Dual divergence estimators and tests: robustness results

Aida Toma, Michel Broniatowski

Dual divergence estimators and tests: robustness results

Aida Toma [a,b,*], Michel Broniatowski [c]

[a] *Mathematics Department, Academy of Economic Studies, Piaţa Romană 6, Bucharest, Romania, e-mail: aida_toma@yahoo.com*

[b] *"Gheorghe Mihoc - Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Calea 13 Septembrie 13, Bucharest, Romania*

[c] *LSTA, Université Paris 6, 175 Rue du Chevaleret, 75013 Paris, France, e-mail: michel.broniatowski@upmc.fr*

Abstract

The class of dual ϕ-divergence estimators (introduced in Broniatowski and Keziou (2009) [6]) is explored with respect to robustness through the influence function approach. For scale and location models, this class is investigated in terms of robustness and asymptotic relative efficiency. Some hypothesis tests based on dual divergence criterions are proposed and their robustness properties are studied. The empirical performances of these estimators and tests are illustrated by Monte Carlo simulation for both noncontaminated and contaminated data.

Key words: Location model, minimum divergence estimators, robust estimation, robust test, scale model

1 Introduction

Minimum divergence estimators and related methods have received considerable attention in statistical inference because of their ability to reconcile efficiency and robustness. Among others, Beran [3], Tamura and Boos [22], Simpson [20,21] and Toma [23] proposed families of parametric estimators minimizing the Hellinger distance between a nonparametric estimator of the observations density and the model. They showed that those estimators are both asymptotically efficient and robust. Generalizing earlier work based on

* Corresponding author. Mathematics Department, Academy of Economic Studies, Piaţa Romană 6, Bucharest, Romania. E-mail: aida_toma@yahoo.com

the Hellinger distance, Lindsay [17], Basu and Lindsay [2], Morales et al. [18] have investigated minimum divergence estimators, for both discrete and continuous models. Some families of estimators based on approximate divergence criterions have also been considered; see Basu et al. [1].

Broniatowski and Keziou [6] have introduced a new minimum divergence estimation method based on a dual representation of the divergence between probability measures. Their estimators are defined in an unified way for both continuous and discrete models. They do not require any prior smoothing and include the classical maximum likelihood estimators as a benchmark. A special case for the Kullback-Leibler divergence is presented in Broniatowski [4]. The present paper presents robustness studies for the classes of estimators generated by the minimum dual ϕ-divergence method, as well as for some tests based on corresponding estimators of the divergence criterion.

We give general results that allow to identify robust estimators in the class of dual ϕ-divergence estimators. We apply this study for the Cressie-Read divergences and state explicit robustness results for scale models and location models. Gain in robustness is often paid by some loss in efficiency. This is discussed for some scale and location models. Our main remarks are as follows. All the relevant information pertaining to the model and the true value of the parameter to be estimated should be used in order to define, when possible, robust and nearly efficient procedures. Some models allow for such procedures. The example provided by the scale normal model shows that the choice of a good estimation criterion is heavily dependent on the acceptable loss in efficiency in order to achieve a compromise with the robustness requirement. When sampling under the model is overspread (typically for Cauchy and logistic models), non surprisingly the maximum likelihood estimator is both efficient and robust and therefore should be prefered (see subsection 3.2).

On the other hand, these estimation results constitute the premises to construct some robust tests. The purpose of robust testing is twofold. First, the level of a test should be stable under small arbitrary departures from the null hypothesis (i.e. robustness of validity). Second, the test should have a good power under small arbitrary departures from specified alternatives (i.e. robustness of efficiency). To control the test stability against outliers in the aforementioned senses, we compute the asymptotic level of the test under a sequence of contaminated null distributions, as well as the asymptotic power of the test under a sequence of contaminated alternatives. These quantities are seen to be controlled by the influence function of the test statistic. In this way, the robustness of the test is a consequence of the robustness of the test statistic based on a dual ϕ-divergence estimator. In many cases, this requirement is met when the dual ϕ-divergence estimator itself is robust.

The paper is organized as follows: in Section 2 we present the classes of estimators generated by the minimum dual ϕ-divergence method. In Section 3, for these estimators, we compute the influence functions and give the Fisher consistency. We particularize this study for the Cressie-Read divergences and state robustness results for scale models and location models. Section 4 is devoted to hypothesis testing. We give general convergence results for contaminated observations and use it to compute the asymptotic level and the asymptotic power for the tests that we propose. In Section 5, the performances of the estimators and tests are illustrated by Monte Carlo simulation studies. In Section 6 we shortly presents a proposal for the adaptive choice of tuning parameters.

2 Minimum divergence estimators

2.1 Minimum divergence estimators

Let φ be a non-negative convex function defined from $(0, \infty)$ onto $[0, \infty]$ and satisfying $\varphi(1) = 0$. Also extend φ at 0 defining $\varphi(0) = \lim_{x\downarrow 0} \varphi(x)$. Let $(\mathcal{X}, \mathcal{B})$ be a measurable space and P be a probability measure (p.m.) defined on $(\mathcal{X}, \mathcal{B})$. Following Rüschendorf [19], for any p.m. Q absolutely continuous (a.c.) w.r.t. P, the ϕ-divergence between Q and P is defined by

$$\phi(Q, P) := \int \varphi\left(\frac{dQ}{dP}\right) dP. \tag{1}$$

When Q is not a.c. w.r.t. P, we set $\phi(Q, P) = \infty$. We refer to Liese and Vajda [16] for an overview on the origin of the concept of divergence in Statistics.

A commonly used family of divergences is the so-called "power divergences", introduced by Cressie and Read [9] and defined by the class of functions

$$x \in \mathbb{R}_+^* \mapsto \varphi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)} \tag{2}$$

for $\gamma \in \mathbb{R} \setminus \{0, 1\}$ and $\varphi_0(x) := -\log x + x - 1$, $\varphi_1(x) := x \log x - x + 1$ with $\varphi_\gamma(0) = \lim_{x\downarrow 0} \varphi_\gamma(x)$, $\varphi_\gamma(\infty) = \lim_{x\to\infty} \varphi_\gamma(x)$, for any $\gamma \in \mathbb{R}$. The Kullback-Leibler divergence (KL) is associated with φ_1, the modified Kullback-Leibler (KL_m) to φ_0, the χ^2 divergence to φ_2, the modified χ^2 divergence (χ^2_m) to φ_{-1} and the Hellinger distance to $\varphi_{1/2}$.

Let $\{P_\theta : \theta \in \Theta\}$ be some identifiable parametric model with Θ a subset of $\mathbb{R}^d$. Consider the problem of estimation of the unknown true value of the parameter θ_0 on the basis of an i.i.d. sample $X_1, \ldots, X_n$ with p.m. P_{θ_0}.

When all p.m. P_θ share the same finite support S which is independent upon the parameter θ, the ϕ-divergence between P_θ and P_{θ_0} has the form

$$\phi(P_\theta, P_{\theta_0}) = \sum_{j\in S} \varphi\left(\frac{P_\theta(j)}{P_{\theta_0}(j)}\right) P_{\theta_0}(j).$$

In this case, Liese and Vajda [15], Lindsay [17] and Morales et al. [18] investigated the so-called "minimum ϕ-divergence estimators" (minimum disparity estimators in Lindsay [17]) of the parameter θ_0 defined by

$$\widehat{\theta}_n := \arg\inf_{\theta\in\Theta} \phi(P_\theta, P_n), \tag{3}$$

where $\phi(P_\theta, P_n)$ is the plug-in estimator of $\phi(P_\theta, P_{\theta_0})$

$$\phi(P_\theta, P_n) = \sum_{j\in S} \varphi\left(\frac{P_\theta(j)}{P_n(j)}\right) P_n(j),$$

P_n being the empirical measure associated to the sample. The interest on these estimators is motivated by the fact that a suitable choice of the divergence may leads to an estimator more robust than the maximum likelihood one (see also Jiménez and Shao [14]). For continuous models, the estimators in (3) are not defined. Basu and Lindsay [2], among others, proposed smoothed versions of (3) in this case.

In the following, for notational clearness we write $\phi(\alpha, \theta)$ for $\phi(P_\alpha, P_\theta)$ for α and θ in Θ. We assume that for any $\theta \in \Theta$, P_θ has density p_θ with respect to some dominating σ-finite measure λ.

The divergence $\phi(\alpha, \theta_0)$ can be represented as resulting from an optimization procedure. This result has been obtained independently by Liese and Vajda [16] and Broniatowski and Keziou [5] who called it the dual form of a divergence, due to its connection with convex analysis. Assuming the strict convexity and the differentiability of the function φ, it holds

$$\varphi(t) \geq \varphi(s) + \varphi'(s)(t-s) \tag{4}$$

with equality only for $s = t$. Let α and θ_0 be fixed and put $t = p_\alpha(x)/p_{\theta_0}(x)$ and $s = p_\alpha(x)/p_\theta(x)$ in (4) and then integrate with respect to P_{θ_0}. This gives

$$\phi(\alpha, \theta_0) = \int \varphi\left(\frac{p_\alpha}{p_{\theta_0}}\right) dP_{\theta_0} = \sup_{\theta\in\Theta} \int m(\theta, \alpha) dP_{\theta_0} \tag{5}$$

with $m(\theta, \alpha) : x \mapsto m(\theta, \alpha, x)$ and

$$m(\theta, \alpha, x) := \int \varphi'\left(\frac{p_\alpha}{p_\theta}\right) dP_\alpha - \left\{\varphi'\left(\frac{p_\alpha}{p_\theta}(x)\right) \frac{p_\alpha}{p_\theta}(x) - \varphi\left(\frac{p_\alpha}{p_\theta}(x)\right)\right\}. \tag{6}$$

The supremum in (5) is unique and is attained in $\theta = \theta_0$, independently upon the value of α. Naturally, a class of estimators of θ_0, called "dual ϕ-divergence estimators" (DϕE's), is defined by

$$\widehat{\theta}_n(\alpha) := \arg\sup_{\theta \in \Theta} \int m(\theta, \alpha) dP_n, \quad \alpha \in \Theta. \tag{7}$$

Formula (7) defines a family of M-estimators indexed by some instrumental value of the parameter α and by the function φ defining the divergence. The choice of α appears as a major feature in the estimation procedure. Its value is strongly dependent upon some a priori knowledge on the value of the parameter to be estimated. In some examples in subsection 3.2, it even appears that a sharp a priori knowledge on the order of θ_0 leads to nearly efficient and robust estimates. This plays in favor of using the available information pertaining to the model and the data. Section 6 shortly presents some proposal for the adaptive choice of α.

For each $\alpha \in \Theta$, the divergence $\phi(P_\alpha, P_{\theta_0})$ between P_α and P_{θ_0} is estimated by

$$\widehat{\phi}_n(\alpha, \theta_0) := \int m(\widehat{\theta}_n(\alpha), \alpha) dP_n = \sup_{\theta \in \Theta} \int m(\theta, \alpha) dP_n. \tag{8}$$

Further, since

$$\inf_{\alpha \in \Theta} \phi(\alpha, \theta_0) = \phi(\theta_0, \theta_0) = 0,$$

and since the infimum in the above display is unique due to the strict convexity of φ, a natural definition of estimators of θ_0, called "minimum dual ϕ-divergence estimators" (MDϕE's), is provided by

$$\widehat{\alpha}_n := \arg\inf_{\alpha \in \Theta} \widehat{\phi}_n(\alpha, \theta_0) = \arg\inf_{\alpha \in \Theta} \sup_{\theta \in \Theta} \int m(\theta, \alpha) dP_n. \tag{9}$$

The DϕE's enjoy the same invariance property as the maximum likelihood estimator does. Invariance with respect to a reparametrization (one to one transformation of the parameter space) holds with direct substitution in (7). Also, consider a one to one differentiable transformation of the observations, say $Y = T(X)$ and the Jacobian $J(x) = \frac{d}{dx}T(x)$. Let $\widehat{\theta}_n(\alpha)$ defined in (7), based on the X_i's. Let $f_\theta(y)$ denote the density of the transformed variable Y and $\widehat{\theta}_n^*(\alpha)$ be the DϕE based on the Y_i's in the transformed model (with the same parameter θ). Specifically,

$$\widehat{\theta}_n^*(\alpha) = \arg\sup_{\theta \in \Theta} \left\{ \int \varphi' \left(\frac{f_\alpha}{f_\theta}(y) \right) f_\alpha(y) dy - \frac{1}{n} \sum_{i=1}^n \left(\varphi' \left(\frac{f_\alpha}{f_\theta}(Y_i) \right) \frac{f_\alpha}{f_\theta}(Y_i) - \varphi \left(\frac{f_\alpha}{f_\theta}(Y_i) \right) \right) \right\}.$$

Since

$$f_\theta(y) = p_\theta(T^{-1}(y)) |J(T^{-1}(y))|^{-1}$$

for all $\theta \in \Theta$, it follows that $\widehat{\theta}_n^*(\alpha) = \widehat{\theta}_n(\alpha)$, which is to say that the DϕE's are invariant estimators under any regular transformation of the observation space. The same invariance properties hold for MDϕE's.

Broniatowski and Keziou [6] have proved both the weak and the strong consistency, as well as the asymptotic normality for the estimators $\widehat{\theta}_n(\alpha)$ and $\widehat{\alpha}_n$. In the next sections, we study robustness properties for these classes of estimators and robustness of some tests based on dual ϕ-divergence estimators.

2.2 Some comments on robustness

The special form of divergence based estimators to be studied in this paper leads us to handle robustness characteristics through the influence function approach. An alternative and appealing robustness analysis in the minimum divergence methods is provided by the Residual Adjustment Function (RAF) (introduced in Lindsay [17]), which explains the incidence of non typical Pearson residuals, corresponding to over or sub-sampling, in the stability of the estimates. This method is quite natural for finitely supported models. In the case when the densities in the model are continuous, the Pearson residuals are estimated non parametrically which appears to cause quite a number of difficulties when adapted to minimum dual divergence estimation. This motivates the present choice in favor of the influence function approach.

Let α be fixed. For the Cressie-Read divergences, the equation whose solution is $\widehat{\theta}_n(\alpha)$ defined by (7) is

$$-\int \left(\frac{p_\alpha}{p_\theta}\right)^\gamma \dot{p}_\theta d\lambda + \frac{1}{n}\sum_{i=1}^n \left(\frac{p_\alpha(X_i)}{p_\theta(X_i)}\right)^\gamma \frac{\dot{p}_\theta(X_i)}{p_\theta(X_i)} = 0, \tag{10}$$

where $\dot{p}_\theta$ is the derivative with respect to θ of p_θ. Starting from the definition given by (7), this equation is obtained by equalizing with zero the derivative with respect to θ of $\int m(\theta, \alpha) dP_n$.

