



# Robust Empirical Likelihood

Amor Keziou<sup>1(✉)</sup> and Aida Toma<sup>2,3</sup>

<sup>1</sup> Université de Reims Champagne-Ardenne, Laboratoire de Mathématiques de Reims (LMR) - CNRS UMR 9008, U.F.R. Sciences Exactes et Naturelles Moulin de la Housse, BP 1039, 51687 Reims Cedex 2, France

[amor.keziou@univ-reims.fr](mailto:amor.keziou@univ-reims.fr)

<sup>2</sup> Department of Applied Mathematics, Bucharest University of Economic Studies, Piața Romană No. 6, Bucharest, Romania

[aida.toma@csie.ase.ro](mailto:aida.toma@csie.ase.ro)

<sup>3</sup> Gheorghe Mihoc-Caius Iacob Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, Bucharest, Romania

<http://www.univ-reims.fr/labo-maths>

**Abstract.** In this paper, we present a robust version of the empirical likelihood estimator for semiparametric moment condition models. This estimator is obtained by minimizing the modified Kullback-Leibler divergence, in its dual form, using truncated orthogonality functions. Some asymptotic properties regarding the limit laws of the estimators are stated.

**Keywords:** Moment condition models · Robustness · Divergence measures

## 1 Introduction

We consider a moment condition model, namely a family  $\mathcal{M}$  of probability measures  $Q$ , all defined on the same measurable space  $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ , such that  $\int_{\mathbb{R}^m} g(x, \theta) dQ(x) = 0$ . The unknown parameter  $\theta$  belongs to the interior of a compact set  $\Theta \subset \mathbb{R}^d$ , and the function  $g := (g_1, \dots, g_\ell)^\top$ , with  $\ell \geq d$ , is defined on the set  $\mathbb{R}^m \times \Theta$ , each  $g_i$  being a real valued function. Denoting by  $M$  the set of all probability measures on  $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$  and defining the sets

$$\mathcal{M}_\theta := \left\{ Q \in M \text{ s.t. } \int_{\mathbb{R}^m} g(x, \theta) dQ(x) = 0 \right\}, \theta \in \Theta,$$

then the moment condition model  $\mathcal{M}$  can be written under the form

$$\mathcal{M} = \bigcup_{\theta \in \Theta} \mathcal{M}_\theta. \quad (1)$$

---

This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS/CCCDI – UEFISCDI, project number PN-III-P4-ID-PCE-2020-1112, within PNCDI III.

© Springer Nature Switzerland AG 2021

F. Nielsen and F. Barbaresco (Eds.): GSI 2021, LNCS 12829, pp. 841–848, 2021.

[https://doi.org/10.1007/978-3-030-80209-7\\_90](https://doi.org/10.1007/978-3-030-80209-7_90)

Let  $X_1, \dots, X_n$  be an i.i.d. sample with unknown probability measure  $P_0$ . We assume that the equation  $\int_{\mathbb{R}^m} g(x, \theta) dP_0(x) = 0$  has a unique solution (in  $\theta$ ) which will be denoted  $\theta_0$ . We consider the estimation problem of the true unknown value  $\theta_0$ .

Among the most known methods for estimating the parameter  $\theta_0$ , we recall the Generalized Method of Moments (GMM) of [6], the Continuous Updating (CU) estimator of [7], the Empirical Likelihood (EL) estimator of [8, 14, 15], the Exponential Tilting (ET) of [11], as well as the Generalized Empirical Likelihood (GEL) class of estimators of [13] that contains the EL, ET and CU estimators in particular. Some alternative methods have been proposed in order to improve the finite sample accuracy or the robustness under misspecification of the model, for example in [4, 8, 11, 12, 16].

The authors in [3] have developed a general methodology for estimation and testing in moment condition models. Their approach is based on minimizing divergences in dual form and allows the asymptotic study of the estimators (called minimum empirical divergence estimators) and of the associated test statistics, both under the model and under misspecification of the model. Using the approach based on the influence function, [18] studied robustness properties for these classes of estimators and test statistics, showing that the minimum empirical divergence estimators of the parameter  $\theta_0$  of the model are generally not robust. This approach based on divergences and duality was initially used in the case of parametric models, the results being published in the articles, [2, 19, 20].

