DIVERSIONES AND DUALITY FOR ESTIMATION AND TEST UNDER MOMENT CONDITION MODELS

MICHEL BRONIATOWSKI\textsuperscript{1} AND AMOR KEZIOU\textsuperscript{1,2}

\textbf{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 approach. We treat the problems of existence and characterization of the divergence projections of probability distributions on sets of signed finite measures. We give a precise characterization of duality, for the proposed class of estimates and test statistics, which is used to derive their limiting distributions (including the EL estimate and the EL ratio statistic) both under the null hypotheses and under alternatives or misspecification. An approximation to the power function is deduced as well as the sample size which ensures a desired power for a given alternative.

\textbf{Keywords:} Empirical likelihood; Generalized Empirical likelihood; Minimum divergence; Efficiency; Power function; Duality; Divergence projection.


\textbf{Contents}

1. Introduction and notation \hfill 2
2. Statistical divergences \hfill 6
3. Minimum divergence estimates \hfill 8
4. Dual representation of $\phi$–divergences under constraints \hfill 11
5. Asymptotic properties of the estimates of the parameter and the divergences \hfill 19
5.1. Asymptotic properties under the model \hfill 19

\textit{Date:} February 29, 2012, à paraître dans JSPI.
1. Introduction and notation

Statistical models are often defined through estimating equations

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

where \( \mathbb{E}[\cdot] \) denotes the mathematical expectation, \( g := (g_1, \ldots, g_l)^\top \in \mathbb{R}^l \) is some specified vector valued function of a random vector \( X \in \mathbb{R}^m \) and a parameter vector \( \theta \in \Theta \subset \mathbb{R}^d \). 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. Denote \( P_0 \) the probability distribution of the random vector \( X \). Then the above estimating equations can be written as

\[ \int_{\mathbb{R}^m} g(x, \theta) \, dP_0(x) = 0. \]

Denoting \( M^1 \) the collection of all probability measures (p.m.) on the measurable space \( (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m)) \), the submodel \( \mathcal{M}_\theta^1 \), associated to a given value \( \theta \) of the parameter, consists of all distributions \( Q \) satisfying \( l \) linear constraints induced by the vector valued function \( g(\cdot, \theta) \), namely

\[ \mathcal{M}_\theta^1 := \left\{ Q \in M^1 \text{ such that } \int g(x, \theta) \, dQ(x) = 0 \right\}. \]
with \( l \geq d \). The statistical model which we consider can be written as

\[
M_1 := \bigcup_{\theta \in \Theta} \mathcal{M}_\theta^1.
\]  

(1.1)

Let \( X_1, \ldots, X_n \) denote an i.i.d. sample of \( X \) with unknown distribution \( P_0 \). We denote \( \theta_0 \), if it exists, the value of the parameter such that \( P_0 \) belongs to \( \mathcal{M}_{\theta_0}^1 \), namely the value satisfying

\[
E[g(X, \theta_0)] = 0,
\]

and we assume obviously that \( \theta_0 \) is unique. This paper addresses the two following natural questions:

**Problem 1:** Does \( P_0 \) belong to the model \( M_1 \)?

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

When \( m = d = l \), and \( g(x, \theta) = x - \theta \), then the model is the same as those of Owen (1988) and Owen (1990), and in this case our interest concerns interval estimation or confidence areas construction for the parameter \( \theta \). The main interest, however, is the case where \( l > d \). We introduce some examples for illustration; see Qin and Lawless (1994), Guggenberger and Smith (2005) and Owen (2001).

**Example 1.1.** Sometimes we have information relating the first and second moments of a random variable \( X \) (see e.g. Godambe and Thompson (1989) and McCullagh and Nelder (1983)). Let \( X_1, \ldots, X_n \) be an i.i.d. sample of a random variable \( X \) with mean \( E(X) = \theta \), and assume that \( E(X^2) = h(\theta) \), where \( h(\cdot) \) is a known function. Our aim is to estimate \( \theta \). The information about the distribution \( P_0 \) of \( X \) can be expressed in the form of (1.1) by taking \( g(x, \theta) := (x - \theta, x^2 - h(\theta))^\top \).

**Example 1.2.** Let \( (X_{1,1}, X_{2,1}), \ldots, (X_{1,n}, X_{2,n}) \) be an i.i.d. sample of a bivariate random vector \( X := (X_1, X_2)^\top \) with \( E(X_1) = E(X_2) = \theta \). In this case, we can take \( g(x, \theta) = \)
\((x_1 - \theta, x_2 - \theta)^\top\). A somewhat similar problem is when \(\mathbb{E}(X_1) = c\) is known and \(\mathbb{E}(X_2) = \theta\) is to be estimated, by taking \(g(x, \theta) = (x_1 - c, x_2 - \theta)^\top\). Such problems are common in survey sampling (see e.g. Kuk and Mak (1989) and Chen and Qin (1993)).

**Example 1.3.** Let \(F_0\) be a continuous distribution function of a random variable \(X\) with probability distribution \(P_0\) that is symmetric about zero, namely \(F_0(t) = 1 - F_0(-t)\) for all \(t \in \mathbb{R}\). Consider estimation of the parameter \(\theta = F_0(t)\), for a given \(t \in \mathbb{R}\), from an i.i.d. sample \(X_1, \ldots, X_n\) of \(X\). This problem can be handled in the context of model (1.1) by taking \(g(x, \theta) = (1_{[-\infty, t]}(x) - \theta, 1_{[-t, +\infty]}(x) - \theta)^\top\).

We note that the problems 1 and 2 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 estimate. The recent literature in econometrics focuses on such models; Smith (1997), Newey and Smith (2004) provided a class of estimates called generalized empirical likelihood (GEL) estimates which contains the EL and the CU ones. Schennach (2007) discussed the asymptotic properties of the empirical likelihood estimate under misspecification; the author showed the important fact that the EL estimate may cease to be root \(n\) consistent when the functions \(g_j\) defining the moments conditions and the support of \(P_0\) 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. Moreover, 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 \(\theta_0\) have been considered in the seminal paper by Owen (1990). Problems 1 and 2 have been handled via EL and GEL approaches in Qin and Lawless (1994), Smith (1997) and Newey and Smith (2004) under the null hypothesis \(\mathcal{H}_0 : P_0 \in \mathcal{M}^1\); the limiting distributions of the GEL estimates and the GEL test statistics have been obtained only under the model
and under the null hypotheses; Imbens (1997) discusses the asymptotic properties of the EL and exponential tilting estimates under misspecification and give the formula of the asymptotic variance, using dual characterizations, without presenting the hypotheses under which their results hold. Chen et al. (2007) give the limiting distribution of the EL estimate under misspecification as well as the EL ratio statistic between a parametric model and a moment condition model. The paper by Kitamura (2007) gives a discussion of duality for GEL estimates under moment condition models. Bertail (2006) uses duality to study, under the model, the asymptotic properties of the EL ratio statistic and its Bartlett correctability; the author extends his results to semiparametric problems with infinite-dimensional parameters.

The main contribution of the present paper is the precise characterization of duality for a large class of estimates and test statistics (including GEL and EL ones) and its use in deriving the limiting properties of both the estimates and the test statistics under misspecification and under alternative hypotheses. Moreover,

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 propose a wide class of estimates, test statistics and confidence regions for the parameter $\theta_0$ as well as various test statistics for Problems 1 and 2, all depending on the choice of the divergence.

2) The limiting 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 limiting distributions of the proposed estimates and test statistics (including the EL ones) both under the null hypotheses, under alternatives and under misspecification.

3) The limiting 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 divergence.

From the point of view of the statistical criterion under consideration, the main advantage, of using a divergence based approach and duality, lays in the fact that it leads to asymptotic 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; Keziou (2003) and Broniatowski and Keziou (2009) stated the consistency and obtained the limiting 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 for the proposed class of estimates and test statistics (including the EL ones) in the context of semiparametric models (1.1).

The rest of the paper is organized as follows. Section 2 describes the statistical divergences used in the sequel. Section 3 is devoted to the description of the proposed estimation and test procedures. In Section 3, we adapt the Lagrangian duality formalism 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 the test statistics under the model and under misspecification. 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. on the measurable space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$. Denote by $M$ the space of all signed finite measures (s.f.m.) on
Let \( \varphi \) be a convex function from \( \mathbb{R} \) onto \([0, +\infty]\) with \( \varphi(1) = 0 \), and such that its domain, \( \text{dom}\varphi := \{x \in \mathbb{R} \text{ such that } \varphi(x) < \infty \} =: (a, b) \), is an interval, with endpoints satisfying \( a < 1 < b \), which may be bounded or unbounded, open or not. We assume that \( \varphi \) is closed; the closedness of \( \varphi \) means that if \( a \) or \( b \) are finite then
\[
\varphi(x) \to \varphi(a) \quad \text{when } x \downarrow a,
\]
and
\[
\varphi(x) \to \varphi(b) \quad \text{when } x \uparrow b.
\]
Note that, this is equivalent to the fact that the level sets \( \{x \in \mathbb{R}; \varphi(x) \leq \alpha \} \), \( \forall \alpha \in \mathbb{R} \), are closed in \( \mathbb{R} \) endowed with the usual topology.

For any s.f.m. \( Q \in M \), the \( \varphi \)-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
\[
D_\varphi(Q, P) := \int_{\mathbb{R}^m} \varphi\left(\frac{dQ}{dP}(x)\right) dP(x),
\]
(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 \( D_\varphi(Q, P) := +\infty \). For any p.m. \( P \), the mapping \( Q \in M \mapsto D_\varphi(Q, P) \) is convex and takes nonnegative values. When \( Q = P \) then \( D_\varphi(Q, P) = 0 \). Furthermore, if the function \( x \mapsto \varphi(x) \) is strictly convex on a neighborhood of \( x = 1 \), then
\[
D_\varphi(Q, P) = 0 \quad \text{if and only if } Q = P.
\]
(2.2)

All the above properties are presented in Csiszár (1963), Csiszár (1967) and in Chapter 1 of Liese and Vajda (1987), for \( \varphi \)-divergences defined on the set of all p.m.’s \( M^1 \). When the \( \varphi \)-divergences are extended to \( M \), then the same arguments as developed on \( M^1 \) hold.

