

Article

Robust Z-Estimators for Semiparametric Moment Condition Models

Aida Toma ^{1,2}
¹ Department of Applied Mathematics, Bucharest University of Economic Studies, 010374 Bucharest, Romania; aida.toma@csie.ase.ro

² “Gheorghe Mihoc-Caius Iacob” Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, 050711 Bucharest, Romania

Abstract: In the present paper, we introduce a class of robust Z-estimators for moment condition models. These new estimators can be seen as robust alternatives for the minimum empirical divergence estimators. By using the multidimensional Huber function, we first define robust estimators of the element that realizes the supremum in the dual form of the divergence. A linear relationship between the influence function of a minimum empirical divergence estimator and the influence function of the estimator of the element that realizes the supremum in the dual form of the divergence led to the idea of defining new Z-estimators for the parameter of the model, by using robust estimators in the dual form of the divergence. The asymptotic properties of the proposed estimators were proven, including here the consistency and their asymptotic normality. Then, the influence functions of the estimators were derived, and their robustness is demonstrated.

Keywords: moment condition models; divergences; robustness

1. Introduction

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

$$\int g(x, \theta) dQ(x) = 0, \text{ for all } Q \in \mathcal{M}^1. \quad (1)$$

The parameter θ belongs to $\Theta \subset \mathbb{R}^d$; the function $g := (g_1, \dots, g_l)^\top$ is defined on $\mathbb{R}^m \times \Theta$, each of the g_i 's being real-valued, $l \geq d$, and the functions $g_1, \dots, g_l$ and 1_X are supposed to be linearly independent. Denote by $\mathcal{M}^1$ the set of all probability measures on $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ and

$$\mathcal{M}_\theta^1 := \{Q \in \mathcal{M}^1 : \int g(x, \theta) dQ(x) = 0\}, \quad (2)$$

such that

$$\mathcal{M}^1 = \bigcup_{\theta \in \Theta} \mathcal{M}_\theta^1. \quad (3)$$

Let $X_1, \dots, X_n$ be an i.i.d. sample on the random vector X with unknown probability distribution P_0 . We considered the problem of the estimation of the parameter θ_0 for which the constraints of the model are satisfied:

$$\int g(x, \theta_0) dP_0(x) = 0. \quad (4)$$

We supposed that θ_0 is the unique solution of Equation (4). Thus, we assumed that information about θ_0 and P_0 is available in the form of $l \geq d$ functionally independent unbiased estimating functions, and we used this information to estimate θ_0 .

Among the best-known estimation methods for moment condition models, we mention the generalized method of moments (GMM) [1], the continuous updating (CU) estimator [2],



Citation: Toma, A. Robust Z-Estimators for Semiparametric Moment Condition Models. *Entropy* **2023**, *25*, 1013. <https://doi.org/10.3390/e25071013>

Academic Editor: Leandro Pardo

Received: 8 May 2023

Revised: 28 June 2023

Accepted: 29 June 2023

Published: 30 June 2023



Copyright: © 2023 by the author. 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/>).

the empirical likelihood (EL) estimator [3,4], the exponential tilting (ET) estimator [5], and the generalized empirical likelihood (GEL) estimators [6]. Although the EL estimator is superior to other estimators in terms of higher-order asymptotic properties, these properties hold only under the correct specification of the moment conditions. In [7] was proposed the exponentially tilted empirical likelihood (ETEL) estimator, which has the same higher-order property as the EL estimator under the correct specification, while maintaining the usual asymptotic properties such as the consistency and asymptotic normality under misspecification. The so-called information and entropy econometric techniques have been proposed to improve the finite sample performance of the GMM-estimators and tests (see, e.g., [4,5]).

Some recent methods for the estimation and testing of moment condition models are based on using divergences. Divergences between probability measures are widely used in statistics and data science in order to perform inference in models of various kinds, parametric or semiparametric. Statistical methods based on divergence minimization extend the likelihood paradigm and often have the advantage of providing a trade-off between efficiency and robustness [8–11]. A general methodology for the estimation and testing of moment condition models was developed in [12]. This approach is based on minimizing divergences in their 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. The approach based on minimizing dual forms of divergences was initially used in the case of parametric models, the results being published in a series of articles [13–16]. The broad class of minimum empirical divergence estimators contains in particular the EL estimator, the CU estimator, as well as the ET estimator mentioned above. Using the influence function as the robustness measure, it has been shown that the minimum empirical divergence estimators are not robust, because the corresponding influence functions are generally not bounded [17]. On the other hand, the minimum empirical divergence estimators have the same efficiency of first order, and moreover, the EL estimator, which belong to this class, is superior in higher-order efficiency. Therefore, proposing robust versions of the minimum empirical divergence estimators would bring a trade-off between robustness and efficiency. These aspects motivated our studies in the present paper.

Some robust estimation methods for moment condition models have been proposed in the literature, for example in [18–22]. In the present paper, we introduce a class of robust Z-estimators for moment condition models. These new estimators can be seen as robust alternatives for the minimum empirical divergence estimators. By using the multidimensional Huber function, we first define robust estimators of the element that realizes the supremum in the dual form of the divergence. A linear relationship between the influence function of a minimum empirical divergence estimator and the influence function of the estimator of the element that realizes the supremum in the dual form of the divergence led to the idea of defining new Z-estimators for the parameter of the model, by using robust estimators in the dual form of the divergence. The asymptotic properties of the proposed estimators were proven, including here the consistency and their asymptotic normality. Then, the influence functions of the estimators were derived, and their robustness is demonstrated.

The paper is organized as follows. In Section 2, we briefly recall the context and the definitions of the minimum empirical divergence estimators, these being necessary for defining the new estimators. In Section 3, the new Z-estimators for moment condition models are defined. The asymptotic properties of these estimators were proven, including here the consistency and their asymptotic normality. Then, the influence functions of the estimators were derived, and their robustness is demonstrated. The proofs of the theoretical results are deferred in Appendix A.

2. Minimum Empirical Divergence Estimators

2.1. Statistical Divergences

Let φ be a convex function defined on $\mathbb{R}$ and $[0, \infty]$ -valued, such that $\varphi(1) = 0$, and let $P \in M^1$ be some probability measure. For any signed finite measure Q defined on the same measurable space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, absolutely continuous (a.c.) with respect to P , the φ divergence between Q and P is defined by

$$D_\varphi(Q, P) := \int \varphi\left(\frac{dQ}{dP}(x)\right) dP(x). \quad (5)$$

When Q is not a.c. with respect to P , we set $D_\varphi(Q, P) = \infty$. This extension, for the case when Q is not absolutely continuous with respect to P , was considered in order to have a unique definition of divergences, appropriate for both cases—that of continuous probability laws and that of discrete probability laws.

This definition extends the one of divergences between probability measures [23], and the necessity of working with signed finite measures will be explained in Section 2.2.

Largely used in information theory, the Kullback–Leibler divergence is associated with the real convex function $\varphi(x) := x \log x - x + 1$ and is defined by

$$KL(Q, P) := \int \log\left(\frac{dQ}{dP}\right) dQ.$$

The modified Kullback–Leibler divergence is associated with the convex function $\varphi(x) := -\log x + x - 1$ and is defined through

$$KL_m(Q, P) := \int -\log\left(\frac{dQ}{dP}\right) dP.$$

Other divergences, largely used in inferential statistics, are the χ^2 and the modified χ^2 divergences, namely

$$\begin{aligned} \chi^2(Q, P) &:= \frac{1}{2} \int \left(\frac{dQ}{dP} - 1\right)^2 dP, \\ \chi_m^2(Q, P) &:= \frac{1}{2} \int \frac{\left(\frac{dQ}{dP} - 1\right)^2}{\frac{dQ}{dP}} dP, \end{aligned}$$

these being associated with the convex functions $\varphi(x) := \frac{1}{2}(x - 1)^2$ and $\varphi(x) := \frac{1}{2}(x - 1)^2/x$, respectively. The Hellinger distance and the L_1 distance are also φ divergences. They are associated with the convex functions $\varphi(x) := 2(\sqrt{x} - 1)^2$ and $\varphi(x) := |x - 1|$, respectively.