Let x be some outlier. The role of x in (10) is handled in the term

$$\left(\frac{p_\alpha(x)}{p_\theta(x)}\right)^\gamma \frac{\dot{p}_\theta(x)}{p_\theta(x)}. \tag{11}$$

The more stable this term, the more robust the estimate. In the classical case of the maximum likelihood estimator (which corresponds to $\widehat{\theta}_n(\alpha)$ with $\gamma = 0$ and independent on α), this term writes as $\frac{\dot{p}_\theta(x)}{p_\theta(x)}$ which is the likelihood score function associated to x. It is well known that, for most models, this term is usually unbounded when x belongs to $\mathbb{R}$, saying that the maximum likelihood estimator is not robust. In this respect, (11) appears as a weighted likelihood score function. In our approach, for several models, such as the normal scale, (11) is a bounded function of x, although $\frac{\dot{p}_\theta(x)}{p_\theta(x)}$ itself is not. Thus, in estimating equation (10), the score function is downweighted for large observations. The robustness of $\widehat{\theta}_n(\alpha)$ comes as a downweight effect of the quantity $\frac{\dot{p}_\theta(X_i)}{p_\theta(X_i)}$ through the multiplicative term $\left(\frac{p_\alpha(X_i)}{p_\theta(X_i)}\right)^\gamma$ which depends on the choice of the divergence. This choice is dictated by the form of $\frac{p_\alpha(x)}{p_\theta(x)}$ for large x and α fixed. For the models we'll consider as examples, for large x and α fixed, the quantity $\frac{p_\alpha(x)}{p_\theta(x)}$ can be large, close to zero, or close to one. Then we appropriately choose γ to be negative, respectively positive in order to obtain the downweight effect. In the next section we study in detail these robustness properties by the means of the influence function.

Some alternative choice has been proposed in literature. Basu et al. [1] proposed to alter the likelihood score factor by the multiplicative term $p_\theta^\beta(x)$, where $\beta > 0$. This induces an estimating procedure which is connected to the minimization of a density power divergence. Both their approach and the present one are adaptive in the sense that the downweight likelihood score factor is calibrated on the data.

Robustness as handled in the present paper is against the bias due to the presence of very few outliers in the data set. Bias due to misspecification of the model is not considered. It has been observed that DϕE's are biased under misspecification even in simple situations (for example when estimating the mean in a normal model with assumed variance 1, whereas the true variance is not 1); see Broniatowski and Vajda [8]; similar bias are unavoidable in parametric inference and can only be reduced through adaptive specific procedures, not studied here. For alternative robust M-estimation methods using divergences we refer to Toma [24].

3 Robustness of the estimators

3.1 Fisher consistency and influence functions

In order to measure the robustness of an estimator it is common to compute the influence function of the corresponding functional.

A map T which sends an arbitrary probability measure into the parameter space is a statistical functional corresponding to an estimator T_n of the parameter θ whenever $T(P_n) = T_n$.

This functional is called Fisher consistent for the parametric model $\{P_\theta : \theta \in \Theta\}$ if $T(P_\theta) = \theta$, for all $\theta \in \Theta$.

The influence function of the functional T in P measures the effect on T of adding a small mass at x and is defined as

$$\mathrm{IF}(x;T,P)=\lim_{\varepsilon\to 0}\frac{T(\widetilde{P}_{\varepsilon x})-T(P)}{\varepsilon} \tag{12}$$

where $\widetilde{P}_{\varepsilon x}=(1-\varepsilon)P+\varepsilon\delta_x$ and δ_x is the Dirac measure putting all its mass at x.

The gross error sensitivity measures approximately the maximum contribution to the estimation error that can be produced by a single outlier and is defined as

$$\sup_x \|\mathrm{IF}(x;T,P)\|.$$

Whenever the gross error sensitivity is finite, the estimator associated with the functional T is called B-robust.

Let $X_1,\ldots,X_n$ be an i.i.d. sample with p.m. P.

Let α be fixed and consider the dual ϕ-divergence estimators $\widehat{\theta}_n(\alpha)$ defined in (7). The functional associated to an estimator $\widehat{\theta}_n(\alpha)$ is

$$T_\alpha(P):=\arg\sup_{\theta\in\Theta}\int m(\theta,\alpha,y)dP(y). \tag{13}$$

The functional T_α is Fisher consistent. Indeed, the function $\theta\mapsto\int m(\theta,\alpha)dP_{\theta_0}$ has a unique maximizer $\theta=\theta_0$. Therefore $T_\alpha(P_\theta)=\theta$, for all $\theta\in\Theta$.

We denote $m'(\theta,\alpha)=\frac{\partial}{\partial\theta}m(\theta,\alpha)$ the d-dimensional column vector with entries $\frac{\partial}{\partial\theta_i}m(\theta,\alpha)$ and $m''(\theta,\alpha)$ the $d\times d$ matrix with entries $\frac{\partial^2}{\partial\theta_i\partial\theta_j}m(\theta,\alpha)$.

In the rest of the paper, for each α, we suppose that the function $\theta\mapsto m(\theta,\alpha)$ is twice continuously differentiable and that the matrix $\int m''(\theta_0,\alpha)dP_{\theta_0}$ exists and is invertible. We also suppose that, for each α, all the partial derivatives of order 1 and 2 of the function $\theta\mapsto m(\theta,\alpha)$ are respectively dominated on some neighborhoods of θ_0 by P_{θ_0}-integrable functions. This justifies the subsequent interchanges of derivation with respect to θ and integration.

Proposition 1 *The influence function of the functional T_α corresponding to an estimator $\widehat{\theta}_n(\alpha)$ is given by*

$$\mathrm{IF}(x;T_\alpha,P_{\theta_0})=\left[\int m''(\theta_0,\alpha)dP_{\theta_0}\right]^{-1}\left\{\int\varphi''\left(\frac{p_\alpha}{p_{\theta_0}}\right)\frac{p_\alpha}{p_{\theta_0}^2}\dot{p}_{\theta_0}dP_\alpha-\right.$$
$$\left.-\varphi''\left(\frac{p_\alpha}{p_{\theta_0}}(x)\right)\frac{p_\alpha^2(x)}{p_{\theta_0}^3(x)}\dot{p}_{\theta_0}(x)\right\}.$$

Particularizing $\alpha = \theta_0$ in Proposition 1 yields

$$\text{IF}(x; T_{\theta_0}, P_{\theta_0}) = -\left[\int m''(\theta_0, \theta_0) dP_{\theta_0}\right]^{-1} \varphi''(1) \frac{\dot{p}_{\theta_0}(x)}{p_{\theta_0}(x)}$$

and taking into account that

$$-\left[\int m''(\theta_0, \theta_0) dP_{\theta_0}\right]^{-1} = \frac{1}{\varphi''(1)} I_{\theta_0}^{-1}$$

it holds

$$\text{IF}(x; T_{\theta_0}, P_{\theta_0}) = I_{\theta_0}^{-1} \frac{\dot{p}_{\theta_0}(x)}{p_{\theta_0}(x)} \tag{14}$$

where I_{θ_0} is the information matrix $I_{\theta_0} = \int \frac{\dot{p}_{\theta_0} \dot{p}_{\theta_0}^t}{p_{\theta_0}} d\lambda$.

We now look at the corresponding estimators of the ϕ-divergence. For fixed α, the divergence $\phi(P_\alpha, P)$ between the probability measures P_α and P is estimated by (8). The statistical functional associated to $\widehat{\phi}_n(P_\alpha, P_{\theta_0})$ is

$$U_\alpha(P) := \int m(T_\alpha(P), \alpha, y) dP(y). \tag{15}$$

The functional U_α has the property that $U_\alpha(P_\theta) = \phi(\alpha, \theta)$, for any $\theta \in \Theta$. Indeed, using the fact that T_α is a Fisher consistent functional,

$$U_\alpha(P_\theta) = \int m\left(T_\alpha(P_\theta), \alpha, y\right) dP_\theta(y) = \int m(\theta, \alpha, y) dP_\theta(y) = \phi(\alpha, \theta)$$

for all $\theta \in \Theta$.

Proposition 2 *The influence function of the functional U_α corresponding to the estimator $\widehat{\phi}_n(P_\alpha, P)$ is given by*

$$\text{IF}(x; U_\alpha, P_{\theta_0}) = -\phi(\alpha, \theta_0) + m(\theta_0, \alpha, x). \tag{16}$$

For a minimum dual ϕ-divergence estimator $\widehat{\alpha}_n$ defined in (9), the corresponding functional is

$$V(P) := \arg \inf_{\alpha \in \Theta} U_\alpha(P) = \arg \inf_{\alpha \in \Theta} \int m(T_\alpha(P), \alpha, y) dP(y). \tag{17}$$

The statistical functional V is Fisher consistent. Indeed,

$$V(P_\theta) = \arg \inf_{\alpha \in \Theta} U_\alpha(P_\theta) = \arg \inf_{\alpha \in \Theta} \phi(\alpha, \theta) = \theta$$

for all $\theta \in \Theta$.

In the following proposition, we suppose that the function $m(\theta, \alpha)$ admits partial derivatives of order 1 and 2 with respect to θ and α and also we suppose that conditions permitting to derivate $m(\theta, \alpha)$ under the integral sign hold. The following result states that, unlike $\widehat{\theta}_n(\alpha)$, an estimator $\widehat{\alpha}_n$ is generally not robust. Indeed, it has the same robustness properties as the maximum likelihood estimator, since it has its influence function which in most cases is unbounded. Whatever the divergence, the estimators $\widehat{\alpha}_n$ have the same influence function.

Proposition 3 *The influence function of the functional V corresponding to an estimator $\widehat{\alpha}_n$ is given by*

$$\text{IF}(x; V, P_{\theta_0}) = I_{\theta_0}^{-1} \frac{\dot{p}_{\theta_0}(x)}{p_{\theta_0}(x)}. \tag{18}$$

3.2 Robustness of the estimators for scale models and location models

In this subsection, examining the expressions of the influence functions, we give conditions for attaining the B-robustness of the dual ϕ-divergence estimators $\widehat{\theta}_n(\alpha)$, as well as of the corresponding divergence estimators. The case of interest in our B-robustness study is $\alpha \neq \theta_0$ since, as observed above, the choice $\alpha = \theta_0$ generally leads to unbounded influence functions. For the Cressie-Read family of divergences (2) it holds

$$\text{IF}(x; T_\alpha, P_{\theta_0}) = [\int m''(\theta_0, \alpha) dP_{\theta_0}]^{-1} \left\{ \int \left(\frac{p_\alpha}{p_{\theta_0}} \right)^\gamma \dot{p}_{\theta_0} d\lambda - \left(\frac{p_\alpha(x)}{p_{\theta_0}(x)} \right)^\gamma \frac{\dot{p}_{\theta_0}(x)}{p_{\theta_0}(x)} \right\} \tag{19}$$

and

$$\begin{aligned} \text{IF}(x; U_\alpha, P_{\theta_0}) &= -\phi(\alpha, \theta_0) + m(\theta_0, \alpha, x) \\ &= -\phi(\alpha, \theta_0) + \frac{1}{\gamma - 1} \left\{ \int \left(\frac{p_\alpha}{p_{\theta_0}} \right)^{\gamma - 1} dP_\alpha - 1 \right\} - \frac{1}{\gamma} \left\{ \left(\frac{p_\alpha(x)}{p_{\theta_0}(x)} \right)^\gamma - 1 \right\}. \end{aligned}$$

3.2.1 Scale models

For a given density p, it holds $p_\theta(x) = \frac{1}{\theta} p(\frac{x}{\theta})$ and $\dot{p}_\theta(x) = -\frac{1}{\theta^2} \left[p\left(\frac{x}{\theta}\right) + \frac{x}{\theta} \dot{p}\left(\frac{x}{\theta}\right) \right]$. Consider the following conditions:

(A.1) $\int |u \dot{p}(u)| du < \infty$.

(A.2) $\sup_x \frac{p(\alpha^{-1} x)}{p(\theta_0^{-1} x)} < \infty$.

(A.3) $\sup_x \frac{p(\theta_0^{-1}x)}{p(\alpha^{-1}x)} < \infty$.

(A.4) $\sup_x \left| \frac{\partial}{\partial\theta}[\log p(\theta_0^{-1}x)] \left(\frac{p(\alpha^{-1}x)}{p(\theta_0^{-1}x)}\right)^\gamma \right| < \infty$.

Proposition 4 *For scale models, if the conditions* (A.2) (*for the case* $\gamma > 0$) *or* (A.3) (*for the case* $\gamma < 0$) *together with* (A.1) *and* (A.4) *are satisfied, then* $\widehat{\theta}_n(\alpha)$ *is B-robust.*

As a particular case, consider the problem of robust estimation of the parameter $\theta_0 = \sigma$ of the univariate normal model, when the mean m is known, intending to use an estimator $\widehat{\theta}_n(\overline{\sigma})$ with $\overline{\sigma} \neq \sigma$. We are interested on those divergences from the Cressie-Read family and those possible values of $\overline{\sigma}$ for which $\widehat{\theta}_n(\overline{\sigma})$ is B-robust. We have

$$\text{IF}(x; T_{\overline{\sigma}}, P_\sigma) = [\int m''(\sigma, \overline{\sigma}) dP_\sigma]^{-1} \left\{ \int \left(\frac{p_{\overline{\sigma}}}{p_\sigma}\right)^\gamma \frac{\dot{p}_\sigma}{p_\sigma} dP_\sigma - \left(\frac{p_{\overline{\sigma}}(x)}{p_\sigma(x)}\right)^\gamma \frac{\dot{p}_\sigma(x)}{p_\sigma(x)} \right\}.$$

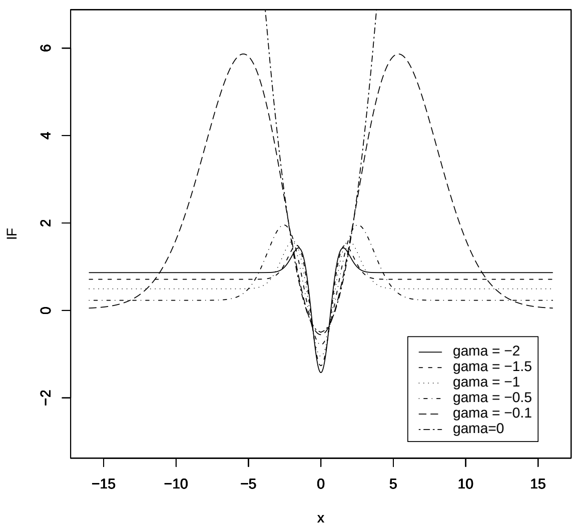

Fig. 1. Influence functions $\text{IF}(x; T_{\overline{\sigma}}, P_\sigma)$ for normal scale model, when $m = 0$, the true scale parameter is $\sigma = 1$ and $\overline{\sigma} = 1.9$.