When defined on \( M^1 \), the Kullback-Leibler (\( KL \)), modified Kullback-Leibler (\( KL_m \)), \( \chi^2 \), modified \( \chi^2 \) (\( \chi^2_m \)), Hellinger (\( H \)), and \( L^1 \) divergences are respectively associated to the convex functions \( \varphi(x) = x \log x - x + 1 \), \( \varphi(x) = -\log x + x - 1 \), \( \varphi(x) = \frac{1}{2}(x-1)^2 \), \( \varphi(x) = \frac{1}{2}(x-1)^2/x \), \( \varphi(x) = 2(\sqrt{x} - 1)^2 \) and \( \varphi(x) = |x - 1| \). All these divergences except the \( L^1 \) one, belong to the class of the so called power divergences introduced in Cressie and Read (1984) (see also Liese and Vajda (1987) and Pardo (2006)). They are defined through the class of convex functions
\[
x \in \mathbb{R}^+_m \mapsto \varphi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)}
\]
(2.3)
if $\gamma \in \mathbb{R} \setminus \{0, 1\}$, $\varphi_0(x) := -\log x + x - 1$ and $\varphi_1(x) := x \log x - x + 1$. So, the $KL$–divergence is associated to $\varphi_1$, the $KL$ $m$ to $\varphi_0$, the $\chi^2$ to $\varphi_2$, the $\chi^2_m$ to $\varphi_{-1}$ and the Hellinger distance to $\varphi_{1/2}$. We extend the definition of the power divergences functions $Q \in M^1 \mapsto D_{\varphi_\gamma}(Q, P)$ onto the whole set of signed finite measures $M$ as follows. When the function $x \mapsto \varphi_\gamma(x)$ is not defined on $]-\infty, 0[$ or when $\varphi_\gamma$ is defined on $\mathbb{R}$ but is not convex (for instance when $\gamma = 3$), we extend the definition of $\varphi_\gamma$ as follows

$$x \in \mathbb{R} \mapsto \varphi_\gamma(x) \mathbb{I}_{[0, +\infty]}(x) + (+\infty) \mathbb{I}_{]-\infty, 0[}(x). \quad (2.4)$$

Note that for $\chi^2$-divergence, the corresponding $\varphi$ function $\varphi(x) = \frac{1}{2}(x - 1)^2$ is convex and defined on whole $\mathbb{R}$. In this paper, for technical considerations, we assume that the functions $\varphi$ are strictly convex on their domain $(a, b)$, twice continuously differentiable on $]a, b[$, the interior of their domain. Hence, $\varphi'(1) = 0$, and for all $x \in ]a, b[$, $\varphi''(x) > 0$. Here, $\varphi'$ and $\varphi''$ are used to denote respectively the first and the second derivative functions of $\varphi$. Note that the above assumptions on $\varphi$ are not restrictive, and that all the power functions $\varphi_\gamma$, see (2.4), satisfy the above conditions, including all standard divergences.

**Definition 2.1.** Let $\Omega$ be some subset of $M$. The $\varphi$–divergence between the set $\Omega$ and a p.m. $P$ is defined by

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

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

$$D_{\varphi}(Q^*, P) \leq D_{\varphi}(Q, P) \quad \text{for all } Q \in \Omega,$$

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

**3. Minimum divergence estimates**

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

$$P_n(\cdot) := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}(\cdot),$$
where $\delta_x(\cdot)$ denotes the Dirac measure at point $x$, for all $x$. We will endow our statistical approach in the global context of s.f.m.’s with total mass 1 satisfying $l$ linear constraints, namely,

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

and

$$\mathcal{M} := \bigcup_{\theta \in \Theta} \mathcal{M}_\theta, \quad (3.2)$$

sets of signed finite measures that replace $\mathcal{M}_1$ 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; see also Remark 4.6 below. The “plug-in” estimate of $D_\varphi(\mathcal{M}_\theta, P_0)$ is

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

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

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

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

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

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

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

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

$$\hat{\theta}_\varphi := \text{arg inf}_{\theta \in \Theta} \inf_{Q \in \mathcal{M}^{(n)}_\theta} \frac{1}{n} \sum_{i=1}^n \varphi \left( nQ(X_i) \right). \quad (3.7)$$
Enhancing $\mathcal{M}^1$ to $\mathcal{M}$ and accordingly extensions in the definitions of the $\varphi$ functions on $\mathbb{R}$ and the $\varphi$-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 $\varphi$ is included in $[0, +\infty[$ then minimizing over $\mathcal{M}^1$ or over $\mathcal{M}$ leads to the same estimates and test statistics. It is the case of the $KL_m$, $KL$, modified $\chi^2$ and Hellinger divergences.

- Let $\theta$ be a given value in $\Theta$. Denote $Q_{\theta}^{(1,n)}$ and $Q_{\theta}^{(n)}$, respectively, the projection of $P_n$ on $\mathcal{M}^1_{\theta}$ and on $\mathcal{M}_{\theta}$. If $Q_{\theta}^{(1,n)}$ satisfies $0 < Q_{\theta}^{(1,n)}(X_i) < 1$, for all $i = 1, \ldots, n$, then $Q_{\theta}^{(1,n)} = Q_{\theta}^{(n)}$. Therefore, in this case, both approaches leads also to the same estimates and test statistics.

- It may occur that for some $\theta$ in $\Theta$ and some $i = 1, \ldots, n$, $Q_{\theta}^{(1,n)}(X_i)$ is a boundary value of $[0, 1]$, hence the first order conditions are not met which makes a real difficulty for computing the estimates over the sets of p.m. $\mathcal{M}^1_{\theta}$ 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 $\text{dom}\varphi = \mathbb{R}$, which is the case for the $\chi^2$-divergence. Other arguments are given in Remark 4.6 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 $\varphi$—divergences with $\varphi(x) = \varphi_0(x) := -\log x + x - 1$, namely on $KL_m$—divergence. Indeed, it is straightforward to see that the empirical log-likelihood ratio statistic for testing $\mathcal{H}_0 : P_0 \in \mathcal{M}$ against $\mathcal{H}_1 : P_0 \notin \mathcal{M}$, in the context of $\varphi$-divergences, can be written as $2n\hat{D}_{KL_m}(\mathcal{M}, P_0)$; and that the EL estimate of $\theta_0$ can be written as $\hat{\theta}_{KL_m} = \arg\inf_{\theta \in \Theta} \hat{D}_{KL_m}(\mathcal{M}_{\theta}, P_0)$; see Remark 4.4 below. In the case of the power functions $\varphi = \varphi_\gamma$, the corresponding estimates (3.7) belong to the class of GEL estimates introduced by Smith (1997) and Newey and Smith (2004), and (3.5) in this case are the empirical Cressie-Read statistics introduced by Baggerly (1998) and Corcoran (1998); see Remark 4.5 below.
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 below from Rockafellar (1970). On the other hand, the obtaining of asymptotic statistical results of the estimates and the test statistics, under misspecification or under alternative hypotheses, requires handle existence conditions and characterization of the projection of $P_0$ on the submodel $\mathcal{M}_\theta$ or on the model $\mathcal{M}$. This also will be considered through duality, along the following Section.

4. Dual representation of $\varphi$–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

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

and its empirical counterpart (3.5), make use of the so-called Fenchel-Legendre transform of $\varphi$, defined by

$\psi : t \in \mathbb{R} \mapsto \psi(t) := \sup_{x \in \mathbb{R}} \{ tx - \varphi(x) \}.$

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

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

and

$\sup_{t \in \mathbb{R}^{1+l}} \left\{ t_0 - \int_{\mathbb{R}^m} \psi(t_0 + \sum_{j=1}^l t_j g_j(x, \theta)) \, dP_0(x) \right\} = \sup_{t \in \mathbb{R}^{1+l}} \left\{ t_0 - \frac{1}{n} \sum_{i=1}^n \psi(t_0 + \sum_{j=1}^l t_j g_j(X_i, \theta)) \right\}.$ \hspace{1cm} (4.4)

In the following Propositions 4.1 and 4.2, 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). We give also sufficient conditions for the existence of the “primal optimal” solutions (i.e.,
the projection $Q_0^\theta$ of $P_0$ on the set $\mathcal{M}_\theta$, and the projection $Q_0^{(n)}$ of $P_n$ on the set $\mathcal{M}_\theta^{(n)}$, which will be related to the “dual attainment” problem, namely the existence of the supremum (in $t \in \mathbb{R}^{1+I}$) in the dual problems (4.3) and (4.4).

First, recall some properties of the convex conjugate $\psi$ of $\varphi$. For the proofs we can refer to Section 26 in Rockafellar (1970). Theses properties will be used to determine the convex conjugates $\psi$ of some standard divergence functions $\varphi$; see Table 1 below. The function $\psi$ is convex and closed, its domain is an interval $(a^*, b^*)$ with endpoints

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

(4.5)

satisfying $a^* < 0 < b^*$ with $\psi(0) = 0$. Note that the interval

$$(\tilde{a}, \tilde{b}) := \left( \lim_{x \downarrow a} \frac{\varphi(x)}{x}, \lim_{x \uparrow b} \frac{\varphi(x)}{x} \right)$$

can be different from $(a^*, b^*)$, the real domain of $\psi$ given by (4.5). This holds when $a$ or $b$ is finite and $\varphi'(a)$ or $\varphi'(b)$ is finite, respectively. For example, for the convex function

$$\varphi(x) = \frac{1}{2}(x^2 - 1)^2 1_{\mathbb{R}^+}(x) + (+\infty) 1_{[0, +\infty)}(x),$$

we have $\text{dom} \varphi = [0, +\infty]$ and $\varphi'(0) = -1$, and we can see that the domain of $\psi$ is $(a^*, b^*) = (-\infty, +\infty)$ which is different from $(\tilde{a}, \tilde{b}) = \left( \lim_{x \downarrow 0} \frac{\varphi(x)}{x}, \lim_{x \uparrow +\infty} \frac{\varphi(x)}{x} \right) = (-1, +\infty)$. The two intervals $(a^*, b^*)$ and $(\tilde{a}, \tilde{b})$ coincide if the function $\varphi$ is “essentially smooth”, i.e., differentiable with

$$\lim_{t \downarrow a} \varphi'(t) = -\infty \quad \text{if} \quad a \text{ is finite},$$
$$\lim_{t \uparrow b} \varphi'(t) = +\infty \quad \text{if} \quad b \text{ is finite}.$$ (4.6)

The strict convexity of $\varphi$ on its domain $(a, b)$ is equivalent to the condition that its conjugate $\psi$ is essentially smooth, i.e., differentiable with

$$\lim_{t \downarrow a^*} \psi'(t) = -\infty \quad \text{if} \quad a^* \text{ is finite},$$
$$\lim_{t \uparrow b^*} \psi'(t) = +\infty \quad \text{if} \quad b^* \text{ is finite}.$$ (4.7)
Conversely, $\varphi$ is essentially smooth on its domain $(a, b)$ if and only if $\psi$ is strictly convex on its domain $(a^*, b^*)$.

