Statistica Sinica Preprint No: SS-2015-0320R2	
Title	Robust Bounded Influence Tests for Independent Non-Homogeneous Observations
Manuscript ID	SS-2015.0320.R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202015.0320
Complete List of Authors	Ayanendranath Basu and Abhik Ghosh
Corresponding Author	Ayanendranath Basu
E-mail	ayanbasu@isical.ac.in
Notice: Accepted version subject to English editing.	

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

Robust Bounded Influence Tests for Independent Non-Homogeneous Observations*

Abhik Ghosh and **Ayanendranath Basu**
Indian Statistical Institute, Kolkata, India

January 9, 2017

Abstract

Real-life experiments often yield non-identically distributed data which have to be analyzed using statistical modelling techniques. Tests of hypothesis under such set-ups are generally performed using the likelihood ratio test, which is highly 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 on the simple and generalized linear regression models with fixed covariates.

Keywords: Robust Testing of Hypothesis, Non-Homogeneous Observation, Linear Regression, Generalized Linear Model, Influence Function.

1 Introduction

One of the most important paradigms of parametric statistical inference is testing of hypotheses. Arguably the most popular hypothesis testing procedure in a general situation is the likelihood ratio test (LRT). However, just like the maximum likelihood estimator (MLE), the LRT may lead to highly unstable inference under 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. The robust hypothesis testing problem in case of non-identically distributed data has

*This is a part of the Ph.D. dissertation of the first author.

received limited attention in the literature though there have been few attempts for some of the special cases like the fixed-carrier linear regression model etc.

In this paper, we consider the general case of non-identically distributed data. Mathematically, 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 $i = 1, 2, \ldots, n$. Also let G_i and $F_i(\cdot, \boldsymbol{\theta})$ be the distribution functions corresponding to g_i and $f_i(\cdot;\boldsymbol{\theta})$. Even though the Y_is have possibly different densities, all of them share the common parameter $\boldsymbol{\theta}$. Throughout the paper, we will refer to this set-up as the set-up of independent non-homogeneous observations or simply as the I-NH set-up.

The most prominent application of this set-up is the regression model with fixed 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 $\boldsymbol{\beta}$, i.e., $y_i \sim f_i(\cdot, \boldsymbol{x}_i, \boldsymbol{\beta})$. This set-up models many real-life applications. However, it is different from the usual regression set-up with stochastic covariates, which has been explored in relatively 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 et al., 2005; Maronna et al., 2006; Wang and Qu, 2007; Salibian-Barrera et al., 2014). Our set-up treats the regression problem from a design point of view where we generally pre-fix the covariate levels; examples of such situations include the clinical trials with pre-fixed treatment levels, any planned experiment etc. This general I-NH set-up also includes the heteroscedastic regression model provided we know the type of heteroscedasticity in residuals, eg. the i-th residual has variance proportional to the covariate value $\boldsymbol{x}_i$. The robustness literature under this general I-NH set-up is limited; some scattered attempts have been made in some simple particular cases like normal regression (Huber, 1983; Muller, 1998).

In this context, 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. This approach has been used in Ghosh (2016) to obtain a robust alternative for the tail index estimation under suitable assumptions of an exponential regression model. Here, we exploit the properties of this estimation approach of Ghosh and Basu (2013) to develop a general class of robust tests of hypotheses under I-NH data.

For the sake of completeness, we start with a brief description of the MDPDE under I-NH set-up in Section 2. Then, we consider the case of both the simple and composite null hypotheses in Sections 3 and 4 respectively. Several useful asymptotic and robustness properties including the boundedness of the influence functions of the proposed tests are derived. To illustrate the applicability of these general tests, the standard linear regression model and the generalized linear model (GLM) with fixed covariates are discussed in Sections 5 and 6 respectively. Section 7 presents some numerical illustrations; many more are provided in the online supplement. Some comments on the choice of tuning parameters and on comparison with existing tests are provided in Section 8. The paper ends with a short overall discussion in Section 9. Proofs of all the results and some more details and examples are presented in the online supplement.