It is easily seen that $\text{IF}(x; T_{\overline{\sigma}}, P_\sigma)$ is bounded whenever the function $\left(\frac{p_{\overline{\sigma}}(x)}{p_\sigma(x)}\right)^\gamma \frac{\dot{p}_\sigma(x)}{p_\sigma(x)}$ is bounded. Since

$$\left(\frac{p_{\overline{\sigma}}(x)}{p_\sigma(x)}\right)^\gamma \frac{\dot{p}_\sigma(x)}{p_\sigma(x)} = \frac{\sigma^{\gamma-1}}{\overline{\sigma}^\gamma} \left\{ \left(\frac{x-m}{\sigma}\right)^2 - 1 \right\} \left(\exp\left(-\frac{1}{2} \left\{ \left(\frac{x-m}{\overline{\sigma}}\right)^2 - \left(\frac{x-m}{\sigma}\right)^2 \right\} \right) \right)^\gamma \quad (20)$$

boundedness of $\text{IF}(x; T_{\overline{\sigma}}, P_\sigma)$ holds when $\gamma > 0$ and $\overline{\sigma} < \sigma$ or when $\gamma < 0$ and $\overline{\sigma} > \sigma$, cases in which the conditions of Proposition 4 are satisfied. A simple calculation shows that these choices of γ and $\overline{\sigma}$ assure that $\int m''(\sigma, \overline{\sigma})dP_\sigma$ is finite and non zero. However, when using the modified Kullback-Leibler divergence (γ=0), none of the estimators $\widehat{\theta}_n(\overline{\sigma})$ is B-robust, the function (20) being unbounded. These aspects can also be observed in Figure 1, which presents influence functions for different divergences when $\sigma = 1$ and $\overline{\sigma} = 1.9$. The negative values of the influence function in a neighborhood of 0 is explained by the decrease of the variance estimate when oversampling close to the mean.

The asymptotic relative efficiency of an estimator is the ratio of the asymptotic variance of the maximum likelihood estimator to that of the estimator in question. For the scale normal model, the choice of $\overline{\sigma}$ close to σ assures a good efficiency of $\widehat{\theta}_n(\overline{\sigma})$ and also the B-robustness property. Then, the bigger is the value of $|\gamma|$, the smaller is the gross error sensitivity of the estimator. For example, for $\sigma = 1$ and $\overline{\sigma} = 0.99$, the efficiency of $\widehat{\theta}_n(\overline{\sigma})$ is 0.9803 when $\gamma = 0.5$, 0.9615 when $\gamma = 1$, 0.9266 when $\gamma = 2$ and 0.8947 when $\gamma = 3$, the most B-robust estimator corresponding to $\gamma = 3$. As can be inferred from Figure 1, the curves $\text{IF}^2(x; T_{\overline{\sigma}}, P_\sigma)$ are ordered decreasingly with respect to $|\gamma|$. Therefore, large values of $|\gamma|$ lead to small gross error sensitivities and low efficiencies, since the asymptotic variance of $\widehat{\theta}_n(\overline{\sigma})$ is $[\int \text{IF}^2(x; T_{\overline{\sigma}}, P_\sigma)dP_\sigma]^{-1}$ (see also Hampel et al. [11] for this formula).

For scale models, conditions of Proposition 4 assure that $\widehat{\theta}_n(\alpha)$ and the corresponding divergence estimator $\widehat{\phi}_n(\alpha, \theta_0)$ are B-robust.

3.2.2 Location models

It holds $p_\theta(x) = p(x - \theta)$.

Proposition 5 *For location models, if the condition*

$$\sup_x \left| \left(\frac{p(x-\alpha)}{p(x-\theta_0)} \right)^\gamma \frac{\partial}{\partial \theta} \log p(x - \theta_0) \right| < \infty \tag{21}$$

is satisfied, then $\widehat{\theta}_n(\alpha)$ is B-robust.

For the Cauchy density the maximum likelihood estimator exists, it is consistent, efficient and B-robust and all the estimators $\widehat{\theta}_n(\alpha)$ exist and are B-robust. Indeed, condition (21) writes

$$\sup_x 2 \left| \left(\frac{1 + (x-\theta_0)^2}{1 + (x-\alpha)^2} \right)^\gamma \frac{x - \theta_0}{1 + (x-\theta_0)^2} \right| < \infty$$

and is fulfilled for any γ and any α. Also, the integral $\int m''(\theta_0, \alpha)dP_{\theta_0}$ exists and is different to zero for any γ and any α. This is quite natural since sampling of the Cauchy law makes equivalent outliers and large sample points due

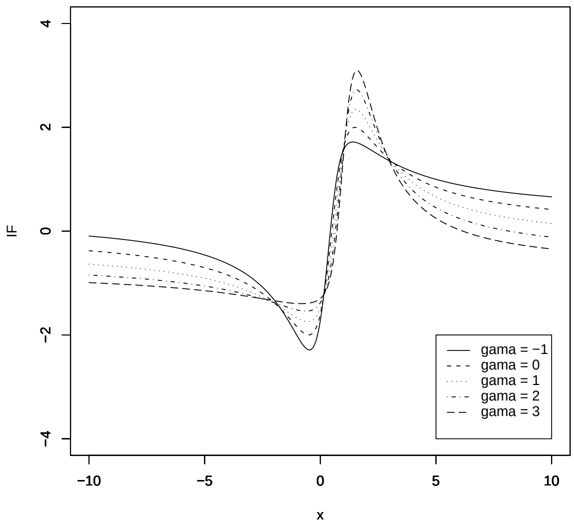

Fig. 2. Influence functions $\mathrm{IF}(x;T_\alpha,P_{\theta_0})$ for the Cauchy location model, when the true location parameter is $\theta_0 = 0.5$ and $\alpha = 0.8$.

to heavy tails. However it is known that the likelihood equation for Cauchy distribution has multiple roots. The number of solutions behaves asymptotically as two times a Poisson$(1/\pi)$ variable plus 1 (see van der Vaart [25] p. 74). The possible selection rule for the estimate is to check the nearly common estimates for different α and ϕ-divergences. Figure 2 presents influence functions $\mathrm{IF}(x;T_\alpha,P_{\theta_0})$, when $\gamma \in \{-1,0,1,2,3,\}$, $\theta_0 = 0.5$ and $\alpha = 0.8$. For these choices of θ_0 and α, the efficiency of $\widehat{\theta}_n(\alpha)$ is 0.9775 when $\gamma = 1$, 0.9208 when $\gamma = 2$, 0.8508 when $\gamma = 3$. Here, when γ increases, the decrease of the efficiency is worsened by a loss in B-robustness. In this respect, the maximum likelihood estimator appears as a good choice in terms of robustness and efficiency.

In the case of the logistic location model, a simple calculation shows that the condition (21) is fulfilled for any γ and any α. Also, the integral $\int m''(\theta_0,\alpha)dP_{\theta_0}$ exists and is different from zero for any γ and any α. These conditions entail the fact that all the estimators $\widehat{\theta}_n(\alpha)$ are B-robust. Figure 4 presents influence functions $\mathrm{IF}(x;T_\alpha,P_{\theta_0})$, when $\gamma \in \{-1,0,0.5,1,2,3,\}$, $\theta_0 = 1$ and $\alpha = 1.5$. As in the case of the Cauchy model, when γ increases, the decrease of the efficiency is worsened by the increase of the gross error sensitivity, such that the maximum likelihood estimator appears again as a good choice in terms of robustness and efficiency.

On the other hand, for the mean of the normal law, none of the estimators $\widehat{\theta}_n(\alpha)$ is B-robust, their influence functions being always unbounded.

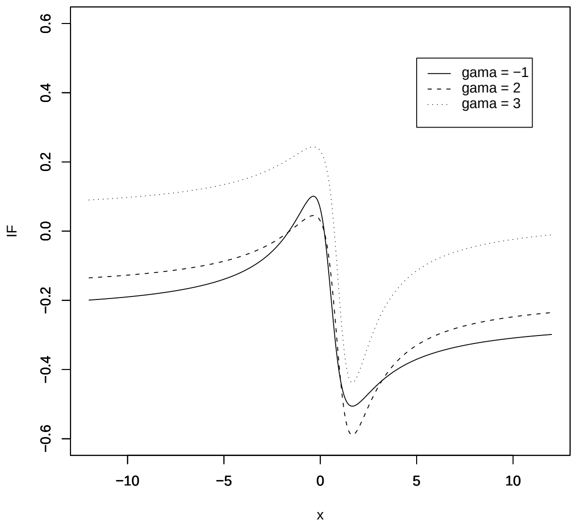

Fig. 3. Influence functions $\mathrm{IF}(x; U_\alpha, P_{\theta_0})$ for the Cauchy location model, when the true location parameter is $\theta_0 = 0.5$ and $\alpha = 0.8$.

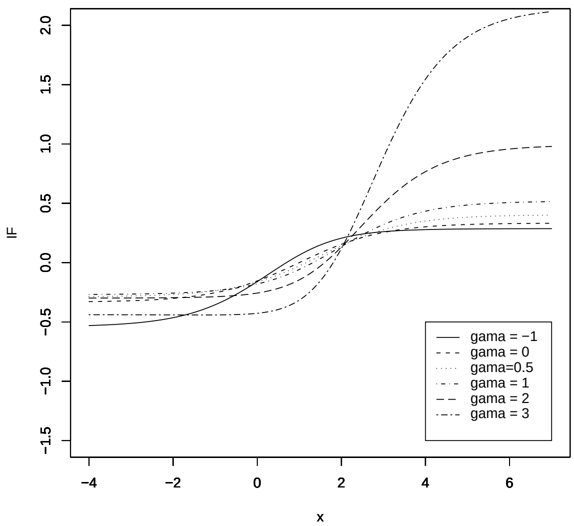

Fig. 4. Influence functions $\mathrm{IF}(x; T_\alpha, P_{\theta_0})$ for the logistic location model, when the true location parameter is $\theta_0 = 1$ and $\alpha = 1.5$.

In the case of the Cauchy model, as well as in the case of the logistic model, $\mathrm{IF}(x; U_\alpha, P_{\theta_0})$ is bounded for any γ and any α. In Figure 3, respectively in Figure 5, we present such influence functions for different choices of γ. Thus, for these two location models, all the estimators $\widehat{\phi}_n(\alpha, \theta_0)$ are B-robust.

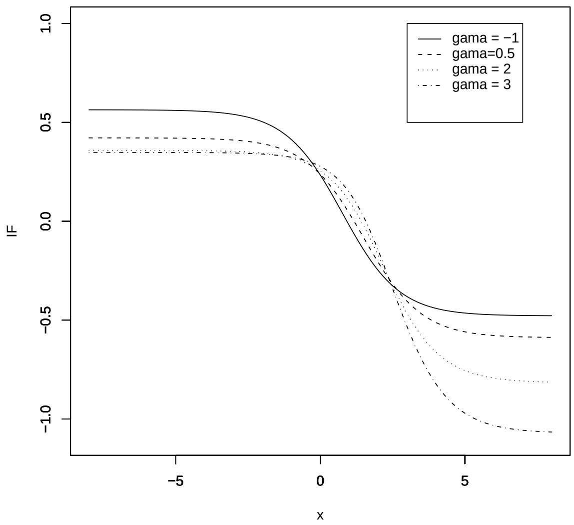

Fig. 5. Influence functions $\mathrm{IF}(x; U_\alpha, P_{\theta_0})$ for the logistic location model, when the true location parameter is $\theta_0 = 1$ and $\alpha = 1.5$.

4 Robust tests based on divergence estimators

4.1 Asymptotic results for contaminated observations

This subsection presents some asymptotic results that are necessary in order to analyze the robustness of some tests based on divergence estimators. These asymptotic results are obtained for contaminated observations, namely $X_1, \ldots, X_n$ are i.i.d. with

$$P^P_{n,\varepsilon,x} := \left(1 - \frac{\varepsilon}{\sqrt{n}}\right) P_{\theta_n} + \frac{\varepsilon}{\sqrt{n}} \delta_x \tag{22}$$

where $\theta_n = \theta_0 + \frac{\Delta}{\sqrt{n}}$, Δ being an arbitrary vector from $\mathbb{R}^d$.

For α fixed consider the following conditions:

(C.1) The function $\theta \mapsto m(\theta, \alpha)$ is C^3 for all x and all partial derivatives of order 3 of $\theta \mapsto m(\theta, \alpha)$ are dominated by some P_{θ_n}-integrable function $x \mapsto H(x)$ with the property $\int H^2 dP_{\theta_n}$ is finite, for any n and any Δ.

(C.2) $\int m(\theta_0, \alpha) dP_{\theta_n}$ and $\int m^2(\theta_0, \alpha) dP_{\theta_n}$ are finite, for any n and any Δ.

(C.3) $\int m'(\theta_0, \alpha) dP_{\theta_n}$ and $\int m'(\theta_0, \alpha) m'(\theta_0, \alpha)^t dP_{\theta_n}$ exist, for any n and any Δ.

(C.4) $\int m''(\theta_0, \alpha) dP_{\theta_n}$ and $\int m''(\theta_0, \alpha)^2 dP_{\theta_n}$ exist, for any n and any Δ.

The estimators $\widehat{\theta}_n(\alpha)$ have good properties with respect to contamination in terms of consistency.

Proposition 6 *If the conditions* (C.1), (C.3) *and* (C.4) *are satisfied, then*

$$\sqrt{n}(\widehat{\theta}_n(\alpha) - T_\alpha(P^P_{n,\varepsilon,x})) = O_P(1).$$

Also, $\widehat{\phi}_n(\alpha, \theta_0)$ enjoys normal convergence under (22).

Proposition 7 *If* $\alpha \neq \theta_0$ *and the conditions* (C.1) $-$ (C.4) *are satisfied, then*

$$\frac{\sqrt{n}(\widehat{\phi}_n(\alpha, \theta_0) - U_\alpha(P^P_{n,\varepsilon,x}))}{[\int \mathrm{IF}^2(y; U_\alpha, P^P_{n,\varepsilon,x}) dP^P_{n,\varepsilon,x}(y)]^{1/2}}$$

converges in distribution to a normal standard variable.

