

Article

A Robust Version of the Empirical Likelihood Estimator

Amor Keziou ^{1,†} and Aida Toma ^{2,3,*}

¹ Laboratoire de Mathématiques de Reims, UMR9008 CNRS et Université de Reims Champagne-Ardenne, UFR SEN, Moulin de la Housse, B.P. 1039, 51687 Reims, France; amor.keziou@univ-reims.fr

² Department of Applied Mathematics, Bucharest University of Economic Studies, Piața Romană no. 6, 010374 Bucharest, Romania

³ “Gheorghe Mihoc-Caius Iacob” Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, Calea 13 Septembrie no. 13, 050711 Bucharest, Romania

* Correspondence: aida.toma@csie.ase.ro

† These authors contributed equally to this work.

Abstract: In this paper, we introduce 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. We prove the robustness and the consistency of the new estimator. The performance of the robust empirical likelihood estimator is illustrated through examples based on Monte Carlo simulations.

Keywords: moment condition models; estimation; robustness; empirical likelihood

1. Introduction

A moment condition model is a family $\mathcal{M}^{(1)}$ of probability measures (p.m.), all defined on the same measurable space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, such that (s.t.)

$$\int_{\mathbb{R}^m} g(x, \theta) dQ(x) = 0 \quad \text{for all } Q \in \mathcal{M}^{(1)},$$

where $\mathcal{B}(\mathbb{R}^m)$ is the Borel σ -field. The parameter of interest θ belongs to 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 $\mathbb{R}^m \times \Theta$, each component g_i being real valued function. Denote by $M^{(1)}$ the set of all probability measures on $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, and for each $\theta \in \Theta$ define

$$\mathcal{M}_\theta^{(1)} := \left\{ Q \in M^{(1)} \text{ s.t. } \int_{\mathbb{R}^m} g(x, \theta) dQ(x) = 0 \right\},$$

so that

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

Let X_1, \dots, X_n be an i.i.d. sample with unknown p.m. P_0 . We assume that the equation $\int g(x, \theta) dP_0(x) = 0$ has a unique solution (in θ), which will be denoted θ_0 . We consider the estimation problem of θ_0 from the data X_1, \dots, X_n . The traditional way of estimating the parameter θ_0 is given by the Generalized Method of Moments (GMM) [1]. The GMM estimators are consistent and have asymptotically normal distribution. Despite their desirable asymptotic properties, the finite sample performance of the GMM estimators is not satisfactory. Some other alternative methods have been proposed in the literature. The Continuous Updating (CU) estimator [2], the Empirical Likelihood (EL) estimator [3–5], and the Exponential Tilting (ET) [6], are three of the most known examples. Imbens [7] showed that EL and ET estimators are characterized by lower bias than GMM in nonlinear models. Newey and Smith [8] studied the theoretical properties of EL, ET, and CU estimators, by including them into the Generalized Empirical Likelihood (GEL) family of



Citation: Keziou, A.; Toma, A. A Robust Version of the Empirical Likelihood Estimator. *Mathematics* **2021**, *9*, 829. <https://doi.org/10.3390/math9080829>

Academic Editor: Vlad Stefan Barbu

Received: 1 March 2021

Accepted: 7 April 2021

Published: 10 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

estimators, and showed that all GEL estimators are characterized by lower asymptotic bias than GMM. The information and entropy econometric (IEE) techniques have been proposed in order to improve the finite sample performance of the GMM estimators and tests [4,6]. Ronchetti and Trojani [9] and Lo and Ronchetti [10] have proposed robust alternatives for GMM estimators and tests, respectively, for the IEE techniques, keeping also the finite sample accuracy. Felipe et al. [11] proposed empirical divergence test statistics based on exponentially tilted empirical likelihood estimator with good robustness properties. Broniatowski and Keziou [12] proposed a general approach for estimation and testing in moment condition models, which includes some of the above-mentioned methods. This approach, based on minimizing divergences with their dual forms, allows the asymptotic study of the estimators (called minimum empirical divergence estimators) and of the associated test statistics, both under the model and under mis-specification of the model. The approach based on divergences and duality was firstly considered in the case of parametric models, for example, in [13–15]. Applications of the minimum dual divergence estimators in model selection problems are considered in [16].

The EL paradigm enters as a special case of the general methodology from Broniatowski and Keziou [12], namely, when using the modified Kullback–Leibler divergence. Although the EL estimator is preferable to other estimators due to higher-order asymptotic properties, these properties are valid only in the case of the correct specification of the moment conditions. On the other hand, when the support of the p.m. corresponding to the model and the orthogonality functions are not bounded, the EL estimator may cease to be root n consistent (Schennach [17]) under mis-specification. It is a known fact that the EL estimator for moment condition models is not robust. This fact is also justified by the results from [18], where it is shown that the influence function of a minimum dual divergence estimator, and particularly that of the EL estimator, is linear related to the orthogonality function g corresponding to the model. Thus, the influence function of the EL estimator is bounded if and only if the function g corresponding to the underlying model is bounded. Hence, the EL estimator is usually not robust, since the function g could be unbounded on the observation. For this reason, in practice, the classical EL estimator, as well as the minimum dual divergence estimators and also GMM estimators, is unstable even under small deviations from the assumed model.

As examples in this context, we mention models for which the orthogonality functions are unbounded [9]. The autoregressive models with heteroscedastic errors [19] can be written under the form of moment condition models, but the orthogonality functions defining the orthogonality conditions are unbounded. Moreover, the nonlinear empirical asset pricing models [20] can be written under the form of moment condition models and have natural orthogonality conditions (given by the given asset pricing equations), which are given by unbounded orthogonality functions. We also recall the following classical example, which is used in the last section of the paper, in the Monte Carlo simulation study.

Example 1 ([3] p. 302). *Consider a random variable X with unbounded support (\mathbb{R} or \mathbb{R}_+ for instance). Let $\mathbb{E}(X) = \theta$, and assume that $\mathbb{E}(X^2) = h(\theta)$, with $h(\cdot)$ being a known function. The aim is to estimate the parameter θ using an i.i.d. sample X_1, \dots, X_n of X . The information on the probability distribution P_0 of X can be expressed in the context of model (1), with $d = 1$ and $\ell = 2$, by taking $g(x, \theta) =: (g_1(x, \theta), g_2(x, \theta))^T = (x - \theta, x^2 - h(\theta))^T$. One can see that the orthogonality functions $g_1(\cdot, \theta)$ and $g_2(\cdot, \theta)$ are unbounded (with respect to x).*

For such models, the lack of robustness of the EL estimator, as well as the lack of robustness of other classical estimators, represents the motivation to study some robust alternatives.

In the present paper, we propose a locally robust version of the EL estimator for moment condition models. Locally robust in the sense that the functional associated with this estimator is locally approximated by the means of the influence function, and then, the boundedness of the influence function will imply the fact that, in the neighborhood of the model, the asymptotic bias of the estimator cannot become arbitrarily large (see [21]).

The new estimator is defined by minimization of an empirical version of the modified Kullback–Leibler divergence in dual form, using truncated orthogonality functions. This leads to a robust EL estimate. Moreover, we prove the consistency of this estimator. Finally, we present an example based on Monte Carlo simulations illustrating the performance of the robust EL estimator in the case of contaminated data.

2. A Robust Version of the Empirical Likelihood Estimator

2.1. Statistical Divergences

Let φ be a convex function on \mathbb{R} onto $[0, +\infty]$ valued, satisfying $\varphi(1) = 0$. Let P be some p.m. on the measurable space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$. For any signed finite measure Q , on the same measurable space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, absolutely continuous with respect to P , the φ -divergence (sometimes we simply say divergence) between Q and P is defined by

$$D_\varphi(Q, P) := \int_{\mathbb{R}^m} \varphi\left(\frac{dQ}{dP}(x)\right) dP(x), \tag{2}$$

where $\frac{dQ}{dP}$ is the Radon–Nikodym derivative. When Q is not absolutely continuous with respect to P , we set $D_\varphi(Q, P) = +\infty$. This definition extends the one given in [22] for divergences between p.m.’s. A known class of divergences between p.m.’s is the class of Cressie–Read divergences, introduced in [23] and defined by the functions

$$\varphi_\gamma : x \in \mathbb{R}_+ \mapsto \varphi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)}, \tag{3}$$

for $\gamma \in \mathbb{R} \setminus \{0, 1\}$, $\varphi_0(x) := -\log x + x - 1$ and $\varphi_1(x) := x \log x - x + 1$. For any $\gamma \in \mathbb{R}$, if $\varphi_\gamma(0)$ is not defined, we set $\varphi_\gamma(0) := \lim_{x \rightarrow 0^+} \varphi_\gamma(x)$, which may be finite or infinite. The Kullback–Leibler divergence (KL) is associated to φ_1 , the modified Kullback–Leibler (KL_m) to φ_0 , the χ^2 divergence to φ_2 , the modified χ^2 divergence (χ_m^2) to φ_{-1} and the Hellinger (H) distance to $\varphi_{1/2}$. The φ -divergence between some set Ω of probability measures and a probability measure P is defined by