All the preceding examples, except the L_1 distance, belong to the class of power divergences introduced by Cressie and Read [24] and defined by the convex functions:

$$x \in \mathbb{R}_+^* \mapsto \varphi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)}, \quad (6)$$

for $\gamma \in \mathbb{R} \setminus \{0, 1\}$ and $\varphi_0(x) := -\log x + x - 1$, $\varphi_1(x) := x \log x - x + 1$. The Kullback–Leibler divergence is associated with φ_1 , the modified Kullback–Leibler with φ_0 , the χ^2 divergence with φ_2 , the modified χ^2 divergence with φ_{-1} , and the Hellinger distance with $\varphi_{1/2}$. When φ_γ is not defined on $(-\infty, 0)$ or when φ_γ is not convex, the definition of the corresponding power divergence function $Q \in M^1 \mapsto D_{\varphi_\gamma}(Q, P)$ can be extended to the whole set of signed finite measures by taking the following extension of φ_γ :

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

The φ divergence between some set Ω of signed finite measures and a probability measure P is defined by

$$D_\varphi(\Omega, P) = \inf_{Q \in \Omega} D_\varphi(Q, P). \quad (7)$$

Assuming that $D_\varphi(\Omega, P)$ is finite, a measure $Q^* \in \Omega$ is called a φ -projection of P on Ω if

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

2.2. Minimum Empirical Divergence Estimators

Let $X_1, \dots, X_n$ be an i.i.d. sample on the random vector X with the probability distribution P_0 . The “plug-in” estimator of the φ divergence $D_\varphi(\mathcal{M}_\theta^1, P_0)$ between the set $\mathcal{M}_\theta^1$ and the probability measure P_0 is defined by replacing P_0 with the empirical measure associated with the sample. More precisely,

$$\hat{D}_\varphi(\mathcal{M}_\theta^1, P_0) = \inf_{Q \in \mathcal{M}_\theta^1} D_\varphi(Q, P_n) = \inf_{Q \in \mathcal{M}_\theta^1} \int \varphi\left(\frac{dQ}{dP_n}(x)\right) dP_n(x), \quad (8)$$

where $P_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ is the empirical measure associated with the sample, δ_x being the Dirac measure putting all mass at x . If the projection of the measure P_n on $\mathcal{M}_\theta^1$ exists, it is a law a.c. with respect to P_n . Then, it is natural to consider

$$\mathcal{M}_\theta^{(n)} = \{Q \in M^1 : Q \text{ a.c. with respect to } P_n \text{ and } \sum_{i=1}^n g(X_i, \theta) Q(X_i) = 0\}, \quad (9)$$

and then, the plug-in estimator (8) can be written as

$$\hat{D}_\varphi(\mathcal{M}_\theta^1, P_0) = \inf_{Q \in \mathcal{M}_\theta^{(n)}} \frac{1}{n} \sum_{i=1}^n \varphi(nQ(X_i)). \quad (10)$$

The infimum in the above expression (10) may be achieved at a point situated on the frontier of the set $\mathcal{M}_\theta^{(n)}$, a case in which the Lagrange method for characterizing the infimum and computing $\hat{D}_\varphi(\mathcal{M}_\theta^1, P_0)$ cannot be applied. In order to avoid this difficulty, Broniatowski and Keziou [12,25] proposed to work on sets of signed finite measures and defined

$$\mathcal{M}_\theta := \{Q \in M : \int dQ = 1 \text{ and } \int g(x, \theta) dQ(x) = 0\}, \quad (11)$$

where M denotes the set of all signed finite measures on the measurable space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, and

$$\mathcal{M} := \bigcup_{\theta \in \Theta} \mathcal{M}_\theta. \quad (12)$$

They showed that, if Q_1^* the projection of P_n on $\mathcal{M}_\theta^1$ is an interior point of $\mathcal{M}_\theta^1$ and Q^* the projection of P_n on $\mathcal{M}_\theta$ is an interior point of $\mathcal{M}_\theta$, then both approaches based on signed finite measures, respectively on probability measures, for defining minimum divergence estimators coincide. On the other hand, in the case when Q_1^* is a frontier point of $\mathcal{M}_\theta^1$, the estimator of the parameter θ_0 defined using the context of signed finite measures converges to θ_0 . These aspects justify the substitution of $\mathcal{M}_\theta^1$ by $\mathcal{M}_\theta$.

In the following, we briefly recall the definitions of the estimators for the moment condition proposed in [12] in the context of signed finite measure sets.

Denote by $\bar{g}$ the function defined on $\mathbb{R}^m \times \Theta$ and $\mathbb{R}^{l+1}$ -valued:

$$\bar{g}(x, \theta) := (\mathbf{1}_X(x), g_1(x, \theta), \dots, g_l(x, \theta))^\top. \quad (13)$$

Given a φ divergence, when the function φ is strictly convex on its domain, denote

$$\varphi^*(u) := u\varphi'^{-1}(u) - \varphi(\varphi'^{-1}(u)),$$

the convex conjugate of the function φ . For a given probability measure $P \in M^1$ and a fixed $\theta \in \Theta$, define

$$\Lambda_\theta(P) := \{t \in \mathbb{R}^{l+1} : \int |\varphi^*(t_0 + \sum_{j=1}^l t_j g_j(x, \theta))| dP(x) < \infty\}. \quad (14)$$

We also use the notations Λ_θ for $\Lambda_\theta(P_0)$ and $\Lambda_\theta^{(n)}$ for $\Lambda_\theta(P_n)$.

Supposing that P_0 admits a projection Q_θ^* on $\mathcal{M}_\theta$ with the same support as P_0 and that the function φ is strictly convex on its domain, then the φ divergence $D_\varphi(\mathcal{M}_\theta, P_0)$ admits the dual representation:

$$D_\varphi(\mathcal{M}_\theta, P_0) = \sup_{t \in \Lambda_\theta} \int m(x, \theta, t) dP_0(x), \quad (15)$$

where $m(x, \theta, t) := t_0 - \varphi^*(t^\top \bar{g}(x, \theta))$.

The supremum in (15) is unique and is reached at a point that we denote as $t_\theta = t_\theta(P_0)$:

$$t_\theta := \arg \sup_{t \in \Lambda_\theta} \int m(x, \theta, t) dP_0(x). \quad (16)$$

Then, $D_\varphi(\mathcal{M}_\theta, P_0)$, t_θ , $D_\varphi(\mathcal{M}, P_0)$ and θ_0 can be estimated respectively by

$$\hat{D}_\varphi(\mathcal{M}_\theta, P_0) := \sup_{t \in \Lambda_\theta^{(n)}} \int m(x, \theta, t) dP_n(x), \quad (17)$$

$$\hat{t}_\theta := \arg \sup_{t \in \Lambda_\theta^{(n)}} \int m(x, \theta, t) dP_n(x), \quad (18)$$

$$\hat{D}_\varphi(\mathcal{M}, P_0) := \inf_{\theta \in \Theta} \sup_{t \in \Lambda_\theta^{(n)}} \int m(x, \theta, t) dP_n(x), \quad (19)$$

$$\hat{\theta}_\varphi := \arg \inf_{\theta \in \Theta} \sup_{t \in \Lambda_\theta^{(n)}} \int m(x, \theta, t) dP_n(x). \quad (20)$$

The estimators defined in (20) are called minimum empirical divergence estimators. We refer to [12] for the complete study of the existence and of the asymptotic properties of the above estimators.

The influence functions of these estimators and corresponding robustness properties were studied in [17]. According to those results, for $\theta \in \Theta$ fixed, the influence function of the estimator $\hat{t}_\theta$ is given by

$$\text{IF}(x; t_\theta, P_0) = - \left[\int \frac{\partial^2}{\partial^2 t} m(y, \theta, t_\theta(P_0)) dP_0(y) \right]^{-1} \frac{\partial}{\partial t} m(x, \theta, t_\theta(P_0)), \quad (21)$$

where

$$\frac{\partial}{\partial t} m(x, \theta, t) = (1, 0_l)^\top - \varphi'^{-1}(t^\top \bar{g}(x, \theta)) \bar{g}(x, \theta), \quad (22)$$

$$\frac{\partial^2}{\partial^2 t} m(x, \theta, t) = - \frac{1}{\varphi''(\varphi'^{-1}(t^\top \bar{g}(x, \theta)))} \bar{g}(x, \theta) \bar{g}(x, \theta)^\top, \quad (23)$$

with the particular case $\theta = \theta_0$:

$$\text{IF}(x; t_{\theta_0}, P_0) = -\varphi''(1) \left[\int \bar{g}(y, \theta_0) \bar{g}(y, \theta_0)^\top dP_0(y) \right]^{-1} (0, g(x, \theta_0)^\top)^\top. \quad (24)$$

On the other hand, the influence function of the estimator $\hat{\theta}_\varphi$ is given by

$$\begin{aligned} \text{IF}(x; T_\varphi, P_0) = & - \left\{ \left[\int \frac{\partial}{\partial \theta} g(y, \theta_0) dP_0(y) \right]^\top \left[\int g(y, \theta_0) g(y, \theta_0)^\top dP_0(y) \right]^{-1} \int \frac{\partial}{\partial \theta} g(y, \theta_0) dP_0(y) \right\}^{-1} \\ & \cdot \left[\int \frac{\partial}{\partial \theta} g(y, \theta_0) dP_0(y) \right]^\top \left[\int g(y, \theta_0) g(y, \theta_0)^\top dP_0(y) \right]^{-1} g(x, \theta_0). \end{aligned} \quad (25)$$

Since the function $x \mapsto g(x, \theta)$ is usually not bounded, for example, when we have linear constraints, the influence function $\text{IF}(x; T_\varphi, P_0)$ is not bounded; therefore, the minimum empirical divergence estimators $\hat{\theta}_\varphi$ defined in (20) are generally not robust.