The classical EL estimator represents a particular case of the class of estimators from [3], namely, when using the modified Kullback-Leibler divergence. Although the EL estimator is superior to other above mentioned estimators in what regards higher-order asymptotic efficiency, this property is valid only in the case of the correct specification of the moment conditions. It is a known fact that the EL estimator and the EL ratio test for moment condition models are not robust with respect to the presence of outliers in the sample. Also, [17] showed that, when the support of the p.m. corresponding to the model and the orthogonality functions are not bounded, the EL estimator is not root  $n$  consistent under misspecification.

In this paper, we present a robust version of the EL estimator for moment condition models. This estimator is defined by minimizing an empirical version of the modified Kullback-Leibler divergence in dual form, using truncated orthogonality functions. For this estimator, we present some asymptotic properties regarding both consistency and limit laws. The robust EL estimator is root  $n$  consistent, even under misspecification, which gives a solution to the problem noticed by [17] for the EL estimator.

## 2 A Robust Version of the Empirical Likelihood Estimator

Let  $\{P_\theta; \theta \in \Theta\}$  be a reference identifiable model, containing probability measures such that, for each  $\theta \in \Theta$ ,  $P_\theta \in \mathcal{M}_\theta$ , meaning that  $\int_{\mathbb{R}^m} g(x, \theta) dP_\theta(x) = 0$ ,

and  $\theta$  is the unique solution of the equation. We assume that the p.m.  $P_0$  of the data, corresponding to the true unknown value  $\theta_0$  of the parameter to be estimated, belongs to this reference model. The reference model will be associated to the truncated orthogonality function  $g_c$ , defined hereafter, that will be used in the definition of the robust version of the EL estimator of the parameter  $\theta_0$ . We use the notation  $\|\cdot\|$  for the Euclidean norm. Similarly as in [16], using the reference model  $\{P_\theta; \theta \in \Theta\}$ , define the function  $g_c : \mathbb{R}^m \times \Theta \rightarrow \mathbb{R}^\ell$ ,

$$g_c(x, \theta) := H_c(A_\theta [g(x, \theta) - \tau_\theta]), \tag{2}$$

where  $H_c : \mathbb{R}^\ell \rightarrow \mathbb{R}^\ell$  is the Huber's function

$$H_c(y) := \begin{cases} y \cdot \min\left(1, \frac{c}{\|y\|}\right) & \text{if } y \neq 0, \\ 0 & \text{if } y = 0, \end{cases} \tag{3}$$

and  $A_\theta, \tau_\theta$  are determined by the solutions of the system of implicit equations

$$\begin{cases} \int g_c(x, \theta) dP_\theta(x) = 0, \\ \int g_c(x, \theta) g_c(x, \theta)^\top dP_\theta(x) = I_\ell, \end{cases} \tag{4}$$

where  $I_\ell$  is the  $\ell \times \ell$  identity matrix and  $c > 0$  is a given positive constant. Therefore, we have  $\|g_c(x, \theta)\| \leq c$ , for all  $x$  and  $\theta$ . We also use the function

$$h_c(x, \theta, A, \tau) := H_c(A [g(x, \theta) - \tau]), \tag{5}$$

when needed to work with the dependence on the  $\ell \times \ell$  matrix  $A$  and on the  $\ell$ -dimensional vector  $\tau$ . Then,

$$g_c(x, \theta) = h_c(x, \theta, A_\theta, \tau_\theta), \tag{6}$$

where  $A_\theta$  and  $\tau_\theta$  are the solution of (4). For given  $P_\theta$  from the reference model, the triplet  $(\theta, A_\theta, \tau_\theta)$  is the unique solution of the system

$$\begin{cases} \int g(x, \theta) dP_\theta(x) = 0, \\ \int g_c(x, \theta) dP_\theta(x) = 0, \\ \int g_c(x, \theta) g_c(x, \theta)^\top dP_\theta(x) = I_\ell. \end{cases} \tag{7}$$

The uniqueness is justified in [16], p. 48.

