ROBUST BOUNDED INFLUENCE TESTS FOR INDEPENDENT NON-HOMOGENEOUS OBSERVATIONS

Abhik Ghosh and Ayanendranath Basu

Indian Statistical Institute

Abstract: Experiments often yield non-identically distributed data for statistical analysis. Tests of hypothesis under such set-ups are generally performed using the likelihood ratio test, which is non-robust with respect to outliers and model misspecification. In this paper, we consider the set-up of non-identically but independently distributed observations and develop a general class of test statistics for testing parametric hypothesis based on the density power divergence. The proposed tests have bounded influence functions, are highly robust with respect to data contamination, have high power against contiguous alternatives, and are consistent at any fixed alternative. The methodology is illustrated by the simple and generalized linear regression models with fixed covariates.

Key words and phrases: Generalized linear model, influence function, linear regression, non-homogeneous observation, robust testing of hypothesis.

1. Introduction

One of the most important paradigms of statistical inference is hypothesis testing; arguably the most common test is the likelihood ratio test (LRT). However, like the maximum likelihood estimator (MLE), the LRT may lead to unstable inference in the presence of outliers. Attempts to rectify this (Simpson (1989); Lindsay (1994); Basu et al. (2013a,b)) have mostly been in the context of independent and identically distributed (i.i.d.) data. In this paper, we consider the general case of non-identically distributed data. Suppose the observed data $Y_1, \ldots, Y_n$ are independent, but for each i, $Y_i \sim g_i$ with $g_1, \ldots, g_n$ being possibly different densities with respect to some common dominating measure. We model g_i by the family $\mathcal{F}_{i,\boldsymbol{\theta}} = \{f_i(\cdot; \boldsymbol{\theta}) | \boldsymbol{\theta} \in \Theta\}$ for all $i = 1, 2, \ldots, n$. Let G_i and $F_i(\cdot, \boldsymbol{\theta})$ be their respective distribution functions. Even though the Y_is have possibly different densities, they share the common parameter $\boldsymbol{\theta}$. We will refer to this set-up as the independent non-homogeneous (I-NH) set-up.

The most prominent application of the I-NH set-up is the regression model with non-stochastic covariates, where f_i is a known density depending on the

given predictors $\boldsymbol{x}_i$, error distribution, and a common regression parameter $\boldsymbol{\beta}$, $y_i \sim f_i(\cdot, \boldsymbol{x}_i, \boldsymbol{\beta})$. This differs from the usual regression set-up with stochastic covariates that has been explored in greater detail in the literature (Ronchetti and Rousseeuw (1980); Schrader and Hettmansperger (1980); Ronchetti (1982a,b, 1987); Sen (1982); Markatou and Hettmansperger (1990); Markatou and He (1994); Markatou and Manos (1996); Cantoni and Ronchetti (2001); Liu, Markatou, and Tsai (2005); Maronna, Martin and Yohai (2006); Wang and Qu (2007); Hosseinian (2009); Salibian-Barrera, Aelst and Yohai (2016)). Our set-up treats the regression problem from a design point of view where we pre-fix the covariate levels. The robustness literature under this general I-NH set-up is limited; some scattered attempts have been made in particular cases like normal regression (Huber (1983); Muller (1998)).

Ghosh and Basu (2013) proposed a global approach for estimating $\boldsymbol{\theta}$ under the I-NH set-up by minimizing the average density power divergence (DPD) measure (originally introduced by Basu et al. (1998) for i.i.d. data) between the data and the model density; the proposed minimum DPD estimator (MDPDE) has excellent efficiency and robustness properties in the normal regression model. The approach is also implemented in the context of generalized linear models by Ghosh and Basu (2016); it provides a competitive alternative to existing robust methods. The approach has been used in Ghosh (2017) to obtain a robust alternative for tail index estimation under suitable assumptions of an exponential regression model. Here, we exploit the properties of the Ghosh and Basu (2013) estimator to develop a general class of robust tests of hypotheses for I-NH data.

The specific advantages of the proposed methods are as follows. (1) The method is completely general in that it works for any set-up involving independent and non-homogeneous data. (2) The proposal is simple to implement with minimal addition in computational complexity compared to likelihood based methods. (3) The testing procedure is based on the minimization of a bona-fide objective function and the selection of the proper root of the estimating equation is simple as it must correspond to the global minimum. (4) Our methods have bounded influence for the test statistics, and the bounded level and power influence functions. (5) The proposed tests are consistent at any fixed alternative, and have high power against any contiguous alternative.

In this paper, we assume Conditions (A1)–(A7) of Ghosh and Basu (2013), which we refer to as the "Ghosh-Basu conditions", and Assumptions A, B, C and D of Lehmann (1983, p. 429), which we refer to as the "Lehmann conditions". These conditions are listed in Section S1 of the Online Supplement for completeness.

2. The MDPDE under the I-NH Set-up

Under the I-NH set-up, Ghosh and Basu (2013) proposed the estimation of $\boldsymbol{\theta}$ by minimizing the average DPD measure between the data and the model, or equivalently

$$H_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left[\int f_i(y;\boldsymbol{\theta})^{1+\tau}dy - \left(1+\frac{1}{\tau}\right)f_i(Y_i;\boldsymbol{\theta})^{\tau}\right] = \frac{1}{n}\sum_{i=1}^{n}V_i(Y_i;\boldsymbol{\theta}). \quad (2.1)$$

The corresponding estimating equation is given by

$$\sum_{i=1}^{n}\left[f_i(Y_i;\boldsymbol{\theta})^{\tau}\boldsymbol{u}_i(Y_i;\boldsymbol{\theta}) - \int f_i(y;\boldsymbol{\theta})^{1+\tau}\boldsymbol{u}_i(y;\boldsymbol{\theta})dy\right] = 0, \quad (2.2)$$

where ∇ represents the gradient with respect to $\boldsymbol{\theta}$, and $\boldsymbol{u}_i(y;\boldsymbol{\theta}) = \nabla\ln f_i(y;\boldsymbol{\theta})$ is the likelihood score function for i-th model density (Similarly, ∇^2 will represent the second order derivative with respect to $\boldsymbol{\theta}$). When $\tau = 0$, the MDPDE is seen to coincide with the non-robust maximum likelihood estimator (MLE); as τ increases the robustness increases significantly at the cost of a slight loss in asymptotic efficiency.

With $\underline{\boldsymbol{G}} = (G_1, \cdots, G_n)$, the minimum DPD functional $\boldsymbol{\theta}^g = \boldsymbol{U}_\tau(\underline{\boldsymbol{G}})$ for the I-NH observations is defined by

$$\frac{1}{n}\sum_{i=1}^{n}d_\tau(g_i(.), f_i(.; \boldsymbol{U}_\tau(\underline{\boldsymbol{G}}))) = \min_{\boldsymbol{\theta}\in\Theta}\frac{1}{n}\sum_{i=1}^{n}d_\tau(g_i(.), f_i(.; \boldsymbol{\theta})), \quad (2.3)$$

where $d_\tau(f_1, f_2)$ denotes the DPD measure between two densities f_1 and f_2 with the tuning parameter τ, as given by Basu et al. (1998),

$$d_\tau(f_1, f_2) = \begin{cases} \int\left[f_2^{1+\tau} - \left(1+\frac{1}{\tau}\right)f_2^{\tau}f_1 + \frac{1}{\tau}f_1^{1+\tau}\right], & \text{for } \tau > 0, \\ \int f_1\log(\frac{f_1}{f_2}), & \text{for } \tau = 0. \end{cases} \quad (2.4)$$

Equivalently, $\boldsymbol{U}_\tau(\underline{\boldsymbol{G}})$ is the minimizer of $1/n\sum_{i=1}^{n}H^{(i)}(\boldsymbol{\theta})$, with respect to $\boldsymbol{\theta}\in\Theta$, where $H^{(i)}(\boldsymbol{\theta}) = \int f_i(y;\boldsymbol{\theta})^{1+\tau}dy - (1+1/\tau)\int f_i(y;\boldsymbol{\theta})^{\tau}g_i(y)dy$.

Ghosh and Basu (2013) derived the asymptotic distribution of the MDPDE $\hat{\boldsymbol{\theta}}_n$, under this set-up. Under the Ghosh-Basu conditions, we have the following.

(i) There exists a consistent sequence $\hat{\boldsymbol{\theta}}_n$ of roots of (2.2).

(ii) The asymptotic distribution of $\boldsymbol{\Omega}_n^{\tau}(\boldsymbol{\theta}^g)^{-1/2}\boldsymbol{\Psi}_n^{\tau}(\boldsymbol{\theta}^g)[\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}^g)]$ is p-dimensional normal with (vector) mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{I}_p$, the p-dimensional identity matrix, where $\boldsymbol{\Psi}_n^{\tau}(\boldsymbol{\theta}^g) = 1/n\sum_{i=1}^{n}\boldsymbol{J}^{(i)}(\boldsymbol{\theta}^g)$, with

$$\boldsymbol{J}^{(i)}(\boldsymbol{\theta}^g) = \int \boldsymbol{u}_i(y;\boldsymbol{\theta}^g)\boldsymbol{u}_i^{\ T}(y;\boldsymbol{\theta}^g)f_i^{1+\tau}(y;\boldsymbol{\theta}^g)dy$$

$$- \int \{\nabla \boldsymbol{u}_i(y;\boldsymbol{\theta}^g) + \tau \boldsymbol{u}_i(y;\boldsymbol{\theta}^g)\boldsymbol{u}_i^{\ T}(y;\boldsymbol{\theta}^g)\}\{g_i(y) - f_i(y;\boldsymbol{\theta}^g)\}f_i(y;\boldsymbol{\theta}^g)^{\tau}dy,$$

and

$$\boldsymbol{\Omega}_n^{\tau}(\boldsymbol{\theta}^g) = \frac{1}{n}\sum_{i=1}^{n} \left[\int \boldsymbol{u}_i(y;\boldsymbol{\theta}^g)\boldsymbol{u}_i^{\ T}(y;\boldsymbol{\theta}^g)f_i(y;\boldsymbol{\theta}^g)^{2\tau}g_i(y)dy - \boldsymbol{\xi}_i\boldsymbol{\xi}_i^T\right],$$

with

$$\boldsymbol{\xi}_i = \int \boldsymbol{u}_i(y;\boldsymbol{\theta}^g)f_i(y;\boldsymbol{\theta}^g)^{\tau}g_i(y)dy. \tag{2.5}$$

3. Testing Simple Hypothesis

We start with the simple hypothesis testing problem with a fully specified null under the I-NH set-up. Let $\boldsymbol{\theta}_0$ be a fixed point in the parameter space Θ. We want to test

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{against} \quad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0. \tag{3.1}$$

When the model is correctly specified and the null hypothesis is correct, $f_i(\cdot;\boldsymbol{\theta}_0)$ is the data generating density for the i-th observation. We can test for the hypothesis in (3.1) by using the DPD measure between $f_i(\cdot;\boldsymbol{\theta}_0)$ and $f_i(\cdot;\widehat{\boldsymbol{\theta}})$ for any estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. We consider the MDPDE $\boldsymbol{\theta}_n^{\tau}$ of $\boldsymbol{\theta}$ as defined in Section 2. Since there are n divergence measures corresponding to each i, we consider the total divergence measure over the n data points for testing (3.1) and define the DPD based test statistic (DPDTS) as

$$T_{\gamma}(\boldsymbol{\theta}_n^{\tau}, \boldsymbol{\theta}_0) = 2\sum_{i=1}^{n} d_{\gamma}(f_i(.;\boldsymbol{\theta}_n^{\tau}), f_i(.;\boldsymbol{\theta}_0)),$$

where $d_{\gamma}(f_1, f_2)$ is defined in (2.4). In case of i.i.d. data, this DPDTS coincides with the test statistic in Basu et al. (2013a).

3.1. Asymptotic properties

Consider the matrices $\boldsymbol{\Psi}_n^{\tau}$ and $\boldsymbol{\Omega}_n^{\tau}$ defined in Section 2 and let $\boldsymbol{A}_n^{\gamma}(\boldsymbol{\theta}) = 1/n\sum_{i=1}^{n} \boldsymbol{A}_{\gamma}^{(i)}(\boldsymbol{\theta})$ with $\boldsymbol{A}_{\gamma}^{(i)}(\boldsymbol{\theta}_0) = \nabla^2 d_{\gamma}(f_i(.;\boldsymbol{\theta}), f_i(.;\boldsymbol{\theta}_0))|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$. For some $p \times p$ matrices $\boldsymbol{J}_{\tau}$, $\boldsymbol{V}_{\tau}$, A_{τ}, and $\boldsymbol{\theta} \in \Theta$, consider the assumptions.

