



**HAL**  
open science

# UNIFORM MINIMUM RISK EQUIVARIANT ESTIMATES FOR MOMENT CONDITION MODELS

Michel Broniatowski, Jana Jureckova, Amor Keziou

► **To cite this version:**

Michel Broniatowski, Jana Jureckova, Amor Keziou. UNIFORM MINIMUM RISK EQUIVARIANT ESTIMATES FOR MOMENT CONDITION MODELS. 2019. hal-02105790

**HAL Id: hal-02105790**

**<https://hal.science/hal-02105790>**

Preprint submitted on 24 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# UNIFORM MINIMUM RISK EQUIVARIANT ESTIMATES FOR MOMENT CONDITION MODELS

MICHEL BRONIATOWSKI<sup>1</sup>, JANA JUREČKOVÁ<sup>2</sup> AND AMOR KEZIOU<sup>3</sup>

ABSTRACT. We consider semiparametric moment condition models invariant to transformation groups. The parameter of interest is estimated by minimum empirical divergence approach, introduced by Broniatowski and Keziou (2012). It is shown that the minimum empirical divergence estimates, including the empirical likelihood one, are equivariants. The minimum risk equivariant estimate is then identified to be any one of the minimum empirical divergence estimates minus its expectation conditionally to maximal invariant statistic of the considered group of transformations. An asymptotic approximation to the conditional expectation, is obtained, using the result of Jurečková and Picek (2009).

*AMS Subject Classifications:* 62F10, 62F30, 62F99

*Keywords :* Pitman estimator, semiparametric model, equivariant estimator

## 1. INTRODUCTION

The semiparametric moment condition models are defined through estimating equations

$$\mathbb{E}(f_j(\mathbf{X}, \boldsymbol{\theta}_T)) = 0 \text{ for all } j = 1, \dots, \ell,$$

where  $\mathbb{E}(\cdot)$  denotes the mathematical expectation,  $\mathbf{X} \in \mathbb{R}^m$  is a random vector,  $\boldsymbol{\theta}_T \in \Theta \subseteq \mathbb{R}^d$  is the unknown true value of the parameter of interest which is assumed to be unique, and  $\mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) := (f_1(\mathbf{x}, \boldsymbol{\theta}), \dots, f_\ell(\mathbf{x}, \boldsymbol{\theta}))^\top$  is some specified measurable  $\mathbb{R}^\ell$ -valued function defined on  $\mathbb{R}^m \times \Theta$ . Such models are popular in statistics and econometrics, see e.g., Qin and Lawless (1994), Haberman (1984), Sheehy (1987), McCullagh and Nelder (1983), Owen (2001) and the references therein. Denoting  $P_{\mathbf{X}}(\cdot)$  the probability distribution of the random vector  $\mathbf{X}$ , then the above estimating equations can be written as

$$\int_{\mathbb{R}^m} \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}_T) dP_{\mathbf{X}}(\mathbf{x}) = \mathbf{0}.$$

Let  $M^1$  be the collection of all signed finite measures (s.f.m.)  $Q$  on the Borel  $\sigma$ -field  $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$  such that  $Q(\mathbb{R}^m) = 1$ . The submodel  $\mathcal{M}_{\boldsymbol{\theta}}$ , associated to a given value  $\boldsymbol{\theta} \in \Theta$ , consists of all

---

*Date:* April, 2019.

s.f.m.'s  $Q \in M^1$  satisfying  $\ell$  linear constraints induced by the vector valued function  $\mathbf{f}(\cdot, \boldsymbol{\theta}) := (f_1(\cdot, \boldsymbol{\theta}), \dots, f_\ell(\cdot, \boldsymbol{\theta}))^\top$ , namely,

$$\begin{aligned} \mathcal{M}_\theta &:= \left\{ Q \in M^1 \text{ such that } \int_{\mathbb{R}^m} f_j(\mathbf{x}, \boldsymbol{\theta}) dQ(\mathbf{x}) = 0, \forall j = 1, \dots, \ell \right\} \\ &= \left\{ Q \in M^1 \text{ such that } \int_{\mathbb{R}^m} \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) dQ(\mathbf{x}) = \mathbf{0} \right\}, \end{aligned}$$

with  $\ell > d$ . The statistical model which we consider can be written as

$$\mathcal{M} := \bigcup_{\boldsymbol{\theta} \in \Theta} \mathcal{M}_\theta := \bigcup_{\boldsymbol{\theta} \in \Theta} \left\{ Q \in M^1 \text{ such that } \int_{\mathbb{R}^m} \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) dQ(\mathbf{x}) = \mathbf{0} \right\}. \quad (1)$$

Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be an i.i.d. sample of the random vector  $\mathbf{X} \in \mathbb{R}^m$  with unknown probability distribution  $P_{\mathbf{X}}(\cdot)$ . The problems of testing the model  $\mathcal{H}_0 : P_{\mathbf{X}} \in \mathcal{M}$ , confidence region and point estimations of  $\boldsymbol{\theta}_T$ , have been widely investigated in the literature. Hansen (1982) considered generalized method of moments (GMM) in order to estimate  $\boldsymbol{\theta}_T$ . Hansen *et al.* (1996) introduced the continuous updating (CU) estimate. Asymptotic confidence regions for the parameter  $\boldsymbol{\theta}_T$  have been obtained by Owen (1988) and Owen (1990), introducing the empirical likelihood (EL) approach. It has been used, in the context of model (1), by Qin and Lawless (1994) and Imbens (1997) introducing the EL estimate for the parameter  $\boldsymbol{\theta}_T$ . The recent literature in econometrics focusses 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. Among other results pertaining to EL, Newey and Smith (2004) stated that EL estimate enjoys asymptotic optimality properties in term of efficiency when bias corrected among all GEL estimates including the GMM one. Broniatowski and Keziou (2012) proposed a general approach through empirical divergences and duality technique which includes the above methods in the general context of signed finite measures under moment condition models (1). These approach allows the asymptotic study of the estimates and associated test statistics both under the model and under misspecification, leading to new results, in particular, for the EL approach. Note that all the proposed estimates including the EL one are generally biased, and that the problem of their finite sample efficiency, at our knowledge, have not yet been studied.

The aim of the present paper is to investigate the finite-sample optimality property estimation in the context of semiparametric model (1). We will discuss the problem of constructing minimum risk equivariant estimates (MRE) for the parameter  $\boldsymbol{\theta}_T$ , as well as the problem of the numerical calculation of these estimates.

We recall in the following lines, for the above estimation problem, the notions of group transformations on the random vector space, model invariance and the induced group of transformations on the parameter space, loss invariance and equivariance estimation; we refer to the unpublished preprint of Hoff (2012) for an excellent presentation of the above notions, and the book of Lehmann and Casella (1998).

Let  $\mathcal{G}$  be a collection, of one-to-one transformations from the vector space  $\mathbb{R}^m$  in  $\mathbb{R}^m$ , which we assume to be a “group”, in the sense that, it should be closed under both composition and inversion, namely,

$$\forall g_1, g_2 \in \mathcal{G}, g_1 \circ g_2 \in \mathcal{G} \quad \text{and} \quad \forall g \in \mathcal{G}, g^{-1} \in \mathcal{G}.$$

The group  $\mathcal{G}$  can be extended to a group of transformations on the sample space,  $\mathbb{R}^{mn}$  onto  $\mathbb{R}^{mn}$ , which will be denoted  $\mathcal{G}_n$ , as follows

$$\mathcal{G}_n := \{ \mathbf{g} : (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^{mn} \mapsto \mathbf{g}(\mathbf{x}_1, \dots, \mathbf{x}_n) := (g(\mathbf{x}_1), \dots, g(\mathbf{x}_n)) \in \mathbb{R}^{mn}; g \in \mathcal{G} \}.$$

We will consider two kinds of transformation groups,

- “additive”

$$\mathcal{G} := \{ \mathbf{x} \in \mathbb{R}^m \mapsto \mathbf{x} + \mathbf{a} \in \mathbb{R}^m; \mathbf{a} \in S \}, \quad (2)$$

where  $S$  is some subset of  $\mathbb{R}^m$ ,

or

- “multiplicative”

$$\mathcal{G} := \{ \mathbf{x} \in \mathbb{R}^m \mapsto \text{diag}(\lambda_1, \dots, \lambda_m) \mathbf{x} \in \mathbb{R}^m; \lambda \in \mathbb{R}^m \text{ or } \lambda \in \mathbb{R}^{*m} \}, \quad (3)$$

where  $\text{diag}(\lambda_1, \dots, \lambda_m)$  is diagonal matrix, with entries  $\lambda_1, \dots, \lambda_m \in \mathbb{R}^*$  or  $\lambda_1, \dots, \lambda_m \in \mathbb{R}_+^*$  with possibly some entries  $\lambda_i$  equal to one.

