Identification of Beliefs in the Presence of Disaster Risk and Misspecification[*]

Saraswata Chaudhuri[†] Eric Renault[‡] and Oscar Wahlstrom[§]

November 8, 2021

Abstract

We discuss the econometric underpinnings of Barro [2006]'s defense of the rare disaster model as a way to bring back an asset pricing model "into the right ballpark for explaining the equity-premium and related asset-market puzzles". Arbitrarily low-probability economic disasters can restore the validity of model-implied moment conditions only if the amplitude of disasters may be arbitrary large in due proportion. We prove an impossibility theorem that in case of potentially unbounded disasters, there is no such thing as a population empirical likelihood (EL)-based model-implied probability distribution. That is, one cannot identify some belief distortions for which the EL-based implied probabilities in sample, as computed by Julliard and Ghosh [2012], could be a consistent estimator. This may lead to consider alternative statistical discrepancy measures to avoid the problem with EL. Indeed, we prove that, under sufficient integrability conditions, power divergence Cressie-Read measures with positive power coefficients properly define a unique population model-implied probability measure. However, when this computation is useful because the reference asset pricing model is misspecified, each power divergence will deliver a different model-implied beliefs distortion. One way to provide economic underpinnings to the choice of a particular belief distortion is to see it as the endogenous result of investor's choice when optimizing a recursive multiple-priors utility à la Chen and Epstein [2002]. Jeong, Kim, and Park [2015]'s econometric study confirms that this way of accommodating ambiguity aversion may help to address the Equity Premium puzzle.

[*]We thank the coeditor Simon Lee, an anonymous referee, Alastair Hall, Lars Peter Hansen and Essie Maasoumi for their very helpful comments.

[†]Department of Economics, McGill University, Canada. Email: saraswata.chaudhuri@mcgill.ca.

[‡]Corresponding author. Department of Economics, University of Warwick, United Kingdom. Email: Eric.Renault@warwick.ac.uk.

[§]Department of Economics, Brown University, United States. Email: oscar_wahlstrom@brown.edu.

1 Introduction

The absence of arbitrage opportunities implies the existence of a stochastic discount factor (SDF), such that the equilibrium price of a traded security can be represented as the conditional expectation of the future payoff discounted by the SDF. Thus, a typical asset pricing equation is:

$$E[SR - e_n \,|I] = 0$$

where R denotes an n-dimensional vector of gross returns corresponding to payoffs on financial assets over some investment horizon, S denotes the corresponding SDF for this horizon, and I stands for the information set of the representative investor. e_n is a n-dimensional vector of ones.

Hereafter, the interval $[t, t+1]$ is the period of the investment between date t (today) and date $(t+1)$ where gross returns, denoted by R_{t+1}, are observed. Besides the horizon time $(t+1)$, the SDF S_{t+1} may depend on unknown parameters θ, giving rise to the set of conditional moment restrictions:

$$E[g\left(X_{t+1}, \theta\right) |I(t)] = 0, \theta \in \Theta \subset \mathbb{R}^p \tag{1}$$

where the function $g\left(X_{t+1}, \theta\right) = S_{t+1}\left(\theta\right) R_{t+1} - e_n$ captures the parameter dependence of the SDF $S_{t+1}\left(\theta\right)$ along with random variables X_{t+1} observed by the econometrician and used to construct the payoffs, prices, and the SDF.

The unknown parameters θ typically describe the preferences of the representative investor and are identified by the fact that for a true unknown value θ^0 of the parameters, $g\left(X_{t+1}, \theta^0\right)$ should be a martingale difference sequence. However, standard asset pricing models lead to the so-called equity premium puzzle (EPP, see Mehra and Prescott [1985]), that is the failure of the representative agent model to fit historical averages of the equity premium and the risk free rate.

An alternative view put forward in our current paper is that part of the premium can be accommodated only by considering in (1) a distortion of the expectation operator that reflects the lack of investor confidence in the assignment of probabilities to future events. We set the focus on models that capture this departure from rational expectations by one of the two following channels: distorted subjective beliefs or ambiguity aversion.

As clearly discussed by Chen and Epstein [2002] (see their Section 1.2.), there is some observational equivalence between models with aversion for ambiguity captured by a multiple-priors recursive utility (and a minimax type of value function) and models of belief distortion which only relax the rational expectations hypothesis that the agent knows the true probability distribution. As a matter of fact, with the recursive multiple-priors utility, the investor's optimization ultimately delivers a distorted probability measure selected endogenously from the agent's set of priors.

This latter remark suggests an econometric procedure to look for a distorted probability distribution that is minimally distorted with respect to the Data Generating Process (DGP) while bringing a plausible solution to the EPP. This econometric issue has been addressed by Jeong, Kim, and Park [2015] for the ambiguity aversion approach and by Ghosh, Otsu, and Roussellet [2021] for the distorted subjective beliefs. In both cases the goal is to state the asymptotic theory of a minimum distance approach to estimate both preference parameters and distorted beliefs.

Before developing the econometric methodology, it is worth understanding why we expect that these estimation procedures will deliver estimators of preference parameters and belief distortions that will improve upon the solution of the EPP.

In the case of a model of ambiguity aversion (the so-called κ-Ignorance model of Chen and Epstein [2002], Jeong, Kim, and Park [2015] end up with a three factor CAPM. The risk premium of a given asset is determined not only by the covariance of its return with consumption growth and with aggregate wealth (as in Epstein and Zin [1989]'s recursive utility model) but also by covariance with the density generator that defines the multiple-priors recursive utility model. The latter covariance adds some significant ambiguity compensation to the traditional risk premium, allowing the risk premium to be consistent with more realistic levels of relative risk aversion.

As far as the subjective beliefs distortion is concerned, Barro [2006] has revisited Rietz

[1988]'s original way to address the EPP by bringing in low-probability economic disasters. According to Barro [2006], "the major reason for scepticism about Rietz's argument is the belief that it depends on counterfactually high probabilities and sizes of economic disasters". This has been the main motivation of Barro [2006]'s empirical analysis focused on "the measurement of the frequency and size of economic disasters that occurred during the twentieth century". The goal is to "calibrate the model using the observed probability distribution for economic disasters in the twentieth century" and the conclusion is that "the model's solution gets into the right ballpark for explaining the equity-premium and related asset market puzzles". Following this observation, Julliard and Ghosh [2012] have promoted an information theoretic empirical strategy that could reconcile a given asset pricing model with the observed data, and the asymptotic theory of this strategy has been settled by Ghosh, Otsu, and Roussellet [2021].

In light of the above discussion, the contribution of our current paper is threefold.

First, we provide some mathematical arguments to confirm the rather general validity of Julliard and Ghosh [2012]'s empirical observation that the estimated belief distortion makes economic sense because "a priori, we would expect that the rare events distribution needed to rationalize the EPP assigns relatively higher weights to a few particular bad states of the economy (..) this is exactly what the estimated (probabilities) do".

Second, besides economic sense, we ask whether the estimated belief distortion also makes statistical sense because it may consistently estimate a well-defined population probability distribution. Since "higher weights to a few particular bad states"seems to give some support to the disaster risk theory as advocated by Barro [2006] ("counterfactually high probabilities and sizes of economic disasters"), one would expect to see positive probabilities assigned to unbounded disaster events. Unfortunately, we point out that when one does not maintain the rather restrictive assumption that the possible disasters are of bounded amplitude, a population distorted belief defined by minimization of a population statistical divergence function may not exist. In particular, we show that with a natural scheme of unbounded disasters, the population distorted beliefs do not exist when estimated by maximum empirical likelihood, as in Julliard and Ghosh [2012]. Of course, the non-existence issue is at stake only when there is a need to distort subjective beliefs because the historical probability distribution does not satisfy the moment conditions (1).

Finally, we prove that sufficient conditions for existence of distorted subjective population beliefs fulfilling moment restrictions (1) are given either by assuming that all possible disasters are of bounded amplitude, or by using a discrepancy function that, by contrast with empirical likelihood, is an increasing convex function.

1.1 Relation to the existing literature

There is a vast literature on disaster risk and its implications for empirical asset pricing (see Tsai and Wachter [2015] and the references therein). In terms of link to the data, the current paper is particularly focused on the empirical results of Julliard and Ghosh [2012]. Our goal is not to add to this important empirical evidence but rather to discuss its methodological underpinnings. As explained above, although we are able to confirm mathematically the main intuition of these authors that the subjective empirical beliefs deliver a rare events distribution that assigns relatively higher weights to a few particular bad states, we question the statistical meaning of this observation by proving an impossibility theorem.

This theorem which puts forward rather realistic circumstances in which the possibility of disasters of unbounded amplitude precludes the existence of population distorted beliefs is a minor extension of a result first proved by Chen, Hansen, and Hansen [2021]. It confirms the problematic asymptotic behaviour of the empirical likelihood estimator in the presence of misspecification as documented by Schennach [2007]. The latter paper shows that, even when postulating the existence of a pseudo-true value, there does not exist a root-T consistent estimator of it. Our impossibility theorem even stresses that a unique pseudo-true value may not exist.

By contrast, we apply Csiszar [1995]'s "generalized projections for non-negative functions"to provide sufficient conditions for the existence of a population distorted belief solution of minimization of a general ϕ-divergence function. While boundedness of disasters amplitude is a sufficient condition (under standard regularity conditions), the boundedness assumption can be relaxed if we consider only increasing ϕ-divergence functions. Recent work by Cerreia-Vioglio, Hansen, Maccheroni, and Marinacci [2021] provides a first step to extend the min-max analysis under model ambiguity by also considering ϕ-divergence functions to acknowledge that the model used in decision-making is a simplified approximation. Both our impossibility theorem and our existence theorems set the focus on unconditional mean restrictions obtained by integrating out the conditional moment restrictions (1). This is conformable to the empirical strategy of Julliard and Ghosh [2012] and not restrictive as explained by Hansen and Jagannathan [1997] through the concept of actively managed portfolios (see Section 2.1. below). An alternative approach would be to refer to Komunjer and Ragusa [2016]'s "conditional density projections" to define directly conditional distorted subjective beliefs.

The extant literature also suggests some interesting connections to make between the ambiguity approach as developed by Jeong, Kim, and Park [2015] and asset pricing under disaster risk as discussed above. First we note that, similarly to disasters of bounded amplitude, the multiple-priors recursive utility model with κ-ignorance only considers a bounded set of possible scenarios (κ is an upper bound for the density generator of different priors). The model identifies the true unknown value of the parameters by imposing the martingale condition for pricing error. As noted by Jeong, Kim, and Park [2015], "the spirit of the methodology is therefore somewhat similar to the GMM estimation for the nonlinear Euler equation models", or more generally to the minimization of ϕ-divergence subject to the conditional moment restrictions (1).