4.2 Robust tests based on divergence estimators

In this subsection we propose tests based on dual ϕ-divergence estimators and study their robustness properties. We mention that the use of the dual form of a divergence to derive robust tests was discussed in a different context by Broniatowski and Leorato [7] in the case of the Neyman χ^2 divergence.

For testing the hypothesis $\theta = \theta_0$ against the alternative $\theta \neq \theta_0$, consider the test of level α_0 defined by the test statistic $\widehat{\phi}_n := \widehat{\phi}_n(\alpha, \theta_0)$ with $\alpha \neq \theta_0$ and by the critical region

$$C := \left\{ \left| \frac{\sqrt{n}(\widehat{\phi}_n - \phi(\alpha, \theta_0))}{[\int \mathrm{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2}} \right| \geq q_{1-\frac{\alpha_0}{2}} \right\}$$

where $q_{1-\frac{\alpha_0}{2}}$ is the $(1 - \frac{\alpha_0}{2})$-quantile of the standard normal distribution.

Due to the asymptotic normality of $\widehat{\phi}_n$, for n large, the level writes as

$$\alpha_0 \simeq P_{\theta_0} \left(\left| \frac{\sqrt{n}(\widehat{\phi}_n - \phi(\alpha, \theta_0))}{[\int \mathrm{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2}} \right| \geq q_{1-\frac{\alpha_0}{2}} \right) \quad (23)$$

$$= P_{\theta_0}(|\widehat{\phi}_n - \phi(\alpha, \theta_0)| \geq (\sqrt{n})^{-1} [\int \mathrm{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2} q_{1-\frac{\alpha_0}{2}}) \quad (24)$$

$$= 2P_{\theta_0}(\widehat{\phi}_n \geq k_n(\alpha_0)) \quad (25)$$

where $k_n(\alpha_0) = (\sqrt{n})^{-1} [\int \mathrm{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2} q_{1-\frac{\alpha_0}{2}} + \phi(\alpha, \theta_0)$.

We work with the form (25) of the level and consequently of the probability to reject the null hypothesis, this being easier to handle in the proofs of the results that follows.

Consider the sequence of contiguous alternatives $\theta_n = \theta_0 + \Delta n^{-1/2}$, where Δ is any vector from $\mathbb{R}^d$. When θ_n tends to θ_0, the contamination must converge to 0 at the same rate, to avoid the overlapping between the neighborhood of the hypothesis and that of the alternative (see Hampel et al. [11], p.198 and Heritier and Ronchetti [12]). Therefore we consider the contaminated distributions

$$P^L_{n,\varepsilon,x} = \left(1 - \frac{\varepsilon}{\sqrt{n}}\right) P_{\theta_0} + \frac{\varepsilon}{\sqrt{n}}\delta_x \tag{26}$$

for the level and

$$P^P_{n,\varepsilon,x} = \left(1 - \frac{\varepsilon}{\sqrt{n}}\right) P_{\theta_n} + \frac{\varepsilon}{\sqrt{n}}\delta_x \tag{27}$$

for the power.

The asymptotic level (the asymptotic power) under (26) (under (27)) will be evaluated now.

Let $\beta_0 = \lim_{n\to\infty} 2P_{\theta_n}(\widehat{\phi}_n \geq k_n(\alpha_0))$ be the asymptotic power of the test under the family of alternatives P_{θ_n}. The test is robust with respect to the power if the limit of the powers under the contaminated alternatives stays in a bounded neighborhood of β_0, so that the role of the contamination is somehow controlled. Also, the test is robust with respect to the level if the limit of the level under the contaminated null distributions stays in a bounded neighborhood of α_0.

Let $P_{n,\varepsilon,x} = 2P^P_{n,\varepsilon,x}(\widehat{\phi}_n \geq k_n(\alpha_0))$. In the same vein as in Dell'Aquilla and Ronchetti [10] it holds:

Proposition 8 *If the conditions* (C.1) – (C.4) *are fulfilled, then the asymptotic power of the test under $P^P_{n,\varepsilon,x}$ is given by*

$$\lim_{n\to\infty} P_{n,\varepsilon,x} = 2 - 2\Phi\left(\Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right) - \Delta\frac{c}{[\int \mathrm{IF}^2(y;U_\alpha,P_{\theta_0})dP_{\theta_0}(y)]^{1/2}} - \varepsilon\frac{\mathrm{IF}(x;U_\alpha,P_{\theta_0})}{[\int \mathrm{IF}^2(y;U_\alpha,P_{\theta_0})dP_{\theta_0}(y)]^{1/2}}\right) \tag{28}$$

where $c = \int m(\theta_0,\alpha,y)\frac{\dot{p}_{\theta_0}(y)}{p_{\theta_0}(y)}dP_{\theta_0}(y)$ and Φ is the cumulative distribution function of the standard normal.

A Taylor expansion with respect to ε yields

$$\lim_{n\to\infty} P_{n,\varepsilon,x} = 2 - 2\Phi\left(\Phi^{-1}\left(1-\frac{\alpha_0}{2}\right) - \Delta\frac{c}{[\int \mathrm{IF}^2(y;U_\alpha,P_{\theta_0})dP_{\theta_0}(y)]^{1/2}}\right) +$$
$$+2\varepsilon f\left(\Phi^{-1}\left(1-\frac{\alpha_0}{2}\right)-\Delta\frac{c}{[\int \mathrm{IF}^2(y;U_\alpha,P_{\theta_0})dP_{\theta_0}(y)]^{1/2}}\right)\frac{\mathrm{IF}(x;U_\alpha,P_{\theta_0})}{[\int \mathrm{IF}^2(y;U_\alpha,P_{\theta_0})dP_{\theta_0}(y)]^{1/2}}+o(\varepsilon)$$
$$= \beta_0+2\varepsilon f\left(\Phi^{-1}\left(1-\frac{\alpha_0}{2}\right)-\Delta\frac{c}{[\int \mathrm{IF}^2(y;U_\alpha,P_{\theta_0})dP_{\theta_0}(y)]^{1/2}}\right)\frac{\mathrm{IF}(x;U_\alpha,P_{\theta_0})}{[\int \mathrm{IF}^2(y;U_\alpha,P_{\theta_0})dP_{\theta_0}(y)]^{1/2}}+$$
$$+o(\varepsilon)$$

where β_0 is the asymptotic power for the non contaminated model and f is the density of the standard normal distribution.

In order to limit the bias in the power of the test it is sufficient to bound the influence function $\mathrm{IF}(x;U_\alpha,P_{\theta_0})$. Bounding the influence function is therefore enough to maintain the power in a pre-specified band around β_0.

Let $L_{n,\varepsilon,x} = 2P^L_{n,\varepsilon,x}(\widehat{\phi}_n \geq k_n(\alpha_0))$. Putting $\Delta = 0$ in (28) yields:

Proposition 9 *If the conditions* (C.1) – (C.4) *are fulfilled, then the asymptotic level of the test under* $P^L_{n,\varepsilon,x}$ *is given by*

$$\lim_{n\to\infty} L_{n,\varepsilon,x} = 2 - 2\Phi\left(\Phi^{-1}\left(1-\frac{\alpha_0}{2}\right) - \varepsilon\frac{\mathrm{IF}(x;U_\alpha,P_{\theta_0})}{[\int \mathrm{IF}^2(y;U_\alpha,P_{\theta_0})dP_{\theta_0}(y)]^{1/2}}\right)$$
$$= \alpha_0 + \varepsilon f\left(\Phi^{-1}\left(1-\frac{\alpha_0}{2}\right)\right)\frac{\mathrm{IF}(x;U_\alpha,P_{\theta_0})}{[\int \mathrm{IF}^2(y;U_\alpha,P_{\theta_0})dP_{\theta_0}(y)]^{1/2}} + o(\varepsilon).$$

Hence, when $\mathrm{IF}(x;U_\alpha,P_{\theta_0})$ is bounded, $L_{n,\varepsilon,x}$ remains between pre-specified bounds of α_0.

As the Proposition 8 and Proposition 9 show, both the asymptotic power of the test under $P^P_{n,\varepsilon,x}$ and the asymptotic level of the test under $P^L_{n,\varepsilon,x}$ are controlled by the influence function of the test statistic. Hence, the robustness of the test statistic $\widehat{\phi}_n$, as discussed in the previous section, assures the stability of the test under small arbitrary departures from the null hypothesis, as well as a good power under small arbitrary departures from specified alternatives. Figures 3 and 5 provide some specific values of γ and α inducing robust tests for θ_0 corresponding to those models.

5 Simulation results

Simulation were run in order to examine empirically the performances of the robust dual ϕ-divergence estimators and tests. The considered parametric model was the scale normal model with known mean. We worked with data generated from the model, as well as with contaminated data.

To make some comparisons, beside dual ϕ-divergence estimators, we considered minimum density power divergence estimators of Basu et al. [1] (MDPDE's) and the maximum likelihood estimator (MLE). Recall that a MDPDE of a parameter θ is obtained as solution of the equation

$$\int \dot{p}_\theta(z) p_\theta^\beta(z) dz - \frac{1}{n} \sum_{i=1}^{n} \dot{p}_\theta(X_i) p_\theta^{\beta-1}(X_i) = 0 \tag{29}$$

with respect to θ, where $\beta > 0$ and $X_1, \ldots, X_n$ is a sample from P_θ. In the case of the scale normal model $\mathcal{N}(m, \sigma)$, equation (29) writes as

$$\begin{aligned} &\int \frac{1}{\sigma^{\beta+2}(\sqrt{2\pi})^{\beta+1}} \left(e^{-\frac{1}{2}\left(\frac{z-m}{\sigma}\right)^2}\right)^{\beta+1} \left[-1 + \left(\frac{z-m}{\sigma}\right)^2\right] dz - \\ &-\frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sigma^{\beta+1}(\sqrt{2\pi})^{\beta}} \left(e^{-\frac{1}{2}\left(\frac{X_i-m}{\sigma}\right)^2}\right)^{\beta} \left[-1 + \left(\frac{X_i-m}{\sigma}\right)^2\right] = 0 \end{aligned}$$

and the MDPDE of the parameter σ is robust for any $\beta > 0$.

In a first Monte Carlo experiment the data were generated from the scale normal model $\mathcal{N}(0, 1)$ with mean $m = 0$ known, $\sigma = 1$ being the parameter of interest. We considered different choices for the tuning parameter α and for the Cressie-Read divergence to compute DϕE's, and different choices for the tuning parameter β in order to compute MDPDE's. For each set of configurations considered, 5000 samples of size $n = 100$ were generated from the model, and for each sample DϕE's, MDPDE's and MLE were obtained.

In Table 1 we present the results of the simulations, showing simulation based estimates of the bias and MSE given by

$$\widehat{\text{Bias}} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\widehat{\sigma}_i - \sigma), \quad \widehat{\text{MSE}} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\widehat{\sigma}_i - \sigma)^2,$$

where n_s denotes the number of samples (5000 in our case) and $\widehat{\sigma}_i$ denotes an estimate of σ for the ith sample. Examination of the table shows that DϕE's give as good results as MDPDE's or MLE.

In a second Monte Carlo experiment, we first generated samples with 100 observations, namely 98 coming from $\mathcal{N}(0, 1)$ and 2 outliers $x = 10$ and then we generated samples with 100 observations, namely 96 from $\mathcal{N}(0, 1)$ and 4 outliers $x = 10$. The tuning parameters were the same as in the non contaminated case and also $n_s = 5000$. The simulation results are given in Table 2. As can be seen, the results for DϕE's and MDPDE's are comparable, they being better than the results for MLE in both cases.

A close look at the results of the simulations show the DϕE performs well under the model, when no outliers are generated; indeed the best results are obtained when $\gamma = -0.1$, whatever $\overline{\sigma} = 1.5$ or $\overline{\sigma} = 1.9$. The performance of the estimator under the model is comparable to that of some MDPDE's in terms of empirical MSE ($\widehat{\text{MSE}}$): indeed the $\widehat{\text{MSE}}$ for DϕE with $\gamma = -0.1$ parallels MDPDE's for small β. It is also slightly shorter than the one obtained through the MLE. Under contamination, the DϕE with $\gamma = -0.5$ yields clearly the most robust estimate and the empirical MSE is very small, indicating a strong stability of the estimate. It compares favorably with MDPE for all β, whatever $\overline{\sigma} = 1.5$ or $\overline{\sigma} = 1.9$. The simulation with 4 outliers at $x = 10$ provide a clear evidence of the properties of the DϕE with $\gamma = -0.5$. Also small values of β give similar results as large negative values of γ, whatever $\overline{\sigma}$, under contamination. Although $\gamma = -0.1$ is a good alternative to MLE under the model, $\gamma = -0.5$ behaves quite well in terms of bias while keeping short empirical MSE under the model or under contamination. These results are in full accordance with Figure 1; indeed the influence function is constant close to 0 for large values of x.

Thus, the DϕE is shown to be an attractive alternative to both the MLE and MDPDE in these settings.

In order to test the hypothesis $\sigma = 1$ with respect to the alternative $\sigma \neq 1$, we considered the test statistic

$$\frac{\sqrt{n}(\widehat{\phi}_n - \phi(\alpha, \theta_0))}{[\int \text{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2}}$$

(here $\theta_0 = \sigma = 1$). Under the null hypothesis, this test statistic is asymptotically $\mathcal{N}(0, 1)$. We worked with data generated from the model $\mathcal{N}(0, 1)$, as well as with contaminated data. In each case, we simulated 5000 samples and we computed the actual levels

$$P\left(\left|\frac{\sqrt{n}(\widehat{\phi}_n - \phi(\alpha, \theta_0))}{[\int \text{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2}}\right| \geq q_{1-\frac{\alpha_0}{2}}\right)$$

corresponding to the nominal levels $\alpha_0 = 0.01, 0.02, \ldots, 0.1$. We reported the corresponding relative errors

$$\left(P\left(\left|\frac{\sqrt{n}(\widehat{\phi}_n - \phi(\alpha, \theta_0))}{[\int \text{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2}}\right| \geq q_{1-\frac{\alpha_0}{2}}\right) - \alpha_0\right) / \alpha_0.$$

In Figure 6 we present relative errors for the robust tests applied to the scale normal model $\mathcal{N}(0, 1)$, when the data are generated from the model. The sample size is $n = 100$, the tuning parameter is $\overline{\sigma} = 1.9$ and the Cressie-Read divergences correspond to $\gamma \in \{-1.5, -1, -0.5, -0.1\}$. The approximation of the level is good for all the considered divergences.