$$D_\varphi(\Omega, P) := \inf_{Q \in \Omega} D_\varphi(Q, P). \tag{4}$$

2.2. Definition of the Estimator

We consider a reference identifiable model $\{P_\theta; \theta \in \Theta\}$ of probability measures such that, for each $\theta \in \Theta$, $P_\theta \in \mathcal{M}_\theta$, which means that $\int_{\mathbb{R}^m} g(x, \theta) dP_\theta(x) = 0$, and assume that θ is the unique solution of the equation. We assume that the p.m. P_0 of the data corresponding to the true unknown value θ_0 of the parameter to be estimated belongs to this reference model. The reference model will be associated with the truncated orthogonality function that will be used to define the robust version of the EL estimator of the parameter θ_0 . We will use the notation $\|\cdot\|$ for the Euclidean norm. Similarly as in [9], using the reference model $\{P_\theta; \theta \in \Theta\}$, we define the function $g^c : \mathbb{R}^m \times \Theta \rightarrow \mathbb{R}^\ell$,

$$g^c(x, \theta) := H_c(A_\theta [g(x, \theta) - \tau_\theta]), \tag{5}$$

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

$$H_c(y) := \begin{cases} y \cdot \min\left(1, \frac{c}{\|y\|}\right) & \text{if } y \neq 0, \\ 0 & \text{if } y = 0, \end{cases} \tag{6}$$

and A_θ, τ_θ are, respectively, $\ell \times \ell$ -matrix and ℓ -vector, defined as the solutions of the system of implicit equations

$$\begin{cases} \int g^c(x, \theta) dP_\theta(x) = 0, \\ \int g^c(x, \theta) g^c(x, \theta)^\top dP_0(x) = I_\ell, \end{cases} \tag{7}$$

where I_ℓ 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 θ . We also use the function

$$h^c(x, \theta, A, \tau) := H_c(A [g(x, \theta) - \tau]), \tag{8}$$

when needed to work with the dependence on the matrix A and on the vector τ . Therefore,

$$g^c(x, \theta) = h^c(x, \theta, A_\theta, \tau_\theta). \tag{9}$$

For given P_θ from the reference model, the triplet $(\theta, A_\theta, \tau_\theta)$ is the unique solution of the system

$$\begin{cases} \int g(x, \theta) dP_\theta(x) = 0, \\ \int g^c(x, \theta) dP_\theta(x) = 0, \\ \int g^c(x, \theta) g^c(x, \theta)^\top dP_\theta(x) = I_\ell \end{cases} \tag{10}$$

see [9], p. 48.

Consider the estimating problem of the triplet $(\theta_0, A_{\theta_0}, \tau_{\theta_0})$ on the basis of a sample $X_1, \dots, X_n \sim P_0, P_0 \in \mathcal{M}_{\theta_0}$. For each $\theta \in \Theta$, using the p.m. P_θ from the reference model, we define \hat{A}_θ and $\hat{\tau}_\theta$, solutions of the system

$$\begin{cases} \int h^c(x, \theta, \hat{A}_\theta, \hat{\tau}_\theta) dP_\theta(x) = 0, \\ \int h^c(x, \theta, \hat{A}_\theta, \hat{\tau}_\theta) h^c(x, \theta, \hat{A}_\theta, \hat{\tau}_\theta)^\top dP_n(x) = I_\ell, \end{cases} \tag{11}$$

where $P_n(\cdot)$ is the empirical measure associated with the sample,

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

with $\delta_x(\cdot)$ being the Dirac measure at the point x , for any x . We denote

$$g_n^c(x, \theta) := h^c(x, \theta, \hat{A}_\theta, \hat{\tau}_\theta) = H_c(\hat{A}_\theta [g(x, \theta) - \hat{\tau}_\theta]). \tag{12}$$

Note that $g_n^c(x, \theta)$ depends on both the data and the reference probability P_θ . We now consider the moment condition model associated to the function $g_n^c(x, \theta)$, namely,

$$\mathcal{M}^{c,n} := \bigcup_{\theta \in \Theta} \mathcal{M}_\theta^{c,n}, \tag{13}$$

where

$$\mathcal{M}_\theta^{c,n} := \left\{ Q \in M^{(1)} \text{ s.t. } \int g_n^c(x, \theta) dQ(x) = 0 \right\}, \forall \theta \in \Theta. \tag{14}$$

The p.m. P_0 belongs to $\mathcal{M}_{\theta_0}^{c,n}$.

In what follows, we consider the modified Kullback–Leibler divergence, which corresponds to the strictly convex function $\varphi(x) := -\log x + x - 1$, if $x > 0$, respectively $\varphi(x) := +\infty$, if $x \leq 0$. The convex conjugate, called also the Fenchel–Legendre transform, of any function $f : x \in \mathbb{R} \mapsto f(x) \in \overline{\mathbb{R}}$, is the function defined by $f^*(u) := \sup_{x \in \mathbb{R}} \{ux - f(x)\}$, for all $u \in \mathbb{R}$. A straightforward calculus shows that the convex conjugate of the convex function φ , denote it ψ , is given by $\psi(u) = -\log(1 - u)$ if $u < 1$, respectively $\psi(u) = +\infty$, if $u \geq 1$. For given $\theta \in \Theta$, we define the set

$$\overline{\Lambda}_\theta^{c,n}(P_0) := \left\{ \bar{t} := (t_0, t_1, \dots, t_\ell)^\top \in \mathbb{R}^{1+\ell} \text{ s.t. } \int |\psi(\bar{t}^\top \overline{g}_n^c(x, \theta))| dP_0(x) < \infty \right\}, \tag{15}$$

where $\overline{g}_n^c := (\mathbf{1}_{\mathbb{R}^m}, g_n^c)^\top$. We denote $\overline{\Lambda}_\theta^{c,n} := \overline{\Lambda}_\theta^{c,n}(P_0)$ and $\overline{\Lambda}_\theta^{c,n}(P_n) := \overline{\Lambda}_\theta^{c,n}(P_n)$.

Since $g_n^c(x, \theta)$ is bounded (with respect to x), on the basis of Theorem 1.1 in [24] and Proposition 4.2 in [12], the following dual representation of KL_m -divergence holds

$$KL_m(\mathcal{M}_\theta^{c,n}, P_0) = \sup_{\bar{t} \in \bar{\Lambda}_\theta^{c,n}} \int m_n^c(x, \theta, \bar{t}) dP_0(x), \tag{16}$$

where

$$m_n^c(x, \theta, \bar{t}) := t_0 - \psi(\bar{t}^\top \bar{g}_n^c(x, \theta)) = t_0 + \log(1 - \bar{t}^\top \bar{g}_n^c(x, \theta)), \tag{17}$$

and the supremum in (16) is reached, provided that $KL_m(\mathcal{M}_\theta^{c,n}, P_0)$ is finite. Note that $m_n^c(x, \theta, \bar{t})$ depends on the reference p.m. P_θ , since $g_n^c(x, \theta)$ depends on P_θ . We denote then $\bar{t}_\theta^{c,n} = \bar{t}_\theta^{c,n}(P_0)$ any vector such that

$$\bar{t}_\theta^{c,n} := \arg \sup_{\bar{t} \in \bar{\Lambda}_\theta^{c,n}} \int m_n^c(x, \theta, \bar{t}) dP_0(x). \tag{18}$$

Furthermore, according to Proposition 4.2 from [12], for each $\theta \in \Theta$, the condition

$$P_0(\{x \in \mathbb{R}^m \text{ s.t. } \bar{t}^\top \bar{g}_n^c(x, \theta) \neq 0\}) > 0, \text{ for all } \bar{t} \in \mathbb{R}^{1+\ell} \setminus \{0\}, \tag{19}$$

ensures that $\bar{t}_\theta^{c,n}$, defined as solution of optimization problem (18), is unique. Notice that the linear independence of the functions $\mathbf{1}_{\mathbb{R}^m}, g_{n,1}^c(\cdot, \theta), \dots, g_{n,\ell}^c(\cdot, \theta)$ implies condition (19) whenever P_0 is not degenerate.