Through the calculations, it can be seen that there is a connection between the influence functions $\text{IF}(x; t_{\theta_0}, P_0)$ and $\text{IF}(x; T_\varphi, P_0)$, namely the relation

$$\left[\int \frac{\partial}{\partial \theta} \bar{g}(y, \theta_0) dP_0(y) \right]^\top \cdot \left\{ \frac{\partial}{\partial \theta} t(\theta_0, P_0) \text{IF}(x; T_\varphi, P_0) + \text{IF}(x; t_{\theta_0}, P_0) \right\} = 0.$$

Since $\text{IF}(x; T_\varphi, P_0)$ is linearly related to $\text{IF}(x; t_{\theta_0}, P_0)$, using a robust estimator of $t_\theta = t_\theta(P_0)$ in the original duality Formula (15) would lead to a new robust estimator of θ_0 . This is the idea at the basis of our proposal in this paper, for constructing new robust estimators for moment condition models.

3. Robust Estimators for Moment Condition Models

3.1. Definitions of New Estimators

In this section, we define robust versions of the estimators $\hat{t}_\theta$ from (18) and robust versions of minimum empirical divergence estimators $\hat{\theta}_\varphi$ from (20). First, we define robust estimators of t_θ , by using a truncated version of the function $x \mapsto \frac{\partial}{\partial t} m(x, \theta, t)$, and then, we insert such a robust estimator in the estimating equation corresponding to the minimum empirical divergence estimator. The truncated function is based on the multidimensional Huber function and contains a shift vector τ_θ and a scale matrix A_θ to calibrate t_θ and, thus, t_θ , which realizes the supremum in the duality formula, will also be the solution of a new equation based on the new truncated function.

For simplicity, for fixed $\theta \in \Theta$, we also use the notation $m_\theta(x, t) := m(x, \theta, t)$. With this notation, $t_\theta = t_\theta(P_0)$ defined in (16) is the unique solution of the equation:

$$\int \frac{\partial}{\partial t} m_\theta(x, t_\theta(P_0)) dP_0(x) = 0. \quad (26)$$

Consider the system

$$\int \frac{\partial}{\partial t} m_\theta(y, t) dP_0(y) = 0 \quad (27)$$

$$\int H_c(A[\frac{\partial}{\partial t} m_\theta(y, t) - \tau]) dP_0(y) = 0 \quad (28)$$

$$\int H_c(A[\frac{\partial}{\partial t} m_\theta(y, t) - \tau]) H_c(A[\frac{\partial}{\partial t} m_\theta(y, t) - \tau])^\top dP_0(y) = I_{\ell+1} \quad (29)$$

where

$$H_c(y) := \begin{cases} y \cdot \min\left(1, \frac{c}{\|y\|}\right) & \text{if } y \neq 0 \\ 0 & \text{if } y = 0 \end{cases} \quad (30)$$

is the multidimensional Huber function, with $c > 0$, $I_{\ell+1}$ the identity matrix, A is a $(l+1) \times (l+1)$ matrix, and $\tau \in \mathbb{R}^{l+1}$. For fixed θ , this system admits a unique solution $(t, A, \tau) = (t_\theta(P_0), A_\theta(P_0), \tau_\theta(P_0))$ (according to [18], p.17).

The multidimensional Huber function is useful to define robust estimators; it transforms each point outside a hypersphere of c radius to the nearest point of it and leaves those inside unchanged (see [26], p. 239, [27]). By applying the multidimensional Huber function to the function $y \mapsto \frac{\partial}{\partial t} m_\theta(y, t)$, together with considering the scale matrix A_θ and the shift vector τ_θ , a modification is produced there, where the norm exceeds the bound c , and in the meantime, the original t_θ remains the solution of the equation based on the new truncated function. For parametric models, the multidimensional Huber function was also used in other contexts, for example to define optimal B_s -robust estimators or optimal B_i -robust estimators (see [26], p. 244).

The above arguments can be used for each probability measure P from the moment condition model $\mathcal{M}^1$. This context allows defining the truncated version of the function $y \mapsto \frac{\partial}{\partial t} m_\theta(y, t)$, which we denote by $\psi_\theta(y, t)$, such that the original $t_\theta(P_0)$, the solution of Equation (26), is also the solution of the equation $\int \psi_\theta(y, t_\theta(P_0)) dP_0(y) = 0$.

For θ fixed and P a probability measure, the equation $\int \frac{\partial}{\partial t} m_\theta(y, t) dP(y) = 0$ has a unique solution $t = t_\theta(P) \in \Lambda_\theta(P)$ assuring the supremum in the dual form of the divergence $D_\varphi(\mathcal{M}_\theta, P)$ (see [12]). For each t , we define the $A_\theta(t)$ and $\tau_\theta(t)$ solutions of the system:

$$\int H_c(A_\theta(t) [\frac{\partial}{\partial t} m_\theta(y, t) - \tau_\theta(t)]) dP(y) = 0 \quad (31)$$

$$\int H_c(A_\theta(t) [\frac{\partial}{\partial t} m_\theta(y, t) - \tau_\theta(t)]) H_c(A_\theta(t) [\frac{\partial}{\partial t} m_\theta(y, t) - \tau_\theta(t)])^\top dP(y) = I_{\ell+1}. \quad (32)$$

We define a new estimator $\hat{t}_\theta^c$ of $t_\theta = t_\theta(P_0)$, as a Z-estimator corresponding to the ψ -function:

$$\psi_\theta(x, t) := H_c(A_\theta(t) [\frac{\partial}{\partial t} m_\theta(x, t) - \tau_\theta(t)]); \quad (33)$$

more precisely, $\hat{t}_\theta^c$ is defined by

$$\int \psi_\theta(y, \hat{t}_\theta^c) dP_n(y) = 0 \quad \text{or} \quad \sum_{i=1}^n H_c(A_\theta(\hat{t}_\theta^c) [\frac{\partial}{\partial t} m_\theta(X_i, \hat{t}_\theta^c) - \tau_\theta(\hat{t}_\theta^c)]) = 0, \quad (34)$$

the theoretical counterpart of this estimating equation being

$$\int \psi_\theta(y, t_\theta(P_0)) dP_0(y) = 0. \quad (35)$$

For a given probability measure P , the statistical functional $t_\theta^c(P)$ associated with the estimator $\hat{t}_\theta^c$, whenever it exists, is defined by

$$\int \psi_\theta(y, t_\theta^c(P)) dP(y) = \int H_c(A_\theta(t_\theta^c(P)) [\frac{\partial}{\partial t} m_\theta(y, t_\theta^c(P)) - \tau_\theta(t_\theta^c(P))]) dP(y) = 0. \quad (36)$$

Note that

$$t_\theta^c(P_0) = t_\theta(P_0), \quad (37)$$

by construction.

Remark 1. We notice a similarity between the Z-estimator defined in (34) and the classical optimal B_s -robust estimator for parametric models from [26]. In the case of the parametric models, the M-estimator corresponding to the ψ -function (33), but defined for the classical score function $\frac{\partial}{\partial t} (\ln f_t(x)) = \frac{\frac{\partial}{\partial t} f_t(x)}{f_t(x)}$ instead of the function $\frac{\partial}{\partial t} m_\theta(x, t)$ (inclusively in the system (31) and (32) defining $A_\theta(t)$ and $\tau_\theta(t)$), is the classical optimal B_s -robust estimator ($f_t(x)$ denotes the density

corresponding to a parametric model indexed by the parameter t). The classical optimal Bs-robust estimator for parametric models has the optimal property that minimizes a measure of the asymptotic mean-squared error, among all the Fisher-consistent estimators with a self-standardized sensitivity smaller than the positive constant c .

In the following, for a given divergence, using the estimators $\hat{t}_\theta^c$ for $t_\theta(P_0)$, we constructed new estimators of the parameter θ_0 of the model. In Section 3.3, we prove that all the estimators $\hat{t}_\theta^c$ are robust, and this property will be transferred to the new estimators that we define for the parameter θ_0 .

To define new estimators for θ_0 , we used the dual representation (15) of the divergence $D_\varphi(\mathcal{M}_\theta, P_0)$. Since

$$\theta_0 = \arg \inf_{\theta \in \Theta} D_\varphi(\mathcal{M}_\theta, P_0) = \arg \inf_{\theta \in \Theta} \sup_{t \in \Lambda_\theta} \int m_\theta(y, t) dP_0(y) \quad (38)$$

$$= \arg \inf_{\theta \in \Theta} \int m_\theta(y, t_\theta(P_0)) dP_0(y), \quad (39)$$

$\theta = \theta_0$ is the solution of the equation:

$$\int \frac{\partial}{\partial \theta} [m(y, \theta, t(\theta, P_0))] dP_0(y) = 0, \quad (40)$$

where we used the notation $t(\theta, P) := t_\theta(P)$. Equation (40) may be written as

$$\int \frac{\partial}{\partial \theta} m(y, \theta_0, t(\theta_0, P_0)) dP_0(y) + \frac{\partial}{\partial \theta} t(\theta_0, P_0)^\top \int \frac{\partial}{\partial t} m(y, \theta_0, t(\theta_0, P_0)) dP_0(y) = 0. \quad (41)$$

On the basis of the definition of $t_\theta(P_0) = t(\theta, P_0)$, for $\theta = \theta_0$, we have

$$\int \frac{\partial}{\partial t} m(y, \theta_0, t(\theta_0, P_0)) dP_0(y) = 0; \quad (42)$$

therefore, we deduce that $\theta = \theta_0$ is the solution of equation:

$$\int \frac{\partial}{\partial \theta} m(y, \theta, t(\theta, P_0)) dP_0(y) = 0. \quad (43)$$

Using (37), namely $t^c(\theta, P_0) = t(\theta, P_0)$, we obtain that $\theta = \theta_0$ is in fact the solution of equation:

$$\int \frac{\partial}{\partial \theta} m(y, \theta, t^c(\theta, P_0)) dP_0(y) = 0. \quad (44)$$

Then, we define a new estimator $\hat{\theta}_\varphi^c$ of θ_0 , as a plug-in estimator solution of the equation:

$$\int \frac{\partial}{\partial \theta} m(y, \hat{\theta}_\varphi^c, t^c(\hat{\theta}_\varphi^c, P_n)) dP_n(y) = 0. \quad (45)$$

For a probability measure P , the statistical functional T^c corresponding to the estimator $\hat{\theta}_\varphi^c$, whenever it exists, is defined by

$$\int \frac{\partial}{\partial \theta} m(y, T^c(P), t^c(T^c(P), P)) dP(y) = 0. \quad (46)$$

The functional T^c is Fisher-consistent, because

$$T^c(P_0) = \theta_0. \quad (47)$$

This equality is obtained by using (46) for $P = P_0$, the fact that $t^c(T^c(P_0), P_0) = t(T^c(P_0), P_0)$, and the definition of $t_\theta(P_0) = t(\theta, P_0)$ for $\theta = T^c(P_0)$, all these leading to

$$\int \frac{\partial}{\partial \theta} m(y, T^c(P_0), t(T^c(P_0), P_0)) dP_0(y) + \frac{\partial}{\partial \theta} t(T^c(P_0), P_0)^\top \int \frac{\partial}{\partial t} m(y, T^c(P_0), t(T^c(P_0), P_0)) dP_0(y) = 0. \quad (48)$$

Since θ_0 is the unique solution of Equation (41) and, according to (48), $T^c(P_0)$ would be another solution to the same equation, we deduce (47).

From (34) and (45), we have

$$\begin{aligned} \int \psi_{\hat{\theta}_\varphi^c}(y, t^c(\hat{\theta}_\varphi^c, P_n)) dP_n(y) &= 0, \\ \int \frac{\partial}{\partial \theta} m(y, \hat{\theta}_\varphi^c, t^c(\hat{\theta}_\varphi^c, P_n)) dP_n(y) &= 0, \end{aligned}$$

and then,

$$\begin{aligned} \int \psi(y, \hat{\theta}_\varphi^c, \hat{t}_{\hat{\theta}_\varphi^c}^c) dP_n(y) &= 0, \\ \int \frac{\partial}{\partial \theta} m(y, \hat{\theta}_\varphi^c, \hat{t}_{\hat{\theta}_\varphi^c}^c) dP_n(y) &= 0, \end{aligned}$$

with $\psi(y, \theta, t) := \psi_\theta(y, t)$. The couple of estimators $\hat{\theta}_\varphi^c, \hat{t}_{\hat{\theta}_\varphi^c}^c$ can be viewed as a Z-estimator solution of the above system. Denoting

$$\Psi(y, \theta, t) := (\psi(y, \theta, t)^\top, (\frac{\partial}{\partial \theta} m(y, \theta, t))^\top)^\top, \quad (49)$$

the Z-estimators $\hat{\theta}_\varphi^c, \hat{t}_{\hat{\theta}_\varphi^c}^c$ are the solutions of the system:

$$\int \Psi(y, \hat{\theta}_\varphi^c, \hat{t}_{\hat{\theta}_\varphi^c}^c) dP_n(y) = 0, \quad (50)$$

and the theoretical counterpart is given by

$$\int \Psi(y, \theta_0, t_{\theta_0}) dP_0(y) = 0. \quad (51)$$

3.2. Asymptotic Properties

In this section, we establish the consistency and the asymptotic distributions for the estimators $\hat{\theta}_\varphi^c$ and $\hat{t}_{\hat{\theta}_\varphi^c}^c$. In order to prove the consistency of the estimators, we adopted the results from the general theory of Z-estimators as presented for example in [28]. Then, using the consistency of the estimators, as well as supplementary conditions, we proved that the asymptotic distributions of the estimators are multivariate normal:

Assumption 1.

(a) There exist compact neighbourhoods V_{θ_0} of θ_0 and $V_{t_{\theta_0}}$ of t_{θ_0} such that

$$\int \sup_{\theta \in V_{\theta_0}, t \in V_{t_{\theta_0}}} \|\Psi(y, \theta, t)\| dP_0(y) < \infty.$$

(b) For any positive ε , the following condition holds

$$\inf_{(\theta, t) \in M} \left\| \int \Psi(y, \theta, t) dP_0(y) \right\| > 0 = \left\| \int \Psi(y, \theta_0, t_{\theta_0}) dP_0(y) \right\|,$$

where $M := \{(\theta, t) \text{ s.t. } \|(\theta, t) - (\theta_0, t_{\theta_0})\| > \varepsilon\}$.

Proposition 1. Under Assumption 1, $\hat{\theta}_\varphi^c$ converges in probability to θ_0 and $\hat{t}_{\hat{\theta}_\varphi^c}^c$ converges in probability to t_{θ_0} :

Assumption 2.

- (a) Both estimators $\hat{\theta}_\varphi^c$ and $\hat{t}_{\hat{\theta}_\varphi^c}^c$ converge in probability to θ_0 and t_{θ_0} , respectively.
- (b) The function $(\theta, t) \mapsto \psi(x, \theta, t)$ is C^2 on some neighbourhood $V(\theta_0, t_{\theta_0})$ for all x (P_0 a.s.), and the partial derivatives of order two of the functions $\{(\theta, t) \mapsto \psi(x, \theta, t); (\theta, t) \in V(\theta_0, t_{\theta_0})\}$ are dominated by some P_0 -integrable function $H_1(x)$.
- (c) The function $(\theta, t) \mapsto m(x, \theta, t)$ is C^3 on some neighbourhood $U(\theta_0, t_{\theta_0})$ for all x (P_0 a.s.), and the partial derivatives of order three of the functions $\{(\theta, t) \mapsto m(x, \theta, t); (\theta, t) \in U(\theta_0, t_{\theta_0})\}$ are dominated by some P_0 -integrable function $H_2(x)$.
- (d) $\int \|\frac{\partial}{\partial \theta} m(y, \theta_0, t_{\theta_0})\|^2 dP_0(y)$ is finite, and the matrix:

$$S := \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \quad (52)$$

with $S_{11} := (\int \frac{\partial}{\partial t} \psi(y, \theta_0, t_{\theta_0}) dP_0(y))^\top$, $S_{12} := (\int \frac{\partial}{\partial \theta} \psi(y, \theta_0, t_{\theta_0}) dP_0(y))^\top$, $S_{21} := (\int \frac{\partial^2}{\partial \theta \partial t} m(y, \theta_0, t_{\theta_0}) dP_0(y))^\top$ and $S_{22} := \int \frac{\partial^2}{\partial^2 \theta} m(y, \theta_0, t_{\theta_0}) dP_0(y)$, exists and is invertible.

Proposition 2. Let P_0 belong to the model $\mathcal{M}^1$, and suppose that Assumption 2 holds. Then, both $\sqrt{n}(\hat{\theta}_\varphi^c - \theta_0)$ and $\sqrt{n}(\hat{t}_{\hat{\theta}_\varphi^c}^c - t_{\theta_0})$ converge in distribution to a centred multivariate normal variable with covariance matrices given by

$$[[S_{21}S_{11}^{-1}S_{12}]^{-1}S_{21}S_{11}^{-1}] \times [[S_{21}S_{11}^{-1}S_{12}]^{-1}S_{21}S_{11}^{-1}]^\top, \quad (53)$$

and

$$[S_{11}^{-1} - S_{11}^{-1}S_{12}[S_{21}S_{11}^{-1}S_{12}]^{-1}S_{21}S_{11}^{-1}] \times [S_{11}^{-1} - S_{11}^{-1}S_{12}[S_{21}S_{11}^{-1}S_{12}]^{-1}S_{21}S_{11}^{-1}]^\top. \quad (54)$$

The condition of Type (a) from Assumption 1 is usually considered to apply the uniform law of large numbers. For many choices of divergence (for example, those from the Cressie–Read family), the function Ψ is continuous in (θ, t) , and consequently, this condition is verified. The second condition from Assumption 1 is imposed for the uniqueness of (θ_0, t_{θ_0}) as a solution of the equation and is verified, for example, whenever Ψ is continuous and the parameter space is compact ([28], p. 46). Furthermore, the conditions of Type (b)–(d), included in Assumption 2, are often imposed in order to apply the law of large numbers or the central limit theorem and can be verified for the functions appearing in the definitions of estimators proposed in the present paper.

3.3. Influence Functions and Robustness

In this section, we derive the influence functions of the estimators $\hat{t}_\theta^c$ and $\hat{\theta}_\varphi^c$ and prove their B-robustness. The corresponding statistical functionals are defined by (36) and (46), respectively.

Recall that, a map T , defined on a set of probability measures and parameter-space-valued, is a statistical functional corresponding to an estimator $\hat{\theta}$ of the parameter θ_0 from the model P_0 , if $\hat{\theta} = T(P_n)$, P_n being the empirical measure corresponding to the sample. The influence function of T at P_0 is defined by

$$\text{IF}(x; T, P_0) := \left. \frac{\partial T(\tilde{P}_{\varepsilon x})}{\partial \varepsilon} \right|_{\varepsilon=0},$$

where $\tilde{P}_{\varepsilon x} := (1 - \varepsilon)P_0 + \varepsilon\delta_x$, δ_x being the Dirac measure. An unbounded influence function implies an unbounded asymptotic bias of a statistic under single-point contamination of the model. Therefore, a natural robustness requirement on a statistical functional is the

boundedness of its influence function. Whenever the influence function is bounded with respect to x , the corresponding estimator is called B-robust [26].

Proposition 3. For fixed θ , the influence function of the functional t_θ^c is given by

$$\text{IF}(x; t_\theta^c, P_0) = - \left\{ \int \frac{\partial}{\partial t} \psi_\theta(y, t_\theta(P_0)) dP_0(y) \right\}^{-1} \cdot \psi_\theta(x, t_\theta(P_0)). \quad (55)$$

Proposition 4. The influence function of the functional T^c is given by

$$\begin{aligned} \text{IF}(x; T^c, P_0) = & \left\{ \left[\frac{\partial}{\partial \theta} \bar{g}(y, \theta_0) dP_0(y) \right]^\top \left[\int \bar{g}(y, \theta_0) \bar{g}(y, \theta_0)^\top dP_0(y) \right]^{-1} \left[\frac{\partial}{\partial \theta} \bar{g}(y, \theta_0) dP_0(y) \right] \right\}^{-1} \\ & \cdot \left[\frac{\partial}{\partial \theta} \bar{g}(y, \theta_0) dP_0(y) \right]^\top \frac{1}{\varphi''(1)} \text{IF}(x; t_{\theta_0}^c, P_0). \end{aligned} \quad (56)$$

On the basis of Proposition 3 and Proposition 4, since $x \mapsto \psi_\theta(x, t_\theta(P_0))$ is bounded, all the estimators $\hat{\theta}_\varphi^c$ are B-robust.

4. Conclusions

We introduced a class of robust Z-estimators for moment condition models. These new estimators can be seen as robust alternatives for the minimum empirical divergence estimators. By using truncated functions based on the multidimensional Huber function, we defined robust estimators of the element that realizes the supremum in the dual form of the divergence, as well as new robust estimators for the parameter of the model. The asymptotic properties were proven, including the consistency and the limit laws. The influence functions for all the proposed estimators are bounded; therefore, these estimators are B-robust. The truncated function that we used to define the new robust Z-estimators contains functions implicitly defined, for which analytic forms are not available. The implementation of the estimation method will be addressed in a future research study. The idea of using the multidimensional Huber function, together with a scale matrix and a shift vector, to create a bounded version of the function corresponding to the estimating equation for the parameter of interest, could be considered in other contexts as well and would lead to new robust Z-estimators. As one of the Referees suggested, some other bounded functions could be used to define new robust Z-estimators for moment condition models. For example, the Tukey biweight function used together with a norm inside, in order to be appropriate to be applied to functions with vector values, could also be considered. Again, the original parameter of interest should remain the solution of the estimating equation based on the new bounded function. Such an idea is interesting to be analysed in future studies, in order to provide new robust versions of minimum empirical divergence estimators or robust Z-estimators in other contexts.

Funding: 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.

Data Availability Statement: Not applicable.

Acknowledgments: We are very grateful to the Referees for their helpful comments and suggestions.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

i.i.d.	independent and identically distributed
a.c.	absolutely continuous
GMM	generalized method of moments
CU	continuous updating
EL	empirical likelihood
ET	exponential tilting
GEL	generalized empirical likelihood
ETEL	exponentially tilted empirical likelihood

Appendix A

Proof of Proposition 1. Since $(\theta, t) \mapsto \Psi(y, \theta, t)$ is continuous, by the uniform law of large numbers, Assumption 1 (a) implies

$$\sup_{\theta \in V_{\theta_0}, t \in V_{t_{\theta_0}}} \left\| \int \Psi(y, \theta, t) dP_n(y) - \int \Psi(y, \theta, t) dP_0(y) \right\| \rightarrow 0, \quad (\text{A1})$$

in probability. This result together with Assumption 1 (b) ensures the convergence in probability of the estimators $\hat{\theta}_\varphi^c$ and $\hat{t}_{\hat{\theta}_\varphi^c}^c$ toward θ_0 and t_{θ_0} , respectively. The proof is the same as the one for Theorem 5.9 from [28], p. 46. $\square$

Proof of Proposition 2. By the definitions of $\hat{\theta}_\varphi^c$ and $\hat{t}_{\hat{\theta}_\varphi^c}^c$, they both satisfy

$$\int \psi(y, \hat{\theta}_\varphi^c, \hat{t}_{\hat{\theta}_\varphi^c}^c) dP_n(y) = 0 \quad (\text{E1})$$

$$\int \frac{\partial}{\partial \theta} m(y, \hat{\theta}_\varphi^c, \hat{t}_{\hat{\theta}_\varphi^c}^c) dP_n(y) = 0 \quad (\text{E2})$$

Using a Taylor expansion in (E1), there exists $(\tilde{\theta}_\varphi^c, \tilde{t}_\varphi^c)$ inside the segment that links $(\hat{\theta}_\varphi^c, \hat{t}_{\hat{\theta}_\varphi^c}^c)$ and (θ_0, t_{θ_0}) such that

$$0 = \int \psi(y, \theta_0, t_{\theta_0}) dP_n(y) + \left[\left(\int \frac{\partial}{\partial t} \psi(y, \theta_0, t_{\theta_0}) dP_n(y) \right)^\top, \left(\int \frac{\partial}{\partial \theta} \psi(y, \theta_0, t_{\theta_0}) dP_n(y) \right)^\top \right] \cdot a_n + \frac{1}{2} a_n^\top A_n a_n, \quad (\text{A2})$$

where

$$a_n := ((\hat{t}_{\hat{\theta}_\varphi^c}^c - t_{\theta_0})^\top, (\hat{\theta}_\varphi^c - \theta_0)^\top)^\top, \quad (\text{A3})$$

and

$$A_n := \begin{pmatrix} \int \frac{\partial^2}{\partial t^2} \psi(y, \tilde{\theta}_\varphi^c, \tilde{t}_\varphi^c) dP_n(y) & \int \frac{\partial^2}{\partial \theta \partial t} \psi(y, \tilde{\theta}_\varphi^c, \tilde{t}_\varphi^c) dP_n(y) \\ \int \frac{\partial^2}{\partial t \partial \theta} \psi(y, \tilde{\theta}_\varphi^c, \tilde{t}_\varphi^c) dP_n(y) & \int \frac{\partial^2}{\partial \theta^2} \psi(y, \tilde{\theta}_\varphi^c, \tilde{t}_\varphi^c) dP_n(y) \end{pmatrix}. \quad (\text{A4})$$

By Assumption 2 (b), the law of large numbers implies that $A_n = O_P(1)$. Then, using Assumption 2 (a), the last term in (A2) can be written $o_P(1)a_n$. On the other hand, by Assumption 2 (d), using the law of large numbers, we can write

$$\begin{aligned} & \left[\left(\int \frac{\partial}{\partial t} \psi(y, \theta_0, t_{\theta_0}) dP_n(y) \right)^\top, \left(\int \frac{\partial}{\partial \theta} \psi(y, \theta_0, t_{\theta_0}) dP_n(y) \right)^\top \right] = \\ & = \left[\left(\int \frac{\partial}{\partial t} \psi(y, \theta_0, t_{\theta_0}) dP_0(y) \right)^\top, \left(\int \frac{\partial}{\partial \theta} \psi(y, \theta_0, t_{\theta_0}) dP_0(y) \right)^\top \right] + o_P(1). \end{aligned}$$

Consequently, (A2) becomes

$$\begin{aligned} & - \int \psi(y, \theta_0, t_{\theta_0}) dP_n(y) = \\ & = \left[\left(\int \frac{\partial}{\partial t} \psi(y, \theta_0, t_{\theta_0}) dP_0(y) \right)^\top + o_P(1), \left(\int \frac{\partial}{\partial \theta} \psi(y, \theta_0, t_{\theta_0}) dP_0(y) \right)^\top + o_P(1) \right] \cdot a_n. \quad (\text{A5}) \end{aligned}$$

In the same way, using a Taylor expansion in (E2), there exists $(\bar{\theta}_\varphi^c, \bar{t}_\varphi^c)$ inside the segment that links $(\hat{\theta}_\varphi^c, \hat{t}_{\hat{\theta}_\varphi^c}^c)$ and (θ_0, t_{θ_0}) such that

$$0 = \int \frac{\partial}{\partial \theta} m(y, \theta_0, t_{\theta_0}) dP_n(y) + \left[\left(\int \frac{\partial^2}{\partial \theta \partial t} m(y, \theta_0, t_{\theta_0}) dP_n(y) \right)^\top, \right. \\ \left. \left(\int \frac{\partial^2}{\partial^2 \theta} m(y, \theta_0, t_{\theta_0}) dP_n(y) \right)^\top \right] \cdot a_n + \frac{1}{2} a_n^\top B_n a_n, \quad (\text{A6})$$

where

$$B_n := \begin{pmatrix} \int \frac{\partial^3}{\partial \theta \partial^2 t} m(y, \bar{\theta}_\varphi^c, \bar{t}_\varphi^c) dP_n(y) & \int \frac{\partial^3}{\partial^2 \theta \partial t} m(y, \bar{\theta}_\varphi^c, \bar{t}_\varphi^c) dP_n(y) \\ \int \frac{\partial^3}{\partial \theta \partial t \partial \theta} m(y, \bar{\theta}_\varphi^c, \bar{t}_\varphi^c) dP_n(y) & \int \frac{\partial^3}{\partial^3 \theta} m(y, \bar{\theta}_\varphi^c, \bar{t}_\varphi^c) dP_n(y) \end{pmatrix}. \quad (\text{A7})$$

Similarly, as in (A5), we obtain

$$- \int \frac{\partial}{\partial \theta} m(y, \theta_0, t_{\theta_0}) dP_n(y) = \\ = \left[\left(\int \frac{\partial^2}{\partial \theta \partial t} m(y, \theta_0, t_{\theta_0}) dP_0(y) \right)^\top + o_P(1), \int \frac{\partial^2}{\partial^2 \theta} m(y, \theta_0, t_{\theta_0}) dP_0(y) + o_P(1) \right] \cdot a_n. \quad (\text{A8})$$

Using (A5) and (A8), we obtain

$$\sqrt{n} a_n = \sqrt{n} \begin{pmatrix} \left(\int \frac{\partial}{\partial t} \psi(y, \theta_0, t_{\theta_0}) dP_0(y) \right)^\top & \left(\int \frac{\partial}{\partial \theta} \psi(y, \theta_0, t_{\theta_0}) dP_0(y) \right)^\top \\ \left(\int \frac{\partial^2}{\partial \theta \partial t} m(y, \theta_0, t_{\theta_0}) dP_n(y) \right)^\top & \int \frac{\partial^2}{\partial^2 \theta} m(y, \theta_0, t_{\theta_0}) dP_n(y) \end{pmatrix}^{-1} \times \\ \times \begin{pmatrix} - \int \psi(y, \theta_0, t_{\theta_0}) dP_n(y) \\ - \int \frac{\partial}{\partial \theta} m(y, \theta_0, t_{\theta_0}) dP_n(y) \end{pmatrix} + o_P(1). \quad (\text{A9})$$

Consider S the $(l+1+d) \times (l+1+d)$ matrix:

$$S := \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \quad (\text{A10})$$

with $S_{11} := \left(\int \frac{\partial}{\partial t} \psi(y, \theta_0, t_{\theta_0}) dP_0(y) \right)^\top$, $S_{12} := \left(\int \frac{\partial}{\partial \theta} \psi(y, \theta_0, t_{\theta_0}) dP_0(y) \right)^\top$, $S_{21} := \left(\int \frac{\partial^2}{\partial \theta \partial t} m(y, \theta_0, t_{\theta_0}) dP_0(y) \right)^\top$, and $S_{22} := \int \frac{\partial^2}{\partial^2 \theta} m(y, \theta_0, t_{\theta_0}) dP_0(y)$. Through calculations, we have

$$S_{21} = -[0_d, \int \frac{\partial}{\partial \theta} g(y, \theta_0) dP_0(y)], \quad (\text{A11})$$

$$S_{22} = [0_d, \dots, 0_d]. \quad (\text{A12})$$

From (A9), we deduce that

$$\sqrt{n} \begin{pmatrix} \hat{t}_{\hat{\theta}_\varphi^c}^c - t_{\theta_0} \\ \hat{\theta}_\varphi^c - \theta_0 \end{pmatrix} = S^{-1} \sqrt{n} \begin{pmatrix} - \int \psi(y, \theta_0, t_{\theta_0}) dP_n(y) \\ 0_d \end{pmatrix} + o_P(1). \quad (\text{A13})$$

On the other hand, under assumption Assumption 2 (d), using the central limit theorem,

$$\sqrt{n} \begin{pmatrix} - \int \psi(y, \theta_0, t_{\theta_0}) dP_n(y) \\ 0_d \end{pmatrix} \quad (\text{A14})$$

converges in distribution to a centred multivariate normal variable with covariance matrix:

$$M := \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \quad (\text{A15})$$

with

$$M_{11} := \text{cov}[\psi(X, \theta_0, t_{\theta_0})], \quad M_{12} := \begin{pmatrix} 0_d^\top \\ \vdots \\ 0_d^\top \end{pmatrix}, \quad M_{21} := \begin{pmatrix} 0 & 0_l^\top \\ \vdots & \vdots \\ 0 & 0_l^\top \end{pmatrix}, \quad M_{22} := \begin{pmatrix} 0_d^\top \\ \vdots \\ 0_d^\top \end{pmatrix}.$$

Since $E[\psi(X, \theta_0, t_{\theta_0})] = 0$ by the construction of ψ , we obtain

$$M_{11} = \text{cov}[\psi(X, \theta_0, t_{\theta_0})] = \int \psi(y, \theta_0, t_{\theta_0}) \psi(y, \theta_0, t_{\theta_0})^\top dP_0(y) = I_{l+1}, \quad (\text{A16})$$

on the basis of (29) for $\theta = \theta_0$.

Using then (A13) and the Slutsky theorem, we obtain that

$$\sqrt{n} \begin{pmatrix} \hat{t}_{\hat{\theta}_\varphi^c}^c - t_{\theta_0} \\ \hat{\theta}_\varphi^c - \theta_0 \end{pmatrix} \quad (\text{A17})$$

converges in distribution to a centred multivariate normal variable with the covariance matrix given by

$$C = S^{-1} M [S^{-1}]^\top. \quad (\text{A18})$$

If we denote

$$C := \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, \quad (\text{A19})$$

through calculation, we obtain

$$C_{11} = [S_{11}^{-1} - S_{11}^{-1} S_{12} [S_{21} S_{11}^{-1} S_{12}]^{-1} S_{21} S_{11}^{-1}] \times [S_{11}^{-1} - S_{11}^{-1} S_{12} [S_{21} S_{11}^{-1} S_{12}]^{-1} S_{21} S_{11}^{-1}]^\top, \quad (\text{A20})$$

$$C_{12} = [S_{11}^{-1} - S_{11}^{-1} S_{12} [S_{21} S_{11}^{-1} S_{12}]^{-1} S_{21} S_{11}^{-1}] \times [[S_{21} S_{11}^{-1} S_{12}]^{-1} S_{21} S_{11}^{-1}]^\top, \quad (\text{A21})$$

$$C_{21} = [[S_{21} S_{11}^{-1} S_{12}]^{-1} S_{21} S_{11}^{-1}] \times [S_{11}^{-1} - S_{11}^{-1} S_{12} [S_{21} S_{11}^{-1} S_{12}]^{-1} S_{21} S_{11}^{-1}]^\top, \quad (\text{A22})$$

$$C_{22} = [[S_{21} S_{11}^{-1} S_{12}]^{-1} S_{21} S_{11}^{-1}] \times [[S_{21} S_{11}^{-1} S_{12}]^{-1} S_{21} S_{11}^{-1}]^\top. \quad (\text{A23})$$

□

Proof of Proposition 3. For the contaminated model $\tilde{P}_{\varepsilon x} = (1 - \varepsilon)P_0 + \varepsilon\delta_x$, whenever it exists, $t_\theta^c(\tilde{P}_{\varepsilon x})$ is defined as the solution of equation:

$$\int H_c(A_\theta(t_\theta^c(\tilde{P}_{\varepsilon x}))) \left[\frac{\partial}{\partial t} m_\theta(y, t_\theta^c(\tilde{P}_{\varepsilon x})) - \tau_\theta(t_\theta^c(\tilde{P}_{\varepsilon x})) \right] d\tilde{P}_{\varepsilon x}(y) = 0. \quad (\text{A24})$$

It follows that

$$(1 - \varepsilon) \int H_c(A_\theta(t_\theta^c(\tilde{P}_{\varepsilon x}))) \left[\frac{\partial}{\partial t} m_\theta(y, t_\theta^c(\tilde{P}_{\varepsilon x})) - \tau_\theta(t_\theta^c(\tilde{P}_{\varepsilon x})) \right] dP_0(y) + \varepsilon H_c(A_\theta(t_\theta^c(\tilde{P}_{\varepsilon x}))) \left[\frac{\partial}{\partial t} m_\theta(x, t_\theta^c(\tilde{P}_{\varepsilon x})) - \tau_\theta(t_\theta^c(\tilde{P}_{\varepsilon x})) \right] = 0. \quad (\text{A25})$$

Derivation with respect to ε in (A25) yields

$$\begin{aligned} & - \int H_c(A_\theta(t_\theta(P_0))) \left[\frac{\partial}{\partial t} m_\theta(y, t_\theta(P_0)) - \tau_\theta(t_\theta(P_0)) \right] dP_0(y) + \\ & + \int \frac{\partial}{\partial t} [H_c(A_\theta(t)) \left[\frac{\partial}{\partial t} m_\theta(y, t) - \tau_\theta(t) \right]]_{t=t_\theta(P_0)} dP_0(y) \text{IF}(x; t_\theta^c, P_0) + \\ & + H_c(A_\theta(t_\theta(P_0))) \left[\frac{\partial}{\partial t} m_\theta(x, t_\theta(P_0)) - \tau_\theta(t_\theta(P_0)) \right]. \end{aligned} \quad (\text{A26})$$

Since the first integral in (A26) equals zero, we obtain

$$\begin{aligned} \text{IF}(x; t_\theta^c, P_0) &= - \left\{ \int \frac{\partial}{\partial t} [H_c(A_\theta(t)) \left[\frac{\partial}{\partial t} m_\theta(y, t) - \tau_\theta(t) \right]]_{t=t_\theta(P_0)} dP_0(y) \right\}^{-1} \cdot \\ & \cdot H_c(A_\theta(t_\theta(P_0))) \left[\frac{\partial}{\partial t} m_\theta(x, t_\theta(P_0)) - \tau_\theta(t_\theta(P_0)) \right] \\ &= - \left\{ \int \frac{\partial}{\partial t} \psi_\theta(y, t_\theta(P_0)) dP_0(y) \right\}^{-1} \cdot \psi_\theta(x, t_\theta(P_0)). \end{aligned} \quad (\text{A27})$$

For each θ , the influence function (55) is bounded with respect to x ; therefore, the estimators $\hat{t}_\theta^c$ are B-robust.

□

Proof of Proposition 4. For the contaminated model $\tilde{P}_{\varepsilon x} = (1 - \varepsilon)P_0 + \varepsilon\delta_x$, $T^c(\tilde{P}_{\varepsilon x})$ is defined as the solution of equation:

$$\int \frac{\partial}{\partial \theta} m(y, T^c(\tilde{P}_{\varepsilon x}), t^c(T^c(\tilde{P}_{\varepsilon x}), \tilde{P}_{\varepsilon x})) d\tilde{P}_{\varepsilon x}(y) = 0, \quad (\text{A28})$$

whenever this solution exists. Then,

$$(1 - \varepsilon) \int \frac{\partial}{\partial \theta} m(y, T^c(\tilde{P}_{\varepsilon x}), t^c(T^c(\tilde{P}_{\varepsilon x}), \tilde{P}_{\varepsilon x})) dP_0(y) + \varepsilon \frac{\partial}{\partial \theta} m(x, T^c(\tilde{P}_{\varepsilon x}), t^c(T^c(\tilde{P}_{\varepsilon x}), \tilde{P}_{\varepsilon x})) = 0. \quad (\text{A29})$$

Derivation with respect to ε in (A29) yields

$$\begin{aligned} & - \int \frac{\partial}{\partial \theta} m(y, \theta_0, t^c(\theta_0, P_0)) dP_0(y) + \int \frac{\partial^2}{\partial^2 \theta} m[y, \theta, t]_{\theta=\theta_0, t=t_{\theta_0}(P_0)} dP_0(y) \text{IF}(x; T^c, P_0) + \\ & + \int \frac{\partial^2}{\partial \theta \partial t} m[y, \theta, t]_{\theta=\theta_0, t=t_{\theta_0}(P_0)} dP_0(y) \cdot \left\{ \frac{\partial}{\partial \theta} t^c(\theta_0, P_0) \text{IF}(x; T^c, P_0) + \right. \\ & \left. + \text{IF}(x; t_{\theta_0}^c, P_0) \right\} + \frac{\partial}{\partial \theta} [m(x, \theta, t)]_{\theta=\theta_0, t=t_{\theta_0}(P_0)} = 0. \end{aligned} \quad (\text{A30})$$

Some calculations show that

$$\frac{\partial}{\partial \theta} [m(x, \theta, t)]_{\theta=\theta_0, t=t_{\theta_0}(P_0)} = 0 \quad \text{and} \quad \frac{\partial^2}{\partial^2 \theta} [m(x, \theta, t)]_{\theta=\theta_0, t=t_{\theta_0}(P_0)} = 0, \quad (\text{A31})$$

for any x ; therefore, (A30) reduces to

$$\int \frac{\partial^2}{\partial \theta \partial t} m[y, \theta, t]_{\theta=\theta_0, t=t_{\theta_0}(P_0)} dP_0(y) \cdot \left\{ \frac{\partial}{\partial \theta} t^c(\theta_0, P_0) \text{IF}(x; T^c, P_0) + \text{IF}(x; t_{\theta_0}^c, P_0) \right\} = 0. \quad (\text{A32})$$

On the other hand,

$$\begin{aligned} \int \frac{\partial^2}{\partial \theta \partial t} m[y, \theta, t]_{\theta=\theta_0, t=t_{\theta_0}(P_0)} dP_0(y) &= -\psi'(\varphi'(1)) \int \frac{\partial}{\partial \theta} \bar{g}(y, \theta_0) dP_0(y) \\ &= - \int \frac{\partial}{\partial \theta} \bar{g}(y, \theta_0) dP_0(y), \end{aligned}$$

since $\psi'(u) = \varphi'^{-1}(u)$.

Taking into account that $t^c(\theta, P_0) = t(\theta, P_0)$ and $t(\theta, P_0)$ verifies

$$\int \frac{\partial}{\partial t} m(y, \theta, t(\theta, P_0)) dP_0(y) = 0, \quad (\text{A33})$$

and the derivation with respect to θ yields

$$\int \frac{\partial^2}{\partial t \partial \theta} m(y, \theta, t(\theta, P_0)) dP_0(y) + \int \frac{\partial^2}{\partial^2 t} m(y, \theta, t(\theta, P_0)) dP_0(y) \cdot \frac{\partial}{\partial \theta} t(\theta, P_0) = 0, \quad (\text{A34})$$

which implies

$$\begin{aligned} \frac{\partial}{\partial \theta} t^c(\theta_0, P_0) &= \frac{\partial}{\partial \theta} t(\theta_0, P_0) = - \left\{ \int \frac{\partial^2}{\partial^2 t} m(y, \theta_0, t(\theta_0, P_0)) dP_0(y) \right\}^{-1} \int \frac{\partial^2}{\partial t \partial \theta} m(y, \theta_0, t(\theta_0, P_0)) dP_0(y) \\ &= -\varphi''(1) \left\{ \int \bar{g}(y, \theta_0) \bar{g}(y, \theta_0)^\top dP_0(y) \right\}^{-1} \int \frac{\partial}{\partial \theta} \bar{g}(y, \theta_0) dP_0(y), \end{aligned}$$

because

$$\int \frac{\partial^2}{\partial^2 t} m(y, \theta_0, t(\theta_0, P_0)) dP_0(y) = -\frac{1}{\varphi''(1)} \int \bar{g}(y, \theta_0) \bar{g}(y, \theta_0)^\top dP_0(y). \quad (\text{A35})$$

Then, (A32) becomes

$$\left[\int \frac{\partial}{\partial \theta} \bar{g}(y, \theta_0) dP_0(y) \right]^\top \left[-\varphi''(1) \left\{ \int \bar{g}(y, \theta_0) \bar{g}(y, \theta_0)^\top dP_0(y) \right\}^{-1} \int \frac{\partial}{\partial \theta} \bar{g}(y, \theta_0) dP_0(y) \text{IF}(x; T^c, P_0) + \right. \\ \left. + \text{IF}(x; t_{\theta_0}^c, P_0) \right] = 0,$$

and consequently,

$$\begin{aligned} \text{IF}(x; T^c, P_0) = & \left\{ \left[\frac{\partial}{\partial \theta} \bar{g}(y, \theta_0) dP_0(y) \right]^\top \left[\int \bar{g}(y, \theta_0) \bar{g}(y, \theta_0)^\top dP_0(y) \right]^{-1} \left[\frac{\partial}{\partial \theta} \bar{g}(y, \theta_0) dP_0(y) \right] \right\}^{-1} \\ & \cdot \left[\frac{\partial}{\partial \theta} \bar{g}(y, \theta_0) dP_0(y) \right]^\top \frac{1}{\varphi''(1)} \text{IF}(x; t_{\theta_0}^c, P_0). \end{aligned} \quad (\text{A36})$$

□

References

- Hansen, L. P. Large sample properties of generalized method of moments estimators. *Econometrica* **1982**, *50*, 1029–1054.
- Hansen, L.; Heaton, J.; Yaron, A. Finite-sample properties of some alternative gmm 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.
- Imbens, G.W. One-step estimators for over-identified generalized method of moments models. *Rev. Econ. Stud.* **1997**, *64*, 359–383.
- Kitamura, Y.; Stutzer, M. An information-theoretic alternative to generalized method of moments estimation. *Econometrica* **1997**, *65*, 861–874.
- Newey, W.K.; Smith, R.J. Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **2004**, *72*, 219–255.
- Schennach, S.M. Point estimation with exponentially tilted empirical likelihood. *Ann. Stat.* **2007**, *35*, 634–672.
- Pardo, L. *Statistical Inference Based on Divergence Measures*; Chapman & Hall: London, UK, 2006.
- Basu, A.; Shioya, H.; Park, C. *Statistical Inference: The Minimum Distance Approach*; Chapman & Hall: London, UK, 2011.
- Pardo, L.; Martín, N. Robust procedures for estimating and testing in the framework of divergence measures. *Entropy* **2021**, *23*, 430.
- Riani, M.; Atkinson, A.C.; Corbellini, A.; Perrotta, D. Robust regression with density power divergence: Theory, comparisons, and data analysis. *Entropy* **2020**, *22*, 399.
- Broniatowski, M.; Keziou, A. Divergences and duality for estimation and test under moment condition models. *J. Stat. Plan. Inference* **2012**, *142*, 2554–2573.
- Broniatowski, M.; Keziou, A. Parametric estimation and tests through divergences and the duality technique. *J. Multivar. Anal.* **2009**, *100*, 16–36.
- Toma, A.; Broniatowski, M. Dual divergence estimators and tests: Robustness results. *J. Multivar. Anal.* **2011**, *102*, 20–36.
- Toma, A.; Leoni-Aubin, S. Robust tests based on dual divergence estimators and saddlepoint approximations. *J. Multivar. Anal.* **2010**, *101*, 1143–1155.
- Toma, A. Model selection criteria using divergences. *Entropy* **2014**, *16*, 2686–2698.
- Toma, A. Robustness of dual divergence estimators for models satisfying linear constraints. *C. R. Math. Acad. Sci. Paris* **2013**, *351*, 311–316.
- Ronchetti, E.; Trojani, F. Robust inference with GMM estimators. *J. Econom.* **2001**, *101*, 37–69.
- Lô, S.N.; Ronchetti, E. Robust small sample accurate inference in moment condition models. *Comput. Stat. Data Anal.* **2012**, *56*, 3182–3197.
- Felipe, A.; Martín, N.; Miranda, P.; Pardo, L. Testing with exponentially tilted empirical likelihood. *Methodol. Comput. Appl. Probab.* **2018**, *20*, 1319–1358.
- Keziou, A.; Toma, A. A robust version of the empirical likelihood estimator. *Mathematics* **2021**, *9*, 829.
- Keziou, A.; Toma, A. Robust Empirical Likelihood. In *Geometric Science of Information, Proceedings of the 5th International Conference, GSI 2021, Paris, France, 21–23 July 2021*; Springer International Publishing: Cham, Switzerland, 2021.
- Rüschendorf, L. On the minimum discrimination information theorem. *Stat. Decis.* **1984**, 263–283.
- Cressie, N.; Read, T.R.C. Multinomial goodness-of-fit tests. *J. R. Stat. Soc. Ser. B* **1984**, *46*, 440–464.
- Broniatowski, M.; Keziou, A. Minimization of ϕ divergences on sets of signed measures. *Stud. Sci. Math. Hungar.* **2006**, *43*, 403–442.
- Hampel, F.R.; Ronchetti, E.; Rousseeuw, P.J.; Stahel, W. *Robust Statistics: The Approach Based on Influence Functions*; Wiley: New York, NY, USA, 1986.
- Ronchetti, E. M.; Huber, P.J. *Robust Statistics*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
- van der Vaart, A. W. *Asymptotic Statistics*; Cambridge Series in Statistical and Probabilistic Mathematics; Cambridge University Press: Cambridge, UK, 1998.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.