(C1) $\boldsymbol{\Psi}_n^{\tau}(\boldsymbol{\theta}) \to \boldsymbol{J}_{\tau}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}_n^{\tau}(\boldsymbol{\theta}) \to \boldsymbol{V}_{\tau}(\boldsymbol{\theta})$ element-wise as $n \to \infty$

(C2) $\boldsymbol{A}_n^{\gamma}(\boldsymbol{\theta}) \to \boldsymbol{A}_{\gamma}(\boldsymbol{\theta})$ element-wise as $n \to \infty$.

Theorem 1. *Suppose the model density satisfies the Lehmann and Ghosh-Basu*

conditions and conditions (C1) *and* (C2) *hold at* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. *Then, the asymptotic null distribution of the DPDTS* $T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0)$ *is the distribution of* $\sum_{i=1}^r \zeta_i^{\gamma,\tau}(\boldsymbol{\theta}_0) Z_i^2$, *where* $Z_1, \cdots, Z_r$ *are independent standard normal variables and* $\zeta_1^{\gamma,\tau}(\boldsymbol{\theta}_0), \cdots,$ $\zeta_r^{\gamma,\tau}(\boldsymbol{\theta}_0)$ *are the nonzero eigenvalues of* $\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_\tau(\boldsymbol{\theta}_0)$ *with* $\boldsymbol{\Sigma}_\tau(\boldsymbol{\theta}) = \boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta})$ $\boldsymbol{V}_\tau(\boldsymbol{\theta})\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta})$ *and* $r = rank(\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0)\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0)\boldsymbol{V}_\tau(\boldsymbol{\theta}_0))$.

The null distribution of the proposed DPDTS has the same form as that in Basu et al. (2013a,b) for i.i.d. observations. The critical region of our proposal can be easily determined from the relevant discussion in Basu et al. (2013a,b).

Next we present an approximation to its power function. Let $\boldsymbol{M}_n^\gamma(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^n \boldsymbol{M}_\gamma^{(i)}(\boldsymbol{\theta})$ with $\boldsymbol{M}_\gamma^{(i)}(\boldsymbol{\theta}) = \nabla d_\gamma(f_i(.;\boldsymbol{\theta}), f_i(.;\boldsymbol{\theta}_0))$, and assume

(C3) $\boldsymbol{M}_n^\gamma(\boldsymbol{\theta}) \to \boldsymbol{M}_\gamma(\boldsymbol{\theta})$ element-wise as $n \to \infty$ for some p-vector $\boldsymbol{M}_\gamma(\boldsymbol{\theta})$.

Theorem 2. *Suppose the model density satisfies the Lehmann and Ghosh-Basu conditions and take any* $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$ *in* Θ *for which* (C1) *and* (C3) *hold. Then, an approximation to the power function of the test* $\{T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0) > t_\alpha^{\tau,\gamma}\}$ *for testing the hypothesis in* (3.1) *at the significance level* α *is given by*

$$\pi_{n,\alpha}^{\tau,\gamma}(\boldsymbol{\theta}^*) = 1 - \Phi\left(\frac{1}{\sqrt{n}\sigma_{\tau,\gamma}(\boldsymbol{\theta}^*)}\left(\frac{t_\alpha^{\tau,\gamma}}{2} - \sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}^*), f_i(.;\boldsymbol{\theta}_0))\right)\right),$$

where $t_\alpha^{\tau,\gamma}$ *is the* $(1-\alpha)$-*th quantile of the asymptotic null distribution of* $T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0)$ *and* $\sigma_{\tau,\gamma}^2(\boldsymbol{\theta}) = \boldsymbol{M}_\gamma(\boldsymbol{\theta})^T\boldsymbol{\Sigma}_\tau(\boldsymbol{\theta})\boldsymbol{M}_\gamma(\boldsymbol{\theta})$.

Corollary 1. *For any* $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$, *the probability of rejecting the null hypothesis* H_0 *at any fixed significance level* $\alpha > 0$ *with the rejection rule* $\{T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0) > t_\alpha^{\tau,\gamma}\}$ *tends to 1 as* $n \to \infty$, *provided* $1/n\sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}^*), f_i(.;\boldsymbol{\theta}_0)) = O(1)$.

Theorem 2 can be used to obtain the sample size required to achieve a pre-specified power η. For this we just need to solve the equation

$$\eta = 1 - \Phi\left(\frac{1}{\sqrt{n}\sigma_{\tau,\gamma}(\boldsymbol{\theta}^*)}\left(\frac{t_\alpha^{\tau,\gamma}}{2} - \sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}^*), f_i(.;\boldsymbol{\theta}_0))\right)\right)$$

in terms of n. If n^* denotes the solution, then the required sample size is the least integer greater than or equal to n^*.

3.2. Robustness properties

3.2.1. Influence functions of the test statistics

We illustrate the robustness of the proposed DPDTS using some extensions of Hampel's influence function (IF), as in Huber (1983) and Ghosh and Basu (2013).

Ignoring the multiplier 2 in the DPDTS, we consider the functional

$$T_{\gamma,\tau}^{(1)}(\underline{\mathbf{G}}) = \sum_{i=1}^{n} d_{\gamma}(f_i(\cdot; \boldsymbol{U}_{\tau}(\underline{\mathbf{G}})), f_i(\cdot; \boldsymbol{\theta}_0)),$$

where $\underline{\mathbf{G}} = (G_1, \cdots, G_n)$ and $\boldsymbol{U}_{\tau}(\underline{\mathbf{G}})$ is the minimum DPD functional under the I-NH set-up. Unlike the i.i.d. case, here the functional depends on the sample size n, and so does the corresponding IF. We refer to it as the fixed-sample IF. Consider the contaminated distribution $G_{i,\epsilon} = (1 - \epsilon)G_i + \epsilon \wedge_{t_i}$, where $\wedge_{t_i}$ is the degenerate distribution at the point of contamination t_i in the i-th direction for all $i = 1, \ldots, n$. Thus, we can have contamination in some fixed direction or in all directions.

Consider a contamination only in the i_0-th direction and take $\underline{\mathbf{G}}_{i_0,\epsilon} = (G_1, \cdots, G_{i_0-1}, G_{i_0,\epsilon}, \cdots, G_n)$. Then the corresponding first order IF of the test functional $T_{\gamma,\tau}^{(1)}(\underline{\mathbf{G}})$ can be defined as

$$IF_{i_0}(t_{i_0}, T_{\gamma,\tau}^{(1)}, \underline{\mathbf{G}}) = \frac{\partial}{\partial \epsilon} T_{\gamma,\tau}^{(1)}(\underline{\mathbf{G}}_{i_0,\epsilon})\Big|_{\epsilon=0} = \sum_{i=1}^{n} \boldsymbol{M}_{\gamma}^{(i)}(\boldsymbol{U}_{\tau}(\underline{\mathbf{G}}))^T IF_{i_0}(t_{i_0}, \boldsymbol{U}_{\tau}, \underline{\mathbf{G}}),$$

where $IF_{i_0}(t_{i_0}, \boldsymbol{U}_{\tau}, \underline{\mathbf{G}})$ is the IF of $\boldsymbol{U}_{\tau}$ in Ghosh and Basu (2013),

$$IF_{i_0}(t_{i_0}, \boldsymbol{U}_{\tau}, \underline{\mathbf{G}}) = \frac{1}{n} \boldsymbol{\Psi}_n^{\tau}(\boldsymbol{\theta}^g)^{-1} \boldsymbol{D}_{\tau,i_0}(t_{i_0}; \boldsymbol{\theta}^g), \tag{3.2}$$

where $\boldsymbol{D}_{\tau,i}(t; \boldsymbol{\theta}) = [f_i(t; \boldsymbol{\theta})^{\tau} \boldsymbol{u}_i(t; \boldsymbol{\theta}) - \boldsymbol{\xi}_i]$ with $\boldsymbol{\xi}_i$ defined at (2.5). In general, the IF of a test is evaluated at the null distribution $G_i(\cdot) = F_i(\cdot, \boldsymbol{\theta}_0)$ for all i. Letting $\underline{\mathbf{F}}_{\boldsymbol{\theta}_0} = (F_1(\cdot, \boldsymbol{\theta}_0), \cdots, F_n(\cdot, \boldsymbol{\theta}_0))$, we get $\boldsymbol{U}_{\tau}(\underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = \boldsymbol{\theta}_0$ and $\boldsymbol{M}_{\gamma}^{(i)}(\boldsymbol{\theta}_0) = \mathbf{0}$ so that Hampel's first-order IF of the DPDTS is zero at H_0.

The second order IF of the DPDTS can be defined similarly as

$$IF_{i_0}^{(2)}(t_{i_0}, T_{\gamma,\tau}^{(1)}, \underline{\mathbf{G}}) = \frac{\partial^2}{\partial^2 \epsilon} T_{\gamma,\tau}^{(1)}(G_1, \cdots, G_{i_0-1}, G_{i_0,\epsilon}, \cdots, G_n)\Big|_{\epsilon=0}.$$

In particular, at the null distribution $\underline{\mathbf{G}} = \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}$, it simplifies to

$$IF_{i_0}^{(2)}(t_{i_0}, T_{\gamma,\tau}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = n \cdot IF_{i_0}(t_{i_0}, \boldsymbol{U}_{\tau}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})^T \boldsymbol{A}_n^{\gamma}(\boldsymbol{\theta}_0) IF_{i_0}(t_{i_0}, \boldsymbol{U}_{\tau}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}).$$

Thus the IF of the test at the null is bounded for any fixed sample size if and only if the IF of the corresponding minimum DPD functional is bounded. Using the form of the IF of the MDPDE from (3.2), the IF of the test is

$$IF_{i_0}^{(2)}(t_{i_0}, T_{\gamma,\tau}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = \frac{1}{n} \boldsymbol{D}_{\tau,i_0}(t_{i_0}; \boldsymbol{\theta}_0)^T [\boldsymbol{\Psi}_n^{\tau}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{A}_n^{\gamma}(\boldsymbol{\theta}_0) \boldsymbol{\Psi}_n^{\tau}(\boldsymbol{\theta}_0)^{-1}] \boldsymbol{D}_{\tau,i_0}(t_{i_0}; \boldsymbol{\theta}_0).$$

Here $\boldsymbol{D}_{\tau,i}(t; \boldsymbol{\theta})$ is bounded in t if the parametric model satisfies $f_i(t; \boldsymbol{\theta})^{\tau} \boldsymbol{u}_i(t; \boldsymbol{\theta})$ is bounded in t. For most parametric models, this holds at $\tau > 0$ (but not at $\tau = 0$) implying that the $\boldsymbol{D}_{\tau,i}(t; \boldsymbol{\theta})$, and therefore the IF, is bounded whenever

$\tau > 0$, but unbounded at $\tau = 0$. Note that $\boldsymbol{D}_{\tau,i}(t; \boldsymbol{\theta})$ does not depend on the tuning parameter γ, hence its boundedness and that of the IF of the proposed test is independent of the choice of γ.

If we consider contamination in all directions at the contamination point $\mathbf{t} = (t_1, \cdots, t_n)$, we can derive the corresponding IF of the proposed DPDTS in a similar manner. Again, at the null distribution, its first order IF is zero and its second order IF simplifies to

$$IF^{(2)}(\mathbf{t}, T_{\gamma,\tau}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = n \cdot IF(\mathbf{t}, \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})^T A_n^\gamma IF(\mathbf{t}, \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})$$

$$= \frac{1}{n} \left(\sum_{i=1}^n \boldsymbol{D}_{\tau,i}(t_i; \boldsymbol{\theta}_0) \right)^T [\boldsymbol{\Psi}_n^\tau(\boldsymbol{\theta}_0)^{-1} A_n^\gamma(\boldsymbol{\theta}_0) \boldsymbol{\Psi}_n^\tau(\boldsymbol{\theta}_0)^{-1}] \left(\sum_{i=1}^n \boldsymbol{D}_{\tau,i}(t_i; \boldsymbol{\theta}_0) \right).$$

This influence function is bounded for most parametric models when $\tau > 0$ and unbounded if $\tau = 0$. Thus, whatever be the contamination direction, the proposed DPDTS is always robust for $\tau > 0$ and non-robust for $\tau = 0$. Here, robustness refers to local robustness of the test statistics under infinitesimal contamination.