Moreover, using again Proposition 4.2 and Remark 4.4 from [12], for each $\theta \in \Theta$, one can show that the first component of the optimal solution $\bar{t}_\theta^{c,n}$ in (18) equals to zero. One can then omit the first component of the vector \bar{t} in displays ((15)–(18)). Therefore, they will be replaced by

$$\Lambda_\theta^{c,n}(P_0) := \left\{ t := (t_1, \dots, t_\ell)^\top \in \mathbb{R}^\ell \text{ s.t. } \int |\psi(t^\top g_n^c(x, \theta))| dP_0(x) < \infty \right\}, \tag{20}$$

$$KL_m(\mathcal{M}_\theta^{c,n}, P_0) = \sup_{t \in \Lambda_\theta^{c,n}} \int m_n^c(x, \theta, t) dP_0(x), \tag{21}$$

where

$$m_n^c(x, \theta, t) := -\psi(t^\top g_n^c(x, \theta)) = \log(1 - t^\top g_n^c(x, \theta)) \tag{22}$$

and

$$t_\theta^{c,n} := \arg \sup_{t \in \Lambda_\theta^{c,n}} \int m_n^c(x, \theta, t) dP_0(x). \tag{23}$$

Denote

$$\Lambda_\theta^{c,n,n} := \Lambda_\theta^{c,n}(P_n) = \left\{ t := (t_1, \dots, t_\ell)^\top \in \mathbb{R}^\ell \text{ s.t. } \frac{1}{n} \sum_{i=1}^n \left| \log(1 - \sum_{j=1}^\ell t_j g_{n,j}^c(X_i, \theta)) \right| < \infty \right\}.$$

In view of relation (23), a natural estimator of $t_\theta^{c,n}$, is defined by

$$\hat{t}_\theta^c := \arg \sup_{t \in \Lambda_\theta^{c,n,n}} \int m_n^c(x, \theta, t) dP_n(x). \tag{24}$$

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

$$\begin{aligned} \widehat{KL}_m(\mathcal{M}_\theta^{c,n}, P_0) &:= \sup_{t \in \Lambda_\theta^{c,n,n}} \int m_n^c(x, \theta, t) dP_n(x) \\ &= \sup_{(t_1, \dots, t_\ell) \in \Lambda_\theta^{c,n,n}} \left\{ \int \log \left(1 - \sum_{j=1}^\ell t_j g_{n,j}^c(x, \theta) \right) dP_n(x) \right\} \\ &= \sup_{(t_1, \dots, t_\ell) \in \mathbb{R}^\ell} \left\{ \frac{1}{n} \sum_{i=1}^n \overline{\log} \left(1 - \sum_{j=1}^\ell t_j g_{n,j}^c(X_i, \theta) \right) \right\}, \end{aligned} \tag{25}$$

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 $u \leq 0$, for any $u \in \mathbb{R}$. Finally, we define the following estimator of θ_0

$$\begin{aligned} \widehat{\theta}^c &:= \arg \inf_{\theta \in \Theta} \sup_{t \in \Lambda_\theta^{c,n,n}} \int m_n^c(x, \theta, t) dP_n(x) \\ &= \arg \inf_{\theta \in \Theta} \sup_{(t_1, \dots, t_\ell) \in \mathbb{R}^\ell} \left\{ \frac{1}{n} \sum_{i=1}^n \overline{\log} \left(1 - \sum_{j=1}^\ell t_j g_{n,j}^c(X_i, \theta) \right) \right\}, \end{aligned} \tag{26}$$

which can be seen as a “robust” version of the well-known EL estimator.

Recall that the EL estimator can be written as (see, e.g., [5])

$$\widehat{\theta} = \arg \inf_{\theta \in \Theta} \sup_{(t_1, \dots, t_\ell) \in \mathbb{R}^\ell} \left\{ \frac{1}{n} \sum_{i=1}^n \overline{\log} \left(1 - \sum_{j=1}^\ell t_j g_j(X_i, \theta) \right) \right\}. \tag{27}$$

For establishing asymptotic properties of the proposed estimators, we need the following additional notations. Consider the moment condition model associated with the truncated function

$$g^c(x, \theta) = H_c(A_\theta[g(x, \theta) - \tau_\theta]),$$

where A_θ and τ_θ are the solution to the system (7). Note that $g^c(x, \theta)$ depends only on the reference model P_θ and not on the data. This model is defined by

$$\mathcal{M}^c := \bigcup_{\theta \in \Theta} \mathcal{M}_\theta^c, \tag{28}$$

where

$$\mathcal{M}_\theta^c := \left\{ Q \in M^{(1)} \text{ s.t. } \int g^c(x, \theta) dQ(x) = 0 \right\}, \forall \theta \in \Theta. \tag{29}$$

Let

$$\Lambda_\theta^c(P_0) := \left\{ t := (t_1, \dots, t_\ell)^\top \in \mathbb{R}^\ell \text{ s.t. } \int |\psi(t^\top g^c(x, \theta))| dP_0(x) < \infty \right\}. \tag{30}$$

Therefore, as above, we have the following dual representation for $KL_m(\mathcal{M}_\theta^c, P_0)$

$$KL_m(\mathcal{M}_\theta^c, P_0) = \sup_{t \in \Lambda_\theta^c} \int m^c(x, \theta, t) dP_0(x), \tag{31}$$

where

$$m^c(x, \theta, t) := -\psi(t^\top g^c(x, \theta)) = \log(1 - t^\top g^c(x, \theta)), \tag{32}$$

and the supremum in (31) is reached, provided that $KL_m(\mathcal{M}_\theta^c, P_0)$ is finite. Moreover, the supremum in (31) is unique under the following assumption

$$P_0\left(\{x \in \mathbb{R}^m \text{ s.t. } \bar{t}^\top g^c(x, \theta) \neq 0\}\right) > 0, \text{ for all } \bar{t} \in \mathbb{R}^{1+\ell} \setminus \{0\}, \tag{33}$$

which is satisfied if the functions $\mathbf{1}_{\mathbb{R}^m}, g_1^c(\cdot, \theta), \dots, g_\ell^c(\cdot, \theta)$ are linearly independent and P_0 is not degenerate. We denote then

$$t_\theta^c := \arg \sup_{t \in \Lambda_\theta^c} \int m^c(x, \theta, t) dP_0(x). \tag{34}$$

Finally, we have

$$\begin{aligned} \theta_0 &= \arg \inf_{\theta \in \Theta} KL_m(\mathcal{M}_\theta^c, P_0) \\ &= \arg \inf_{\theta \in \Theta} \sup_{t \in \Lambda_\theta^c} \int m^c(x, \theta, t) dP_0(x) = \arg \inf_{\theta \in \Theta} \int m^c(x, \theta, t_\theta^c) dP_0(x). \end{aligned} \tag{35}$$

We also use the function

$$n^c(x, \theta, t, A, \tau) := \log(1 - t^\top h^c(x, \theta, A, \tau)), \tag{36}$$

when needed to work with the dependence on matrix A and on vector τ . We have, then, with the above notation, $m^c(x, \theta, t) = n^c(x, \theta, t, A_\theta, \tau_\theta)$, where A_θ and τ_θ are the solution of the system of Equation (7).

2.3. Robustness Property

In order to prove the robustness of the estimator $\hat{\theta}^c$, we use the following well-known tools from the theory of robust statistics; see, e.g., [21]. A functional T , defined on a set of probability measures and parameter space valued, is called a statistical functional associated with an estimator $\hat{\theta}$ of the parameter θ from the model P_θ , if $\hat{\theta} = T(P_n)$. The influence function of T at P_θ is defined by

$$IF(x; T, P_\theta) := \left. \frac{\partial T(\tilde{P}_{\varepsilon x})}{\partial \varepsilon} \right|_{\varepsilon=0},$$

where $\tilde{P}_{\varepsilon x} := (1 - \varepsilon) P_\theta + \varepsilon \delta_x$. A natural robustness requirement on the statistical functional corresponding to the estimator is the boundedness of its influence function.

The statistical functional corresponding to the estimator $\hat{\theta}^c$ given by (26), is defined by

$$T^c(P) := \arg \inf_{\theta \in \Theta} \sup_{t \in \Lambda_\theta^c(P)} \int m^c(y, \theta, t, P) dP(y), \tag{37}$$

where

$$\Lambda_\theta^c(P) := \left\{ t \in \mathbb{R}^\ell \text{ s.t. } \int |\psi(t^\top g^c(x, \theta, P))| dP(x) < \infty \right\}, \tag{38}$$

$g^c(x, \theta, P) := h^c(x, \theta, A_\theta(P), \tau_\theta(P))$, with $A_\theta(P)$ and $\tau_\theta(P)$ solutions of the system

$$\begin{cases} \int h^c(x, \theta, A_\theta(P), \tau_\theta(P)) dP_\theta(x) = 0 \\ \int h^c(x, \theta, A_\theta(P), \tau_\theta(P)) h^c(x, \theta, A_\theta(P), \tau_\theta(P))^\top dP(x) = I_\ell \end{cases} \tag{39}$$

and

$$m^c(x, \theta, t, P) := -\psi(t^\top g^c(x, \theta, P)) = \log(1 - t^\top g^c(x, \theta, P)). \tag{40}$$

Note that, for a given θ , the function $g^c(x, \theta, P)$, as well as $m^c(x, \theta, t, P)$, both depend on the p.m. P_θ . In addition, note that $g^c(x, \theta, P_\theta)$ coincides with $g^c(x, \theta)$ defined in the preceding section. We denote

$$t_\theta^c(P) := \arg \sup_{t \in \Lambda_\theta^c(P)} \int m^c(y, \theta, t, P) dP(y). \tag{41}$$

Then

$$T^c(P) = \arg \inf_{\theta \in \Theta} \int m^c(y, \theta, t_\theta^c(P), P) dP(y). \tag{42}$$

Proposition 1. *The influence function of the estimator $\hat{\theta}^c$ is given by*

$$\begin{aligned} \text{IF}(x; T^c, P_0) &= \left\{ \left[\int \frac{\partial}{\partial \theta} g^c(y, \theta_0) dP_0(y) \right]^\top \int \frac{\partial}{\partial \theta} g^c(y, \theta_0) dP_0(y) \right\}^{-1} \\ &\cdot \left[\int \frac{\partial}{\partial \theta} g^c(y, \theta_0) dP_0(y) \right]^\top g^c(x, \theta_0). \end{aligned} \tag{43}$$