In what follows, we will use the so-called modified Kullback-Leibler divergence between probability measures, say  $Q$  and  $P$ , defined by

$$KL_m(Q, P) := \int_{\mathbb{R}^m} \varphi\left(\frac{dQ}{dP}(x)\right) dP(x), \tag{8}$$

if  $Q$  is absolutely continuous with respect to  $P$ , and  $KL_m(Q, P) := +\infty$ , elsewhere. The strictly convex function  $\varphi$  is defined by  $\varphi(x) := -\log x + x - 1$ , if  $x > 0$ , respectively  $\varphi(x) := +\infty$ , if  $x \leq 0$ . Straightforward calculus show that

the convex conjugate<sup>1</sup> of the convex function  $\varphi$  is  $\psi(u) = -\log(1 - u)$  if  $u < 1$ , respectively  $\psi(u) = +\infty$ , if  $u \geq 1$ . Recall that the Kullback-Leibler divergence, between any probability measures  $Q$  and  $P$ , is defined by

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

if  $Q$  is absolutely continuous with respect to  $P$ , and  $KL_m(Q, P) := +\infty$ , elsewhere. Here, the strictly convex function  $\varphi$  is defined by  $\varphi(x) = x \log x - x + 1$ , if  $x \geq 0$ , and  $\varphi(x) = +\infty$ , if  $x < 0$ . Notice also that  $KL_m(Q, P) = KL(P, Q)$ , for all probability measures  $Q$  and  $P$ . For any subset  $\Omega$  of  $M$ , we define the  $KL_m$ -divergence, between  $\Omega$  and any probability measure  $P$ , by

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

Define the moment condition model

$$\mathcal{M}_c := \bigcup_{\theta \in \Theta} \mathcal{M}_{c,\theta} := \bigcup_{\theta \in \Theta} \left\{ Q \in M \text{ s.t. } \int_{\mathbb{R}^m} g_c(x, \theta) dQ(x) = 0 \right\}. \tag{9}$$

For any  $\theta \in \Theta$ , define the set

$$\Lambda_{c,\theta} := \Lambda_{c,\theta}(P_0) := \left\{ t \in \mathbb{R}^\ell \text{ s.t. } \int_{\mathbb{R}^m} |\psi(t^\top g_c(x, \theta))| dP_0(x) < \infty \right\}.$$

Since  $g_c(x, \theta)$  is bounded (with respect to  $x$ ), then on the basis of Theorem 1.1 in [1] and Proposition 4.2 in [3], the following dual representation of divergence holds

$$KL_m(\mathcal{M}_{c,\theta}, P_0) = \sup_{t \in \Lambda_{c,\theta}} \int_{\mathbb{R}^m} m_c(x, \theta, t) dP_0(x), \tag{10}$$

where

$$m_c(x, \theta, t) := -\psi(t^\top g_c(x, \theta)) = \log(1 - t^\top g_c(x, \theta)), \tag{11}$$

and the supremum in (10) is reached, provided that  $KL_m(\mathcal{M}_{c,\theta}, P_0)$  is finite. Moreover, the supremum in (10) is unique under the following assumption

$$P_0 \left( \{x \in \mathbb{R}^m \text{ s.t. } \bar{t}^\top \bar{g}_c(x, \theta) \neq 0\} \right) > 0, \text{ for all } \bar{t} \in \mathbb{R}^{1+\ell} \setminus \{0\}, \tag{12}$$

where  $\bar{t} := (t_0, t_1, \dots, t_\ell)^\top$  and  $\bar{g}_c := (g_0, g_{c,1}, \dots, g_{c,\ell})^\top$ . This last condition is satisfied if the functions  $g_0(\cdot) := \mathbf{1}_{\mathbb{R}^m}(\cdot), g_{c,1}(\cdot, \theta), \dots, g_{c,\ell}(\cdot, \theta)$  are linearly independent and  $P_0$  is not degenerate. The empirical measure, associated to the sample  $X_1, \dots, X_n$ , is defined by

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

---

<sup>1</sup> The convex conjugate, called also Fenchel-Legendre transform, of  $\varphi$ , is the function defined on  $\mathbb{R}$  by  $\psi(u) := \sup_{x \in \mathbb{R}} \{ux - \varphi(x)\}, \forall u \in \mathbb{R}$ .