In Jeong, Kim, and Park [2015], the main trick for estimation, following the general method of "Martingales Regressions for Conditional Mean Models"developed by Park [2021], is based on the theorem of Dambis, Dubins and Schwarz, that allows to convert (by a well-suited time change) any continuous martingale into Brownian motion. The actual martingale estimator is defined as a minimum distance estimator based on the discrepancy between the empirical distributions of normalized pricing errors after time change and the standard normal distribution. The boundedness assumption on the conditional mean in the κ-ignorance model makes easy the application of the martingale regression for estimation of the ambiguity model similarly to the estimation of population subjective beliefs in the case of bounded disasters. There is an obvious analogy between considering only disasters of bounded amplitude and only bounded worst case scenarios in the κ-ignorance model of ambiguity.

This connection between asset pricing with disaster risk and ambiguity aversion, as captured by multiple-priors recursive utility, paves the way for a potentially unified framework. For instance, it would be worth checking that, as in the distorted beliefs framework of Julliard and Ghosh [2012], the min-max approach to multiple priors also leads to distorted beliefs that make economic sense because "the rare events distribution (...) assigns relatively higher weights to a few particular bad states of the economy".

1.2 Outline of the paper

Section 2 defines the so-called model-implied probabilities, that are empirical probabilities computed by the minimization of a ϕ-divergence with respect to the empirical distribution, as characterizing our empirical distorted subjective beliefs. We briefly discuss the choice of a specific ϕ-divergence function. While Chaudhuri and Renault [2020] had shown an asymptotic equivalence result between implied probability distributions corresponding to different ϕ-divergence functions in case of a well-specified asset pricing model (1), we show that it cannot be the case when the asset pricing model is misspecified. Therefore, it is only in the case of misspecified models that it is worth considering distorted subjective beliefs. For two of the most popular ϕ-divergence functions, Empirical Likelihood (EL) and Euclidean Empirical Likelihood (EEL), we show that, if we assume that the distorted subjective beliefs converge in distribution towards a population distribution, then this distribution should confirm the main intuition of the rare event hypothesis, namely that "the rare events distribution needed to rationalize the EPP assigns relatively higher weights to a few particular bad states of the economy".

In Section 3, we argue that in the case of possibly unbounded risk, the model-implied empirical probability distribution cannot be safely interpreted as estimator of a meaningful population probability distribution because such a model-implied probability distribution may not exist if the asset pricing model is misspecified. The non-existence of a model-implied population probability distribution is caused by the conjunction of two effects:

(i) Unbounded vector $g(X_{t+1}, \theta)$ of pricing errors,

(ii) Decreasing divergence function ϕ.

We then conclude that the researcher is faced with the following vicious circle:

(i) Either the asset pricing model is well-specified, and then the model-implied population probability distribution coincides with the naive historical distribution.

(ii) Or the asset pricing model is misspecified, and then the model-implied population probability distribution depends on the choice of a ϕ-divergence function. Since, following Chen, Hansen, and Hansen [2021]'s terminology, empirical likelihood is a problematic ϕ-divergence function (the associated model-implied population probability distribution may not exist), there is no such thing as a natural choice to replace the naive historical distribution by a well-suited population distribution of distorted subjective beliefs.

We prove in Section 4 two theorems of existence based on the assumption that either (i) or (ii) above does not hold, i.e. based respectively on the assumption of bounded pricing errors $g(X_{t+1}, \theta)$ or on the assumption of increasing divergence function ϕ.

Section 5 concludes and discusses possible alternative strategies because the results of the current paper lead us to share with Chen, Hansen, and Hansen [2021] the opinion that "we do not see why the subjective beliefs of market participants must appear to the econometrician to have minimal divergence relative to rational expectations".

2 Why Model-Implied Probabilities may confirm the Rare Events Hypothesis?

We want to allow for the beliefs that are revealed by the market to differ from the rational expectations beliefs (the historical distribution) implied by infinite histories of the data, assuming that the observed process $X_t, t = 1, 2, ...$, is strictly stationary and ergodic for the historical distribution. We will do so for any given possible value $\theta \in \Theta$ of the parameters.

2.1 Model-implied Empirical Probabilities

The information theoretic approaches to inference in moment condition models have become popular in econometrics since the seminal paper of Imbens, Spady, and Johnson [1998], by applying Corcoran [1998]'s minimum contrast inference strategy to the Cressie-Read family of power divergences. For a given observed sample $X_{t+1}, t = 1, 2, ..., T$, and a divergence function ϕ, we follow Corcoran [1998] by considering the minimization program over T-dimensional vectors $\pi_T = (\pi_{t,T})_{1 \leq t \leq T}$:

$$\min_{\pi_T \in \mathbb{R}^T} \sum_{t=1}^{T} \phi\left(T\pi_{t,T}\right) \tag{2}$$

subject to:

$$\sum_{t=1}^{T} \pi_{t,T} = 1, \sum_{t=1}^{T} \pi_{t,T} g\left(X_{t+1}, \theta\right) = 0 \tag{3}$$

where the function ϕ is a given strictly convex function for which $\phi(1) = 0$. The strict convexity of the function ϕ allows us to apply Jensen's inequality to conclude that:

$$\frac{1}{T} \sum_{t=1}^{T} \phi\left(T\pi_{t,T}\right) \geq \phi\left[\sum_{t=1}^{T} \frac{1}{T} T\pi_{t,T}\right] = \phi(1) = 0$$

with a strict inequality except if $\pi_{t,T}$ is independent of t, that is if and only if:

$$\pi_{t,T} = \frac{1}{T}, \forall t = 1, ..., T. \tag{4}$$

Therefore, the solution $\hat{\pi}_T(\theta) = (\hat{\pi}_{t,T}(\theta))_{1 \leq t \leq T}$ of (2)/(3)) is given by:

$$\hat{\pi}_{t,T}(\theta) = \frac{1}{T}, \forall t = 1, ..., T \tag{5}$$

if and only if the moment restrictions are fulfilled in sample:

$$\frac{1}{T} \sum_{t=1}^{T} g\left(X_{t+1}, \theta\right) = 0.$$

More generally, the vector $\hat{\pi}_T(\theta) = (\hat{\pi}_{t,T}(\theta))_{1 \leq t \leq T}$ can be interpreted as a distorted empirical probability distribution which, by construction, ensures the validity of the moment restrictions:

$$\sum_{t=1}^{T} \hat{\pi}_{t,T}(\theta)\, g(X_{t+1}, \theta) = 0 \tag{6}$$

while being the closest possible (in the sense of the divergence function ϕ) to the empirical distribution (4). Note that we interpret as an expectation operator denoted $\hat{E}_{\phi}[.]$ the operator which associates to any numerical function $\psi(.)$ the real number

$$\hat{E}_{T}^{\theta,\phi}[\psi(X_{t+1})] = \sum_{t=1}^{T} \hat{\pi}_{t,T}(\theta)\, \psi(X_{t+1}).$$

This interpretation may actually be an abuse of notation since no non-negativity constraint for $\hat{\pi}_{t,T}(\theta)$ is maintained in the minimization program (2)/(3). This abuse of notation will be at stake throughout the paper without explicit mention of it.

The bottom line is that the T numbers $\hat{\pi}_{t,T}(\theta), t = 1, ..., T$ are model-implied empirical probabilities which ensure in sample the validity of the asset pricing equation (1) for a given value θ of the parameters:

$$\hat{E}_{T}^{\theta,\phi}[g(X_{t+1}, \theta)] = 0. \tag{7}$$

Several remarks are in order:

First, empirical expectation (7) takes into account conditional moment restrictions (1) only through its unconditional implication:

$$E_0[g(X_{t+1}, \theta)] = 0 \tag{8}$$

where $E_0[]$ stands for the expectation operator of the true (historical) unknown distribution of the stationary process (X_t).

We have known since Hansen and Jagannathan [1997] that this is not restrictive since the vector R_{t+1} of gross returns can include not only some primitive assets but also actively managed portfolio returns built on these primitive assets. Since the shares of investment in primitive assets to define the actively managed portfolios can include any function of the conditioning information set $I(t)$, the unconditional moment conditions (8) may arguably summarize all the conditional information provided by the asset pricing model of interest and observation of asset returns.

Second, having conditional moment restrictions (1) in the background, we do not consider the possibility to pre-average the consecutive values $g\left(X_{t+h},\theta\right)$ for $h=\pm1,2,...,H_T$ with a convenient bandwidth H_T to take care of serial dependence in moment functions (see, e.g., Kitamura and Stutzer [1997]). While (1) tells us that $g\left(X_{t+1},\theta^0\right)$ is a martingale difference sequence for the true unknown value θ^0 of θ (if it exists), it is not the case for other values of $\theta\in\Theta$ and then, pre-averaging may be relevant. However, as explained later, it would not change the substance of our empirical discussions.

Third, it is precisely because $\hat{\pi}_{t,T}\left(\theta\right)$ may differ from the sample distribution (5) that some belief distortions are at stake. Revisiting the analysis of Julliard and Ghosh [2012] we will discuss later how these distortions can be interpreted.

Fourth, when the asset pricing model (1) is well-specified and $\theta=\theta^0$, the true value of the parameters, we expect that even the distorted empirical distribution converges weakly towards the true unknown probability distribution, meaning that for any bounded function $\psi\left(X_t\right)$, we have:

$$\operatorname*{plim}_{T\to\infty}\hat{E}_T^{\theta,\phi}\left[\psi\left(X_{t+1}\right)\right]=\operatorname*{plim}_{T\to\infty}\left[\frac{1}{T}\sum_{t=1}^{T}\psi\left(X_{t+1}\right)\right]=E_0\left[\psi\left(X_{t+1}\right)\right].$$

Fifth, when the moment conditions (8) are not fulfilled for the given value θ, we may hope that the model-implied empirical probabilities $\hat{\pi}_{t,T}\left(\theta\right)$ still define asymptotically a population distribution, but it will be a distribution of distorted subjective beliefs, that will differ from the true one. We will write:

$$\operatorname*{plim}_{T\to\infty}\hat{E}_T^{\theta,\phi}\left[\psi\left(X_{t+1}\right)\right]=\tilde{E}^{\theta,\phi}\left[\psi\left(X_{t+1}\right)\right]\tag{9}$$

where $\tilde{E}_{\phi}^{\theta}$ is an expectation operator different from E_0. The discussion of existence and properties of these population distorted beliefs conformable to (9) is one of the main focuses of interest of this paper. We first discuss in the next subsection what would be a population analog of the minimization program (2)/(3).