To sum up we list, in the following, the specific advantages of the proposed methods. Some of these are matched by some of its competitors, but there are few, if any, tests which combine all these properties.

1. The method is completely general in that it works for any set-up involving independent non-homogeneous data. Other scenarios such as linear regression, generalized linear model etc., with fixed covariate, emerge as specific sub-cases of our approach, but the proposal is by no means limited to these or specific to them.

2. The proposal is very simple to implement with minimal addition in computational complexity compared to likelihood based methods. In this sense, the method distinguishes itself from some of its competitors having strong theoretical properties but high computational burden.

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 level and power influence functions. Boundedness of the level and power influence functions are rarely considered even in case of i.i.d. data. We extend the concept of the level and the power influence functions in the case of independent but non-homogeneous data.

5. The proposed tests are consistent at any fixed alternative. Further they also have high power against any contiguous alternative which makes them even more competitive with other powerful tests.

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 1 of the online supplement for completeness.

2 Background: The MDPDE under the I-NH Set-up

Consider the general I-NH set-up as described in Section 1. Under this set-up, Ghosh and Basu (2013) proposed the estimation of $\boldsymbol{\theta}$ by minimizing the average DPD measure between the data and the model, i.e., equivalently by minimizing the objective function

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

where $V_i(\cdot;\boldsymbol{\theta})$ is the indicated term within the square brackets in the above equation. The corresponding estimating equation is then given by

$$\sum_{i=1}^{n}\left[f_i(Y_i;\boldsymbol{\theta})^{\tau}\boldsymbol{u}_i(Y_i;\boldsymbol{\theta}) - \int f_i(y;\boldsymbol{\theta})^{1+\tau}\boldsymbol{u}_i(y;\boldsymbol{\theta})dy\right] = 0, \quad (2)$$

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 $\boldsymbol{\theta}$). In the particular case $\tau = 0$, 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.

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

$$\frac{1}{n}\sum_{i=1}^{n} d_\tau(g_i(.), f_i(.;\boldsymbol{U}_\tau(\underline{\boldsymbol{G}}))) = \min_{\boldsymbol{\theta}\in\Theta}\frac{1}{n}\sum_{i=1}^{n} d_\tau(g_i(.), f_i(.;\boldsymbol{\theta})), \quad (3)$$

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

$$d_\tau(f_1, f_2) = \begin{cases} \displaystyle\int \left[f_2^{1+\tau} - \left(1+\frac{1}{\tau}\right) f_2^{\tau} f_1 + \frac{1}{\tau} f_1^{1+\tau}\right], & \text{for } \tau > 0, \\ \displaystyle\int f_1 \log(f_1/f_2), & \text{for } \tau = 0. \end{cases} \quad (4)$$

Equivalently, $\boldsymbol{U}_\tau(\underline{\boldsymbol{G}})$ is also the minimizer of $\frac{1}{n}\sum_{i=1}^{n} H^{(i)}(\boldsymbol{\theta})$, with respect to $\boldsymbol{\theta} \in \Theta$, where

$$H^{(i)}(\boldsymbol{\theta}) = \int f_i(y;\boldsymbol{\theta})^{1+\tau}dy - \left(1+\frac{1}{\tau}\right)\int f_i(y;\boldsymbol{\theta})^{\tau} g_i(y)dy. \quad (5)$$

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

(i) There exists a consistent sequence $\widehat{\boldsymbol{\theta}}_n$ of roots to the minimum density power divergence estimating equation (2).