In Figure 7 are represented relative errors of the robust tests applied to the scale normal model $\mathcal{N}(0,1)$, for samples with $n = 100$ data, namely 98 data generated from $\mathcal{N}(0,1)$ and 2 outliers $x = 10$. We considered $\overline{\sigma} = 1.9$ and $\gamma \in \{-2, -1.5\}$. Again, the approximation of the level of the test is good for all the considered divergences.

In Figure 8 we present relative errors of the robust tests applied to the scale normal model $\mathcal{N}(0,1)$, for samples with $n = 100$ data, namely 96 data generated from $\mathcal{N}(0,1)$ and 4 outliers $x = 10$. We considered $\overline{\sigma} = 1.9$ and $\gamma \in \{-2, -1.5\}$.

Observe that the tests give good results for values of γ close to zero when the data are not contaminated, respectively for large negative values of γ when the data are contaminated.

Thus, the numerical results show that dual ϕ-divergence estimates and corresponding tests are stable in the presence of some outliers in the sample.

Table 1.

Simulation results for DϕE, MDPDE and MLE of the parameter $\sigma = 1$ when the data are generated from the model $\mathcal{N}(0,1)$.

	$\widehat{\sigma}$	$\widehat{\text{Bias}}$	$\widehat{\text{MSE}}$
DϕE			
$\overline{\sigma}$=1.5 $\gamma = -2$	0.99770	-0.00229	0.00917
$\overline{\sigma}$=1.5 $\gamma = -1.5$	0.99735	-0.00264	0.00822
$\overline{\sigma}$=1.5 $\gamma = -1$	0.99760	-0.00239	0.00698
$\overline{\sigma}$=1.5 $\gamma = -0.5$	0.99833	-0.00166	0.00563
$\overline{\sigma}$=1.5 $\gamma = -0.1$	0.99799	-0.00200	0.00492
$\overline{\sigma}$=1.9 $\gamma = -2$	0.99892	-0.00107	0.01029
$\overline{\sigma}$=1.9 $\gamma = -1.5$	0.99841	-0.00158	0.00924
$\overline{\sigma}$=1.9 $\gamma = -1$	0.99824	-0.00175	0.00773
$\overline{\sigma}$=1.9 $\gamma = -0.5$	0.99839	-0.00160	0.00588
$\overline{\sigma}$=1.9 $\gamma = -0.1$	0.99768	-0.00231	0.00473
MDPDE			
$\beta = 0.1$	0.99894	-0.00105	0.00514
$\beta = 0.5$	0.99986	-0.00013	0.00686
$\beta = 1$	1.00005	0.00005	0.00927
$\beta = 1.5$	1.00074	0.00074	0.01077
$\beta = 2$	1.00150	0.00150	0.01165
$\beta = 2.5$	1.00294	0.00294	0.01266
MLE	0.99743	-0.00256	0.00501

Table 2.

Simulation results for DϕE, MDPDE and MLE of the parameter $\sigma = 1$ when 98 data are generated from the model $\mathcal{N}(0,1)$ and 2 outliers $x = 10$ are added, respectively when 96 data are generated from the model $\mathcal{N}(0,1)$ and 4 outliers $x = 10$ are added.

	2 outliers			4 outliers		
	$\widehat{\sigma}$	$\widehat{\text{Bias}}$	$\widehat{\text{MSE}}$	$\widehat{\sigma}$	$\widehat{\text{Bias}}$	$\widehat{\text{MSE}}$
DϕE						
$\overline{\sigma}$=1.5 $\gamma = -2$	1.01186	0.01186	0.00914	1.02540	0.02540	0.00946
$\overline{\sigma}$=1.5 $\gamma = -1.5$	1.00850	0.00850	0.00816	1.01911	0.01911	0.00833
$\overline{\sigma}$=1.5 $\gamma = -1$	1.00499	0.00499	0.00697	1.01210	0.01210	0.00707
$\overline{\sigma}$=1.5 $\gamma = -0.5$	1.00171	0.00171	0.00572	1.00526	0.00526	0.00583
$\overline{\sigma}$=1.5 $\gamma = -0.1$	1.09661	0.09661	0.01641	0.99766	-0.00233	0.00088
$\overline{\sigma}$=1.9 $\gamma = -2$	1.01589	0.01589	0.01059	1.03547	0.03547	0.01182
$\overline{\sigma}$=1.9 $\gamma = -1.5$	1.01236	0.01236	0.00942	1.02840	0.02840	0.01027
$\overline{\sigma}$=1.9 $\gamma = -1$	1.00785	0.00785	0.00785	1.01912	0.01912	0.00838
$\overline{\sigma}$=1.9 $\gamma = -0.5$	1.00274	0.00274	0.00598	1.00842	0.00842	0.00637
$\overline{\sigma}$=1.9 $\gamma = -0.1$	1.06708	0.06708	0.02241	1.10531	0.10531	0.02083
MDPDE						
$\beta = 0.1$	1.01117	0.01117	0.00646	1.02676	0.02676	0.00891
$\beta = 0.5$	1.00700	0.00700	0.00712	1.01417	0.01417	0.00743
$\beta = 1$	1.01406	0.01406	0.00975	1.02892	0.02892	0.01062
$\beta = 1.5$	1.01916	0.01916	0.01148	1.03876	0.03876	0.01297
$\beta = 2$	1.02233	0.02233	0.01254	1.04448	0.04447	0.01450
$\beta = 2.5$	1.02450	0.02450	0.01342	1.04771	0.04771	0.01556
MLE	1.72587	0.72587	0.52852	2.22720	1.22720	1.50701

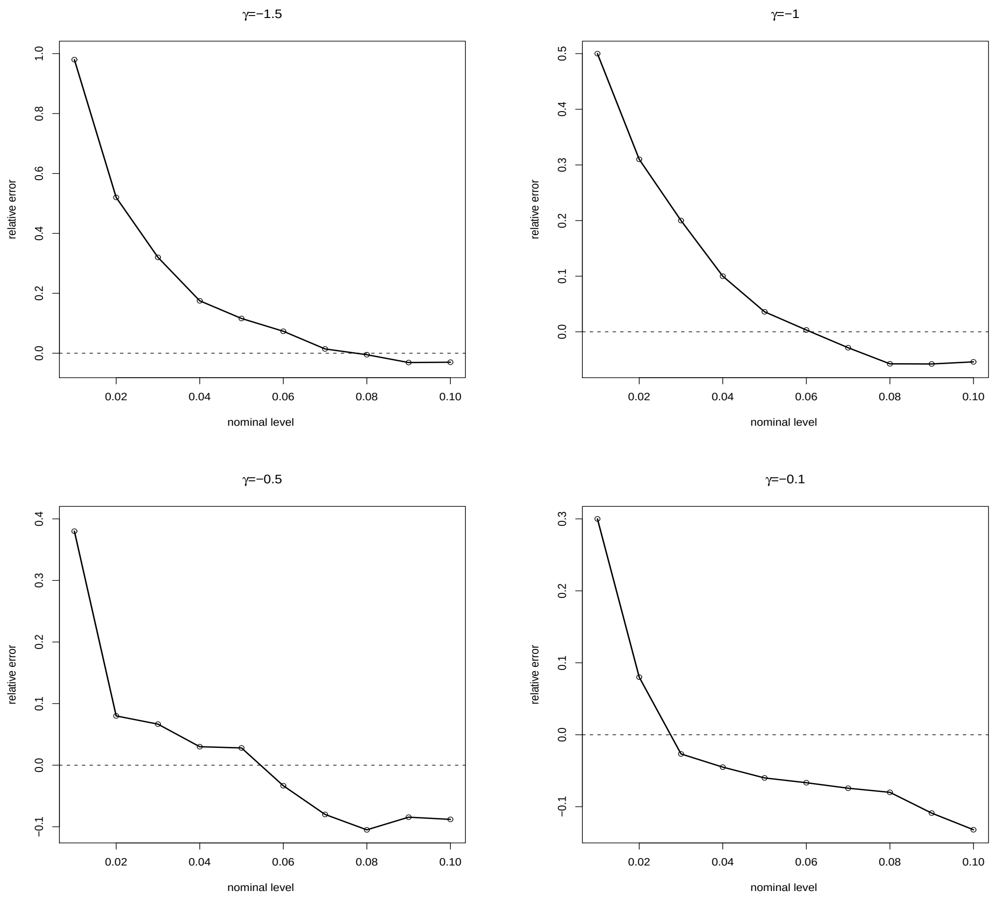

Fig. 6. Relative errors of the robust tests applied to the scale normal model $\mathcal{N}(0, 1)$, when $\overline{\sigma} = 1.9$ and 100 data are generated from model.

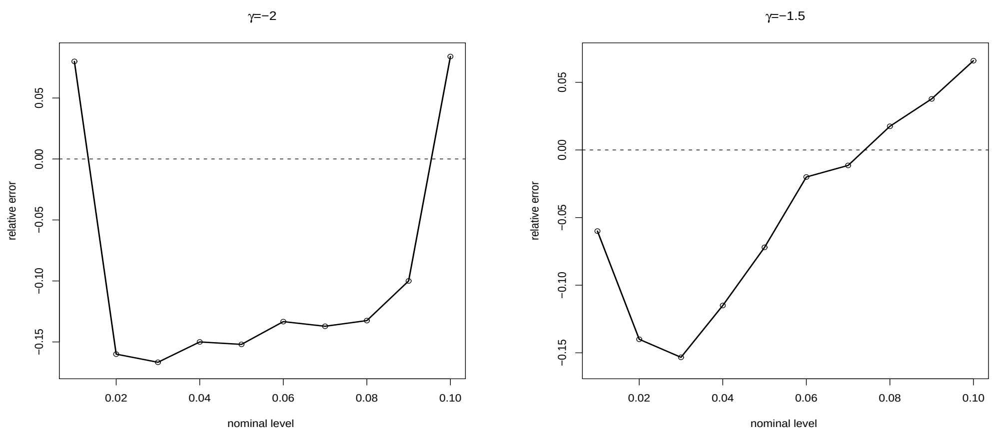

Fig. 7. Relative errors of the robust tests applied to the scale normal model $\mathcal{N}(0, 1)$, when $\overline{\sigma} = 1.9$, 98 data are generated from model and 2 outliers x=10 are added.

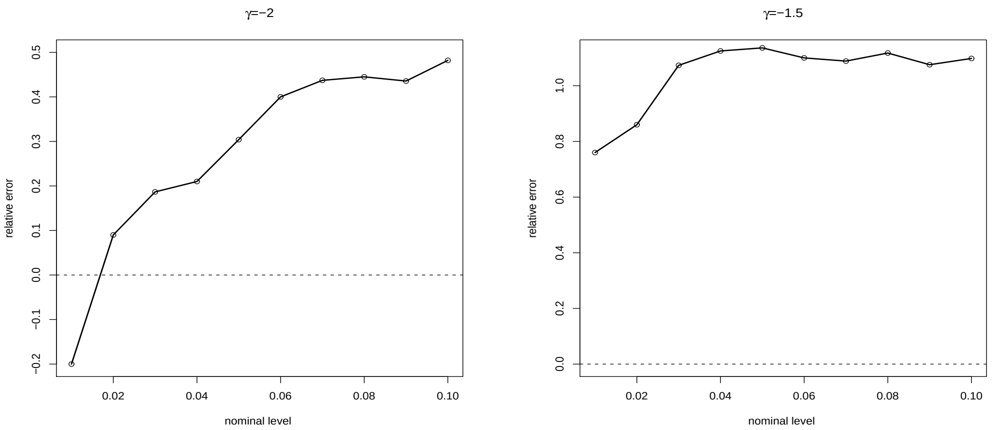

Fig. 8. Relative errors of the robust tests applied to the scale normal model $\mathcal{N}(0,1)$, when $\overline{\sigma} = 1.9$, 96 data are generated from model and 4 outliers x=10 are added.

6 An adaptive choice of the tuning parameter

At the present stage we only present some heuristic and defer the formal treatment of this proposal, which lays beyond the scope of the present work.

According to the model and the parameter to be estimated, the choice of γ should be considered with respect to the expression (11) which has to be bounded. We refer to the examples given in subsection 3.2 for some scale and location model.

Given a set of observations $X_1, \ldots, X_n$ an adaptive choice for α would aim at reducing the estimated maximal bias caused by an extraneous data. Define $\widehat{\theta}_n(\alpha,\gamma)$ the DϕE of θ_0 on the entire set of observation. For $1 \leq i \leq n$, let $\widehat{\theta}^i_{n-1}(\alpha,\gamma)$ be the DϕE of θ_0 built on the leave one out data set $X_1, X_2, \ldots, X_{i-1}$, $X_{i+1}, \ldots, X_n$. Define

$$B_n(\alpha,\gamma) := \max_i |\widehat{\theta}_n(\alpha,\gamma) - \widehat{\theta}^i_{n-1}(\alpha,\gamma)|$$

which measures the maximal bias caused by a single outlier and

$$\alpha^*(\gamma) := \arg\inf_\alpha B_n(\alpha,\gamma).$$

7 Proofs

Proof of Proposition 1

For fixed α, $\widehat{\theta}_n(\alpha)$ are M-estimators. In accordance with the theory regarding the M-estimators (see for example van der Vaart [25]), the so called ψ-function corresponding to $\widehat{\theta}_n(\alpha)$ is

$$\psi_\alpha(x,\theta) = m'(\theta,\alpha,x)$$

and the influence function of T_α is

$$\text{IF}(x;T_\alpha,P_{\theta_0}) = [M(\psi_\alpha,P_{\theta_0})]^{-1}\,\psi_\alpha\left(x,T_\alpha(P_{\theta_0})\right) \tag{30}$$

where

$$M\left(\psi_\alpha,P_{\theta_0}\right) = -\int \frac{\partial}{\partial\theta}\left[\psi_\alpha(y,\theta)\right]_{\theta_0} dP_{\theta_0}(y) = -\int m''(\theta_0,\alpha,y)dP_{\theta_0}(y).$$