In all the sequel, we assume additionally that $\varphi$ is essentially smooth. Hence, $\psi$ is strictly convex on its domain $(a^*, b^*)$, and it holds that

$$a^* = \lim_{x \to -\infty} \frac{\varphi(x)}{x} = \lim_{x \downarrow a} \frac{\varphi(x)}{x} = \lim \varphi'(x), \quad b^* = \lim_{x \to +\infty} \frac{\varphi(x)}{x} = \lim_{x \uparrow b} \frac{\varphi(x)}{x} = \lim \varphi'(x),$$

and

$$\psi(t) = t \varphi'^{-1}(t) - \varphi\left(\varphi'^{-1}(t)\right), \quad \text{for all } t \in]a^*, b^[, \quad (4.8)$$

where $\varphi'^{-1}$ denotes the inverse function of $\varphi'$. It holds also that $\psi$ is twice continuously differentiable on $]a^*, b^[$ with

$$\psi'(t) = \varphi'^{-1}(t) \quad \text{and} \quad \psi''(t) = \frac{1}{\varphi''\left(\varphi'^{-1}(t)\right)}. \quad (4.9)$$

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

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

which may be finite or infinite. Hence, by closedness of $\psi$, we have

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

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

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

$$a < Q_0(X_i) < b, \quad \text{for all } i = 1, \ldots, n, \quad (4.10)$$
Table 1. Convex conjugates for some standard divergences.

<table>
<thead>
<tr>
<th>$D_{\varphi}$</th>
<th>$\varphi(x)$</th>
<th>$\text{dom}\varphi$</th>
<th>$\text{dom}\psi$</th>
<th>$\psi(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{KL,m}$</td>
<td>$\varphi(x) := -\log x + x - 1$</td>
<td>$]0, +\infty[$</td>
<td>$]-\infty, 1[$</td>
<td>$\psi(t) = -\log(1 - t)$</td>
</tr>
<tr>
<td>$D_{KL}$</td>
<td>$\varphi(x) := x \log x - x$</td>
<td>$]0, +\infty[$</td>
<td>$\mathbb{R}$</td>
<td>$\psi(t) = e^t - 1$</td>
</tr>
<tr>
<td>$D_{\chi_2}^m$</td>
<td>$\varphi(x) := \frac{1}{2} (x - 1)^2$</td>
<td>$]0, +\infty[$</td>
<td>$]-\infty, \frac{1}{2}]$</td>
<td>$\psi(t) = 1 - \sqrt{1 - 2t}$</td>
</tr>
<tr>
<td>$D_{\chi_2}$</td>
<td>$\varphi(x) := \frac{1}{2} (x - 1)^2$</td>
<td>$\mathbb{R}$</td>
<td>$\mathbb{R}$</td>
<td>$\psi(t) = \frac{1}{2} t^2 + t$</td>
</tr>
<tr>
<td>$D_H$</td>
<td>$\varphi(x) := 2(\sqrt{x} - 1)^2$</td>
<td>$[0, +\infty[$</td>
<td>$]-\infty, 2[$</td>
<td>$\psi(t) = \frac{2t}{2 - t}$</td>
</tr>
<tr>
<td>$D_{\varphi_\gamma}$</td>
<td>$\varphi(x) := \frac{x^{\gamma-2} - x^{\gamma-1}}{\gamma(\gamma-1)}$</td>
<td>$-$</td>
<td>$-$</td>
<td>$\psi(t) = \frac{1}{\gamma} (\gamma t - t + 1)^{\gamma-1} - \frac{1}{\gamma}$</td>
</tr>
</tbody>
</table>

Then

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

(4.11)

with dual attainment. Conversely, if there exists some dual optimal solution $\hat{t} := (\hat{t}_0, \hat{t}_1, \ldots, \hat{t}_l)^\top \in \mathbb{R}^{1+l}$ such that

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

(4.12)

then the equality (4.11) holds, and the unique optimal solution of the primal problem

$$\inf_{Q \in \mathcal{M}_\theta^{(n)}} D_{\varphi}(Q, P_n),$$

namely the projection of $P_n$ on $\mathcal{M}_\theta^{(n)}$, is given by

$$Q_\theta^{(n)}(X_i) = \frac{1}{n} \varphi^{-1}(\hat{t}_0 + \sum_{j=1}^l \hat{t}_j g_j(X_i, \theta)), \quad i = 1, \ldots, n,$$

where $\hat{t} := (\hat{t}_0, \hat{t}_1, \ldots, \hat{t}_l)^\top$ is solution of the system of equations

$$\begin{cases}
1 - \frac{1}{n} \sum_{i=1}^n \varphi^{-1}(\hat{t}_0 + \sum_{j=1}^l \hat{t}_j g_j(X_i, \theta)) = 0, \\
-\frac{1}{n} \sum_{i=1}^n g_j(X_i, \theta) \varphi^{-1}(\hat{t}_0 + \sum_{j=1}^l \hat{t}_j g_j(X_i, \theta)) = 0, \quad j = 1, \ldots, l.
\end{cases}$$

(4.13)

Remark 4.1. For the $\chi^2$-divergence, we have $a = -\infty$ and $b = +\infty$. Hence, condition (4.10) holds whenever $\mathcal{M}_\theta^{(n)}$ is not void. More generally, the above Proposition holds for any $\varphi$-divergence with $\text{dom}\varphi = \mathbb{R}$. 

Remark 4.2. Assume that \( g(x, \theta) := (x - \theta)\top \). So, for any divergence \( D_\phi \) with \( \text{dom}\phi = [0, +\infty[ \), which is the case of the modified \( \chi^2 \) divergence and the modified Kullback-Leibler divergence (or equivalently EL method), condition (4.10) means that \( \theta \) is an interior point of the convex hull of the data \((X_1, \ldots, X_n)\). 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 \( \theta \) be a given value in \( \Theta \). 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

\[
a < \inf_{x \in \mathbb{R}^m} \frac{dQ_0}{dP_0}(x) \leq \sup_{x \in \mathbb{R}^m} \frac{dQ_0}{dP_0}(x) < b, \quad P_0 - a.s., \tag{4.14}
\]

then

\[
\inf_{Q \in \mathcal{M}_\theta} D_\phi(Q, P_0) = \sup_{t \in \mathbb{R}^{1+l}} \left\{ t_0 - \int_{\mathbb{R}^m} \psi(t_0 + \sum_{j=1}^l t_j g_j(x, \theta)) \, dP_0(x) \right\} \tag{4.15}
\]

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

\[
\left\{ t \in \mathbb{R}^{1+l} \text{ such that } \int_{\mathbb{R}^m} |\psi(t_0 + \sum_{j=1}^l t_j g_j(x, \theta))| \, dP_0(x) < \infty \right\}, \tag{4.16}
\]

then the dual equality (4.15) holds, and the unique optimal solution \( Q^*_\theta \) 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

\[
\frac{dQ^*_\theta}{dP_0}(x) = \varphi^{-1}(t^*_0 + \sum_{j=1}^l t^*_j g_j(x, \theta)), \tag{4.17}
\]

where \( t^* := (t^*_0, t^*_1, \ldots, t^*_l)\top \) is solution of the system of equations

\[
\begin{align*}
1 - \int \varphi^{-1}(t^*_0 + \sum_{j=1}^l t^*_j g_j(x, \theta)) \, dP_0(x) &= 0, \\
- \int g_j(x, \theta) \varphi^{-1}(t^*_0 + \sum_{j=1}^l t^*_j g_j(x, \theta)) \, dP_0(x) &= 0, \quad j = 1, \ldots, l. \tag{4.18}
\end{align*}
\]

\(^1\)The strict inequalities (4.14) mean that \( P_0 \left\{ x \in \mathbb{R}^m \mid \frac{dQ_0}{dP_0}(x) \leq a \right\} = P_0 \left\{ x \in \mathbb{R}^m \mid \frac{dQ_0}{dP_0}(x) \geq b \right\} = 0.\)
Furthermore, \( t^* \) is unique if the functions \( 1_{\mathbb{R}^m}, g_1(., \theta), \ldots, g_l(., \theta) \) are linearly independent in the sense that \( P_0 \left\{ x \in \mathbb{R}^m \mid t_0 + \sum_{j=1}^l t_j g_j(x, \theta) \neq 0 \right\} > 0 \) for all \( t \in \mathbb{R}^m \) with \( t \neq 0 \).

**Remark 4.3.** Broniatowski and Keziou (2006) obtained the dual equality (4.15), results about the existence of the projection of \( P_0 \) on the sets \( \mathcal{M}_\theta \) and the characterization (4.17) under different assumptions; see Theorem 5.1, Corollary 5.2 and Proposition 5.3.

For sake of brevity and clearness, we must introduce some additional notations. In all the sequel, \( \|x\| \) denotes the norm of \( x \) defined by \( \|x\| := \sup_i |x_i| \) for any vector \( x := (x_1, \ldots, x_k) ^\top \in \mathbb{R}^k \), and for any matrix \( A \), the norm of \( A \) is defined by \( \|A\| := \sup_{i,j} |a_{i,j}| \). Denote by \( \bar{g} \) the vector valued function \( \bar{g} := (1_{\mathbb{R}^m}, g_1, \ldots, g_l)^\top \in \mathbb{R}^{1+l} \). For any p.m. \( P \) on \( (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m)) \) and any measurable function \( f \) from \( (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m)) \) to \( (\mathbb{R}, \mathcal{B}(\mathbb{R})) \), denote

\[
P f := \int_{\mathbb{R}^m} f(x) \, dP(x).
\]

Let

\[
t ^\top \bar{g}(x, \theta) := t_0 + \sum_{j=1}^l t_j g_j(x, \theta)
\]

and

\[
m(x, \theta, t) := t_0 - \psi(t ^\top \bar{g}(x, \theta)), \quad \text{for all } x \in \mathbb{R}^m, \theta \in \Theta \subset \mathbb{R}^d, t \in \mathbb{R}^{1+l}.
\]

Note that the sup in (4.11) and (4.15) can be restricted, respectively, to the sets

\[
\Lambda_\theta^{(n)} := \left\{ t \in \mathbb{R}^{1+l} \mid a^* < t ^\top \bar{g}(X_i, \theta) < b^*, \quad \text{for all } i = 1, \ldots, n \right\}
\]

and

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

In view of the above two Propositions 4.1 and 4.2, we redefine the estimates (3.5), (3.6) and (3.7) as follows

\[
\hat{D}_\varphi (\mathcal{M}_\theta, P_0) := \sup_{t \in \Lambda_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n m(X_i, \theta, t) := \sup_{t \in \Lambda_\theta} P_n m(\theta, t),
\]
\[ \hat{D}_\phi (\mathcal{M}, P_0) := \inf_{\theta \in \Theta} \sup_{t \in \Lambda(n)} \frac{1}{n} \sum_{i=1}^{n} m(X_i, \theta, t) := \inf_{\theta \in \Theta} \sup_{t \in \Lambda(n)} P_n m(\theta, t) \]  

(4.23)

and

\[ \hat{\theta}_\phi := \arg \inf_{\theta \in \Theta} \sup_{t \in \Lambda(n)} \frac{1}{n} \sum_{i=1}^{n} m(X_i, \theta, t) := \arg \inf_{\theta \in \Theta} \sup_{t \in \Lambda(n)} P_n m(\theta, t). \]  

(4.24)

**Remark 4.4.** 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 \( X_1, \ldots, X_n \).

The statistic \( 2n\hat{D}_{KLm}(\mathcal{M}, P_0) \), see (3.6), coincides with the empirical likelihood ratio statistic associated to the null hypothesis \( \mathcal{H}_0 : P_0 \in \mathcal{M} \) against the alternative \( \mathcal{H}_1 : P_0 \not\in \mathcal{M} \). The dual representation of \( \hat{D}_{KLm}(\mathcal{M}, P_0) \), see (4.23) and (4.11), is

\[ \hat{D}_{KLm}(\mathcal{M}, P_0) = \inf_{\theta \in \Theta} \sup_{t \in \Lambda(n)} \left\{ t_0 + \frac{1}{n} \sum_{i=1}^{n} \log(1 - \hat{t}_0 - \sum_{j=1}^{l} t_j g_j(X_i, \theta)) \right\} . \]

For a given \( \theta \in \Theta \), the KL\(_m\)-projection \( Q_\theta^{(n)} \), of \( P_n \) on \( \mathcal{M}_\theta \), is given by (see Proposition 4.1)

\[ \frac{1}{Q_\theta^{(n)}(X_i)} = n \left(1 - \hat{t}_0 - \sum_{j=1}^{l} \hat{t}_j g_j(X_i, \theta)\right), \quad i = 1, \ldots, n, \]

which, multiplying by \( Q_\theta^{(n)}(X_i) \) and summing upon \( i = 1, \ldots, n \), yields \( \hat{t}_0 = 0 \). Therefore, \( t_0 \) can be omitted, and the above representation can be rewritten as follows

\[ \hat{D}_{KLm}(\mathcal{M}, P_0) = \inf_{\theta \in \Theta} \sup_{t_1, \ldots, t_l \in \mathbb{R}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \log(1 + \sum_{j=1}^{l} t_j g_j(X_i, \theta)) \right\} \]

and then

\[ \hat{\theta}_{KLm} = \hat{\theta}_{EL} = \arg \inf_{\theta \in \Theta} \sup_{t_1, \ldots, t_l \in \mathbb{R}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \log(1 + \sum_{j=1}^{l} t_j g_j(X_i, \theta)) \right\} \]  

(4.25)

in which the sup is taken over the set

\[ \left\{ (t_1, \ldots, t_l)^\top \in \mathbb{R}^l \text{ such that } -1 < \sum_{j=1}^{l} t_j g_j(X_i, \theta) < +\infty, \text{ for all } i = 1, \ldots, n \right\} . \]
The formula (4.25) is the ordinary dual representation of the EL estimate; see Qin and Lawless (1994) and Owen (2001).

**Remark 4.5.** Consider the power divergences, associated to the power functions $\varphi_\gamma$; see (2.3) and (2.4). We will show that the estimates $\hat{\theta}_{\varphi_\gamma}$ belong to the class of GEL estimators introduced by Smith (1997) and Newey and Smith (2004). The projection $Q_{\varphi}^{(n)}$ of $P_n$ on $\mathcal{M}_{\theta}$ is given by

$$Q_{\varphi}^{(n)}(X_i) = \left( (\gamma - 1)(\hat{t}_0 + \sum_{j=1}^{l} \hat{t}_j g(X_i, \theta)) + 1 \right)^{1/(\gamma - 1)}, \quad i = 1, \ldots, n.$$  

Using the constraint $\sum_{i=1}^{n} Q_{\varphi}^{(n)}(X_i) = 1$, we can explicit $\hat{t}_0$ in terms of $\hat{t}_1, \ldots, \hat{t}_l$, and hence the sup in the dual representation (4.24) can be reduced to a subset of $\mathbb{R}^l$, as in Newey and Smith (2004). When $\varphi(x) = \frac{1}{2}(x - 1)^2$, it is straightforward to see that the corresponding estimate $\hat{\theta}_{\varphi}$ coincides with the continuous updating estimator of Hansen et al. (1996).

**Remark 4.6.** (Numerical calculation of the estimates and the specific role of the $\chi^2$-divergence). The computation of $\hat{t}(\theta)$ for fixed $\theta \in \Theta$ as defined in (4.13) is difficult when handling a generic divergence. In the particular case of $\chi^2$-divergence, i.e., when $\varphi(x) = \frac{1}{2}(x - 1)^2$, optimizing on all s.f.m’s, the system (4.13) is linear; we thus easily obtain an explicit form for $\hat{t}(\theta)$, which in turn allows for a single gradient descent when optimizing upon $\Theta$. This procedure is useful in order to compute 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. Moreover, Hjort et al. (2009) extend the EL approach, to more general moment condition models, allowing the number of constraints to increase with growing sample size. In this case, the computation of EL estimate is more complex, and the same idea as above can help to solve the problem.
5. Asymptotic properties of the estimates of the parameter and the divergences

5.1. Asymptotic properties under the model. This Section addresses Problems 1 and 2, aiming to test the null hypothesis \( H_0 : P_0 \in M \) against the alternative \( H_1 : P_0 \notin M \). We derive the limiting distributions of the proposed test statistics which are the estimated divergences between the model \( M \) and \( P_0 \). We also derive the limiting distributions of the estimates of \( \theta_0 \). The following two Theorems 5.1 and 5.2 extend Theorems 3.1 and 3.2 in Newey and Smith (2004) to the context of divergence based approach. The Assumptions which we will consider match those of Theorems 3.1 and 3.2 in Newey and Smith (2004).

Assumption 1. a) \( P_0 \in M \) and \( \theta_0 \in \Theta \) is the unique solution of \( \mathbb{E}[g(X,\theta)] = 0 \); 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 \( \Omega := \mathbb{E}[g(X,\theta_0)g(X,\theta_0)^\top] \) is nonsingular.

Theorem 5.1. Under Assumption 1, with probability approaching one as \( n \to \infty \), the estimate \( \hat{\theta}_\phi \) exists, and converges to \( \theta_0 \) in probability. 
\[
\frac{1}{n} \sum_{i=1}^n g(X_i, \hat{\theta}_\phi) = O_P(1/\sqrt{n}),
\]
\[
\hat{t}(\hat{\theta}_\phi) := \arg \sup_{t \in \Lambda_{\hat{\theta}_\phi}^{(n)}} P_n m(\hat{\theta}_\phi, t) \text{ exists and belongs to } int(\Lambda_{\hat{\theta}_\phi}^{(n)}) \text{ with probability approaching one as } n \to \infty, \text{ and } \hat{t}(\hat{\theta}_\phi) = O_P(1/\sqrt{n}).
\]

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 \( N_{\theta_0} \) of \( \theta_0 \), and \( \mathbb{E}\left[\sup_{\theta \in N_{\theta_0}} ||\partial g(X,\theta)/\partial \theta||\right] < \infty \); c) \( \text{rank}(G) = d \).

Theorem 5.2. Assume that Assumptions 1 and 2 hold. Then,

1) \( \sqrt{n} \left( \hat{\theta}_\phi - \theta_0 \right) \) converges in distribution to a centered normal random vector with covariance matrix 
\[
V := \left[G\Omega^{-1}G^\top\right]^{-1}.
\]
2) If \( l > d \), then the statistic \( 2n \hat{D}_\varphi(M, P_0) \) converges in distribution to a \( \chi^2 \) random variable with \((l - d)\) degrees of freedom.

**Remark 5.1.** Observe that the estimates \( \hat{\theta}_\varphi \) are asymptotically equivalent in term of efficiency, in the sense that the limiting variance is the same as that of EL case, and it does not depend on the choice of the divergence.

**Remark 5.2.** The above Theorem allows to perform statistical tests (of the model) with asymptotic level \( \alpha \in ]0, 1[ \). Consider the null hypothesis

\[
H_0 : P_0 \in \mathcal{M} \quad \text{against the alternative} \quad H_1 : P_0 \notin \mathcal{M}. \tag{5.1}
\]

The critical region is then

\[
C_\varphi := \left\{ 2n \hat{D}_\varphi(M, P_0) > q_{(1-\alpha)} \right\}
\]

where \( q_{(1-\alpha)} \) is the \((1 - \alpha)\)-quantile of the \( \chi^2(l - d) \) distribution. When \( \varphi(x) = -\log x + x - 1 \), it is straightforward to see that 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 problem of the null hypothesis \( H_0 : P_0 \in \mathcal{M}_\theta \) against two different families of alternative hypotheses: \( H_1 : P_0 \notin \mathcal{M}_\theta \) and \( H'_1 : P_0 \in \mathcal{M} \setminus \mathcal{M}_\theta \). Those two tests address different situations since \( H_1 \) may include misspecification of the model. We give two different test statistics each pertaining to one of the situations and derive their limiting distributions both under \( H_0 \) and under the alternatives. As a by product, we also derive confidence areas for the true value \( \theta_0 \) of the parameter. We will first state the convergence in probability of \( \hat{D}_\varphi(\mathcal{M}_\theta, P_0) \) to \( D_\varphi(\mathcal{M}_\theta, P_0) \), and then we obtain the limiting distribution of \( \hat{D}_\varphi(\mathcal{M}_\theta, P_0) \) 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 \( \theta_0 = \theta \) since the true value \( \theta_0 \) of the parameter is assumed to be unique.

**Assumption 3.** a) \( P_0 \in \mathcal{M}_\theta \) and \( \theta \) is the unique solution of \( \mathbb{E}[g(X, \theta)] = 0 \); b) \( \mathbb{E}[\|g(X, \theta)\|^\alpha] < \infty \) for some \( \alpha > 2 \); c) the matrix \( \Omega := \mathbb{E}[g(X, \theta)g(X, \theta)^\top] \) is nonsingular.
Theorem 5.3. Under Assumption 3, we have

1) \( \hat{t}(\theta) := \text{arg sup}_{t \in \Lambda_\theta} P_n m(\theta, t) \) exists and belongs to \( \text{int}(\Lambda_\theta^{(n)}) \) with probability approaching one as \( n \to \infty \), and \( \hat{t}(\theta) = O_P(1/\sqrt{n}) \).

2) The statistic \( 2n \hat{D}_\varphi(M_\theta, P_0) \) converges in distribution to a \( \chi^2(l) \) random variable.

Remark 5.3. Assumption 4.c above ensures the strict concavity of the function \( \mathcal{H}_1 : P_0 \not\in \mathcal{M}_\theta \), including misspecification, the following Assumption is needed.

Assumption 4. a) \( P_0 \not\in \mathcal{M}_\theta \), and \( t^*(\theta) := \text{arg sup}_{t \in \Lambda_\theta} \mathbb{E} [m(X, \theta, t)] \) exists and is an interior point of \( \Lambda_\theta \); b) \( \mathbb{E} \left[ \sup_{t \in N_{t^*(\theta)}} |m(X, \theta, t)| \right] < \infty \) for some compact set \( N_{t^*(\theta)} \subset \Lambda_\theta \) such that \( t^*(\theta) \in \text{int}(N_{t^*(\theta)}) \); c) the functions \( 1_{\mathbb{R}^m}, g_1, \ldots, g_l \) are linearly independent in the sense that: \( P_0 \left\{ x \in \mathbb{R}^m \mid t_0 + \sum_{j=1}^l t_j g_j(x, \theta) \neq 0 \right\} > 0 \) for all \( t \in \mathbb{R}^{1+l} \) with \( t \neq 0 \).

Theorem 5.4. Under Assumption 4, when \( P_0 \not\in \mathcal{M}_\theta \), we have

1) \( \hat{t}(\theta) \) converges in probability to \( t^*(\theta) \).

2) \( \hat{D}_\varphi(M_\theta, P_0) \) converges in probability to \( D_\varphi(M_\theta, P_0) \).

We now give the limiting distribution of the test statistic under \( \mathcal{H}_1 \). We need the following additional condition.

Assumption 5. a) There exists \( N_{t^*(\theta)} \subset \Lambda_\theta \), some compact neighborhood of \( t^*(\theta) \), such that

\[
\mathbb{E} \left[ \sup_{t \in N_{t^*(\theta)}} \| \partial m(X, \theta, t^*(\theta))/\partial t \| \right] < \infty, \quad \mathbb{E} \left[ \sup_{t \in N_{t^*(\theta)}} \| \partial^2 m(X, \theta, t^*(\theta))/\partial t^2 \| \right] < \infty;
\]

b) as \( \delta \to 0 \),

\[
\mathbb{E} \left\{ \sup_{\{t : \|t - t^*(\theta)\| \leq \delta\}} \| \partial^2 m(X, \theta, t)/\partial t^2 - \partial^2 m(X, \theta, t^*(\theta))/\partial t^2 \| \right\} \to 0;
\]
Remark 5.4. Assumption 5.b is used here to relax the condition on the third derivatives (in $t$) of the function $t \mapsto m(X, \theta, t)$.

Theorem 5.5. Under Assumptions 4 and 5, we have

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

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

2) $\sqrt{n}(\hat{D}_\varphi(M_\theta, P_0) - D_\varphi(M_\theta, P_0))$ converges in distribution to a centered normal random variable with variance

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

Remark 5.5. Let $\theta$ be a given value in $\Theta$. Consider the test of the null hypothesis

$$\mathcal{H}_0 : P_0 \in M_\theta \quad \text{against} \quad \mathcal{H}_1 : P_0 /\in M_\theta.$$  \hspace{1cm} (5.2)

In view of Theorem 5.3 part 2, we reject $\mathcal{H}_0$ against $\mathcal{H}_1$, at asymptotic level $\alpha \in [0, 1[$, when $2n\hat{D}_\varphi(M_\theta, P_0)$ 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

$$P_0 \notin M_\theta \mapsto \beta(P_0) := P_0 \left[2n\hat{D}_\varphi(M_\theta, P_0) > q(1-\alpha)\right].$$

We obtain then the following approximation

$$\beta(P_0) \approx 1 - F_N \left( \frac{\sqrt{n}}{\sigma(\theta, t^*(\theta))} \left[ \frac{q(1-\alpha)}{2n} - D_\varphi(M_\theta, P_0) \right] \right),$$  \hspace{1cm} (5.3)

where $F_N$ is the cumulative distribution function of the standard normal distribution. From this approximation, we can give the approximate sample size that ensures a desired power $\beta$ for a given alternative $P_0 \notin M_\theta$. Let $n_0$ be the positive root of the equation

$$\beta = 1 - F_N \left( \frac{\sqrt{n}}{\sigma(\theta, t^*(\theta))} \left( \frac{q(1-\alpha)}{2n} - D_\varphi(M_\theta, P_0) \right) \right).$$
i.e.,
\[ n_0 = \frac{(a + b) - \sqrt{a(a + 2b)}}{2D\varphi(\mathcal{M}_\theta, P_0)^2} \]
with \( a := \sigma^2(\theta, t^*(\theta)) \left[ F_N^{-1}(1 - \beta) \right]^2 \) and \( b := q_{(1-\alpha)}D\varphi(\mathcal{M}_\theta, P_0) \). The required sample size is then \( \lceil n_0 \rceil + 1 \), where \( \lceil n_0 \rceil \) denotes the integer part of \( n_0 \).

**Remark 5.6. (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
\[ 2nS_n^\varphi := 2n \left[ \hat{D}\varphi(\mathcal{M}_\theta, P_0) - \inf_{\theta \in \Theta} \hat{D}\varphi(\mathcal{M}_\theta, P_0) \right], \tag{5.4} \]
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 \( \alpha \) when \( 2nS_n^\varphi > q_{(1-\alpha)} \), 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
\[ \sqrt{n} \left( S_n^\varphi - D\varphi(\mathcal{M}_\theta, P_0) \right) \tag{5.5} \]
converges to a centered normal random variable with variance
\[ \sigma^2(\theta, t^*(\theta)) := \mathbb{E}(m(X, \theta, t^*(\theta))^2) - (\mathbb{E}m(X, \theta, t^*(\theta)))^2. \]

So, as in the above Remark, we obtain the following approximation
\[ \beta(P_0) \approx 1 - F_N \left( \frac{\sqrt{n}}{\sigma(\theta, t^*(\theta))} \left[ \frac{q_{1-\alpha}}{2n} - D\varphi(\mathcal{M}_\theta, P_0) \right] \right) \tag{5.6} \]
to the power function \( P_0 \in \mathcal{M}/\mathcal{M}_\theta \mapsto \beta(P_0) := P_0 \left[ 2nS_n^\varphi > q_{(1-\alpha)} \right]. \) The approximated sample size required to achieve a desired power for a given alternative can be obtained in a similar way.

**Remark 5.7. (Confidence region for the parameter).** For a fixed level \( \alpha \in ]0, 1[ \), using convergence (5.4), the set
\[ \{ \theta \in \Theta \text{ such that } 2nS_n^\varphi \leq q_{(1-\alpha)} \} \]
is an asymptotic confidence region for $\theta_0$ where $q_{(1-\alpha)}$ is the $(1-\alpha)$-quantile of the $\chi^2(d)$-distribution. It is straightforward to see that the confidence region obtained for the $KL_m$-divergence coincides with that of Owen (1991) and Qin and Lawless (1994).

5.3. Asymptotic properties under misspecification. We address Problem 1 stating the limiting distribution of the proposed test statistics under the alternative $H_1: P_0 \notin \mathcal{M}$. 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 $\theta^*$ as well as the existence of the projection $Q^*_{\theta^*}$ of $P_0$ on $\mathcal{M}$, and states some necessary other regularity conditions. Proposition 4.2 above states the existence and characterization of the projection $Q^*_{\theta^*}$ of $P_0$ on $\mathcal{M}_{\theta}$, for a given $\theta \in \Theta$.