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

$$\boldsymbol{\Psi}_n^{\tau}(\boldsymbol{\theta}^g) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{J}^{(i)}(\boldsymbol{\theta}^g), \quad (6)$$

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

$$\boldsymbol{\Omega}_n^\tau(\boldsymbol{\theta}^g) = \frac{1}{n}\sum_{i=1}^n \left[\int \boldsymbol{u}_i(y;\boldsymbol{\theta}^g)\boldsymbol{u}_i^T(y;\boldsymbol{\theta}^g) f_i(y;\boldsymbol{\theta}^g)^{2\tau} g_i(y)dy - \boldsymbol{\xi}_i\boldsymbol{\xi}_i^T\right], \tag{7}$$

with

$$\boldsymbol{\xi}_i = \int \boldsymbol{u}_i(y;\boldsymbol{\theta}^g) f_i(y;\boldsymbol{\theta}^g)^\tau g_i(y)dy. \tag{8}$$

3 Testing Simple Hypothesis under the I-NH Set-up

We start with the simple hypothesis testing problem with a fully specified null. We adopt the notations of Section 1 for the I-NH set-up and take a fixed point $\boldsymbol{\theta}_0$ in the parameter space Θ. Based on the observed data, we want to test

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{against} \quad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0. \tag{9}$$

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 (9) 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. However, since there are n divergence measures corresponding to each i, we consider the total divergence measure over the n data points for testing (9). Thus, we 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 $d_\gamma(f_1, f_2)$ denotes the DPD measure between two densities f_1 and f_2 as defined in (4). In case of i.i.d. data, this DPDTS coincides with the test statistic in Basu et al. (2013a). Here the tuning parameter τ specifies the choice of the estimator used and γ refers to the divergence measure used in constructing the test statistics. More details on their selection and implications will be discussed in Section 8.

3.1 Asymptotic Properties

Consider the matrices $\boldsymbol{\Psi}_n^\tau$ and $\boldsymbol{\Omega}_n^\tau$ as defined in Equations (6) and (7) respectively and define $\boldsymbol{A}_n^\gamma(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{A}_\gamma^{(i)}(\boldsymbol{\theta})$ with $\boldsymbol{A}_\gamma^{(i)}(\boldsymbol{\theta}_0) = \nabla^2 d_\gamma(f_i(.;\boldsymbol{\theta}), f_i(.;\boldsymbol{\theta}_0))\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$. Also, 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 3.1. *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 $T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0)$ coincides with 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*

$$r = rank(\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)\boldsymbol{J}_\tau{}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\boldsymbol{J}_\tau{}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)).$$

Note that the null distribution of the proposed DPDTS has the same form as that was in Basu et al. (2013a,b) for i.i.d. observations. So, we can easily find the critical region of the our proposal also from the relevant discussions in Basu et al. (2013a,b).

Next we present an approximation to its power function. Define

$$\boldsymbol{M}_n^\gamma(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^n \boldsymbol{M}_\gamma^{(i)}(\boldsymbol{\theta})$$

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 3.2. *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 (9) 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 $\sigma_{\tau,\gamma}(\boldsymbol{\theta}^*)$ is defined by $\sigma_{\tau,\gamma}^2(\boldsymbol{\theta}) = \boldsymbol{M}_\gamma(\boldsymbol{\theta})^T\boldsymbol{\Sigma}_\tau(\boldsymbol{\theta})\boldsymbol{M}_\gamma(\boldsymbol{\theta})$.

Corollary 3.3. *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 $\frac{1}{n}\sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}^*), f_i(.;\boldsymbol{\theta}_0)) = O(1)$. So, the proposed DPD based test statistic is consistent.*

Theorem 3.2 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^* denote the solution of the above equation, then the required sample size is the least integer greater than or equal to n^*.

3.2 Robustness Properties

3.2.1 Influence Function of the Test Statistics

Now we illustrate the robustness of the proposed DPDTS; first we consider Hampel's influence function (IF) of the test statistics (Rousseeuw and Ronchetti, 1979, 1981; Hampel et al., 1986). However, in the case of I-NH observations, we cannot define the IF exactly as in the i.i.d. cases. Suitable extensions can be found in Huber (1983) and Ghosh and Basu (2013). Here we will use a similar idea to define the IF of the DPDTS.