Using the Fisher consistency of the functional T_α,

$$\begin{aligned}\psi_\alpha\left(x,T_\alpha(P_{\theta_0})\right) &= \psi_\alpha(x,\theta_0)\\ &= -\int \varphi''\left(\frac{p_\alpha}{p_{\theta_0}}\right)\frac{p_\alpha}{p_{\theta_0}^2}\dot{p}_{\theta_0}dP_\alpha + \varphi''\left(\frac{p_\alpha}{p_{\theta_0}}(x)\right)\frac{p_\alpha^2(x)}{p_{\theta_0}^3(x)}\dot{p}_{\theta_0}(x)\end{aligned}$$

which substituted in (30) leads to the announced result. □

Proof of Proposition 2

Let $\varepsilon > 0$ and $\widetilde{P_{\theta_0\varepsilon x}} = (1-\varepsilon)P_{\theta_0} + \varepsilon\delta_x$ be the contaminated model. Then

$$\begin{aligned}U_\alpha\left(\widetilde{P_{\theta_0\varepsilon x}}\right) &= \int m\left(T_\alpha\left(\widetilde{P_{\theta_0\varepsilon x}}\right),\alpha,y\right)d\widetilde{P_{\theta_0\varepsilon x}}(y)\\ &= (1-\varepsilon)\int m\left(T_\alpha\left(\widetilde{P_{\theta_0\varepsilon x}}\right),\alpha,y\right)dP_{\theta_0}(y) + \varepsilon m\left(T_\alpha\left(\widetilde{P_{\theta_0\varepsilon x}}\right),\alpha,x\right)\end{aligned}$$

and derivation yields

$$\begin{aligned}\text{IF}(x;U_\alpha,P_{\theta_0}) &= \frac{\partial}{\partial\varepsilon}\left[U_\alpha\left(\widetilde{P_{\theta_0\varepsilon x}}\right)\right]_{\varepsilon=0} =\\ &= -\int m(\theta_0,\alpha,y)dP_{\theta_0}(y) + \text{IF}(x;T_\alpha,P_{\theta_0})^t\int m'(\theta_0,\alpha,y)dP_{\theta_0}(y) + m(\theta_0,\alpha,x)\\ &= -\phi(\alpha,\theta_0) + m(\theta_0,\alpha,x).\end{aligned}$$

□

Proof of Proposition 3

For notational clearness, define $T:\Theta\times\mathcal{M}\to\Theta$,

$$T(\alpha,P) := T_\alpha(P).$$

For each $\alpha \in \Theta$, the definition of $T(\alpha, P)$ leads to

$$\int m'\left(T(\alpha, P), \alpha, y\right) dP(y) = 0.$$

By the very definition of $V(P)$ and $T(V(P), P)$, they both obey

$$\begin{cases} \int m'\left(T(V(P), P), V(P), y\right) dP(y) = 0 \\ \int \frac{\partial}{\partial \alpha}\left[m\left(T(\alpha, P), \alpha, y\right)\right]_{V(P)} dP(y) = 0 \end{cases}. \tag{31}$$

Denoting $n(\theta, \alpha, y) = \frac{\partial}{\partial \alpha} m(\theta, \alpha, y)$

$$\begin{aligned} n(\theta, \alpha, y) = & \int \varphi''\left(\frac{p_\alpha}{p_\theta}\right) \frac{\dot{p}_\alpha}{p_\theta} dP_\alpha + \int \varphi'\left(\frac{p_\alpha}{p_\theta}\right) \frac{\dot{p}_\alpha}{p_\alpha} dP_\alpha - \\ & - \left\{\varphi''\left(\frac{p_\alpha}{p_\theta}(y)\right) \frac{p_\alpha}{p_\theta^2}(y) \dot{p}_\alpha(y) + \varphi'\left(\frac{p_\alpha}{p_\theta}(y)\right) \frac{\dot{p}_\alpha(y)}{p_\theta(y)} - \varphi'\left(\frac{p_\alpha}{p_\theta}(y)\right) \frac{\dot{p}_\alpha(y)}{p_\theta(y)}\right\} \\ = & \int \left\{\varphi''\left(\frac{p_\alpha}{p_\theta}\right) \frac{1}{p_\theta} + \varphi'\left(\frac{p_\alpha}{p_\theta}\right) \frac{1}{p_\alpha}\right\} \dot{p}_\alpha dP_\alpha - \varphi''\left(\frac{p_\alpha}{p_\theta}(y)\right) \frac{p_\alpha}{p_\theta^2}(y) \dot{p}_\alpha(y). \end{aligned}$$

From (31)

$$\begin{cases} \int m'\left(T(V(P), P), V(P), y\right) dP(y) = 0 \\ \frac{\partial}{\partial \alpha}\left[T(\alpha, P)\right]_{V(P)} \int m'\left(T(V(P), P), V(P), y\right) dP(y) + \int n\left(T(V(P), P), V(P), y\right) dP(y) = 0 \end{cases}$$

and consequently

$$\begin{cases} \int m'\left(T(V(P), P), V(P), y\right) dP(y) = 0 \\ \int n\left(T(V(P), P), V(P), y\right) dP(y) = 0 \end{cases}.$$

For the contaminated model

$$\int n\left(T\left(V\left(\widetilde{P_{\theta_0 \varepsilon x}}\right), \widetilde{P_{\theta_0 \varepsilon x}}\right), V\left(\widetilde{P_{\theta_0 \varepsilon x}}\right), y\right) d\widetilde{P_{\theta_0 \varepsilon x}}(y) = 0$$

and so

$$\begin{aligned} &(1 - \varepsilon) \int n\left(T\left(V\left(\widetilde{P_{\theta_0 \varepsilon x}}\right), \widetilde{P_{\theta_0 \varepsilon x}}\right), V\left(\widetilde{P_{\theta_0 \varepsilon x}}\right), y\right) dP_{\theta_0}(y) + \\ &+ \varepsilon n\left(T\left(V\left(\widetilde{P_{\theta_0 \varepsilon x}}\right), \widetilde{P_{\theta_0 \varepsilon x}}\right), V\left(\widetilde{P_{\theta_0 \varepsilon x}}\right), x\right) = 0. \end{aligned}$$

Now derivation yields

$$-\int n(\theta_0,\theta_0,y)dP_{\theta_0}(y)+\int \tfrac{\partial}{\partial\theta}n\left(T(\theta_0,P_{\theta_0}),\theta_0,y\right)dP_{\theta_0}(y)\left\{\tfrac{\partial}{\partial\alpha}T(\theta_0,P_{\theta_0})\mathrm{IF}(x;V,P_{\theta_0})+\right.$$
$$\left.+\mathrm{IF}(x;T_{\theta_0},P_{\theta_0})\right\}+\int \tfrac{\partial}{\partial\alpha}n(\theta_0,\theta_0,y)dP_{\theta_0}(y)\mathrm{IF}(x;V,P_{\theta_0})+n(\theta_0,\theta_0,x)=0. \tag{32}$$

Since

$$\frac{\partial}{\partial\theta}n(\theta,\alpha,y)=\int\left\{-2\varphi''\left(\frac{p_\alpha}{p_\theta}\right)\frac{\dot p_\theta}{p_\theta^2}-\varphi'''\left(\frac{p_\alpha}{p_\theta}\right)\frac{p_\alpha}{p_\theta^3}\dot p_\theta\right\}\dot p_\alpha^t dP_\alpha+$$
$$+\left\{\varphi'''\left(\frac{p_\alpha}{p_\theta}(y)\right)+2\varphi''\left(\frac{p_\alpha}{p_\theta}(y)\right)\frac{p_\theta(y)}{p_\alpha(y)}\right\}\frac{p_\alpha^2(y)}{p_\theta^4(y)}\dot p_\theta(y)\dot p_\alpha(y)^t$$

and particularly

$$\frac{\partial}{\partial\theta}n(\theta_0,\theta_0,y)=\{2\varphi''(1)+\varphi'''(1)\}\left\{\frac{\dot p_{\theta_0}(y)\dot p_{\theta_0}(y)^t}{p_{\theta_0}^2(y)}-\int\frac{\dot p_{\theta_0}\dot p_{\theta_0}^t}{p_{\theta_0}^2}dP_{\theta_0}\right\},$$

deduce that

$$\int\frac{\partial}{\partial\theta}n(\theta_0,\theta_0,y)dP_{\theta_0}(y)=0. \tag{33}$$

On the other hand

$$\frac{\partial}{\partial\alpha}n(\theta,\alpha,y)=\int\left\{\varphi'''\left(\frac{p_\alpha}{p_\theta}\right)\frac{\dot p_\alpha}{p_\theta^2}-\varphi'\left(\frac{p_\alpha}{p_\theta}\right)\frac{\dot p_\alpha}{p_\alpha^2}+\varphi''\left(\frac{p_\alpha}{p_\theta}\right)\frac{\dot p_\alpha}{p_\alpha p_\theta}\right\}\dot p_\alpha^t dP_\alpha+$$
$$+\int\left\{\varphi''\left(\frac{p_\alpha}{p_\theta}\right)\frac{1}{p_\theta}+\varphi'\left(\frac{p_\alpha}{p_\theta}\right)\frac{1}{p_\alpha}\right\}\left(\ddot p_\alpha+\frac{\dot p_\alpha\dot p_\alpha^t}{p_\alpha}\right)dP_\alpha-$$
$$-\varphi'''\left(\frac{p_\alpha}{p_\theta}(y)\right)\frac{p_\alpha(y)}{p_\theta^3(y)}\dot p_\alpha(y)\dot p_\alpha(y)^t-$$
$$-\varphi''\left(\frac{p_\alpha}{p_\theta}(y)\right)\frac{\dot p_\alpha(y)\dot p_\alpha(y)^t}{p_\theta^2(y)}-\varphi''\left(\frac{p_\alpha}{p_\theta}(y)\right)\frac{p_\alpha(y)}{p_\theta^2(y)}\ddot p_\alpha(y)$$

and particularly

$$\frac{\partial}{\partial\alpha}n(\theta_0,\theta_0,y)=\int\{\varphi'''(1)+2\varphi''(1)\}\frac{\dot p_{\theta_0}\dot p_{\theta_0}^t}{p_{\theta_0}^2}dP_{\theta_0}+\int\varphi''(1)\frac{\ddot p_{\theta_0}}{p_{\theta_0}}dP_{\theta_0}$$
$$-\{\varphi'''(1)+\varphi''(1)\}\frac{\dot p_{\theta_0}(y)\dot p_{\theta_0}(y)^t}{p_{\theta_0}^2(y)}-\varphi''(1)\frac{\ddot p_{\theta_0}(y)}{p_{\theta_0}(y)}.$$

As a consequence

$$\int \frac{\partial}{\partial \alpha} n(\theta_0, \theta_0, y) dP_{\theta_0}(y) = \varphi''(1) \int \frac{\dot{p}_{\theta_0} \dot{p}_{\theta_0}^t}{p_{\theta_0}^2} dP_{\theta_0} = \varphi''(1) I_{\theta_0}. \tag{34}$$

Also

$$n(\theta_0, \theta_0, x) = -\varphi''(1) \frac{\dot{p}_{\theta_0}(x)}{p_{\theta_0}(x)}. \tag{35}$$

Using the Fisher consistency of the functional T_α and substituting (33), (34) and (35) in (32) it holds

$$\mathrm{IF}(x; V, P_{\theta_0}) = I_{\theta_0}^{-1} \frac{\dot{p}_{\theta_0}(x)}{p_{\theta_0}(x)} \tag{36}$$

and this completes the proof. □

Proof of Proposition 4

By replacing $\dot{p}_\theta$,

$$\begin{aligned} \left(\frac{p_\alpha(x)}{p_{\theta_0}(x)}\right)^\gamma \dot{p}_{\theta_0}(x) &= -\frac{\theta_0^{\gamma-2}}{\alpha^\gamma} \left\{ \frac{p(\alpha^{-1}x)^\gamma}{p(\theta_0^{-1}x)^{\gamma-1}} + \frac{x}{\theta_0} \left(\frac{p(\alpha^{-1}x)}{p(\theta_0^{-1}x)}\right)^\gamma \dot{p}(\theta_0^{-1}x) \right\} \\ &= -\frac{\theta_0^{\gamma-2}}{\alpha^\gamma} \left\{ \frac{p(\alpha^{-1}x)^\gamma}{p(\theta_0^{-1}x)^{\gamma-1}} - \theta_0 \frac{p(\alpha^{-1}x)^\gamma}{p(\theta_0^{-1}x)^{\gamma-1}} \frac{\partial}{\partial \theta} \log p(\theta_0^{-1}x) \right\} \end{aligned} \tag{37}$$

and similarly

$$\begin{aligned} \left(\frac{p_\alpha(x)}{p_{\theta_0}(x)}\right)^\gamma \frac{\dot{p}_{\theta_0}(x)}{p_{\theta_0}(x)} &= -\frac{\theta_0^{\gamma-1}}{\alpha^\gamma} \left\{ \left(\frac{p(\alpha^{-1}x)}{p(\theta_0^{-1}x)}\right)^\gamma + \frac{x}{\theta_0} \left(\frac{p(\alpha^{-1}x)}{p(\theta_0^{-1}x)}\right)^\gamma \frac{\dot{p}(\theta_0^{-1}x)}{p(\theta_0^{-1}x)} \right\} \\ &= -\frac{\theta_0^{\gamma-1}}{\alpha^\gamma} \left\{ \left(\frac{p(\alpha^{-1}x)}{p(\theta_0^{-1}x)}\right)^\gamma - \theta_0 \left(\frac{p(\alpha^{-1}x)}{p(\theta_0^{-1}x)}\right)^\gamma \frac{\partial}{\partial \theta} \log p(\theta_0^{-1}x) \right\}. \end{aligned}$$

The condition (A.1) together with one of the conditions (A.2) or (A.3) (depending on the choice of γ) entails that the function (37) is integrable. On the other hand (A.2) or (A.3) together with (A.4) assure that the function in the above display is bounded. Then $\mathrm{IF}(x; T_\alpha, P_{\theta_0})$ as it is expressed by (19) is a bounded function. □

Proof of Proposition 5

It holds

$$\left(\frac{p_\alpha(x)}{p_{\theta_0}(x)}\right)^\gamma \dot{p}_{\theta_0}(x) = \frac{p(x-\alpha)^\gamma}{p(x-\theta_0)^{\gamma-1}} \frac{\partial}{\partial \theta} \log p(x-\theta_0)$$

and

$$\left(\frac{p_\alpha(x)}{p_{\theta_0}(x)}\right)^\gamma \frac{\dot{p}_{\theta_0}(x)}{p_{\theta_0}(x)} = \left(\frac{p(x-\alpha)}{p(x-\theta_0)}\right)^\gamma \frac{\partial}{\partial\theta} \log p(x-\theta_0).$$

Then the condition (21) allows to conclude that $\mathrm{IF}(x; T_\alpha, P_{\theta_0})$, as it is expressed by (19), is bounded. □

Proof of Proposition 6

First prove that $\widehat{\theta}_n(\alpha) - T_\alpha(P^P_{n,\varepsilon,x}) = o_P(1)$.

