Journal of Multivariate Analysis 101 (2010) 1143–1155

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Robust tests based on dual divergence estimators and saddlepoint approximations

Aida Toma [a,b,*], Samuela Leoni-Aubin [c]

[a] Mathematics Department, Academy of Economic Studies, Piaţa Romană 6, Bucharest, Romania

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

[c] INSA Lyon, ICJ, 20, Rue Albert Einstein, 69621 Villeurbanne Cedex, France

ARTICLE INFO

Article history:
Received 17 January 2009
Available online 5 November 2009

AMS subject classifications:
primary 62F03
secondary 62F35
62F10

Keywords:
Robust testing
Saddlepoint approximations
Divergences
M-estimators

ABSTRACT

This paper is devoted to robust hypothesis testing based on saddlepoint approximations in the framework of general parametric models. As is known, two main problems can arise when using classical tests. First, the models are approximations of reality and slight deviations from them can lead to unreliable results when using classical tests based on these models. Then, even if a model is correctly chosen, the classical tests are based on first order asymptotic theory. This can lead to inaccurate p-values when the sample size is moderate or small. To overcome these problems, robust tests based on dual divergence estimators and saddlepoint approximations, with good performances in small samples, are proposed.

1. Introduction

The nonrobustness of classical estimators and tests for parametric models is a well known problem and alternative robust methods have been proposed in the literature. Usually, the robust methods are based on first order asymptotic theory and their accuracy in small samples is often an open issue. In this paper, we propose test statistics which have both robustness and small sample properties. We combine robust dual divergence estimators [1,2] and saddlepoint approximations – as presented in Robinson et al. [3] – and obtain robust test statistics which are asymptotically χ^2 distributed with a relative error of order $\mathcal{O}(n^{-1})$, where n is the sample size.

Recently introduced in Broniatowski and Keziou [1] for general parametric models, the class of dual divergence estimators is based on the optimization of a new dual form of a divergence. It represents a class of M-estimators indexed by a tuning parameter and by the used divergence. Toma and Broniatowski [2] have proved that this class contains robust and efficient estimators and proposed robust test statistics based on divergence estimators. However, these robust testing methods are based on first order asymptotic theory and have absolute error of order $\mathcal{O}(n^{-1/2})$. This can lead to inaccurate p-values when the sample size is moderate to small. For this reason, we investigate new test statistics which combine robustness with good accuracy for small sample sizes, by using saddlepoint approximations.

Saddlepoint approximations have been widely studied and used in different areas in recent years due to their excellent performances. Their main property is that it provides very accurate approximations of the exact distribution of a statistic with a relative error of order $\mathcal{O}(n^{-1})$. For details on the statistical importance and applications of saddlepoint approximations

* Corresponding author at: Mathematics Department, Academy of Economic Studies, Piaţa Romană 6, Bucharest, Romania.
E-mail addresses: aida_toma@yahoo.com (A. Toma), samuela.leoni@insa-lyon.fr (S. Leoni-Aubin).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2009.11.001

we refer to the books Field and Ronchetti [4], Jensen [5] and to the papers Field [6], Field and Hampel [7], Almudevar et al. [8], Field et al. [9] for M-estimators.

The paper is organized as follows. In Section 2 we present the class of dual divergence estimators. In Section 3 we introduce test statistics for hypotheses testing based on dual divergence estimators and saddlepoint approximations. In Section 4 we show that when using robust dual divergence estimators, the corresponding saddlepoint test statistics are robust too. For several parametric models, we indicate simple specific conditions on the dual divergence estimators such that the corresponding test statistics are robust and asymptotically χ^2-distributed with a relative error of order $\mathcal{O}(n^{-1})$. In Section 5, for the scale normal model and for the Cauchy location model, we present Monte Carlo simulation studies to show the performance of the proposed tests from both robustness and small sample accuracy points of view.

2. Dual divergence estimators

The class of dual divergence estimators has been recently introduced by Broniatowski and Keziou [1]. In the following, we shortly recall their context and definition.

2.1. A class of divergence measures

Let φ be a non-negative convex function defined from $(0, \infty)$ onto $[0, \infty]$ and satisfying $\varphi(1) = 0$. Let $(\mathcal{X}, \mathcal{B})$ be a measurable space and P be a probability measure defined on $(\mathcal{X}, \mathcal{B})$. Following Rüschendorf [10], for any probability measure Q absolutely continuous with respect to P, the ϕ-divergence between Q and P is defined by

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

When Q is not absolutely continuous with respect to P, we set $\phi(Q, P) = \infty$.

The Kullback–Leibler (KL), modified Kullback–Leibler (KL_m), χ^2, modified χ^2 (χ^2_m), Hellinger (H) and L_1 divergences are respectively associated to the convex functions $\varphi(x) = x\log x - x + 1$, $\varphi(x) = -\log x + x - 1$, $\varphi(x) = \frac{1}{2}(x-1)^2$, $\varphi(x) = \frac{1}{2}(x-1)^2/x$, $\varphi(x) = 2(\sqrt{x}-1)^2$ and $\varphi(x) = |x-1|$. All these divergences excepting the L_1 one, belong to the class "power divergences" introduced in Cressie and Read [11] and defined through the class of convex functions

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

for $\gamma \in \mathbb{R} \setminus \{0, 1\}$, where $\mathbb{R}_+^*$ is the set of all non-zero positive real numbers, and otherwise by $\varphi_0(x) := -\log x + x - 1$, $\varphi_1(x) := x\log x - x + 1$. The KL divergence is associated with φ_1, the KL_m to φ_0, the χ^2 to φ_2, the χ^2_m to φ_{-1} and the Hellinger distance to $\varphi_{1/2}$.

2.2. Dual divergence estimators

Let $\{P_\theta : \theta \in \Theta\}$ be some identifiable parametric model with Θ an open subset of $\mathbb{R}^d$. Assume that for any $\theta \in \Theta$, P_θ has density p_θ with respect to some dominating σ-finite measure μ. Consider the problem of estimating the unknown true value of the parameter θ_0 on the basis of an i.i.d. sample $X_1, \ldots, X_n$ with probability measure P_{θ_0}.

Let ϕ be a divergence as defined in (1) and suppose that the corresponding function φ is C^2, strictly convex and satisfies

$$\int \left|\varphi'\left(\frac{p_\alpha}{p_\theta}\right)\right| \mathrm{d}P_\alpha < \infty, \quad \alpha, \theta \in \Theta. \tag{3}$$

With this hypothesis, using Fenchel duality technique, Broniatowski and Keziou [12] have proved the following dual representation of divergences:

$$\phi(P_\alpha, P_{\theta_0}) = \sup_{\theta\in\Theta} \int m(\theta, \alpha, x)\mathrm{d}P_{\theta_0}(x), \tag{4}$$

with

$$m(\theta, \alpha, x) := \int \varphi'\left(\frac{p_\alpha}{p_\theta}\right)\mathrm{d}P_\alpha - \left\{\varphi'\left(\frac{p_\alpha(x)}{p_\theta(x)}\right)\frac{p_\alpha(x)}{p_\theta(x)} - \varphi\left(\frac{p_\alpha(x)}{p_\theta(x)}\right)\right\}.$$

The supremum in (4) is unique and is attained in $\theta = \theta_0$, independently upon the value of α. Naturally, an estimator of the divergence between P_α and P_{θ_0} is given by

$$\phi(P_\alpha, P_n) := \sup_{\theta\in\Theta} \int m(\theta, \alpha, x)\mathrm{d}P_n(x),$$

where P_n is the empirical measure placing mass $1/n$ in each sample observation, namely $P_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$, with δ_x denoting the Dirac measure.

Minimum divergence estimators of the parameter θ_0 have been defined by minimizing $\phi(P_\alpha, P_n)$ with respect to α on the parameter space. Also, various parametric tests for simple and composite hypotheses, using $\phi(P_\alpha, P_n)$ to construct the test statistic, have been proposed; see Broniatowski and Keziou [1] and Toma and Broniatowski [2] for details on these estimation and testing procedures.

In this paper, our interest focuses on the class of estimators of θ_0, called dual divergence estimators, which are defined as follows. For a given $\alpha \in \Theta$, a dual divergence estimator of θ_0 is defined by

$$\widehat{\theta}_{\varphi,n}(\alpha) := \arg\sup_{\theta\in\Theta} \int m(\theta, \alpha, x)\mathrm{d}P_n(x). \tag{5}$$

Formula (5) determines a class of M-estimators indexed by some instrumental value of the parameter α and by the function φ specifying the divergence. The choices of α and φ represent a major feature of the estimation procedure, since they induce efficiency and robustness properties.