2.2 Model-Implied Population Probability Distribution

We consider throughout this subsection a given value θ of the parameters and, taking advantage of the stationarity of the historical distribution of X_{t+1}, we simplify the notations by writing for any function ξ such that $\xi\left[g\left(X_{t+1},\theta\right)\right]$ is integrable:

$$E_0\left[\xi\left[g\left(X_{t+1},\theta\right)\right]\right]=E_0\left[\xi\left[Y\left(\theta\right)\right]\right].$$

The population analog of the minimization program (2)/(3) is written as:

$$\min_M E_0\left[\phi\left(M\left[Y\left(\theta\right)\right]\right)\right] \tag{10}$$

subject to:

$$E_0\left[M\left[Y\left(\theta\right)\right]\right]=1, E_0\left[M(Y\left(\theta\right))Y\left(\theta\right)\right]=0. \tag{11}$$

In particular, if the historical distribution of $Y\left(\theta\right)$ is characterized by a probability density function $f_Y(.\left|\theta\right)$ with respect to some σ-finite measure λ, then:

$$E_0\left[M\left(Y\left(\theta\right)\right)Y\left(\theta\right)\right]=E_M\left[Y\left(\theta\right)\right]=\int y f_Y^M(y\left|\theta\right)d\lambda(y)$$

where the distorted probability density function $f_Y^M(y\left|\theta\right)$ of $Y\left(\theta\right)$ is defined by its Radon-Nikodym derivative $M\left(y\right)$ with respect to the historical distribution:

$$f_Y^M(y\left|\theta\right)=M(y)f_Y(y\left|\theta\right).$$

The rationale for this minimization is an obvious implication of Jensen's inequality (jointly with $\phi(1)=0$), telling us that on the one hand:

$$E_0\left[M\left[Y\left(\theta\right)\right]\right]=1\Longrightarrow E_0\left[\phi(M\left[Y\left(\theta\right)\right])\right]\geq 0$$

and on the other hand when not only $E_0\left[M\left[Y\left(\theta\right)\right]\right]=1$ but also $E_0\left[M(Y\left(\theta\right))Y\left(\theta\right)\right]=0$, then:

$$E_0\left[\phi(M)\right]=0 \Longleftrightarrow E_0\left[Y\left(\theta\right)\right]=0.$$

In particular, the value of the minimization programm (10)/(11) is zero if and only if $E_0\left[Y\left(\theta\right)\right]=0$, that is when the pricing model (1) is well-specified and θ is a true value of the parameters. In this case, the minimum is reached at only one (up to an almost sure equality) change of measure $M\left[Y\left(\theta\right)\right]$ which is identical to the constant 1.

Otherwise, if it exists, a solution $M_\phi^\theta\left[Y(\theta)\right]$ is a non-degenerate random variable, and when the historical probability distribution is defined by a density function $f_Y(y\left|\theta\right.)$, a density function for distorted beliefs is given by:

$$f_Y^{M_\phi^\theta}(y\left|\theta\right.)=M_\phi^\theta\left(y\right)f_Y^0(y\left|\theta\right.).$$

In any case, when it exists, the change of measure given by $M_\phi^\theta\left[Y(\theta)\right]$ defines distorted expectations of bounded functions $\xi\left[Y(\theta)\right]$:

$$E^{\theta,\phi}\left[\xi\left[Y(\theta)\right]\right]=E_0\left[M_\phi^\theta\left[Y\left(\theta\right)\right]\xi\left[Y(\theta)\right]\right].$$

In particular, a maintained assumption is that distorted beliefs are absolutely continuous with respect to historical ones. They cannot assign positive probabilities to events that would almost surely not happen historically.

Obviously, since (10)/(11) is the population analog of (2)/(3) we expect that the solutions of the programs, when existing, should be related asymptotically. More precisely, we expect that for any bounded function $\xi\left[g\left(X_{t+1},\theta\right)\right]$:

$$\operatorname*{plim}_{T\to\infty}\hat{E}_T^{\theta,\phi}\left[\xi\left[g\left(X_{t+1},\theta\right)\right]\right]=\tilde{E}^{\theta,\phi}\left[\xi\left[g\left(X_{t+1},\theta\right)\right]\right]=E^{\theta,\phi}\left[\xi\left[g\left(X_{t+1},\theta\right)\right]\right]. \qquad (12)$$

However, it is worth keeping in mind that while the finite sample solution $\hat{E}_T^{\theta,\phi}\left[\xi\left[g\left(X_{t+1},\theta\right)\right]\right]$ always exists, while its limit $\tilde{E}^{\theta,\phi}\left[\xi\left[g\left(X_{t+1},\theta\right)\right]\right]$ may exist under convenient regularity conditions, we do not know yet whether the population solution $E^{\theta,\phi}\left[\xi\left[g\left(X_{t+1},\theta\right)\right]\right]$ may exist. As already announced, we will present in Sections 3 and 4 respectively, situations where it does not (resp. it does) exist.

2.3 The Cases of Empirical Likelihood

Empirical Likelihood (EL) is a ϕ-divergence corresponding to:

$$\phi_L(m) = -\log(m).$$

Owen [2001] has dubbed Euclidean Empirical Likelihood (EEL) the quadratic approximation of EL in the neighborhood of $m = 1$:

$$\phi_Q(m) = \frac{m^2 - 1}{2}.$$

This approach is underpinned by the fact that with EEL, one is led to minimize $E_0[\phi_Q(m)]$, which is nothing but the first term in the Taylor series expansion of $E_0[\phi_L(m)]$ in the neighborhood of $m = 1$. The fact that $\phi_Q(m)$ is a quadratic function allows us to see the program (2)/(3) as a quadratic program subject to linear restrictions so that we get a solution in closed form that can be written (see, e.g., Chaudhuri and Renault [2020]) as a function of the first two empirical moments:

$$\hat{\pi}_{t,T}^{Q}(\theta) = \frac{1}{T} - \bar{Y}_T(\theta)' \left[V_T(Y_t(\theta))\right]^{-1} \frac{1}{T}\left[Y_t(\theta) - \bar{Y}_T(\theta)\right] \tag{13}$$

where:

$$\begin{aligned} Y_t(\theta) &= g(X_{t+1}, \theta), \bar{Y}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} Y_t(\theta), \\ V_T(Y_t(\theta)) &= \frac{1}{T}\sum_{t=1}^{T} Y_t(\theta)\left[Y_t(\theta) - \bar{Y}_T(\theta)\right]'. \end{aligned}$$

Even though no closed form formula is available in the case of genuine EL, Chaudhuri and Renault [2020] have shown that an equation formally similar to (13) is available for the solution of the EL program:

$$\hat{\pi}_{t,T}^{L}(\theta) = \frac{1}{T} - \bar{Y}_T(\theta)' \left[V_T^{\theta,L}(Y_t(\theta)) \right]^{-1} \hat{\pi}_{t,T}^{L}(\theta) Y_t(\theta) \tag{14}$$

where:

$$V_T^{\theta,L}(Y_t(\theta)) = \sum_{t=1}^{T} \hat{\pi}_{t,T}^{L}(\theta) Y_t(\theta) Y_t(\theta)'. \tag{15}$$

Obviously, (15) can be interpreted as an empirical variance but where the sample distribution (4) has been replaced by the model-implied probabilities $\hat{\pi}_{t,T}^{L}(\theta)$. It is worth noting that in spite of the similarities of formulas, (14) does not deliver, by contrast with (13) for EEL implied probabilities, closed form formulas for EL implied probabilities $\hat{\pi}_{t,T}^{L}(\theta)$. Not only the implied probability $\hat{\pi}_{t,T}^{L}(\theta)$ is explicitly on the RHS of (14), but even solving for it would not give it in closed form since all the implied probabilities $\hat{\pi}_{\tau,T}^{L}(\theta), \tau = 1, 2, ..., T$, are hidden within the matrix $\left[V_T^{L}(Y_t(\theta)) \right]$.

The distorted subjective belief distributions obtained by maximization of EL and minimization of quadratic divergence EEL respectively are equivalently defined by associated expectation operators for any bounded function $\xi[Y_t(\theta)]$:

$$\hat{E}_T^{\theta,A}[\xi[Y_t(\theta)]] = \sum_{t=1}^{T} \hat{\pi}_{t,T}^{A}(\theta)\xi[Y_t(\theta)], A \in \{Q, L\}.$$

We deduce from (13) and (14) that:

$$\hat{E}_T^{\theta,Q}[\xi[Y_t(\theta)]] = \frac{1}{T}\sum_{t=1}^{T} \xi[Y_t(\theta)] - \bar{Y}_T(\theta)' [V_T[Y_t(\theta)]]^{-1} Cov_T[Y_t(\theta), \xi[Y_t(\theta)]] \tag{16}$$

and:

$$\hat{E}_T^{\theta,L}[\xi[Y_t(\theta)]] = \frac{1}{T}\sum_{t=1}^{T} \xi[Y_t(\theta)] - \bar{Y}_T(\theta)' \left[V_T^{\theta,L}[Y_t(\theta)] \right]^{-1} Cov_T^{\theta,L}[Y_t(\theta), \xi[Y_t(\theta)]] \tag{17}$$

where:

$$
\begin{aligned}
Cov_T\left[Y_t(\theta), \xi\left[Y_t(\theta)\right]\right] &= \frac{1}{T}\sum_{t=1}^{T} Y_t(\theta)\,\xi\left[Y_t(\theta)\right]' - \bar{Y}_T(\theta)\left(\frac{1}{T}\sum_{t=1}^{T}\xi\left[Y_t(\theta)\right]\right)' \\
Cov_T^{\theta,L}\left[Y_t(\theta), \xi\left[Y_t(\theta)\right]\right] &= \sum_{t=1}^{T}\hat{\pi}_{t,T}^{L}(\theta)Y_t(\theta)\,\xi\left[Y_t(\theta)\right]'.
\end{aligned}
$$