3.2.2. Level and power influence functions

The performance of any testing procedure is generally measured by its level and power. We consider the effect of contamination on level and power of the proposed DPDTS through the level and power influence functions (Hampel et al. (1986); Heritier and Ronchetti (1994); Toma and Broniatowski (2010)). Since the exact level and power of the proposed test are difficult to obtain, we work with their asymptotic versions.

Since the proposed DPDTS is consistent, we examine its asymptotic power under the contiguous alternative $H_{1,n} : \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2} \boldsymbol{\Delta}$ with $\boldsymbol{\Delta} \in \mathbb{R}^p - \{0\}$. Here we consider contamination over these alternatives. As argued in Hampel et al. (1986), we consider contaminations such that their effects tend to zero as $\boldsymbol{\theta}_n$ tends to $\boldsymbol{\theta}_0$ at the same rate to avoid the confusion between the null and alternative neighborhoods. Consider the contaminated distributions

$$\underline{\mathbf{F}}_{n,\epsilon,\mathbf{t}}^L = \left(1 - \frac{\epsilon}{\sqrt{n}} \right) \underline{\mathbf{F}}_{\boldsymbol{\theta}_0} + \frac{\epsilon}{\sqrt{n}} \wedge_{\mathbf{t}}, \text{ and } \underline{\mathbf{F}}_{n,\epsilon,\mathbf{t}}^P = \left(1 - \frac{\epsilon}{\sqrt{n}} \right) \underline{\mathbf{F}}_{\boldsymbol{\theta}_n} + \frac{\epsilon}{\sqrt{n}} \wedge_{\mathbf{t}},$$

for the level and power, respectively, where $\mathbf{t} = (t_1, \cdots, t_n)^T$, $\mathbf{F}_{n,\epsilon,\mathbf{t}}^P = (F_{i,n,\epsilon,t_i}^P)_{i=1,\cdots,n}$ and $\underline{\mathbf{F}}_{n,\epsilon,\mathbf{t}}^L = (F_{i,n,\epsilon,t_i}^L)_{i=1,\cdots,n}$. Then the level influence function (LIF) and the power influence function (PIF) are

$$LIF(\mathbf{t}; T_\gamma^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = \lim_{n \to \infty} \frac{\partial}{\partial \epsilon} P_{\underline{\mathbf{F}}_{n,\epsilon,\mathbf{t}}^L}(T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0) > t_\alpha^{\tau,\gamma})\big|_{\epsilon=0},$$

$$PIF(\mathbf{t}; T_\gamma^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = \lim_{n \to \infty} \frac{\partial}{\partial \epsilon} P_{\underline{\mathbf{F}}_{n,\epsilon,\mathbf{t}}^P}(T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0) > t_\alpha^{\tau,\gamma})\big|_{\epsilon=0}.$$

We start with the asymptotic power under the contaminated distribution $\underline{\mathbf{F}}_{n,\epsilon,\mathbf{y}}^P$ and examine some special cases by substituting specific values of $\boldsymbol{\Delta}$ and ϵ.

Theorem 3. *Suppose that the Lehmann and Ghosh-Basu conditions hold for the model density and (C1)-(C2) hold at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Then for any $\boldsymbol{\Delta} \in \mathbb{R}^p$ and $\epsilon \geq 0$, we have the following.*

(i) The asymptotic distribution of the proposed DPDTS under $\underline{\mathbf{F}}_{n,\epsilon,\mathbf{t}}^P$ is the distribution of the quadratic form $\boldsymbol{W}^T \boldsymbol{A}_\gamma(\boldsymbol{\theta}_0) \boldsymbol{W}$, where $\boldsymbol{W} \sim N_p(\widetilde{\boldsymbol{\Delta}}, \boldsymbol{\Sigma}_\tau(\boldsymbol{\theta}_0))$ with $\widetilde{\boldsymbol{\Delta}} = \left[\boldsymbol{\Delta} + \epsilon IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})\right]$ and $\boldsymbol{\Sigma}_\tau$ defined as in Theorem 1.

(ii) The asymptotic power of the proposed DPDTS under $\underline{\mathbf{F}}_{n,\epsilon,\mathbf{t}}^P$ is given by

$$P_{\tau,\gamma}(\boldsymbol{\Delta}, \epsilon; \alpha) = \lim_{n \to \infty} P_{\underline{\mathbf{F}}_{n,\epsilon,\mathbf{t}}^L}(T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0) > t_\alpha^{\tau,\gamma}),$$

$$= \sum_{v=0}^\infty C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \widetilde{\boldsymbol{\Delta}}) P\left(\chi_{r+2v}^2 > \frac{t_\alpha^{\tau,\gamma}}{\zeta_{(1)}^{\gamma,\tau}(\boldsymbol{\theta}_0)}\right), \qquad (3.3)$$

where χ_p^2 denotes a chi-square random variable with p degrees of freedom, $\zeta_{(1)}^{\gamma,\tau}(\boldsymbol{\theta}_0)$ is the minimum of $\zeta_i^{\gamma,\tau}(\boldsymbol{\theta}_0)$s for $i = 1, \ldots, r$ defined in Theorem 1, and

$$C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \widetilde{\boldsymbol{\Delta}}) = \frac{1}{v!} \left(\prod_{j=1}^r \frac{\zeta_{(1)}^{\gamma,\tau}(\boldsymbol{\theta}_0)}{\zeta_j^{\gamma,\tau}(\boldsymbol{\theta}_0)}\right)^{1/2} e^{-1/2 \sum_{j=1}^r \delta_j} E\left[\left(\widehat{Q}\right)^v\right],$$

with

$$\widehat{Q} = \frac{1}{2} \sum_{j=1}^r \left[\left(1 - \frac{\zeta_{(1)}^{\gamma,\tau}(\boldsymbol{\theta}_0)}{\zeta_j^{\gamma,\tau}(\boldsymbol{\theta}_0)}\right)^{1/2} Z_j + \sqrt{\delta_j} \left(\frac{\zeta_{(1)}^{\gamma,\tau}(\boldsymbol{\theta}_0)}{\zeta_j^{\gamma,\tau}(\boldsymbol{\theta}_0)}\right)^{1/2}\right]^2,$$

for r independent standard normal random variables $Z_1, \ldots, Z_r$, and δ_is defined as in Remark 1 below.

Corollary 2. *The asymptotic power under the contiguous alternatives $H_{1,n}$: $\boldsymbol{\theta} = \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\Delta}$ is*

$$P_{\tau,\gamma}(\boldsymbol{\Delta}, 0; \alpha) = \sum_{v=0}^\infty C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \boldsymbol{\Delta}) P\left(\chi_{r+2v}^2 > \frac{t_\alpha^{\tau,\gamma}}{\zeta_{(1)}^{\gamma,\tau}(\boldsymbol{\theta}_0)}\right).$$

Corollary 3. *The asymptotic level under the probability distribution $\underline{\mathbf{F}}_{n,\epsilon,\mathbf{t}}^L$ is*

$$\alpha_\epsilon = P_{\tau,\gamma}(\mathbf{0}, \epsilon; \alpha) = \sum_{v=0}^\infty C_v^{\gamma,\tau}\left(\boldsymbol{\theta}_0, \epsilon IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})\right) P\left(\chi_{r+2v}^2 > \frac{t_\alpha^{\tau,\gamma}}{\zeta_{(1)}^{\gamma,\tau}(\boldsymbol{\theta}_0)}\right).$$

Remark 1. The asymptotic distribution of $T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0)$ under $\underline{\mathbf{F}}_{n,\epsilon,\mathbf{t}}^P$ is also the same as that of $\sum_{i=1}^r \zeta_i^{\gamma,\tau}(\boldsymbol{\theta}_0) \chi_{1,\delta_i}^2$, where the $\zeta_i^{\gamma,\tau}(\boldsymbol{\theta}_0)$ are given in Theorem 1 and

χ^2_{1,δ_i}s are independent non-central chi-square variables having degree of freedom one and non-centrality parameters δ_is, respectively, with $\left(\sqrt{\delta_1}, \ldots, \sqrt{\delta_p}\right)^T = \widetilde{\boldsymbol{P}}_{\tau,\gamma}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_\tau^{-1/2}(\boldsymbol{\theta}_0)\widetilde{\boldsymbol{\Delta}}$ and $\widetilde{\boldsymbol{P}}_{\tau,\gamma}(\boldsymbol{\theta}_0)$ being the matrix of normalized eigenvectors of $\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_\tau(\boldsymbol{\theta}_0)$.

Remark 2. The expressions of asymptotic level and power under contiguous alternative with contamination can be approximated by truncating the series to a finite number (N) terms. The error incurred by such a truncation can be made smaller than any pre-specific limit by choosing N suitably large. If we truncate at the N-th term of (3.3), assuming $c_v = C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \widetilde{\boldsymbol{\Delta}})$, the error can be bounded by

$$e_N = \sum_{v=N+1}^{\infty} c_v \cdot P\left(\chi^2_{r+2v} > \frac{t_\alpha^{\tau,\gamma}}{\zeta_{(1)}^{\gamma,\tau}(\boldsymbol{\theta}_0)}\right) \leq \sum_{v=N+1}^{\infty} c_v = 1 - \sum_{v=0}^{N} c_v.$$

See Kotz, Johnson, and Boyd (1967a,b) for more accurate error bounds for such approximations.

Starting with $P_{\tau,\gamma}(\boldsymbol{\Delta}, \epsilon; \alpha)$ at (3.3) and differentiating, we get the power influence function $PIF(\cdot)$.

Theorem 4. *Assume that the Lehmann and Ghosh-Basu conditions hold for the model density and (C1)-(C2) hold at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. If the influence function $IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})$ of the MDPDE is bounded, then for any $\boldsymbol{\Delta} \in \mathbb{R}^p$, the power influence function of the proposed DPDTS is given by*

$$PIF(\mathbf{t}; T_{\gamma,\lambda}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})^T \boldsymbol{K}_{\gamma,\tau}(\boldsymbol{\theta}_0, \boldsymbol{\Delta}, \alpha),$$

where $\boldsymbol{K}_{\gamma,\tau}(\boldsymbol{\theta}_0, \boldsymbol{\Delta}, \alpha) = (\sum_{v=0}^{\infty}[(\partial C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \boldsymbol{d}))/(\partial \boldsymbol{d})|_{\boldsymbol{d}=\boldsymbol{\Delta}}]P(\chi^2_{r+2v} > t_\alpha^{\tau,\gamma}/(\zeta_{(1)}^{\gamma,\tau}(\boldsymbol{\theta}_0)))).$

The level influence function of the proposed DPDTS can be derived by putting $\boldsymbol{\Delta} = \boldsymbol{0}$ in this expression of the PIF, yielding $LIF(\mathbf{t}; T_{\gamma,\lambda}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})^T \boldsymbol{K}_{\gamma,\tau}(\boldsymbol{\theta}_0, \boldsymbol{0}, \alpha)$ whenever the IF of the MDPDE used is bounded. Thus asymptotically the level and power of the DPDTS are unaffected by contiguous contaminations for all $\tau > 0$.