Note that the maximum likelihood estimator belongs to the class (5). It is obtained for $\varphi(x) = -\log x + x - 1$, that is as dual KL_m divergence estimator.

2.3. Influence functions

Robustness properties of dual divergence estimators have been studied by Toma and Broniatowski [2] by means of the influence function.

We recall that a map T which sends an arbitrary probability measure into the parameter space is a statistical functional corresponding to an estimator $\widehat{\theta}_n$ of the parameter θ_0 whenever $T(P_n) = \widehat{\theta}_n$. 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

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

where $\widetilde{P}_{\varepsilon x} = (1-\varepsilon)P + \varepsilon\delta_x$. When the influence function is bounded, the corresponding estimator is called robust.

The statistical functional $T_{\varphi,\alpha}$ associated to a dual divergence estimator $\widehat{\theta}_{\varphi,n}(\alpha)$ is defined by

$$T_{\varphi,\alpha}(P) := \arg\sup_{\theta\in\Theta} \int m(\theta, \alpha, x)\mathrm{d}P(x).$$

Let $m''(\theta, \alpha, x)$ be the $d \times d$ matrix with entries $\frac{\partial^2}{\partial\theta_i\partial\theta_j} m(\theta, \alpha, x)$ and $\dot{p}_\theta$ be the derivative with respect to θ of p_θ. Suppose that, for each α, the function $\theta \to m(\theta, \alpha, x)$ is C^2 and the matrix $\int m''(\theta_0, \alpha, y)\mathrm{d}P_{\theta_0}(y)$ exists and is invertible. For example, these conditions are satisfied in the cases of the parametric models considered in Section 4.3, for the choices indicated in each case.

Then, the influence function of a dual divergence estimator is given by

$$\mathrm{IF}(x; T_{\varphi,\alpha}, P_{\theta_0}) = \left[M(\psi_{\varphi,\alpha}, P_{\theta_0})\right]^{-1} \psi_{\varphi,\alpha}(x, \theta_0), \tag{6}$$

where

$$M\left(\psi_{\varphi,\alpha}, P_{\theta_0}\right) = -\int \frac{\partial}{\partial\theta}\left[\psi_{\varphi,\alpha}(y, \theta)\right]_{\theta_0} \mathrm{d}P_{\theta_0}(y) = -\int m''(\theta_0, \alpha, y)\mathrm{d}P_{\theta_0}(y) \tag{7}$$

and

$$\psi_{\varphi,\alpha}(x, \theta_0) = -\int \varphi''\left(\frac{p_\alpha}{p_{\theta_0}}\right)\frac{p_\alpha}{p_{\theta_0}^2}\dot{p}_{\theta_0}\mathrm{d}P_\alpha + \varphi''\left(\frac{p_\alpha(x)}{p_{\theta_0}(x)}\right)\frac{p_\alpha^2(x)}{p_{\theta_0}^3(x)}\dot{p}_{\theta_0}(x). \tag{8}$$

When using Cressie–Read divergences, formula (6) writes as

$$\mathrm{IF}(x; T_{\varphi_\gamma,\alpha}, P_{\theta_0}) = \left[-\int \left(\frac{p_\alpha}{p_{\theta_0}}\right)^\gamma \frac{\dot{p}_{\theta_0}\dot{p}_{\theta_0}^t}{p_{\theta_0}}\mathrm{d}\mu\right]^{-1} \left\{\int \left(\frac{p_\alpha}{p_{\theta_0}}\right)^\gamma \dot{p}_{\theta_0}\mathrm{d}\mu - \left(\frac{p_\alpha(x)}{p_{\theta_0}(x)}\right)^\gamma \frac{\dot{p}_{\theta_0}(x)}{p_{\theta_0}(x)}\right\}. \tag{9}$$

For scale models and location models, Toma and Broniatowski [2] have indicated conditions on α and on the Cressie–Read divergence such that the influence function $\mathrm{IF}(x; T_{\varphi_\gamma,\alpha}, P_{\theta_0})$ is a bounded function of x, and consequently the corresponding $\widehat{\theta}_{\varphi_\gamma,n}(\alpha)$ is a robust estimator of θ_0. In Section 4.3, for several parametric models, we indicate choices of α and of the Cressie–Read divergence such that $\widehat{\theta}_{\varphi_\gamma,n}(\alpha)$ is a robust estimator of θ_0 that can be used to construct test statistics with both robustness and small sample properties.

3. Saddlepoint test statistics based on dual divergence estimators

In order to test the hypothesis $H_0 : \theta = \theta_0$ in $\mathbb{R}^d$ against the alternatives $H_1 : \theta \neq \theta_0$, we consider test statistics based on dual divergence estimators $\widehat{\theta}_{\varphi,n}(\alpha)$ defined in (5).

Let $\alpha \in \Theta$ be fixed. Note that a dual divergence estimator of the parameter θ_0 is an M-estimator obtained as solution of the equation

$$\sum_{i=1}^{n} \psi_{\varphi,\alpha}(X_i, \widehat{\theta}_{\varphi,n}(\alpha)) = 0, \tag{10}$$

where $\psi_{\varphi,\alpha}(x, \theta)$ is defined as in (8).

Suppose that the cumulant generating function of the vector of scores $\psi_{\varphi,\alpha}(X, \theta)$ defined by

$$K_{\psi_{\varphi,\alpha}}(\lambda, \theta) := \log \mathrm{E}\{e^{\lambda^t \psi_{\varphi,\alpha}(X,\theta)}\}, \tag{11}$$

where the expectation is taken with respect to P_{θ_0}, exists.

We consider the test statistics $h_{\varphi,\alpha}(\widehat{\theta}_{\varphi,n}(\alpha))$, where

$$h_{\varphi,\alpha}(\theta) := \sup_{\lambda}(-K_{\psi_{\varphi,\alpha}}(\lambda, \theta)),$$

with $K_{\psi_{\varphi,\alpha}}$ defined in (11).

Following Robinson et al. [3], we will use an approximation to the p-value

$$p = P_{H_0}(h_{\varphi,\alpha}(\widehat{\theta}_{\varphi,n}(\alpha)) \geq h_{\varphi,\alpha}(\theta_{\varphi,n}(\alpha))) \tag{12}$$

of the test based on the test statistic $h_{\varphi,\alpha}(\widehat{\theta}_{\varphi,n}(\alpha))$, where $\theta_{\varphi,n}(\alpha)$ is the observed value of $\widehat{\theta}_{\varphi,n}(\alpha)$.

In order to derive the approximation of the p-value (12), we assume that the density of $\widehat{\theta}_{\varphi,n}(\alpha)$ exists and has the saddlepoint approximation

$$f_{\widehat{\theta}_{\varphi,n}(\alpha)}(t) = (2\pi/n)^{-d/2} e^{nK_{\psi_{\varphi,\alpha}}(\lambda_{\varphi,\alpha}(t),t)} |B_{\varphi,\alpha}(t)||\Sigma_{\varphi,\alpha}(t)|^{-1/2}(1 + \mathcal{O}(n^{-1})), \tag{13}$$

where $\lambda_{\varphi,\alpha}(t)$ is the saddlepoint satisfying

$$K'_{\psi_{\varphi,\alpha}}(\lambda, t) := \frac{\partial}{\partial \lambda} K_{\psi_{\varphi,\alpha}}(\lambda, t) = 0, \tag{14}$$

$|\cdot|$ denotes the determinant,

$$B_{\varphi,\alpha}(t) := e^{-K_{\psi_{\varphi,\alpha}}(\lambda_{\varphi,\alpha}(t),t)} \mathrm{E}\left\{\frac{\partial}{\partial t}\psi_{\varphi,\alpha}(X, t) e^{\lambda^t \psi_{\varphi,\alpha}(X,t)}\right\}$$

and

$$\Sigma_{\varphi,\alpha}(t) := e^{-K_{\psi_{\varphi,\alpha}}(\lambda_{\varphi,\alpha}(t),t)} \mathrm{E}\left\{\psi_{\varphi,\alpha}(X, t)\psi_{\varphi,\alpha}(X, t)^t e^{\lambda^t \psi_{\varphi,\alpha}(X,t)}\right\}.$$

The saddlepoint approximation of the form (13) was introduced in Field [6] for a general M-estimator and has subsequently been considered by Skovgaard [13] and Almudevar et al. [8]. Conditions which imply this saddlepoint approximation are given in [8].