$\delta_x(\cdot)$  being the Dirac measure at the point  $x$ . Denote

$$\begin{aligned} \Lambda_{c,\theta,n} &:= \Lambda_{c,\theta}(P_n) = \left\{ t \in \mathbb{R}^\ell \text{ s.t. } \int_{\mathbb{R}^m} |\psi(t^\top g_c(x, \theta))| dP_n(x) < \infty \right\} \\ &= \left\{ t \in \mathbb{R}^\ell \text{ s.t. } \frac{1}{n} \sum_{i=1}^n \left| \log \left( 1 - \sum_{j=1}^\ell t_j g_{c,j}(X_i, \theta) \right) \right| < \infty \right\}. \end{aligned} \tag{13}$$

In view of relation (10), for given  $\theta \in \Theta$ , a natural estimator of

$$t_{c,\theta} := \arg \sup_{t \in \Lambda_{c,\theta}} \int m_c(x, \theta, t) dP_0(x), \tag{14}$$

can be defined by “plug-in” as follows

$$\widehat{t}_{c,\theta} := \arg \sup_{t \in \Lambda_{c,\theta,n}} \int m_c(x, \theta, t) dP_n(x). \tag{15}$$

A “dual” plug-in estimator of the modified Kullback-Leibler divergence, between  $\mathcal{M}_{c,\theta}$  and  $P_0$ , can then be defined by

$$\begin{aligned} \widehat{KL}_m(\mathcal{M}_{c,\theta}, P_0) &:= \sup_{t \in \Lambda_{c,\theta,n}} \int m_c(x, \theta, t) dP_n(x) \\ &= \sup_{(t_1, \dots, t_\ell) \in \mathbb{R}^\ell} \left\{ \frac{1}{n} \sum_{i=1}^n \overline{\log} \left( 1 - \sum_{j=1}^\ell t_j g_{c,j}(X_i, \theta) \right) \right\}, \end{aligned} \tag{16}$$

where  $\overline{\log}(\cdot)$  is the extended logarithm function, i.e., the function defined by  $\overline{\log}(u) = \log(u)$  if  $u > 0$ , and  $\overline{\log}(u) = -\infty$  if  $u \leq 0$ . Hence,

$$KL_m(\mathcal{M}_c, P_0) := \inf_{\theta \in \Theta} KL_m(\mathcal{M}_{c,\theta}, P_0) \tag{17}$$

can be estimated by

$$\begin{aligned} \widehat{KL}_m(\mathcal{M}_c, P_0) &:= \inf_{\theta \in \Theta} \widehat{KL}_m(\mathcal{M}_{c,\theta}, P_0) \\ &= \inf_{\theta \in \Theta} \sup_{(t_1, \dots, t_\ell) \in \mathbb{R}^\ell} \left\{ \frac{1}{n} \sum_{i=1}^n \overline{\log} \left( 1 - \sum_{j=1}^\ell t_j g_{c,j}(X_i, \theta) \right) \right\}. \end{aligned} \tag{18}$$