Proof. Using the definitions of $T^c(P)$ and $t_\theta^c(P)$, we have

$$\int \frac{\partial}{\partial t} m^c(y, \theta, t_\theta^c(P), P) dP(y) = 0 \tag{44}$$

$$\int \frac{\partial}{\partial t} m^c(y, T^c(P), t_{T^c(P)}^c(P), P) dP(y) = 0 \tag{45}$$

$$\int \frac{\partial}{\partial \theta} [m^c(y, \theta, t_\theta^c(P), P)]_{\theta=T^c(P)} dP(y) = 0. \tag{46}$$

Using (46), since

$$m^c(y, \theta, t_\theta^c(P), P) = \log \left(1 - t_\theta^c(P)^\top h^c(y, \theta, A_\theta(P), \tau_\theta(P)) \right), \tag{47}$$

$T^c(P)$ is solution of the equation

$$\left[\frac{\partial}{\partial \theta} t_\theta^c(P) \right]^\top \int \frac{\partial}{\partial t} m^c(y, \theta, t_\theta^c(P), P) dP(y) - \int \frac{\left[\frac{\partial}{\partial \theta} h^c(y, \theta, A_\theta(P), \tau_\theta(P)) \right]^\top t_\theta^c(P)}{1 - t_\theta^c(P)^\top h^c(y, \theta, A_\theta(P), \tau_\theta(P))} dP(y) = 0. \tag{48}$$

Since the first integral in the above display is 0, according to (44), the above equation simplifies to

$$\int \frac{\left[\frac{\partial}{\partial \theta} h^c(y, \theta, A_\theta(P), \tau_\theta(P)) \right]^\top t_\theta^c(P)}{1 - t_\theta^c(P)^\top h^c(y, \theta, A_\theta(P), \tau_\theta(P))} dP(y) = 0. \tag{49}$$

By replacing P with the contaminated model $\tilde{P}_{\varepsilon x} := (1 - \varepsilon)P_0 + \varepsilon\delta_x$ in Equation (49) and then derivating with respect to ε , the resulting equation, we obtain

$$\int \left[\frac{\partial}{\partial \theta} h^c(y, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0)) \right]^\top dP_0(y) \cdot \frac{\partial}{\partial \varepsilon} \left[t_{T^c(\tilde{P}_{\varepsilon x})}^c(\tilde{P}_{\varepsilon x}) \right]_{\varepsilon=0} = 0. \tag{50}$$

On the other hand,

$$\frac{\partial}{\partial \varepsilon} \left[t_{T^c(\tilde{P}_{\varepsilon x})}^c(\tilde{P}_{\varepsilon x}) \right]_{\varepsilon=0} = \frac{\partial}{\partial \theta} t_{\theta_0}^c(P_0) \cdot \text{IF}(x; T^c, P_0) + \text{IF}(x; t_{\theta_0}^c, P_0). \tag{51}$$

Combining (50) and (51), we get

$$\int \left[\frac{\partial}{\partial \theta} h^c(y, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0)) \right]^\top dP_0(y) \cdot \left\{ \frac{\partial}{\partial \theta} t_{\theta_0}^c(P_0) \cdot \text{IF}(x; T^c, P_0) + \text{IF}(x; t_{\theta_0}^c, P_0) \right\} = 0. \tag{52}$$

Differentiating with respect to θ of (44) leads to

$$\frac{\partial}{\partial \theta} t_{\theta_0}^c(P_0) = - \left[\int \frac{\partial^2}{\partial^2 t} m^c(y, \theta_0, t_{\theta_0}^c(P_0), P_0) dP_0(y) \right]^{-1} \int \frac{\partial}{\partial \theta \partial t} m^c(y, \theta_0, t_{\theta_0}^c(P_0), P_0) dP_0(y). \tag{53}$$

Some simple calculations yield

$$\frac{\partial}{\partial t} m^c(y, \theta, t, P) = - \frac{h^c(y, \theta, A_{\theta}(P), \tau_{\theta}(P))}{1 - t^{\top} h^c(y, \theta, A_{\theta}(P), \tau_{\theta}(P))} \tag{54}$$

$$\frac{\partial^2}{\partial^2 t} m^c(y, \theta, t, P) = - \frac{h^c(y, \theta, A_{\theta}(P), \tau_{\theta}(P)) h^c(y, \theta, A_{\theta}(P), \tau_{\theta}(P))^{\top}}{(1 - t^{\top} h^c(y, \theta, A_{\theta}(P), \tau_{\theta}(P)))^2}$$

and

$$\begin{aligned} \frac{\partial}{\partial \theta \partial t} m^c(y, \theta, t, P) &= \frac{- \frac{\partial}{\partial \theta} h^c(y, \theta, A_{\theta}(P), \tau_{\theta}(P)) (1 - t^{\top} h^c(y, \theta, A_{\theta}(P), \tau_{\theta}(P)))}{(1 - t^{\top} h^c(y, \theta, A_{\theta}(P), \tau_{\theta}(P)))^2} \\ &+ \frac{h^c(y, \theta, A_{\theta}(P), \tau_{\theta}(P)) (1 - t^{\top} \frac{\partial}{\partial \theta} h^c(y, \theta, A_{\theta}(P), \tau_{\theta}(P)))}{(1 - t^{\top} h^c(y, \theta, A_{\theta}(P), \tau_{\theta}(P)))^2}. \end{aligned}$$

Then, we obtain

$$\int \frac{\partial^2}{\partial^2 t} m^c(y, \theta_0, t_{\theta_0}^c(P_0), P_0) dP_0(y) = -I_{\ell} \tag{55}$$

and

$$\int \frac{\partial}{\partial \theta \partial t} m^c(y, \theta_0, t_{\theta_0}^c(P_0), P_0) dP_0(y) = - \int \frac{\partial}{\partial \theta} h^c(y, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0)) dP_0(y). \tag{56}$$

By replacing in (53), we obtain

$$\frac{\partial}{\partial \theta} t_{\theta_0}^c(P_0) = - \int \frac{\partial}{\partial \theta} h^c(y, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0)) dP_0(y). \tag{57}$$

Using (52) and (57), we obtain

$$\begin{aligned} & - \int \left[\frac{\partial}{\partial \theta} h^c(y, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0)) \right]^{\top} dP_0(y) \\ & \cdot \int \frac{\partial}{\partial \theta} h^c(y, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0)) dP_0(y) \cdot \text{IF}(x; T^c, P_0) \\ & + \int \frac{\partial}{\partial \theta} h^c(y, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0))^{\top} dP_0(y) \cdot \text{IF}(x; t_{\theta_0}^c, P_0) = 0. \end{aligned} \tag{58}$$

In order to calculate $\text{IF}(x; t_{\theta}^c, P_0)$, we first write Equation (44) for $P = \tilde{P}_{\epsilon x}$ and also use (54). Thus, we obtain

$$\begin{aligned} (1 - \epsilon) \int \frac{h^c(y, \theta, A_{\theta}(\tilde{P}_{\epsilon x}), \tau_{\theta}(\tilde{P}_{\epsilon x}))}{1 - t_{\theta}^c(\tilde{P}_{\epsilon x})^{\top} h^c(y, \theta, A_{\theta}(\tilde{P}_{\epsilon x}), \tau_{\theta}(\tilde{P}_{\epsilon x}))} dP_0(y) \\ + \epsilon \frac{h^c(x, \theta, A_{\theta}(\tilde{P}_{\epsilon x}), \tau_{\theta}(\tilde{P}_{\epsilon x}))}{1 - t_{\theta}^c(\tilde{P}_{\epsilon x})^{\top} h^c(x, \theta, A_{\theta}(\tilde{P}_{\epsilon x}), \tau_{\theta}(\tilde{P}_{\epsilon x}))} = 0. \end{aligned} \tag{59}$$

By differentiation with respect to ε , and taking $\theta = \theta_0$, we get

$$\int \frac{\partial}{\partial \varepsilon} [h^c(y, \theta_0, A_{\theta_0}(\tilde{P}_{\varepsilon x}), \tau_{\theta_0}(\tilde{P}_{\varepsilon x}))]_{\varepsilon=0} dP_0(y) - \int h^c(y, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0)) h^c(y, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0))^\top dP_0(y) \text{IF}(x; t_{\theta_0}^c, P_0) + h^c(x, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0)) = 0.$$

Consequently,

$$\text{IF}(x; t_{\theta_0}^c, P_0) = \left\{ \int h^c(y, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0)) h^c(y, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0))^\top dP_0(y) \right\}^{-1} \cdot h^c(x, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0)) = h^c(x, \theta_0, A_{\theta_0}(P_0), \tau_{\theta_0}(P_0)). \tag{60}$$

By combining (58) with (60), we obtain (43). \square

Remark 1. The classical empirical likelihood estimator of the parameter θ_0 of the moment condition model can be obtained as a particular case of the class of minimum empirical divergence estimators introduced by Broniatowski and Keziou [12]. Toma [18] proved that, in the case when P_0 belongs to the model $\mathcal{M}^{(1)}$, the influence functions for the estimators from this class, so particularly the influence function of the EL estimator, are all of the form

$$\begin{aligned} \text{IF}(x; T, P_0) = & - \left\{ \left[\int \frac{\partial}{\partial \theta} g(y, \theta_0) dP_0(y) \right]^\top \left[\int g(y, \theta_0) g(y, \theta_0)^\top dP_0(y) \right]^{-1} \right. \\ & \cdot \left. \left[\int \frac{\partial}{\partial \theta} g(y, \theta_0) dP_0(y) \right] \right\}^{-1} \cdot \left[\int \frac{\partial}{\partial \theta} g(y, \theta_0) dP_0(y) \right]^\top \\ & \cdot \left[\int g(y, \theta_0) g(y, \theta_0)^\top dP_0(y) \right]^{-1} g(x, \theta_0), \end{aligned} \tag{61}$$

irrespective of the used divergence. This influence function also coincides with the influence function of the GMM estimator obtained by Ronchetti and Trojani [9] and is linearly related to the function $g(x, \theta)$ of the model. When the orthogonality function $g(x, \theta)$ is not bounded in x , the minimum empirical divergence estimators, and particularly the EL estimator of θ_0 , are not robust. For many moment condition models, the orthogonality functions are linear and hence unbounded; therefore, these estimation methods are generally not robust. This is also the case of other known estimators, such as the least squares estimators, the GMM estimators, and the exponential tilting estimator for moment condition models. Instead, for the new estimator defined in the present paper, the influence function is linearly related to the function $g^c(x, \theta)$, which is bounded; therefore, this estimator can be seen as a robust version of the classical EL estimator.

An important feature of the robust version of the EL estimator is that its robustness can be controlled by a positive constant c . This constant appears in the Huber function used. In addition, an advantage of this approach based on using the Huber function is that we can require the bound of influence function of the estimator to be satisfied in a norm that is self-standardized with respect to the covariance matrix of the estimator; this norm measures the influence of the estimator relative to its variability expressed by its covariance matrix. Such an approach is also suitable to induce stable testing procedures (see [9]), which could be useful in future studies regarding robust testing. The robust estimator proposed in the present paper has the self-standardized influence function bounded by the constant c appearing in the Huber function. In a similar manner as in [9], the constant c controls the degree of robustness of the estimator, and in practice, we could take a value close to the lower bound $\sqrt{\ell}$ in order to enforce a maximum amount of robustness. On the other hand, the density power divergences [25] combined with the minimum divergence approach have proved useful for construction of robust estimators in different contexts,

in particular for parameter density estimation. Other approaches for robust estimation in regression models is proposed by [26]. Such approaches could be considered in future research studies in order to be adapted to the context of the moment condition models.

2.4. Consistency of the Estimators

In this subsection, we establish consistency of the estimator \hat{t}_θ^c of t_θ^c , for any fixed $\theta \in \Theta$, and the consistency of the estimator $\hat{\theta}^c$ of θ_0 . First, for any fixed $\theta \in \Theta$, we state the consistency of the estimators \hat{A}_θ and $\hat{\tau}_\theta$ defined by the system (11).

2.4.1. Consistency of the Estimators \hat{A}_θ and $\hat{\tau}_\theta$, for Fixed $\theta \in \Theta$

The estimators \hat{A}_θ and $\hat{\tau}_\theta$, of A_θ and τ_θ defined by the theoretical system of Equation (7), are Z-estimators. We consider the following notations

$$\Psi_1(\theta, A, \tau) := \int h^c(y, \theta, A, \tau) dP_\theta(y) \tag{62}$$

$$\Psi_2(x, \theta, A, \tau) := h^c(x, \theta, A, \tau) h^c(x, \theta, A, \tau)^\top - I_\ell \tag{63}$$

and $\Psi(x, \theta, A, \tau) := (\Psi_1(\theta, A, \tau)^\top, \text{vec}(\Psi_2(x, \theta, A, \tau))^\top)^\top$, where “vec(·)” is the operator that transforms the matrix into vector, by putting all the columns of the matrix one under the other. Notice that, $\Psi_1(\theta, A, \tau)$ is a constant function with respect to x . With these notations, for a given $\theta \in \Theta$, the Z-estimators \hat{A}_θ and $\hat{\tau}_\theta$ are solutions of the system (see (11)),

$$\int \Psi(x, \theta, \hat{A}_\theta, \hat{\tau}_\theta) dP_n(x) = 0, \tag{64}$$

and their theoretical counterparts are A_θ and τ_θ solution of the system (see (7)),

$$\int \Psi(x, \theta, A_\theta, \tau_\theta) dP_0(x) = 0. \tag{65}$$

In the rest of the paper, we consider matrix A in its “vec” form, as defined above. This is necessary in order to apply some classical results, for example the uniform weak law of large numbers (UWLLN) or results regarding Z-estimators. Therefore, the argument A of the functions h^c , n^c and Ψ will be in fact $\text{vec}A$. For simplicity, we write A instead of $\text{vec}A$. The same is valid for \hat{A}_θ and A_θ .

Assumption 1. (a) *There exists compact neighborhood N_θ of (A_θ, τ_θ) such that*

$$\int \sup_{(A, \tau) \in N_\theta} \|\Psi(x, \theta, A, \tau)\| dP_0(x) < \infty; \tag{66}$$

(b) *for any positive ε , the following condition holds*

$$\inf_{(A, \tau) \in M_\theta} \left\| \int \Psi(x, \theta, A, \tau) dP_0(x) \right\| > 0 = \left\| \int \Psi(x, \theta, A_\theta, \tau_\theta) dP_0(x) \right\|,$$

where $M_\theta := \{(A, \tau) \text{ s.t. } \|(A, \tau) - (A_\theta, \tau_\theta)\| > \varepsilon\}$.

Proposition 2. *For each $\theta \in \Theta$, under Assumption 1, $(\hat{A}_\theta, \hat{\tau}_\theta)$ converges in probability to (A_θ, τ_θ) .*

Proof. Since $(A, \tau) \mapsto \Psi(x, \theta, A, \tau)$ is continuous, by the UWLLN, Assumption 1(a) implies

$$\sup_{(A, \tau) \in N_\theta} \left\| \int \Psi(x, \theta, A, \tau) dP_n(x) - \int \Psi(x, \theta, A, \tau) dP_0(x) \right\| \rightarrow 0, \tag{67}$$

in probability. This result, together with Assumption 1(b) ensure the convergence in probability of the estimators \widehat{A}_θ and $\widehat{\tau}_\theta$ toward A_θ and τ_θ , respectively. The arguments are the same as those from [27], Theorem 5.9, p. 46. \square

2.4.2. Consistency of the Estimator \widehat{t}_θ^c of t_θ^c , for Fixed $\theta \in \Theta$

We state the consistency of the estimators \widehat{t}_θ^c under the following assumptions.

Assumption 2. (a) $t_\theta^c := \arg \sup_{t \in \Lambda_\theta^c(P_0)} \int m^c(y, \theta, t) dP_0(y)$ exists, unique and interior point of

$\Lambda_\theta^c(P_0)$;

(b) there exists a compact neighborhood $N_{t_\theta^c} \subset \Lambda_\theta^c(P_0)$ of t_θ^c such that $t_\theta^c \in \text{int}(N_{t_\theta^c})$, and there exists compact neighborhood N_θ of (A_θ, τ_θ) such that

$$\int \sup_{t \in N_{t_\theta^c}, (A, \tau) \in N_\theta} |n^c(x, \theta, t, A, \tau)| dP_0(x) < \infty;$$

(c) there exists compact neighborhood $N_{t_\theta^c}$ of t_θ^c , and there exists a sequence $B_n = O_P(1)$, such that, for all $t, t' \in N_{t_\theta^c}$, it holds

$$\left| \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_0(x) - \int n^c(x, \theta, t', \widehat{A}_\theta, \widehat{\tau}_\theta) dP_0(x) \right| \leq B_n \|t - t'\|. \tag{68}$$

Proposition 3. Under Assumptions 1 and 2, we have

- (1) \widehat{t}_θ^c converges in probability to t_θ^c ;
- (2) $\widehat{KL}_m(\mathcal{M}_\theta^{c,n}, P_0)$ converges in probability to $KL_m(\mathcal{M}_\theta^c, P_0)$.