4. Testing Composite Hypothesis

In the I-NH set-up of Section 1, take a fixed (proper) subspace Θ_0 of Θ. Based on the observed data, we want to test the hypothesis

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad \text{against} \quad H_1 : \boldsymbol{\theta} \notin \Theta_0. \tag{4.1}$$

When the model is correctly specified and H_0 is correct, $f_i(\cdot; \boldsymbol{\theta}_0)$ is the data generating density for the i-th observation, where $\boldsymbol{\theta}_0 \in \Theta_0$. We can test this

hypothesis by using the DPD measure between $f_i(\cdot; \widetilde{\boldsymbol{\theta}})$ and $f_i(\cdot; \widehat{\boldsymbol{\theta}})$ for any two estimators $\widetilde{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ under H_0 and $H_0 \cup H_1$, respectively. For $\widehat{\boldsymbol{\theta}}$, we take the MDPDE $\boldsymbol{\theta}_n^\tau$ of $\boldsymbol{\theta}$ with tuning parameter τ, and for $\widetilde{\boldsymbol{\theta}}$, we take the estimator $\widetilde{\boldsymbol{\theta}}_n^\tau$ obtained by minimizing the DPD with tuning parameter τ over the subspace Θ_0 only; we refer to $\widetilde{\boldsymbol{\theta}}_n^\tau$ as the restricted MDPDE (RMDPDE). Thus, our test statistic (DPDTS$_C$) for the hypothesis (4.1) based on the DPD with parameter γ is

$$S_\gamma(\boldsymbol{\theta}_n^\tau, \widetilde{\boldsymbol{\theta}}_n^\tau) = 2 \sum_{i=1}^n d_\gamma(f_i(.; \boldsymbol{\theta}_n^\tau), f_i(.; \widetilde{\boldsymbol{\theta}}_n^\tau)). \tag{4.2}$$

4.1. Properties of the RMDPDE under the I-NH set-up

The restricted minimum density power divergence estimators (RMDPDE) $\widetilde{\boldsymbol{\theta}}_n^\tau$ of $\boldsymbol{\theta}$ is the minimizer of the DPD objective function $H_n(\boldsymbol{\theta})$ given at (2.1) with tuning parameter τ subject to a set of r restrictions of the form $\boldsymbol{v}(\boldsymbol{\theta}) = \mathbf{0}$, where $\boldsymbol{v} : \mathbb{R}^p \mapsto \mathbb{R}^r$ is some vector valued function. For the null hypothesis in (4.1), such restrictions are given by the definition of the null parameter space Θ_0. We assume that the $p \times r$ matrix $\boldsymbol{\Upsilon}(\boldsymbol{\theta}) = (\partial \boldsymbol{v}(\boldsymbol{\theta}))/(\partial \boldsymbol{\theta})$ exists and is continuous in $\boldsymbol{\theta}$ with rank r. Then, the RMDPDE has to satisfy

$$\nabla H_n(\boldsymbol{\theta}) + \boldsymbol{\Upsilon}(\boldsymbol{\theta})\boldsymbol{\lambda}_n = 0, \qquad \boldsymbol{v}(\boldsymbol{\theta}) = 0, \tag{4.3}$$

where $\boldsymbol{\lambda}_n$ is an r-vector of Lagrangian multipliers. Further, the restricted minimum DPD functional $\widetilde{\boldsymbol{\theta}}^g = \widetilde{U}_\tau(\mathbf{G})$ at the true distribution $\mathbf{G}$ is the minimizer of $n^{-1} \sum_{i=1}^n d_\tau(g_i(.), f_i(.; \boldsymbol{\theta}))$ subject to $\boldsymbol{v}(\boldsymbol{\theta}) = \mathbf{0}$.

Theorem 5. *Assume that the Ghosh-Basu conditions are satisfied with respect to Θ_0 (instead of Θ). Then the following hold.*
(i) There exists a consistent sequence $\widetilde{\boldsymbol{\theta}}_n^\tau$ of roots of (4.3).
(ii) Asymptotically, $\boldsymbol{\Omega}_n^\tau(\widetilde{\boldsymbol{\theta}}^g)^{-1/2} \boldsymbol{P}_n^\tau(\widetilde{\boldsymbol{\theta}}^g)^{-1}[\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g)] \sim N_p(\mathbf{0}, \boldsymbol{I}_p)$ where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix, $\boldsymbol{\Upsilon}_n^*(\boldsymbol{\theta}) = \boldsymbol{\Upsilon}(\boldsymbol{\theta})^T [\nabla^2 H_n(\boldsymbol{\theta})]^{-1} \boldsymbol{\Upsilon}(\boldsymbol{\theta})$, and

$$\boldsymbol{P}_n^\tau(\boldsymbol{\theta}) = \left[\frac{\nabla^2 H_n(\boldsymbol{\theta})}{(1+\tau)} \right]^{-1} \left[\boldsymbol{I}_p - \boldsymbol{\Upsilon}(\boldsymbol{\theta}) \left[\boldsymbol{\Upsilon}_n^*(\boldsymbol{\theta}) \right]^{-1} \boldsymbol{\Upsilon}(\boldsymbol{\theta})^T [\nabla^2 H_n(\boldsymbol{\theta})]^{-1} \right].$$

In the following, we need a further assumption.

(C4) $\boldsymbol{P}_n^\tau(\widetilde{\boldsymbol{\theta}}^g) \to \boldsymbol{P}_\tau(\widetilde{\boldsymbol{\theta}}^g)$ ($p \times p$ invertible) element-wise as $n \to \infty$.

Corollary 4. *If the assumptions of Theorem 5 hold, and (C1) and (C4) hold at $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}^g$, then, asymptotically, $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g) \sim N_p(\mathbf{0}, \boldsymbol{P}_\tau(\widetilde{\boldsymbol{\theta}}^g) \boldsymbol{V}_\tau(\widetilde{\boldsymbol{\theta}}^g) \boldsymbol{P}_\tau(\widetilde{\boldsymbol{\theta}}^g)).$*

Next, we explore the robustness properties of the RMDPDEs in terms of

their influence function. First consider the contamination in only one (i_0-th) direction. Suppose the given restrictions are such that they can be substituted explicitly in the expression of average DPD before taking its derivative with respect to $\boldsymbol{\theta}$; then the final derivative should be zero at $\boldsymbol{\theta} = \widetilde{\boldsymbol{U}}_\tau(\underline{\mathbf{G}}_{i_0,\epsilon})$ and $g_{i_0} = g_{i_0,\epsilon}$, the density corresponding to $G_{i_0,\epsilon}$. Standard differentiation of the resulting equation with respect to ϵ at $\epsilon = 0$ yields the IF of the RMDPDE, $IF_{i_0}(t_{i_0}; \widetilde{\boldsymbol{U}}_\tau; \underline{\mathbf{G}}) = (\partial \widetilde{\boldsymbol{U}}_\tau)/(\partial \epsilon)(\underline{\mathbf{G}}_{i_0,\epsilon})\big|_{\epsilon=0}$ as a solution of

$$\boldsymbol{\Psi}_n^{(0)}(\widetilde{\boldsymbol{\theta}^g})IF_{i_0}(t_{i_0}, \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{G}}) - \frac{1}{n}\boldsymbol{D}_{\tau,i_0}^{(0)}(t_{i_0}; \widetilde{\boldsymbol{\theta}^g}) = 0, \tag{4.4}$$

where $\boldsymbol{D}_{\tau,i}^{(0)}(t; \boldsymbol{\theta}) = [f_i(t; \boldsymbol{\theta})^\tau \boldsymbol{u}_i^{(0)}(t; \boldsymbol{\theta}) - \boldsymbol{\xi}_i^{(0)}(\boldsymbol{\theta})]$ and $\boldsymbol{\Psi}_n^{(0)}(\boldsymbol{\theta})$, $\boldsymbol{\xi}_i^{(0)}(\boldsymbol{\theta})$, $\boldsymbol{u}_i^{(0)}(y; \boldsymbol{\theta})$ are the same as $\boldsymbol{\Psi}_n(\boldsymbol{\theta})$, $\boldsymbol{\xi}_i(\boldsymbol{\theta})$, $\boldsymbol{u}_i(y; \boldsymbol{\theta})$ respectively, under the additional restriction $\boldsymbol{v}(\boldsymbol{\theta}) = 0$. Also, $\widetilde{\boldsymbol{U}}_\tau(\underline{\mathbf{G}}_{i_0,\epsilon})$ must satisfy $\boldsymbol{v}(\widetilde{\boldsymbol{U}}_\tau(\underline{\mathbf{G}}_{i_0,\epsilon})) = \boldsymbol{0}$, from which we get

$$\boldsymbol{\Upsilon}(\widetilde{\boldsymbol{\theta}^g})^T IF_{i_0}(t_{i_0}, \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{G}}) = \boldsymbol{0}. \tag{4.5}$$

Solving (4.4) and (4.5), (as done for the i.i.d. case in Ghosh (2015)), we get a general expression for the IF of the RMDPDE as

$$IF_{i_0}(t_{i_0}, \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{G}}) = \frac{1}{n}\boldsymbol{Q}(\widetilde{\boldsymbol{\theta}^g})^{-1}\boldsymbol{\Psi}_n^{(0)}(\widetilde{\boldsymbol{\theta}^g})^T \boldsymbol{D}_{\tau,i_0}^{(0)}(t_{i_0}; \widetilde{\boldsymbol{\theta}^g}), \tag{4.6}$$

where $\boldsymbol{Q}(\boldsymbol{\theta}) = [\boldsymbol{\Psi}_n^{(0)}(\boldsymbol{\theta})^T \boldsymbol{\Psi}_n^{(0)}(\boldsymbol{\theta}) + \boldsymbol{\Upsilon}(\boldsymbol{\theta})\boldsymbol{\Upsilon}(\boldsymbol{\theta})^T]$. Clearly, this IF is bounded in t_{i_0} whenever $f_{i_0}(t_{i_0}; \widetilde{\boldsymbol{\theta}^g})^\tau \boldsymbol{u}_{i_0}^{(0)}(t_{i_0}; \widetilde{\boldsymbol{\theta}^g})$ is bounded, and this is the case at $\tau > 0$ for most parametric models and common parametric restrictions.

If the contamination is in all the directions at the points $\mathbf{t} = (t_1, \cdots, t_n)$, the IF of the RMDPDE is given by

$$IF_o(\mathbf{t}; \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{G}}) = \boldsymbol{Q}(\widetilde{\boldsymbol{\theta}^g})^{-1}\boldsymbol{\Psi}_n^{(0)}(\widetilde{\boldsymbol{\theta}^g})^T \left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{D}_{\tau,i}^{(0)}(t_i; \widetilde{\boldsymbol{\theta}^g})\right].$$

4.2. Asymptotic properties of the proposed test

Assume that Θ_0 is a proper subset of the parameter space Θ which can be defined in terms of r restrictions $\boldsymbol{v}(\boldsymbol{\theta}) = 0$ such that the $p \times r$ matrix $\boldsymbol{\Upsilon}(\boldsymbol{\theta}) = (\partial \boldsymbol{v}(\boldsymbol{\theta}))/(\partial \boldsymbol{\theta})$ exists and is a continuous function of $\boldsymbol{\theta}$ with rank r.

Theorem 6. *Suppose the model density satisfies the Lehmann and Ghosh-Basu conditions, H_0 is true with $\boldsymbol{\theta}_0 \in \Theta_0$ being the true parameter value, and (C1), (C2) and (C4) hold at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. With $\widetilde{\boldsymbol{\Sigma}}_\tau(\boldsymbol{\theta}_0) = [\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)]\boldsymbol{V}_\tau(\boldsymbol{\theta}_0) [\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)]$, the asymptotic null distribution of the DPDTS$_C$ $S_\gamma(\boldsymbol{\theta}_n^\tau, \widetilde{\boldsymbol{\theta}}_n^\tau)$ is the distribution of $\sum_{i=1}^r \widetilde{\zeta_i^{\gamma,\tau}}(\boldsymbol{\theta}_0)Z_i^2$, where $r = rank(\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)[\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)]$*

$\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)[\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)]\boldsymbol{V}_\tau(\boldsymbol{\theta}_0))$, $Z_1, \cdots, Z_r$ *are independent standard normals and* $\widetilde{\zeta_1^{\gamma,\tau}}(\boldsymbol{\theta}_0)$, ..., $\widetilde{\zeta_r^{\gamma,\tau}}(\boldsymbol{\theta}_0)$ *are the nonzero eigenvalues of* $\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\widetilde{\boldsymbol{\Sigma}}_\tau(\boldsymbol{\theta}_0)$.

We can find approximate critical values of the asymptotic null distribution from the discussions in Basu et al. (2013a,b). Next, we derive an asymptotic power approximation of the proposed DPDTS$_C$ at any point $\boldsymbol{\theta}^* \notin \Theta_0$, which can be used to determine minimum sample size requirement to attain any desired power as in the case of a simple hypothesis. If $\boldsymbol{\theta}^* \notin \Theta_0$ is the true parameter value, then $\boldsymbol{\theta}_n^\tau \xrightarrow{\mathcal{P}} \boldsymbol{\theta}^*$ and $\widetilde{\boldsymbol{\theta}}_n^\tau \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$ for some $\boldsymbol{\theta}_0 \in \Theta_0$ and $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$. Then, assuming the Lehman conditions and Ghosh-Basu conditions along with (C1) and (C4) at $\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\theta}^*$, we can show that

$$\sqrt{n}\left(\begin{array}{c} \boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}^* \\ \widetilde{\boldsymbol{\theta}}_n^\tau - \boldsymbol{\theta}_0 \end{array}\right) \to N_{2p}\left(\left[\begin{array}{c} \boldsymbol{0} \\ \boldsymbol{0} \end{array}\right], \left[\begin{array}{cc} \boldsymbol{\Sigma}_\tau(\boldsymbol{\theta}^*) & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{12}^T & \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)\boldsymbol{P}_\tau(\boldsymbol{\theta}_0) \end{array}\right]\right),$$

for a $p \times p$ matrix $\boldsymbol{A}_{12} = \boldsymbol{A}_{12}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)$. Take $\boldsymbol{M}_{1,\gamma}^{(i)}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) = \nabla d_\gamma(f_i(.;\boldsymbol{\theta}), f_i(.;\boldsymbol{\theta}_0))|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ and $\boldsymbol{M}_{2,\gamma}^{(i)}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) = \nabla d_\gamma(f_i(.;\boldsymbol{\theta}^*), f_i(.;\boldsymbol{\theta}))|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$.