We assume that the model  $\mathcal{M}$  given in (1) is invariant under the considered group of transformations  $\mathcal{G}$ , in the sense that,

$$\text{for any random vector } \mathbf{X}, \text{ if } P_{\mathbf{X}} \in \mathcal{M}, \text{ then } P_{g(\mathbf{X})} \in \mathcal{M}, \forall g \in \mathcal{G}. \quad (4)$$

The induced group of transformations on the parameter space,  $\Theta$  onto  $\Theta$ , denoted  $\overline{\mathcal{G}}$  hereafter, will be defined as follows. Let  $g$  be any transformation belonging to  $\mathcal{G}$ , and consider any random vector  $\mathbf{X}$  such that  $P_{\mathbf{X}} \in \mathcal{M}$ . Then, by identifiability assumption, there exists a unique  $\boldsymbol{\theta} \in \Theta$  such that  $P_{\mathbf{X}} \in \mathcal{M}_{\boldsymbol{\theta}}$ . By invariance assumption (4), of the model  $\mathcal{M}$  to the group  $\mathcal{G}$ , the distribution  $P_{g(\mathbf{X})}$  belongs to  $\mathcal{M}$ . Therefore, there exists a unique (by indentifiability)  $\overline{\boldsymbol{\theta}} \in \Theta$

such that  $P_{g(\mathbf{x})} \in \mathcal{M}_{\bar{g}}$ . Denote then by  $\bar{g}$  the bijection induced by  $g$  on the parameter space  $\Theta$  onto  $\Theta$ , defined by

$$\bar{g} : \boldsymbol{\theta} \in \Theta \mapsto \bar{g}(\boldsymbol{\theta}) := \bar{\boldsymbol{\theta}} \in \Theta.$$

The induced group on the parameter space,  $\Theta$  onto  $\Theta$ , is then defined to be

$$\bar{\mathcal{G}} := \{\bar{g} \text{ such that } g \in \mathcal{G}\}.$$

Two points  $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$  are said equivalent iff  $\boldsymbol{\theta}_2 = \bar{g}(\boldsymbol{\theta}_1)$  for some  $\bar{g} \in \bar{\mathcal{G}}$ . The orbit  $\Theta(\boldsymbol{\theta}_0)$ , of a point  $\boldsymbol{\theta}_0 \in \Theta$ , is defined to be the set of equivalent points:

$$\Theta(\boldsymbol{\theta}_0) := \{\bar{g}(\boldsymbol{\theta}_0); \bar{g} \in \bar{\mathcal{G}}\}.$$

We will assume that there is only one orbit of  $\Theta$ , i.e.,

$$\Theta(\boldsymbol{\theta}_0) = \Theta \tag{5}$$

which means that the group of transformation  $\bar{\mathcal{G}}$  is rich enough allowing to go from any point in  $\Theta$  to another via some transformation  $\bar{g} \in \bar{\mathcal{G}}$ . In such case, the group  $\bar{\mathcal{G}}$  is said to be “transitive” over  $\Theta$ .

We give here some examples for illustration. In all the examples below, we can see that the group  $\bar{\mathcal{G}}$  is transitive over  $\Theta$ .

**Example 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  $X_1, \dots, X_n$  be an i.i.d. sample of a random variable  $X \in \mathbb{R}$  with mean  $\mathbb{E}(X) = \theta_T$ , and assume that  $\mathbb{E}(X^2) = h(\theta_T)$ , where  $h(\cdot)$  is a known function. Our aim is to estimate  $\theta_T$ . The information about the distribution  $P_X$  of  $X$  can be expressed in the form of (1) by taking  $\mathbf{f}(x, \theta) := (x - \theta, x^2 - h(\theta))^\top$ . If we take the parameter space to be  $\Theta = \mathbb{R}$ , then it is straightforward to see that the model  $\mathcal{M}$  is invariant to the additive group of transformations*

$$\mathcal{G} := \{g : x \in \mathbb{R} \mapsto g(x) := x + a; a \in \mathbb{R}\},$$

*if  $h(\theta) := \theta^2 + c$  for some  $c \geq 0$ , and invariant to the multiplicative group*

$$\mathcal{G} := \{g : x \in \mathbb{R} \mapsto g(x) := \lambda x; \lambda \in \mathbb{R}_+^*\},$$

*if  $h(\theta) := c\theta^2$  for some  $c > 0$ . The induced groups on the parameter space  $\Theta := \mathbb{R}$  are, respectively,*

$$\bar{\mathcal{G}} = \{\bar{g} : \theta \in \mathbb{R} \mapsto \bar{g}(\theta) = \theta + a; a \in \mathbb{R}\}$$

*and*

$$\bar{\mathcal{G}} = \{\bar{g} : \theta \in \mathbb{R} \mapsto \bar{g}(\theta) = \lambda\theta; \lambda \in \mathbb{R}_+^*\}.$$

**Example 2.** Let  $\mathbf{X}_1 := (X_{1,1}, X_{2,1})^\top, \dots, \mathbf{X}_n := (X_{1,n}, X_{2,n})^\top$  be an i.i.d. sample of a bivariate random vector  $\mathbf{X} := (X_1, X_2)^\top$  with  $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \theta_T$ . In this case, we can take  $\mathbf{f}(\mathbf{x}, \theta) = (x_1 - \theta, x_2 - \theta)^\top$ . If we consider  $\Theta = \mathbb{R}$ , then the model  $\mathcal{M}$  is invariant with respect to the groups

$$\mathcal{G} := \{g : \mathbf{x} \in \mathbb{R}^2 \mapsto g(\mathbf{x}) := \mathbf{x} + a\mathbf{1}; a \in \mathbb{R}\}$$

or

$$\mathcal{G} := \{g : \mathbf{x} \in \mathbb{R}^2 \mapsto g(\mathbf{x}) := \lambda\mathbf{x}; \lambda \in \mathbb{R}_+^*\}.$$

The induced groups on  $\Theta := \mathbb{R}$  are, respectively,

$$\bar{\mathcal{G}} = \{\bar{g} : \theta \in \mathbb{R} \mapsto \bar{g}(\theta) = \theta + a; a \in \mathbb{R}\}$$

and

$$\bar{\mathcal{G}} = \{\bar{g} : \theta \in \mathbb{R} \mapsto \bar{g}(\theta) = \lambda\theta; \lambda \in \mathbb{R}_+^*\}.$$

A somewhat similar problem is when  $\mathbb{E}(X_1) = c$  is known, and  $\mathbb{E}(X_2) = \theta_T$  is to be estimated, by taking  $\mathbf{f}(\mathbf{x}, \theta) = (x_1 - c, x_2 - \theta)^\top$ . Such problems are common in survey sampling (see e.g. [Kuk and Mak \(1989\)](#) and [Chen and Qin \(1993\)](#)). Taking  $\Theta = \mathbb{R}$ , the model  $\mathcal{M}$  is then invariant with respect to the groups

$$\mathcal{G} := \{g : \mathbf{x} \in \mathbb{R}^2 \mapsto g(\mathbf{x}) := \mathbf{x} + (0, a)^\top \in \mathbb{R}^2; a \in \mathbb{R}\}$$

or

$$\mathcal{G} := \{g : \mathbf{x} \in \mathbb{R}^2 \mapsto g(\mathbf{x}) := (x_1, \lambda x_2)^\top \in \mathbb{R}^2; \lambda \in \mathbb{R}_+^*\}.$$

The induced groups on  $\Theta := \mathbb{R}$  are, respectively,

$$\bar{\mathcal{G}} = \{\bar{g} : \theta \in \mathbb{R} \mapsto \bar{g}(\theta) = \theta + a; a \in \mathbb{R}\}$$

and

$$\bar{\mathcal{G}} = \{\bar{g} : \theta \in \mathbb{R} \mapsto \bar{g}(\theta) = \lambda\theta; \lambda \in \mathbb{R}_+^*\}.$$

**Example 3.** Let  $X_1, \dots, X_n$  be an i.i.d. sample of a random variable  $X$  with distribution  $P_X$  such that  $\mathbb{E}(f_i(X - \theta_T)) = 0$ , where  $f_i : x \in \mathbb{R} \mapsto f_i(x) := \mathbf{1}_{(a_i, b_i)}(x) - c_i, \forall i = 1, \dots, \ell$ . The known intervals  $(a_i, b_i)$  may be bounded or unbounded, and  $c_1, \dots, c_\ell$  are known nonnegative numbers. The information about  $P_X$  can be written under the form of model (1) taking  $\mathbf{f}(x, \theta) := (f_1(x - \theta), \dots, f_\ell(x - \theta))^\top$  and  $\theta \in \Theta := \mathbb{R}$ . The model  $\mathcal{M}$  in this case is invariant to the groups

$$\mathcal{G} := \{g : x \in \mathbb{R} \mapsto g(x) := x + a; a \in \mathbb{R}\}$$

or

$$\mathcal{G} := \{g : x \in \mathbb{R} \mapsto g(x) := \lambda x; \lambda \in \mathbb{R}_+^*\},$$

and the induced groups on the parameter space  $\Theta$  are, respectively,

$$\overline{\mathcal{G}} = \{\overline{g} : \theta \in \mathbb{R} \mapsto \overline{g}(\theta) = \theta + a; a \in \mathbb{R}\}$$

and

$$\overline{\mathcal{G}} = \{\overline{g} : \theta \in \mathbb{R} \mapsto \overline{g}(\theta) = \lambda\theta; \lambda \in \mathbb{R}_+^*\}.$$

**Example 4.** Let  $X_1, \dots, X_n$  be an i.i.d. sample of a random variable  $X \in \mathbb{R}$  with continuous distribution  $P_X$  such that  $\mathbb{E}(X) = \theta_{1T}$ ,  $\mathbb{E}(X^2) = 1 + \theta_{1T}^2$  and  $\mathbb{E}(\mathbf{1}_{]-\infty, \theta_{2T}[}(X)) = \alpha$ , where  $\alpha \in ]0, 1[$  is known and  $\boldsymbol{\theta}_T := (\theta_{1T}, \theta_{2T})^\top$  is to be estimated. Note that  $\theta_{2T}$  is the quantile of order  $\alpha$  of the variable  $X$ , and that the variance of  $X$  is assumed to be known and equal to one. This problem can be written under the form of model (1) taking  $\mathbf{f}(x, \boldsymbol{\theta}) := (x - \theta_1, x^2 - 1 - \theta_1^2, \mathbf{1}_{]-\infty, 0]}(x - \theta))^\top$  and  $\boldsymbol{\theta} \in \Theta := \mathbb{R}^2$ . The model  $\mathcal{M}$  in this case is invariant with respect to the additive group

$$\mathcal{G} := \{g : x \in \mathbb{R} \mapsto g(x) := x + a; a \in \mathbb{R}\}$$

and the induced group on  $\Theta$  is

$$\overline{\mathcal{G}} = \{\overline{g} : \boldsymbol{\theta} \in \mathbb{R}^2 \mapsto \overline{g}(\boldsymbol{\theta}) = \boldsymbol{\theta} + a\mathbf{1}; a \in \mathbb{R}\}.$$