It holds $\int m'(\theta_0, \alpha) dP_{\theta_0} = 0$ and

$$\int m''(\theta_0, \alpha) dP_{\theta_0} = -\int \varphi'' \left(\frac{p_\alpha}{p_{\theta_0}}\right) \frac{p_\alpha^2}{p_{\theta_0}^3} \dot{p}_{\theta_0} \dot{p}_{\theta_0}^t d\lambda = -S. \tag{38}$$

The matrix S is symmetric and positive since φ'' is positive by the convexity of φ. Using (C.3) in connection with the Lindeberg-Feller theorem for triangular arrays we have $\sqrt{n} \int m'(\theta_0, \alpha) dP_n = O_P(1)$. Using (C.4) in connection with the Lindeberg-Feller theorem for triangular arrays yields $\int m''(\theta_0, \alpha) dP_n + S = o_P(1)$.

Now, for any $\theta = \theta_0 + un^{-1/3}$ with $\|u\| \leq 1$, a Taylor expansion of $\int m(\theta, \alpha) dP_n$ around θ_0 under (C.1) yields

$$\begin{aligned} &n \int m(\theta, \alpha) dP_n - n \int m(\theta_0, \alpha) dP_n = \\ &= n^{2/3} u^t \int m'(\theta_0, \alpha) dP_n + 2^{-1} n^{1/3} u^t \int m''(\theta_0, \alpha) dP_n u + O_P(1) \end{aligned}$$

uniformly on u with $\|u\| \leq 1$. Hence

$$n \int m(\theta, \alpha) dP_n - n \int m(\theta_0, \alpha) dP_n = O_P(n^{1/6}) - 2^{-1} n^{1/3} u^t S u + O_P(1)$$

uniformly on u with $\|u\| \leq 1$. Hence uniformly on u with $\|u\| = 1$,

$$n \int m(\theta, \alpha) dP_n - n \int m(\theta_0, \alpha) dP_n \leq O_P(n^{1/6}) - 2^{-1} c n^{1/3} + O_P(1) \tag{39}$$

where c is the smallest eigenvalue of the matrix S. Note that c is positive since S is positive definite. In view of (39), by the continuity of $\theta \to \int m(\theta, \alpha) dP_n$, it holds that as $n \to \infty$, with probability one, $\theta \to \int m(\theta, \alpha) dP_n$ attains its maximum at some point $\widehat{\theta}_n(\alpha)$ in the interior of the ball $\{\theta : \|\theta - \theta_0\| \leq n^{-1/3}\}$, and therefore

$$\widehat{\theta}_n(\alpha) - \theta_0 = o_P(1). \tag{40}$$

On the other hand,

$$T_\alpha(P^P_{n,\varepsilon,x}) = \theta_0 + \frac{\varepsilon}{\sqrt{n}}\mathrm{IF}(x;T_\alpha,P_{\theta_0}) + \frac{\Delta}{\sqrt{n}}\mathbf{1} + \rho\left(\frac{\varepsilon}{\sqrt{n}},\frac{\Delta}{\sqrt{n}}\right)\frac{\sqrt{\varepsilon^2+\Delta^2}}{n}$$

where $\mathbf{1} := (1,\ldots,1)^t$ above coincides with $\frac{\partial}{\partial\widetilde{\Delta}}[T_\alpha(P_{\theta_0+\widetilde{\Delta}})]_{\widetilde{\Delta}=0}$ by the Fisher consistency of the functional T_α and the function ρ satisfies $\lim_{n\to\infty}\rho\left(\frac{\varepsilon}{\sqrt{n}},\frac{\Delta}{\sqrt{n}}\right)=0$. Then $T_\alpha(P^P_{n,\varepsilon,x})-\theta_0$ converges to zero in probability as $n\to\infty$. Combining this with (40) we obtain that $\widehat{\theta}_n(\alpha)-T_\alpha(P^P_{n,\varepsilon,x})$ converges to zero in probability.

In the following, we prove that $\sqrt{n}(\widehat{\theta}_n(\alpha)-T_\alpha(P^P_{n,\varepsilon,x})) = O_P(1)$.

By Taylor expansion, there exists $\overline{\theta}_n$ inside the segment that links $T_\alpha(P^P_{n,\varepsilon,x})$ and $\widehat{\theta}_n(\alpha)$ such that

$$\begin{aligned}
0 =& \int m'(\widehat{\theta}_n(\alpha),\alpha)dP_n \\
=& \int m'(T_\alpha(P^P_{n,\varepsilon,x}),\alpha)dP_n + \int m''(T_\alpha(P^P_{n,\varepsilon,x}),\alpha)dP_n(\widehat{\theta}_n(\alpha)-T_\alpha(P^P_{n,\varepsilon,x})) + \\
&+\frac{1}{2}(\widehat{\theta}_n(\alpha)-T_\alpha(P^P_{n,\varepsilon,x}))^t\int m'''(\overline{\theta}_n,\alpha)dP_n(\widehat{\theta}_n(\alpha)-T_\alpha(P^P_{n,\varepsilon,x})). \qquad (41)
\end{aligned}$$

By condition (C.1), using the sup-norm

$$\|\int m'''(\overline{\theta}_n,\alpha)dP_n\| = \|\frac{1}{n}\sum_{k=1}^n m'''(\overline{\theta}_n,\alpha)(X_k)\| \le \frac{1}{n}\sum_{k=1}^n H(X_k).$$

Applying the Lindeberg-Feller theorem for triangular arrays yields $\int m'''(\overline{\theta}_n,\alpha)dP_n = O_P(1)$. Then the last term in (41) writes $o_P(1)(\widehat{\theta}_n(\alpha)-T_\alpha(P^P_{n,\varepsilon,x}))$.

Under (C.1) and (C.4), by applying a Taylor expansion and repeatedly the Lindeberg-Feller theorem for triangular arrays,

$$\begin{aligned}
&\int m''(T_\alpha(P^P_{n,\varepsilon,x}),\alpha)dP_n = \frac{1}{n}\sum_{k=1}^n m''(T_\alpha(P^P_{n,\varepsilon,x}),\alpha,X_k) \\
=& \frac{1}{n}\sum_{k=1}^n m''(\theta_0,\alpha,X_k) + \frac{\varepsilon}{n\sqrt{n}}\sum_{k=1}^n m'''(\theta_0,\alpha,X_k)\mathrm{IF}(x;T_\alpha,P_{\theta_0}) + \\
&+\frac{\Delta}{n\sqrt{n}}\sum_{k=1}^n m'''(\theta_0,\alpha,X_k)\mathbf{1} + \rho\left(\frac{\varepsilon}{\sqrt{n}},\frac{\Delta}{\sqrt{n}}\right)\frac{\sqrt{\varepsilon^2+\Delta^2}}{n} \\
=& P_{\theta_0}m''(\theta_0,\alpha) + o_P(1)
\end{aligned}$$

where $\mathbf{1} := (1,\ldots,1)^t$ above coincides with $\frac{\partial}{\partial\widetilde{\Delta}}[T_\alpha(P_{\theta_0+\widetilde{\Delta}})]_{\widetilde{\Delta}=0}$ by the Fisher consistency of the functional T_α and the function ρ satisfies $\lim_{n\to\infty} \rho\left(\frac{\varepsilon}{\sqrt{n}}, \frac{\Delta}{\sqrt{n}}\right) = 0$.

Therefore (41) becomes

$$-\int m'(T_\alpha(P^P_{n,\varepsilon,x}), \alpha) dP_n = (\int m''(\theta_0, \alpha) dP_{\theta_0} + o_P(1))(\widehat{\theta}_n(\alpha) - T_\alpha(P^P_{n,\varepsilon,x})). \tag{42}$$

We prove that $\sqrt{n}\int m'(T_\alpha(P^P_{n,\varepsilon,x}), \alpha) dP_n$ is $O_P(1)$. By Taylor expansion,

$$\int m'(T_\alpha(P^P_{n,\varepsilon,x}), \alpha) dP_n = \frac{1}{n}\sum_{k=1}^n m'(\theta_0, \alpha, X_k) + \frac{\varepsilon}{n\sqrt{n}}\sum_{k=1}^n m''(\theta_0, \alpha, X_k)\mathrm{IF}(x; T_\alpha, P_{\theta_0}) +$$
$$+\frac{\Delta}{n\sqrt{n}}\sum_{k=1}^n m''(\theta_0, \alpha, X_k)\mathbf{1} + \rho\left(\frac{\varepsilon}{\sqrt{n}}, \frac{\Delta}{\sqrt{n}}\right)\frac{\sqrt{\varepsilon^2+\Delta^2}}{n}$$

and therefore

$$\sqrt{n}\int m'(T_\alpha(P^P_{n,\varepsilon,x}), \alpha) dP_n = \frac{1}{\sqrt{n}}\sum_{k=1}^n m'(\theta_0, \alpha, X_k) +$$
$$+\frac{\varepsilon}{n}\sum_{k=1}^n m''(\theta_0, \alpha, X_k)\mathrm{IF}(x; T_\alpha, P_{\theta_0}) + \frac{\Delta}{n}\sum_{k=1}^n m''(\theta_0, \alpha, X_k)\mathbf{1} + \rho\left(\frac{\varepsilon}{\sqrt{n}}, \frac{\Delta}{\sqrt{n}}\right)\sqrt{\frac{\varepsilon^2+\Delta^2}{n}}.$$

Under (C.3) and (C.4), by applying the Lindeberg-Feller theorem for triangular arrays it holds $\sqrt{n}\int m'(T_\alpha(P^P_{n,\varepsilon,x}), \alpha) dP_n = O_P(1)$. Then from (42)

$$\sqrt{n}(\widehat{\theta}_n(\alpha) - T_\alpha(P^P_{n,\varepsilon,x})) = O_P(1).$$

□

Proof of Proposition 7

By Taylor expansion, there exists $\overline{\theta}_n$ inside the segment that links $T_\alpha(P^P_{n,\varepsilon,x})$ and $\widehat{\theta}_n(\alpha)$ such that

$$\widehat{\phi}_n(\alpha, \theta_0) = \int m(\widehat{\theta}_n(\alpha), \alpha) dP_n$$
$$= \int m(T_\alpha(P^P_{n,\varepsilon,x}), \alpha) dP_n + \int m'(T_\alpha(P^P_{n,\varepsilon,x}), \alpha)^t dP_n(\widehat{\theta}_n(\alpha) - T_\alpha(P^P_{n,\varepsilon,x})) +$$
$$+\frac{1}{2}(\widehat{\theta}_n(\alpha) - T_\alpha(P^P_{n,\varepsilon,x}))^t \int m''(T_\alpha(P^P_{n,\varepsilon,x}), \alpha) dP_n(\widehat{\theta}_n(\alpha) - T_\alpha(P^P_{n,\varepsilon,x})) +$$
$$+\frac{1}{3!}\sum_{1\leq i,j,k\leq d}(\widehat{\theta}_n(\alpha) - T_\alpha(P^P_{n,\varepsilon,x}))_i(\widehat{\theta}_n(\alpha) - T_\alpha(P^P_{n,\varepsilon,x}))_j(\widehat{\theta}_n(\alpha) - T_\alpha(P^P_{n,\varepsilon,x}))_k \int \frac{\partial^3 m(\overline{\theta}_n, \alpha)}{\partial\theta_i\partial\theta_j\partial\theta_k} dP_n.$$

Then

$$\frac{\sqrt{n}(\widehat{\phi}_n(\alpha,\theta_0)-U_\alpha(P^P_{n,\varepsilon,x}))}{[\int \mathrm{IF}^2(y;U_\alpha,P^P_{n,\varepsilon,x})dP^P_{n,\varepsilon,x}(y)]^{1/2}}=\frac{\sqrt{n}(\int m(T_\alpha(P^P_{n,\varepsilon,x}),\alpha)dP_n-U_\alpha(P^P_{n,\varepsilon,x}))}{[\int \mathrm{IF}^2(y;U_\alpha,P^P_{n,\varepsilon,x})dP^P_{n,\varepsilon,x}(y)]^{1/2}}+$$

$$+\frac{(\int m'(T_\alpha(P^P_{n,\varepsilon,x}),\alpha)dP_n)^t\sqrt{n}(\widehat{\theta}_n(\alpha)-T_\alpha(P^P_{n,\varepsilon,x}))}{[\int \mathrm{IF}^2(y;U_\alpha,P^P_{n,\varepsilon,x})dP^P_{n,\varepsilon,x}(y)]^{1/2}}+$$

$$+\frac{\sqrt{n}(\widehat{\theta}_n(\alpha)-T_\alpha(P^P_{n,\varepsilon,x}))^t\int m''(T_\alpha(P^P_{n,\varepsilon,x}),\alpha)dP_n(\widehat{\theta}_n(\alpha)-T_\alpha(P^P_{n,\varepsilon,x}))}{2[\int \mathrm{IF}^2(y;U_\alpha,P^P_{n,\varepsilon,x})dP^P_{n,\varepsilon,x}(y)]^{1/2}}+$$

$$\frac{\sqrt{n}\sum_{1\leq i,j,k\leq d}(\widehat{\theta}_n(\alpha)-T_\alpha(P^P_{n,\varepsilon,x}))_i(\widehat{\theta}_n(\alpha)-T_\alpha(P^P_{n,\varepsilon,x}))_j(\widehat{\theta}_n(\alpha)-T_\alpha(P^P_{n,\varepsilon,x}))_k\int\frac{\partial^3 m(\overline{\theta}_n,\alpha)}{\partial\theta_i\partial\theta_j\partial\theta_k}dP_n}{3![\int \mathrm{IF}^2(y;U_\alpha,P^P_{n,\varepsilon,x})dP^P_{n,\varepsilon,x}(y)]^{1/2}}. \tag{43}$$

In the following we analyze each term in the above display. It holds

$$\int m(T_\alpha(P^P_{n,\varepsilon,x}),\alpha)dP_n-U_\alpha(P^P_{n,\varepsilon,x})=\frac{1}{n}\sum_{k=1}^n\{m(T_\alpha(P^P_{n,\varepsilon,x}),\alpha,X_k)-P^P_{n,\varepsilon,x}m(T_\alpha(P^P_{n,\varepsilon,x}),\alpha)\}.$$

Apply the Lindeberg-Feller theorem for the triangular array

$$Z_{n,k}:=m(T_\alpha(P^P_{n,\varepsilon,x}),\alpha,X_k)-P^P_{n,\varepsilon,x}m(T_\alpha(P^P_{n,\varepsilon,x}),\alpha).$$