Proof. (1) Using Assumption 2(b) and the continuity of the function $n^c(x, \theta, t, A, \tau)$ with respect to t, A , and τ , by the uniform weak law of large numbers (UWLLN), we get

$$\sup_{t \in N_{t_\theta^c}, (A, \tau) \in N_\theta} \left| \int n^c(x, \theta, t, A, \tau) dP_n(x) - \int n^c(x, \theta, t, A, \tau) dP_0(x) \right| \rightarrow 0, \tag{69}$$

in probability. The following inequality holds

$$\begin{aligned} \sup_{t \in N_{t_\theta^c}} \left| \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, t, A_\theta, \tau_\theta) dP_0(x) \right| \leq \\ \sup_{t \in N_{t_\theta^c}} \left| \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_0(x) \right| + \\ + \sup_{t \in N_{t_\theta^c}} \left| \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_0(x) - \int n^c(x, \theta, t, A_\theta, \tau_\theta) dP_0(x) \right|. \tag{70} \end{aligned}$$

The first term in the right hand side of the Inequality (70), tends to 0 in probability, on the basis of the result (69). The second term in the right hand side of (70) also tends to 0 in probability. Indeed, using the convergence in probability $(\widehat{A}_\theta, \widehat{\tau}_\theta) \rightarrow (A_\theta, \tau_\theta)$, by Assumption 2(b), we get the pointwise convergence for each t . Then, according to Corollary 2.1 from [28], using the point-wise convergence together with Assumption 2(c), we obtain the uniform convergence

$$\sup_{t \in N_{t_\theta^c}} \left| \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_0(x) - \int n^c(x, \theta, t, A_\theta, \tau_\theta) dP_0(x) \right| = o_P(1). \tag{71}$$

Consequently

$$\sup_{t \in N_{t_\theta^c}} \left| \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, t, A_\theta, \tau_\theta) dP_0(x) \right| \rightarrow 0, \tag{72}$$

in probability. Using (72) and the fact that t_θ^c is unique and belongs to $\text{int}(N_{t_\theta^c})$ and the strict concavity of the function $t \mapsto \int n^c(y, \theta, t, A_\theta, \tau_\theta) dP_0(y)$, on the basis of Theorem 5.7 in [27], we conclude that any value

$$\underline{t} := \arg \sup_{t \in N_{t_\theta^c}} \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) \tag{73}$$

converges in probability to t_θ^c . We show that \widehat{t}_θ^c belongs to $\text{int}(N_{t_\theta^c})$ with probability one as $n \rightarrow \infty$, and consequently it converges to t_θ^c . Since for n sufficiently large any $\underline{t} \in \text{int}(N_{t_\theta^c})$, the concavity of the criterion function $t \mapsto \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x)$ ensures that no other point t in the complement of $\text{int}(N_{t_\theta^c})$ can maximize $\int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x)$ over $t \in \mathbb{R}^\ell$; hence \widehat{t}_θ^c belongs to $\text{int}(N_{t_\theta^c})$.

(2) We have

$$\begin{aligned} \widehat{KL}_m(\mathcal{M}_\theta^{c,n}, P_0) - KL_m(\mathcal{M}_\theta^c, P_0) &= \int n^c(x, \theta, \widehat{t}_\theta^c, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \\ &\int n^c(x, \theta, t_\theta^c, A_\theta, \tau_\theta) dP_0(x). \end{aligned}$$

Note that

$$\begin{aligned} \int n^c(x, \theta, \widehat{t}_\theta^c, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, t_\theta^c, A_\theta, \tau_\theta) dP_0(x) &\leq \\ \widehat{KL}_m(\mathcal{M}_\theta^{c,n}, P_0) - KL_m(\mathcal{M}_\theta^{c,n}, P_0) &\leq \\ \int n^c(x, \theta, \widehat{t}_\theta^c, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, \widehat{t}_\theta^c, A_\theta, \tau_\theta) dP_0(x). \end{aligned}$$

Both the right hand side and the left hand side in the above display tend to 0 in probability using (72). Hence, $\widehat{KL}_m(\mathcal{M}_\theta^{c,n}, P_0)$ converges to $KL_m(\mathcal{M}_\theta^c, P_0)$, in probability. \square

2.4.3. Consistency of the Estimator $\widehat{\theta}^c$

Assumption 3. (a) the function $(\theta, t, A, \tau) \mapsto h^c(X, \theta, t, A, \tau)$ is continuous, with probability 1;

(b) for each $\theta \in \Theta$, there exist compact neighborhood $N_{t_\theta^c}$ of t_θ^c and compact neighborhood N_θ of (A_θ, τ_θ) such that

$$\int \sup_{\theta \in \Theta} \sup_{t \in N_{t_\theta^c}, (A, \tau) \in N_\theta} |n^c(x, \theta, t, A, \tau)| dP_0(x) < \infty; \tag{74}$$

(c) Let $t_n^*(\theta) := \arg \sup_{t \in N_{t_\theta^c}} \left| \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_0(x) - \int n^c(x, \theta, t, A_\theta, \tau_\theta) dP_0(x) \right|$. There exists a sequence $B_n = O_P(1)$ such that, for all $\theta, \theta' \in \Theta$, it holds

$$\begin{aligned} \left| \int n^c(x, \theta, t_n^*(\theta), \widehat{A}_\theta, \widehat{\tau}_\theta) dP_0(x) - \right. \\ \left. - \int n^c(x, \theta', t_n^*(\theta'), \widehat{A}_{\theta'}, \widehat{\tau}_{\theta'}) dP_0(x) \right| \leq B_n \|\theta - \theta'\|; \tag{75} \end{aligned}$$

- (d) the function $\theta \mapsto \int n^c(x, \theta, t_\theta^c, A_\theta, \tau_\theta) dP_0(x)$ is continuous on Θ ;
- (e) $\theta_0 := \arg \inf_{\theta \in \Theta} \int n^c(x, \theta, t_\theta^c, A_\theta, \tau_\theta) dP_0(x)$ exists, unique and interior point of Θ .

Proposition 4. Under Assumptions 1–3, we have

- (1) $\|\hat{t}_\theta^c - t_\theta^c\| \rightarrow 0$ in probability, uniformly with respect to $\theta \in \Theta$;
- (2) $\|\hat{\theta}^c - \theta_0\| \rightarrow 0$ in probability.

Proof. (1) By Assumption 3(a), $n^c(x, \theta, t, A, \tau)$ is continuous in θ, t, A, τ . Using also Assumption 3(b), by applying UWLLN, we obtain the uniform convergence in probability

$$\sup_{(\theta, t, A, \tau) \in C} \left| \int n^c(x, \theta, t, A, \tau) dP_n(x) - \int n^c(x, \theta, t, A, \tau) dP_0(x) \right| \rightarrow 0 \tag{76}$$

over the compact set $C := \{(\theta, t, A, \tau) \text{ s.t. } \theta \in \Theta, t \in N_{t_\theta^c}, (A, \tau) \in N_\theta\}$. We will prove the uniform convergence in probability

$$\sup_{\theta \in \Theta} \|\hat{t}_\theta^c - t_\theta^c\| \rightarrow 0. \tag{77}$$

Let $\eta > 0$. We show that $P_0(\sup_{\theta \in \Theta} \|\hat{t}_\theta^c - t_\theta^c\| \geq \eta) \rightarrow 0$ for any

$$\tilde{t}_\theta^c := \arg \sup_{t \in N_{t_\theta^c}} \int n^c(x, \theta, t, \hat{A}_\theta, \hat{\tau}_\theta) dP_n(x). \tag{78}$$

We show that \hat{t}_θ^c belongs to $\text{int}(N_{t_\theta^c})$ with probability one, as $n \rightarrow \infty$.

Let $\eta > 0$ be such that $\sup_{\theta \in \Theta} \|\tilde{t}_\theta^c - t_\theta^c\| \geq \eta$. Since Θ is compact, by continuity, there exists $\bar{\theta} \in \Theta$, such that $\sup_{\theta \in \Theta} \|\tilde{t}_\theta^c - t_\theta^c\| = \|\tilde{t}_{\bar{\theta}}^c - t_{\bar{\theta}}^c\| \geq \eta$. Hence, there exists $\varepsilon > 0$ such that

$$\int n^c(x, \bar{\theta}, t_{\bar{\theta}}^c, A_{\bar{\theta}}, \tau_{\bar{\theta}}) dP_0(x) - \int n^c(x, \bar{\theta}, \tilde{t}_{\bar{\theta}}^c, A_{\bar{\theta}}, \tau_{\bar{\theta}}) dP_0(x) > \varepsilon.$$

Therefore,

$$P_0\left(\sup_{\theta \in \Theta} \|\tilde{t}_\theta^c - t_\theta^c\| \geq \eta\right) \leq P_0\left(\int n^c(x, \bar{\theta}, t_{\bar{\theta}}^c, A_{\bar{\theta}}, \tau_{\bar{\theta}}) dP_0(x) - \int n^c(x, \bar{\theta}, \tilde{t}_{\bar{\theta}}^c, A_{\bar{\theta}}, \tau_{\bar{\theta}}) dP_0(x) > \varepsilon\right). \tag{79}$$

On the other hand, using (71) and (72), we can write

$$\begin{aligned} & \int n^c(x, \bar{\theta}, t_{\bar{\theta}}^c, A_{\bar{\theta}}, \tau_{\bar{\theta}}) dP_0(x) - \int n^c(x, \bar{\theta}, \tilde{t}_{\bar{\theta}}^c, A_{\bar{\theta}}, \tau_{\bar{\theta}}) dP_0(x) = \\ & \int n^c(x, \bar{\theta}, t_{\bar{\theta}}^c, \hat{A}_{\bar{\theta}}, \hat{\tau}_{\bar{\theta}}) dP_n(x) - \int n^c(x, \bar{\theta}, \tilde{t}_{\bar{\theta}}^c, A_{\bar{\theta}}, \tau_{\bar{\theta}}) dP_0(x) + o_P(1) \leq \\ & \leq \int n^c(x, \bar{\theta}, \tilde{t}_{\bar{\theta}}^c, \hat{A}_{\bar{\theta}}, \hat{\tau}_{\bar{\theta}}) dP_n(x) - \int n^c(x, \bar{\theta}, \tilde{t}_{\bar{\theta}}^c, \hat{A}_{\bar{\theta}}, \hat{\tau}_{\bar{\theta}}) dP_0(x) + o_P(1) \leq \\ & \leq \sup_C \left| \int n^c(x, \theta, t, A, \tau) dP_n(x) - \int n^c(x, \theta, t, A, \tau) dP_0(x) \right| + o_P(1). \tag{80} \end{aligned}$$