**Example 5.** Let  $X_1, \dots, X_n$  be an i.i.d. sample of a random variable  $X \in \mathbb{R}$  with continuous distribution  $P_X$  such that  $\mathbb{E}(X) = 0$  and  $\mathbb{E}(\mathbf{1}_{]-\infty, \theta_T]}(X)) = \alpha$ , where  $\alpha \in ]0, 1[$  is known and  $\theta_T$  is to be estimated. Note that  $\theta_T$  is the quantile of order  $\alpha$  of the variable  $X$ . This problem can be written under the form of model (1) taking  $\mathbf{f}(x, \theta) := (x, \mathbf{1}_{]-\infty, 0]}(x - \theta))$  and  $\theta \in \Theta := \mathbb{R}$ . The model  $\mathcal{M}$  in this case is invariant with respect to the multiplicative group

$$\mathcal{G} := \{g : x \in \mathbb{R} \mapsto g(x) := \lambda x; \lambda \in \mathbb{R}_+^*\}$$

and the induced group on  $\Theta$  is

$$\overline{\mathcal{G}} = \{\overline{g} : \theta \in \mathbb{R} \mapsto \overline{g}(\theta) = \lambda\theta; \lambda \in \mathbb{R}_+^*\}.$$

**Example 6.** Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be an i.i.d. sample of a random vector  $\mathbf{X} \in \mathbb{R}^m$  with continuous distribution  $P_{\mathbf{X}}$  such that  $\mathbb{E}(\mathbf{h}(\mathbf{X} - \boldsymbol{\theta}_T)) = 0$ , where  $\mathbf{h} : \mathbb{R}^m \mapsto \mathbb{R}^\ell$  is some specified measurable function, and  $\boldsymbol{\theta}_T \in \mathbb{R}^m$  is to be estimated. We can consider also the case where some components of  $\boldsymbol{\theta}_T$  are known and that the other components are to be estimated. It is clear that the corresponding model  $\mathcal{M}$  defined in (1), taking  $\mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) := \mathbf{h}(\mathbf{x} - \boldsymbol{\theta})$  and  $\Theta = \mathbb{R}^m$ , is invariant to the additive group

$$\mathcal{G} := \{g : \mathbf{x} \in \mathbb{R}^m \mapsto g(\mathbf{x}) := \mathbf{x} + \mathbf{a}; \mathbf{a} \in \mathbb{R}^m\}$$

and the induced group on the parameter is

$$\overline{\mathcal{G}} = \{\overline{g} : \boldsymbol{\theta} \in \mathbb{R}^m \mapsto \overline{g}(\boldsymbol{\theta}) = \boldsymbol{\theta} + \mathbf{a}; \mathbf{a} \in \mathbb{R}^m\}.$$

Likewise, if the data are such that  $\mathbb{E}\left(\mathbf{h}\left(\frac{\mathbf{X}}{\boldsymbol{\theta}_T}\right)\right) = 0$ , where  $\mathbf{h} : \mathbb{R}^m \mapsto \mathbb{R}^\ell$  is some specified measurable function, and  $\boldsymbol{\theta}_T \in \Theta \subset \mathbb{R}^m$  is to be estimated, then the corresponding model  $\mathcal{M}$ , taking  $\mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) := \mathbf{h}\left(\frac{\mathbf{x}}{\boldsymbol{\theta}}\right)$ , is invariant to the multiplicative group

$$\mathcal{G} := \{g : \mathbf{x} \in \mathbb{R}^m \mapsto g(\mathbf{x}) := \text{diag}(\lambda_1, \dots, \lambda_m)\mathbf{x}; \boldsymbol{\lambda} \in \mathbb{R}^m \text{ or } \boldsymbol{\lambda} \in \mathbb{R}^{*m}\}$$

and the induced group on the parameter is

$$\overline{\mathcal{G}} = \{\overline{g} : \boldsymbol{\theta} \in \Theta \mapsto \overline{g}(\boldsymbol{\theta}) = \text{diag}(\lambda_1, \dots, \lambda_m)\boldsymbol{\theta}; \boldsymbol{\lambda} \in \mathbb{R}^m \text{ or } \boldsymbol{\lambda} \in \mathbb{R}^{*m}\}.$$

In all the sequel, without loss of generality, we assume that the model (1) and the group of transformation  $\mathcal{G}$  are such that

$$\mathbf{f}(g(\mathbf{x}), \overline{g}(\boldsymbol{\theta})) = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}), \quad \forall \mathbf{x} \in \mathbb{R}^m, \forall \boldsymbol{\theta} \in \Theta, \forall g \in \mathcal{G}. \quad (6)$$

Note that this assumption implies the condition (4) that the model  $\mathcal{M}$  is invariant under  $\mathcal{G}$ .

In all the following, when estimating  $\boldsymbol{\theta}_T$  by an estimate  $\tilde{\boldsymbol{\theta}}_n := T(\mathbf{X}_1, \dots, \mathbf{X}_n)$ , we consider the quadratic loss function

$$L_2(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_T) := \mathbb{E}\left(\left\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_T\right\|^2\right) \quad (7)$$

if the model is invariant with respect to additive group, and the loss function is taken to be relative quadratic

$$L_r(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_T) := \mathbb{E}\left(\left\|\frac{\tilde{\boldsymbol{\theta}}_n}{\boldsymbol{\theta}_T} - \mathbf{1}\right\|^2\right) \quad (8)$$

if the model is invariant with respect to the multiplicative group.

**Definition 7.** (*invariant loss under a group of transformations*). A loss function  $L(\cdot, \cdot) : (\boldsymbol{\theta}_T, d) \in \Theta \times D \mapsto L(\boldsymbol{\theta}_T, d) \in \mathbb{R}_+$ , where  $D$  denotes the set of the parameter estimates (called decision space), is invariant under a transformation  $g \in \mathcal{G}$  iff for any estimate  $d \in D$ , there exists a unique  $d' \in D$  such that  $L(\boldsymbol{\theta}_T, d) = L(\overline{g}(\boldsymbol{\theta}_T), d')$ . We denote then by  $\tilde{g}(\cdot)$  the bijection, from  $D$  onto  $D$ , such that  $d' = \tilde{g}(d)$ . Hence, we have

$$L(\boldsymbol{\theta}_T, d) = L(\overline{g}(\boldsymbol{\theta}_T), \tilde{g}(d)), \quad \forall d \in D.$$

We denote by  $\tilde{\mathcal{G}}$  the induced group on the decision space  $D$ .



**Definition 8.** Assume that the estimation problem  $(\mathcal{M}, D, L)$  is invariant under the group  $\mathcal{G}$ . Let  $\overline{\mathcal{G}}$  and  $\widetilde{\mathcal{G}}$  be, respectively, the induced groups on the parameter space  $\Theta$  and the decision space  $D$ . An estimate  $\widetilde{\theta}_n := T(\mathbf{X}_1, \dots, \mathbf{X}_n) \in D$  is said to be equivariant iff

$$T(g(\mathbf{X}_1), \dots, g(\mathbf{X}_n)) = \widetilde{g}(T(\mathbf{X}_1, \dots, \mathbf{X}_n)), \quad \forall g \in \mathcal{G}.$$

We will see, under condition (6), that the empirical minimum divergence estimates, introduced in Broniatowski and Keziou (2012), are equivariant for the above models, using results on the existence and characterization of the distribution  $P_{\mathbf{X}}$  on the sets  $\mathcal{M}_{\theta}$ . First, we recall the definition of  $D_{\varphi}$ -divergences and some of their properties. 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  $\varphi(x) \rightarrow \varphi(a)$  when  $x \downarrow a$ , and  $\varphi(x) \rightarrow \varphi(b)$  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  $D_{\varphi}$ -divergence between  $Q$  and a probability distribution  $P$ , when  $Q$  is absolutely continuous with respect to (a.c.w.r.t)  $P$ , is defined through

$$D_{\varphi}(Q, P) := \int_{\mathbb{R}^m} \varphi \left( \frac{dQ}{dP}(\mathbf{x}) \right) dP(\mathbf{x}), \quad (9)$$

where  $\frac{dQ}{dP}(\cdot)$  is the Radon-Nikodym derivative of  $Q$  w.r.t.  $P$ . When  $Q$  is not a.c.w.r.t.  $P$ , we set  $D_{\varphi}(Q, P) := +\infty$ . For any probability distribution  $P$ , 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 we have

$$D_{\varphi}(Q, P) = 0 \text{ if and only if } Q = P. \quad (10)$$

All the above properties are presented in Csiszár (1963), Csiszár (1967) and in Chapter 1 of Liese and Vajda (1987), for  $D_{\varphi}$ -divergences defined on the set of all probability distributions  $M^1$ . When the  $D_{\varphi}$ -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  $\chi^2$  ( $\chi_m^2$ ), 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

$$x \in \mathbb{R}_+^* \mapsto \varphi_{\gamma}(x) := \frac{x^{\gamma} - \gamma x + \gamma - 1}{\gamma(\gamma - 1)} \quad (11)$$

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  $\chi_m^2$  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 if  $\gamma = 3$ ), we extend the definition of  $\varphi_\gamma$  as follows

$$x \in \mathbb{R} \mapsto \varphi_\gamma(x) \mathbb{1}_{[0, +\infty[}(x) + (+\infty) \mathbb{1}_{]-\infty, 0[}(x). \quad (12)$$

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 (12), satisfy the above conditions, including all standard divergences.

