of economic models with recursive preferences. Working Paper. At Yale, LSE, and NYU.

Chen, X., Hong, H., Shum, M., 2007. Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models. Journal of Econometrics 141, 109–140.

Chen, X., Ludvigson, S., 2009. Journal of Applied Econometrics 24 (7), 1057–1093.

Corcoran, S.A., 1998. Bartlett adjustment of empirical discrepancy statistics. Biometrika 85, 967–972.

Cressie, N., Read, T.R.C., 1984. Multinomial goodness-of-fit tests. Journal of the Royal Statistical Society: Series B 46, 440–464.

Csiszàr, I., 1975. *I*-divergence geometry of probability distributions and minimization problems. The Annals of Probability 3 (1), 146–158.

Csiszàr, I., 1991. Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems. The Annals of Statistics 19 (4), 2032–2066.

Dittmar, R., 2002. Nonlinear pricing kernels, kurtosis preference, and the cross-section of assets returns. Journal of Finance 57 (1), 369–403.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33 (1), 3–56.

Gospodinov, N., Kan, R., Robotti, C., 2010. On the Hansen–Jagannathan distance with a no-arbitrage constraint. Working Paper. Concordia University, University of Toronto, and FED of Atlanta.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054.

Hansen, L.P., Heaton, J., Luttmer, E.G.J., 1995. Econometric evaluation of asset pricing models. Review of Financial Studies 8 (2), 237–274.

Hansen, L., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14, 262–280.

Hansen, L.P., Jagannathan, R., 1991. Implications of security market data for models of dynamic economies. Journal of Political Economy 99 (2), 225–262.

Hansen, L.P., Jagannathan, R., 1997. Assessing specification errors in stochastic discount factor models. Journal of Finance 52 (2), 557–589.

Hansen, L.P., Singleton, K., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50 (5), 1269–1286.

Harrison, M., Kreps, D., 1979. Martingales and arbitrage in multiperiod security markets. Journal of Economic Theory 20, 381–408.

Harvey, C., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance 55 (3), 1263–1295.
Hodrick, R., Zhang, X., 2001. Evaluationg the specification errors of asset pricing models. Journal of Financial Economics 62, 327–376.
Hong, H., Preston, B., Shum, M., 2003. Generalized empirical likelihood-based model selection criteria for moment condition models. Econometric Theory 19, 923–943.
Huang, C.F., Litzenberger, R., 1988. Foundations for Financial Economics. North-Holland.
Imbens, G., Spady, R.H., Johnson, P., 1998. Information theoretic approaches to inference in moment condition models. Econometrica 66 (2), 333–357.
Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. Journal of Finance 51 (1), 3–53.
Kan, R., Robotti, C., 2009. Review of Financial Studies 22 (9), 3449–3490.
Kitamura, Y., 2000. Comparing misspecified dynamic econometric models using nonparametric likelihood. Working Paper. University of Pennsylvania.
Kitamura, Y., 2006a. A likelihood-based approach to the analysis of a class of nested and non-nested models. Working Paper. Yale University.
Kitamura, Y., 2006b. Empirical likelihood methods in econometrics: theory and practice. Working Paper. Yale University.
Kitamura, Y., Stutzer, M., 1997. An information-theoretic alternative to generalized method of moment estimation. Econometrica 65, 861–874.
Kitamura, Y., Stutzer, M., 2002. Connections between entropic and linear projections in asset pricing estimation. Journal of Econometrics 107, 159–174.
Kraus, A., Litzenberger, R.H., 1976. Skewness preference and the valuation of risk assets. Journal of Finance 31 (4), 1085–1100.
Li, H., Xu, Y., Zhang, X., 2010. Evaluating asset pricing models using the second Hansen–Jagannathan distance. Journal of Financial Economics (forthcoming).
Lucas, R., 1978. Asset prices in an exchange economy. Econometrica 46 (6), 1429–1445.
Luenberger, D., 1969. Optimization by Vector Space Methods. John Wiley & Sons.
Mehra, R., Prescott, E., 1985. The equity premium: a puzzle. Journal of Monetary Economics 15, 145–161.
Newey, W., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R., McFadden, D. (Eds.), The Handbook of Econometrics, Volume IV. pp. 2113–2145.
Newey, W., Smith, R., 2004. Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica 72, 219–255.
Newey, W.K., West, K.D., 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75 (2), 237–249.
Ramalho, J.J.S., Smith, R.J., 2002. Generalized empirical likelihood non-nested tests. Journal of Econometrics 107, 99–125.
Royden, H.L., 1988. Real Analysis. Prentice-Hall.
Rubinstein, M., 1973. The fundamental theorem of parameter-preference security valuation. Journal of Financial and Quantitative Analysis 8 (1), 61–69.
Smith, R.J., 1997. Alternative semi-parametric likelihood approaches to generalized method of moments estimation. Economic Journal 107, 503–510.
Smith, R.J., 2004. GEL criteria for moment condition models. Working Paper. University of Warwick.
Smith, R.J., 2007. Efficient information theoretic inference for conditional moment restrictions. Journal of Econometrics 138 (2), 430–460.
Snow, K., 1991. Diagnosing asset pricing models using the distribution of asset returns. Journal of Finance 46, 955–983.
Stutzer, M., 1995. A Bayesian approach to diagnosis of asset pricing models. Journal of Econometrics 68, 367–397.
Vanden, J.M., 2006. Option coskewness and capital asset pricing. Review of Financial Studies 19 (4), 1279–1320.
Wang, Z., Zhang, X., 2005. Empirical evaluation of asset pricing models: arbitrage and pricing errors over contingent claims. Working Paper. Cornell University.