Under the assumption that the density of the dual divergence estimator $\widehat{\theta}_{\varphi,n}(\alpha)$ exists and admits the saddlepoint approximation (13), the p-value (12) admits the approximations (1.5) and (1.6) in [3]. This means that the test statistic $2nh_{\varphi,\alpha}(\widehat{\theta}_{\varphi,n}(\alpha))$ is asymptotically χ^2 with a relative error of order $\mathcal{O}(n^{-1})$.

Our interest is to combine the test accuracy in small samples with the robustness property of the dual divergence estimator. The result of this combination is discussed in Section 4.

4. Robust saddlepoint test statistics based on dual divergence estimators

Parametric models are idealized approximations of reality and slight deviations from them can have significant effects on classical estimators and tests based on these models. In spite of their second order accuracy, classical inference based on saddlepoint approximations can be affected by small deviations from the assumptions on the model (see [14]). Therefore, robust alternatives need to be looked for. To achieve both robustness and small sample properties, we combine results from Toma and Broniatowski [2] and Robinson et al. [3] and obtain robust test statistics for hypotheses testing which are asymptotically χ^2 distributed, with a relative error of order $\mathcal{O}(n^{-1})$.

For testing the hypothesis $H_0 : \theta = \theta_0$ with respect to the alternatives $\theta \neq \theta_0$, we consider the test statistics $h_{\varphi,\alpha}(\widehat{\theta}_{\varphi,n}(\alpha))$, with $\widehat{\theta}_{\varphi,n}(\alpha)$ robust dual divergence estimators of θ_0. As it is shown in the following subsections, these test statistics are also robust and, moreover, their breakdown point is greater or equal to the breakdown point of the corresponding dual divergence estimator.

4.1. Influence functions

The influence function of a test statistic shows the influence of an outlier in the sample on the value of the test statistic and hence on the decision (acceptance or rejection of H_0) which is based on this value. Boundedness of the influence function of the test statistic implies that in a neighborhood of the model, the level of the test does not become arbitrarily close to 1 and the power of the test does not become arbitrarily close to 0.

The statistical functional associated to a test statistic $h_{\varphi,\alpha}(\widehat{\theta}_{\varphi,n}(\alpha))$ is defined by $U_{\varphi,\alpha}(P) := h_{\varphi,\alpha}(T_{\varphi,\alpha}(P))$, where $T_{\varphi,\alpha}$ is the statistical functional corresponding to $\widehat{\theta}_{\varphi,n}(\alpha)$. As it can be deduced from the following result, the test statistic is robust, because $\widehat{\theta}_{\varphi,n}(\alpha)$ itself is a robust estimator of θ_0.

Proposition 1. *The influence function of the test statistic $h_{\varphi,\alpha}(\widehat{\theta}_{\varphi,n}(\alpha))$ is*

$$\mathrm{IF}(x; U_{\varphi,\alpha}, P_{\theta_0}) = h'_{\varphi,\alpha}(\theta_0)^t [M(\psi_{\varphi,\alpha}, P_{\theta_0})]^{-1} \psi_{\varphi,\alpha}(x, \theta_0), \tag{15}$$

where

$$h'_{\varphi,\alpha}(\theta_0) = -\frac{E\{e^{\lambda_{\varphi,\alpha}(\theta_0)^t \psi_{\varphi,\alpha}(X,\theta_0)} \frac{\partial}{\partial\theta}\psi_{\varphi,\alpha}(X,\theta_0)\lambda_{\varphi,\alpha}(\theta_0)\}}{E\{e^{\lambda_{\varphi,\alpha}(\theta_0)^t \psi_{\varphi,\alpha}(X,\theta_0)}\}},$$

the expectation being taken with respect to P_{θ_0}, $M(\psi_{\varphi,\alpha}, P_{\theta_0})$ is given by (7) and $\psi_{\varphi,\alpha}(x, \theta_0)$ is given by (8).

Proof. First, observe that

$$T_{\varphi,\alpha}(P_{\theta_0}) = \arg\sup_{\theta\in\Theta} \int m(\theta, \alpha, x)\mathrm{d}P_{\theta_0}(x) = \theta_0,$$

since the function $\theta \to \int m(\theta, \alpha, x)\mathrm{d}P_{\theta_0}(x)$ has a unique maximizer $\theta = \theta_0$ (see [1]).

Then, for the contaminated model $\widetilde{P_{\theta_0 \varepsilon x}} := (1-\varepsilon)P_{\theta_0} + \varepsilon\delta_x$, where $\varepsilon > 0$, it holds $U_{\varphi,\alpha}(\widetilde{P_{\theta_0 \varepsilon x}}) = h_{\varphi,\alpha}(T_{\varphi,\alpha}(\widetilde{P_{\theta_0 \varepsilon x}}))$ and differentiation with respect to ε yields

$$\mathrm{IF}(x; U_{\varphi,\alpha}, P_{\theta_0}) = \frac{\partial}{\partial\varepsilon}[U_{\varphi,\alpha}(\widetilde{P_{\theta_0 \varepsilon x}})]_{\varepsilon=0} = h'_{\varphi,\alpha}(\theta_0)^t \mathrm{IF}(x; T_{\varphi,\alpha}, P_{\theta_0}). \tag{16}$$

Differentiation with respect to θ yields

$$\begin{aligned} h'_{\varphi,\alpha}(\theta) &= -\frac{\partial}{\partial\lambda}[K_{\psi_{\varphi,\alpha}}(\lambda, \theta)]_{\lambda=\lambda_{\varphi,\alpha}(\theta)}\lambda'_{\varphi,\alpha}(\theta) - \frac{\partial}{\partial\theta}[K_{\psi_{\varphi,\alpha}}(\lambda, \theta)]_{\lambda=\lambda_{\varphi,\alpha}(\theta)} \\ &= -\frac{\partial}{\partial\theta}K_{\psi_{\varphi,\alpha}}(\lambda_{\varphi,\alpha}(\theta), \theta), \end{aligned}$$

by using the definition of $\lambda_{\varphi,\alpha}(\theta)$. Consequently,

$$h'_{\varphi,\alpha}(\theta_0) = -\frac{E\{e^{\lambda_{\varphi,\alpha}(\theta_0)^t \psi_{\varphi,\alpha}(X,\theta_0)} \frac{\partial}{\partial\theta}\psi_{\varphi,\alpha}(X,\theta_0)\lambda_{\varphi,\alpha}(\theta_0)\}}{E\{e^{\lambda_{\varphi,\alpha}(\theta_0)^t \psi_{\varphi,\alpha}(X,\theta_0)}\}}, \tag{17}$$

which exists since $\psi_{\varphi,\alpha}(x, \theta_0)$ is bounded and the integral $\int \frac{\partial}{\partial\theta}\psi_{\varphi,\alpha}(x, \theta_0)p_{\theta_0}(x)\mathrm{d}x$ exists.

Substituting $h'_{\varphi,\alpha}(\theta_0)$ from (17) and $\mathrm{IF}(x; T_{\varphi,\alpha}, P_{\theta_0})$ from (6) in (16) we obtain (15). □

4.2. Breakdown point

The breakdown point, as well as the influence function, provides information about the stability of an estimator or of a test statistic at the presence of outliers. It quantifies how much a small change in the underlying distribution impacts on the distribution of estimators or test statistics.

We prove that a dual divergence estimator with high breakdown point will induce this characteristic to the corresponding test statistic.

The breakdown point of an estimator $\widehat{\theta}_n$ of a parameter θ_0 is the largest amount of contamination that the data may contain, such that $\widehat{\theta}_n$ still gives some information about θ_0. Following Maronna et al. [15] (p. 58), the asymptotic contamination breakdown point of an estimator $\widehat{\theta}_n$ at P_{θ_0}, denoted by $\varepsilon^*(\widehat{\theta}_n, \theta_0)$, is the largest $\varepsilon^* \in (0, 1)$ such that for $\varepsilon < \varepsilon^*$, $T((1-\varepsilon)P_{\theta_0} + \varepsilon P)$ as function of P remains bounded and also bounded away from the boundary $\partial\Theta$ of Θ. Here, $T((1-\varepsilon)P_{\theta_0} + \varepsilon P)$ is the asymptotic value of the estimator at $(1-\varepsilon)P_{\theta_0} + \varepsilon P$ by means of the convergence in probability. The definition means that there exists a bounded and closed set $K \subset \Theta$ such that $K \cap \partial\Theta = \emptyset$ and

$$T((1-\varepsilon)P_{\theta_0} + \varepsilon P) \in K, \quad \forall\varepsilon < \varepsilon^*, \ \forall P.$$

Similarly, it can be defined the asymptotic contamination breakdown point of a test statistic at P_{θ_0}.

Proposition 2. *Assuming that the asymptotic breakdown point of $\widehat{\theta}_{\varphi,n}(\alpha)$ at P_{θ_0} exists, the asymptotic breakdown point of the test statistic $h_{\varphi,\alpha}(\widehat{\theta}_{\varphi,n}(\alpha))$ at P_{θ_0} exists too and verifies*

$$\varepsilon^*(h_{\varphi,\alpha}(\widehat{\theta}_{\varphi,n}(\alpha)), P_{\theta_0}) \geq \varepsilon^*(\widehat{\theta}_{\varphi,n}(\alpha), P_{\theta_0}). \tag{18}$$

Proof. Let $\varepsilon < \varepsilon^*(\widehat{\theta}_{\varphi,n}(\alpha), P_{\theta_0})$. Then there exists a bounded and closed set $K \subset \Theta$ such that $K \cap \partial\Theta = \emptyset$ and

$$T_{\varphi,\alpha}((1-\varepsilon)P_{\theta_0} + \varepsilon P) \in K, \quad \forall P.$$

Since $h_{\varphi,\alpha}$ is continuous as function of θ and the spaces on which it is defined, respectively, takes values are separated, using Borel–Lebesgue theorem, $h_{\varphi,\alpha}(K)$ is compact. Therefore we get the existence of the set $h_{\varphi,\alpha}(K)$ which is bounded and closed in $[0, \infty)$ such that

$$U_{\varphi,\alpha}((1-\varepsilon)P_{\theta_0} + \varepsilon P) = h_{\varphi,\alpha}(T_{\varphi,\alpha}((1-\varepsilon)P_{\theta_0} + \varepsilon P)) \in h_{\varphi,\alpha}(K), \quad \forall P.$$

Consequently, the inequality (18) is verified. □

It is difficult to give a general theoretic result concerning the breakdown point of dual divergence estimators. The fact that the corresponding ψ-function, given by (19) for Cressie–Read divergences, has a part which depends on θ but does not depend on x induces some difficulties when adapting known methods of finding the breakdown point for M-estimators. However, Monte Carlo simulations show that in many cases the dual divergence estimators have high breakdown point. In some particular cases we can confirm that from theoretical point of view.

For example, for location models, when the ψ-function of the dual divergence estimator, as function of $x-\theta$, is increasing and it can be established that $\psi(\infty) = k_2$ and $\psi(-\infty) = -k_1$, the asymptotic breakdown point is

$$\varepsilon^*(\widehat{\theta}_{\varphi,n}(\alpha), P_{\theta_0}) = \frac{\min\{k_1, k_2\}}{k_1 + k_2}$$

(see [15], p. 59). This is the case of some dual divergence estimators corresponding to the KL_m divergence. For instance, for the logistic location model, the ψ-function associated to the dual KL_m divergence estimator of the location parameter θ_0 is

$$\psi(x, \theta) = \frac{1 - e^{\theta - x}}{1 + e^{\theta - x}},$$

hence $k_1 = k_2 = 1$ and $\varepsilon^*(\widehat{\theta}_{\varphi,n}(\alpha), P_{\theta_0}) = 0.5$.

Moreover, the dual divergence estimators of a location parameter that are redescending M-estimators have the asymptotic breakdown point 0.5 (see [15], p. 59). This is the case of the dual KL_m divergence estimator of the location parameter from the Cauchy model, for which

$$\psi(x, \theta) = \frac{2(x-\theta)}{1 + (x-\theta)^2}.$$

4.3. Choices of the tuning parameter and of the Cressie–Read divergence

In the following, for some parametric models, we provide simple conditions such that $\widehat{\theta}_{\varphi,n}(\alpha)$ is a robust estimator of θ_0 and admits a saddlepoint approximation of the density of the form (13). These simple conditions turn out when we check in each case the conditions settled by Almudevar et al. [8] in order to have the saddlepoint approximation of the density of the dual divergence estimator. They assure that the test statistic is robust and asymptotically χ^2 distributed with a relative error of order $\mathcal{O}(n^{-1})$. In Section 5, we use these conditions for the scale normal model and for the Cauchy location model in order to perform some Monte Carlo studies.

We consider the Cressie–Read divergences. For these divergences, the ψ-function corresponding to a dual divergence estimator is

$$\psi_{\varphi_\gamma,\alpha}(x, \theta) = -\int \left(\frac{p_\alpha}{p_\theta}\right)^\gamma \dot{p}_\theta \, d\mu + \left(\frac{p_\alpha(x)}{p_\theta(x)}\right)^\gamma \frac{\dot{p}_\theta(x)}{p_\theta(x)}. \tag{19}$$

Consider the normal distribution $\mathcal{N}(m, \sigma^2)$ with known mean, σ being the parameter of interest. The ψ-function of a dual divergence estimator of σ is

$$\psi_{\varphi_\gamma,\overline{\sigma}}(x, \sigma) = I_{\gamma,\overline{\sigma}}(\sigma) + c_{\gamma,\overline{\sigma}}(x, \sigma), \tag{20}$$

where

$$I_{\gamma,\overline{\sigma}}(\sigma) = \frac{\sigma^{\gamma-2}}{\overline{\sigma}^\gamma \sqrt{2\pi}} \int \left[1 - \left(\frac{x-m}{\sigma}\right)^2\right] e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \left(e^{-\frac{1}{2}\left[\left(\frac{x-m}{\overline{\sigma}}\right)^2 - \left(\frac{x-m}{\sigma}\right)^2\right]}\right)^\gamma dx$$

and

$$c_{\gamma,\overline{\sigma}}(x,\sigma)=\frac{\sigma^{\gamma-1}}{\overline{\sigma}^{\gamma}}\left[\left(\frac{x-m}{\sigma}\right)^2-1\right]\left(\mathrm{e}^{-\frac{1}{2}\left[\left(\frac{x-m}{\overline{\sigma}}\right)^2-\left(\frac{x-m}{\sigma}\right)^2\right]}\right)^{\gamma}.$$

Here, $\overline{\sigma}$ plays the role of α. In [2] it is shown that $\psi_{\varphi_\gamma,\overline{\sigma}}(x,\sigma)$ is bounded as function of x when $\gamma>0$ and $\overline{\sigma}<\sigma$, respectively when $\gamma<0$ and $\overline{\sigma}>\sigma$. In each of these situations, the integral $M(\psi_{\varphi_\gamma,\overline{\sigma}},P_\sigma)$ exists and is invertible, therefore $\widehat{\theta}_{\varphi_\gamma,n}(\overline{\sigma})$ is robust. The conditions of Almudevar et al. [8] are satisfied if additionally $\gamma<0$. Thus we state that whenever $\gamma<0$ and $\overline{\sigma}>\sigma$, $\widehat{\theta}_{\varphi_\gamma,n}(\overline{\sigma})$ is a robust estimator of σ and its density have a saddlepoint approximation of the form (13).

For the exponential distribution, the ψ-function corresponding to a dual divergence estimator of θ_0 is

$$\psi_{\varphi_\gamma,\alpha}(x,\theta)=I_{\gamma,\alpha}(\theta)+c_{\gamma,\alpha}(x,\theta),$$

with

$$I_{\gamma,\alpha}(\theta)=\int_0^\infty\left(\frac{\theta}{\alpha}\right)^{\gamma}(\mathrm{e}^{-x(\frac{1}{\alpha}-\frac{1}{\theta})})^{\gamma}\left(\frac{x}{\theta^3}-\frac{1}{\theta^2}\right)\mathrm{e}^{-\frac{x}{\theta}}\,\mathrm{d}x$$

and

$$c_{\gamma,\alpha}(x,\theta)=\left(\frac{\theta}{\alpha}\right)^{\gamma}(\mathrm{e}^{-x(\frac{1}{\alpha}-\frac{1}{\theta})})^{\gamma}\left(\frac{x}{\theta^2}-\frac{1}{\theta}\right).$$

The dual divergence estimator is robust when $\gamma>0$ and $\alpha<\theta$ or when $\gamma<0$ and $\alpha>\theta$. If moreover $\gamma<0$, the conditions to have the saddlepoint approximation of the density of dual divergence estimator are satisfied. Therefore the choice will be with respect to $\gamma<0$ and $\alpha>\theta$.

For the Cauchy distribution,

$$\psi_{\varphi_\gamma,\alpha}(x,\theta)=I_{\gamma,\alpha}(\theta)+c_{\gamma,\alpha}(x,\theta),\tag{21}$$

where

$$I_{\gamma,\alpha}(\theta)=-\int\left(\frac{1+(x-\theta)^2}{1+(x-\alpha)^2}\right)^{\gamma}\frac{2(x-\theta)}{\pi[1+(x-\theta)^2]^2}\,\mathrm{d}x$$

and

$$c_{\gamma,\alpha}(x,\theta)=\frac{2(x-\theta)}{\pi[1+(x-\theta)^2]}\left(\frac{1+(x-\theta)^2}{1+(x-\alpha)^2}\right)^{\gamma}.$$

Note that $\psi_{\varphi_\gamma,\alpha}(x,\theta)$ is bounded with respect to x for any α and any γ and that $\widehat{\theta}_{\varphi_\gamma,n}(\alpha)$ is always robust. In the case of this distribution, any α together with $\gamma\le 1$ assure the conditions of robust testing procedure.

In the case of the logistic distribution,

$$\psi_{\varphi_\gamma,\alpha}(x,\theta)=I_{\gamma,\alpha}(\theta)+c_{\gamma,\alpha}(x,\theta),$$

where

$$I_{\gamma,\alpha}(\theta)=-(\mathrm{e}^{\alpha-\theta})^{\gamma}\int\left(\frac{1+\mathrm{e}^{-(x-\theta)}}{1+\mathrm{e}^{-(x-\alpha)}}\right)^{2\gamma}\frac{\mathrm{e}^{-(x-\theta)}-\mathrm{e}^{-2(x-\theta)}}{(1+\mathrm{e}^{-(x-\theta)})^3}\,\mathrm{d}x$$

and

$$c_{\gamma,\alpha}(x,\theta)=(\mathrm{e}^{\alpha-\theta})^{\gamma}\left(\frac{1+\mathrm{e}^{-(x-\theta)}}{1+\mathrm{e}^{-(x-\alpha)}}\right)^{2\gamma}\frac{1-\mathrm{e}^{-(x-\theta)}}{1+\mathrm{e}^{-(x-\theta)}}.$$

Here the robust testing methodology works for any choice of α and any $\gamma\le 0$.

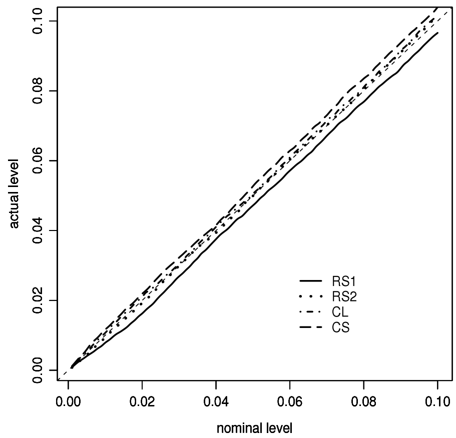
If we consider X distributed as a vector of three independent exponential variables with means θ_i, $i=1,2,3$, the conditions to apply the robust testing procedure reduce to a choice of the Cressie–Read divergence for which $\gamma<0$ and to a choice of α such that $\alpha_i>\theta_i$, where α_i are the components of α.

5. Monte Carlo results

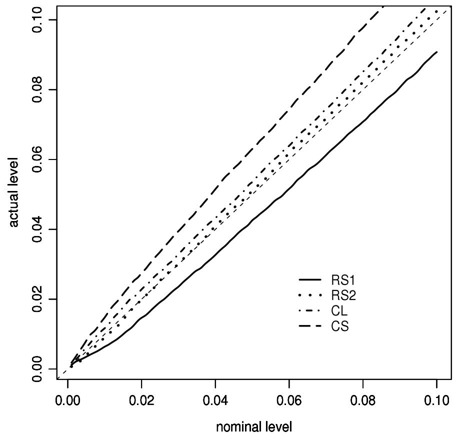
In order to illustrate the behavior of the tests that we propose, we perform Monte Carlo experiments for two parametric models, namely for the scale normal model and for the Cauchy location model. We work with data generated from the considered model and from slight perturbations of it. For some values of the tuning parameter α and for some Cressie–Read divergences, as indicated in the Section 4.3, we compute $\widehat{\theta}_{\varphi_\gamma,n}(\alpha)$ and the corresponding test statistic $2nh_{\varphi_\gamma,\alpha}(\widehat{\theta}_{\varphi_\gamma,n}(\alpha))$. Under the null hypothesis, these test statistics are asymptotically χ_1^2. In each experiment we simulate 50 000 samples and we report the actual levels $P(2nh_{\varphi_\gamma,\alpha}(\widehat{\theta}_{\varphi_\gamma,n}(\alpha))\ge v_{\alpha_0})$ of the tests based on the test statistics $2nh_{\varphi_\gamma,\alpha}(\widehat{\theta}_{\varphi_\gamma,n}(\alpha))$ corresponding to 100 values of the nominal level $\alpha_0=1/1000,2/1000,\ldots,100/1000$, v_{α_0} being the critical value given by $P(\chi_1^2\ge v_{\alpha_0})=\alpha_0$. We also report the relative errors $(P(2nh_{\varphi_\gamma,\alpha}(\widehat{\theta}_{\varphi_\gamma,n}(\alpha))\ge v_{\alpha_0})-\alpha_0)/\alpha_0$.

5.1. The scale normal model

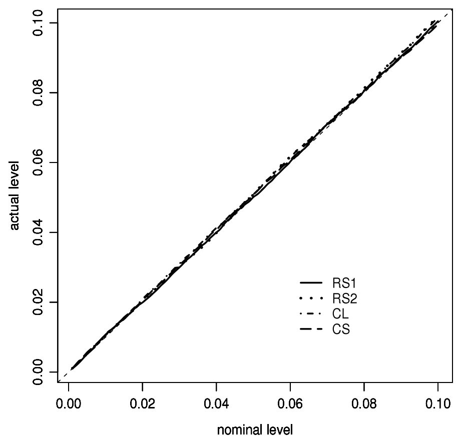
In the first Monte Carlo experiment, we consider the scale normal model with known mean, σ being the parameter of interest. The null hypothesis is $H_0 : \sigma = 1$.



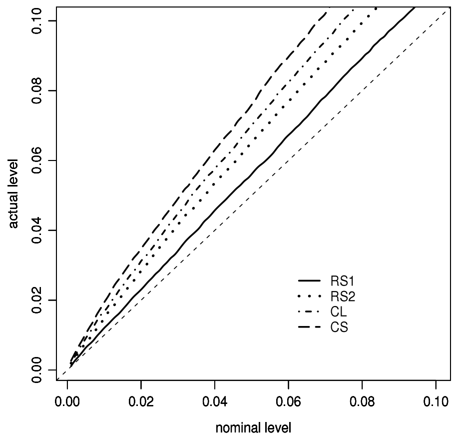
(a) $n = 20$.



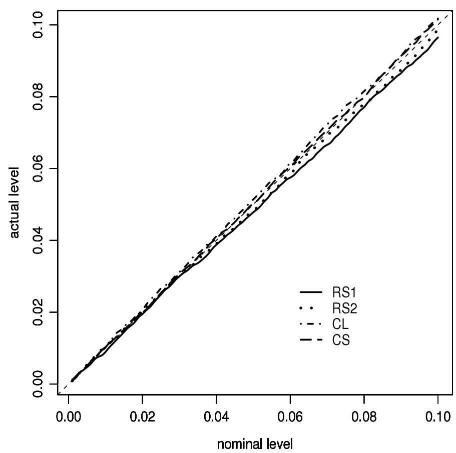
(b) $n = 20$.



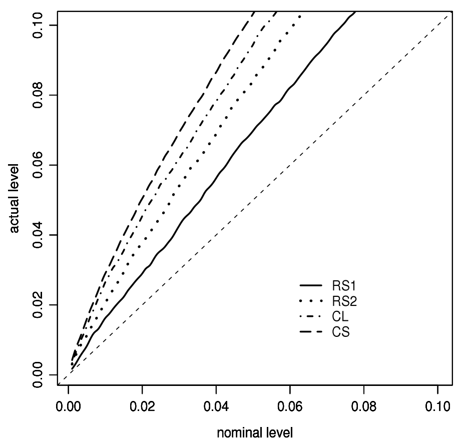
(c) $n = 50$.



(d) $n = 50$.



(e) $n = 100$.



(f) $n = 100$.

Fig. 1. Comparison of the actual level and the nominal level of the tests RS1, RS2, CL and CS.

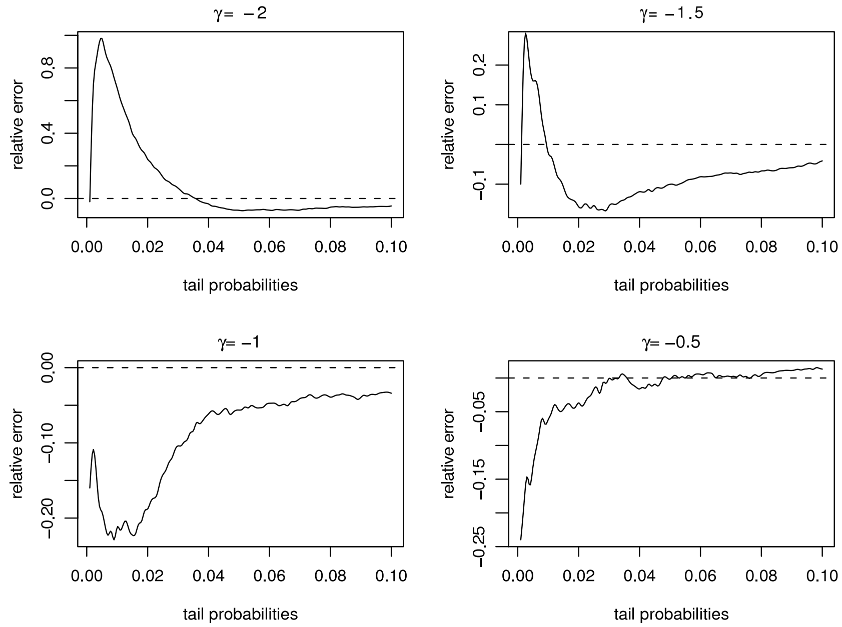


Fig. 2. Relative errors of the robust tests applied to the scale normal model $\mathcal{N}(0, 1)$, when the data are generated from the model, the tuning parameter is $\overline{\sigma} = 1.7$ and $\gamma = -2, -1.5, -1, -0.5$.

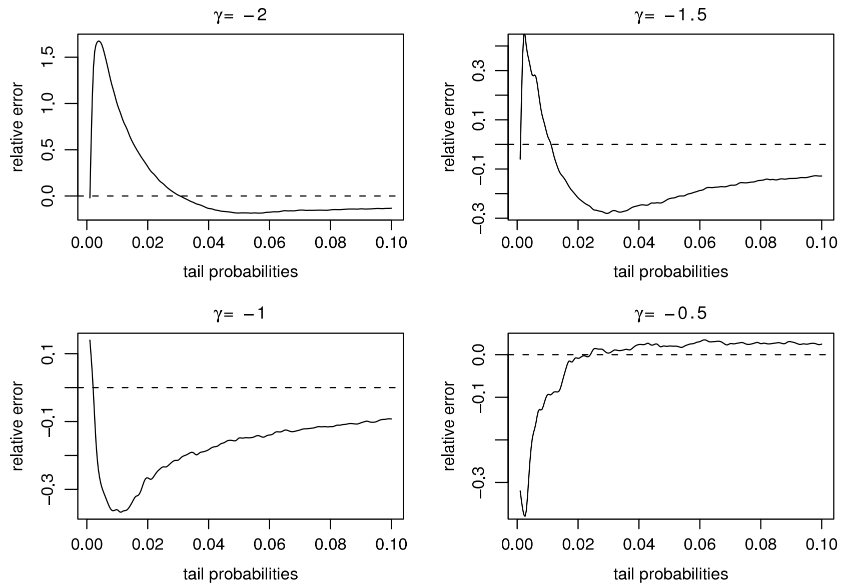


Fig. 3. Relative errors of the robust tests applied to the scale normal model $\mathcal{N}(0, 1)$, when the data are generated from $0.95\mathcal{N}(0, 1) + 0.05\mathcal{N}(0, 2)$, the tuning parameter is $\overline{\sigma} = 1.7$ and $\gamma = -2, -1.5, -1, -0.5$.

In order to show the robustness of the proposed test, we consider data generated from model, and then from small neighborhoods of it. More precisely, first, the data are generated from the normal distribution $\mathcal{N}(0, 1)$ and then from the perturbed distributions of the form $(1 - \varepsilon)\mathcal{N}(0, 1) + \varepsilon\mathcal{N}(0, 2)$, where $\varepsilon = 0.05, 0.1$.

The choice of the tuning parameter is $\overline{\sigma} = 1.7$ and the Cressie–Read divergences that we consider correspond to $\gamma \in \{-2, -1.5, -1, -0.5\}$. These choices are in accordance with the remarks in the Section 4.3, referring to the conditions to achieve robustness and small sample accuracy of the testing procedure. For each of these divergences, we compute the dual divergence estimator of the parameter $\sigma = 1$ (as solution of the Eq. (10) with the ψ-function (20)) and the corresponding test statistic.

In addition to these robust saddlepoint test statistics, we consider the classical χ^2 test statistic $\sum_{i=1}^{n} X_i^2$, as well as the saddlepoint test statistic corresponding to the ψ-function $\psi(x, \theta) = x^2 - \theta^2$. The last ψ-function is associated to the maximum likelihood estimator of $\sigma = 1$.

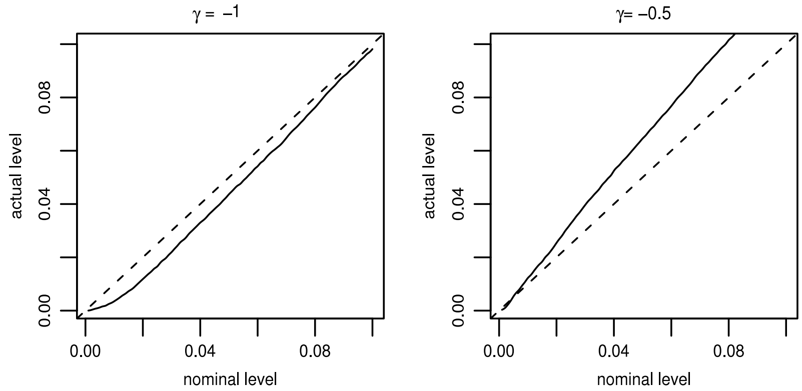


Fig. 4. Comparison of the actual level and the nominal level of the robust tests applied to the scale normal model $\mathcal{N}(0, 1)$, when the data are generated from $0.9\mathcal{N}(0, 1) + 0.1\mathcal{N}(0, 2)$, the tuning parameter is $\overline{\sigma} = 1.7$ and $\gamma = -1, -0.5$.

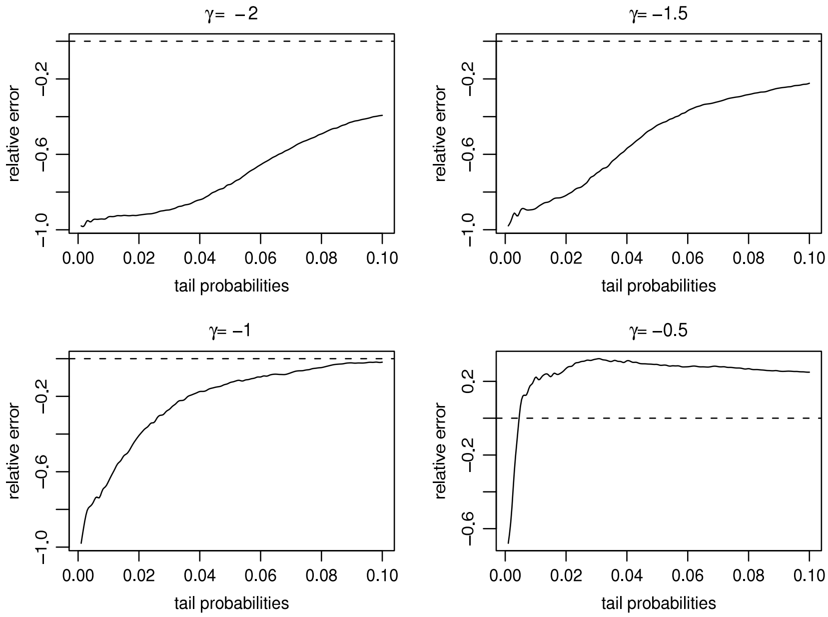


Fig. 5. Relative errors of the robust tests applied to the scale normal model $\mathcal{N}(0, 1)$, when the data are generated from $0.9\mathcal{N}(0, 1) + 0.1\mathcal{N}(0, 2)$, the tuning parameter is $\overline{\sigma} = 1.7$ and $\gamma = -2, -1.5, -1, -0.5$.

In Fig. 1 we compare the tests corresponding to these test statistics (RS1 and RS2 are the robust saddlepoint tests for $\gamma = -1$ and $\gamma = -0.5$, CL is the classical χ^2 test and CS is the classical saddlepoint test). We represent the actual levels of the tests to be compared with the nominal levels. The data are generated from the model $\mathcal{N}(0, 1)$ in (a), (c) and (e), respectively from the perturbed model $0.95\mathcal{N}(0, 1) + 0.05\mathcal{N}(0, 2)$ in (b), (d) and (f). The sample size is $n = 20, 50, 100$. When the data are not contaminated, all the tests give very good results, for all the sample sizes. When the data are contaminated, the robust saddlepoint tests give better results than the classical χ^2 test or the classical saddlepoint test. The best results are obtained by the robust saddlepoint test corresponding to $\gamma = -0.5$ when $n = 20$, respectively by the robust saddlepoint test corresponding to $\gamma = -1$ when $n = 50, 100$.

In Fig. 4 we report the actual levels of the tests when the data are generated from $0.9\mathcal{N}(0, 1) + 0.1\mathcal{N}(0, 2)$ and the sample size is $n = 20$.

In Figs. 2, 3 and 5 we report the relative errors of the tests based on robust saddlepoint test statistics, when the data are generated from the model $\mathcal{N}(0, 1)$, respectively from the perturbed models $0.95\mathcal{N}(0, 1) + 0.05\mathcal{N}(0, 2)$ and $0.9\mathcal{N}(0, 1) + 0.1\mathcal{N}(0, 2)$. In each case the sample size is $n = 20$. When the data are not contaminated, the approximation of the level is very good for all the considered divergences. For contaminated data, the results for the proposed test statistics are slightly modified when comparing with the results for the noncontaminated data, the best being obtained for $\gamma = -0.5$.

5.2. *The Cauchy location model*

Another Monte Carlo experiment involves the Cauchy location model $Cau(\theta)$. The null hypothesis is $H_0 : \theta = 0$.

First, data are generated from the model $Cau(0)$ and then from the perturbed distributions $(1 - \varepsilon)Cau(0) + \varepsilon Cau(1)$, where $\varepsilon = 0.05, 0.1$. The sample size is in all cases $n = 20$.

According to the results from Section 4.3, we choose $\alpha = -1$ and Cressie–Read divergences corresponding to $\gamma \in \{-2, -1.5, 0, 0.5\}$. For these choices we expect to achieve robustness and small sample accuracy for the testing procedure. For each of the mentioned divergences, we compute the dual divergence estimator of the parameter $\theta = 0$ (as solution of the Eq. (10) with the ψ-function (21)) and the corresponding test statistic.

In Figs. 6 and 7 we represent the actual levels of the tests, respectively the relative errors, when the data are generated from $Cau(0)$. The approximation of the level is very good for all the considered divergences.

In the second case, we consider data corresponding to the contaminated model $0.95Cau(0) + 0.05Cau(1)$. With this perturbation, the results for the proposed test statistics are slightly modified if compared with the results for the noncontaminated data. This can be seen from Figs. 8 and 9, the best results being obtained for $\gamma = 0.5$ (the Hellinger divergence).

In the third case, the data are generated from $0.9Cau(0) + 0.1Cau(1)$. As it can be inferred from Figs. 10 and 11, despite of larger deviations from the model, the results are still good, the best being again obtained for the Hellinger divergence.

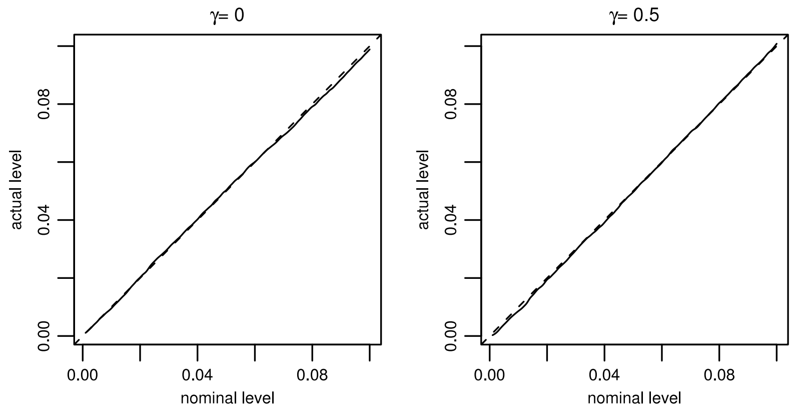


Fig. 6. Comparison of the actual level and the nominal level of the robust tests applied to the model $Cau(0)$, when the data are generated from the model, the tuning parameter is $\alpha = -1$ and $\gamma = 0, 0.5$.

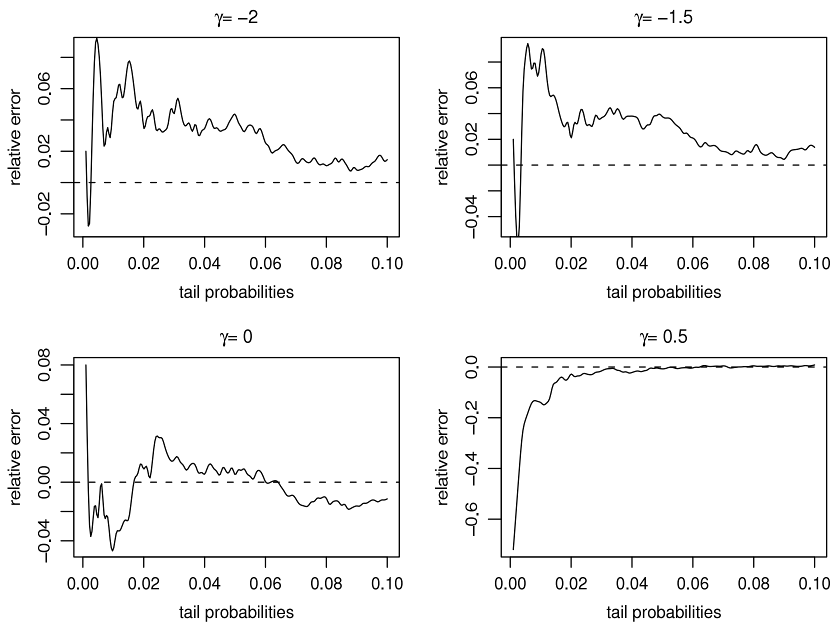


Fig. 7. Relative errors of the robust tests applied to the model $Cau(0)$, when the data are generated from the model, the tuning parameter is $\alpha = -1$ and $\gamma = -2, -1.5, 0, 0.5$.

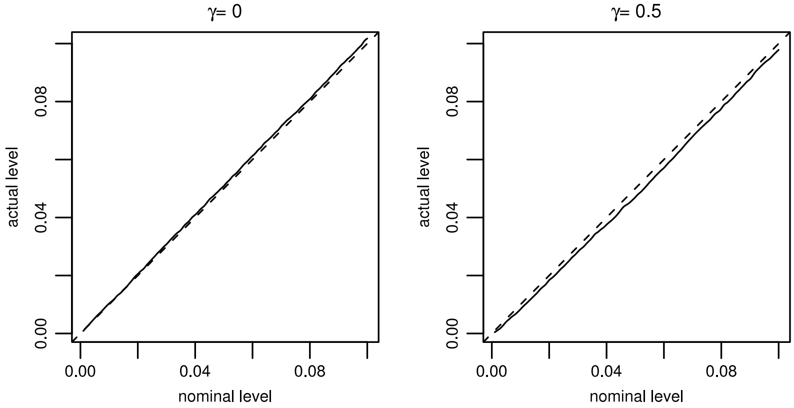


Fig. 8. Comparison of the actual level and the nominal level of the robust tests applied to the model $Cau(0)$, when the data are generated from $0.95Cau(0) + 0.05Cau(1)$, the tuning parameter is $\alpha = -1$ and $\gamma = 0, 0.5$.

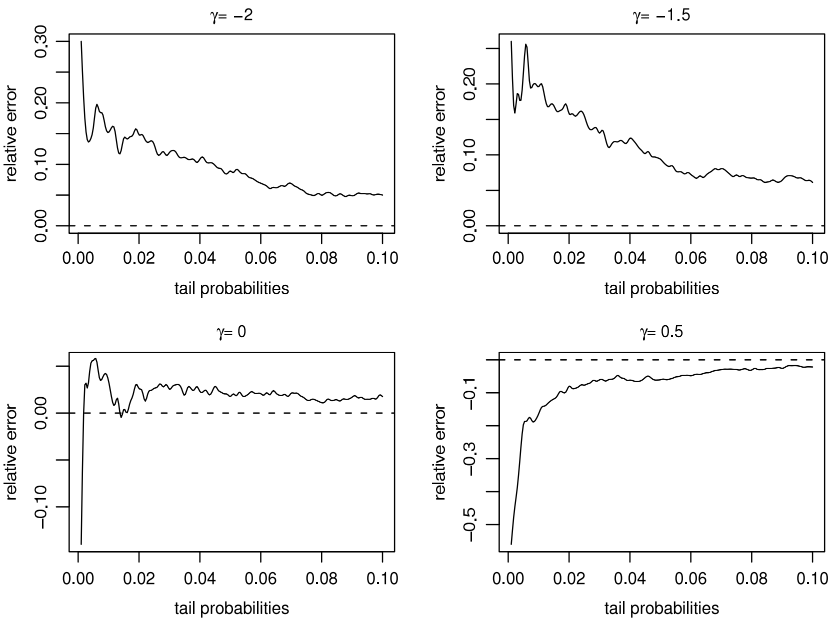


Fig. 9. Relative errors of the robust tests applied to the model $Cau(0)$, when the data are generated from $0.95Cau(0) + 0.05Cau(1)$, the tuning parameter is $\alpha = -1$ and $\gamma = -2, -1.5, 0, 0.5$.

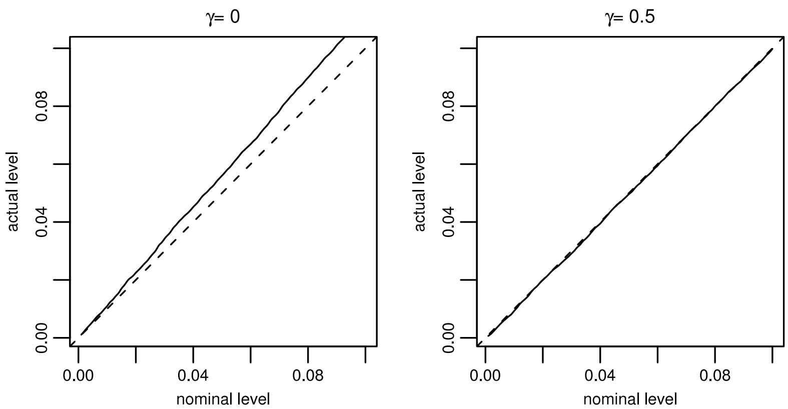


Fig. 10. Comparison of the actual level and the nominal level of the robust tests applied to the model $Cau(0)$, when the data are generated from $0.9Cau(0) + 0.1Cau(1)$, the tuning parameter is $\alpha = -1$ and $\gamma = 0, 0.5$.

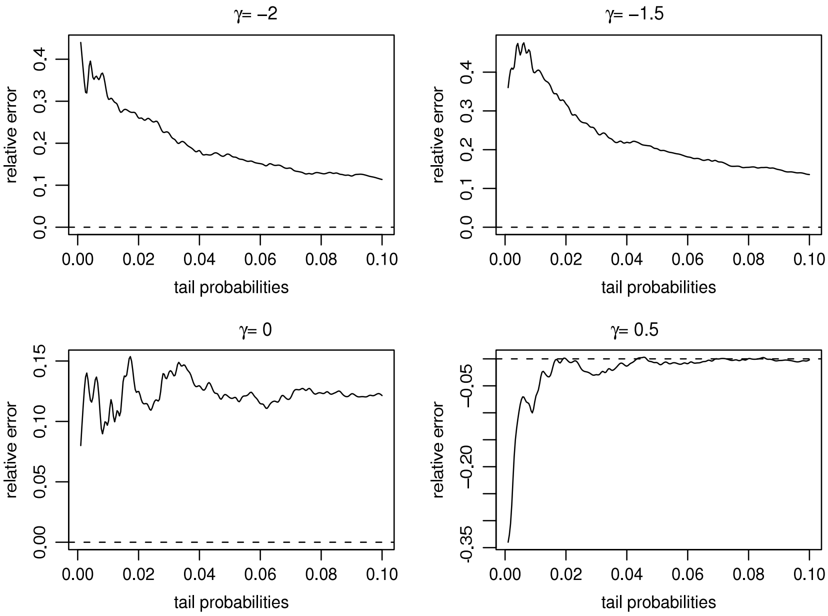


Fig. 11. Relative errors of the robust tests applied to the model $Cau(0)$, when the data are generated from $0.9Cau(0) + 0.1Cau(1)$, the tuning parameter is $\alpha = -1$ and $\gamma = -2, -1.5, 0, 0.5$.

Thus, the numerical results show that the proposed tests are stable in the presence of small deviations from the underlying model and, in the meantime, very accurate in small samples.

Acknowledgments

The authors are grateful to Professor E. Ronchetti for the useful discussion on saddlepoint test statistics and to the referees for a careful reading of the paper which led to an improved version.

References

[1] M. Broniatowski, A. Keziou, Parametric estimation and tests through divergences and duality technique, J. Multivariate Anal. 100 (1) (2009) 16–36.
[2] A. Toma, M. Broniatowski, Minimum divergence estimators and tests: Robustness results, J. Multivariate Anal. (under revision), http://www.lsta.upmc.fr.
[3] J. Robinson, E. Ronchetti, G.A. Young, Saddlepoint approximations and tests based on multivariate M-estimates, Ann. Statist. 31 (4) (2003) 1154–1169.
[4] C.A. Field, E. Ronchetti, Small Sample Asymptotics, IMS, Hayward, 1990.
[5] J.L. Jensen, Saddlepoint Approximations, Oxford University Press, 1995.
[6] C.A. Field, Small sample asymptotic expansion for multivariate M-estimates, Ann. Statist. 10 (3) (1982) 672–689.
[7] C.A. Field, F.R. Hampel, Small-sample asymptotic distributions of M-estimators of location, Biometrika 69 (1) (1982) 29–46.
[8] A. Almudevar, C. Field, J. Robinson, The density of multivariate M-estimates, Ann. Statist. 28 (1) (2000) 275–297.
[9] C.A. Field, J. Robinson, E. Ronchetti, Saddlepoint approximations for multivariate M-estimates with applications to bootstrap accuracy, Ann. Inst. Statist. Math. 60 (1) (2008) 205–224.
[10] L. Rüschendorf, On the minimum discrimination information theorem, Statist. Decisions (Suppl 1) (1984) 163–283.
[11] N. Cressie, T.R.C. Read, Multinomial goodness-of-fit tests, J. R. Statist. Soc. Ser. B 46 (3) (1984) 440–464.
[12] M. Broniatowski, A. Keziou, Minimization of ϕ-divergences on sets of signed measures, Studia Sci. Math. Hungar. 43 (4) (2006) 403–442.
[13] I.M. Skovgaard, On the density of minimum contrast estimators, Ann. Statist. 18 (2) (1990) 779–789.
[14] E. Ronchetti, L. Ventura, Between stability and higher-order asymptotics, Statist. Comput. 11 (1) (2001) 67–73.
[15] R.A. Maronna, R.D. Martin, V.J. Yohai, Robust Statistics — Theory and Methods, Wiley, New York, 2006.