## 2. MINIMUM EMPIRICAL DIVERGENCE ESTIMATES

Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  denote an i.i.d. sample of a random vector  $\mathbf{X} \in \mathbb{R}^m$  with probability distribution  $P_{\mathbf{X}}$ . Let  $P_n(\cdot)$  be the associated empirical measure, namely,

$$P_n(\cdot) := \frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{X}_i}(\cdot),$$

where  $\delta_{\mathbf{x}}(\cdot)$  denotes the Dirac measure at point  $\mathbf{x}$ , for all  $\mathbf{x}$ . For a given  $\theta \in \Theta$ , the ‘‘plug-in’’ estimate of  $D_\varphi(\mathcal{M}_\theta, P_{\mathbf{X}})$  is

$$\tilde{D}_\varphi(\mathcal{M}_\theta, P_{\mathbf{X}}) := \inf_{Q \in \mathcal{M}_\theta} D_\varphi(Q, P_n) = \inf_{Q \in \mathcal{M}_\theta} \int_{\mathbb{R}^m} \varphi \left( \frac{dQ}{dP_n}(\mathbf{x}) \right) dP_n(\mathbf{x}). \quad (13)$$

If the projection  $Q_\theta^{(n)}$  of  $P_n$  on  $\mathcal{M}_\theta$  exists, then it is clear that  $Q_\theta^{(n)}$  is a s.f.m. (or possibly a probability distribution) a.c.w.r.t.  $P_n$ ; this means that the support of  $Q_\theta^{(n)}$  must be included in the set  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ . So, define the set

$$\mathcal{M}_\theta^{(n)} := \left\{ Q \in M \text{ s.t. } Q \text{ a.c.w.r.t. } P_n, \sum_{i=1}^n Q(\mathbf{X}_i) = 1 \text{ and } \sum_{i=1}^n Q(\mathbf{X}_i) \mathbf{f}(\mathbf{X}_i, \theta) = \mathbf{0} \right\}, \quad (14)$$

which may be seen as a subset of  $\mathbb{R}^n$ . Then, the plug-in estimate (13) can be written as

$$\tilde{D}_\varphi(\mathcal{M}_\theta, P_{\mathbf{X}}) := \inf_{Q \in \mathcal{M}_\theta} D_\varphi(Q, P_n) = \inf_{Q \in \mathcal{M}_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n \varphi(nQ(\mathbf{X}_i)). \quad (15)$$

In the same way,

$$D_\varphi(\mathcal{M}, P_{\mathbf{X}}) := \inf_{\boldsymbol{\theta} \in \Theta} \inf_{Q \in \mathcal{M}_\theta} D_\varphi(Q, P_{\mathbf{X}})$$

can be estimated by

$$\tilde{D}_\varphi(\mathcal{M}, P_{\mathbf{X}}) := \inf_{\boldsymbol{\theta} \in \Theta} \tilde{D}_\varphi(\mathcal{M}_\theta, P_{\mathbf{X}}) = \inf_{\boldsymbol{\theta} \in \Theta} \inf_{Q \in \mathcal{M}_\theta} D_\varphi(Q, P_n) = \inf_{\boldsymbol{\theta} \in \Theta} \inf_{Q \in \mathcal{M}_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n \varphi(nQ(\mathbf{X}_i)). \quad (16)$$

By uniqueness of  $\arg \inf_{\boldsymbol{\theta} \in \Theta} D_\varphi(\mathcal{M}_\theta, P_{\mathbf{X}})$  and since the infimum is reached in  $\boldsymbol{\theta} = \boldsymbol{\theta}_T$ , we estimate  $\boldsymbol{\theta}_T$  through

$$\tilde{\boldsymbol{\theta}}_\varphi := \arg \inf_{\boldsymbol{\theta} \in \Theta} \inf_{Q \in \mathcal{M}_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n \varphi(nQ(\mathbf{X}_i)). \quad (17)$$

The expression of the estimate  $\tilde{D}(\mathcal{M}_\theta, P_{\mathbf{X}})$ , given in (15), is the solution of a convex optimization problem under convex constrained subset in  $\mathbb{R}^n$ . In order to transform this problem to an unconstrained one, we will make use of the Fenchel-Legendre transform, denoted  $\varphi^*(\cdot)$ , of the convex function  $\varphi(\cdot)$ , as well as some other duality arguments. It is defined by

$$\varphi^* : t \in \mathbb{R} \mapsto \varphi^*(t) := \sup_{x \in \mathbb{R}} \{tx - \varphi(x)\}. \quad (18)$$

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

$$a^* = \lim_{x \rightarrow -\infty} \frac{\varphi(x)}{x}, \quad b^* = \lim_{x \rightarrow +\infty} \frac{\varphi(x)}{x} \quad (19)$$

satisfying  $a^* < 0 < b^*$  with  $\varphi^*(0) = 0$ . 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  $\varphi^*$  given by (19). This holds when  $a$  or  $b$  is finite and  $\varphi'(a)$  or  $\varphi'(b)$  is finite, respectively. For example, for the convex function

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

we have  $\text{dom} \varphi = [0, +\infty[$  and  $\varphi'(0) = -1$ , and we can see that the domain of the corresponding  $\varphi^*$ -function 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

$$\begin{aligned} \lim_{t \downarrow a} \varphi'(t) &= -\infty & \text{if } a \text{ is finite,} \\ \lim_{t \uparrow b} \varphi'(t) &= +\infty & \text{if } b \text{ is finite.} \end{aligned} \quad (20)$$

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

$$\begin{aligned} \lim_{t \downarrow a^*} \varphi^{*'}(t) &= -\infty & \text{if } a^* \text{ is finite,} \\ \lim_{t \uparrow b^*} \varphi^{*'}(t) &= +\infty & \text{if } b^* \text{ is finite.} \end{aligned} \quad (21)$$

Conversely,  $\varphi$  is essentially smooth on its domain  $(a, b)$  if and only if  $\varphi^*$  is strictly convex on its domain  $(a^*, b^*)$ .

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

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

and

$$\varphi^*(t) = t\varphi'^{-1}(t) - \varphi(\varphi'^{-1}(t)), \quad \text{for all } t \in ]a^*, b^*[ , \quad (22)$$

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

$$\varphi^{*'}(t) = \varphi'^{-1}(t) \quad \text{and} \quad \varphi^{*''}(t) = \frac{1}{\varphi''(\varphi'^{-1}(t))}. \quad (23)$$

In particular,  $\varphi^{*'}(0) = 1$  and  $\varphi^{*''}(0) = 1$ . 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  $\varphi^*$ , likewise we have

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

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  $\varphi^*$  in  $a^*$  and  $b^*$  are defined in a similar way. In Table 1, using the above properties, we give the convex conjugates  $\varphi^*$  of some standard divergence functions  $\varphi$ , associated to standard divergences. We determine also their domains, respectively,  $(a, b)$  and  $(a^*, b^*)$ .

Using some duality arguments, see Broniatowski and Keziou (2012), we can show that, for any  $\theta \in \Theta$ , if there exists  $Q_0$  in  $\mathcal{M}_\theta^{(n)}$  such that

$$a < nQ_0(\mathbf{X}_i) < b, \quad \text{for all } i = 1, \dots, n, \quad (24)$$

TABLE 1. Convex conjugates  $\varphi^*$  of some standard divergence functions  $\varphi$ .

$D_\varphi$	$\text{dom}\varphi =: (a, b)$	$\varphi$	$\text{dom}\varphi^* =: (a^*, b^*)$	$\varphi^*$
$D_{KLm}$	$]0, +\infty[$	$\varphi(x) := -\log x + x - 1$	$] -\infty, 1[$	$\varphi^*(t) = -\log(1 - t)$
$D_{KL}$	$[0, +\infty[$	$\varphi(x) := x \log x - x + 1$	$\mathbb{R}$	$\varphi^*(t) = e^t - 1$
$D_{\chi_m^2}$	$]0, +\infty[$	$\varphi(x) := \frac{1}{2} \frac{(x-1)^2}{x}$	$] -\infty, \frac{1}{2}]$	$\varphi^*(t) = 1 - \sqrt{1 - 2t}$
$D_{\chi^2}$	$\mathbb{R}$	$\varphi(x) := \frac{1}{2} (x - 1)^2$	$\mathbb{R}$	$\varphi^*(t) = \frac{1}{2} t^2 + t$
$D_H$	$[0, +\infty[$	$\varphi(x) := 2(\sqrt{x} - 1)^2$	$] -\infty, 2[$	$\varphi^*(t) = \frac{2t}{2-t}$

then

$$\tilde{D}_\varphi(\mathcal{M}_\theta, P_{\mathbf{X}}) := \inf_{Q \in \mathcal{M}_\theta} D_\varphi(Q, P_n) = \sup_{(t_0, t_1, \dots, t_\ell)^\top \in \mathbb{R}^{1+\ell}} \left\{ t_0 - \frac{1}{n} \sum_{i=1}^n \varphi^*(t_0 + \sum_{j=1}^{\ell} t_j f_j(\mathbf{X}_i, \theta)) \right\} \quad (25)$$

with dual attainment. Conversely, if there exists some dual optimal solution

$$\tilde{\mathbf{t}}(\theta) := (\tilde{t}_0(\theta), \tilde{t}_1(\theta), \dots, \tilde{t}_\ell(\theta))^\top \in \mathbb{R}^{1+\ell}$$

such that

$$a^* < \tilde{t}_0(\theta) + \sum_{j=1}^{\ell} \tilde{t}_j(\theta) f_j(\mathbf{X}_i, \theta) < b^*, \quad \text{for all } i = 1, \dots, n, \quad (26)$$

then the equality (25) 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

$$Q_\theta^{(n)}(\mathbf{X}_i) = \frac{1}{n} \varphi'^{-1} \left( \tilde{t}_0(\theta) + \sum_{j=1}^{\ell} \tilde{t}_j(\theta) f_j(\mathbf{X}_i, \theta) \right), \quad i = 1, \dots, n, \quad (27)$$

where  $\tilde{\mathbf{t}}(\theta) := (\tilde{t}_0(\theta), \tilde{t}_1(\theta), \dots, \tilde{t}_\ell(\theta))^\top \in \mathbb{R}^{1+\ell}$  is solution of the system of equations

$$\begin{cases} 1 - \frac{1}{n} \sum_{i=1}^n \varphi'^{-1}(t_0 + \sum_{j=1}^{\ell} t_j f_j(\mathbf{X}_i, \theta)) & = 0, \\ -\frac{1}{n} \sum_{i=1}^n f_j(\mathbf{X}_i, \theta) \varphi'^{-1}(t_0 + \sum_{j=1}^{\ell} t_j f_j(\mathbf{X}_i, \theta)) & = 0, \quad j = 1, \dots, \ell. \end{cases} \quad (28)$$