(C5) $\boldsymbol{M}_n^{j,\gamma}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) = n^{-1}\sum_{i=1}^n \boldsymbol{M}_{j,\gamma}^{(i)}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) \to \boldsymbol{M}_{j,\gamma}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)$ element-wise as $n \to \infty$ for some p-vectors $\boldsymbol{M}_{j,\gamma}$ $(j = 1, 2)$.

Theorem 7. *Suppose the model density satisfies the Lehmann and Ghosh-Basu conditions and* $\boldsymbol{\theta}^* \notin \Theta_0$ *for which* (C1), (C4), *and* (C5) *hold. Then, an approximation to the power function of the DPDTS$_C$ for testing* (4.1) *at the significance level* α *is given by*

$$\pi_{n,\alpha}^{\tau,\gamma}(\boldsymbol{\theta}^*) = 1 - \Phi\left(\frac{1}{\sqrt{n}\sigma_{\tau,\gamma}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0)}\left(\frac{s_\alpha^{\tau,\gamma}}{2} - \sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}^*), f_i(.;\boldsymbol{\theta}_0))\right)\right),$$

where $s_\alpha^{\tau,\gamma}$ *is* $(1-\alpha)$-*th quantile of the asymptotic null distribution of* $S_\gamma(\boldsymbol{\theta}_n^\tau, \widetilde{\boldsymbol{\theta}}_n^\tau)$,

$\sigma_{\tau,\gamma}^2(\boldsymbol{\theta}^*, \boldsymbol{\theta}_0) = \boldsymbol{M}_{1,\gamma}^T \boldsymbol{\Sigma}_\tau \boldsymbol{M}_{1,\gamma} + \boldsymbol{M}_{1,\gamma}^T \boldsymbol{A}_{12} \boldsymbol{M}_{2,\gamma} + \boldsymbol{M}_{2,\gamma}^T \boldsymbol{A}_{12}^T \boldsymbol{M}_{1,\gamma} + \boldsymbol{M}_{2,\gamma}^T \boldsymbol{P}_\tau \boldsymbol{V}_\tau \boldsymbol{P}_\tau \boldsymbol{M}_{2,\gamma}$.

Corollary 5. *For* $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$, *the probability of rejecting* H_0 *in* (4.1) *at level* $\alpha > 0$ *based on the DPDTS$_C$ tends to* 1 *as* $n \to \infty$, *provided* $1/n\sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}^*), f_i(.;\boldsymbol{\theta}_0)) = O(1)$.

4.3. Robustness properties of the test

Using the functional form of $\boldsymbol{\theta}_n^\tau$ and $\widetilde{\boldsymbol{\theta}}_n^\tau$ and ignoring the multiplier 2 in our test statistic, the functional corresponding to the DPDTS$_C$ is

$$S_{\gamma,\tau}^{(1)}(\mathbf{G}) = \sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{U}_\tau(\mathbf{G})), f_i(.;\widetilde{\boldsymbol{U}}_\tau(\mathbf{G}))).$$

Clearly, this depends on the sample size n, implying the same dependency in its IF. Consider the contaminated distribution $G_{i,\epsilon}$ defined in Section 3.2.1 and assume contamination in only one fixed direction-i_0. The first order IF of $S_{\gamma,\tau}^{(1)}(\underline{\mathbf{G}})$ is

$$IF_{i_0}(t_{i_0}, S_{\gamma,\tau}^{(1)}, \underline{\mathbf{G}}) = \frac{\partial}{\partial\epsilon}S_{\gamma,\tau}^{(1)}(\underline{\mathbf{G}}_{i_0,\epsilon})\big|_{\epsilon=0} = n\boldsymbol{M}_n^{1,\gamma}(\boldsymbol{U}_\tau(\underline{\mathbf{G}}), \widetilde{\boldsymbol{U}}_\tau(\underline{\mathbf{G}}))^T IF_{i_0}(t_{i_0}, \boldsymbol{U}_\tau, \underline{\mathbf{G}})$$
$$+ n\boldsymbol{M}_n^{2,\gamma}(\boldsymbol{U}_\tau(\underline{\mathbf{G}}), \widetilde{\boldsymbol{U}}_\tau(\underline{\mathbf{G}}))^T IF_{i_0}(t_{i_0}, \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{G}}),$$

where $IF_{i_0}(t_{i_0}, \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{G}})$ is the IF of the RMDPD functional $\widetilde{\boldsymbol{U}}_\tau$ under H_0. If the null hypothesis is true with $\underline{\mathbf{G}} = \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}$ for some $\boldsymbol{\theta}_0 \in \Theta_0$, then $\boldsymbol{U}_\tau(\underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = \widetilde{\boldsymbol{U}}_\tau(\underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = \boldsymbol{\theta}_0$ and $\boldsymbol{M}_{j,\gamma}^{(i)}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) = \mathbf{0}$ for $j = 1, 2$. Hence Hampel's first-order IF of the DPDTS$_C$ is again zero at the composite null.

Similarly, at $\underline{\mathbf{G}} = \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}$, the second order IF of the DPDTS$_C$ functional $S_{\gamma,\tau}^{(1)}$ is

$$IF_{i_0}^{(2)}(t_{i_0}, S_{\gamma,\tau}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = n\boldsymbol{D}_{\tau,i_0}(t_{i_0}, \boldsymbol{\theta}_0)^T \boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_0)\boldsymbol{D}_{\tau,i_0}(t_{i_0}, \boldsymbol{\theta}_0),$$

where $\boldsymbol{D}_{\tau,i_0}(t_{i_0}, \boldsymbol{\theta}_0) = [IF_{i_0}(t_{i_0}, \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) - IF_{i_0}(t_{i_0}, \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})]$. Clearly, this IF is bounded if the corresponding MDPDEs over Θ_0 and Θ both have bounded IFs. However, the boundedness of the IF of the MDPDE over Θ implies the same under any restricted subspace Θ_0 and this holds for most parametric models if $\tau > 0$, but the IF is unbounded at $\tau = 0$.

Next, considering the contamination in all directions at $\mathbf{t} = (t_1, \ldots, t_n)$, the first order IF of the proposed DPDTS$_C$ is again zero at any point inside Θ_0 and its second order IF at the null is given by

$$IF_o^{(2)}(\mathbf{t}, T_{\gamma,\tau}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = n\boldsymbol{D}_{\tau,o}(\mathbf{t}, \boldsymbol{\theta}_0)^T \boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_0)\boldsymbol{D}_{\tau,o}(\mathbf{t}, \boldsymbol{\theta}_0),$$

where $\boldsymbol{D}_{\tau,o}(\mathbf{t}, \boldsymbol{\theta}_0) = [IF_o(\mathbf{t}, \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) - IF_o(\mathbf{t}, \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})]$. This implies the robustness for $\tau > 0$.

The level and power influence functions of the proposed test for this case are similar to that in the simple hypothesis case (Section 3.2.2).

5. Application (I): Normal Linear Regression

Consider the linear regression model $y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \epsilon_i$, for $i = 1, \ldots, n$, where the error ϵ_i's are assumed to be i.i.d. normal with mean zero and variance σ^2; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ and $\boldsymbol{x}_i^T = (x_{i1}, \ldots, x_{i,p})$ denote the regression coefficients and the i-th observation for the covariates, respectively. Here, we assume $\boldsymbol{x}_i$ to be fixed so that $y_i \sim N(\boldsymbol{x}_i^T\boldsymbol{\beta}, \sigma^2)$ for each i. Clearly the y_i's are independent but not identically distributed. The MDPDEs of $\boldsymbol{\beta}$ and σ^2 and their properties are described in Section S2.1 of the Online Supplement.

5.1. Testing for the regression coefficients with known σ

The simple hypothesis on the regression coefficient $\boldsymbol{\beta}(=\boldsymbol{\theta})$ assuming the error variance σ^2 to be known, say $\sigma^2 = \sigma_0^2$, is

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0, \qquad \text{against} \qquad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0, \tag{5.1}$$

for some pre-specified $\boldsymbol{\beta}_0(=\boldsymbol{\theta}_0)$.

Consider the test statistics $T_\gamma(\boldsymbol{\beta}_n^\tau, \boldsymbol{\beta}_0)$ for testing (5.1), where $\boldsymbol{\beta}_n^\tau$ is the MD-PDE of $\boldsymbol{\beta}$ with tuning parameter τ and known $\sigma = \sigma_0$. Using the normal density, we get

$$T_\gamma(\boldsymbol{\beta}_n^\tau, \boldsymbol{\beta}_0) = \frac{2\sqrt{1+\gamma}}{\gamma(\sqrt{2\pi}\sigma_0)^\gamma} \left[n - \sum_{i=1}^n e^{-(\gamma(\boldsymbol{\beta}_n^\tau - \boldsymbol{\beta}_0)^T(\boldsymbol{x}_i \boldsymbol{x}_i^T)(\boldsymbol{\beta}_n^\tau - \boldsymbol{\beta}_0))/(2(\gamma(\sigma_n^\tau)^2 + \sigma_0^2))} \right],$$

if $\gamma > 0$, and $T_0(\boldsymbol{\beta}_n^\tau, \boldsymbol{\beta}_0) = \dfrac{(\boldsymbol{\beta}_n^\tau - \boldsymbol{\beta}_0)^T(\boldsymbol{X}^T\boldsymbol{X})(\boldsymbol{\beta}_n^\tau - \boldsymbol{\beta}_0)}{\sigma_0^2},$

where $\boldsymbol{X} = [\boldsymbol{x}_1 \cdots \boldsymbol{x}_n]^T$. The estimator $\boldsymbol{\beta}_n^{(0)}$, the MDPDE with $\tau = 0$, is indeed the MLE of $\boldsymbol{\beta}$. The usual LRT statistics for this problem is $-2\log[(\prod_{i=1}^n \phi(y_i, \boldsymbol{x}_i^T\boldsymbol{\beta}_0, \sigma_0))/(\prod_{i=1}^n \phi(y_i; \boldsymbol{x}_i^T\boldsymbol{\beta}_n^{(0)}, \sigma_0))]$; after simplification, this statistic is the same as $T_0(\boldsymbol{\beta}_n^{(0)}, \boldsymbol{\beta}_0)$. Hence the proposed test is nothing but a robust generalization of the likelihood ratio test. Here $\phi(\cdot, \mu, \sigma)$ refers to the $N(\mu, \sigma^2)$ density.

5.1.1. Asymptotic properties

Suppose that the conditions (R1)–(R2) of Ghosh and Basu (2013) hold true and assume

(C6) $1/n(\boldsymbol{X}^T\boldsymbol{X})$ converges point-wise to some positive definite matrix $\boldsymbol{\Sigma}_x$ as $n \to \infty$.

Then, the corresponding limiting matrices simplify to $\boldsymbol{J}_\tau(\boldsymbol{\beta}_0) = \zeta_\tau\boldsymbol{\Sigma}_x$, $\boldsymbol{V}_\tau(\boldsymbol{\beta}_0) = \zeta_{2\tau}\boldsymbol{\Sigma}_x$ and $\boldsymbol{A}_\gamma(\boldsymbol{\beta}_0) = (1+\gamma)\zeta_\gamma\boldsymbol{\Sigma}_x$, where $\zeta_\tau = (2\pi)^{-\tau/2}\sigma^{-(\tau+2)}(1+\tau)^{-3/2}$.

Theorem 1 gives the asymptotic null distribution of $T_\gamma(\boldsymbol{\beta}_n^\tau, \boldsymbol{\beta}_0)$ under $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, which turns out to be a scalar multiple of a χ_p^2 distribution with the multiplier $\zeta_1^{\gamma,\tau} = (\sqrt{2\pi}\sigma_0)^{-\gamma}(1+\gamma)^{-1/2}(1+\tau^2/(1+2\tau))^{3/2}$. The critical region for testing (5.1) at the significance level α is given by $\left\{ T_\gamma(\boldsymbol{\beta}_n^\tau, \boldsymbol{\beta}_0) > \zeta_1^{\gamma,\tau}\chi_{p,\alpha}^2 \right\}$, where $\chi_{p,\alpha}^2$ is the $(1-\alpha)$-th quantile of the χ_p^2 distribution. At $\gamma = \tau = 0$, we have $\zeta_1^{0,0} = 1$ so that $T_0(\boldsymbol{\theta}_n^{(0)}, \boldsymbol{\theta}_0)$ asymptotically follows a χ_p^2 distribution under H_0, as expected from its relation to the LRT.