Note that the last formula for covariance is justified by the fact that, by definition:

$$
E_T^{\theta,L}\left[Y_t(\theta)\right] = 0.
$$

Assuming at this stage, for sake of simplified interpretation, that all the quantities considered in (12) exist for both ϕ-divergences ϕ_Q and ϕ_L, we are led to the definition of population probability distributions for which the expectation operators are defined (with obvious notations) as:

$$
E^{\theta,Q}\left[\xi\left(Y_t(\theta)\right)\right] = E\left[\xi\left(Y_t(\theta)\right)\right] - E\left[Y_t(\theta)\right]'\left[Var\left[Y_t(\theta)\right]\right]^{-1} Cov\left[Y_t(\theta), \xi\left[Y_t(\theta)\right]\right] \tag{18}
$$

and:

$$
E^{\theta,L}\left[\xi\left(Y_t(\theta)\right)\right] = E\left[\xi\left(Y_t(\theta)\right)\right] - E\left[Y_t(\theta)\right]'\left[Var^{\theta,L}\left[Y_t(\theta)\right]\right]^{-1} Cov^{\theta,L}\left[Y_t(\theta), \xi\left[Y_t(\theta)\right]\right] \tag{19}
$$

where:

$$
\begin{aligned}
Var^{\theta,L}\left[Y_t(\theta)\right] &= E^{\theta,L}\left[Y_t(\theta)\,Y_t(\theta)'\right] \\
Cov^{\theta,L}\left[Y_t(\theta), \xi\left[Y_t(\theta)\right]\right] &= E^{\theta,L}\left[Y_t(\theta)\,\xi\left[Y_t(\theta)\right]'\right].
\end{aligned}
$$

Note that by contrast with (18), (19) does not give an explicit definition of a probability distribution. The operator $E^{\theta,L}[.]$ is defined as implicit solution of the equation (19), while it shows up not only on the LHS but also twice on the RHS (for the definition of $Var^{\theta,L}$ and $Cov^{\theta,L}$). Assuming that both the historical, the EEL and the EEL probability distributions are absolutely continuous with respect to the same measure on $\mathbb{R}^n$, we deduce from (18) and (19) that their respective density functions $f_Y(.\,|\theta)$, $f_Y^{\theta,Q}(.\,|\theta)$ and $f_Y^{\theta,L}(.\,|\theta)$ are related as follows:

$$f_Y^{\theta,Q}(y\,|\theta) = f_Y(y\,|\theta) - E\left[Y(\theta)\right]' \left[Var\left[Y(\theta)\right]\right]^{-1} \left[y - E\left[Y(\theta)\right]\right] f_Y(y\,|\theta) \qquad (20)$$

and:

$$f_Y^{\theta,L}(y\,|\theta) = f_Y(y\,|\theta) - E\left[Y(\theta)\right]' \left[Var^{\theta,L}\left[Y_t(\theta)\right]\right]^{-1} y f_Y^{\theta,L}(y\,|\theta). \qquad (21)$$

Note that (21) defines implicitly the density function $f_Y^{\theta,L}(.\,|\theta)$ as solution of an equation that contains it once in the LHS and twice in the RHS (one of them to compute $Var^{\theta,L}$).

2.4 The rare disaster interpretation

Both, the explicit formula (20) and the implicit formula (21) allow us to find some theoretical underpinnings for the common intuition that disaster risk may help to rationalize the Equity Premium Puzzle. To follow the counterfactual analysis in Julliard and Ghosh [2012] means: (i) fixing the risk aversion parameter or more generally our vector θ of parameters to a “reasonable” value (relative risk aversion parameter fixed to 10 in their case), and (ii) then asking the EEL and EL “estimation procedures to identify the distribution of the data that would solve the EPP in the historical sample”. According to Julliard and Ghosh [2012], “the first question to ask is whether the implied state probabilities make economic sense. A priori, we would expect that the rare events distribution needed to rationalize the EPP assigns relatively higher weights to a few particular bad states of the economy. Figure 4 suggests that this is exactly what the estimated” probabilities do.

To address this issue, we first note that:

$$E\left[Y(\theta)\right] = Cov\left[S(\theta), R\right] + E\left[S(\theta)\right] E[R] - e_n.$$

Therefore, if for instance, for the sake of expositional simplicity, we consider that the risk free asset is properly priced:

$$E\left[S\left(\theta\right)\right]=\frac{1}{R_F}$$

where R_F stands for the risk-free return in any unit period (interest rate risk is overlooked for expositional simplicity), then we will have:

$$E\left[Y\left(\theta\right)\right]=Cov\left[S\left(\theta\right),R\right]+\frac{1}{R_F}EPR$$

where:

$$EPR=E[R]-R_F e_n$$

is the equity premium vector for the vector of n risky assets under consideration. Therefore, if the equity premium of a given asset $R_{i,t+1}$ is larger than the covariance with the counterfactual SDF $S_{t+1}\left(\theta\right)$ can explain, we will have:

$$E\left[Y_{i,t+1}\left(\theta\right)\right]>0.$$

Let us then consider a collection of asset returns $R_{i,t+1}, i=1,..,n$, some of the assets being exactly priced ($E\left[Y_{i,t+1}\left(\theta\right)\right]=0$) while the other ones displaying a risk premium higher than the model prediction ($E\left[Y_{i,t+1}\left(\theta\right)\right]>0$). We assume in addition that the joint precision matrix $C(\theta)=\left[Var\left[Y\left(\theta\right)\right]\right]^{-1}$ of the pricing errors $Y_{i,t+1}\left(\theta\right), i=1,..,n$ has only non-negative coefficients $c_{ij}\left(\theta\right), i,j=1,.,n$. Note that this is in particular true if the n pricing errors $Y_{i,t+1}\left(\theta\right), i=1,..,n$ are pairwise non-correlated. By revisiting the popular concept of multivariate total positivity of order 2 (see, e.g., Chapter 2, Joe [1997]), we could say that we assume more generally a kind of multivariate total negativity of the pricing errors (see Example 2.2 and Exercise 2.19, Joe [1997]). It turns out that without this assumption, the comparison between the historical distribution density $f_Y(y\,|\theta\,)$ and the EEL distribution density $f_Y^{\theta,Q}(y\,|\theta\,)$ would be ambiguous. As shown in Example 1 below, positive dependence between different asset returns will give rise to more likely disasters through contagion, even though their likelihood is not captured by a high value of $f_Y^{\theta,Q}(y\,|\theta)$.

To figure out the comparison between $f_Y(y\,|\theta\,)$and $f_Y^{\theta,Q}(y\,|\theta\,)$ in a bad state y of the economy, we first deduce from (20) that:

$$f_Y^{\theta,Q}(y\,|\theta\,) = f_Y(y\,|\theta\,) - \left\{ \sum_{1\leq i,j\leq n} c_{ij}\,(\theta)\, E\left[Y_{j,t+1}\,(\theta)\right]\left\{y_i - E\left[Y_{i,t+1}\,(\theta)\right]\right\} \right\} f_Y(y\,|\theta\,).$$

This formula shows that if $y = (y_i)_{1\leq i\leq n}$ is a "bad state" because:

$$y_i < E\left[Y_{i,t+1}\,(\theta)\right], \forall i = 1, ..., n$$

then:

$$f_Y^{\theta,Q}(y\,|\theta\,) > f_Y(y\,|\theta\,).$$

Let us consider a simple example to further illustrate the underlying ideas. For notational brevity, we will in general suppress in this example the dependence of the concerned quantities on the given θ.

Example 1: Assume for sake of notational simplicity that $n = 2$ and let ρ stand for the correlation between the two asset pricing errors with $-1 < \rho < 1$, so that the variance matrix can be inverted. The analysis could be easily extended to more than two assets by considering instead partial correlations between two pricing errors given the other ones. Let us assume that the first asset is accurately priced while the second one displays some overly high risk premium:

$$E(Y_1) = 0, E(Y_2) = \mu_2 > 0.$$

Then, recalling that $C = \left[Var\,(Y_1, Y_2)'\right]$, with elements denoted by c_{ij}, is the joint precision matrix of the pricing errors, we obtain that:

$$\begin{aligned}
&\sum_{1\leq i,j\leq 2} c_{ij}\,(\theta)\, E\left[Y_{j,t+1}\,(\theta)\right]\left\{y_i - E\left[Y_{i,t+1}\,(\theta)\right]\right\} \\
&= \quad c_{12}\mu_2 y_1 + c_{22}\mu_2\,(y_2 - \mu_2) \\
&= \quad \frac{\mu_2}{(1-\rho^2)Var\,(Y_2)}\left\{ y_2 - \mu_2 - \rho\left[\frac{Var\,(Y_2)}{Var\,(Y_1)}\right]^{1/2} y_1 \right\}.
\end{aligned}$$

By abuse of notation, let us denote the affine regression of Y_2 on Y_1 as a conditional expectation. It is correct in case of joint normality and all statements below remain valid in case of affine regression. Then the above formula shows that the relative difference between the historical density $f_Y(y\,|\theta\,)$ and the EEL implied one $f_Y^{\theta,Q}(y\,|\theta\,)$ is:

$$\frac{f_Y^{\theta,Q}(y\,|\theta\,) - f_Y(y\,|\theta\,)}{f_Y(y\,|\theta\,)} = \frac{-\mu_2}{(1-\rho^2)Var\,(Y_2)}\left\{y_2 - E[Y_2\,|Y_1 = y_1]\right\}. \tag{22}$$

While $c_{22} = \left[\left(1-\rho^2\right)Var(Y_1)\right]^{-1}$ is positive by definition, the sign of $c_{12} = -\rho\left(1-\rho^2\right)^{-1}$ $[Var(Y_1)Var(Y_2)]^{-1/2}$ will be crucial to figure out the impact on implied probabilities of a bad state, such that:

$$y_1 < E(Y_1) = 0, y_2 < E(Y_2) = \mu_2.$$

We see with (22) that the sign of the relative difference between densities is the opposite of the sign of the conditional surprise on Y_2:

$$v_2 = y_2 - E[Y_2\,|Y_1 = y_1]\,.$$

Then, if $c_{12} \geq 0$, meaning that $\rho \leq 0$ and $v_2 \leq y_2 - \mu_2 < 0$ since:

$$E[Y_2\,|Y_1 = y_1] = \mu_2 + \rho\left[\frac{Var\,(Y_2)}{Var\,((Y_1)}\right]^{1/2} y_1 \geq \mu_2.$$

By contrast, if $\rho > 0$, then $E[Y_2\,|Y_1 = y_1] < E(Y_2)$ so that, even though (y_1, y_2) is a bad state, it is still possible that $v_2 = y_2 - E[Y_2\,|Y_1 = y_1] > 0$ so that:

$$f_Y^{\theta,Q}(y\,|\theta\,) < f_Y(y\,|\theta\,).$$

The contagion (or correlation) effect makes the conditionally pricing error $E[Y_2\,|Y_1 = y_1]$ worse than the unconditionally expected pricing error $E(Y_2)$, so that it may be that y_2 is not conditionally a bad state (it exceeds the conditional expectation given other assets). Hence, while (y_1, y_2) had been defined unconditionally as a bad state, it is not true any-more that the rare events distribution needed to rationalize the EPP assigns relatively higher weight to this state. ■

Overall, at least for EEL, it is true that, when precluding overly influential contagion effects, "the rare events distribution needed to rationalize the EPP assigns relatively higher weights to a few particular bad states of the economy".