In view of the last results, using the notations

$$\underline{\mathbf{t}} := (t_0, t_1, \dots, t_\ell)^\top \in \mathbb{R}^{1+\ell}, \quad \underline{\mathbf{f}}(\cdot, \theta) := (\mathbf{1}_{\mathbb{R}^m}(\cdot), f_1(\cdot, \theta), \dots, f_\ell(\cdot, \theta))^\top$$

and

$$\tilde{\mathbf{t}}(\boldsymbol{\theta}) := \arg \sup_{\mathbf{t} \in \mathbb{R}^{1+\ell}} \left\{ t_0 - \frac{1}{n} \sum_{i=1}^n \varphi^* (\mathbf{t}^\top \mathbf{f}(X_i, \boldsymbol{\theta})) \right\}, \quad \forall \boldsymbol{\theta} \in \Theta,$$

we obtain the following equivalent expressions to the estimates  $\tilde{D}_\varphi(\mathcal{M}_\theta, P_{\mathbf{X}})$ ,  $\tilde{D}_\varphi(\mathcal{M}, P_{\mathbf{X}})$  and  $\tilde{\boldsymbol{\theta}}$ , see (13), (16) and (17),

$$\tilde{D}_\varphi(\mathcal{M}_\theta, P_{\mathbf{X}}) = \sup_{\mathbf{t} \in \mathbb{R}^{1+\ell}} \left\{ t_0 - \frac{1}{n} \sum_{i=1}^n \varphi^* (\mathbf{t}^\top \mathbf{f}(X_i, \boldsymbol{\theta})) \right\} \quad (29)$$

$$= \tilde{t}_0(\boldsymbol{\theta}) - \frac{1}{n} \sum_{i=1}^n \varphi^* (\tilde{\mathbf{t}}(\boldsymbol{\theta})^\top \mathbf{f}(X_i, \boldsymbol{\theta})), \quad (30)$$

$$\tilde{D}_\varphi(\mathcal{M}, P_{\mathbf{X}}) := \inf_{\boldsymbol{\theta} \in \Theta} \tilde{D}_\varphi(\mathcal{M}_\theta, P_{\mathbf{X}}) = \inf_{\boldsymbol{\theta} \in \Theta} \sup_{\mathbf{t} \in \mathbb{R}^{1+\ell}} \left\{ t_0 - \frac{1}{n} \sum_{i=1}^n \varphi^* (\mathbf{t}^\top \mathbf{f}(\mathbf{X}_i, \boldsymbol{\theta})) \right\} \quad (31)$$

$$= \inf_{\boldsymbol{\theta} \in \Theta} \left\{ \tilde{t}_0(\boldsymbol{\theta}) - \frac{1}{n} \sum_{i=1}^n \varphi^* (\tilde{\mathbf{t}}(\boldsymbol{\theta})^\top \mathbf{f}(X_i, \boldsymbol{\theta})) \right\}, \quad (32)$$

and

$$\tilde{\boldsymbol{\theta}}_\varphi = \arg \inf_{\boldsymbol{\theta} \in \Theta} \left\{ \tilde{t}_0(\boldsymbol{\theta}) - \frac{1}{n} \sum_{i=1}^n \varphi^* (\tilde{\mathbf{t}}(\boldsymbol{\theta})^\top \mathbf{f}(X_i, \boldsymbol{\theta})) \right\}. \quad (33)$$

**Remark 9.** The empirical likelihood estimate  $\tilde{\boldsymbol{\theta}}_{EL}$  is obtained for the particular choice of the modified Kullback-Leibler divergence  $KL_m$ , namely, when  $\varphi(x) = \varphi_0(x) = -\log x + x - 1$ . Moreover, straightforward computation shows that  $\tilde{t}_0(\boldsymbol{\theta}) = 0, \forall \boldsymbol{\theta} \in \Theta$ . Therefore,  $t_0$  can be omitted, and the above expression can be simplified to

$$\tilde{D}_{KL_m}(\mathcal{M}, P_{\mathbf{X}}) = \inf_{\boldsymbol{\theta} \in \Theta} \sup_{\mathbf{t} \in \mathbb{R}^\ell} \frac{1}{n} \sum_{i=1}^n \log (1 - \mathbf{t}^\top \mathbf{f}(\mathbf{X}_i, \boldsymbol{\theta}))$$

and

$$\tilde{\boldsymbol{\theta}}_{EL} = \tilde{\boldsymbol{\theta}}_{\varphi_0} = \arg \inf_{\boldsymbol{\theta} \in \Theta} \sup_{\mathbf{t} \in \mathbb{R}^\ell} \frac{1}{n} \sum_{i=1}^n \log (1 - \mathbf{t}^\top \mathbf{f}(\mathbf{X}_i, \boldsymbol{\theta})).$$

We will show that for any divergence  $D_\varphi$ , the estimate  $\tilde{\boldsymbol{\theta}}_\varphi := \tilde{\boldsymbol{\theta}}_\varphi(\mathbf{X}_1, \dots, \mathbf{X}_n)$  is invariant with respect to  $L_2$  loss for the additive group, and invariant with respect to  $L_r$  loss for the multiplicative group. First, we expose the asymptotic counterpart of the estimates (29), (31) and (33). In particular, we give results about existence and characterization of the projection of  $P_{\mathbf{X}}$  on the model  $\mathcal{M}$ . The characterization of the projection will be of great importance in computing

the minimum risk equivariant estimate. We have; see Theorem 1 in Broniatowski and Keziou (2006):

**Proposition 10.** *Let  $\boldsymbol{\theta}$  be a given value in  $\Theta$ . Assume that  $\int_{\mathbb{R}^m} |f_j(\mathbf{x}, \boldsymbol{\theta})| dP_{\mathbf{X}}(\mathbf{x}) < \infty$  for all  $j = 1, \dots, \ell$ , and that there exists  $Q_0$  in  $\mathcal{M}_{\boldsymbol{\theta}}$  such that  $D_{\varphi}(Q_0, P_{\mathbf{X}}) < \infty$  and<sup>1</sup>*

$$a < \inf_{\mathbf{x} \in \mathbb{R}^m} \frac{dQ_0}{dP_{\mathbf{X}}}(\mathbf{x}) \leq \sup_{\mathbf{x} \in \mathbb{R}^m} \frac{dQ_0}{dP_{\mathbf{X}}}(\mathbf{x}) < b, \quad P_{\mathbf{X}} - a.s. \quad (34)$$

Then, we have

$$\inf_{Q \in \mathcal{M}_{\boldsymbol{\theta}}} D_{\varphi}(Q, P_{\mathbf{X}}) = \sup_{\underline{\mathbf{t}} \in \mathbb{R}^{1+\ell}} \left\{ t_0 - \int_{\mathbb{R}^m} \varphi^*(\underline{\mathbf{t}}^{\top} \mathbf{f}(\mathbf{x}, \boldsymbol{\theta})) dP_{\mathbf{X}}(\mathbf{x}) \right\} \quad (35)$$

with dual attainment. Conversely, if there exists a dual optimal solution

$$\underline{\mathbf{t}}(\boldsymbol{\theta}) := (t_0(\boldsymbol{\theta}), t_1(\boldsymbol{\theta}), \dots, t_{\ell}(\boldsymbol{\theta}))^{\top} \in \mathbb{R}^{1+\ell}$$

belonging to the interior (in  $\mathbb{R}^{1+\ell}$ ) of the set

$$\left\{ \underline{\mathbf{t}} \in \mathbb{R}^{1+\ell} \text{ such that } \int_{\mathbb{R}^m} |\varphi^*(\underline{\mathbf{t}}^{\top} \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}))| dP_{\mathbf{X}}(\mathbf{x}) < \infty \right\}, \quad (36)$$

then the dual equality (35) holds, and the unique optimal solution  $Q_{\boldsymbol{\theta}}^*$  of the primal problem  $\inf_{Q \in \mathcal{M}_{\boldsymbol{\theta}}} D_{\varphi}(Q, P_{\mathbf{X}})$ , namely, the projection of  $P_{\mathbf{X}}$  on  $\mathcal{M}_{\boldsymbol{\theta}}$ , is given by

$$\frac{dQ_{\boldsymbol{\theta}}^*}{dP_{\mathbf{X}}}(\mathbf{x}) = \varphi'^{-1} \left( t_0(\boldsymbol{\theta}) + \sum_{j=1}^{\ell} t_j(\boldsymbol{\theta}) f_j(\mathbf{x}, \boldsymbol{\theta}) \right),$$

where  $\underline{\mathbf{t}}(\boldsymbol{\theta}) := (t_0(\boldsymbol{\theta}), t_1(\boldsymbol{\theta}), \dots, t_{\ell}(\boldsymbol{\theta}))^{\top} \in \mathbb{R}^{1+\ell}$  is the solution of the system of equations

$$\begin{cases} 1 - \int \varphi'^{-1}(t_0 + \sum_{j=1}^{\ell} t_j f_j(\mathbf{x}, \boldsymbol{\theta})) dP_{\mathbf{X}}(\mathbf{x}) & = 0 \\ - \int f_j(\mathbf{x}, \boldsymbol{\theta}) \varphi'^{-1}(t_0 + \sum_{j=1}^{\ell} t_j f_j(\mathbf{x}, \boldsymbol{\theta})) dP_{\mathbf{X}}(\mathbf{x}) & = 0, \quad j = 1, \dots, \ell. \end{cases} \quad (37)$$

Furthermore, the solution  $\underline{\mathbf{t}}(\boldsymbol{\theta})$  is unique if the functions  $\mathbf{1}_{\mathbb{R}^m}(\cdot), f_1(\cdot, \boldsymbol{\theta}), \dots, f_{\ell}(\cdot, \boldsymbol{\theta})$  are linearly independent in the sense that  $P_{\mathbf{X}} \left( \left\{ \mathbf{x} \in \mathbb{R}^m; t_0 + \sum_{j=1}^{\ell} t_j f_j(\mathbf{x}, \boldsymbol{\theta}) \neq 0 \right\} \right) > 0$  for all  $\underline{\mathbf{t}} := (t_0, t_1, \dots, t_{\ell})^{\top} \in \mathbb{R}^{1+\ell}$  with  $(t_0, t_1, \dots, t_{\ell})^{\top} \neq \mathbf{0}$ .