Using (76), (79), and (80), we deduce that $\sup_{\theta \in \Theta} \|\tilde{t}_\theta^c - t_\theta^c\| \rightarrow 0$ in probability. In particular, for large n , $\tilde{t}_\theta^c \in \text{int}(N_{t_\theta^c})$ uniformly in θ . Since $t \mapsto \int n^c(x, \theta, t, \hat{A}_\theta, \hat{\tau}_\theta) dP_n(x)$ is

concave, the maximizer \widehat{t}_θ^c belongs to $\text{int}(N_{t_\theta^c})$ for sufficiently large n . Then, $\sup_{\theta \in \Theta} \|\widehat{t}_\theta^c - t_\theta^c\| \rightarrow 0$ in probability. (2) For large n , we can write

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \int n^c(x, \theta, \widehat{t}_\theta^c, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, t_\theta^c, A_\theta, \tau_\theta) dP_0(x) \right| \\ &= \sup_{\theta \in \Theta} \left| \int n^c(x, \theta, \widehat{t}_\theta^c, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, t_\theta^c, A_\theta, \tau_\theta) dP_0(x) \right| =: \sup_{\theta \in \Theta} |B|. \end{aligned} \tag{81}$$

Note that

$$\begin{aligned} & \int n^c(x, \theta, t_\theta^c, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, t_\theta^c, A_\theta, \tau_\theta) dP_0(x) \leq \\ & \leq B \leq \int n^c(x, \theta, \widehat{t}_\theta^c, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, \widehat{t}_\theta^c, A_\theta, \tau_\theta) dP_0(x). \end{aligned} \tag{82}$$

In order to prove that $\sup_{\theta \in \Theta} |B| \rightarrow 0$ in probability, we first prove that

$$\sup_{\theta \in \Theta} \sup_{t \in N_{t_\theta^c}} \left| \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, t, A_\theta, \tau_\theta) dP_0(x) \right| \rightarrow 0, \tag{83}$$

in probability. Notice that

$$\begin{aligned} & \sup_{\theta \in \Theta} \sup_{t \in N_{t_\theta^c}} \left| \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, t, A_\theta, \tau_\theta) dP_0(x) \right| \leq \\ & \sup_{\theta \in \Theta} \sup_{t \in N_{t_\theta^c}} \left| \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_0(x) \right| + \\ & \sup_{\theta \in \Theta} \sup_{t \in N_{t_\theta^c}} \left| \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_0(x) - \int n^c(x, \theta, t, A_\theta, \tau_\theta) dP_0(x) \right|. \end{aligned}$$

The first term in the right-hand side of the above inequality tends to 0 in probability, using (76). Regarding the second term, we have the convergence (71), and combining this with Assumption 3(c), we obtain

$$\sup_{\theta \in \Theta} \sup_{t \in N_{t_\theta^c}} \left| \int n^c(x, \theta, t, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_0(x) - \int n^c(x, \theta, t, A_\theta, \tau_\theta) dP_0(x) \right| \rightarrow 0, \tag{84}$$

in probability. Consequently, (83) holds. Then, (81) and (82) lead to

$$\sup_{\theta \in \Theta} \left| \int n^c(x, \theta, \widehat{t}_\theta^c, \widehat{A}_\theta, \widehat{\tau}_\theta) dP_n(x) - \int n^c(x, \theta, t_\theta^c, A_\theta, \tau_\theta) dP_0(x) \right| \rightarrow 0. \tag{85}$$

Assumptions 3(d) and (e) ensure that θ_0 is well-separated in the sense that, $\forall \varepsilon > 0$,

$$\inf_{\{\theta; \|\theta - \theta_0\| \geq \varepsilon\}} \int n^c(x, \theta, t_\theta^c, A_\theta, \tau_\theta) dP_0(x) > 0 = \int n^c(x, \theta_0, t_{\theta_0}^c, A_{\theta_0}, \tau_{\theta_0}) dP_0(x). \tag{86}$$

Finally, (85) and (86) imply that $\widehat{\theta}^c \rightarrow \theta_0$ in probability, on the basis of Theorem 5.7 p. 45 from [27]. \square

3. Simulation Results

In order to compare the performance of the proposed robust EL estimate (26) with that of the EL estimator (27) in the case of contaminated data, we consider the moment condition model presented in Example 1 in Section 1.

Let X be a random variable with probability distribution $P_0 = \chi_1^2$, a chi-square distribution with one degree of freedom. Then, it holds that the equation $\int_{\mathbb{R}} g(x, \theta) dP_0(x) = 0$, with $g(x, \theta) = (x - \theta, x^2 - \theta^2 - 2\theta)^\top$, has a unique solution $\theta = \theta_0 = 1$. This is a particular case of the model from Example 1, namely when $h(\theta) = \theta^2 + \theta$. Observe that, for this model, $g(x, \theta)$ is unbounded (in x).

We use i.i.d. data generated from a slight deviation of the model P_0 , namely from the model

$$(1 - \epsilon) \chi_1^2 + \epsilon \chi_{10}^2,$$

with $\epsilon = 0.05$, respectively $\epsilon = 0.10$. The considered sample sizes are $n = 100$ and $n = 500$. All the simulations are repeated 1000 times. The obtained estimates are compared through bias, variance, and mean square error, computed on the basis of the 1000 replications. We give the corresponding box-plots in Figures 1–4, where the true value of the parameter $\theta_0 = 1$ is presented with horizontal dashed line. For computing the proposed Robust EL estimate (26), we use the truncated functions $g^c(x, \theta)$ and $g_n^c(x, \theta)$ with $c = 2$. The algorithm for computing the estimate was obtained by adapting the one from [10] (Appendix A.1., p. 3196). Namely, for each iteration, the estimation of the parameter θ_0 , corresponding to the new orthogonality function in step iii, is computed using Uzawa algorithm for the saddle-point optimum in (26). The obtained results are presented in Tables 1–4 and Figures 1–4. All these results illustrate the fact that, in the case of contaminated data, the robust EL estimator outperforms the classical EL estimator.

Table 1. Robust Empirical Likelihood (EL) versus EL, for $n = 100$ and $\epsilon = 0.05$.

	Estimate	Squared Bias	Variance	Mean Square Error
Robust EL	1.0477	0.0023	0.0034	0.0056
EL	1.0861	0.0074	0.0208	0.0282

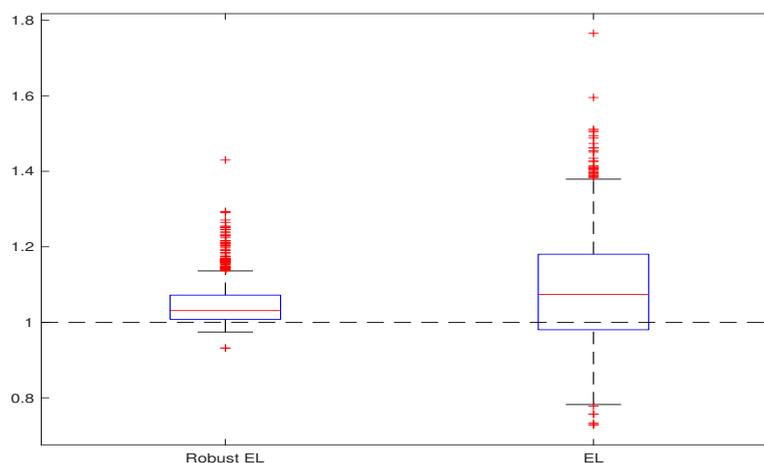


Figure 1. Robust EL versus EL, for $n = 100$ and $\epsilon = 0.05$.

Table 2. Robust EL versus EL, for $n = 100$ and $\epsilon = 0.10$.

	Estimate	Squared Bias	Variance	Mean Square Error
Robust EL	1.1281	0.0164	0.0116	0.0281
EL	1.2052	0.0421	0.0342	0.0763

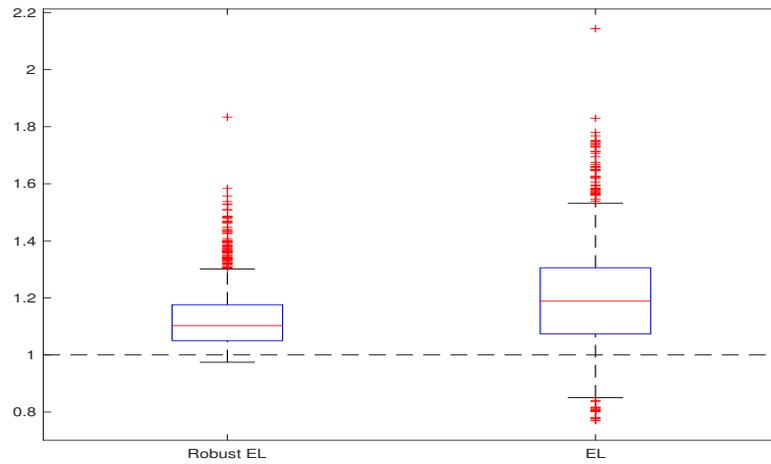


Figure 2. Robust EL versus EL, for $n = 100$ and $\epsilon = 0.10$.

Table 3. Robust EL versus EL, for $n = 500$ and $\epsilon = 0.05$.

	Estimate	Squared Bias	Variance	Mean Square Error
Robust EL	1.0336	0.0011	0.0005	0.0016
EL	1.0750	0.0056	0.0039	0.0095

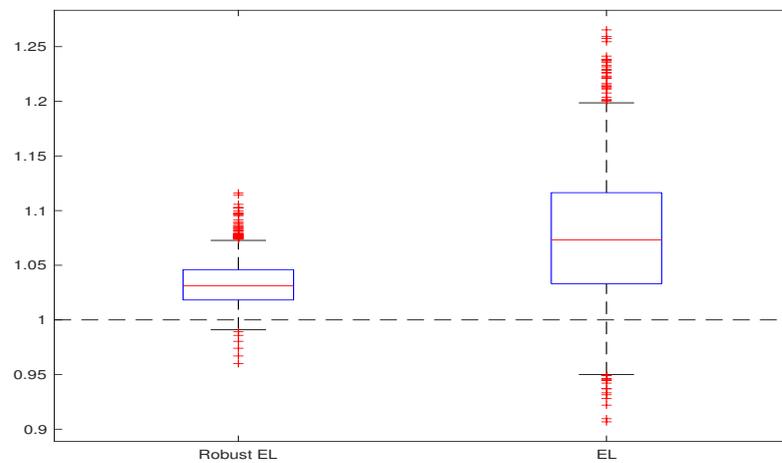


Figure 3. Robust EL versus EL, for $n = 500$ and $\epsilon = 0.05$.

Table 4. Robust EL versus EL, for $n = 500$ and $\epsilon = 0.10$.

	Estimate	Squared Bias	Variance	Mean Square Error
Robust EL	1.1033	0.0107	0.0019	0.0126
EL	1.1835	0.0337	0.0065	0.0402

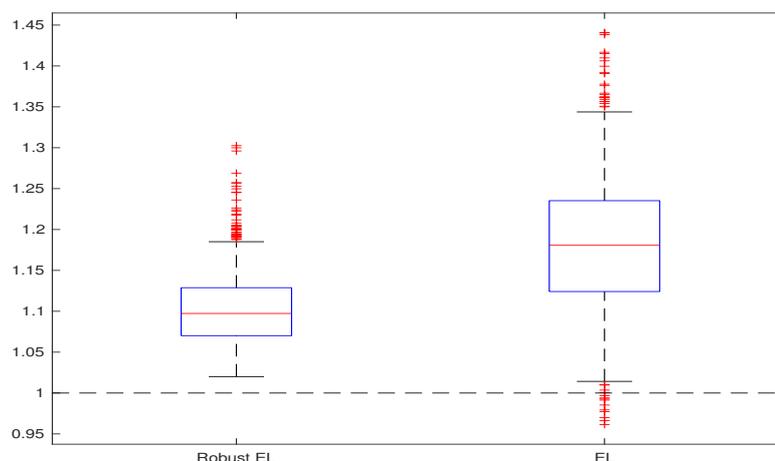


Figure 4. Robust EL versus EL, for $n = 500$ and $\epsilon = 0.10$.

4. Conclusions

We proposed a robust version of the EL estimator for moment condition models. This estimator is defined through the minimization of an empirical version of the modified Kullback–Leibler divergence in dual form, using truncated orthogonality functions based on multivariate Huber function. We proved the robustness by means of the influence function, as well the consistency of the new estimator. The results of the Monte Carlo simulation study show that, in the case of contaminated data, the robust EL estimator outperforms the classical EL estimator.

Author Contributions: Conceptualization, A.K. and A.T.; methodology, A.K. and A.T.; investigation, A.K. and A.T.; writing the manuscript, A.K. and A.T. All authors have read and agreed to the final version of the manuscript.

Funding: This work was supported by a grant of the Romanian Ministry of Education and Research, CNCS—UEFISCDI, project number PN-III-P4-ID-PCE-2020-1112, within PNCDI III.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hansen, L.P. Large sample properties of generalized method of moments estimators. *Econometrica* **1982**, *50*, 1029–1054. [\[CrossRef\]](#)
- Hansen, L.P.; Heaton, J.; Yaron, A. Finite-sample properties of some alternative generalized method of moments estimators. *J. Bus. Econ. Stat.* **1996**, *14*, 262–280.
- Qin, J.; Lawless, J. Empirical likelihood and general estimating equations. *Ann. Stat.* **1994**, *22*, 300–325. [\[CrossRef\]](#)
- Imbens, G.W. One-step estimators for over-identified generalized method of moments models. *Rev. Econ. Stud.* **1997**, *64*, 359–383. [\[CrossRef\]](#)
- Owen, A. *Empirical Likelihood*; Chapman and Hall: New York, NY, USA, 2001.
- Kitamura, Y.; Stutzer, M. An information-theoretic alternative to generalized method of moments estimation. *Econometrica* **1997**, *65*, 861–874. [\[CrossRef\]](#)
- Imbens, G.W. Generalized method of moments and empirical likelihood. *J. Bus. Econ. Stat.* **2002**, *20*, 493–506. [\[CrossRef\]](#)
- Newey, W.K.; Smith, R.J. Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **2004**, *72*, 219–255. [\[CrossRef\]](#)
- Ronchetti, E.; Trojani, F. Robust inference with GMM estimators. *J. Econom.* **2001**, *101*, 37–69. [\[CrossRef\]](#)
- Lô, S.N.; Ronchetti, E. Robust small sample accurate inference in moment condition models. *Comput. Stat. Data Anal.* **2012**, *56*, 3182–3197. [\[CrossRef\]](#)

11. Felipe, A.; Martin, N.; Miranda, P.; Pardo, L. Testing with exponentially tilted empirical likelihood. *Methodol. Comput. Appl. Probab.* **2018**, *20*, 1–40. [[CrossRef](#)]
12. Broniatowski, M.; Keziou, A. Divergences and duality for estimation and test under moment condition models. *J. Stat. Plan. Inference* **2012**, *142*, 2554–2573. [[CrossRef](#)]
13. Broniatowski, M.; Keziou, A. Parametric estimation and tests through divergences and the duality technique. *J. Multivar. Anal.* **2009**, *100*, 16–36. [[CrossRef](#)]
14. Toma, A.; Leoni-Aubin, S. Robust tests based on dual divergence estimators and saddlepoint approximations. *J. Multivar. Anal.* **2010**, *101*, 1143–1155. [[CrossRef](#)]
15. Toma, A.; Broniatowski, M. Dual divergence estimators and tests: Robustness results. *J. Multivar. Anal.* **2011**, *102*, 20–36. [[CrossRef](#)]
16. Toma, A. Model selection criteria using divergences. *Entropy* **2014**, *16*, 2686–2698. [[CrossRef](#)]
17. Schennach, S.M. Point estimation with exponentially tilted empirical likelihood. *Ann. Stat.* **2007**, *35*, 634–672. [[CrossRef](#)]
18. Toma, A. Robustness of dual divergence estimators for models satisfying linear constraints. *C. R. Math.* **2013**, *351*, 311–316. [[CrossRef](#)]
19. Engle, R.F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **1982**, *50*, 987–1007. [[CrossRef](#)]
20. Bansal, R.; Hsieh, D.; Viswanathan, S. No arbitrage and arbitrage pricing: A new approach. *J. Financ.* **1993**, *48*, 1719–1747. [[CrossRef](#)]
21. 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; John Wiley & Sons Inc.: New York, NY, USA, 1986.
22. Rüschemdorf, L. On the minimum discrimination information theorem. *Stat. Decis.* **1984**, *1984* (Suppl. 1), 263–283.
23. Cressie, N.; Read, T.R.C. Multinomial goodness-of-fit tests. *J. R. Stat. Soc. Ser. B* **1984**, *46*, 440–464. [[CrossRef](#)]
24. Broniatowski, M.; Keziou, A. Minimization of ϕ -divergences on sets of signed measures. *Stud. Sci. Math. Hung.* **2006**, *43*, 403–442.
25. Basu, A.; Harris, I.R.; Hjort, N.L.; Jones, M.C. Robust and efficient estimation by minimizing a density power divergence. *Biometrika* **1998**, *85*, 549–559. [[CrossRef](#)]
26. She, Y.; Owen, A. Outlier detection using nonconvex penalized regression. *J. Am. Stat. Assoc.* **2011**, *106*, 626–639. [[CrossRef](#)]
27. van der Vaart, A. *Asymptotic Statistics*; Cambridge University Press: Cambridge, UK, 1998.
28. Newey, W.K. Uniform convergence in probability and stochastic equicontinuity. *Econometrica* **1991**, *59*, 1161–1167. [[CrossRef](#)]