However, in their empirical study, Julliard and Ghosh [2012] do not use EEL but EL. Formula (21) shows that EL should lead to the same kind of conclusion as EEL, at least if we define "bad state" by:

$$y_i < 0, \forall i = 1, ..., n$$

and the maintained assumption about implied precision matrix (assumption of non-negative coefficients) is about the EL implied precision matrix $\left[Var^{\theta,L}\left[Y\left(\theta\right)\right]\right]^{-1}$.

3 Disaster Risk and Problematic Characterization of Distorted Beliefs

3.1 Unidentified beliefs under disaster risk

We show in this subsection that when an excess of equity premium (or more generally $E\left[Y(\theta)\right] \neq 0$) not explained by the asset pricing model (under historical distribution) is matched by a disaster risk, it may lead to the non-existence of a well defined model-implied probability distribution.

To see that, we first introduce the notation $E_\lambda(Z\,|B\,)$ for any Borel set B such that $0 < \lambda(B) < +\infty$:

$$E_\lambda(Z\,|B) = \frac{1}{\lambda(B)} \int_B z d\lambda(z).$$

In other words, $E_\lambda(Z\,|B\,)$ is the expectation of the probability distribution on $\mathbb{R}^n$ defined by:

$$P_\lambda(A\,|B\,) = \frac{\lambda(A \cap B)}{\lambda\,(B)}.$$

For instance, if λ is the Lebesgue measure on $\mathbb{R}^n$, $P_\lambda(.\,|B\,)$ is the uniform probability distribution on B.

We maintain in this subsection the following assumption:

Assumption "Undetected Misfit" (mispricing matched by unbounded disaster risk):

For a given $\theta \in \Theta$ such that $E\left[Y(\theta)\right] \neq 0$, there exists a sequence B_j of Borel sets of $\mathbb{R}^n$ and a sequence α_j of positive real numbers such that for all $j = 1, 2, ..$:

$$\begin{aligned} 0 &< \lambda\left(B_j\right) < +\infty \\ E_\lambda(Z\left|B_j\right.) &= -\alpha_j E\left[Y(\theta)\right], \lim_{j\to\infty} \alpha_j = +\infty \\ f_Y(y\left|\theta\right.) &> 0, \forall y \in B_j. \ \blacksquare \end{aligned}$$

For instance, if λ is the Lebesgue measure on $\mathbb{R}^n$, B_j may be a ball with center $\left[-\alpha_j E\left[Y(\theta)\right]\right]$. We can then prove the following theorem:

Theorem 1: If ϕ is a decreasing function on $\mathbb{R}^+$, continuous at $m = 1$ (with $\phi(1) = 0$), under assumption "Undetected Misfit", there exists a sequence of changes of measure $M^{(j)}(y\left|\theta\right.)$ such that for all $j = 1, 2, ..$:

$$E\left[M^{(j)}(Y(\theta)\left|\theta\right.)\right] = 1,\ E\left[M^{(j)}(Y(\theta)\left|\theta\right.)Y\left(\theta\right)\right] = 0$$

and:

$$\lim_{j\to\infty} E\left\{\phi\left[M^{(j)}(Y(\theta)\left|\theta\right.)\right]\right\} = 0. \ \blacksquare$$

The interpretation of Theorem 1 is clear. It proves that under assumption "Undetected Misfit", with a decreasing contrast function ϕ, the minimization problem (10) subject to (11) does not admit a solution. We can find changes of measure M, fulfilling the constraints (11) and making $E\left[\phi(M)\right]$ arbitrarily close to zero but the lower bound zero cannot be reached since the pricing model is misspecified. This impossibility theorem is important since it can be applied in particular to empirical likelihood that corresponds to a divergence function:

$$\phi_L(M) = -\log(M).$$

It means in particular that the empirical exercise performed by Julliard and Ghosh [2012], as described in Section 2 above, may not be meaningful because there would be no such thing as a model-implied population probability distribution $f_Y^{\theta,L}(y\,|\theta\,)$ for which the empirical model-implied distribution would be a consistent estimator.

It is worth looking at the proof of Theorem 1 to get convinced that, insofar as we reckon the possibility of unbounded disasters, the case of impossibility put forward by Theorem 1 is not unrealistic. This proof is a slight generalization of a proof first proposed in Chen, Hansen, and Hansen [2021]. The trick is to define a sequence of changes of measure as follows:

$$M^{(j)}(Y(\theta)\,|\theta\,) = 1 - \pi_j + \frac{\pi_j}{\lambda\,(B_j)} \frac{1_{B_j}\,[Y(\theta)]}{f_Y(Y(\theta)\,|\theta\,)}.$$

This variable is well-defined since by assumption "Undetected Misfit":

$$Y(\theta) \in B_j \Longrightarrow f_Y(Y(\theta)\,|\theta\,) > 0.$$

By construction:

$$E\left[M^{(j)}(Y(\theta)\,|\theta\,)\right] = 1 - \pi_j + \frac{\pi_j}{\lambda\,(B_j)} \int_{B_j} d\lambda(y) = 1$$

while:

$$\begin{aligned} E\left[M^{(j)}(Y(\theta)\,|\theta\,)Y\,(\theta)\right] &= (1-\pi_j)\,E\,[Y\,(\theta)] + \frac{\pi_j}{\lambda\,(B_j)} \int_{B_j} y d\lambda(y) \\ &= (1-\pi_j)\,E\,[Y\,(\theta)] + \pi_j E_\lambda(Z\,|B_j) \\ &= (1-\pi_j)\,E\,[Y\,(\theta)] - \pi_j \alpha_j E\,[Y(\theta)] \\ &= 0 \Longleftrightarrow \pi_j = \frac{1}{1+\alpha_j}. \end{aligned}$$

Thus, by applying assumption "Undetected Misfit", we conclude that the moment matching implies that:

$$\lim_{j \to \theta} \pi_j = 0.$$

In other words, it is precisely because the disaster risk is unbounded that we have been able to match moments with a sequence of changes of measure that converge to the unit constant, which implies since the divergence measure ϕ is decreasing:

$$0 \leq E \left\{ \phi \left[M^{(j)}(Y(\theta) \,|\theta) \right] \right\} \leq E \left\{ \phi \left(1 - \pi_j \right) \right\} = \phi \left(1 - \pi_j \right)$$

which converges to zero since $\phi(1) = 0$ and ϕ is continuous at $m = 1$. QED

3.2 About the choice of a ϕ-divergence

We will argue that this issue of the choice of a ϕ-divergence is dramatically different depending on whether the asset pricing model is well-specified and studied at the true unknown value θ^0 of the parameters (or at a consistent estimator of θ^0) or the asset pricing model is misspecified (or studied at a calibrated value of the parameters that is not a consistent estimator of the true one).

3.2.1 Case of a well-specified asset pricing model

We assume in this subsection that the asset pricing model is well specified and identified, such that there is a unique true unknown value θ^0 such that:

$$E \left[Y_t \left(\theta^0 \right) \right] = 0.$$

Moreover, we assume that $Y_t \left(\theta^0 \right)$ is a stationary martingale difference sequence. This allows us to apply the results of Chaudhuri and Renault [2020] and Antoine, Bonnal, and Renault [2007]. Even though their results have been proved for i.i.d. data, we can be sure that, as usual for inference based on moment conditions, procedures that are valid for i.i.d. data remain valid when the vector of moments is, at the true unknown value of the parameters, a stationary martingale difference sequence.

For the choice of a ϕ-divergence, we set the focus on the family of power divergence functions introduced by Cressie and Read [1984] that we define as follows for any given real number γ:

$$\phi_\gamma(m) = \left\{ \begin{array}{c} \frac{1}{\gamma(\gamma+1)}\left[m^{\gamma+1}-1\right], \gamma < 0 \\ \frac{1}{\gamma(\gamma+1)}\left[m^{\gamma+1}-m\right], \gamma \geq 0 \end{array} \right\}. \tag{23}$$

Note that we adopt the trick of Chen, Hansen, and Hansen [2021] to modify the classical definition of power divergence functions by replacing the term (-1) (used when $\gamma < 0$) by $(-m)$ (used when $\gamma \geq 0$). This change is immaterial when minimizing $E\left[\phi_\gamma(M)\right]$ subject to restriction $E(M) = 1$ but it helps to figure out the two limit cases $\gamma \to (-1)$ and $\gamma \to 0$, just by application of L'Hopital's rule to obtain two special cases of interest:

$$\begin{aligned} \lim_{\gamma \to (-1)} \phi_\gamma(m) &= -\log(m) = \phi_L(m) \\ \lim_{\gamma \to 0} \phi_\gamma(m) &= m\log(m) = \phi_E(m) \end{aligned}$$

where we recognize the negative log-likelihood $\phi_L(.)$and $\phi_E(.)$ is by definition the relative entropy. For $\gamma = 1$, we get:

$$\phi_1(m) = \frac{m^2 - m}{2}$$

which, as already mentioned, leads to the same minimization program as EEL:

$$\phi_Q(m) = \frac{m^2 - 1}{2}.$$

More generally, for any real number γ, $\phi_\gamma(.)$ is obviously a valid ϕ-divergence, that is a strictly convex function such that $\phi(1) = 0$.