**Remark 11.** *By minimizing  $D_{\varphi}(Q, P_{\mathbf{X}})$ , upon  $Q \in \mathcal{M}_{\boldsymbol{\theta}}$ ,  $\boldsymbol{\theta} \in \Theta$ , we obtain the semiparametric model of densities*

$$p(\mathbf{x}, \boldsymbol{\theta}) := \varphi'^{-1} \left( t_0(\boldsymbol{\theta}) + \sum_{j=1}^{\ell} t_j(\boldsymbol{\theta}) f_j(\mathbf{x}, \boldsymbol{\theta}) \right) p_{\mathbf{X}}(\mathbf{x}); \quad \boldsymbol{\theta} \in \Theta,$$

---

<sup>1</sup>The strict inequalities mean that  $P_{\mathbf{X}} \left( \left\{ \mathbf{x} \in \mathbb{R}^m; \frac{dQ_0}{dP_{\mathbf{X}}}(\mathbf{x}) \leq a \right\} \right) = P_{\mathbf{X}} \left( \left\{ \mathbf{x} \in \mathbb{R}^m; \frac{dQ_0}{dP_{\mathbf{X}}}(\mathbf{x}) \geq b \right\} \right) = 0$ .

where  $\underline{\mathbf{t}}(\boldsymbol{\theta}) := (t_0(\boldsymbol{\theta}), t_1(\boldsymbol{\theta}), \dots, t_\ell(\boldsymbol{\theta}))^\top \in \mathbb{R}^{1+\ell}$  is the solution of the system of equations (37). For the particular case of the KL-divergence, namely, when  $\varphi(x) = \varphi_1(x) = x \log x - x + 1$ ,  $t_0(\boldsymbol{\theta})$  can be explicitly computed, and the obtained model is the semiparametric exponential family of probability densities

$$\begin{aligned} p(\mathbf{x}, \boldsymbol{\theta}) &:= \frac{\exp \left\{ \sum_{j=1}^{\ell} t_j(\boldsymbol{\theta}) f_j(\mathbf{x}, \boldsymbol{\theta}) \right\} p_{\mathbf{X}}(\mathbf{x})}{\int_{\mathbb{R}^m} \exp \left\{ \sum_{j=1}^{\ell} t_j(\boldsymbol{\theta}) f_j(\mathbf{x}, \boldsymbol{\theta}) \right\} dP_{\mathbf{X}}(\mathbf{x})} \\ &:= \frac{\exp \left\{ \mathbf{t}(\boldsymbol{\theta})^\top \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \right\} p_{\mathbf{X}}(\mathbf{x})}{\int_{\mathbb{R}^m} \exp \left\{ \mathbf{t}(\boldsymbol{\theta})^\top \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \right\} dP_{\mathbf{X}}(\mathbf{x})}; \boldsymbol{\theta} \in \Theta, \end{aligned} \quad (38)$$

where, for all  $\boldsymbol{\theta} \in \Theta$ ,  $\mathbf{t}(\boldsymbol{\theta}) := (t_1(\boldsymbol{\theta}), \dots, t_\ell(\boldsymbol{\theta}))^\top \in \mathbb{R}^\ell$  is the solution in  $\mathbf{t} \in \mathbb{R}^\ell$  of the system of equations

$$\int_{\mathbb{R}^m} f_j(\mathbf{x}, \boldsymbol{\theta}) \exp \left\{ \mathbf{t}^\top \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \right\} dP_{\mathbf{X}}(\mathbf{x}) = 0, \quad j = 1, \dots, \ell. \quad (39)$$

or equivalently

$$\int_{\mathbb{R}^m} \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \exp \left\{ \mathbf{t}^\top \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \right\} dP_{\mathbf{X}}(\mathbf{x}) = \mathbf{0}.$$

**Proposition 12.** Assume that condition (6) holds. Then, the minimum empirical  $\phi$ -divergence estimates (33) are equivariant

- to the additive group of transformations with respect to the  $L_2$  loss;
- to the multiplicative group of transformations with respect to the  $L_r$  loss.

Moreover, in both cases, the induced group of transformations  $\tilde{\mathcal{G}}$  on the space of estimates is equal to  $\bar{\mathcal{G}}$ , the group of transformations on the parameter space  $\Theta$ , in the sense that

$$\tilde{\boldsymbol{\theta}}_\varphi(g(\mathbf{X}_1), \dots, g(\mathbf{X}_n)) = \bar{g}(\tilde{\boldsymbol{\theta}}_\varphi(\mathbf{X}_1, \dots, \mathbf{X}_n)), \quad \forall g \in \mathcal{G}.$$

**Corollary 13.** For any estimate  $\tilde{\boldsymbol{\theta}}_\varphi$ , the corresponding loss function  $\boldsymbol{\theta}_T \in \Theta \mapsto L(\boldsymbol{\theta}_T, \tilde{\boldsymbol{\theta}}_\varphi)$  is constant.

In view of the above corollary, for the additive group, in order to obtain the uniform minimum risk estimate, we can compute the risk  $L_2(\mathbf{0}, \tilde{\boldsymbol{\theta}}_\varphi)$  of any estimate  $\tilde{\boldsymbol{\theta}}_\varphi$  under the particular value  $\boldsymbol{\theta}_T = \mathbf{0}$ , and then select the estimate that minimizes the risk. Likewise, if a multiplicative group is considered, to obtain the uniform minimum risk estimate, we can compute the risk  $L_r(\mathbf{1}, \tilde{\boldsymbol{\theta}}_\varphi)$  of any estimate  $\tilde{\boldsymbol{\theta}}_\varphi$  under the particular value  $\boldsymbol{\theta}_T = \mathbf{1}$ , and then select the estimate that minimizes the risk. To do this, we will first characterize the equivariant estimates.

**Definition 14.** A functional  $U : (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^{mn} \mapsto U_{(\mathbf{x}_1, \dots, \mathbf{x}_n)}(\cdot) \in \tilde{\mathcal{G}}$  is “invariant” iff

$$U_{(g(\mathbf{x}_1), \dots, g(\mathbf{x}_n))}(\cdot) = U_{(\mathbf{x}_1, \dots, \mathbf{x}_n)}(\cdot), \quad \forall g \in \mathcal{G}, \forall (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^{mn}.$$



**Definition 15.** A functional  $\mathcal{U} : (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^{mn} \mapsto \mathcal{U}_{(\mathbf{x}_1, \dots, \mathbf{x}_n)}(\cdot) \in \tilde{\mathcal{G}}$  is a “maximal invariant” iff it is invariant and satisfies  $\forall(\mathbf{x}_1, \dots, \mathbf{x}_n), \forall(\mathbf{y}_1, \dots, \mathbf{y}_n) \in \mathbb{R}^{mn}$ ,

$$\mathcal{U}_{(\mathbf{x}_1, \dots, \mathbf{x}_n)}(\cdot) = \mathcal{U}_{(\mathbf{y}_1, \dots, \mathbf{y}_n)}(\cdot) \Rightarrow (\mathbf{y}_1, \dots, \mathbf{y}_n) = (g(\mathbf{x}_1), \dots, g(\mathbf{x}_n)), \text{ for some } g \in \mathcal{G}.$$

**Remark 16.** For the additive group, we have that a functional  $\mathcal{U}_{(\mathbf{x}_1, \dots, \mathbf{x}_n)}(\cdot) := f_{(\mathbf{0}, \mathbf{x}_2 - \mathbf{x}_1, \dots, \mathbf{x}_n - \mathbf{x}_1)}(\cdot)$ , a function of  $(\mathbf{0}, \mathbf{x}_2 - \mathbf{x}_1, \dots, \mathbf{x}_n - \mathbf{x}_1)$ , is maximal invariant. Likewise, for the multiplicative groupe, a functional  $\mathcal{U}_{(\mathbf{x}_1, \dots, \mathbf{x}_n)}(\cdot) := f_{\left(\frac{\mathbf{x}_1}{|\mathbf{x}_1|}, \frac{\mathbf{x}_2}{|\mathbf{x}_2|}, \dots, \frac{\mathbf{x}_n}{|\mathbf{x}_n|}\right)}(\cdot)$ , a function of  $\left(\frac{\mathbf{x}_1}{|\mathbf{x}_1|}, \frac{\mathbf{x}_2}{|\mathbf{x}_2|}, \dots, \frac{\mathbf{x}_n}{|\mathbf{x}_n|}\right)$ , is maximal invariant.

**Proposition 17.** Assume that the estimation problem  $(\mathcal{M}, D, L)$  is invariant under the group  $\mathcal{G}$ . Let  $\bar{\mathcal{G}}$  and  $\tilde{\mathcal{G}}$  be, respectively, the induced groups on the parameter space  $\Theta$  and the decision space  $D$ . Let  $\tilde{\theta}_0(\mathbf{X}_1, \dots, \mathbf{X}_n)$  be any equivariant estimate. Then, an estimate  $\tilde{\theta}(\mathbf{X}_1, \dots, \mathbf{X}_n)$  is equivariant iff

$$\tilde{\theta}(\mathbf{X}_1, \dots, \mathbf{X}_n) = U_{(\mathbf{X}_1, \dots, \mathbf{X}_n)} \left( \tilde{\theta}_0(\mathbf{X}_1, \dots, \mathbf{X}_n) \right),$$

for some invariant functional  $U_{(\mathbf{X}_1, \dots, \mathbf{X}_n)}(\cdot) \in \tilde{\mathcal{G}}$ , i.e.,

$$U_{(g(\mathbf{x}_1), \dots, g(\mathbf{x}_n))}(\cdot) = U_{(\mathbf{x}_1, \dots, \mathbf{x}_n)}(\cdot), \forall g \in \mathcal{G}, \forall(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^{mn}.$$

**Proposition 18.** (Hoff (2012), Theorem 3). A functional  $U : (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^{mn} \mapsto U_{(\mathbf{x}_1, \dots, \mathbf{x}_n)}(\cdot) \in \tilde{\mathcal{G}}$  is invariant iff it is a function of a maximal invariant functional  $\mathcal{U}$ .