Since  $\theta_0 = \arg \inf_{\theta \in \Theta} KL_m(\mathcal{M}_{c,\theta}, P_0)$ , where the infimum is unique, we propose then to estimate  $\theta_0$  by

$$\begin{aligned} \widehat{\theta}_c &:= \arg \inf_{\theta \in \Theta} \sup_{t \in \Lambda_{c,\theta,n}} \int m_c(x, \theta, t) dP_n(x) \\ &= \arg \inf_{\theta \in \Theta} \sup_{(t_1, \dots, t_\ell) \in \mathbb{R}^\ell} \left\{ \frac{1}{n} \sum_{i=1}^n \overline{\log} \left( 1 - \sum_{j=1}^\ell t_j g_{c,j}(X_i, \theta) \right) \right\}, \end{aligned} \tag{19}$$

which can be seen as a “robust” version of the classical EL estimator. Recall that the EL estimator, see e.g. [14], can be written as

$$\hat{\theta} = \arg \inf_{\theta \in \Theta} \sup_{(t_1, \dots, t_\ell) \in \mathbb{R}^\ell} \left\{ \frac{1}{n} \sum_{i=1}^n \overline{\log} \left( 1 - \sum_{j=1}^{\ell} t_j g_j(X_i, \theta) \right) \right\}.$$

A slightly different definition of an estimator for the parameter  $\theta_0$  was introduced in [10], where robustness and consistency properties are also stated. However, the limiting distribution of the estimator in [10] is not standard, and not easy to be obtained, due to the fact that the used bounded orthogonality functions depend on both  $\theta$  and the data. The present version is simpler and does not present this difficulty. We give in the following sections, the influence function of the estimator (19), and state both consistency and the limiting distributions of all the proposed estimators (15), (16), (19) and (18).

### 2.1 Robustness Property

The classical EL estimator of the parameter  $\theta_0$  of a moment condition model can be obtained as a particular case of the class of minimum empirical divergence estimators introduced by [3]. [18] showed that the influence functions for the estimators from this class, so particularly the influence function of the EL estimator, are each proportional to the orthogonality function  $g(x, \theta_0)$  of the model. These influence functions also coincide with the influence function of the GMM estimator obtained by [16]. Therefore, when  $g(x, \theta)$  is not bounded in  $x$ , all these estimators, and particularly the EL estimator of  $\theta_0$ , are not robust.

Denote  $T_c(\cdot)$  the statistical functional associated to the estimator  $\hat{\theta}_c$ , so that  $\hat{\theta}_c = T_c(P_n)$ . The influence function  $IF(x; T_c, P_0)$  of  $T_c$  at  $P_0$  is defined by

$$IF(x; T_c, P_0) := \left. \frac{\partial}{\partial \varepsilon} T_c(P_{\varepsilon, x}) \right|_{\varepsilon=0},$$

where  $P_{\varepsilon, x}(\cdot) = (1 - \varepsilon) P_0(\cdot) + \varepsilon \delta_x(\cdot)$ ,  $\varepsilon \in ]0, 1[$ ; see e.g. [5]. The influence function  $IF(x; T_c, P_0)$  of the estimator  $\hat{\theta}_c$  presented in this paper is linearly related to the bounded function  $g_c(x, \theta)$ , more precisely, the following result holds

$$IF(x; T_c, P_0) = \left\{ \left[ \int \frac{\partial}{\partial \theta} g_c(y, \theta_0) dP_0(y) \right]^\top \int \frac{\partial}{\partial \theta} g_c(y, \theta_0) dP_0(y) \right\}^{-1} \cdot \left[ \int \frac{\partial}{\partial \theta} g_c(y, \theta_0) dP_0(y) \right]^\top g_c(x, \theta_0),$$

which implies the robustness of the estimator  $\hat{\theta}_c$  of the parameter  $\theta_0$ . The proof of this result is similar to the one presented in [10].

## 2.2 Asymptotic Properties

In this subsection, we give the limiting distributions of the proposed estimators, under some regularity assumptions similar to those used by [3].

**Proposition 1.** *For any fixed  $\theta \in \Theta$ , under some regularity assumptions, we have*

- (1)  $\sqrt{n}(\widehat{t}_{c,\theta} - t_{c,\theta})$  converges in distribution to a centered normal random vector;
- (2) If  $P_0 \notin \mathcal{M}_{c,\theta}$ , then  $\sqrt{n}(\widehat{KL}_m(\mathcal{M}_{c,\theta}, P_0) - KL_m(\mathcal{M}_{c,\theta}, P_0))$  converges in distribution to a centered normal random variable;
- (3) If  $P_0 \in \mathcal{M}_{c,\theta}$ , then  $2n\widehat{KL}_m(\mathcal{M}_{c,\theta}, P_0)$  converges in distribution to a  $\chi^2(\ell)$  random variable.

**Proposition 2.** *Under some regularity assumptions, we have*

- (1)  $\sqrt{n}(\widehat{\theta}_c - \theta_0)$  converges in distribution to a centered normal random vector;
- (2) If  $\ell > d$ , then  $2n\widehat{KL}_m(\mathcal{M}_c, P_0)$  converges in distribution to a  $\chi^2(\ell - d)$  random variable.

## References

1. Broniatowski, M., Keziou, A.: Minimization of  $\phi$ -divergences on sets of signed measures. *Stud. Sci. Math. Hung.* **43**(4), 403–442 (2006)
2. Broniatowski, M., Keziou, A.: Parametric estimation and tests through divergences and the duality technique. *J. Multivar. Anal.* **100**(1), 16–36 (2009)
3. Broniatowski, M., Keziou, A.: Divergences and duality for estimation and test under moment condition models. *J. Stat. Plann. Infer.* **142**(9), 2554–2573 (2012)
4. Felipe, A., Martin, N., Miranda, P., Pardo, L.: Testing with exponentially tilted empirical likelihood. *Methodol. Comput. Appl. Probab.* **20**, 1–40 (2018)
5. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A.: *Robust Statistics: The Approach Based on Influence Functions*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Wiley, New York (1986)
6. Hansen, L.P.: Large sample properties of generalized method of moments estimators. *Econometrica* **50**(4), 1029–1054 (1982)
7. Hansen, L.P., Heaton, J., Yaron, A.: Finite-sample properties of some alternative generalized method of moments estimators. *J. Bus. Econ. Stat.* **14**, 262–280 (1996)
8. Imbens, G.W.: One-step estimators for over-identified generalized method of moments models. *Rev. Econ. Stud.* **64**(3), 359–383 (1997)
9. Imbens, G.W.: Generalized method of moments and empirical likelihood. *J. Bus. Econ. Stat.* **20**(4), 493–506 (2002)
10. Keziou, A., Toma, A.: A robust version of the empirical likelihood estimator. *Mathematics* **9**(8), 829 (2021)
11. Kitamura, Y., Stutzer, M.: An information-theoretic alternative to generalized method of moments estimation. *Econometrica* **65**(4), 861–874 (1997)
12. Lô, S.N., Ronchetti, E.: Robust small sample accurate inference in moment condition models. *Comput. Stat. Data Anal.* **56**(11), 3182–3197 (2012)

13. Newey, W.K., Smith, R.J.: Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72**(1), 219–255 (2004)
14. Owen, A.: *Empirical likelihood*. Chapman and Hall, New York (2001)
15. Qin, J., Lawless, J.: Empirical likelihood and general estimating equations. *Ann. Stat.* **22**(1), 300–325 (1994)
16. Ronchetti, E., Trojani, F.: Robust inference with GMM estimators. *J. Econ.* **101**(1), 37–69 (2001)
17. Schennach, S.M.: Point estimation with exponentially tilted empirical likelihood. *Ann. Stat.* **35**(2), 634–672 (2007)
18. Toma, A.: Robustness of dual divergence estimators for models satisfying linear constraints. *Comptes Rendus Mathématiques* **351**(7–8), 311–316 (2013)
19. Toma, A., Leoni-Aubin, S.: Robust tests based on dual divergence estimators and saddlepoint approximations. *J. Multivar. Anal.* **101**(5), 1143–1155 (2010)
20. Toma, A., Broniatowski, M.: Dual divergence estimators and tests: robustness results. *J. Multivar. Anal.* **102**(1), 20–36 (2011)