Next we study the asymptotic power of the proposed test. We derive its asymptotic power under the contiguous alternatives $H_{1,n}$ using Corollary 2. The asymptotic distribution of $T_\gamma(\boldsymbol{\beta}_n^\tau, \boldsymbol{\beta}_0)$ under $H_{1,n}$ is $\zeta_1^{\gamma,\tau}\chi_{p,\delta}^2$ with $\delta = 1/v_\tau^{\boldsymbol{\beta}}\boldsymbol{\Delta}^T\boldsymbol{\Sigma}_x$

(a) $\mathbf{\Delta}^{\mathrm{T}}\Sigma_x\mathbf{\Delta} = 5$ (b) $\mathbf{\Delta}^{\mathrm{T}}\Sigma_x\mathbf{\Delta} = 20$
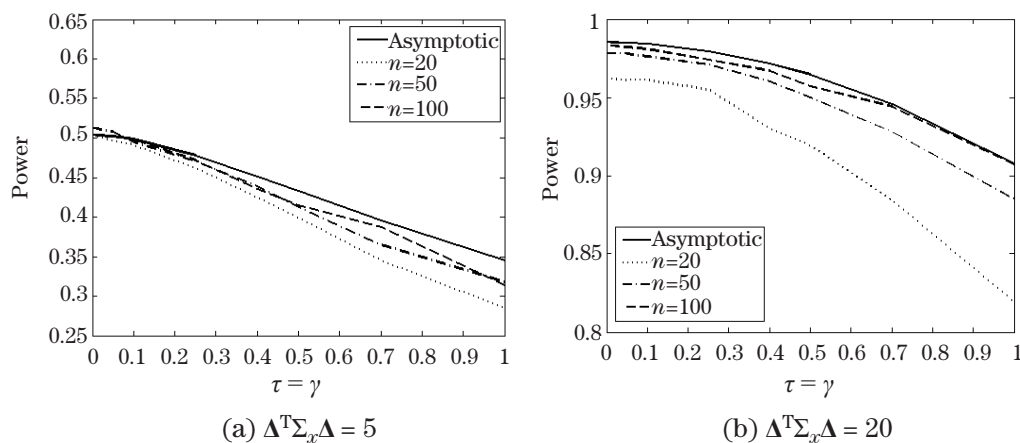
Figure 1. Asymptotic contiguous power of the simple DPD based test of $\boldsymbol{\beta}$ for different values of $t = \mathbf{\Delta}^T\Sigma_x\mathbf{\Delta}$ and p, the number of explanatory variables.

$\mathbf{\Delta}$, where $v_\tau^{\boldsymbol{\beta}} = \sigma^2(1 + \tau^2/(1 + 2\tau))^{3/2}$. Thus its asymptotic contiguous power is $P_{\tau,\gamma}(\mathbf{\Delta}, 0; \alpha) = P(\zeta_1^{\gamma,\tau} W_{p,\delta} > \zeta_1^{\gamma,\tau} \chi_{p,\alpha}^2) = 1 - G_{p,\delta}(\chi_{p,\alpha}^2)$, where $G_{p,\delta}$ is the distribution function of $\chi_{p,\delta}^2$. Figure 1 shows the asymptotic power over the tuning parameters $\gamma = \tau$ for different values of $\mathbf{\Delta}^T\Sigma_x\mathbf{\Delta}(= t,$ say); it does not depend on the tuning parameter γ. The power depends on the distance ($\mathbf{\Delta}$) of the contiguous alternatives from null and the limiting second order moments (Σ_x) of the covariates through $t = \mathbf{\Delta}^T\Sigma_x\mathbf{\Delta}$; for any fixed $\tau = \gamma$ it increases as the value of t increases. It also depends on the number (p) of explanatory variables used. In Figure 1, we show the cases of p is 2 and 10 as well as $p = 50$ and 200. For fixed values of $\gamma = \tau$ and $\mathbf{\Delta}_1^T\Sigma_x\mathbf{\Delta}_1$, the power decreases as p increases; this is expected as the number of components of $\boldsymbol{\beta}$ increases with p. The power against any contiguous alternative and any model is seen to decrease slightly with increasing τ which brings in the non-centrality parameter δ and hence the asymptotic variance $v_\tau^{\boldsymbol{\beta}}$ of each element of $(\mathbf{X}^T\mathbf{X})^{1/2}\boldsymbol{\beta}_n^\tau$. As $v_\tau^{\boldsymbol{\beta}}$ increases slightly with τ, the efficiency of the MDPDE and the asymptotic contiguous power of the DPDTS decrease slightly.

We simulated finite-sample situations with different sample sizes n and values of $t = \mathbf{\Delta}^T\Sigma_x\mathbf{\Delta}$ and p. The convergence of the finite-sample power to the asymptotic value depends on the convergence rate of $1/n(\mathbf{X}^T\mathbf{X})$ in Condition (C6); we chose X to be a p-variate normal distribution with mean 0 and covariance matrix $\sigma_x^2\mathbf{I}_p$ so that (C6) holds with $\Sigma_x = \sigma_x^2\mathbf{I}_p$. Results from two such simulations with $p = 2$ and $\sigma_x^2 = 5$ are presented in Figure 2; other cases have
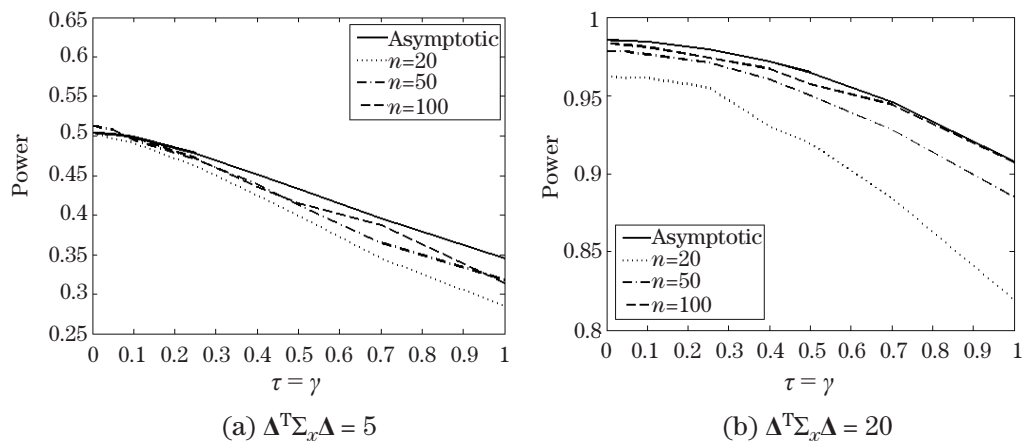
(a) $\mathbf{\Delta}^{\mathrm{T}}\Sigma_x\mathbf{\Delta} = 5$ (b) $\mathbf{\Delta}^{\mathrm{T}}\Sigma_x\mathbf{\Delta} = 20$

Figure 2. Comparison of finite-sample empirical power at different sample sizes n with asymptotic contiguous power for the simple DPDTS of $\boldsymbol{\beta}$ for $p = 2$ and $\sigma_x^2 = 5$.

similar patterns. Clearly the finite-sample powers are close to the asymptotic power in moderate sample sizes like $n = 100$ and the convergence rate is little slower for larger $t = \mathbf{\Delta}^T\mathbf{\Sigma}_x\mathbf{\Delta}$.

5.1.2. Robustness results

As the first order IF of DPDTS $T_\gamma(\boldsymbol{\beta}_n^\tau, \boldsymbol{\beta}_0)$ is zero at any simple null, we measure the stability of the proposed test by the second order IF. For contamination in a single direction (i_0^{th} direction), the second order IF at the null $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ simplifies to

$$
IF_{i_0}^{(2)}(t_{i_0}, T_{\gamma,\tau}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})
$$
$$
= (1+\gamma)\zeta_\gamma(1+\tau)^3 n[\boldsymbol{x}_{i_0}{}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_{i_0}](t_{i_0} - \boldsymbol{x}_{i_0}^T\boldsymbol{\beta}_0)^2 e^{-(\tau(t_{i_0}-\boldsymbol{x}_{i_0}^T\boldsymbol{\beta}_0)^2)/\sigma_0^2}.
$$

As expected, the IF depends on the outliers and the leverage points through $(t_{i_0} - \boldsymbol{x}_{i_0}^T\boldsymbol{\beta}_0)$ and $[\boldsymbol{x}_{i_0}{}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_{i_0}]$. The LIF and PIF of the proposed DPDTS under contiguous alternatives is presented in Section S2.2 of the Online Supplement.

Both the IF and the PIF are bounded with respect to the contamination point t_{i_0} for any $\tau > 0$ implying their stability against contamination. Both are unbounded at t_{i_0} for the LRT at $\gamma = \tau = 0$ indicating its non-robustness. The LIF of this test is identically zero for all $\tau, \gamma \geq 0$ implying no asymptotic influence of contiguous contamination on its size.

6. Application (II): Generalized Linear Model

In a generalized linear model (GLM), the response variables Y_i are inde-

pendent and have an exponential family distribution with density $f(y_i; \theta_i, \phi) = \exp\{(y_i\theta_i - b(\theta_i))/a(\phi) + c(y_i, \phi)\}$; the canonical parameter θ_i depends on the predictor $\boldsymbol{x}_i$ and ϕ is a nuisance scale parameter. The mean μ_i of Y_i satisfies $g(\mu_i) = \eta_i = \boldsymbol{x}_i^T\boldsymbol{\beta}$, for a monotone differentiable link function g and linear predictor $\eta_i = \boldsymbol{x}_i^T\boldsymbol{\beta}$.

The GLMs with fixed predictors fit the general I-NH set-up. The properties of the MDPDEs of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi)$ in the GLM were derived in Ghosh and Basu (2016) and are presented in Section S3 of the Online Supplement.

Suppose we have a sample of size n from a GLM with parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi) \in \Theta = \mathbb{R}^p \times [0, \infty)$ and want to test the hypothesis

$$H_0 : \boldsymbol{L}^T\boldsymbol{\beta} = \boldsymbol{l}_0 \qquad \text{against} \qquad H_1 : \boldsymbol{L}^T\boldsymbol{\beta} \neq \boldsymbol{l}_0, \tag{6.1}$$

where $\boldsymbol{L}$ is a $p \times r$ known matrix with $p \geq r$ and $\boldsymbol{l}_0$ is a real r-vector. We assume that the nuisance parameter ϕ is unknown.

The DPD based test statistic (DPDTS$_C$) for testing this problem is

$$S_\gamma(\boldsymbol{\theta}_n^\tau, \widetilde{\boldsymbol{\theta}}_n^\tau) = 2\sum_{i=1}^n d_\gamma(f_i(.; (\widehat{\boldsymbol{\beta}}_n^\tau, \widehat{\phi}_n^\tau)), f_i(.; (\widetilde{\boldsymbol{\beta}}_n^\tau, \widetilde{\phi}_n^\tau))),$$

where $\boldsymbol{\theta}_n^\tau = (\widehat{\boldsymbol{\beta}}_n^\tau, \widehat{\phi}_n^\tau)$ is the unrestricted MDPDE, $\widetilde{\boldsymbol{\theta}}_n^\tau = (\widetilde{\boldsymbol{\beta}}_n^\tau, \widetilde{\phi}_n^\tau)$ is the restricted MDPDE under H_0 corresponding to the tuning parameter τ.