Combining the above results, we obtain

**Proposition 19.** Let  $\tilde{\theta}_0(\mathbf{X}_1, \dots, \mathbf{X}_n)$  be any equivariant estimate. Then  $\tilde{\theta}(\mathbf{X}_1, \dots, \mathbf{X}_n)$  is equivariant iff

$$\tilde{\theta}(\mathbf{X}_1, \dots, \mathbf{X}_n) = \mathcal{H}_{\mathcal{U}_{(\mathbf{X}_1, \dots, \mathbf{X}_n)}}(\tilde{\theta}_0),$$

where  $\mathcal{H}_{\mathcal{U}_{(\mathbf{X}_1, \dots, \mathbf{X}_n)}}(\cdot) \in \tilde{\mathcal{G}}$  is some function of the maximal invariant functional  $\mathcal{U}_{(\mathbf{X}_1, \dots, \mathbf{X}_n)}$ .

**Remark 20.** Notice that  $\mathcal{H}_{\mathcal{U}_{(\mathbf{X}_1, \dots, \mathbf{X}_n)}}(\cdot)$  acts additively for additive group, and multiplicatively for multiplicative group, i.e.,

$$\mathcal{H}_{\mathcal{U}_{(\mathbf{X}_1, \dots, \mathbf{X}_n)}}(\tilde{\theta}_0) = \tilde{\theta}_0 + \mathcal{H}(\mathcal{U}_{(\mathbf{X}_1, \dots, \mathbf{X}_n)}),$$

when an additive group is considered, and

$$\mathcal{H}_{\mathcal{U}_{(\mathbf{X}_1, \dots, \mathbf{X}_n)}}(\tilde{\theta}_0) = \mathcal{H}(\mathcal{U}_{(\mathbf{X}_1, \dots, \mathbf{X}_n)}) \cdot \tilde{\theta}_0,$$

for multiplicative group.

## 3. UMRE ESTIMATE FOR ADDITIVE GROUP

Let  $\tilde{\boldsymbol{\theta}}_n := \tilde{\boldsymbol{\theta}}_n(\mathbf{X}_1, \dots, \mathbf{X}_n)$  be any one of the equivariant estimates  $\tilde{\boldsymbol{\theta}}_\varphi$ , and assume that  $\mathbb{E} \left( \left\| \tilde{\boldsymbol{\theta}}_n \right\|^2 \right) < \infty$ . Consider the  $L_2$  loss. In view of the above statements, the UMRE estimate of  $\boldsymbol{\theta}_T$  is then given by

$$\hat{\boldsymbol{\theta}}_n = \tilde{\boldsymbol{\theta}}_n - \mathbb{E}_0 \left( \tilde{\boldsymbol{\theta}}_n \mid \mathbf{T} \right), \quad (40)$$

where  $\mathbf{T} := (\mathbf{X}_2 - \mathbf{X}_1, \dots, \mathbf{X}_n - \mathbf{X}_1)$ , and  $\mathbb{E}_0 \left( \tilde{\boldsymbol{\theta}}_n \mid \mathbf{T} \right)$  is the conditional expectation of  $\tilde{\boldsymbol{\theta}}_n$  given  $\mathbf{T}$ , under the assumption that  $\boldsymbol{\theta}_T = \mathbf{0}$ . We give in the following an asymptotic approximation to the conditional expectation

$$\mathbb{E}_0 \left( \tilde{\boldsymbol{\theta}}_n \mid \mathbf{T} \right),$$

using the result of [Jurečková and Picek \(2009\)](#). Straightforward calculus, shows that the score function, of the semiparametric exponential family (38), can be written as

$$\psi_p(\mathbf{x}, \boldsymbol{\theta}) := \frac{\partial}{\partial \boldsymbol{\theta}} \log p(\mathbf{x}, \boldsymbol{\theta}) = \frac{(\partial / \partial \boldsymbol{\theta}) p(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{x}, \boldsymbol{\theta})} \quad (41)$$

$$\begin{aligned} &= \mathbf{t}'(\boldsymbol{\theta}) \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) + \mathbf{f}'(\mathbf{x}, \boldsymbol{\theta}) \mathbf{t}(\boldsymbol{\theta}) \\ &\quad - \frac{\int_{\mathbb{R}^m} [\mathbf{t}'(\boldsymbol{\theta}) \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) + \mathbf{f}'(\mathbf{x}, \boldsymbol{\theta}) \mathbf{t}(\boldsymbol{\theta})] \exp \{ \mathbf{t}(\boldsymbol{\theta})^\top \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \} dP_{\mathbf{X}}(\mathbf{x})}{\int_{\mathbb{R}^m} \exp \{ \mathbf{t}(\boldsymbol{\theta})^\top \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \} dP_{\mathbf{X}}(\mathbf{x})}, \end{aligned} \quad (42)$$

where  $\mathbf{f}'(\mathbf{x}, \boldsymbol{\theta})$  and  $\mathbf{t}'(\boldsymbol{\theta})$  are, respectively, the derivative w.r.t.  $\boldsymbol{\theta}$ , of  $\mathbf{f}(\mathbf{x}, \boldsymbol{\theta})^\top$  and  $\mathbf{t}(\boldsymbol{\theta})^\top$  the solution of the system (39). The derivative  $\mathbf{t}'(\boldsymbol{\theta})$  can be derived by the implicit function theorem. Denote  $\mathbf{h}(\boldsymbol{\theta}, \mathbf{t}) := (h_1(\boldsymbol{\theta}, \mathbf{t}), \dots, h_\ell(\boldsymbol{\theta}, \mathbf{t}))^\top$  with

$$h_j(\boldsymbol{\theta}, \mathbf{t}) := \int_{\mathbb{R}^m} f_j(\mathbf{x}, \boldsymbol{\theta}) \exp \{ \mathbf{t}^\top \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \} dP_{\mathbf{X}}(\mathbf{x}), \quad \forall j = 1, \dots, \ell.$$

Let

$$J_h^\theta(\boldsymbol{\theta}, \mathbf{t}) := \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{h}(\boldsymbol{\theta}, \mathbf{t})^\top \quad \text{and} \quad J_h^t(\boldsymbol{\theta}, \mathbf{t}) := \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{h}(\boldsymbol{\theta}, \mathbf{t})^\top.$$

Then, by the implicit function theorem, we have

$$\mathbf{t}'(\boldsymbol{\theta}) := \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{t}(\boldsymbol{\theta})^\top = -J_h^\theta(\boldsymbol{\theta}, \mathbf{t}(\boldsymbol{\theta})) [J_h^t(\boldsymbol{\theta}, \mathbf{t}(\boldsymbol{\theta}))]^{-1}. \quad (43)$$

Notice that, for true value  $\boldsymbol{\theta} = \boldsymbol{\theta}_T$ , since  $\mathbf{t}(\boldsymbol{\theta}_T) = \mathbf{0}$ , we obtain for the true value  $\boldsymbol{\theta}_T$  the more simpler expression

$$\psi_p(\mathbf{x}, \boldsymbol{\theta}_T) = \mathbf{t}'(\boldsymbol{\theta}_T) \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}_T) =: t'_1(\boldsymbol{\theta}_T) f_1(\mathbf{x}, \boldsymbol{\theta}_T) + \dots + t'_\ell(\boldsymbol{\theta}_T) f_\ell(\mathbf{x}, \boldsymbol{\theta}_T). \quad (44)$$

Let

$$J_\psi(\mathbf{x}, \boldsymbol{\theta}) := \left[ (\partial^2 / \partial \theta_i \partial \theta_j) \log p(\mathbf{x}, \boldsymbol{\theta}) \right]_{i,j=1,\dots,d}. \quad (45)$$

Under some integrability assumptions, by dominated convergence theorem, we obtain

$$\mathbb{E}(J_\psi(\mathbf{X}, \boldsymbol{\theta}_T)) = -\mathbb{E}(\psi_p(\mathbf{X}, \boldsymbol{\theta}_T)\psi_p(\mathbf{X}, \boldsymbol{\theta}_T)^\top) =: -I(\boldsymbol{\theta}_T), \quad (46)$$

which is the opposite of the Fisher information matrix.

**Theorem 21.** *Under some regularity conditions, we have*

$$\mathbb{E}_0(\tilde{\boldsymbol{\theta}}_\varphi | \mathbf{T}) = -I(\boldsymbol{\theta}_T)^{-1} \frac{1}{n} \sum_{i=1}^n \psi_p(\mathbf{X}_i, \tilde{\boldsymbol{\theta}}_\varphi) + o_P(n^{-1/2}) \quad (47)$$

$$\approx -\hat{I}(\tilde{\boldsymbol{\theta}}_\varphi)^{-1} \frac{1}{n} \sum_{i=1}^n \hat{\psi}_p(\mathbf{X}_i, \tilde{\boldsymbol{\theta}}_\varphi) \quad (48)$$

which gives the following approximation of the UMRE estimate

$$\hat{\boldsymbol{\theta}} \approx \tilde{\boldsymbol{\theta}}_\varphi + \hat{I}(\tilde{\boldsymbol{\theta}}_\varphi)^{-1} \frac{1}{n} \sum_{i=1}^n \hat{\psi}_p(\mathbf{X}_i, \tilde{\boldsymbol{\theta}}_\varphi), \quad (49)$$

where  $\hat{I}(\tilde{\boldsymbol{\theta}}_\varphi)$  is the empirical estimate of the Fisher information matrix  $I(\boldsymbol{\theta}_T)$ , given by

$$\hat{I}(\tilde{\boldsymbol{\theta}}_\varphi) := \frac{1}{n} \sum_{i=1}^n \hat{\psi}_p(\mathbf{X}_i, \hat{\boldsymbol{\theta}}_\varphi) \hat{\psi}_p(\mathbf{X}_i, \hat{\boldsymbol{\theta}}_\varphi)^\top,$$

with  $\forall \mathbf{x} \in \mathbb{R}^m, \forall \boldsymbol{\theta} \in \Theta$ ,

$$\begin{aligned} \hat{\psi}_p(\mathbf{x}, \boldsymbol{\theta}) &:= \tilde{\mathbf{t}}'(\boldsymbol{\theta}) \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) + \mathbf{f}'(\mathbf{x}, \boldsymbol{\theta}) \hat{\mathbf{t}}(\boldsymbol{\theta}) \\ &= \frac{\int_{\mathbb{R}^m} \left[ \tilde{\mathbf{t}}'(\boldsymbol{\theta}) \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) + \mathbf{f}'(\mathbf{x}, \boldsymbol{\theta}) \hat{\mathbf{t}}(\boldsymbol{\theta}) \right] \exp \left\{ \hat{\mathbf{t}}(\boldsymbol{\theta})^\top \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \right\} dP_n(\mathbf{x})}{\int_{\mathbb{R}^m} \exp \left\{ \hat{\mathbf{t}}(\boldsymbol{\theta})^\top \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \right\} dP_n(\mathbf{x})}, \end{aligned} \quad (50)$$