Therefore, we can define model-implied empirical probabilities $\hat{\pi}_T^{(\gamma)}(\theta) = \left(\hat{\pi}_{t,T}^{(\gamma)}(\theta)\right)_{1 \leq t \leq T}$, $\gamma \in \mathbb{R}$, as solution of the minimization program:

$$\min_{\pi_T \in \mathbb{R}^T} \sum_{t=1}^{T} \phi_\gamma(T\pi_{t,T})$$

subject to:

$$\sum_{t=1}^{T} \pi_{t,T} = 1, \sum_{t=1}^{T} \pi_{t,T} g\left(X_{t+1}, \theta\right) = 0.$$

By doing so, we define a vast family of model-implied probabilities indexed by $\gamma \in \mathbb{R}$, with particular cases:

$$\hat{\pi}_T^{(1)}(\theta) = \hat{\pi}_T^{Q}(\theta), \hat{\pi}_T^{(-1)}(\theta) = \hat{\pi}_T^{L}(\theta).$$

Chaudhuri and Renault [2020] prove that implied probabilities, irrespective of the Cressie-Read divergence that one uses, are asymptotically equivalent at the convenient order, meaning that for all $t = 1, ..., T$:

$$\left| \hat{\pi}_{t,T}^{(\gamma)}(\theta^0) \; - \hat{\pi}_{t,T}^{Q}(\theta^0) \right| = o_P\left(\frac{1}{T\sqrt{T}}\right), \forall \gamma \in \mathbb{R}. \tag{24}$$

Note that we dub "convenient" the order of magnitude in the upper bound (24) because, even though it is not uniform over $t = 1, ..., T$, it allows Chaudhuri and Renault [2020] to prove that it makes immaterial the choice of a discrepancy measure for estimating a population expectation. For any integrable real function $\xi\left[Y(\theta)\right]$:

$$\sqrt{T}\left\{\sum_{t=1}^{T} \hat{\pi}_{t,T}^{(\gamma)}(\theta^0)\xi\left(Y_t\left(\theta^0\right)\right) - \sum_{t=1}^{T} \hat{\pi}_{t,T}^{Q}(\theta^0)\xi\left(Y_t\left(\theta^0\right)\right)\right\} = o_P(1), \forall \gamma \neq 0. \tag{25}$$

As far as inference on the population expectation is concerned, (25) implies that we have the same asymptotic normal distribution for any estimator (for all γ):

$$\sqrt{T}\left\{\sum_{t=1}^{T} \hat{\pi}_{t,T}^{(\gamma)}(\theta^0)\xi\left(Y_t\left(\theta^0\right)\right) - E\left[\xi\left(Y_t\left(\theta^0\right)\right)\right]\right\} \xrightarrow{d} \aleph\left(0, \Sigma^0\right)$$

Since inference is not the focus of interest of this paper, we let the reader to check in Antoine, Bonnal, and Renault [2007] what is the value of the asymptotic variance Σ^0 and how it must be modified when, for the purpose of feasible inference, we replace θ^0 in $\sum_{t=1}^{T} \hat{\pi}_{t,T}^{(\gamma)}(\theta^0)\xi\left(Y_t\left(\theta^0\right)\right)$ by an efficient GMM or any GEL estimator $\hat{\theta}_T$. The key point is that the information provided by the well-specified moment conditions allows to get an asymptotic variance smaller than the one of the naive estimator based on the sample mean, i.e.,

$$Var\left[\xi\left(Y_t\left(\theta^0\right)\right)\right]-\Sigma^0 \text{ is positive semi-definite.}$$

It is worth realizing that, since the moment conditions are well specified, the model-implied empirical probabilities are asymptotically matching the sample distribution (same weight $(1/T)$ for all observations $t = 1, ..., T$), but we only have for $t = 1, ...T$:

$$\left|\hat{\pi}_{t,T}^{(\gamma)}(\theta^0)\ -\frac{1}{T}\right| = o_P\left(\frac{1}{T}\right), \forall\gamma\neq 0. \tag{26}$$

Obviously the upper bound in (26) is less tight than the one in (24), precisely because model-implied empirical probabilities are not equivalent to the naive sample frequencies: they provide asymptotically more accurate estimators of a population expectation than the naive sample mean. However, we always have consistent estimators:

$$\operatorname*{plim}_{T\to\theta}\sum_{t=1}^{T}\hat{\pi}_{t,T}^{(\gamma)}(\theta^0)\xi\left(Y_t\left(\theta^0\right)\right) = \operatorname*{plim}_{T\to\theta}\frac{1}{T}\sum_{t=1}^{T}\xi\left(Y_t\left(\theta^0\right)\right) = E\left[\xi\left(Y_t\left(\theta^0\right)\right)\right].$$

Hence there is no point for the purpose of economic interpretation to compute model-implied probabilities $\hat{\pi}_{t,T}^{(\gamma)}(\theta^0)$ (or feasible counterparts $\hat{\pi}_{t,T}^{(\gamma)}(\hat{\theta}_T)$). Up to statistical accuracy of estimators of population expectations, they do not provide any economically meaningful economic information that is not carried by sample frequencies.

Finally, it is worth knowing that while positivity is not granted for some of the implied probabilities, it is always possible to shrink the probabilities to preclude their possible negativity. More precisely, following Antoine, Bonnal, and Renault [2007], we define nonnegative shrunk probabilities as follows:

$$\hat{\pi}_{t,T}^{(\gamma),sh}(\hat{\theta}_T) = \frac{1}{1+\varepsilon_{t,T}^{(\gamma)}(\hat{\theta}_T)}\hat{\pi}_{t,T}^{(\gamma)}(\hat{\theta}_T) + \frac{\varepsilon_{t,T}^{(\gamma)}(\hat{\theta}_T)}{1+\varepsilon_{t,T}^{(\gamma)}(\hat{\theta}_T)}\frac{1}{T}$$

where:

$$\varepsilon_{t,T}^{(\gamma)}(\theta) = -T\min\left[\min_{1\leq t\leq T}\hat{\pi}_{t,T}^{(\gamma)}(\theta), 0\right].$$

By virtue of (26), $\min_{1\leq t\leq T}\hat{\pi}_{t,T}^{(\gamma)}(\hat{\theta}_T)$ is asymptotically nonnegative with probability one ($\varepsilon_{t,T}^{(\gamma)}(\hat{\theta}_T)$ is asymptotically nil with probability one) so that, as proved by Antoine, Bonnal, and Renault [2007], there is no harm for the asymptotic efficiency of estimators to shrink the model-implied empirical probabilities:

$$\sqrt{T}\left\{\sum_{t=1}^{T}\hat{\pi}_{t,T}^{(\gamma),sh}(\theta^0)\xi\left(Y_t\left(\theta^0\right)\right)-\sum_{t=1}^{T}\hat{\pi}_{t,T}^{Q}(\theta^0)\xi\left(Y_t\left(\theta^0\right)\right)\right\}=o_P(1).$$

3.2.2 Case of a misspecified asset pricing model (or counterfactual analysis)

Let us now consider the case of counterfactual analysis based on a given parameter value θ that is not local to the true unknown value (that may even not exist):

$$E\left[Y_t\left(\theta\right)\right]\neq 0. \tag{27}$$

While we have argued that there is no point computing model-implied empirical probabilities in the case of a well specified model with a consistent estimator of θ^0 (since they all estimate the same object as naive sample frequencies), it is obviously different in the case of biased moments as (27). For instance following (18) and (19):

$$\operatorname*{plim}_{T\to\infty}\hat{E}_T^{\theta,Q}\left[\xi\left(Y_t\left(\theta\right)\right)\right]=E\left[\xi\left(Y_t\left(\theta\right)\right)\right]-E\left[Y_t\left(\theta\right)\right]'\left[Var\left[Y_t\left(\theta\right)\right]\right]^{-1}Cov\left[Y_t\left(\theta\right),\xi\left[Y_t\left(\theta\right)\right]\right]$$

while, if the limit exists:

$$\operatorname*{plim}_{T\to\infty}\hat{E}_T^{\theta,L}\left[\xi\left(Y_t\left(\theta\right)\right)\right]=E\left[\xi\left(Y_t\left(\theta\right)\right)\right]-E\left[Y_t\left(\theta\right)\right]'\left[Var^{\theta,L}\left[Y_t\left(\theta\right)\right]\right]^{-1}Cov^{\theta,L}\left[Y_t\left(\theta\right),\xi\left[Y_t\left(\theta\right)\right]\right].$$

We deduce from the comparison of these two formulas that it is hard to believe that there is something economically meaningful to learn from "allowing the probabilities of the states of the economy to differ from their sample frequencies" (Julliard and Ghosh [2012]).

First, there is no reason to believe that the Cressie-Read discrepancies will pick the same "pseudo-true value" of population expectations provided asymptotically by implied probabilities $\hat{\pi}_{t,T}^{(\gamma)}(\theta)$ and $\hat{\pi}_{t,T}^{(\gamma^*)}(\theta)$, $\gamma \neq \gamma^*$. The almost closed form formulas (18) and (19) clearly show the difference. Different probabilities will lead to compute modified variance and covariance which in turn imply that probabilities are different. Of course, this difference is present even asymptotically, precisely because we have (27).

Second the definite proof that implied probabilities depend (even asymptotically) on the discrepancy measure has been given by the impossibility theorem of Section 3.1. And the impossibility is precisely in the case of rare disasters whose potential impact is unbounded, thereby causing the model-implied distribution to not exist for decreasing ϕ-divergences, in particular for Cressie-Read power divergences ϕ_γ for negative powers γ and in particular for the EL case.

The bottom line is that we have no clear argument to claim that the population model-implied probability distribution associated to some particular power value γ has a better economic interpretation than others. The deep reason why there is not much economics in this discussion is that ϕ-divergences are purely statistical objects and there is no compelling argument to relate them to investor's preferences. For instance, Cressie and Read [1984] have introduced power divergences for the purpose of goodness-of-fit tests. In this respect, it makes sense to consider as target of estimation not the minimal divergence beliefs (as in Ghosh, Otsu, and Roussellet [2021]) but rather a set of "plausible beliefs and model parameters consistent with certain levels of divergence from rational expectations (i.e. misspecification sets) and perform sensitivity analysis with respect to the level of divergence" (see Chen, Hansen, and Hansen [2021]).

However, it must be acknowledged that even this coherent approach focused on set identification rests upon the somewhat arbitrary choice of a ϕ-divergence, at least in the set of those that are not "problematic". This choice issue may arguably justify the ambiguity approach in which, as in Jeong, Kim, and Park [2015], the investor's optimization ultimately delivers a distorted probability measure selected endogenously from the investor's set of priors. These distorted subjective beliefs are not the result of a statistical artifact, but produced instead by investor's concern for robustness. A similar comment applies to robust control as promoted by Hansen and Sargent [2011]. Some additional work may be still needed to relate these results to the economic literature on disaster risk. The ultimate goal would be to reconcile "uncertainty outside and inside economic models" (Hansen [2014]).

4 Sufficient Conditions for the Existence of Model-Implied Population Probabilities

We have seen in Section 3 that non-existence of a model-implied population probability distribution was caused by the conjunction of two effects:

(i) Unbounded vector $Y(\theta)$ of pricing errors;

(ii) Decreasing divergence function ϕ.

We prove in this section two results (resp. in Subsections 4.1 and 4.2) of existence based on the assumption that either (i) or (ii) does not hold, i.e., based respectively on the assumption of bounded pricing errors $Y(\theta)$ or on the assumption of increasing divergence function ϕ.

4.1 The case of bounded variables

For a given value $\theta \in \Theta$, the boundedness of the n pricing errors $Y_i(\theta)$ for $i = 1, , n$ allows us to consider $2n$ non-negative random variables a_i:

$$a_i\left(Y(\theta)\right) = L - Y_i(\theta), \quad a_{i+n}\left(Y(\theta)\right) = Y_i\left(\theta\right) - l, \; \forall \; i = 1, ..., n$$

where it is assumed that we have with probability one, for all $i = 1, .., n$:

$$l \leq Y_i\left(\theta\right) \leq L.$$

Then, for any change of measure $M(.)$, we can characterize the moment restrictions

$\{E\,[MY(\theta)] = 0\}$ by the following system of $2n$ inequalities:

$$\begin{aligned} \int a_i(y)M(y)dP_\theta(y) &\leq L, \forall i = 1, ..., n \\ \int a_i(y)M(y)dP_\theta(y) &\leq -l, \forall i = n+1, ..., 2n. \end{aligned} \tag{28}$$

In order to apply the results of Csiszar [1995], we will maintain the following assumption.

Assumption A1: The probability distribution P_θ of the random vector $Y(\theta)$ is absolutely continuous with respect to some σ-finite measure λ:

$$\frac{dP_\theta}{d\lambda}(y) = f_Y\,(y|\theta)$$

and $f_Y\,(y|\theta) > 0$ $\lambda-$almost everywhere. ■

Note that, the strict positivity of $f_Y(y\,|\theta)$ $\lambda-$almost everywhere is hardly restrictive since by definition:

$$[f_Y(y\,|\theta) = 0,\ \forall\ y \in B] \quad \Longrightarrow \quad P_\theta(B) = 0$$

and then, the dominating measure λ can always be chosen such that $\lambda(B) = 0$. In this context, Csiszar [1995] studies the linear inverse problem (28) by looking for a probability density function $s(y)$ with respect to the measure λ solution of a minimization problem:

$$\min_s \int f_Y(y\,|\theta)G\left(\frac{s(y)}{f_Y(y\,|\theta)}\right)d\lambda(y) \tag{29}$$

subject to:

$$\int a_i(y)s(y)d\lambda(y) \leq L,\ \forall\ i = 1, ..., n, \quad \text{and} \quad \int a_i(y)s(y)d\lambda(y) \leq -l,\ \forall\ i = n+1, ..., 2n$$

where G is a given differentiable strictly convex function on $\mathbb{R}_*^+$, satisfying:

$$G(1) = G'(1) = 0.$$

The objective function of (29) is well-defined precisely because $f_Y(y|\theta) > 0$ $\lambda-$almost everywhere. Note that, if we consider a differentiable divergence function ϕ then we get a well-suited function G by considering:

$$G(u) = \phi(u) - \phi'(1)[u-1]].$$

We are then able to apply Theorem 3(i) in Csiszar [1995] by noting that with our notation $M(.)$ for the change of measure, the program (29) can be rewritten as:

$$\min_M \int G[M(y)]\, dP_\theta(y)$$

subject to:

$$\int a_i(y)M(y)dP_\theta(y) \leq L,\ \forall\ i = 1, ..., n, \quad \text{and} \quad \int a_i(y)M(y)dP_\theta(y) \leq -l,\ \forall\ i = n+1, ..., 2n,$$

which can be rewritten as:

$$\min_M\ E[G[M(Y(\theta))]] \quad \text{subject to:}\ \ E[M(Y(\theta))Y(\theta)] = 0.$$

Moreover, it is worth noting that:

$$E[G[M(Y(\theta))]] = E[\phi[M(Y(\theta))]] - \phi'(1)[E[M(Y(\theta))] - 1] = E[\phi[M(Y(\theta))]]$$

since by definition:

$$E[M(Y(\theta))] = 1.$$

In other words, the minimization problem (29) is nothing but the minimization problem of interest with the change of variable $M \mapsto s$.

Regarding the minimization problem (29), Theorem 3(i), p177, in Csiszar [1995] tells us that, thanks to the non-negativity of the random variables $a_i(Y_\theta)$ for all $i = 1, ..., 2n$, and to Assumption A1, a solution s_θ to the problem (29) (the so-called D-projection problem in Csiszar's terminology) always exists. Then, from the above discussion, we do have a solution:

$$M(y\,|\theta\,) = \frac{s_\theta(y)}{f_Y(y\,|\theta\,)},\ \ \lambda - ae$$

to our problem of interest. The function s_θ is the D-projection of $f_Y(y\,|\theta\,)$ on the set of functions s defined by inequalities (28) (with $M(y)$ replaced by $s(y)/f_Y(y\,|\theta\,)$).

4.2 How to deal with unbounded pricing errors?

To figure out why the unboundedness of moment conditions may be harmful for our minimization problem and what conditions on contrast functions may provide a hedge, it is worth looking at the first order conditions of our minimization program. The Lagrangian function can be written as:

$$\mathcal{L} = \int \phi\left[M(y)\right] f_Y(y\,|\theta\,)d\lambda(y) - a\left\{\int M(y) f_Y(y\,|\theta\,)d\lambda(y) - 1\right\} - b'\left\{\int M(y) y f_Y(y\,|\theta\,)d\lambda(y)\right\}$$

where $a \in \mathbb{R}$ and $b \in \mathbb{R}^n$ are the Lagrange multipliers. Then, under very general conditions with a differentiable contrast function, the first order conditions can be written (for $\lambda - a.e.$ value of y) after differentiation with respect to $M(y)$:

$$\phi'\left[M(y)\right] f_Y(y\,|\theta\,) - a f_Y(y\,|\theta\,) - b' y f_Y(y\,|\theta) = 0. \tag{30}$$

By right-multiplying by $M(y)$ (resp. $M(y)y'$), integrating with respect to y, and using the constraints of the program, we get 1 equation (resp. n equations) to determine the Lagrange multipliers a and b respectively as:

$$\begin{aligned} a^* &= E\left[M(Y(\theta))\phi\left[M(Y(\theta))\right]\right] \\ E\left[M(Y(\theta))Y(\theta)\,Y(\theta)'\right] b^* &= E\left[M(Y(\theta))Y(\theta)\,\phi'\left[M(Y(\theta))\right]\right]. \end{aligned}$$

By plugging these values of a and b in (30), we get the optimal value of $M(y)$ for all y (up to $\lambda - a.e.$ equality) by inverting the strictly increasing function ϕ'. Thanks to Assumption

A1, we can deduce from (30) that:

$$\phi'\left[M(y)\right] = a^* + b^{*\prime} y, \quad \lambda - a.e.$$

In particular if a solution $M\left(.|\theta\right)$ exists, we will have almost surely:

$$\phi'\left[M\left(Y(\theta)|\theta\right)\right] = a^* + b^{*\prime} Y\left(\theta\right). \tag{31}$$

The identity (31) displays clearly the challenge we are facing for the existence of $M\left(.|\theta\right)$. If the random variable $Y\left(\theta\right)$ is not bounded, the linear function $\left[a^* + b^{*\prime} Y\left(\theta\right)\right]$ is not bounded either. Since the function ϕ' is strictly increasing, the divergence of $Y(\theta)$ must be coupled with a divergence of the density function $M\left(.|\theta\right)$ of the model-implied population probabilities, leading to the divergence of $\phi'\left[M\left(Y\left(\theta\right)|\theta\right)\right]$ thanks to the maintained Assumption A2.

Assumption A2:

$$\lim_{m \to \infty} \phi'(m) = +\infty. \ \blacksquare$$

In the context of the Cressie-Read family (23) of contrasts:

$$\begin{aligned} \phi'_\gamma(m) &= \frac{m^\gamma}{\gamma}, \ \forall\ \gamma \neq 0 \\ \phi'_0(m) &= \log(m) + 1, \end{aligned}$$

we note that Assumption A2 is fulfilled if and only if $\gamma \geq 0$, while by contrast for all EL-like contrasts ($\gamma < 0$):

$$\lim_{m \to \infty} \phi'_\gamma(m) = 0.$$

While Assumption A2 appears to be necessary for the existence of $M\left(.|\theta\right)$ in case of an unbounded variable $Y\left(\theta\right)$ (as confirmed by the pretty general construction in Section 3 of a counter-example for all decreasing power divergence functions), we can again use Csiszar [1995] to show to what extent it is sufficient.

If we want to relax the boundedness assumption about $Y\left(\theta\right)$, we can simply consider the system (28) of inequalities with arbitrary values of numbers l and L (that are not bounds anymore), for instance $l = L = 0$, and variables $a_i(Y), j = 1, .., 2n$, which are not assumed anymore to be non-negative. As in the former subsection, we still note the equivalence between Csiszar [1995]'s projection problem (29) and our problem of interest, through the change of variable $M \mapsto s$.

Regarding the minimization problem (29), Theorem 3(iii), p177, in Csiszar [1995] tells us that, thanks to Assumptions A1 and A2, and in spite of the fact that the functions $a_i(Y), i = 1, .., 2n$ may take both positive and negative values, a solution s_θ to the problem (29) always exists as soon as:

$$\int G^* \left[\alpha a_i^-(y)\right] f_Y(y\,|\theta) d\lambda(y) < \infty, \forall \alpha > 0,\ \forall\ i = 1, ..., 2n$$

where:

$$a_i^-(y) = \max\left(0, -a_i(y)\right)$$

and G^* denotes the convex conjugate of G:

$$G^*(v) = \sup_u \left[uv - G(u)\right].$$

As reminded in Csiszar [1995] (see formula (3.3), p177), our Assumption A2 allows us to characterize the convex conjugate of $G(u) = \phi(u) - \phi'(1)\left[u - 1\right]]$ as follows:

$$G^*(v) = \int_0^v (G')^{-1}(z)dz = \int_0^v (\phi')^{-1}\left[z + \phi'(1)\right] dz.$$

With our definition:

$$a_i(Y) = -Y_i\left(\theta\right),\ a_{i+n}(Y) = Y_i\left(\theta\right),\ \forall\ i = 1, ..., n,$$

we are then led to maintain the following assumption.

Assumption A3: For all $\alpha > 0$ and all $i = 1, ..., n$:

$$E\left\{G^*\left[\alpha Y_i^+(\theta)\right]\right\} < \infty,\ E\left\{G^*\left[\alpha Y_i^-(\theta)\right]\right\} < \infty$$

where:

$$\begin{aligned} y^+ &= \max(y, 0),\ y^- = \max(-y, 0) \\ G^*(v) &= \int_0^v (\phi')^{-1}\left[z + \phi'(1)\right] dz.\ \blacksquare \end{aligned}$$

Then, from the above discussion, under Assumptions A1, A2 and A3, we do have a solution:

$$M(y\,|\theta) = \frac{s_\theta(y)}{f_Y(y\,|\theta)},\quad \lambda - ae$$

to our problem of interest.

This result ensures very generally the existence of the model-implied population probabilities for any Cressie-Read power divergence ϕ_γ, with $\gamma \geq 0$ (as in particular EEL, $\gamma = 1$), since we can now show the following result in Lemma 1.

Lemma 1 *Let us consider a Cressie-Read contrast function ϕ_γ with $\gamma \geq 0$. Then a necessary and sufficient condition for Assumption A3 with $\phi = \phi_\gamma$ is:*

For $\gamma > 0$, $|Y_i(\theta)|^{\frac{\gamma+1}{\gamma}}$ is integrable for all $j = 1, ..., n$.

For $\gamma = 0$, $Y(\theta)$ has a finite Laplace transform $E\left[\exp\left(t'Y\left(\theta\right)\right)\right]$ for all $t \in \mathbb{R}^n$. $\blacksquare$

Not surprisingly, the smaller the index γ, the more restrictive is the integrability assumption about $Y(\theta)$ that is needed for the existence of the model-implied population probabilities. The condition for $\gamma = 0$ is tantamount to assuming the integrability at any order, which is as expected the limit case (when $\gamma \longrightarrow 0$) of the assumption needed in the case $\gamma > 0$. However, it is worth noting that the necessary and sufficient condition put forward by Lemma 1 is very natural. To see that, we first note that when using the contrast function ϕ_γ , we work with changes of measure $M \geq 0$ such that $\phi_\gamma(M)$ is integrable, meaning (with standard notations) that $M \in L^{\gamma+1}$. Thus, we want that:

$$M \in L^{\gamma+1} \implies MY_i(\theta) \in L^1, \; \forall \; i = 1, .., n$$

in order to be able to impose the constraint $E[MY(\theta)] = 0$. By virtue of the Holder-inequality, this assumption will be fulfilled if:

$$Y_i(\theta) \in L^q, \; \forall \; i = 1, .., n$$

such that:

$$\frac{1}{q} + \frac{1}{\gamma+1} = 1, \quad \text{that is} \quad q = \frac{\gamma+1}{\gamma},$$

and this is exactly the condition put forward for Lemma 1. For instance, with Euclidean Empirical Likelihood ($\gamma = 1$), we need to use changes of measure M with finite variance and the corresponding moment functions, i.e., the components of $Y(\theta)$, must have finite variance as well.

5 Conclusion

In this paper, we address the issue of econometric analysis of a structural dynamic model that is defined by a finite-dimensional set of unconditional moment restrictions. These restrictions are misspecified under rational expectations but valid under agent's subjective beliefs. Our point of view is more general than misspecification of the asset pricing model. It also may be motivated by the willingness to perform a counterfactual analysis for a value of preference parameters that we impose on the basis of prior knowledge (like limited value of risk aversion) while it does not match the rational expectation restrictions.

This empirical strategy has been pervasive in the extant literature on calibration of disaster risk for the purpose of asset pricing. Julliard and Ghosh [2012] have provided arguably the most appealing variant of this strategy by minimizing a ϕ-divergence between the historical distribution and the set of candidate distortions of subjective beliefs. However, we conclude that the theoretical underpinnings of this appealing empirical strategy are problematic for many reasons (see also Chen, Hansen, and Hansen [2021] for related arguments):

First, the minimally divergent belief may not even exist when disaster risk is unbounded, in particular in the case of an empirical likelihood approach, which is rather worrying.

Second, even for ϕ-divergences for which a minimizer exists to get a well-defined subjective beliefs distortion, the approach is problematic since different choices of ϕ-divergences will lead to different minimally divergent beliefs. Since maximization of empirical likelihood does not work in general misspecified models, there is no such thing as a natural choice of the ϕ-divergence.

Third, we stress that ϕ-divergences have been first introduced in statistics, in particular for issues of goodness-of-fit testing but have arguably no economic interpretation. We put forward the work of Chen and Epstein [2002] (and Jeong, Kim, and Park [2015] for econometric implementation) to suggest an alternative way of eliciting a subjective beliefs distortion as the endogenous result of the investor's optimization among a set of priors. Moreover, Chen and Epstein [2002] clearly explain why the model is such that ambiguity will "not disappear eventually as the agent learns about her environment". By eliciting "conditional one-step-ahead beliefs that are independent of history of times", the κ-ignorance model solved by Jeong, Kim, and Park [2015] belongs to what Chen and Epstein [2002] dub IID ambiguity.

Fourth, even though the underlying rational expectations model comes as a set of conditional moment restrictions that ensures that pricing errors are martingale difference sequences, this property is by definition violated in the case of counterfactual analysis. Therefore, an efficient statistical approach should take into account serial dependence in the sequence of pricing errors. That would at least lead to revise the model-implied probabilities computed in the current paper, by replacing the stationary marginal variance and covariances by long run quantities (with HAC estimators) taking account the likely infinite order of moving average dynamics of pricing errors.

Fifth, as far as statistical efficiency is concerned, one should prefer to perform a conditional density projection of the distribution of pricing errors on the set of distributions characterized by the conditional moment restrictions that define the asset pricing model.

While the statistical characterization of these projections has been thoroughly tackled by Komunjer and Ragusa [2016], this is arguably a statistical challenge without a compelling economic interpretation.

Overall, it seems to us that, in spite of its statistical appeal, the concept of ϕ-divergence may not be so appealing if one looks for an economically meaningful interpretation of the investors' beliefs regarding disaster risk. It may be more relevant to contemplate a direct matching of the observed data to the data simulated according to an asset pricing model including some concern for disaster risk. This would take an extension of Indirect Inference to misspecified models as sketched in Dridi, Guay, and Renault [2007] and developed in the case of disaster risk in the recent work of Sonksen and Grammig [2021]. By putting forward a novel strategy to estimate and empirically assess pricing models that allow for multi-period disaster risk, the latter paper is able to increase the likelihood of concern for disaster risk because "the total contraction can pan out over subsequent quarters", avoiding to impose counterfactual severity of the disaster events.

References

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

R. J. Barro. Rare Disasters and Asset Markets in the Twentieth Century. *Quarterly Journal of Economics*, 121: 823–866, 2006.

S. Cerreia-Vioglio, L. P. Hansen, F. Maccheroni, and M. Marinacci. Making Decisions under Model Misspecification. Working paper, University of Chicago, 2021.

S. Chaudhuri and E. Renault. Score Tests in GMM: Why Use Implied Probabilities. *Journal of Econometrics*, 219: 260–280, 2020.

X. Chen, L. P. Hansen, and P. G. Hansen. Robust Inference for Moment Condition Models without Rational Expectations. Working paper, University of Chicago, 2021.

Z. Chen and L. Epstein. Ambiguity, Risk, and Asset Returns in Continuous Time. *Econometrica*, 70: 1403–1444, 2002.

S. A. Corcoran. Bartlett Adjustment of Empirical Discrepancy Statistics. *Biometrika*, 85: 967–972, 1998.

N. Cressie and T. Read. Multinomial Goodness-of-Fit Tests. *Journal of the Royal Statistical Society. Series B (Methological)*, 46: 440–464, 1984.

I. Csiszar. Generalized projections for non-negative functions. *Acta Mathematica Hungarica*, 68: 161–186, 1995.

R. Dridi, A. Guay, and E. Renault. Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models. *Journal of Econometrics*, 136: 397–430, 2007.

L. Epstein and S. Zin. Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. *Econometrica*, 57: 937–969, 1989.

A. Ghosh, T. Otsu, and G. Roussellet. Subjective Beliefs Estimators and Their Properties. Working paper, McGill University, 2021.

L. P. Hansen. Nobel Lecture: Uncertainty Outside and Inside Economic Models. *Journal of Political Economy*, 122: 954–987, 2014.

L. P. Hansen and R. Jagannathan. Assessing Specification Errors in Stochastic Discount Factors Models. *Journal of Finance*, 52: 557–590, 1997.

L. P. Hansen and T. J. Sargent. Robustness and ambiguity in continuous time. *Journal of Economic Theory*, 146: 1195–1223, 2011.

G. W. Imbens, R. H. Spady, and P. Johnson. Information Theoretic Approaches to Inference in Moment Condition Models. *Econometrica*, 66:333–357, 1998.

J. Jeong, H. Kim, and J. Park. Does Ambiguity Matter? Estimating Asset Pricing Models with a Multiple-Priors Recursive Utility. *Journal of Financial Economics*, 115: 361–382, 2015.

H. Joe. *Multivariate Models and Dependence Concepts*, volume 73. Chapman and Hall/CRC, 1997.

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

Y. Kitamura and M. Stutzer. An Information-Theoretic Alternative to Generalized Method of Moments Estimation. *Econometrica*, 65:861–874, 1997.

I. Komunjer and G. Ragusa. Existence and Characterization of Conditional Density Projections. *Econometric Theory*, 32: 947–987, 2016.

R. Mehra and E. C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15: 145–161, 1985.

A. B. Owen. *Empirical Likelihood*. Chapman and Hall, 2001.

J. Park. Martingale Regressions for Conditional Mean Models in Continuous Time. Working paper, Indiana University, 2021.

T. A. Rietz. The Equity Risk Premium: A Solution. *Journal of Monetary Economics*, XXII: 117–131, 1988.

S. M. Schennach. Point estimation with exponentially tilted empirical likelihood. *The Annals of Statistics*, 35:634–672, 2007.

J. Sonksen and J. Grammig. Empirical asset pricing with multi-period disaster risk: A simulation-based approach. *Journal of Econometrics*, 222: 805–832, 2021.

J. Tsai and J. Wachter. Disaster Risk and Its Implications for Asset Pricing. *Annual Review of Financial Economics*, 7: 219–252, 2015.