To derive the asymptotic distribution of the RMDPDE $(\widetilde{\boldsymbol{\beta}}_n^\tau, \widetilde{\phi}_n^\tau)$ of $(\boldsymbol{\beta}, \phi)$ from Theorem 5, some simple matrix algebra leads to

$$\boldsymbol{P}_n^\tau(\boldsymbol{\beta}, \sigma)$$
$$= n\begin{bmatrix} \boldsymbol{\Psi}_{n,11.2}^{-1}[\boldsymbol{I}_p - \boldsymbol{L}\{\boldsymbol{L}^T\boldsymbol{\Psi}_{n,11.2}^{-1}\boldsymbol{L}\}^{-1}\boldsymbol{L}^T\boldsymbol{\Psi}_{n,11.2}^{-1}] & -\boldsymbol{M}_{11}\boldsymbol{X}^T\boldsymbol{\Gamma}_{12}^{(\tau)}\boldsymbol{1}\boldsymbol{\Psi}_{n,22.1}^{-1} \\ -\boldsymbol{\Psi}_{n,22.1}^{-1}\boldsymbol{1}^T\boldsymbol{\Gamma}_{12}^{(\tau)}\boldsymbol{X}\boldsymbol{M}_{11} & \boldsymbol{\Psi}_{n,22.1}^{-1} \end{bmatrix},$$

where, for any $i, j = 1, 2$, $\boldsymbol{\Psi}_{n,ii.j} = \boldsymbol{X}^T\boldsymbol{\Gamma}_{jj}^{(\tau)}\boldsymbol{X} - \boldsymbol{X}^T\boldsymbol{\Gamma}_{ij}^{(\tau)}\boldsymbol{1}(\boldsymbol{1}^T\boldsymbol{\Gamma}_{jj}^{(\tau)}\boldsymbol{1})^{-1}\boldsymbol{1}^T\boldsymbol{\Gamma}_{ji}^{(\tau)}\boldsymbol{X}$ with $\boldsymbol{\Gamma}_{ij}^{(\tau)}$ defined in Section 4 of the online Supplement, and $\boldsymbol{M}_{11} = (\boldsymbol{X}^T\boldsymbol{\Gamma}_{11}^{(\tau)}\boldsymbol{X})^{-1}$.

Corollary 6. *Suppose the Ghosh-Basu conditions hold with respect to Θ_0. The RMDPDE $(\widetilde{\boldsymbol{\beta}}_n, \widetilde{\phi}_n)$ exists and is consistent for $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}^g, \phi^g)$, true parameter value under Θ_0. The asymptotic distribution of $\boldsymbol{\Omega}_n^{-1/2}\boldsymbol{P}_n[\sqrt{n}((\widetilde{\boldsymbol{\beta}}_n, \widetilde{\phi}_n) - (\boldsymbol{\beta}^g, \widetilde{\phi}^g))]$ is $(p+1)$-dimensional normal with mean $\boldsymbol{0}$ and variance $\boldsymbol{I}_{p+1}$, where $\boldsymbol{P}_n = \boldsymbol{P}_n^\tau(\widetilde{\boldsymbol{\beta}}^g, \widetilde{\phi}^g)$ and $\boldsymbol{\Omega}_n = \boldsymbol{\Omega}_n^\tau(\widetilde{\boldsymbol{\beta}}^g, \widetilde{\phi}^g)$, with $\boldsymbol{\Omega}_n(\boldsymbol{\beta}, \phi)$ defined in Section S3 of the Online Supplement.*

As in the unrestricted case, the restricted MDPDE of $\boldsymbol{\beta}$ and ϕ may not be asymptotically independent. They are independent if $\gamma_{12i}^{1+2\tau} = 0, \gamma_{1i}^{1+\tau}\gamma_{2i}^{1+\tau} = 0$, for all i.

To derive asymptotic distribution of the DPDTS_C, we assume fixed covariates $\boldsymbol{x}_i$'s for which the matrices $\boldsymbol{\Psi}_n^\tau(\widetilde{\boldsymbol{\theta}}^g)$ and $\boldsymbol{\Omega}_n^\tau(\widetilde{\boldsymbol{\theta}}^g)$, defined in Section S3 of the Online Supplement, converge element-wise as $n \to \infty$ to some $(p+1) \times (p+1)$ invertible matrices $\boldsymbol{J}_\tau$ and $\boldsymbol{V}_\tau$, respectively. Consider the partitions

$$\boldsymbol{J}_\tau(\boldsymbol{\beta}, \sigma) = \begin{bmatrix} \boldsymbol{J}_{11,\tau} & \boldsymbol{J}_{12,\tau} \\ \boldsymbol{J}_{12,\tau}^T & \boldsymbol{J}_{22,\tau} \end{bmatrix}, \quad \text{and} \quad \boldsymbol{V}_\tau(\boldsymbol{\beta}, \sigma) = \begin{bmatrix} \boldsymbol{V}_{11,\tau} & \boldsymbol{V}_{12,\tau} \\ \boldsymbol{V}_{12,\tau}^T & \boldsymbol{V}_{22,\tau} \end{bmatrix},$$

where $\boldsymbol{J}_{11,\tau}$ and $\boldsymbol{V}_{11,\tau}$ are of order $p \times p$. We suppress τ in above notation whenever it is clear from the context. Then, the asymptotic null distribution of the DPDTS_C $S_\gamma(\boldsymbol{\theta}_n^\tau, \widetilde{\boldsymbol{\theta}}_n^\tau)$ follows directly from Theorem 6 provided the Ghosh-Basu conditions hold for the model under H_0.

Corollary 7. *For the GLM set-up, assume that its density satisfies the Lehmann and Ghosh-Basu conditions under Θ_0. Then the asymptotic null distribution of the DPDTS_C $S_\gamma(\boldsymbol{\theta}_n^\tau, \widetilde{\boldsymbol{\theta}}_n^\tau)$ is the same as that of $\sum_{i=1}^r \zeta_i^{\gamma,\tau}(\boldsymbol{\theta}_0) Z_i^2$, where $Z_1, \cdots, Z_r$ are independent standard normal variables, and $\zeta_1^{\gamma,\tau}(\boldsymbol{\theta}_0), \cdots, \zeta_r^{\gamma,\tau}(\boldsymbol{\theta}_0)$ are r nonzero eigenvalues of the matrix*

$$\boldsymbol{E} = (1+\gamma)\boldsymbol{J}_{11,\gamma}\boldsymbol{J}_{11.2}^{-1}\boldsymbol{L}\boldsymbol{N}_{11}\boldsymbol{L}^T\boldsymbol{J}_{11.2}^{-1}\boldsymbol{V}_{11}\boldsymbol{J}_{11.2}^{-1}\boldsymbol{L}\boldsymbol{N}_{11}\boldsymbol{L}^T\boldsymbol{J}_{11.2}^{-1},$$

where $\boldsymbol{J}_{ii.j} = \boldsymbol{J}_{ii,\tau} - \boldsymbol{J}_{ij,\tau}\boldsymbol{J}_{jj,\tau}^{-1}\boldsymbol{J}_{ji,\tau}^T$ for $i,j = 1,2; i \neq j$ and $\boldsymbol{N}_{11} = (\boldsymbol{L}^T\boldsymbol{J}_{11.2}^{-1}\boldsymbol{L})^{-1}$.

This result can be used to obtain the critical values of the proposed DPD based test. The other asymptotic results regarding power and robustness of the test can be derived by direct application of the general theory of Section 4. For instance, the second order IF of the test statistics at the null hypothesis, when there is contamination in only one fixed direction-i_0, is

$$IF_{i_0}^{(2)}(t_{i_0}, S_{\gamma,\tau}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = n(1+\gamma) \cdot \boldsymbol{W}^T\boldsymbol{\Psi}_n^\gamma\boldsymbol{W}, \tag{6.2}$$

where $\boldsymbol{W} = \boldsymbol{\Psi}_n^{-1}\dfrac{1}{n}\begin{pmatrix} [f_{i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi))^\tau K_{1i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi)) - \gamma_{1i_0}]\boldsymbol{x}_i \\ f_{i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi))^\tau K_{2i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi)) - \gamma_{2i_0} \end{pmatrix}$

$$-\boldsymbol{Q}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Psi}_n^{(0)}(\boldsymbol{\theta}_0)^T\dfrac{1}{n}\begin{pmatrix} f_{i_0}(t_{i_0}; \boldsymbol{\theta}_0)^\tau \boldsymbol{u}_{1i_0}^{(0)}(t_{i_0}; \boldsymbol{\theta}_0) - \gamma_{1i_0}^{(0)} \\ f_{i_0}(t_{i_0}; \boldsymbol{\theta}_0)^\tau \boldsymbol{u}_{2i_0}^{(0)}(t_{i_0}; \boldsymbol{\theta}_0) - \gamma_{2i_0}^{(0)} \end{pmatrix},$$

with $K_{ji_0}(t_{i_0}; (\boldsymbol{\beta}, \phi))$ as defined in Section S3 of the Online Supplement for $j = 1, 2$, $\boldsymbol{u}_{1i}^{(0)}(y_i; (\boldsymbol{\beta}, \phi))$ and $\boldsymbol{u}_{2i}^{(0)}(y_i; (\boldsymbol{\beta}, \phi))$ denoting the restricted derivative of log $f_i(y_i; (\boldsymbol{\beta}, \phi))$ with respect to $\boldsymbol{\beta}$ and ϕ under H_0, and $\boldsymbol{\Psi}_n^{(0)}$ being the matrix $\boldsymbol{\Psi}_n$ constructed using $(\boldsymbol{u}_{1i}^{(0)}, \boldsymbol{u}_{1i}^{(0)})$ in place of $\boldsymbol{u}_i = (\boldsymbol{u}_{1i}, \boldsymbol{u}_{2i})^T$.

Example 1. Consider testing the first r components ($r \leq p$) of the regression

coefficient $\boldsymbol{\beta}$ at a pre-fixed value $\boldsymbol{\beta}_0^{(1)}$, the null hypothesis given by (6.1) with

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{O}_{(p-r) \times r} \end{bmatrix}.$$

We partition the relevant vectors and matrices as $\boldsymbol{\beta} = (\boldsymbol{\beta}_0^{(1)}, \boldsymbol{\beta}_0^{(2)})$, $\boldsymbol{x}_i = (\boldsymbol{x}_i^{(1)}, \boldsymbol{x}_i^{(2)})$ and $\boldsymbol{X} = [\boldsymbol{X}_1 \ \boldsymbol{X}_2]$, where $\boldsymbol{\beta}_0^{(1)}$ and $\boldsymbol{x}_i^{(1)}$ are r-vectors and $\boldsymbol{X}_1$ is the $n \times r$ matrix consisting of the first r columns of $\boldsymbol{X}$. Let

$$\boldsymbol{J}_{11} = \begin{bmatrix} \boldsymbol{J}_{11}^{11} & \boldsymbol{J}_{11}^{12} \\ (\boldsymbol{J}_{11}^{12})^T & \boldsymbol{J}_{11}^{22} \end{bmatrix}, \quad \boldsymbol{V}_{11} = \begin{bmatrix} \boldsymbol{V}_{11}^{11} & \boldsymbol{V}_{11}^{12} \\ (\boldsymbol{V}_{11}^{12})^T & \boldsymbol{V}_{11}^{22} \end{bmatrix}, \quad \boldsymbol{J}_{11.2}^{-1} = \begin{bmatrix} \boldsymbol{J}_{11.2}^{-11} & \boldsymbol{J}_{11.2}^{-12} \\ (\boldsymbol{J}_{11.2}^{-12})^T & \boldsymbol{J}_{11.2}^{-22} \end{bmatrix},$$

where the first block of each partitioned matrix is of order $r \times r$.

Here the asymptotic distribution of the DPD-based test statistics $S_\gamma(\boldsymbol{\theta}_n^\tau, \widetilde{\boldsymbol{\theta}}_n^\tau)$ under the null is that of $\sum_{i=1}^r \zeta_i^{\gamma,\tau}(\boldsymbol{\theta}_0) Z_i^2$, where $Z_1, \cdots, Z_r$ are independent standard normal variables, and $\zeta_1^{\gamma,\tau}(\boldsymbol{\theta}_0), \cdots, \zeta_r^{\gamma,\tau}(\boldsymbol{\theta}_0)$ are r nonzero eigenvalues of the matrix $(1+\gamma)\boldsymbol{J}_{11,\gamma}^{11}\boldsymbol{J}_{11.2}^{-11}\boldsymbol{V}_{11}^{11}\boldsymbol{J}_{11.2}^{-11}$. We have

$$\boldsymbol{W} = \boldsymbol{\Psi}_n^{-1} \frac{1}{n} \begin{pmatrix} \boldsymbol{0}_r \\ [f_{i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi))^\tau K_{1i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi)) - \gamma_{1i_0}] \boldsymbol{x}_i^{(2)} \\ f_{i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi))^\tau K_{2i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi)) - \gamma_{2i_0} \end{pmatrix}.$$

The second order IF follows from (6.2) with this form of $\boldsymbol{W}$. As expected, there is no influence of contamination on the first r components of the RMDPDE.