For this, compute first $\mathrm{Var}(Z_{n,k})$. Observe that

$$\begin{aligned}\mathrm{Var}(Z_{n,k})&=\int m^2(T_\alpha(P^P_{n,\varepsilon,x}),\alpha,y)dP^P_{n,\varepsilon,x}(y)-\left(\int m(T_\alpha(P^P_{n,\varepsilon,x}),\alpha,y)dP^P_{n,\varepsilon,x}(y)\right)^2=\\&=\left(1-\frac{\varepsilon}{\sqrt{n}}\right)\int m^2(T_\alpha(P^P_{n,\varepsilon,x}),\alpha,y)dP_{\theta_n}(y)+\frac{\varepsilon}{\sqrt{n}}m^2(T_\alpha(P^P_{n,\varepsilon,x}),\alpha,x)-\\&\quad-\left\{\left(1-\frac{\varepsilon}{\sqrt{n}}\right)\int m(T_\alpha(P^P_{n,\varepsilon,x}),\alpha,y)dP_{\theta_n}(y)+\frac{\varepsilon}{\sqrt{n}}m(T_\alpha(P^P_{n,\varepsilon,x}),\alpha,x)\right\}^2.\end{aligned}$$

By Taylor expansions

$$\begin{aligned}m(T_\alpha(P^P_{n,\varepsilon,x}),\alpha,y)&=m(\theta_0,\alpha,y)+\frac{\varepsilon}{\sqrt{n}}m'(\theta_0,\alpha,y)^t\mathrm{IF}(x;T_\alpha,P_{\theta_0})+\\&\quad+\frac{\Delta}{\sqrt{n}}m'(\theta_0,\alpha,y)^t\mathbf{1}+\rho\left(\frac{\varepsilon}{\sqrt{n}},\frac{\Delta}{\sqrt{n}}\right)\frac{\sqrt{\varepsilon^2+\Delta^2}}{n}\\m^2(T_\alpha(P^P_{n,\varepsilon,x}),\alpha,y)&=m^2(\theta_0,\alpha,y)+2\frac{\varepsilon}{\sqrt{n}}m(\theta_0,\alpha,y)m'(\theta_0,\alpha,y)^t\mathrm{IF}(x;T_\alpha,P_{\theta_0})+\\&\quad 2\frac{\Delta}{\sqrt{n}}m(\theta_0,\alpha,y)m'(\theta_0,\alpha,y)^t\mathbf{1}+\rho\left(\frac{\varepsilon}{\sqrt{n}},\frac{\Delta}{\sqrt{n}}\right)\frac{\sqrt{\varepsilon^2+\Delta^2}}{n}.\end{aligned}$$

Hence the conditions (C.2) and (C.3) assure that $\mathrm{Var}(Z_{n,k})$ is finite.

We now prove the equality

$$\mathrm{Var}(Z_{n,k}) = \int \mathrm{IF}^2(y; U_\alpha, P^P_{n,\varepsilon,x}) dP^P_{n,\varepsilon,x}(y). \tag{44}$$

By definition,

$$\mathrm{IF}(y; U_\alpha, P^P_{n,\varepsilon,x}) = \frac{\partial}{\partial t}[U_\alpha(\widetilde{P^P_{n,\varepsilon,x_{ty}}})]_{t=0},$$

where $\widetilde{P^P_{n,\varepsilon,x_{ty}}} = (1-t)P^P_{n,\varepsilon,x} + t\delta_y$. Also

$$U_\alpha(\widetilde{P^P_{n,\varepsilon,x_{ty}}}) = (1-t)\int m(T_\alpha(\widetilde{P^P_{n,\varepsilon,x_{ty}}}), \alpha, z) dP^P_{n,\varepsilon,x}(z) + tm(T_\alpha(\widetilde{P^P_{n,\varepsilon,x_{ty}}}), \alpha, y)$$

whence

$$\mathrm{IF}(y; U_\alpha, P^P_{n,\varepsilon,x}) = -\int m(T_\alpha(P^P_{n,\varepsilon,x}), \alpha, z) dP^P_{n,\varepsilon,x}(z) +$$
$$+ \int m'(T_\alpha(P^P_{n,\varepsilon,x}), \alpha, z) dP^P_{n,\varepsilon,x}(z) \mathrm{IF}(y; T_\alpha, P^P_{n,\varepsilon,x}) + m(T_\alpha(P^P_{n,\varepsilon,x}), \alpha, y).$$

By the definition of $T_\alpha(P^P_{n,\varepsilon,x})$, it holds

$$\int m'(T_\alpha(P^P_{n,\varepsilon,x}), \alpha, z) dP^P_{n,\varepsilon,x}(z) = 0$$

and hence (44) holds. Here we observe that $\mathrm{IF}(y; T_\alpha, P^P_{n,\varepsilon,x})$ is finite for any y, for any n and any Δ, since $\mathrm{IF}(y; T_\alpha, P_{\theta_0})$ is.

Thus, by Lindeberg-Feller theorem for triangular arrays, the first term in the expansion (43) converges in distribution to a variable $\mathcal{N}(0, 1)$.

We have $\int m'(T_\alpha(P^P_{n,\varepsilon,x}), \alpha) dP_n = o_P(1)$ since $\sqrt{n}\int m'(T_\alpha(P^P_{n,\varepsilon,x}), \alpha) dP_n = O_P(1)$ (see the proof of Proposition 6). Also, it holds $\int m''(T_\alpha(P^P_{n,\varepsilon,x}), \alpha) dP_n = O_P(1)$ and $\int \frac{\partial^3 m(\overline{\theta}_n, \alpha)}{\partial\theta_i \partial\theta_j \partial\theta_k} dP_n = O_P(1)$.

Consequently, using Proposition 6 we obtain the announced result. □

Proof of Proposition 8

The level α_0 is given by

$$\begin{aligned}\alpha_0 &= 2P_{\theta_0}(\widehat{\phi}_n \geq k_n(\alpha_0)) \\ &= 2P_{\theta_0}\left(\frac{\sqrt{n}(\widehat{\phi}_n - U_\alpha(P_{\theta_0}))}{[\int \mathrm{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2}} \geq \frac{\sqrt{n}(k_n(\alpha_0) - U_\alpha(P_{\theta_0}))}{[\int \mathrm{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2}}\right).\end{aligned}$$

Using the asymptotic normality of $\widehat{\phi}_n$ in the case of uncontaminated observations (see Broniatowski and Keziou [6]),

$$\frac{\sqrt{n}(k_n(\alpha_0) - U_\alpha(P_{\theta_0}))}{[\int \mathrm{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2}} = \Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right) + o(1).$$

Therefore

$$k_n(\alpha_0) = U_\alpha(P_{\theta_0}) + \frac{1}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right)[\int \mathrm{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2} + o\left(\frac{1}{\sqrt{n}}\right). \tag{45}$$

Now we are interested in the value of the asymptotic power, when the underlying distribution deviates slightly from the model. Using (45)

$$\begin{aligned} P_{n,\varepsilon,x} &= 2P^P_{n,\varepsilon,x}(\widehat{\phi}_n \geq k_n(\alpha_0)) \\ &= 2P^P_{n,\varepsilon,x}\left(\frac{\sqrt{n}(\widehat{\phi}_n - U_\alpha(P^P_{n,\varepsilon,x}))}{[\int \mathrm{IF}^2(y; U_\alpha, P^P_{n,\varepsilon,x}) dP^P_{n,\varepsilon,x}(y)]^{1/2}} \geq \frac{\sqrt{n}(k_n(\alpha_0) - U_\alpha(P^P_{n,\varepsilon,x}))}{[\int \mathrm{IF}^2(y; U_\alpha, P^P_{n,\varepsilon,x}) dP^P_{n,\varepsilon,x}(y)]^{1/2}}\right) \\ &= 2P^P_{n,\varepsilon,x}\left(\frac{\sqrt{n}(\widehat{\phi}_n - U_\alpha(P^P_{n,\varepsilon,x}))}{[\int \mathrm{IF}^2(y; U_\alpha, P^P_{n,\varepsilon,x}) dP^P_{n,\varepsilon,x}(y)]^{1/2}} \geq -\frac{\sqrt{n}(U_\alpha(P^P_{n,\varepsilon,x}) - U_\alpha(P_{\theta_0}))}{[\int \mathrm{IF}^2(y; U_\alpha, P^P_{n,\varepsilon,x}) dP^P_{n,\varepsilon,x}(y)]^{1/2}} + \right. \\ &\left. + \Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right)\frac{[\int \mathrm{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2}}{[\int \mathrm{IF}^2(y; U_\alpha, P^P_{n,\varepsilon,x}) dP^P_{n,\varepsilon,x}(y)]^{1/2}} + o(1)\right). \end{aligned}$$

Expand $U_\alpha(P^P_{n,\varepsilon,x})$ around to $U_\alpha(P_{\theta_0})$ to obtain

$$\begin{aligned} \sqrt{n}(U_\alpha(P^P_{n,\varepsilon,x}) - U_\alpha(P_{\theta_0})) &= \varepsilon \mathrm{IF}(x; U_\alpha, P_{\theta_0}) + \Delta \frac{\partial}{\partial \widetilde{\Delta}}[U_\alpha(P_{\theta_0 + \widetilde{\Delta}})]_{\widetilde{\Delta}=0} + \\ &\quad + \rho\left(\frac{\varepsilon}{\sqrt{n}}, \frac{\Delta}{\sqrt{n}}\right)\sqrt{\frac{\varepsilon^2 + \Delta^2}{n}}. \end{aligned}$$

Using the asymptotic normality of the test statistic when the observations are i.i.d. with $P^P_{n,\varepsilon,x}$ and taking into account that

$$\lim_{n\to\infty}[\int \mathrm{IF}^2(y; U_\alpha, P^P_{n,\varepsilon,x}) dP^P_{n,\varepsilon,x}(y)]^{1/2} = [\int \mathrm{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2}$$

it holds

$$\begin{aligned} \lim_{n\to\infty} P_{n,\varepsilon,x} &= 2 - 2\Phi\left(\Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right) - \Delta \frac{\frac{\partial}{\partial \widetilde{\Delta}}[U_\alpha(P_{\theta_0 + \widetilde{\Delta}})]_{\widetilde{\Delta}=0}}{[\int \mathrm{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2}} - \right. \\ &\quad \left. - \varepsilon \frac{\mathrm{IF}(x; U_\alpha, P_{\theta_0})}{[\int \mathrm{IF}^2(y; U_\alpha, P_{\theta_0}) dP_{\theta_0}(y)]^{1/2}}\right). \end{aligned} \tag{46}$$

A simple calculation shows that $\frac{\partial}{\partial \widetilde{\Delta}}[U_\alpha(P_{\theta_0+\widetilde{\Delta}})]_{\widetilde{\Delta}=0} = \int m(\theta_0, \alpha, y)\frac{\dot{p}_{\theta_0}(y)}{p_{\theta_0}(y)}dP_{\theta_0}(y)$. Hence (28) holds. $\square$

References

[1] Basu, A., Harris, I.R., Hjort, N.L. and Jones, M.C., 1998. Robust and efficient estimation by minimising a density power divergence. Biometrika, 85 549-559.

[2] Basu, A. and Lindsay, B.G., 1994. Minimum disparity estimation for continuous models. Annals of the Institute of Statistical Mathematics, 46 683-705.

[3] Beran, R., 1977. Minimum Hellinger distance estimates for parametric models. Annals of Statistics, 5 445-463.

[4] Broniatowski, M., 2003. Estimation through Kullback-Leibler divergence. Mathematical Methods of Statistics, 12 391-409.

[5] Broniatowski, M. and Keziou, A., 2006. Minimization of ϕ-divergences on sets of signed measures. Studia Scientiarum Mathematicarum Hungarica, 43 403-442.

[6] Broniatowski, M. and Keziou, A., 2009. Parametric estimation and tests through divergences and duality technique. J. Multivariate Anal., 100 16-36.

[7] Broniatowski, M. and Leorato, S., 2006. An estimation method for Neyman chi-square divergence with application to test of hypothesis. J. Multivariate Anal., 97 1409-1436.

[8] Broniatowski, M. and Vajda, I., 2009. Several applications of divergence criteria in continuous families. arXiv:0911.0937.

[9] Cressie, N. and Read, T.R.C., 1984. Multinomial goodness of fit tests. J. R. Statist. Soc. Ser. B, 46 440-464.

[10] Dell'Aquilla, R. and Ronchetti, E., 2004. Robust tests of predictive accuracy. Metron - International Journal of Statistics, LXII 2 161-184.

[11] Hampel, F., Ronchetti, E., Rousseeuw, P.J. and Stahel, W., 1986. Robust Statistics: the Approach Based on Influence Functions. Wiley.

[12] Heritier, S. and Ronchetti, E., 1994. Robust bounded-influence tests in general parametric models. Journal of the American Statistical Association, 89 897-904.

[13] Huber, P.J., 1981. Robust Statistics. Wiley, New York.

[14] Jiménez, R. and Shao, Y., 2001. On robustness and efficiency of minimum divergence estimators. Test, 10 241-248.

[15] Liese, F. and Vajda, I., 1987. Convex Statistical Distances. BSB Teubner, Leipzig.

[16] Liese, F. and Vajda, I., 2006. On divergences and informations in statistics and information theory. IEEE Transactions of Information Theory, 52 4394-4412.

[17] Lindsay, B.G., 1994, Efficiency versus robustness: the case of minimum Hellinger distance and related methods. Annals of Statistics, 22 1081-1114.

[18] Morales, D., Pardo, L. and Vajda, I., 1995. Asymptotic divergence of estimates of discrete distribution. Journal of Statistical Planning and Inference, 38 347-369.

[19] Rüschendorf, L., 1984. On the minimum discrimination information theorem. Statistics and Decisions, Suppl 1 163-283.

[20] Simpson, D.G., 1987. Minimum Hellinger distance estimation for the analysis of count data. Journal of the American Statistical Association, 82 802-807.

[21] Simpson, D.G., 1989. Hellinger deviance tests: efficiency, breakdown points, and examples. Journal of the American Statistical Association, 84 104-113.

[22] Tamura, R.N. and Boos, D.D., 1986. Minimum Hellinger distance estimation for multivariate location and covariance. Journal of the American Statistical Association, 81 223-229.

[23] Toma, A., 2008. Minimum Hellinger distance estimators for multivariate distributions from Johnson system. Journal of Statistical Planning and Inference, 183 803-816.

[24] Toma, A., 2009. Optimal robust M-estimators using divergences. Statistics and Probability Letters, 79 1-5.

[25] van der Vaart, A.W., 1998. Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics.