Divergences and duality for estimation and test under moment condition model

Michel Broniatowski, Amor Keziou

ON GENERALIZED EMPIRICAL LIKELIHOOD METHODS

MICHEL BRONIATOWSKI* AND AMOR KEZIOU**

Abstract. We introduce estimation and test procedures through divergence minimization for models satisfying linear constraints with unknown parameter. These procedures extend the empirical likelihood (EL) method and share common features with generalized empirical likelihood (GEL) approach. We treat the problems of existence and characterization of the divergence projections of probability measures on sets of signed finite measures. Our approach allows to obtain the limit distributions of the estimates and test statistics (including the EL ones) under alternatives and misspecification. The asymptotic behavior of the estimates and test statistics are studied both under the model and under alternatives including misspecification, using the dual representation of the divergences and the explicit forms of the divergence projections. An approximation to the power function is deduced as well as the sample size which ensures a desired power for a given alternative.

Keywords: Empirical likelihood; Generalized Empirical likelihood; Minimum divergence; Efficiency; Power function; Duality; Divergence projection.

MSC (2000) Classification: 62G05; 62G10; 62G15; 62G20; 62G35.

Contents

1. Introduction and notation

Statistical models are often defined through estimating equations

$$\mathbb{E}\left[g(X,\theta)\right]=0$$

Date: April 2010.

where $g(X,\theta)$ is some vector valued function of a random vector $X \in \mathbb{R}^m$ and a parameter vector $\theta \in \Theta \subset \mathbb{R}^d$. The function g has l real valued functions g_j as its components. Examples of such models are numerous, see e.g. Qin and Lawless (1994), Haberman (1984), Sheehy (1987), McCullagh and Nelder (1983), Owen (2001) and the references therein. Denoting M^1 the collection of all probability measures (p.m.) on $\mathbb{R}^m$, the submodel $\mathcal{M}_\theta^1$, associated to a given value θ of the parameter, consists of all distributions Q satisfying the linear constraints induced by $g(.,\theta)$, namely

$$\mathcal{M}_\theta^1 := \left\{ Q \in M^1 \ \text{ such that } \ \int g(x,\theta)\, dQ(x) = 0 \right\}.$$

The statistical model which we consider can be written as

$$\mathcal{M}^1 := \bigcup_{\theta\in\Theta} \mathcal{M}_\theta^1. \tag{1.1}$$

Let $X_1, ..., X_n$ denote an i.i.d sample of X with unknown distribution P_0. We denote θ_0, if it exists, the value of the parameter such that P_0 belongs to $\mathcal{M}_{\theta_0}^1$, namely the value satisfying $\mathbb{E}\left[g(X,\theta_0)\right] = 0$, and we assume obviously that θ_0 is unique. This paper addresses the two following natural questions:

Problem 1: Does P_0 belong to the model $\mathcal{M}^1$?

Problem 2: When P_0 is in the model, which is the value θ_0 of the parameter for which $\mathbb{E}\left[g(X,\theta_0)\right] = 0$? Also can we perform tests about θ_0? Can we construct confidence areas for θ_0?

We note that these problems have been investigated by many authors. Hansen (1982) considered generalized method of moments (GMM). Hansen *et al.* (1996) introduced the continuous updating (CU) estimate. The empirical likelihood (EL) approach, developed by Owen (1988) and Owen (1990), has been investigated in the context of model (1.1) by Qin and Lawless (1994) and Imbens (1997) introducing the EL estimator. The recent literature in econometrics focusses on such models; Newey and Smith (2004) provided a class of estimates called generalized empirical likelihood (GEL) estimates which contains the EL and CU estimates. Schennach (2007) discussed the asymptotic properties of the empirical likelihood estimate under misspecification; She showed the important fact that the EL estimate may cease to be root n consistent when the functions defining the moments conditions are unbounded. Among other results pertaining to EL, Newey and Smith (2004) stated that EL estimate enjoys optimality properties in term of efficiency when bias corrected among all GEL estimates including the GMM one. Also Corcoran (1998) and Baggerly (1998) proved that in a class of minimum discrepancy statistics (called power divergence statistics), EL ratio is the only one that is Bartlett correctable. Confidence areas for the parameter θ_0 have been considered in the seminal paper by Owen (1990). Problem 1 and 2 have been handled via EL approach in Qin and Lawless (1994) and in Newey and Smith (2004) under the null hypothesis $\mathcal{H}_0 : P_0 \in \mathcal{M}^1$; however the limit distributions of the EL estimate and the EL test statistic under misspecification have not been obtained so far. Our contribution is as follows:

(1) The approach which we develop is based on *minimum discrepancy estimates*, which extends the EL method and has common features with minimum distance and GEL techniques, using merely divergences. We present a wide class of estimates, test statistics and confidence regions for the parameter θ_0 as well as various test statistics for *Problems* 1 and 2, all depending on the choice of the divergence.

(2) The limit distribution of the EL test statistic under the alternative and under misspecification remains up to date an open problem. The present paper fills this gap; indeed, we give the limit distributions of the proposed estimates and test statistics (including the EL ones) for *Problems* 1 and 2 both under the null hypotheses, under alternatives and under misspecification.

(3) The limit distributions of the test statistics under the alternatives and misspecification are used to give an approximation to the power function and the sample size which ensures a desired power for a given alternative.
(4) We extend confidence region (C.R.) estimation techniques based on EL (see Owen (1990)), providing a wide range of such C.R.'s, each one depending upon a specific criterion.

From the point of view of the statistical criterion under consideration, the main advantage of using a divergence based approach lays in the fact that it leads to all statistical properties of the estimates and test statistics under the alternative, including misspecification, which cannot be achieved through the classical EL context. In the case of parametric models of densities, White (1982) studied the asymptotic properties of the parametric maximum likelihood estimate and the parametric likelihood ratio statistic under misspecification. Broniatowski and Keziou (2009) stated the consistency and obtained the limit distributions of the minimum divergence estimates and the corresponding test statistics (including the parametric likelihood ones) both under the null hypotheses and the alternatives, from which they deduced an approximation to the power function. In this paper, we extend the above results to the case of the semi-parametric models (1.1) in the global context of empirical divergences; including the EL method.

The paper is organized as follows. Section 2 describes the statistical divergences used in the sequel. Section 3 is devoted to the description of estimation and test procedures. In Section 3 we adapt the formalism of Lagrangian duality to the context of statistical divergence, and we use it to give practical formulas (for the study and the numerical computation) of the proposed estimates and test statistics. Section 5 deals with the asymptotic properties of the estimates and test statistics. Simulations results are given in Section 6. All proofs are postponed to the Appendix.

2. Statistical divergences

We first set some general definitions and notations. Let P be some p.m. Denote by M the space of all signed finite measures (s.f.m.) on $\mathbb{R}^m$. Let ϕ be a convex function from $\mathbb{R}$ onto $[0,+\infty]$ with $\phi(1)=0$, and such that its domain $\mathrm{dom}\phi := \{x \in \mathbb{R} \text{ such that } \phi(x) < \infty\}$ is an interval with endpoints $a < 1 < b$ (which may be finite or infinite). We assume that ϕ is closed[1]. For any s.f.m. Q, the ϕ-divergence between Q and the p.m. P, when Q is absolutely continuous with respect to (a.c.w.r.t) P, is defined through

$$D_\phi(Q,P) := \int_{\mathbb{R}^m} \phi\left(\frac{dQ}{dP}(x)\right)\, dP(x). \tag{2.1}$$

in which $\frac{dQ}{dP}(\cdot)$ denotes the Radon-Nikodym derivative. When Q is not a.c.w.r.t. P, we set $D_\phi(Q,P) = +\infty$. For any p.m. P, the mapping $Q \in M \mapsto D_\phi(Q,P)$ is convex and takes nonnegative values. When $Q = P$ then $D_\phi(Q,P) = 0$. Furthermore, if the function $x \mapsto \phi(x)$ is strictly convex on a neighborhood of $x = 1$, then

$$D_\phi(Q,P) = 0 \text{ if and only if } Q = P. \tag{2.2}$$

All the above properties are presented in Csiszár (1963), Csiszár (1967) and Liese and Vajda (1987) chapter 1, for $\phi-$divergences defined on the set of all p.m.'s M^1. When the ϕ-divergences are defined on M, then the same arguments as developed on M^1 hold. When defined on M^1, the Kullback-Leibler (KL), modified Kullback-Leibler (KL_m), χ^2, modified χ^2 (χ^2_m), Hellinger (H), and L^1 divergences are respectively associated to the convex functions $\phi(x) = x\log x - x + 1$, $\phi(x) = -\log x + x - 1$, $\phi(x) = \frac{1}{2}(x-1)^2$, $\phi(x) = \frac{1}{2}(x-1)^2/x$, $\phi(x) = 2(\sqrt{x}-1)^2$ and $\phi(x) = |x-1|$. All these divergences except the L^1 one, belong to the class of power divergences introduced in

[1]The closedness of ϕ means that if a or b are finite then $\varphi(x) \to \varphi(a)$ when $x \downarrow a$, and $\varphi(x) \to \varphi(b)$ when $x \uparrow b$.

Cressie and Read (1984) (see also Liese and Vajda (1987) and Pardo (2006)). They are defined through the class of convex functions

$$x \in \mathbb{R}_+^* \mapsto \phi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)} \tag{2.3}$$

if $\gamma \in \mathbb{R}\setminus\{0,1\}$ and by $\phi_0(x) := -\log x + x - 1$ and $\phi_1(x) := x\log x - x + 1$. So, the $KL-$divergence is associated to ϕ_1, the KL_m to ϕ_0, the χ^2 to ϕ_2, the χ^2_m to ϕ_{-1} and the Hellinger distance to $\phi_{1/2}$. We extend the definition of the power divergences functions $Q \in M^1 \mapsto D_{\phi_\gamma}(Q,P)$ onto the whole set of signed finite measures M as follows. When the function $x \mapsto \phi_\gamma(x)$ is not defined on $(-\infty, 0[$ or when ϕ_γ is defined on $\mathbb{R}$ but is not a convex function we extend the definition of ϕ_γ through

$$x \in \mathbb{R} \mapsto \phi_\gamma(x)\mathbb{1}_{[0,+\infty]}(x) + (+\infty)\mathbb{1}_{[-\infty,0[}(x). \tag{2.4}$$

Note for instance that for χ^2-divergence, the corresponding ϕ function $\phi(x) = \frac{1}{2}(x-1)^2$ is convex and defined on whole $\mathbb{R}$. In this paper, for technical considerations, we assume that the ϕ functions are strictly convex on their domain (a,b), twice continuously differentiable on the interior of their domain and satisfy $\phi(1) = 0$, $\phi'(1) = 0$ and $\phi''(1) = 1$. We assume also that ϕ is "essentially smooth" in the sense that $\lim_{x\downarrow a}\phi'(x) = -\infty$ if $a > -\infty$ and $\lim_{x\uparrow b}\phi'(x) = +\infty$ if $b < +\infty$. Note that all the power functions ϕ_γ, see (2.4), satisfy the above conditions, including all standard divergences.

Definition 2.1. *Let Ω be some subset in M. The $\phi-$divergence between the set Ω and a p.m. P is defined by*

$$D_\phi(\Omega, P) := \inf_{Q\in\Omega} D_\phi(Q,P).$$

A finite measure $Q^* \in \Omega$, such that $D_\phi(Q^*,P) < \infty$ and

$$D_\phi(Q^*,P) \le D_\phi(Q,P) \quad \textit{for all} \quad Q \in \Omega,$$

is called a projection of P on Ω. This projection may not exist, or may be not defined uniquely.

3. Minimum divergence estimates

Let $X_1, ..., X_n$ denote an i.i.d. sample of a random vector $X \in \mathbb{R}^m$ with distribution P_0. Let P_n be the empirical measure pertaining to this sample, namely

$$P_n(\cdot) := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}(\cdot)$$

in which δ_x denotes the Dirac measure at point x. We will endow our statistical approach in the global context of s.f.m's with total mass 1 satisfying linear constraints:

$$\mathcal{M}_\theta := \left\{ Q \in M \ \text{ such that } \ \int_{\mathbb{R}^m} dQ(x) = 1 \text{ and } \int_{\mathbb{R}^m} g(x,\theta)\, dQ(x) = 0 \right\} \tag{3.1}$$

and

$$\mathcal{M} := \bigcup_{\theta\in\Theta} \mathcal{M}_\theta, \tag{3.2}$$

sets of signed finite measures that replace $\mathcal{M}^1_\theta$ and $\mathcal{M}^1$. Enhancing the model (1.1) to the above one (3.2) bears a number of improvements upon existing results; this is argued at the end of the present Section. The "plug-in" estimate of $D_\phi(\mathcal{M}_\theta, P_0)$ is

$$\widehat{D}_\phi(\mathcal{M}_\theta, P_0) := \inf_{Q\in\mathcal{M}_\theta} D_\phi(Q,P_n) = \inf_{Q\in\mathcal{M}_\theta} \int_{\mathbb{R}^m} \phi\left(\frac{dQ}{dP_n}(x)\right)\, dP_n(x). \tag{3.3}$$

If the projection Q_n of P_n on $\mathcal{M}_\theta$ exists, then it is clear that Q_n is a s.f.m. (or possibly a p.m.) a.c.w.r.t. P_n; this means that the support of Q_n must be included in the set $\{X_1, \ldots, X_n\}$. So, define the sets

$$\mathcal{M}_\theta^{(n)} := \left\{ Q \in M \mid Q \text{ a.c.w.r.t. } P_n,\ \sum_{i=1}^n Q(X_i) = 1 \text{ and } \sum_{i=1}^n Q(X_i) g(X_i, \theta) = 0 \right\}, \tag{3.4}$$

which may be seen as subsets of $\mathbb{R}^n$. Then, the plug-in estimate (3.3) can be written as

$$\widehat{D}_\phi(\mathcal{M}_\theta, P_0) = \inf_{Q \in \mathcal{M}_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n \phi\left(nQ(X_i)\right). \tag{3.5}$$

In the same way, $D_\phi(\mathcal{M}, P_0) := \inf_{\theta\in\Theta} \inf_{Q\in\mathcal{M}_\theta} D_\phi(Q, P_0)$ can be estimated by

$$\widehat{D}_\phi(\mathcal{M}, P_0) = \inf_{\theta\in\Theta} \inf_{Q \in \mathcal{M}_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n \phi\left(nQ(X_i)\right). \tag{3.6}$$

By uniqueness of $\arg\inf_{\theta\in\Theta} D_\phi(\mathcal{M}_\theta, P_0)$ and since the infimum is reached at $\theta = \theta_0$ under the model, we estimate θ_0 through

$$\widehat{\theta}_\phi = \arg\inf_{\theta\in\Theta} \inf_{Q \in \mathcal{M}_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n \phi\left(nQ(X_i)\right). \tag{3.7}$$

Enhancing $\mathcal{M}^1$ to $\mathcal{M}$ and accordingly extensions in the definitions of the ϕ functions on $]-\infty, +\infty[$ and of the ϕ-divergences on the whole space of s.f.m's M, is motivated by the following arguments:

- If the domain (a, b) of the function ϕ is included in $[0, +\infty[$ then minimizing over $\mathcal{M}^1$ or over $\mathcal{M}$ leads to the same estimates and test statistics. Hence, both approaches coincide for instance in the case of the divergences KL_m, KL, modified χ^2 and Hellinger.
- Let θ be a given value in Θ. Denote Q_n^1 and Q_n respectively the projection of P_n on $\mathcal{M}_\theta^1$ and on $\mathcal{M}_\theta$. If Q_n^1 satisfies $0 < Q_n(X_i) < 1$ for all $i = 1, \ldots, n$ then it coincides with Q_n, i.e., $Q_n^1 = Q_n$. Therefore, in this case, both approaches leads also to the same estimates and test statistics.
- It may occur that for some θ in Θ and some $i = 1, \ldots, n$, $Q_n^1(X_i)$ is a boundary value of $[0, 1]$, hence the first order conditions are not met which makes a real difficulty for the calculation of the estimates over the sets of p.m. $\mathcal{M}_\theta^1$ and $\mathcal{M}^1$. However, when $\mathcal{M}^1$ is replaced by $\mathcal{M}$, then this problem does not hold any longer in particular when $\mathrm{dom}\phi = \mathbb{R}$, which is the case for instance of the χ^2-divergence. Other arguments are given in remark 4.5 below.

The empirical likelihood paradigm (see Owen (1988), Owen (1990), Qin and Lawless (1994) and Owen (2001)), enters as a special case of the statistical issues related to estimation and tests based on $\phi-$divergences with $\phi(x) = \phi_0(x) = -\log x + x - 1$, namely on KL_m-divergence. Indeed, it is straightforward to see that the empirical log-likelihood ratio statistic for testing $P_0 \in \mathcal{M}$ against $P_0 \notin \mathcal{M}$, in the context of ϕ-divergences, can be written as $2n\widehat{D}_{KL_m}(\mathcal{M}, P_0)$; and that the EL estimate of θ_0 can be written as $\widehat{\theta}_{KL_m} = \arg\inf_{\theta\in\Theta} \widehat{D}_{KL_m}(\mathcal{M}_\theta, P_0)$; see Remark 4.3 below. In the case of the power functions $\phi = \phi_\gamma$, the corresponding estimates (3.7) belong to the class of GEL estimates introduced by Newey and Smith (2004), and (3.5) are the empirical Cressie-Read statistics introduced by Baggerly (1998) and Corcoran (1998).

The constrained optimization problems (3.5), (3.6) and (3.7) can be transformed into unconstrained ones making use of some arguments of "duality" which we briefly state hereunder from Rockafellar (1970). On the other hand, the obtention of asymptotic statistical results of the estimates and the test statistics, under misspecification or under alternative hypotheses, requires to handle existence

conditions and characterization of the projection of P_0 on the submodel $\mathcal{M}_\theta$ or on the entire model $\mathcal{M}$. This also will be considered through duality, along the following Section.

4. Dual representation of $\phi-$divergences under constraints

This Section is central for our purposes. Indeed, it provides the explicit form of the proposed estimates by transforming the constrained problems (3.5) to unconstrained ones, using Lagrangian duality which is a classical tool in optimization theory. This Section adapts this formalism to the context of divergences and the present statistical setting. The Lagrangian "dual" problems, corresponding to the "primal" ones

$$\inf_{Q\in\mathcal{M}_\theta} D_\phi(Q,P_0) \tag{4.1}$$

and its empirical counterpart (3.5), make use of the Fenchel-Legendre transform of ϕ, defined through

$$\psi : t\in\mathbb{R}\mapsto \psi(t) := \sup_{x\in\mathbb{R}}\left\{tx-\phi(x)\right\}. \tag{4.2}$$

The "dual" problems associated to (4.1) and (3.5) are respectively

$$\sup_{t\in\mathbb{R}^{1+l}}\left\{t_0-\int_{\mathbb{R}^m}\psi(t_0+\sum_{j=1}^{l}t_j g_j(x,\theta))\;dP_0(x)\right\}, \tag{4.3}$$

and

$$\sup_{t\in\mathbb{R}^{1+l}}\left\{t_0-\frac{1}{n}\sum_{i=1}^{n}\psi(t_0+\sum_{j=1}^{l}t_j g_j(X_i,\theta))\right\}. \tag{4.4}$$

In the following propositions, we state sufficient conditions under which the primal problems (4.1) and (3.5) coincide respectively with the dual ones (4.3) and (4.4). First, recall some properties of the convex conjugate ψ of ϕ. For the proofs we can refer to Rockafellar (1970) Section 26. The function ψ is convex and closed, its domain is an interval with endpoints

$$a^*=\lim_{x\to-\infty}\frac{\phi(x)}{x},\quad b^*=\lim_{x\to+\infty}\frac{\phi(x)}{x} \tag{4.5}$$

satisfying $a^*<0<b^*$ and $\psi(0)=0$. The strict convexity of ϕ on its domain (a,b) is equivalent to the condition that its conjugate ψ is essentially smooth, i.e., differentiable with

$$\begin{array}{rcll}\lim_{t\downarrow a^*}\psi'(t) & = & -\infty & \text{if}\quad a^*>-\infty,\\ \lim_{t\uparrow b^*}\psi'(t) & = & +\infty & \text{if}\quad b^*<+\infty.\end{array} \tag{4.6}$$

Conversely, ϕ is essentially smooth on its domain (a,b) if and only if ψ is strictly convex on its domain (a^*,b^*). In all the sequel, we assume additionally that ϕ is essentially smooth. Hence, ψ is strictly convex on its domain (a^*,b^*), and it holds that

$$a^*=\lim_{x\downarrow a}\phi'(x),\qquad b^*=\lim_{x\uparrow b}\phi'(x),$$

and

$$\psi(t)=t{\phi'}^{-1}(t)-\phi\left({\phi'}^{-1}(t)\right),\quad \text{for all } t\in]a^*,b^*[. \tag{4.7}$$

It holds also that ψ is twice continuously differentiable on $]a^*,b^*[$,

$$\psi'(t)={\phi'}^{-1}(t)\quad\text{and}\quad \psi''(t)=\frac{1}{\phi''\left({\phi'}^{-1}(t)\right)}. \tag{4.8}$$

In particular, $\psi'(0)=1$ and $\psi''(0)=1$. Obviously, since ϕ is assumed to be closed, we have

$$\phi(a)=\lim_{x\downarrow a}\phi(x) \quad \text{and} \quad \phi(b)=\lim_{x\uparrow b}\phi(x),$$

which may be finite or infinite. Hence, by closedness of ψ, we have

$$\psi(a^*)=\lim_{t\downarrow a^*}\psi(x) \quad \text{and} \quad \psi(b^*)=\lim_{t\uparrow b^*}\psi(t).$$

Finally, the first and second derivatives of ϕ in a and b are defined to be the limits of $\phi'(x)$ and $\phi''(x)$ when $x\downarrow a$ and when $x\uparrow b$. The first and second derivatives of ψ in a^* and b^* are defined in a similar way. In Table 1, we give the convex conjugates ψ of some functions ϕ associated to standard divergences. We determine also their domains, (a,b) and (a^*,b^*).

TABLE 1. Convex conjugates for some standard divergences.

D_ϕ	ϕ	domϕ	domψ	ψ
D_{KL_m}	$\phi(x):=-\log x+x-1$	$]0,+\infty[$	$]-\infty,1[$	$\psi(t)=-\log(1-t)$
D_{KL}	$\phi(x):=x\log x-x+1$	$[0,+\infty[$	$\mathbb{R}$	$\psi(t)=e^t-1$
$D_{\chi^2_m}$	$\phi(x):=\frac{1}{2}\frac{(x-1)^2}{x}$	$]0,+\infty[$	$]-\infty,\frac{1}{2}]$	$\psi(t)=1-\sqrt{1-2t}$
D_{χ^2}	$\phi(x):=\frac{1}{2}(x-1)^2$	$\mathbb{R}$	$\mathbb{R}$	$\psi(t)=\frac{1}{2}t^2+t$
D_H	$\phi(x):=2(\sqrt{x}-1)^2$	$[0,+\infty[$	$]-\infty,2[$	$\psi(t)=\frac{2t}{2-t}$
D_{ϕ_γ}	$\phi(x):=\frac{x^\gamma-\gamma x+\gamma-1}{\gamma(\gamma-1)}$	$--$	$--$	$\psi(t)=\frac{1}{\gamma}(\gamma t-t+1)^{\frac{\gamma}{\gamma-1}}-\frac{1}{\gamma}$

Proposition 4.1. *Let θ be a given value in Θ. If there exists Q_0 in $\mathcal{M}_\theta^{(n)}$ such that*

$$a<Q_0(X_i)<b, \quad \text{for all } i=1,\dots,n. \tag{4.9}$$

Then

$$\inf_{Q\in\mathcal{M}_\theta^{(n)}} D_\phi(Q,P_n)=\sup_{t\in\mathbb{R}^{1+l}}\left\{t_0-\frac{1}{n}\sum_{i=1}^n\psi(t_0+\sum_{j=1}^l t_j g_j(X_i,\theta))\right\} \tag{4.10}$$

with dual attainment. Conversely, if there exists a dual optimal solution $\widehat{t}$ such that

$$a^*<\widehat{t}_0+\sum_{j=1}^l\widehat{t}_j g_j(X_i,\theta)<b^*, \text{ for all } i=1,\dots,n, \tag{4.11}$$

then the equality (4.10) holds, and the unique optimal solution of the primal problem $\inf_{Q\in\mathcal{M}_\theta^{(n)}} D_\phi(Q,P_n)$, namely the projection of P_n on $\mathcal{M}_\theta^{(n)}$, is given by

$$Q_n(X_i)=\frac{1}{n}\phi'^{-1}(\widehat{t}_0+\sum_{j=1}^l\widehat{t}_j g_j(X_i,\theta)), \quad i=1,...,n,$$

where $\widehat{t}$ is solution of the equations

$$\begin{cases} 1-\frac{1}{n}\sum_{i=1}^n\phi'^{-1}(\widehat{t}_0+\sum_{j=1}^l\widehat{t}_j g_j(X_i,\theta))=0 \\ -\frac{1}{n}\sum_{i=1}^n g_j(X_i,\theta)\phi'^{-1}(\widehat{t}_0+\sum_{j=1}^l\widehat{t}_j g_j(X_i,\theta))=0, \quad j=1,...,l. \end{cases}$$

Remark 4.1. For the χ^2-divergence, we have $a=-\infty$ and $b=+\infty$. Hence, condition (4.9) holds whenever $\mathcal{M}_\theta^{(n)}$ is not void. More generally, the above Proposition holds for any ϕ-divergence with ϕ function satisfying dom$\phi=\mathbb{R}$.

Remark 4.2. Assume that $g(x,\theta) := x - \theta$. So, for any divergence D_ϕ with $\mathrm{dom}\phi =]0, +\infty[$, which is the case of the modified χ^2 divergence and the modified Kullback-Leibler divergence (or equivalently EL method), condition (4.9) means that θ is an interior point of the convex hull of the data $(X_1, ..., X_n)$. This is precisely what is checked in Owen (1990), p. 100, for the EL method; see also Owen (2001).

For the asymptotic counterpart of the above results we have; see Theorem 1 in Broniatowski and Keziou (2006):

Proposition 4.2. *Let θ be a given value in Θ. Assume that $\int |g_j(x,\theta)|\ dP_0(x) < \infty$ for all $j = 1, \ldots, l$. If there exists Q_0 in $\mathcal{M}_\theta$ with $D_\phi(Q_0, P_0) < \infty$ and*[2]

$$a < \inf_x \frac{dQ_0}{dP_0}(x) \leq \sup_x \frac{dQ_0}{dP_0}(x) < b, \quad P_0 - a.s. \tag{4.12}$$

Then

$$\inf_{Q\in\mathcal{M}_\theta} D_\phi(Q, P_0) = \sup_{t\in\mathbb{R}^{1+l}} \left\{ t_0 - \int_{\mathbb{R}^m} \psi(t_0 + \sum_{j=1}^{l} t_j g_j(x,\theta))\ dP_0(x) \right\} \tag{4.13}$$

with dual attainment. Conversely, if there exists a dual optimal solution t^* which is an interior point of the set

$$\left\{ t \in \mathbb{R}^{1+l} \text{ such that } \int_{\mathbb{R}^m} |\psi(t_0 + \sum_{j=1}^{l} t_j g_j(x,\theta))|\ dP_0(x) < \infty \right\}, \tag{4.14}$$

then the dual equality (4.13) holds, and the unique optimal solution Q_θ^* of the primal problem $\inf_{Q\in\mathcal{M}_\theta} D_\phi(Q, P_0)$, namely the projection of P_0 on $\mathcal{M}_\theta$, is given by

$$\frac{dQ_\theta^*}{dP_0}(x) = {\phi'}^{-1}(t_0^* + \sum_{j=1}^{l} t_j^* g_j(x,\theta)),$$

where t^* is solution of

$$\begin{cases} 1 - \int {\phi'}^{-1}(t_0^* + \sum_{j=1}^{l} t_j^* g_j(x,\theta))\ dP_0(x) = 0 \\ -\int g_j(x,\theta) {\phi'}^{-1}(t_0^* + \sum_{j=1}^{l} t_j^* g_j(x,\theta))\ dP_0(x) = 0, \quad j = 1, \ldots, l. \end{cases} \tag{4.15}$$

Furthermore, t^* is unique if the functions $\mathbb{1}_{\mathbb{R}^m}, g_1(.,\theta), \ldots, g_l(.,\theta)$ are linearly independent in the sense that $P_0\left\{ x \mid t_0 + \sum_{j=1}^{l} t_j g_j(x,\theta) \neq 0 \right\} > 0$ for all $t \in \mathbb{R}^m$ with $t \neq 0$.

For sake of brevity and clearness, we must introduce some additional notations. Denote by $\overline{g}$ the vector valued function $(\mathbb{1}_{\mathbb{R}^m}, g_1, \ldots, g_l)^T$. For any p.m. P and any measurable function f on $\mathbb{R}^m$, Pf denotes the integral $\int_{\mathbb{R}^m} f(x)\ dP(x)$. Let

$$m(x,\theta,t) := t_0 - \psi(t^T \overline{g}(x,\theta)), \quad \text{for all } x \in \mathbb{R}^m, \theta \in \Theta \subset \mathbb{R}^d, t \in \mathbb{R}^{1+l}. \tag{4.16}$$

Note that the sup in (4.10) and (4.13) can be restricted respectively to the sets

$$\Lambda_n(\theta) := \left\{ t \in \mathbb{R}^{1+l} \mid a^* < t^T \overline{g}(X_i,\theta) < b^*, \text{ for all } i = 1, \ldots, n \right\} \tag{4.17}$$

[2]The strict inequalities in (4.12) mean that $P_0\left\{ x \in \mathbb{R}^m \mid \frac{dQ_0}{dP_0}(x) \leq a \right\} = P_0\left\{ x \mid \frac{dQ_0}{dP_0}(x) \geq b \right\} = 0$.

and

$$\Lambda(\theta) := \left\{ t \in \mathbb{R}^{1+l} \mid \int_{\mathbb{R}^m} |\psi(t_0 + \sum_{j=1}^{l} t_j g_j(x,\theta))| \, dP_0(x) < \infty \right\}. \tag{4.18}$$

In view of the above propositions, we redefine the estimates (3.5), (3.6) and (3.7) as follows

$$\widehat{D}_\phi\left(\mathcal{M}_\theta, P_0\right) := \sup_{t \in \Lambda_n(\theta)} \frac{1}{n} \sum_{i=1}^{n} m(X_i, \theta, t) := \sup_{t \in \Lambda_n(\theta)} P_n m(\theta, t), \tag{4.19}$$

$$\widehat{D}_\phi\left(\mathcal{M}, P_0\right) := \inf_{\theta \in \Theta} \sup_{t \in \Lambda_n(\theta)} \frac{1}{n} \sum_{i=1}^{n} m(X_i, \theta, t) := \inf_{\theta \in \Theta} \sup_{t \in \Lambda_n(\theta)} P_n m(\theta, t) \tag{4.20}$$

and

$$\widehat{\theta}_\phi := \arg \inf_{\theta \in \Theta} \sup_{t \in \Lambda_n(\theta)} \frac{1}{n} \sum_{i=1}^{n} m(X_i, \theta, t) := \arg \inf_{\theta \in \Theta} \sup_{t \in \Lambda_n(\theta)} P_n m(\theta, t). \tag{4.21}$$

Remark 4.3. When $\phi(x) = -\log x + x - 1$, then the estimate (3.7) clearly coincides with the EL one, so it can be seen as the value of the parameter which minimizes the KL_m-divergence between the model $\mathcal{M}$ and the empirical measure P_n of the data. The statistics $2n\widehat{D}_{KL_m}(\mathcal{M}, P_0)$, see (3.6), coincides with the empirical likelihood ratio associated to the null hypothesis $\mathcal{H}_0 : P_0 \in \mathcal{M}$ against the alternative $\mathcal{H}_1 : P_0 \notin \mathcal{M}$. The dual representation of $\widehat{D}_{KL_m}(\mathcal{M}, P_0)$, see (4.20), is

$$\widehat{D}_{KL_m}(\mathcal{M}, P_0) = \inf_{\theta \in \Theta} \sup_{t \in \Lambda_n(\theta)} \left\{ t_0 + \frac{1}{n} \sum_{i=1}^{n} \log(1 - t_0 - \sum_{j=1}^{l} t_j g_j(X_i, \theta)) \right\}.$$

For a given $\theta \in \Theta$, the KL_m-projection Q_n, of P_n on $\mathcal{M}_\theta$, is given by (see proposition 4.1)

$$\frac{1}{Q_n(X_i)} = n \left(1 - t_0^* - \sum_{j=1}^{l} t_j^* g(X_i, \theta) \right), \quad i = 1, \ldots, n,$$

which, multiplying by $Q_n(X_i)$ and summing upon i yields $t_0^* = 0$. Therefore, t_0 can be omitted, and the above representation can be rewritten as follows

$$\widehat{D}_{KL_m}(\mathcal{M}, P_0) = \inf_{\theta \in \Theta} \sup_{t_1, \ldots, t_l} \left\{ \frac{1}{n} \sum_{i=1}^{n} \log(1 + \sum_{j=1}^{l} t_j g_j(X_i, \theta)) \right\}$$

and then

$$\widehat{\theta}_{KL_m} = \widehat{\theta}_{EL} = \arg \inf_{\theta \in \Theta} \sup_{t_1, \ldots, t_l} \left\{ \frac{1}{n} \sum_{i=1}^{n} \log(1 + \sum_{j=1}^{l} t_j g_j(X_i, \theta)) \right\}$$

in which the sup is taken over the set

$$\left\{ (t_1, \ldots, t_l) \in \mathbb{R}^m \mid -1 < \sum_{j=1}^{l} t_j g_j(X_i, \theta) < +\infty, \text{ for all } i = 1, \ldots, n \right\}.$$

This is the ordinary dual representation of the EL estimate; see Qin and Lawless (1994) and Owen (2001).

Remark 4.4. Consider the power divergences, associated to the power functions ϕ_γ; see (2.3) and (2.4). We will show that the estimates $\widehat{\theta}_{\phi_\gamma}$ belong to the class of GEL estimators introduced by Newey and Smith (2004). The projection Q_n of P_n on $\mathcal{M}_\theta$ is given by

$$Q_n(X_i) = \left((\gamma-1)(t_0^* + \sum_{j=1}^{l} t_j^* g(X_i,\theta)) + 1 \right)^{1/(\gamma-1)}, \quad i = 1,\dots,n.$$

Using the constraint $\sum_{i=1}^n Q_n(X_i) = 1$, we can explicit t_0^* in terms of $t_1^*,\dots,t_l^*$, and hence the sup in the dual representation (4.21) can be reduced to a subset of $\mathbb{R}^l$, as in Newey and Smith (2004). When $\phi(x) = \frac{1}{2}(x-1)^2$, then $\widehat{\theta}_\phi$ coincides with the continuous updating estimator of Hansen *et al.* (1996).

Remark 4.5. (**Numerical calculation of the estimates and the specific role of the χ^2-divergence**). The computation of $\widehat{t}(\theta)$ for fixed $\theta \in \Theta$ as defined in (4.15) is difficult when handling a generic divergence. In the case of χ^2-divergence, i.e., when $\phi(x) = \frac{1}{2}(x-1)^2$, optimizing on all s.f.m's, the system (4.15) is linear; we thus easily obtain an explicit form for $\widehat{t}(\theta)$, which in turn allows for a single gradient descent when optimizing upon Θ. This procedure is useful in order to calculate the estimates for all other divergences (for which the corresponding system is non linear) including EL, since it provides an easy starting point for the resulting double gradient descent.

5. Asymptotic properties of the estimates of the parameter and the estimates of the divergences

5.1. **Under the model.** This Section addresses Problems 1 and 2, aiming at testing the null hypothesis $\mathcal{H}_0 : P_0 \in \mathcal{M}$ against the alternative $\mathcal{H}_1 : P_0 \notin \mathcal{M}$. We expose the limit distributions of the proposed test statistics which are the estimated divergences between the model $\mathcal{M}$ and P_0. We also derive the limit distributions of the estimates of θ_0. The following two Theorems extend Theorem 3.1 and 3.2 in Newey and Smith (2004) to the context of divergence based approach. The assumptions which we consider match those of Theorems 3.1 and 3.2 in Newey and Smith (2004).

Assumption 1. (a) $P_0 \in \mathcal{M}$ and $\theta_0 \in \Theta$ is the unique solution to $\mathbb{E}[g(X,\theta)] = 0$; (b) $\Theta \subset \mathbb{R}^d$ is compact; (c) $g(X,\theta)$ is continuous at each $\theta \in \Theta$ with probability one; (d) $\mathbb{E}[\sup_{\theta\in\Theta} \|g(X,\theta)\|^\alpha] < \infty$ for some $\alpha > 2$; (e) the matrix $\Omega := \mathbb{E}\left[g(X,\theta_0)g(X,\theta_0)^T\right]$ is nonsingular.

Theorem 5.1. *Under assumption 1, the estimate $\widehat{\theta}_\phi$ exists and converges to θ_0 in probability, $\frac{1}{n}\sum_{i=1}^n g(X_i,\widehat{\theta}_\phi) = O_P(1/\sqrt{n})$, $\widehat{t}(\widehat{\theta}_\phi) := \arg\sup_{t\in\Lambda_n(\widehat{\theta}_\phi)} P_n m(\widehat{\theta}_\phi, t)$ exists and belongs to int$(\Lambda_n(\widehat{\theta}_\phi))$ with probability approaching one as $n \to \infty$, and $\widehat{t}(\widehat{\theta}_\phi) = O_P(1/\sqrt{n})$.*

In order to obtain asymptotic normality, we need some additional assumptions. Denote by G the matrix $G := \mathbb{E}[\partial g(X,\theta_0)/\partial\theta]$.

Assumption 2. (a) $\theta_0 \in \text{int}(\Theta)$; (b) With probability one $g(X,\theta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of θ_0 and $\mathbb{E}[\sup_{\theta\in\mathcal{N}} \|\partial g(X,\theta)/\partial\theta\|] < \infty$; (c) $\text{rank}(G) = d$.

Theorem 5.2. *Assume that assumptions 1 and 2 hold. Then,*

(1) $\sqrt{n}\left(\widehat{\theta}_\phi - \theta_0\right)$ *converges in distribution to a centered normal vector with covariance matrix*

$$V := \left[G\Omega^{-1}G^T\right]^{-1}.$$

(2) *If $l > d$, the statistic $2n\widehat{D}_\phi(\mathcal{M}, P_0)$ converges in distribution to a χ^2 random variable with $(l-d)$ degrees of freedom.*

Remark 5.1. The above Theorem allows to perform statistical tests (of the model) with asymptotic level α. Consider the null hypothesis

$$\mathcal{H}_0 : P_0 \in \mathcal{M} \quad \text{against the alternative} \quad \mathcal{H}_1 : P_0 \notin \mathcal{M}. \tag{5.1}$$

The critical region is then

$$C_\phi := \left\{ 2n\widehat{D}_\phi(\mathcal{M}, P_0) > q_{(1-\alpha)} \right\}$$

where $q_{(1-\alpha)}$ is the $(1-\alpha)$-quantile of the $\chi^2(l-d)$ distribution. When $\phi(x) = -\log x + x - 1$, the corresponding test is the empirical likelihood ratio one; see Qin and Lawless (1994).

5.2. **Asymptotic properties of the estimates of the divergences for a given value of the parameter.** For a given $\theta \in \Theta$, consider the test problems of the null hypothesis $\mathcal{H}_0 : P_0 \in \mathcal{M}_\theta$ against two different families of alternative hypotheses: $\mathcal{H}_1 : P_0 \notin \mathcal{M}_\theta$ and $\mathcal{H}_1' : P_0 \in \mathcal{M} \setminus \mathcal{M}_\theta$. Those two tests address different situations since $\mathcal{H}_1$ may include misspecification of the model. We present two different test statistics each pertaining to one of the situations and derive their limit distributions both under $\mathcal{H}_0$ and under the alternatives. As a by product we also derive confidence areas for the true value θ_0 of the parameter. We will state the convergence in probability of $\widehat{D}_\phi(\mathcal{M}_\theta, P_0)$ to $D_\phi(\mathcal{M}_\theta, P_0)$, and we will obtain the limit law of $\widehat{D}_\phi(\mathcal{M}_\theta, P_0)$ both when $P_0 \in \mathcal{M}_\theta$ and when $P_0 \notin \mathcal{M}_\theta$. Obviously, when $P_0 \in \mathcal{M}_\theta$, this means that $\theta = \theta_0$ since the true-value θ_0 of the parameter is assumed to be unique.

Assumption 3. (a) $P_0 \in \mathcal{M}_\theta$ and θ is the unique solution to $\mathbb{E}\left[g(X,\theta)\right] = 0$; (b) $\mathbb{E}\left[\|g(X,\theta)\|^\alpha\right] < \infty$ for some $\alpha > 2$; (c) the matrix
$\Omega := \mathbb{E}\left[g(X,\theta)g(X,\theta)^T\right]$ is nonsingular.

Theorem 5.3. *Under assumption 3, we have*

1. $\widehat{t}(\theta) := \arg\sup_{t\in\Lambda(\theta)} P_n m(\theta, t)$ *exists and belongs to int*$(\Lambda(\theta))$ *with probability approaching one as* $n \to \infty$, *and* $\widehat{t}(\theta) = O_P(1/\sqrt{n})$.
2. *The statistic* $2n\widehat{D}_\phi(\mathcal{M}_\theta, P_0)$ *converges in distribution to a* $\chi^2(l)$ *random variable.*

In order to obtain the limit distribution of the test statistic $2n\widehat{D}_\phi\left(\mathcal{M}_\theta, P_0\right)$ under the alternative $\mathcal{H}_1 : P_0 \notin \mathcal{M}_\theta$, including misspecification, the following assumption is needed.

Assumption 4. (a) $P_0 \notin \mathcal{M}_\theta$, and $t^*(\theta) := \arg\sup_{t\in\Lambda(\theta)} \mathbb{E}\left[m(X,\theta,t)\right]$ exists and is an interior point of $\Lambda(\theta)$; (b) $\mathbb{E}\left[\sup_{t\in N} |m(X,\theta,t)|\right] < \infty$ for some compact set $N \subset \Lambda(\theta)$ such that $t^*(\theta) \in \mathrm{int}(N)$; (c) the functions $\mathbb{1}_{\mathbb{R}^m}, g_1, \ldots, g_l$ are linearly independent in the following sense:
$P_0\left\{x \mid t_0 + \sum_{j=1}^l t_j g_j(x,\theta) \neq 0\right\} > 0$ for all $t \in \mathbb{R}^{1+l}$ with $t \neq 0$.

Assumption (c) hereabove ensures the strict concavity of the function $t \in \Lambda(\theta) \mapsto \mathbb{E}\left[m\left(X,\theta,t\right)\right]$; otherwise $t^*(\theta)$ may not be defined uniquely implying possible inconsistency of $\widehat{t}(\theta)$.

Theorem 5.4. *Under assumption 4, when* $P_0 \notin \mathcal{M}_\theta$, *we have*

1. $\widehat{t}(\theta)$ *converges in probability to* $t^*(\theta)$.
2. $\widehat{D}_\phi(\mathcal{M}_\theta, P_0)$ *converges in probability to* $D_\phi(\mathcal{M}_\theta, P_0)$.

We now give the limit distribution of the test statistics under $\mathcal{H}_1$. We need the following additional condition.

Assumption 5. (a) with probability one, the function $t \mapsto m(X,\theta,t)$ is $\mathcal{C}^3$ in a neighborhood $\mathcal{N}(t^*(\theta))$ of $t^*(\theta)$, and all third order partial derivatives (w.r.t. t) of $\{t \mapsto m(X,\theta,t);\ t \in \mathcal{N}\}$ are dominated by some P_0-integrable function;
(b) $\mathbb{E}\left[m(X,\theta,t^*(\theta))^2\right] < \infty$, $\mathbb{E}\left[\|\partial m(X,\theta,t^*(\theta))/\partial t\|^2\right] < \infty$, and the matrix
$\mathbb{E}\left[\partial^2 m(X,\theta,t^*(\theta))/\partial t^2\right]$ exists and nonsingular.

Theorem 5.5. *Under assumptions 4 and 5, we have*

(1) $\sqrt{n}(\widehat{t}(\theta) - t^*(\theta))$ *converges in distribution to a centered normal vector with covariance matrix*

$$\left[\mathbb{E}\left[m''(X,\theta,t^*)\right]\right]^{-1}\mathbb{E}\left[m'(X,\theta,t^*)m'(X,\theta,t^*)^T\right]\left[\mathbb{E}\left[m''(X,\theta,t^*)\right]\right]^{-1}.$$

(2) $\sqrt{n}\left(\widehat{D}_\phi(\mathcal{M}_\theta,P_0) - D_\phi(\mathcal{M}_\theta,P_0)\right)$ *converges in distribution to a centered normal random variable with variance*

$$\sigma^2(\theta) = \mathbb{E}\left[m(X,\theta,t^*(\theta))^2\right] - \left[\mathbb{E}\left[m(X,\theta,t^*(\theta))\right]\right]^2.$$

Remark 5.2. Let θ be a given value in Θ. Consider the test problem of the null hypothesis

$$\mathcal{H}_0 : P_0 \in \mathcal{M}_\theta \quad \text{against} \quad P_0 \notin \mathcal{M}_\theta. \tag{5.2}$$

In view of Theorem 5.3 part 2, we reject $\mathcal{H}_0$ against $\mathcal{H}_1$ at asymptotic level α when $2n\widehat{D}_\phi\left(\mathcal{M}_\theta,P_0\right)$ exceeds the $(1-\alpha)$- quantile of the $\chi^2(l)$ distribution. Theorem 5.5 part 2 is useful to give an approximation to the power function

$$P_0 \notin \mathcal{M}_\theta \mapsto \beta(P_0) := P_0\left[2n\widehat{D}_\phi\left(\mathcal{M}_\theta,P_0\right) > q_{(1-\alpha)}\right].$$

We obtain then the following approximation

$$\beta(P_0) \approx 1 - F_{\mathcal{N}}\left(\frac{\sqrt{n}}{\sigma(\theta)}\left[\frac{q_{1-\alpha}}{2n} - D_\phi(\mathcal{M}_\theta,P_0)\right]\right), \tag{5.3}$$

where $F_{\mathcal{N}}$ is the cumulative distribution of the standard normal distribution. From this approximation, we can give the approximate sample size that ensures a desired power β for a given alternative $P_0 \notin \mathcal{M}_\theta$. Let n_0 be the positive root of the equation

$$\beta = 1 - F_{\mathcal{N}}\left[\frac{\sqrt{n}}{\sigma\left(\theta\right)}\left(\frac{q_{(1-\alpha)}}{2n} - D_\phi\left(\mathcal{M}_\theta,P_0\right)\right)\right]$$

i.e.,

$$n_0 = \frac{(a+b) - \sqrt{a\left(a+2b\right)}}{2D_\phi\left(\mathcal{M}_\theta,P_0\right)^2}$$

with $a := \sigma(\theta^*)^2\left[F_{\mathcal{N}}^{-1}\left(1-\beta\right)\right]^2$ and $b := q_{(1-\alpha)}D_\phi\left(\mathcal{M}_\theta,P_0\right)$. The required sample size is then $\lfloor n_0\rfloor + 1$ where $\lfloor n_0\rfloor$ is the integer part of n_0.

Remark 5.3. (**Generalized empirical likelihood ratio test**). For testing $\mathcal{H}_0 : P_0 \in \mathcal{M}_\theta$ against the alternative $\mathcal{H}_1' : \mathcal{M} \setminus \mathcal{M}_\theta$, we propose to use the statistics

$$2nS_n^\phi := 2n\left[\widehat{D}_\phi\left(\mathcal{M}_\theta,P_0\right) - \inf_{\theta\in\Theta}\widehat{D}_\phi\left(\mathcal{M}_\theta,P_0\right)\right] \tag{5.4}$$

which converge in distribution to a $\chi^2(d)$ random variable under $\mathcal{H}_0$ when assumptions 1 and 2 hold. This can be proved using similar arguments as in Theorems 5.2 and 5.3. We then reject $\mathcal{H}_0$ at asymptotic level α when $2nS_n^\phi > q_{(1-\alpha)}$, the $(1-\alpha)$-quantile of the $\chi^2(d)$-distribution. Under $\mathcal{H}_1'$ and when assumptions 1,2,4 and 5 hold, as in Theorem 5.5, it can be proved that

$$\sqrt{n}\left(S_n^\phi - D_\phi\left(\mathcal{M}_\theta,P_0\right)\right) \tag{5.5}$$

converges to a centered normal random variable with variance

$$\sigma^2(\theta) := \mathbb{E}\left(m(X,\theta,t^*(\theta))^2\right) - \left(\mathbb{E}m(X,\theta,t^*(\theta))\right)^2.$$

So, as in the above remark, we obtain the following approximation

$$\beta(P_0) \approx 1 - F_{\mathcal{N}}\left(\frac{\sqrt{n}}{\sigma(\theta)}\left[\frac{q_{1-\alpha}}{2n} - D_\phi(\mathcal{M}_\theta,P_0)\right]\right) \tag{5.6}$$

to the power function $P_0 \in \mathcal{M}/\mathcal{M}_\theta \mapsto P_0\left[2nS_n^\phi > q_{(1-\alpha)}\right]$. The approximated sample size required to achieve a desired power for a given alternative can be obtained as in the above Remark.

Remark 5.4. (**Confidence region for the parameter**). For a fixed level α, using convergence (5.4), the set

$$\left\{\theta \in \Theta \text{ such that } 2nS_n^\phi \leq q_{(1-\alpha)}\right\}$$

is an asymptotic confidence region for θ_0 where $q_{(1-\alpha)}$ is the $(1-\alpha)$-quantile of the $\chi^2(d)$-distribution.

5.3. **Under misspecification.** We address Problem 1 stating the limit distribution of the proposed test statistics under the alternative $\mathcal{H}_1 : P_0 \notin \mathcal{M}$. This needs the introduction of $Q_{\theta^*}^*$, the projection of P_0 on $\mathcal{M}$. Assumption 6 below ensures the existence of the "pseudo-true" value θ^* as well as the existence of the projection $Q_{\theta^*}^*$ of P_0 on $\mathcal{M}$, and states some necessary other regularity conditions.

Assumption 6. (a) Θ is compact, $\theta^* := \arg\inf_{\theta\in\Theta}\sup_{t\in\Lambda(\theta)}\mathbb{E}\left[m(X,\theta,t)\right]$ exists and is unique; (b) $g(X,\theta)$ is continuous at each $\theta \in \Theta$ with probability one; (c) $\mathbb{E}\left[\sup_{\theta\in\Theta,t\in N(\theta)}|m(X,\theta,t)|\right] < \infty$ where $N(\theta) \subset \Lambda(\theta)$ is a compact set such that $t^*(\theta) \in \operatorname{int}(N(\theta))$; (d) the functions $\mathbb{1}_{\mathbb{R}^m}, g_1, \ldots, g_l$ are linearly independent in the following sense: $P_0\left\{x \mid t_0 + \sum_{j=1}^l t_j g_j(x,\theta) \neq 0\right\} > 0$ for all $t \in \mathbb{R}^{1+l}$ with $t \neq 0$.

Theorem 5.6. *Under assumption 6, we have*

(1) $\|\widehat{t}(\theta) - t^*(\theta)\|$ *converges in probability to* 0 *uniformly in* $\theta \in \Theta$.
(2) $\widehat{\theta}_\phi$ *converges in probability to* θ^*;
(3) $\widehat{D}_\phi(\mathcal{M}, P_0)$ *converges in probability to* $D_\phi(\mathcal{M}, P_0)$.

The asymptotic normality of the test statistics under misspecification requires the following additional conditions.

Assumption 7. (a) $\theta^* \in \operatorname{int}(\Theta)$; (b) with probability one, the function $(\theta,t) \mapsto m(X,\theta,t)$ is $\mathcal{C}^3$ in a neighborhood $\mathcal{N} \subset \Theta \times \Lambda(\Theta)$ of $(\theta^*, t^*(\theta^*))$, and all the third order partial derivative functions are dominated on $\mathcal{N}$ by some P_0-integrable function; (c) $\mathbb{E}\left[m(X,\theta^*,t^*(\theta^*))^2\right]$, $\mathbb{E}\left[\|\partial m(X,\theta^*,t^*(\theta^*))/\partial t\|^2\right]$ and $\mathbb{E}\left[\|\partial m(X,\theta^*,t^*(\theta^*)/\partial\theta\|^2\right]$ are finite, and the matrix

$$S := \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix},$$

exists and is nonsingular, where $S_{11} := \mathbb{E}\left[\partial^2 m(X,\theta^*,t^*(\theta^*))/\partial t^2\right]$, $S_{12} = {S_{21}}^T := \mathbb{E}\left[\partial^2 m(X,\theta^*,t^*(\theta^*))/\partial t\partial\theta\right]$ and $S_{22} := \mathbb{E}\left[\partial^2 m(X,\theta^*,t^*(\theta^*))/\partial\theta^2\right]$.

Theorem 5.7. *Under assumptions 6 and 7, we have*

(1)

$$\sqrt{n}\begin{pmatrix} \widehat{t}(\widehat{\theta}_\phi) - t^*(\theta^*) \\ \widehat{\theta}_\phi - \theta^* \end{pmatrix}$$

converges in distribution to a centered normal vector with covariance matrix

$$W = S^{-1}MS^{-1}$$

where

$$M := \mathbb{E}\left[\begin{bmatrix} \frac{\partial}{\partial t}m\left(X,\theta^*,t^*(\theta^*)\right) \\ \frac{\partial}{\partial\theta}m\left(X,\theta^*,t^*(\theta^*)\right) \end{bmatrix}\begin{bmatrix} \frac{\partial}{\partial t}m\left(X,\theta^*,t^*(\theta^*)\right) \\ \frac{\partial}{\partial\theta}m\left(X,\theta^*,t^*(\theta^*)\right) \end{bmatrix}^T\right];$$

(2) $\sqrt{n}\left(\widehat{D}_\phi(\mathcal{M},P_0)-D_\phi(\mathcal{M},P_0)\right)$ *converges in distribution to a centered normal variable with variance*

$$\sigma^2(\theta^*)=\mathbb{E}\left[m(X,\theta^*,t^*(\theta^*))^2\right]-\left[\mathbb{E}\left[m(X,\theta^*,t^*(\theta^*))\right]\right]^2.$$

Remark 5.5. In the case of EL, i.e., when $\phi(x)=-\log x+x-1$, assumption (6-c) implies that (see 4.12)

$$-\infty<\inf_x t_0+t^Tg(x,\theta)\leq\sup_x t_0+t^Tg(x,\theta)<1$$

P_0-a.s for all $\theta\in N(\theta^*)$ and $t\in N(\theta)$. This imposes a restriction on the model when the support of P_0 is unbounded. Indeed, when the support of P_0 is for example the whole space $\mathbb{R}^m$ condition above does not hold when g is unbounded. At the contrary the same condition may hold for other divergences associated to ϕ functions with $\mathrm{dom}\phi=\mathbb{R}$.

Remark 5.6. Theorem 5.7 is useful for the computation of the power function. For testing the null hypothesis $P_0\in\mathcal{M}$ against the alternative $\mathcal{H}_1:P_0\notin\mathcal{M}$, the power function is

$$P_0\notin\mathcal{M}\mapsto\beta(P_0):=P_0\left[2n\widehat{D}_\phi\left(\mathcal{M},P_0\right)>q_{(1-\alpha)}\right]. \tag{5.7}$$

Using Theorem 5.7 part 2, we obtain the following approximation to the power function (5.7):

$$\beta(P_0)\approx 1-F_{\mathcal{N}}\left[\frac{\sqrt{n}}{\sigma\left(\theta^*\right)}\left(\frac{q_{(1-\alpha)}}{2n}-D_\phi\left(\mathcal{M},P_0\right)\right)\right] \tag{5.8}$$

where $F_{\mathcal{N}}$ is the empirical cumulative distribution of the standard normal distribution. From the proxy value of $\beta(P_0)$ hereabove, the approximate sample size that ensures a given power β for a given alternative $P_0\notin\mathcal{M}$ can be obtained as follows. Let n_0 be the positive root of the equation

$$\beta=1-F_{\mathcal{N}}\left[\frac{\sqrt{n}}{\sigma(\theta^*)}\left(\frac{q_{(1-\alpha)}}{2n}-D_\phi\left(\mathcal{M},P_0\right)\right)\right]$$

i.e.

$$n_0=\frac{(a+b)-\sqrt{a\left(a+2b\right)}}{2D_\phi\left(\mathcal{M},P_0\right)^2}$$

with $a:=\sigma(\theta^*)^2\left[F_{\mathcal{N}}^{-1}\left(1-\beta\right)\right]^2$ and $b:=q_{(1-\alpha)}D_\phi\left(\mathcal{M},P_0\right)$. The required sample size is then $\lfloor n_0\rfloor+1$ where $\lfloor n_0\rfloor$ is the integer part of n_0.

6. Simulation results: Approximation of the power function of the empirical likelihood ratio test

We will illustrate by simulation the accuracy of the power approximation (5.8) in the case of EL method, i.e., when $\phi(x)=-\log x+x-1$. Consider the test problem of the composite null hypothesis

$$\mathcal{H}_0:P_0\in\mathcal{M}\quad\text{against the alternative}\quad\mathcal{H}_1:P_0\notin\mathcal{M}$$

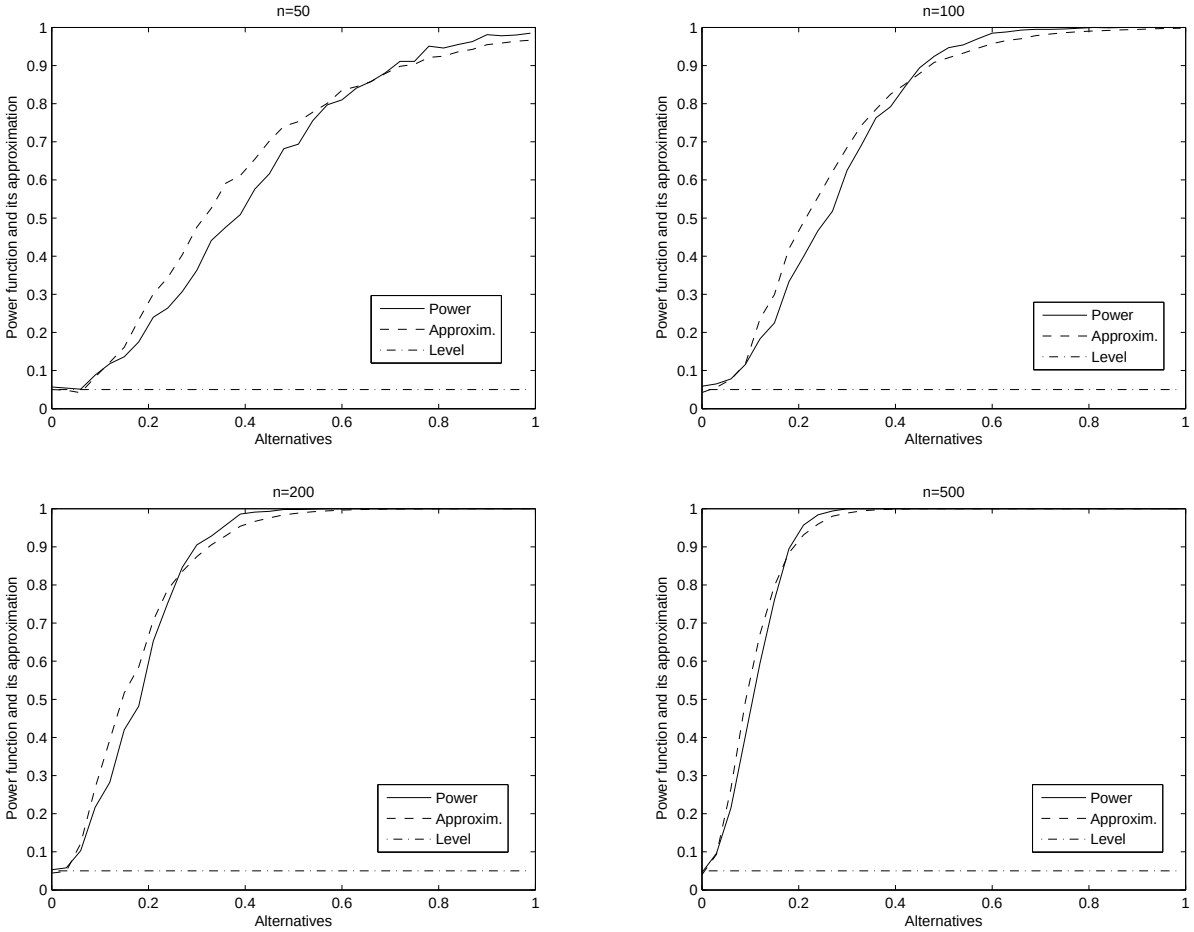
where $\mathcal{M}=\bigcup_{\theta\in\mathbb{R}}\mathcal{M}_\theta$ and $\mathcal{M}_\theta$ is the set of all s.f.m's satisfying the constraints $\int dQ(x)=1$ and $\int g(x,\theta)\ dQ(x)=0$ with $g(x,\theta):=(x,x^2-\theta)$, namely

$$\mathcal{M}_\theta:=\left\{Q\ \text{ such that }\ \int_{\mathbb{R}}dQ(x)=1\text{ and }\int_{\mathbb{R}}g(x,\theta)\ dQ(x)=0\right\},$$

where $\theta\in\mathbb{R}$ is the parameter of interest. We consider the asymptotic level $\alpha=0.05$ and the alternatives $P_0:=\mathcal{U}([-1,1+\epsilon])\notin\mathcal{M}$ for different values of ϵ in the interval $]0,1]$. Note that when $\epsilon=0$ then the uniform distribution $\mathcal{U}([-1,1])$ belongs to the model $\mathcal{M}$. For this model, we can show also that all assumptions of Theorem 5.2 are satisfied when $\epsilon=0$, and all assumptions of Theorem 5.7 are met under alternatives. In figure 1, the power function (5.7) is plotted (with a continuous line), with sample sizes $n=50, n=100$, $n=200$ and $n=500$, for different values of ϵ. Each

power entry was obtained by Monte-Carlo from 1000 independent runs. The approximation (5.8) is plotted (with a dashed line) as a function of ϵ. The estimates $\widehat{\theta}_\phi$ and $\widehat{D}_\phi(\mathcal{M}, P_0)$ are calculated using the Newton algorithm. We observe from figure 1 that the approximation is accurate even for moderate sample sizes.

FIGURE 1. Approximation of the power function



7. Concluding remarks and possible developments

We have proposed new estimates and tests for model satisfying linear constraints with unknown parameter through divergence based methods which generalize the EL approach. This leads to the obtention of the limit distributions of the test statistics and the estimates under alternatives and under misspecification, which can not be obtained through the likelihood point of view. Consistency of the test statistics under the alternatives is the starting point for the study of the optimality of the tests through Bahadur approach; also the generalized Neyman-Pearson optimality of EL test (as developed by Kitamura (2001)) can be adapted for empirical divergence based methods. Many problems remain to be studied in the future such as the choice of the divergence which leads to an optimal (in some sense) estimator or test in terms of efficiency and/or robustness. Preliminary

simulation results show that Hellinger divergence enjoys good properties in terms of efficiency-robustness; see Broniatowski and Keziou (2008). Also comparisons under local alternatives should be developed.

8. Appendix

Proof of Theorem 5.1 The same arguments, used for the proof of Theorem 3.1 in Newey and Smith (2004), hold when their criterion function $(\theta,\lambda)\in\Theta\times\mathbb{R}^l\mapsto\frac{1}{n}\sum_{i=1}^n\rho(\lambda^T g(X,\theta))$ is replaced by our function $(\theta,t)\in\Theta\times\mathbb{R}^{1+l}\mapsto\frac{1}{n}\sum_{i=1}^n m(t^T\overline{g}(X,\theta))$. In particular, we have $\max_{i\leq n}\left|\widehat{t}(\widehat{\theta}_\phi)^T\overline{g}(X_i,\widehat{\theta}_\phi)\right|$ tends to 0 in probability, which implies that $\widehat{t}(\widehat{\theta}_\phi)\in\mathrm{int}(\Lambda_n(\widehat{\theta}_\phi))$ with probability one as $n\to\infty$, since $a^*<0<b^*$.

Proof of Theorem 5.2. The proof is similar to that of Newey and Smith (2004) Theorem 3.2. Hence, it is omitted.

Proof of Theorem 5.3. (1) It is a particular case of Theorem 5.1 taking $\Theta=\{\theta\}$. (2) The first order conditions $P_n\partial m(\theta,\widehat{t})/\partial t=0$ are satisfied with probability one as $n\to\infty$. Hence by a Taylor expansion we obtain

$$\begin{aligned} 0 &= P_n\partial m\left(\theta,\widehat{t}\right)/\partial t \\ &= P_n\partial m\left(\theta,0\right)/\partial t+\frac{1}{2}\left[P_n\partial^2 m\left(\theta,\overline{t}\right)/\partial t^2\right]^T\widehat{t}, \end{aligned} \tag{8.1}$$

where $\overline{t}\in\mathbb{R}^{1+l}$ is a vector inside the segment that links 0 and $\widehat{t}$. By the uniform weak law of large numbers (UWLLN), and dominated convergence Theorem, we have $P_n\partial^2 m\left(\theta,\overline{t}\right)/\partial t^2$ tends in probability to

$$\mathbb{E}\left[\partial^2 m(X,\theta,0)/\partial t^2\right]=-\left[\begin{array}{cc}1 & 0^T\\ 0 & \Omega\end{array}\right]=:-M,$$

which is nonsingular and symmetric. Hence, we can write

$$\sqrt{n}\widehat{t}=M^{-1}\sqrt{n}P_n\partial m(X,\theta,0)/\partial t+o_P(1). \tag{8.2}$$

Using similar arguments, we get also

$$\widehat{D}_\phi(\mathcal{M}_\theta,P_0)=P_n m(\theta,\widehat{t})=\left[P_n\partial m(\theta,0)/\partial t\right]^T\widehat{t}-\frac{1}{2}\widehat{t}^T M\widehat{t}+o_P(1/n).$$

From this, using (8.2), we obtain

$$\widehat{D}_\phi(\mathcal{M}_\theta,P_0)=\frac{1}{2}\left[P_n\partial m(\theta,0)/\partial t\right]^T M^{-1}[P_n\partial m(\theta,0)/\partial t]+o_P(1).$$

This yields to

$$2n\widehat{D}_\phi(\mathcal{M}_\theta,P_0)=\left[P_n\partial m(\theta,0)/\partial t\right]^T M^{-1}[P_n\partial m(\theta,0)/\partial t]+o_P(1). \tag{8.3}$$

In the other hand, direct calculation shows that

$$\mathbb{E}\left[\partial m(X,\theta,0)\partial m(X,\theta,0)^T\right]=M.$$

Combining this with (8.3), we conclude the proof.

Proof of Theorem 5.4. (1) First, note that condition (b) implies that $t^*(\theta)$ is unique since $t\in\Lambda(\theta)\mapsto\mathbb{E}\left[m(X,\theta,t)\right]$ is strictly concave by (c) and $\Lambda(\theta)$ is a convex set. By UWLLN, using continuity of $m(X,\theta,t)$ in t and condition (b), we obtain

$$\left|P_n m(\theta,t)-\mathbb{E}\left[m(X,\theta,t)\right]\right|\to 0, \tag{8.4}$$

in probability uniformly in t over the compact set N. Using this and the fact that $t^*(\theta) := \arg\sup_{t\in\Lambda(\theta)} P_0 m(\theta,t)$ is unique and belongs to $\text{int}(N)$ and the strict concavity of $t \mapsto P_0 m(\theta,t)$, we conclude that any value

$$\overline{t} := \arg\sup_{t\in N} P_n m(\theta,t) \tag{8.5}$$

converges in probability to $t^*(\theta)$; see e.g. Theorem 5.7 in van der Vaart (1998). We end the proof by showing that $\widehat{t}(\theta)$ belongs to $\text{int}(N)$ with probability one as $n \to \infty$, and therefore it converges to $t^*(\theta)$. In fact, since for n sufficiently large any value $\overline{t}$ lies in the interior of N, concavity of $t \mapsto P_n m(\theta,t)$ implies that no other point t in the complement of $\text{int}(N)$ can maximize $P_n m(\theta,t)$ over $t \in \mathbb{R}^{1+l}$, hence $\widehat{t}(\theta)$ must be in $\text{int}(N)$.
(2) We have $\widehat{D}_\phi(\mathcal{M}_\theta, P_0) = P_n m(\theta,\widehat{t}) = P_n m(\theta,\overline{t})$ where the second equality holds for n sufficiently large. Hence we can write

$$\begin{aligned}\left|\widehat{D}_\phi(\mathcal{M}_\theta,P_0) - D_\phi(\mathcal{M}_\theta,P_0)\right| &= \left|P_n m(\theta,\overline{t}) - P_0 m(\theta,t^*)\right| \\ &\leq \left|P_n m(\theta,\overline{t}) - P_0 m(\theta,\overline{t})\right| + \left|P_0 m(\theta,\overline{t}) - P_0 m(\theta,t^*)\right|.\end{aligned}$$

The first term tends to 0 in probability by (8.4), the second term tends to 0 by dominated convergence Theorem using assumption (b).

Proof of Theorem 5.5. (1) By Taylor expansion, there exists $\overline{t} \in \mathbb{R}^{l+1}$ inside the segment that links $\widehat{t}$ and t^* with

$$\begin{aligned} 0 &= P_n m'(\theta,\widehat{t}) \\ &= P_n m'(\theta,t^*) + \left(P_n m''(\theta,t^*)\right)^T \left(\widehat{t}-t^*\right) \\ &\quad +\tfrac{1}{2}\left(\widehat{t}-t^*\right)^T P_n m'''(\theta,\overline{t})\left(\widehat{t}-t^*\right).\end{aligned} \tag{8.6}$$

By condition (a) and the Law of Large Numbers (LLN), we get $P_n m'''(\theta,\overline{t}) = O_P(1)$. Hence, we can write the last term in the right hand side of (8.6) as $o_P(1)\left(\widehat{t}-t^*\right)$. On the other hand, by the WLLN, $P_n m''(\theta,t^*)$ converges in probability to the matrix $P_0 m''(\theta,t^*)$. Write $P_n m''(\theta,t^*)$ as $P_0 m''(\theta,t^*) + o_P(1)$ to obtain from (8.6)

$$-P_n m'(\theta,t^*) = \left(P_0 m''(\theta,t^*) + o_P(1)\right)\left(\widehat{t}-t^*\right). \tag{8.7}$$

By the Central Limit Theorem (CLT), we have $\sqrt{n}P_n m'(\theta,t^*) = O_P(1)$, which by (8.7) implies that $\sqrt{n}\left(\widehat{t}-t^*\right) = O_P(1)$. Hence, from (8.7), we get

$$\sqrt{n}\left(\widehat{t}-t^*\right) = \left[-P_0 m''(\theta,t^*)\right]^{-1}\sqrt{n}P_n m'(\theta,t^*) + o_P(1). \tag{8.8}$$

The CLT concludes the proof of part 1. (2) Using the fact that $\left(\widehat{t}-t^*\right) = O_P(1/\sqrt{n})$ and $P_n m'(\theta,t^*) = P_0 m'(\theta,t^*) + o_P(1) = 0 + o_P(1) = o_P(1)$, we obtain

$$\begin{aligned}\sqrt{n}\left(\widehat{D}_\phi(\mathcal{M}_\theta,P_0) - D_\phi(\mathcal{M}_\theta,P_0)\right) &= \sqrt{n}\left(\widehat{D}_\phi(\mathcal{M}_\theta,P_0) - P_0 m(\theta,t^*)\right) \\ &= \sqrt{n}\left(P_n m(\theta,t^*) - P_0 m(\theta,t^*)\right) + o_P(1),\end{aligned}$$

and the CLT yields to the conclusion of the proof.

Proof of Theorem 5.6. (1) First note that condition (d) implies that the function $t \in \Lambda(\theta) \mapsto \mathbb{E}m(X,\theta,t)$ is strictly concave for all $\theta \in \Theta$. Hence, condition (c) implies that $t^*(\theta)$ is unique for all $\theta \in \Theta$. By UWLLN, using continuity of $m(X,\theta,t)$, in θ and t, and condition (c), we obtain the uniform convergence in probability, over the compact set $\{(\theta,t) \mid \theta \in \Theta, t \in N(\theta)\}$,

$$\sup_{\theta\in\Theta, t\in N(\theta)} |P_n m(\theta,t) - P_0 m(\theta,t)| \to 0. \tag{8.9}$$

We can then prove the convergence in probability $\sup_{\theta\in\Theta}\|\widehat{t}(\theta)-t^*(\theta)\|\to 0$ in two steps. Step 1: let $\eta>0$, we will show that $P_0\left[\sup_{\theta\in\Theta}\|\overline{t}(\theta)-t^*(\theta)\|\geq\eta\right]\to 0$ for any value

$$\overline{t}(\theta):=\arg\sup_{t\in N(\theta)}P_nm(\theta,t). \tag{8.10}$$

Step 2: to conclude the proof we will show that $\widehat{t}(\theta)$ belongs to $\mathrm{int}(N(\theta))$ with probability one as $n\to\infty$ for all $\theta\in\Theta$. Let $\eta>0$ such that $\sup_{\theta\in\Theta}\|\overline{t}(\theta)-t^*(\theta)\|\geq\eta$. Sine Θ is a compact set, by continuity there exists $\overline{\theta}\in\Theta$ such that $\sup_{\theta\in\Theta}\|\overline{t}(\theta)-t^*(\theta)\|=\|\overline{t}(\overline{\theta})-t^*(\overline{\theta})\|\geq\eta$. Hence, there exists $\varepsilon>0$ such that $P_0m(\overline{\theta},t^*(\overline{\theta}))-P_0m(\overline{\theta},\overline{t}(\overline{\theta}))>\varepsilon$. In fact, ε may be defined as follows

$$\varepsilon:=\inf_{\theta\in\Theta}\sup_{t\in N(\theta):\|t-t^*(\theta)\|\geq\eta}\mathbb{E}[m(X,\theta,t^*(\theta))]-\mathbb{E}[m(X,\theta,t)]$$

which is strictly positive by the strict concavity of $\mathbb{E}[m(X,\theta,t)]$ in t for all $\theta\in\Theta$, the uniqueness of $t^*(\theta)\in\mathrm{int}(N(\theta))$ and the fact that Θ is compact. Hence the event $\left[\sup_{\theta\in\Theta}\|\overline{t}(\theta)-t^*(\theta)\|\geq\eta\right]$ implies the event

$$\left[P_0m(\overline{\theta},t^*(\overline{\theta}))-P_0m(\overline{\theta},\overline{t}(\theta))\geq\varepsilon\right],$$

from which we obtain

$$P_0\left[\sup_{\theta\in\Theta}\|\overline{t}(\theta)-t^*(\theta)\|\geq\eta\right]\leq P_0\left[P_0m(\overline{\theta},t^*(\overline{\theta}))-P_0m(\overline{\theta},\overline{t}(\theta))\geq\varepsilon\right]. \tag{8.11}$$

On the other hand, by (8.9), we have

$$\begin{aligned}P_0m(\overline{\theta},t^*(\overline{\theta}))-P_0m(\overline{\theta},\overline{t}(\theta)) &= P_nm(\overline{\theta},t^*(\overline{\theta}))-P_0m(\overline{\theta},\overline{t}(\theta))+o_P(1)\\ &\leq P_nm(\overline{\theta},\overline{t}(\overline{\theta}))-P_0m(\overline{\theta},\overline{t}(\theta))+o_P(1)\\ &\leq \sup_{\theta\in\Theta,t\in N(\theta)}|P_nm(\theta,t)-P_0m(\theta,t)|+o_P(1).\end{aligned}$$

Combining this with (8.11) and (8.9), we conclude that $\sup_{\theta\in\Theta}\|\overline{t}(\theta)-t^*(\theta)\|\to 0$ in probability. In particular, $\overline{t}(\theta)\in\mathrm{int}(N(\theta))$ for sufficiently large n, for all $\theta\in\Theta$. Since $t\mapsto P_nm(\theta,t)$ is concave then $\widehat{t}(\theta)$ must be in $\mathrm{int}(N(\theta))$ for sufficiently large n; hence the same results holds when $\overline{t}$ is replaced by $\widehat{t}$.
(2) From (1), we have for large n,

$$\begin{aligned}\sup_{\theta\in\Theta}|P_nm(\theta,\widehat{t}(\theta))-P_0m(\theta,t^*(\theta))| &= \sup_{\theta\in\Theta}|P_nm(\theta,\overline{t}(\theta))-P_0m(\theta,t^*(\theta))|\\ &\leq \sup_{\theta\in\Theta}|P_nm(\theta,\overline{t}(\theta))-P_0m(\theta,\overline{t}(\theta))|\\ &\quad+\sup_{\theta\in\Theta}|P_0m(\theta,\overline{t}(\theta))-P_0m(\theta,t^*(\theta))|.\end{aligned}$$

Both terms in the above display tend to 0; the first one by (8.9), the second one by Dominated convergence Theorem using assumption (c). Now, since the minimizer θ^* of $\theta\mapsto P_0m(\theta,t^*(\theta))$ over the compact set Θ is unique, by continuity and the above uniform convergence, we conclude that $\widehat{\theta}_\phi$ tends in probability to θ^*.
(3) This holds as a consequence of the uniform convergence in probability $\sup_{\theta\in\Theta}|P_nm(\theta,\widehat{t}(\theta))-P_0m(\theta,t^*(\theta))|\to 0$ proved in part (2) above.

Proof of Theorem 5.7. By the first order conditions, we have

$$\left\{\begin{array}{rcl}P_n\frac{\partial}{\partial t}m\left(\theta,t\right) &=& 0\\ P_n\frac{\partial}{\partial\theta}m\left(\theta,t(\theta)\right) &=& 0,\end{array}\right.$$

i.e.,

$$\left\{\begin{array}{rcl} P_n \frac{\partial}{\partial t} m\left(\widehat{\theta}, \widehat{t}(\widehat{\theta})\right) & = & 0 \\ P_n \frac{\partial}{\partial \theta} m\left(\widehat{\theta}, \widehat{t}(\widehat{\theta})\right) + P_n \frac{\partial}{\partial t} m\left(\widehat{\theta}, \widehat{t}(\widehat{\theta})\right) \frac{\partial}{\partial \theta} \widehat{t}(\widehat{\theta}) & = & 0. \end{array}\right.$$

The second term in the left hand side of the second equation is equal to 0, due to the first equation. Hence $\widehat{t}(\widehat{\theta})$ and $\widehat{\theta}$ are solutions of the somehow simpler system

$$\left\{\begin{array}{rcl} P_n \frac{\partial}{\partial t} m\left(\widehat{\theta}, \widehat{t}(\widehat{\theta})\right) & = & 0 \ (E1) \\ P_n \frac{\partial}{\partial \theta} m\left(\widehat{\theta}, \widehat{t}(\widehat{\theta})\right) & = & 0 \ (E2). \end{array}\right.$$

Use a Taylor expansion in (E1); there exists $\left(\overline{\theta}, \overline{t}\right)$ inside the segment that links $(\widehat{\theta}, \widehat{t}(\widehat{\theta}))$ and $(\theta^*, t^*(\theta^*))$ such that

$$\begin{aligned} 0 &= P_n \frac{\partial}{\partial t} m\left(\theta^*, t^*(\theta^*)\right) + \left[\left(P_n \frac{\partial^2}{\partial t^2} m(\theta^*, t^*(\theta^*))\right)^T, \left(P_n \frac{\partial^2}{\partial \theta \partial t} m(\theta^*, t^*(\theta^*))\right)^T\right] a_n \\ &\quad + \frac{1}{2} a_n^T A_n a_n, \end{aligned} \tag{8.12}$$

with

$$a_n := \left(\left(\widehat{t}(\widehat{\theta}) - t^*(\theta^*)\right)^T, \left(\widehat{\theta} - \theta^*\right)^T\right)^T \tag{8.13}$$

and

$$A_n := \left(\begin{array}{cc} P_n \frac{\partial^3}{\partial t^3} m(\overline{\theta}, \overline{c}) & P_n \frac{\partial^3}{\partial t \partial \theta \partial t} m(\overline{\theta}, \overline{c}) \\ P_n \frac{\partial^3}{\partial \theta \partial t^2} m(\overline{\theta}, \overline{c}) & P_n \frac{\partial^3}{\partial \theta^2 \partial t} m(\overline{\theta}, \overline{c}) \end{array}\right). \tag{8.14}$$

By condition (a) and the WLLN we have $A_n = O_P(1)$. So using (A.4), we can write the last term in right hand side of (8.12) as $o_P(1)a_n$. On the other hand by (A.6), we can write also $\left[\left(P_n \frac{\partial^2}{\partial t^2} m(\theta^*, t^*)\right)^T, \left(P_n \frac{\partial^2}{\partial \theta \partial t} m(\theta^*, t^*)\right)^T\right]$ as $\left[P_0 \frac{\partial^2}{\partial t^2} m(\theta^*, t^*), \left(P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta^*, t^*)\right)^T\right] + o_P(1)$ to obtain from (8.12)

$$-P_n \frac{\partial}{\partial t} m(\theta^*, t^*) = \left[P_0 \frac{\partial^2}{\partial t^2} m(\theta^*, t^*) + o_P(1), \left(P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta^*, t^*)\right)^T + o_P(1)\right] a_n. \tag{8.15}$$

In the same way, using a Taylor expansion in (E2), there exists $(\overline{\theta}, \overline{t})$ inside the segment that links $\left(\widehat{\theta}, \widehat{t}\right)$ and (θ^*, t^*) such that

$$\begin{aligned} 0 &= P_n \frac{\partial}{\partial \theta} m(\theta^*, t^*) + \left[\left(P_n \frac{\partial^2}{\partial t \partial \theta} m(\theta^*, t^*)\right)^T, \left(P_n \frac{\partial^2}{\partial \theta^2} m(\theta^*, t^*)\right)^T\right] a_n \\ &\quad + \frac{1}{2} a_n^t B_n a_n, \end{aligned} \tag{8.16}$$

with

$$B_n := \left[\begin{array}{cc} P_n \frac{\partial^3}{\partial t^2 \partial \theta} m(\overline{\theta}, \overline{t}) & P_n \frac{\partial^3}{\partial t \partial \theta^2} m(\overline{\theta}, \overline{t}) \\ P_n \frac{\partial^3}{\partial \theta \partial t \partial \theta} m(\overline{\theta}, \overline{t}) & P_n \frac{\partial^3}{\partial \theta^3} m(\overline{\theta}, \overline{t}) \end{array}\right].$$

As in (8.15), we obtain

$$-P_n \frac{\partial}{\partial \theta} m(\theta^*, t^*) = \left[\left(P_0 \frac{\partial^2}{\partial t \partial \theta} m(\theta^*, t^*)\right)^T + o_P(1), P_0 \frac{\partial^2}{\partial \theta^2} m(\theta^*, t^*) + o_P(1)\right] a_n. \tag{8.17}$$

From (8.15) and (8.17), we get

$$\begin{aligned}\sqrt{n}a_n &= \sqrt{n}\left(\begin{array}{cc} P_0\frac{\partial^2}{\partial t^2}m(\theta^*,t^*) & \left(P_0\frac{\partial^2}{\partial\theta\partial t}m(\theta^*,t^*)\right)^T \\ \left(P_0\frac{\partial^2}{\partial t\partial\theta}m(\theta^*,t^*)\right)^T & P_0\frac{\partial^2}{\partial\theta^2}m(\theta^*,t^*)\end{array}\right)^{-1}\times \\ &\quad \times\left(\begin{array}{c} -P_n\frac{\partial}{\partial t}m(\theta^*,t^*) \\ -P_n\frac{\partial}{\partial\theta}m(\theta^*,t^*)\end{array}\right)+o_P(1). \qquad (8.18)\end{aligned}$$

Denote S the $(l+1+d)\times(l+1+d)-$matrix defined by

$$S := \left(\begin{array}{cc} S_{11} & S_{12} \\ S_{21} & S_{22}\end{array}\right) := \left(\begin{array}{cc} P_0\frac{\partial^2}{\partial t^2}m(\theta^*,t^*) & \left(P_0\frac{\partial^2}{\partial\theta\partial t}m(\theta^*,t^*)\right)^T \\ \left(P_0\frac{\partial^2}{\partial t\partial\theta}m(\theta^*,t^*)\right)^T & P_0\frac{\partial^2}{\partial\theta^2}m(\theta^*,t^*)\end{array}\right). \qquad (8.19)$$

Hence, we obtain

$$\sqrt{n}\left(\begin{array}{c}\widehat{t}(\widehat{\theta})-t^* \\ \widehat{\theta}-\theta^*\end{array}\right) = \sqrt{n}S^{-1}\left(\begin{array}{c} -P_n\frac{\partial}{\partial t}m(\theta^*,t^*) \\ -P_n\frac{\partial}{\partial\theta}m(\theta^*,t^*)\end{array}\right)+o_P(1),$$

and the CLT concludes the proof.

References

Baggerly, K. A. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika*, **85**(3), 535–547.

Broniatowski, M. and Keziou, A. (2006). Minimization of ϕ-divergences on sets of signed measures. *Studia Sci. Math. Hungar. (arXiv:1003.5457)*, **43**(4), 403–442.

Broniatowski, M. and Keziou, A. (2008). Estimation and tests for models satisfying linear constraints with unknown parameter. *arXiv:0811.3477v1.*

Broniatowski, M. and Keziou, A. (2009). Parametric estimation and tests through divergences and the duality technique. *J. Multivariate Anal.*, **100**(1), 16–36.

Corcoran, S. (1998). Bertlett adjustement of empirical discrepancy statistics. *Biometrika*, **85**, 967–972.

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B*, **46**(3), 440–464.

Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **8**, 85–108.

Csiszár, I. (1967). On topology properties of f-divergences. *Studia Sci. Math. Hungar.*, **2**, 329–339.

Haberman, S. J. (1984). Adjustment by minimum discriminant information. *Ann. Statist.*, **12**(3), 971–988.

Hansen, L., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business and Economic Statistics*, **14**, 462–2800.

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

Imbens, G. W. (1997). One-step estimators for over-identified generalized method of moments models. *Rev. Econom. Stud.*, **64**(3), 359–383.

Kitamura, Y. (2001). Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica*, **69**(6), 1661–1672.

Liese, F. and Vajda, I. (1987). *Convex statistical distances*, volume 95. BSB B. G. Teubner Verlagsgesellschaft, Leipzig.

McCullagh, P. and Nelder, J. A. (1983). *Generalized linear models*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Newey, W. K. and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, **72**(1), 219–255.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18**(1), 90–120.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**(2), 237–249.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall, New York.

Pardo, L. (2006). *Statistical inference based on divergence measures*, volume 185 of *Statistics: Textbooks and Monographs*. Chapman & Hall/CRC, Boca Raton, FL.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.*, **22**(1), 300–325.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton University Press, Princeton, N.J.

Schennach, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. *Ann. Statist.*, **35**(2), 634–672.

Sheehy, A. (1987). Kullback-Leibler constrained estimation of probability measures. *Report, Dept. Statistics, Stanford Univ.*

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**(1), 1–25.

*LSTA-Université Paris 6. E-Mail: michel.broniatowski@upmc.fr

**Laboratoire de Mathématiques, Université de Reims and LSTA-Université Paris 6.
E-Mail: amor.keziou@upmc.fr