7. Numerical Illustration: A Data Example

We consider the multiple regression model using the "Salinity data" (Rousseeuw and Leroy (1987, Chap. 2)). These data were discussed in Ruppert and Carroll (1980), Rousseeuw and Leroy (1987), and Ghosh and Basu (2013). Our analysis shows, except for two potential outliers, that the data is well modeled by a multiple linear regression model, taking salinity as the response variable and the covariates as salinity in two weeks lag (x_1), the number of biweekly periods elapsed since the beginning of spring (x_2), and the volume of river discharge into the sound (x_3). Cases 5 and 16 are outlying observations that correspond to periods of very heavy discharge.

The maximum likelihood estimate of the regression coefficient $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ and the error standard deviation σ for the full data are $(9.6, 0.8, -0.03, -0.3)^T$ and 1.23. After deleting the outlying observations these estimates are $(23.39, 0.70, -0.25, -0.84)^T$ and 0.91, respectively, indicating the dramatic effect of outliers. Ghosh and Basu (2013) showed that the MDPDE with $\tau \geq 0.25$ can successfully generate robust estimators even under presence of the two outlying
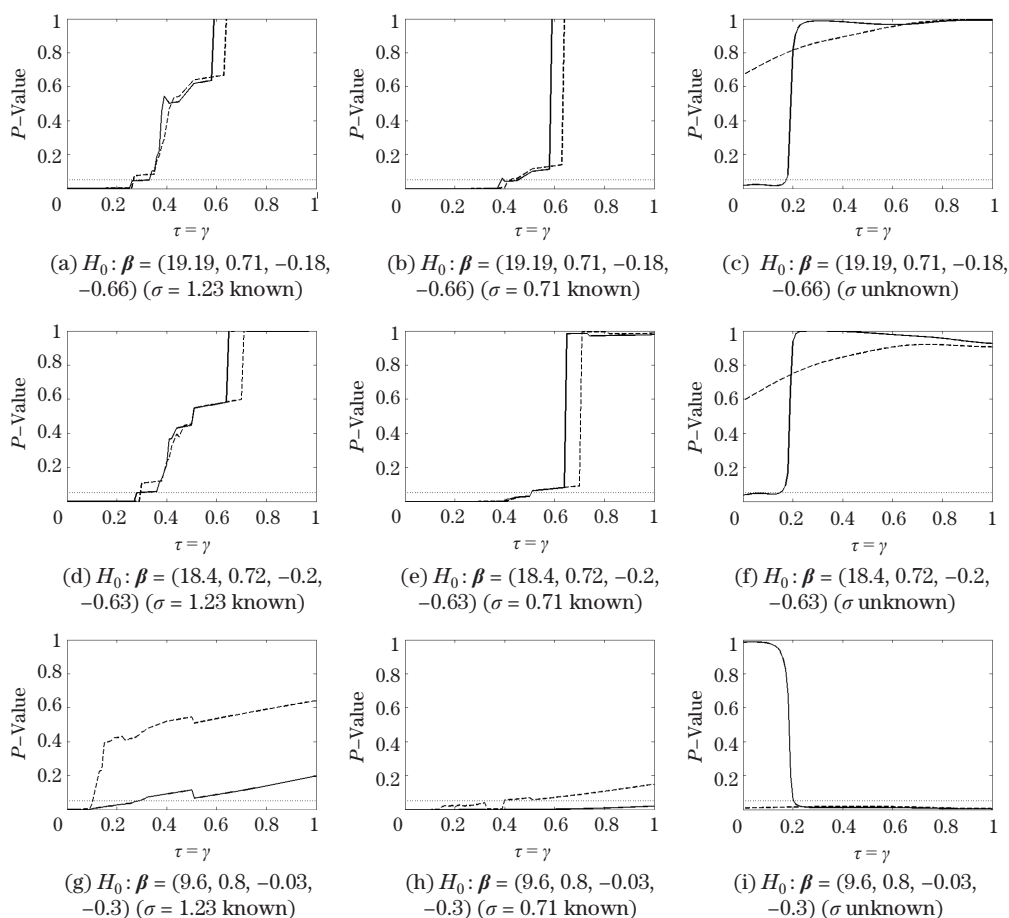
Figure 3. The p-values of the DPD-based tests for different H_0, with known and unknown σ^2, for the Salinity data (Solid line - full data; dashed line - outlier deleted data).

observations. In particular, the MDPDEs at $\tau = 0.5$ and $\tau = 1$ are, respectively, $\widehat{\boldsymbol{\beta}} = (18.4, 0.72, -0.2, -0.63)^T$, $\widehat{\sigma} = 0.87$, and $\widehat{\boldsymbol{\beta}} = (19.19, 0.71, -0.18, -0.66)^T$, $\widehat{\sigma} = 0.87$. These estimates are quite close to the outlier deleted MLE.

We applied the proposed DPD-based test using the full and outlier deleted data. We tested such hypotheses on $\boldsymbol{\beta}$ as $H_0 : \boldsymbol{\beta} = (19.19, 0.71, -0.18, -0.66)^T$, $H_0 : \boldsymbol{\beta} = (18.4, 0.72, -0.2, -0.63)^T$, and $H_0 : \boldsymbol{\beta} = (9.6, 0.8, -0.03, -0.3)^T$. They were chosen at the estimated values for two robust estimators, MDPDE at $\tau = 1$ and 0.5, and the non-robust MLE, respectively. Therefore, a robust test should accept the first two hypotheses while rejecting the third. We considered both simple and composite tests by assuming σ to be known and unknown. For the

known σ case, we assumed two distinct values of σ: 1.23 (a non-robust estimate, MLE) and 0.71 (a robust estimate, MDPDE at $\tau = 1$). The p-values of the proposed DPD based tests for these cases are presented in Figure 3.

When σ is unknown, the DPD-based tests with $\tau = \gamma \geq 0.2$ give robust results by failing to reject the first two hypotheses (Figures 3c, 3f) and by rejecting the third one (Figure 3i) under full data. The performances of the LRT at $\tau = \gamma = 0$ is clearly non-robust under full data. When σ is robustly specified, under full data the DPD-based tests still fail to reject the first two hypotheses at larger $\tau = \gamma \geq 0.5$ but the LRT rejects them (Figures 3b, 3e). All DPD-based tests, including the LRT, successfully reject the third hypothesis under full data for correctly specified robust σ (Figure 3h). When σ is incorrectly specified, the DPD-based tests at $\tau = \gamma \geq 0.5$ still lead to robust inference while the LRT provides incorrect inference for the first two hypotheses (Figures 3a, 3d); the third hypothesis gets accepted by the DPD-based tests at larger $\tau = \gamma$ due to the incorrect specification of σ (Figure 3g).

8. Conclusions

In this paper we have presented a general framework based on density power divergence for performing robust tests of hypothesis in the independent but non-homogeneous case. We have established the wide scope of the test, and numerically demonstrated its applicability to the linear regression problem. Due to the generality of the method and theoretical indicators it is expected that it will be a powerful tool for the practitioner, although further numerical studies would be helpful to explore the performance of these tests in specific situations.

Among possible extensions, we hope to study the multisample problem and the case of heteroscedastic models. Finally, the choice of tuning parameters requires a thorough study (see Section S7 of the Online Supplement for a brief discussion).

Supplementary Materials

The Online Supplement to this paper contains the details of S1. required assumptions; S2. MDPDE under the linear regression model with fixed design along with some additional results for testing; S3. MDPDE under the GLM with fixed design; S4. proofs; S5. simulations under linear regression models; S6. two more data applications from linear and Poisson regression models; and S7. comments on the choice of tuning parameters.

Acknowledgment

The authors thank the referees, the Editor and an associate editor for their constructive comments and suggestions.

References

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.

Basu, A., Mandal, A., Martin, N. and Pardo, L. (2013a). Testing statistical hypotheses based on the density power divergence. *Ann. Inst. Statist. Math.* **65**, 319–348.

Basu, A., Mandal, A., Martin, N. and Pardo, L. (2013b). Density power divergence tests for composite null hypotheses. *ArXiv pre-print*, arXiv:1403.0330 [stat.ME].

Cantoni, E. and Ronchetti, E. (2001). Bounded influence for generalized linear models. *J. Amer. Statist. Assoc.* **96**, 1022–1030.

Ghosh, A. (2015). Influence function of the restricted minimum divergence estimators: A general form. *Electron. J. Stat.* **9**, 1017–1040.

Ghosh, A. (2017). Divergence based robust estimation of tail index with exponential regression model. *Stat. Methods Appl.* **26**, 181–213.

Ghosh, A. and Basu, A. (2013). Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electron. J. Stat.* **7**, 2420–2456.

Ghosh, A. and Basu, A. (2016). Robust estimation in generalised linear models: The density power divergence approach. *TEST* **25**, 269–290.

Hampel, F. R., Ronchetti, E., Rousseeuw, P. J. and Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions.* John Wiley & Sons, New York.

Heritier, S. and Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *J. Amer. Statist. Assoc.* **89**, 897–904.

Hosseinian, S. (2009). *Robust Inference for Generalized Linear Models: Binary and Poisson Regression.* Ph. D. thesis, Ecole Polytechnique Federal de Lausanne.

Huber, P. J. (1983). Minimax aspects of bounded-influence regression (with discussion). *J. Amer. Statist. Assoc.* **78**, 66–80.

Kotz, S., Johnson, N. L. and Boyd, D. W. (1967a). Series representations of distributions of quadratic forms in normal variables. I. *Ann. Math. Statist.* **38**, 823–837.

Kotz, S., Johnson, N. L. and Boyd, D. W. (1967b). Series representations of distributions of quadratic forms in normal variables. II. *Ann. Math. Statist.* **38**, 838–848.

Lehmann, E. L. (1983). *Theory of Point Estimation.* John Wiley & Sons, New York.

Lindsay, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Ann. Statist.* **22**, 1081–1114.

Liu, R. C., Markatou, M. and Tsai, C. L. (2005). Robust estimation and testing in nonlinear regression models. *Int. J. Pure Appl. Math.* **21**, 525–552.

Markatou, M. and He, X. (1994). Bounded-influence and high-breakdown-point testing procedures in linear models. *J. Amer. Statist. Assoc.* **89**, 543–549.

Markatou, M. and Hettmansperger, T. P. (1990). Robust bounded-influence tests in linear

models. *J. Amer. Statist. Assoc.* **85**, 187–190.

Markatou, M. and Manos, G. (1996). Robust tests in nonlinear regression models. *J. Statist. Plann. Inference* **55**, 205–217.

Maronna, R. A., Martin, D. R. and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods.* John Wiley and Sons, New York.

Muller, C. (1998). Optimum robust testing in linear models. *Ann. Statist.* **26**, 1126–1146.

Ronchetti, E. (1982a). Robust alternatives to the F-test for the linear model. In *Probability and Statistical Inference*, W. Grossmann, C. Pflug, and W. Wertz (eds.). Reider, Dortrecht, 329–342.

Ronchetti, E. (1982b). Robust testing in linear models: the infinitesimal approach. *Ph.D. Thesis*, ETH, Zurich.

Ronchetti, E. (1987). Robustness aspect of model choice. *Statist. Sinica* **7**, 327–338.

Ronchetti, E. and Rousseeuw, P. J. (1980). A robust F-test for the linear model. *Abstract Book, 13th European Meeting of Statisticians*, England, 210–211.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* John Wiley & Sons, New York.

Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75**, 828–838.

Salibian-Barrera, M., Van Aelst, S. and Yohai, V. J. (2016). Robust tests for linear regression models based on τ-estimates. *Comput. Statist. Data Anal.* **93**, 436–455.

Schrader, R. M. and Hettmansperger, T. P. (1980). Robust analysis of variance based upon a likelihood ratio criterion. *Biometrika* **67**, 93–101.

Sen, P. K. (1982). On M-tests in linear models. *Biometrika* **69**, 245–248.

Simpson, D. G. (1989). Hellinger deviance test: efficiency, breakdown points, and examples. *J. Amer. Statist. Assoc.* **84**, 107–113.

Toma, A. and Broniatowski, M. (2010). Dual divergence estimators and tests: robustness results. *J. Multivariate Anal.* **102**, 20–36.

Wang, L. and Qu, A. (2007). Robust tests in regression models with omnibus alternatives and bounded influence. *J. Amer. Statist. Assoc.* **102**, 347–358.

Indian Statistical Institute, Kolkata, 203, B. T. Road, Kolkata 700108, West Bengal, India.

E-mail: abhianik@gmail.com

Indian Statistical Institute, Kolkata, 203, B. T. Road, Kolkata 700108, West Bengal, India.

E-mail: ayanbasu@isical.ac.in