$\hat{\mathbf{t}}(\boldsymbol{\theta})$  is the solution of the empirical version of the system (39), i.e., the solution in  $\mathbf{t}$  of

$$\int_{\mathbb{R}^m} f_j(\mathbf{x}, \boldsymbol{\theta}) \exp \left\{ \mathbf{t}^\top \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \right\} dP_n(\mathbf{x}) = 0, \quad j = 1, \dots, \ell, \quad (51)$$

and  $\tilde{\mathbf{t}}'(\boldsymbol{\theta})$  is the gradient of  $\hat{\mathbf{t}}(\boldsymbol{\theta})$  at the point  $\boldsymbol{\theta}$  given by

$$\tilde{\mathbf{t}}'(\boldsymbol{\theta}) := \frac{\partial}{\partial \boldsymbol{\theta}} \hat{\mathbf{t}}(\boldsymbol{\theta})^\top = -\widehat{J}_h^\theta(\boldsymbol{\theta}, \hat{\mathbf{t}}(\boldsymbol{\theta})) \left[ \widehat{J}_h^t(\boldsymbol{\theta}, \hat{\mathbf{t}}(\boldsymbol{\theta})) \right]^{-1}, \quad (52)$$

where

$$\widehat{J}_h^\theta(\boldsymbol{\theta}, \mathbf{t}) := \frac{\partial}{\partial \boldsymbol{\theta}} \widehat{\mathbf{h}}(\boldsymbol{\theta}, \mathbf{t})^\top, \quad \widehat{J}_h^t(\boldsymbol{\theta}, \mathbf{t}) := \frac{\partial}{\partial \boldsymbol{\theta}} \widehat{\mathbf{h}}(\boldsymbol{\theta}, \mathbf{t})^\top, \quad \widehat{\mathbf{h}}(\boldsymbol{\theta}, \mathbf{t})^\top := (\widehat{h}_1(\boldsymbol{\theta}, \mathbf{t}), \dots, \widehat{h}_\ell(\boldsymbol{\theta}, \mathbf{t}))^\top$$

and

$$\widehat{h}_j(\boldsymbol{\theta}, \mathbf{t}) := \int_{\mathbb{R}^m} f_j(\mathbf{x}, \boldsymbol{\theta}) \exp \left\{ \mathbf{t}^\top \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \right\} dP_n(\mathbf{x}), \quad \forall j = 1, \dots, \ell.$$

## 4. UMRE ESTIMATE FOR MULTIPLICATIVE GROUP

Let  $\tilde{\boldsymbol{\theta}}_n := \tilde{\boldsymbol{\theta}}_n(\mathbf{X}_1, \dots, \mathbf{X}_n)$  be any one of the equivariant estimates  $\tilde{\boldsymbol{\theta}}_\varphi$  of  $\boldsymbol{\theta}_T$ , and assume that  $\mathbb{E} \left( \left\| \tilde{\boldsymbol{\theta}}_n \right\|^2 \right) < \infty$ . Consider the  $L_r$  loss. In view of the above statements, the UMRE estimate of  $\boldsymbol{\theta}_T$  is given by

$$\hat{\boldsymbol{\theta}}_n = \frac{\tilde{\boldsymbol{\theta}}_n \cdot \mathbb{E}_1 \left( \tilde{\boldsymbol{\theta}}_n | \mathbf{T} \right)}{\mathbb{E}_1 \left( \tilde{\boldsymbol{\theta}}_n^\top \tilde{\boldsymbol{\theta}}_n | \mathbf{T} \right)}, \quad (53)$$

where  $\mathbf{T} = \left( \frac{\mathbf{X}_1}{|\mathbf{X}_1|}, \frac{\mathbf{X}_2}{\mathbf{X}_1}, \dots, \frac{\mathbf{X}_n}{\mathbf{X}_1} \right)$ , and  $\mathbb{E}_1(\cdot | \mathbf{T})$  is the conditional expectation given  $\mathbf{T}$ , under the assumption that  $\boldsymbol{\theta}_T = \mathbf{1}$ .

## 5. SIMULATION RESULTS

**Example 22.** Consider the model

$$\mathcal{M} := \bigcup_{\theta \in \Theta} \mathcal{M}_\theta := \bigcup_{\theta \in \Theta} \left\{ Q \in M^1 \text{ such that } \int_{\mathbb{R}} \mathbf{f}(x, \theta) dQ(x) = \mathbf{0} \right\},$$

where  $\mathbf{f} : (x, \theta) \in \mathbb{R} \times \Theta \mapsto \mathbf{f}(x, \theta) := (x - \theta, (x - \theta)^2 - 1)^\top$ . Let  $X$  be a random variable with distribution  $P_X := \mathcal{N}(\theta_T, 1)$  with  $\theta_T = 1$ . The model is invariant to the additive group. We compare the mean square errors (MSE) of the EL estimate and the proposed UMRE estimate using the approximation (49), for the sample sizes  $n = 30, 40, 50, 60, 70, 80$ , with 1000 runs. We can see, from figure 5, that the proposed estimate improves the EL one for moderate sample sizes.

## REFERENCES

- Broniatowski, M. and Keziou, A. (2006). Minimization of  $\phi$ -divergences on sets of signed measures. *Studia Sci. Math. Hungar.; arXiv:1003.5457*, **43**(4), 403–442.
- Broniatowski, M. and Keziou, A. (2012). Divergences and duality for estimation and test under moment condition models. *J. Statist. Plann. Inference*, **142**(9), 2554–2573.
- Chen, J. H. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, **80**(1), 107–116.
- 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.

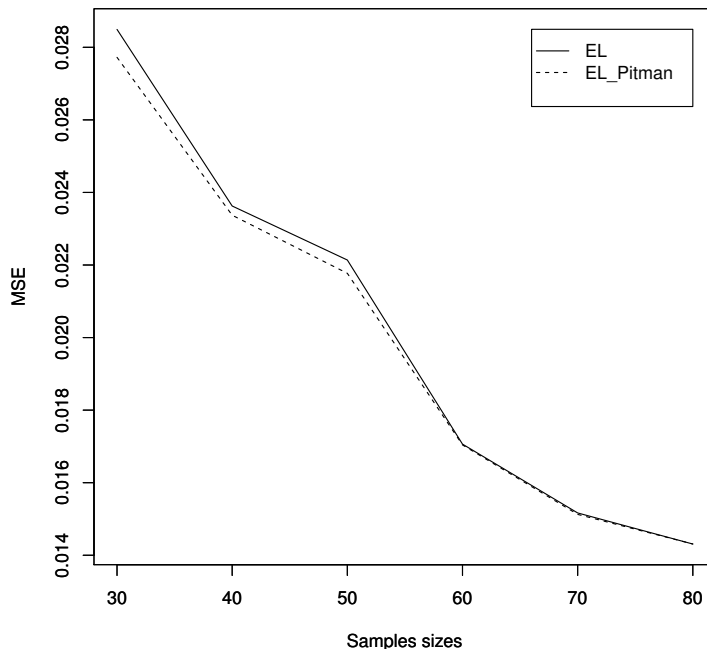


FIGURE 1. MSE v.s. sample sizes

- Csiszár, I. (1967). On topology properties of  $f$ -divergences. *Studia Sci. Math. Hungar.*, **2**, 329–339.
- Godambe, V. P. and Thompson, M. E. (1989). An extension of quasi-likelihood estimation. *J. Statist. Plann. Inference*, **22**(2), 137–172. With discussion and a reply by the authors.
- 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.
- Hoff, P. D. (2012). Equivariant estimation. *Preprint*.
- Imbens, G. W. (1997). One-step estimators for over-identified generalized method of moments models. *Rev. Econom. Stud.*, **64**(3), 359–383.
- Jurečková, J. and Picek, J. (2009). Minimum risk equivariant estimator in linear regression model. *Statist. Decisions*, **27**(1), 37–54.

- Kuk, A. Y. C. and Mak, T. K. (1989). Median estimation in the presence of auxiliary information. *J. Roy. Statist. Soc. Ser. B*, **51**(2), 261–269.
- Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition.
- 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.
- Sheehy, A. (1987). Kullback-Leibler constrained estimation of probability measures. *Report, Dept. Statistics, Stanford Univ.*
- Smith, R. J. (1997). Alternative semi-parametric likelihood approaches to generalized method of moments estimation. *Economic Journal*, **107**, 503–519.

<sup>1</sup> LPSM, SORBONNE-UNIVERSITÉ, PARIS, FRANCE, MICHEL.BRONIATOWSKI@UPMC.FR

<sup>2</sup>THE CZECH ACADEMY OF SCIENCES, INSTITUTE OF INFORMATION THEORY AND AUTOMATION, PRAGUE, CZECH REPUBLIC, JURECKOVA@KARLIN.MFF.CUNI.CZ

<sup>3</sup>LABORATOIRE DE MATHÉMATIQUES DE REIMS, FRANCE, AMOR.KEZIOU@UNIV-REIMS.FR