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

**Remark 5.8.** Assumption 6.d ensures the strict concavity of the function $t \in \Lambda_{\theta} \mapsto E[m(X, \theta, t)]$ on the convex set $\Lambda_{\theta}$, which implies the uniqueness of $t^*(\theta)$, for all $\theta \in \Theta$. This Assumption can be replaced by the following one: for all $\theta \in \Theta$, there exists a neighborhood $N_{t^*}(\theta)$ of $t^*(\theta)$ such that

$$E\left[ \sup_{t \in N_{t^*}(\theta)} \| \partial m(X, \theta, t) / \partial t \| \right] < \infty, \quad E\left[ \sup_{t \in N_{t^*}(\theta)} \| \partial^2 m(X, \theta, t) / \partial t^2 \| \right] < \infty$$

and the matrix $E[\partial^2 m(X, \theta, t^*(\theta)) / \partial t^2] < \infty$ is nonsingular, which implies the uniqueness of $t^*(\theta)$, for all $\theta \in \Theta$.

**Theorem 5.6.** Under Assumption 6, we have

1) $\| \hat{t}(\theta) - t^*(\theta) \|$ converges in probability to 0 uniformly in $\theta \in \Theta$.
2) $\hat{\theta}_\phi$ converges in probability to $\theta^*$;
3) $\hat{D}_\phi(\mathcal{M}, P_0)$ 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) $\theta^* \in \text{int}(\Theta)$; b) there exists $N \subset \Theta \times \Lambda_\Theta$, some compact neighborhood of $(\theta^*, t^*(\theta^*))$, such that with probability one $(\theta, t) \in N \mapsto m(X, \theta, t)$ is $C^2$ and

$$\mathbb{E}\left[ \sup_{(\theta, t) \in N} \left\| \frac{\partial m(X, \theta, t)}{\partial (\theta, t)} \right\| \right] < \infty, \quad \mathbb{E}\left[ \sup_{(\theta, t) \in N} \left\| \frac{\partial^2 m(X, \theta, t)}{\partial (\theta, t)^2} \right\| \right] < \infty;
$$

(c) as $\delta \to 0$,

$$\mathbb{E}\left\{ \sup_{\{(t, \theta) : \|t - (t^*(\theta^*), \theta^*)\| \leq \delta\}} \left\| \frac{\partial^2 m(X, \theta, t)}{\partial (\theta, t)^2} - \frac{\partial^2 m(X, \theta^*, t^*(\theta^*))}{\partial (\theta, t)^2} \right\| \right\} \to 0;
$$

d) $\mathbb{E}[m(X, \theta^*, t^*(\theta^*))^2], \mathbb{E}\left[ \left\| \frac{\partial m(X, \theta^*, t^*(\theta^*))}{\partial t} \right\|^2 \right]$ and $\mathbb{E}\left[ \left\| \frac{\partial m(X, \theta^*, t^*(\theta^*))}{\partial \theta} \right\|^2 \right]$ are finite, and the matrix

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

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

**Remark 5.9.** Assumption 7.c is used here to relax the condition on the third derivatives (in $t$ and $\theta$) of the function $(\theta, t) \mapsto m(X, \theta, t)$.

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

1) $$\sqrt{n} \begin{pmatrix} \hat{t}(\varphi) - t^*(\theta^*) \\ \hat{\varphi} - \theta^* \end{pmatrix}$$

converges in distribution to a centered normal random vector with covariance matrix

$$W := S^{-1} M S^{-1},$$

where $M := \mathbb{E}\left[ \begin{bmatrix} \frac{\partial m}{\partial \theta} m(X, \theta^*, t^*(\theta^*)) \\ \frac{\partial^2 m}{\partial \theta^2} m(X, \theta^*, t^*(\theta^*)) \end{bmatrix} \begin{bmatrix} \frac{\partial m}{\partial \theta} m(X, \theta^*, t^*(\theta^*)) \\ \frac{\partial^2 m}{\partial \theta^2} m(X, \theta^*, t^*(\theta^*)) \end{bmatrix}^\top \right]$. 
2) $\sqrt{n} \left( \hat{D}_\varphi(\mathcal{M}, P_0) - D_\varphi(\mathcal{M}, P_0) \right)$ converges in distribution to a centered normal random variable with variance

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

**Remark 5.10.** In the case of EL, i.e., when $\varphi(x) = -\log x + x - 1$, Assumption 6.c implies that

$$
-\infty < \inf_{x \in \mathbb{R}^m} t_0 + \sum_{i=1}^l t_i g_i(x, \theta) \leq \sup_{x \in \mathbb{R}^m} t_0 + \sum_{i=1}^l t_i g_i(x, \theta) < 1 \quad (5.7)
$$

for all $x \in \mathbb{R}^m - P_0$-a.s., for all $\theta \in \Theta$ and for all $t \in N_t(\theta)$. This imposes a restriction on the model when the support of $P_0$ and at least one function $g_i$ is unbounded. Indeed, when the support of $P_0$ is for example the whole space $\mathbb{R}^m$, the condition above does not hold when at least one function $g_i$ is unbounded. In this case, the EL estimate may cease to be consistent as it is stated by Schennach (2007) under misspecification. This is a potential problem for all divergences associated to $\varphi$-functions with domain of the form $(a, +\infty[, ] - \infty, b)$ or $(a, b)$, where $a$ and $b$ are some finite real numbers; it is the case of modified $\chi^2$, Hellinger, KL and modified KL divergences. At the contrary, Assumption 6.c may be satisfied for other divergences associated to $\varphi$ functions with dom$\varphi = \mathbb{R}$ which is the case of $\chi^2$ divergence for example.

**Remark 5.11.** Theorem 5.7 part 2 is useful for the computation of the power function. For testing the null hypothesis $\mathcal{H}_0 : P_0 \in \mathcal{M}$ against the alternative $\mathcal{H}_1 : P_0 \notin \mathcal{M}$, the power function is

$$
P_0 \notin \mathcal{M} \mapsto \beta(P_0) := P_0 \left[ 2n \hat{D}_\varphi (\mathcal{M}, P_0) > q_{(1-\alpha)} \right]. \quad (5.8)
$$

Using Theorem 5.7 part 2, we obtain the following approximation to the power function (5.8):

$$
\beta(P_0) \approx 1 - F_N \left[ \frac{\sqrt{n}}{\sigma (\theta^*, t^*(\theta^*))} \left( \frac{q_{(1-\alpha)}}{2n} - D_\varphi (\mathcal{M}, P_0) \right) \right] \quad (5.9)
$$
where $F_N$ 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 $\beta$ for a given alternative $P_0 \notin M$ can be obtained in similar way as above.

6. Simulation results: Approximation of the power functions of the empirical likelihood ratio and the empirical chi-square tests

We will illustrate by simulation the accuracy of the power approximation (5.9) in the case of EL method (or $KL_m$-divergence), i.e., when $\varphi(x) = -\log x + x - 1$, and for the $\chi^2$-divergence, i.e., when $\varphi(x) = \frac{1}{2}(x - 1)^2$. We will consider two different models from Examples 1.1 and 1.2.

**Example 6.1.** Consider the model test (see Example 1.1) of the null hypothesis

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

where $\mathcal{M} := \bigcup_{\theta \in \Theta} \mathcal{M}_\theta$ 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 $g(x, \theta) := (x, x^2 - \theta)^\top$, namely

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

where $\theta \in \Theta := \mathbb{R}$ is the parameter of interest to be estimated. 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 $\epsilon$ 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 that all Assumptions of Theorem 5.2 are satisfied when $\epsilon = 0$, and all Assumptions of Theorem 5.7 are met under alternatives for both $KL_m$ and $\chi^2$ divergences. The power function (5.8) is plotted (with a continuous line), with sample sizes $n = 50$, $n = 100$, $n = 200$ and $n = 500$, for different values of $\epsilon \in [0, 1]$. Each power entry was computed by Monte-Carlo from 1000 independent runs. The approximation (5.9) is plotted (with a dashed line) as a function of $\epsilon$; both $\sigma(\hat{\theta}^*, \hat{t}^*(\hat{\theta}^*))$ and $D_\varphi(M, P_0)$ in (5.9) are computed through their empirical counterparts $\hat{\sigma}(\hat{\theta}_\varphi, \hat{t}(\hat{\theta}_\varphi))$ and $\hat{D}_\varphi(M, P_0)$ respectively. The estimates $\hat{\theta}_\varphi$ and $\hat{D}_\varphi(M, P_0)$ are computed using the Newton-Raphson algorithm for $\chi^2$-divergence and the Uzawa algorithm for
KL\textsubscript{m}-divergence. The results are presented in Figure 1 for the EL ratio test and in Figure 2 for the empirical $\chi^2$ test. We observe that the approximation is accurate even for moderate sample sizes for both EL ratio and $\chi^2$ tests, but for small samples sizes ($n = 50$) the approximation is slightly better for the EL ratio test. However, the computation is faster for the $\chi^2$ divergence since the optimization on the parameter $t := (t_0, t_1, t_2)^\top$ is obtained explicitly by solving the system (4.13) which is linear only for the $\chi^2$ divergence. At the contrary, the EL case needs a double gradient descent both on $\theta$ and $t$ (using Uzawa algorithm).

**Figure 1.** Example 6.1 : Approximation of the power function of the EL ratio test ($KL_m$)
Example 6.2. Consider the model test (see Example 1.2) of the null hypothesis

\[ \mathcal{H}_0 : P_0 \in \mathcal{M} \] against the alternative \[ \mathcal{H}_1 : P_0 \notin \mathcal{M}, \]

where \( \mathcal{M} := \bigcup_{\theta \in \Theta} \mathcal{M}_\theta \) 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 \( g(x, \theta) := (\mathbb{1}_{-\infty, 1/2}(x) - \theta, \mathbb{1}_{-1/2, +\infty}(x) - \theta)^\top \), namely

\[ \mathcal{M}_\theta := \left\{ Q \in M \text{ such that } \int \! dQ(x) = 1 \text{ and } \int \! g(x, \theta) \, dQ(x) = 0 \right\}, \]

where \( \theta := F_0(1/2) \in \Theta := [0, 1] \) is the parameter of interest to be estimated. We aim to test the equality \( F_0(1/2) = 1 - F_0(-1/2) \) and estimate the value \( \theta = F_0(1/2) \).
using this symmetry. We consider the asymptotic level $\alpha = 0.05$ and the alternatives $P_0 := U([-1, 1 + \epsilon]) \not\in \mathcal{M}$ for different values of $\epsilon$ in the interval $[0, 1]$. Note that when $\epsilon = 0$ then the uniform distribution $P_0 := U([-1, 1])$ belongs to the model $\mathcal{M}$, since it is symmetric about zero. 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 for both $KL_m$ and $\chi^2$ divergences. The power function (5.8) is plotted (with a continuous line), with sample sizes $n = 50, n = 100, n = 200$ and $n = 500$, for different values of $\epsilon \in [0, 1]$. Each power entry was computed by Monte-Carlo from 1000 independent runs. The approximation (5.9) is plotted (with a dashed line) as a function of $\epsilon$, and it was computed in a similar way as in Example 6.1. The results for EL ratio test are presented in Figure 3, and in Figure 4 for empirical $\chi^2$ test. We obtain similar results as in Example 6.1 for both EL ratio and empirical $\chi^2$ tests.

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. The use of duality leads to the obtaining of the limiting distributions of the test statistics and the estimates of the parameter under alternatives and under misspecification. An approximation to the power function is deduced, as well as the sample size which ensures a desired power for a given alternative, for all the proposed empirical divergence tests including the EL ratio one. 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. The proposed estimates of the parameters are asymptotically equivalent in term of efficiency at first order; 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 second order efficiency and/or robustness. Preliminary simulation results show that Hellinger divergence enjoys good properties in terms of efficiency-robustness; see Broniatowski and Keziou (2008).
Figure 3. Example 6.2: Approximation of the power function of the EL ratio test ($KL_m$)

Comparisons of the robustness of the estimates and tests in term of influence function can be handled in a similar way as the case of parametric models; see Toma and Leoni-Aubin (2010). Also comparisons of the tests 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 $(\theta, \lambda) \in \Theta \times \mathbb{R}^l \mapsto \frac{1}{n} \sum_{i=1}^{n} \rho(\lambda^\top g(X, \theta))$ is replaced by our
Figure 4. Example 6.2: Approximation of the power function of the empirical $\chi^2$ test.

The function $(\theta, t) \in \Theta \times \mathbb{R}^{1+l} \mapsto \frac{1}{n} \sum_{i=1}^{n} m(t \top \bar{g}(X, \theta))$. In particular, we have

$$\max_{1 \leq i \leq n} |\hat{t}(\hat{\theta}_\varphi) \top \bar{g}(X_i, \hat{\theta}_\varphi)| \to 0$$

in probability, which implies that $\hat{t}(\hat{\theta}_\varphi) \in \text{int}(\Lambda_{\hat{\theta}_\varphi}^{(n)})$ 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.
Proof of Theorem 5.3.
It is a particular case of Theorem 5.1 taking $\Theta = \{\theta\}$.

Proof of Theorem 5.4.
1) First, note that $t^*(\theta)$ exists and is unique by Assumption 4. By the uniform weak law of large numbers (UWLLN), using continuity of $m(X, \theta, t)$ in $t$, and Assumption 4.b, we obtain
\[ |P_n m(\theta, t) - \mathbb{E}[m(X, \theta, t)]| \rightarrow 0, \] (8.1)
in probability uniformly in $t$ over the compact set $N_{t^*(\theta)}$. Using this and the fact that $t^*(\theta) := \arg\sup_{t \in \Lambda_\theta} P_0 m(\theta, t)$ is unique and belongs to $\text{int}(N_{t^*(\theta)})$ and the strict concavity of $t \mapsto P_0 m(\theta, t)$, we conclude that any value
\[ \overline{t} := \arg\sup_{t \in N_{t^*(\theta)}} P_n m(\theta, t) \] (8.2)
converges in probability to $t^*(\theta)$; see e.g. Theorem 5.7 in van der Vaart (1998). We end then the proof by showing that $\widehat{\theta}(\theta)$ belongs to $\text{int}(N_{t^*(\theta)})$ with probability one as $n \rightarrow \infty$, and therefore it converges to $t^*(\theta)$. In fact, since for $n$ sufficiently large any value $\overline{t}$ lies in the interior of $N_{t^*(\theta)}$, concavity of $t \mapsto P_n m(\theta, t)$ implies that no other point $t$ in the complement of $\text{int}(N_{t^*(\theta)})$ can maximize $P_n m(\theta, t)$ over $t \in \mathbb{R}^{1+l}$, hence $\widehat{\theta}(\theta)$ must belongs to $\text{int}(N_{t^*(\theta)})$.
2) With probability tending to 1 as $n \rightarrow \infty$, we have $\widehat{D}_\varphi(M_\theta, P_0) = P_n m(\theta, \overline{t}) = P_n m(\theta, \overline{\theta})$. We can then write
\[ \left| \widehat{D}_\varphi(M_\theta, P_0) - D_\varphi(M_\theta, P_0) \right| = \left| P_n m(\theta, \overline{t}) - P_0 m(\theta, t^*(\theta)) \right| =: |A|, \]
and
\[ P_n m(\theta, t^*(\theta)) - P_0 m(\theta, t^*(\theta)) \leq A \leq P_n m(\theta, \overline{t}) - P_0 m(\theta, \overline{t}). \]
Both the RHS and the LHS in the above display tend to 0 in probability by (8.1). Hence,
\[ \left| \widehat{D}_\varphi(M_\theta, P_0) - D_\varphi(M_\theta, P_0) \right| \] tends to 0 in probability as $n \rightarrow \infty$. This ends the proof.
Proof of Theorem 5.5.
1) For $n$ sufficiently large, by a Taylor expansion, there exists $\tilde{t} \in \mathbb{R}^{1+l}$ inside the segment that links $\hat{t}$ and $t^*(\theta)$ with

$$0 = P_n m'(\theta, \hat{t}) = P_n m'(\theta, t^*(\theta)) + (P_n m''(\theta, \tilde{t}))^\top (\hat{t} - t^*(\theta)).$$

(8.3)

By Assumptions 5.a and 5.b, using the fact that $\tilde{t} = t^*(\theta) + o_P(1)$ and the UWLLN, we can prove that

$$P_n m''(\theta, \tilde{t}) = P_0 m''(\theta, t^*(\theta)) + o_P(1).$$

Using this display, one gets from (8.3)

$$-P_n m'(\theta, t^*(\theta)) = (P_0 m''(\theta, t^*(\theta)) + o_P(1)) (\hat{t} - t^*(\theta)).$$

(8.4)

Assumptions 4.a and 5.a imply that $P_0 m'(\theta, t^*(\theta)) = 0$. By the central limit theorem (CLT), we have

$$\sqrt{n} P_n m'(\theta, t^*(\theta)) = O_P(1),$$

which by (8.4) implies that $\sqrt{n} (\hat{t} - t^*(\theta)) = O_P(1)$. From (8.4), we get then

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

(8.5)

The CL and Slutsky theorems conclude the proof of part 1.

2) Using the fact that $(\hat{t} - t^*(\theta)) = O_P(1/\sqrt{n})$ and $P_n m'(\theta, t^*(\theta)) = P_0 m'(\theta, t^*(\theta)) + o_P(1) = 0 + o_P(1) = o_P(1)$, we obtain

$$\sqrt{n} \left( \hat{D}_\varphi(M_\theta, P_0) - D_\varphi(M_\theta, P_0) \right) = \sqrt{n} \left( P_n m(\theta, \hat{t}) - P_0 m(\theta, t^*(\theta)) \right)$$

$$= \sqrt{n} (P_n m(\theta, t^*(\theta)) - P_0 m(\theta, t^*(\theta))) + o_P(1),$$

and the CL and Slutsky theorems conclude the proof.

Proof of Theorem 5.6.
1) First note that Assumption 6.d implies that the function $t \in \Lambda_\theta \mapsto \mathbb{E} m(X, \theta, t)$ is strictly concave for all $\theta \in \Theta$, which implies that $t^*(\theta)$ is unique for all $\theta \in \Theta$. By the
UWLLN, using continuity of \( m(X, \theta, t) \), in \( \theta \) and \( t \), and Assumption 6.c, we obtain the uniform convergence in probability, over the compact set \( \{ (\theta, t) \in \Theta \times \mathbb{R}^{1+t}; \ \theta \in \Theta, t \in N_{t^*} (\theta) \} \),

\[
sup_{\{\theta \in \Theta, t \in N_{t^*} (\theta)\}} |P_n m(\theta, t) - P_0 m(\theta, t)| \to 0. \tag{8.6}
\]

We can then prove the convergence in probability \( \sup_{\theta \in \Theta} \| \hat{t}(\theta) - t^*(\theta) \| \to 0 \) in two steps.

Step 1: Let \( \eta > 0 \). We will show that \( P_0 \left[ \sup_{\theta \in \Theta} \| \hat{t}(\theta) - t^*(\theta) \| \geq \eta \right] \to 0 \) for any value \( \eta > 0 \). Let \( \eta > 0 \) such that \( \sup_{\theta \in \Theta} \| \hat{t}(\theta) - t^*(\theta) \| \geq \eta \). Since \( \Theta \) is a compact set, by continuity there exists \( \bar{\theta} \in \Theta \) such that \( \sup_{\theta \in \Theta} \| \hat{t}(\theta) - t^*(\theta) \| = \| \hat{t}(\bar{\theta}) - t^*(\bar{\theta}) \| \geq \eta \). Hence, there exists \( \varepsilon > 0 \) such that \( P_0 m(\bar{\theta}, t^*(\bar{\theta})) - P_0 m(\bar{\theta}, \hat{t}(\bar{\theta})) > \varepsilon \).

In fact, \( \varepsilon \) may be defined as follows

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

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 \( t^*(\theta) \in \text{int}(N_{t^*}(\theta)) \) and the fact that \( \Theta \) is compact. Hence the event \( \left[ \sup_{\theta \in \Theta} \| \hat{t}(\theta) - t^*(\theta) \| \geq \eta \right] \) implies the event

\[
\left[ P_0 m(\bar{\theta}, t^*(\bar{\theta})) - P_0 m(\bar{\theta}, \hat{t}(\bar{\theta})) \geq \varepsilon \right],
\]

from which we obtain

\[
P_0 \left[ \sup_{\theta \in \Theta} \| \hat{t}(\theta) - t^*(\theta) \| \geq \eta \right] \leq P_0 \left[ P_0 m(\bar{\theta}, t^*(\bar{\theta})) - P_0 m(\bar{\theta}, \hat{t}(\bar{\theta})) \geq \varepsilon \right]. \tag{8.8}
\]

On the other hand, by (8.6), we have

\[
P_0 m(\bar{\theta}, t^*(\bar{\theta})) - P_0 m(\bar{\theta}, \hat{t}(\bar{\theta})) = P_n m(\bar{\theta}, t^*(\bar{\theta})) - P_0 m(\bar{\theta}, \hat{t}(\bar{\theta})) + o_P(1)
\]

\[
\leq P_n m(\bar{\theta}, \hat{t}(\bar{\theta})) - P_0 m(\bar{\theta}, \hat{t}(\bar{\theta})) + o_P(1)
\]

\[
\leq \sup_{\{\theta \in \Theta, t \in N_{t^*} (\theta)\}} |P_n m(\theta, t) - P_0 m(\theta, t)| + o_P(1).
\]
Combining this with (8.8) and (8.6), we conclude that

$$\sup_{\theta \in \Theta} \| \bar{t}(\theta) - t^*(\theta) \| \to 0$$

(8.9)

in probability. In particular, $\bar{t}(\theta) \in \text{int}(N_{t^*(\theta)})$ for sufficiently large $n$, uniformly in $\theta \in \Theta$. Since $t \mapsto P_n m(\theta, t)$ is concave, then the maximizer $\hat{t}(\theta)$ belongs to $\text{int}(N_{t^*(\theta)})$ for sufficiently large $n$; hence the same result (8.9) holds when $\bar{t}(\theta)$ is replaced by $\hat{t}(\theta)$.

2) From part 1, we have for large $n$,

$$\sup_{\theta \in \Theta} |P_n m(\theta, \bar{t}(\theta)) - P_0 m(\theta, t^*(\theta))| = \sup_{\theta \in \Theta} |P_n m(\theta, \bar{t}(\theta)) - P_0 m(\theta, \bar{t}(\theta))| =: |B|.$$  

On the other hand, we have

$$P_n m(\theta, t^*(\theta)) - P_0 m(\theta, t^*(\theta)) \leq B \leq P_n m(\theta, \bar{t}(\theta)) - P_0 m(\theta, \bar{t}(\theta)).$$

By Assumption 6.c, and the convergence in probability $\sup_{\theta \in \Theta} \| \bar{t}(\theta) - t^*(\theta) \| \to 0$, both the RHS and LHS of the above display tends to 0 in probability uniformly in $\theta \in \Theta$, by the UWLLN. Hence, $\sup_{\theta \in \Theta} |P_n m(\theta, \bar{t}(\theta)) - P_0 m(\theta, t^*(\theta))| \to 0$ in probability. Now, since the minimizer $\theta^*$ of $\theta \mapsto P_0 m(\theta, t^*(\theta))$ over the compact set $\Theta$ is unique and interior point of $\Theta$, by continuity and the above uniform convergence, we conclude that $\hat{\theta}_\varphi$ tends in probability to $\theta^*$; see e.g. Theorem 5.7 in *van der Vaart (1998)*.

3) This holds as a consequence of the uniform convergence in probability

$$\sup_{\theta \in \Theta} |P_n m(\theta, \hat{t}(\theta)) - P_0 m(\theta, t^*(\theta))| \to 0$$

(8.10)

proved in part 2 above. In fact, we have for $n$ sufficiently large

$$|\hat{D}_\varphi(\mathcal{M}, P_0) - D_\varphi(\mathcal{M}, P_0)| = |P_n m(\hat{\theta}, \hat{t}(\hat{\theta})) - P_0 m(\theta^*, t^*(\theta^*))| =: |C|,$$

with

$$P_n m(\hat{\theta}, \hat{t}(\hat{\theta})) - P_0 m(\hat{\theta}, \hat{t}(\hat{\theta})) \leq C \leq P_n m(\theta^*, \hat{t}(\theta^*)) - P_0 m(\theta^*, t^*(\theta^*))$$

and both the RHS and LHS tend to 0 in probability by (8.10). This concludes the proof.
Proof of Theorem 5.7.

1) By the first order conditions, with probability tending to one, we have

\[
\begin{align*}
P_n \frac{\partial}{\partial t} m\left( \hat{\theta}, \hat{t}\left(\hat{\theta}\right) \right) &= 0 \\
\left( P_n \frac{\partial}{\partial \theta} m\left( \hat{\theta}, \hat{t}\left(\hat{\theta}\right) \right) + P_n \frac{\partial}{\partial \theta} m\left( \hat{\theta}, \hat{t}\left(\hat{\theta}\right) \right) \frac{\partial}{\partial \theta} \left( \hat{t}\left(\hat{\theta}\right) \right) \right) &= 0.
\end{align*}
\]

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

\[
\begin{align*}
P_n \frac{\partial}{\partial t} m\left( \hat{\theta}, \hat{t}\left(\hat{\theta}\right) \right) &= 0 \quad (8.11) \\
P_n \frac{\partial}{\partial \theta} m\left( \hat{\theta}, \hat{t}\left(\hat{\theta}\right) \right) &= 0. \quad (8.12)
\end{align*}
\]

Using a Taylor expansion in (8.11) in \( (\hat{\theta}, \hat{t}) \) around \( (\theta^*, t^*) \); there exists \( (\theta, t) \) inside the segment that links \( (\hat{\theta}, \hat{t}\left(\hat{\theta}\right)) \) and \( (\theta^*, t^*(\theta^*)) \) such that

\[
0 = P_n \frac{\partial}{\partial t} m(\theta^*, t^*(\theta^*)) + \left[ \left( P_0 \frac{\partial^2}{\partial t^2} m(\theta^*, t^*(\theta^*)) \right) \top \left( P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta^*, t^*(\theta^*)) \right) \top \right] a_n.
\]

with

\[
a_n := \left( \left( \hat{t}\left(\hat{\theta}\right) - t^*(\theta^*) \right) \top, \left( \hat{\theta} - \theta^* \right) \top \right) \top.
\]

By Assumption 7, using the UWLLN, we can write

\[
\left[ P_n \frac{\partial^2}{\partial t^2} m(\theta, \bar{e}), P_n \frac{\partial^2}{\partial \theta \partial t} m(\theta, \bar{e}) \right] = \left[ P_0 \frac{\partial^2}{\partial t^2} m(\theta^*, t^*(\theta^*)), P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta^*, t^*(\theta^*)) \right] + o_P(1),
\]

to obtain from (8.13)

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

In the same way, using a Taylor expansion in (8.12), we obtain

\[
- P_n \frac{\partial}{\partial \theta} m(\theta^*, t^*) = \left[ \left( P_0 \frac{\partial^2}{\partial t \partial \theta} m(\theta^*, t^*) \right) \top + o_P(1), \left( P_0 \frac{\partial^2}{\partial \theta^2} m(\theta^*, t^*) \right) \top + o_P(1) \right] a_n.
\]
From (8.15) and (8.16), we get
\[
\sqrt{na_n} = \sqrt{n} \left( \begin{pmatrix} P_0 \frac{\partial^2}{\partial \theta^2} m(\theta^*, t^*) & P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta^*, t^*) \\ P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta^*, t^*) & P_0 \frac{\partial^2}{\partial t^2} m(\theta^*, t^*) \end{pmatrix} \right)^{-1} \times \\
\times \begin{pmatrix} -P_n \frac{\partial}{\partial t} m(\theta^*, t^*) \\ -P_n \frac{\partial}{\partial \theta} m(\theta^*, t^*) \end{pmatrix} + o_P(1).
\]

(8.17)

Denote $S$ the $(1 + l + d) \times (1 + l + d)$ matrix defined by
\[
S := \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} := \begin{pmatrix} P_0 \frac{\partial^2}{\partial \theta^2} m(\theta^*, t^*) & P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta^*, t^*) \\ P_0 \frac{\partial^2}{\partial \theta \partial t} m(\theta^*, t^*) & P_0 \frac{\partial^2}{\partial t^2} m(\theta^*, t^*) \end{pmatrix}.
\]

(8.18)

Hence, we obtain
\[
\sqrt{n} \begin{pmatrix} \hat{t}(\hat{\theta}) - t^*(\theta^*) \\ \hat{\theta} - \theta^* \end{pmatrix} = \sqrt{n} S^{-1} \begin{pmatrix} -P_n \frac{\partial}{\partial t} m(\theta^*, t^*) \\ -P_n \frac{\partial}{\partial \theta} m(\theta^*, t^*) \end{pmatrix} + o_P(1),
\]

and the CL and Slutsky theorems conclude the proof.

2) Using the fact that
\[
\hat{t}(\hat{\theta}) - t^*(\theta^*) = O_P(1/\sqrt{n}), \quad P_n \frac{\partial m(\theta^*, t^*(\theta^*))}{\partial t} = P_0 \frac{\partial m(\theta^*, t^*(\theta^*))}{\partial t} + o_P(1) = o_P(1)
\]

and
\[
\hat{\theta} - \theta^* = O_P(1/\sqrt{n}), \quad P_n \frac{\partial m(\theta^*, t^*(\theta^*))}{\partial \theta} = P_0 \frac{\partial m(\theta^*, t^*(\theta^*))}{\partial \theta} + o_P(1) = o_P(1),
\]

we can write
\[
\sqrt{n} \left( \hat{D}_\varphi(\mathcal{M}, P_0) - D_\varphi(\mathcal{M}, P_0) \right) = \sqrt{n} \left( P_n m(\hat{\theta}, \hat{t}(\hat{\theta})) - P_0 m(\theta^*, t^*(\theta^*)) \right)
\]
\[
= \sqrt{n} \left( P_n m(\theta^*, t^*(\theta^*)) - P_0 m(\theta^*, t^*(\theta^*)) \right) + o_P(1),
\]

and the CL and Slutsky theorems end the proof.

**Acknowledgments.** The authors wish to thank the reviewers for their constructive comments leading to improvement of this paper.
References


---

1*LSTA, Université Pierre et Marie Curie - Paris 6. E-Mail : michel.broniatowski@upmc.fr
2Laboratoire de Mathématiques de Reims, EA 4535, Université de Reims Champagne-Ardenne. E-Mail : amor.keziou@univ-reims.fr