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 I-NH set-up as defined in Section 2. Note that, unlike the i.i.d. case, here the functional itself depends on the sample size n so that the corresponding IF will also depend on the sample size. 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 $i = 1, \ldots, n$. As in the estimation problem in Ghosh and Basu (2013), here also we can have contamination in some fixed direction or in all the directions.

First, consider the contamination only in the i_0-th direction and define $\underline{\mathbf{G}}_{i_0,\epsilon} = (G_1, \cdots, G_{i_0-1}, G_{i_0,\epsilon}, \cdots, G_n)$. 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 corresponding IF of $\boldsymbol{U}_\tau$ derived in Ghosh and Basu (2013) is given by

$$IF_{i_0}(t_{i_0}, \boldsymbol{U}_\tau, \underline{\mathbf{G}}) = \frac{1}{n} \boldsymbol{\Psi}_n^\tau(\boldsymbol{\theta}^g)^{-1} \boldsymbol{D}_{\tau,i_0}(t_{i_0}; \boldsymbol{\theta}^g) \tag{10}$$

where $\boldsymbol{D}_{\tau,i}(t;\boldsymbol{\theta}) = [f_i(t;\boldsymbol{\theta})^\tau \boldsymbol{u}_i(t;\boldsymbol{\theta}) - \boldsymbol{\xi}_i]$ with $\boldsymbol{\xi}_i$ as defined in Equation (8). In general practice, 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.

So, we need to consider higher order influence functions of this test. 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 Equation (3), the IF of the test becomes

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

Note that, $\boldsymbol{D}_{\tau,i}(t;\boldsymbol{\theta})$ is bounded in t if the parametric models satisfies that $f_i(t;\boldsymbol{\theta})^{\tau}\boldsymbol{u}_i(t;\boldsymbol{\theta})$ 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 $\tau = 0$. However, note that $\boldsymbol{D}_{\tau,i}(t;\boldsymbol{\theta})$ does not depend on the tuning parameter γ and hence its boundedness and therefore the boundedness of the IF of the proposed test is independent of the choice of γ.

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

$$\begin{aligned} &IF^{(2)}(\mathbf{t}, T_{\gamma,\tau}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = n \cdot IF(\mathbf{t}, \boldsymbol{U}_{\tau}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})^T A_n^{\gamma} IF(\mathbf{t}, \boldsymbol{U}_{\tau}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}). \\ &= \frac{1}{n}\left(\sum_{i=1}^{n}\boldsymbol{D}_{\tau,i}(t_i;\boldsymbol{\theta}_0)\right)^T [\boldsymbol{\Psi}_n^{\tau}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{A}_n^{\gamma}(\boldsymbol{\theta}_0)\boldsymbol{\Psi}_n^{\tau}(\boldsymbol{\theta}_0)^{-1}]\left(\sum_{i=1}^{n}\boldsymbol{D}_{\tau,i}(t_i;\boldsymbol{\theta}_0)\right). \end{aligned}$$

This influence function is also 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, the robustness refers to the local robustness of the test statistics under infinitesimal contamination in an outlying point. A bounded IF implies that, even if the outliers is far away from the central data cloud, the resulting test statistic (and hence the inference drawn based on its value) cannot be shifted by an infinite distance from the true value under infinitesimal contamination at that outlying point.

3.2.2 Level and Power under contamination and their Influence Functions

The performance of any testing procedure is generally measured in terms of its level and power. So, the stability of its level and power is essential for proposing any robust test of statistical hypothesis. We consider the effect of contamination on level and power of the proposed DPDTS through its 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 will work with their asymptotic versions and the resulting level and power influence functions then measure, respectively, the bias in the asymptotic level and the power of the test due to an infinitesimal contamination at some outlier. If the level or power influence functions are bounded in the outlying point, the asymptotic level and power, respectively, cannot be displaced indefinitely from their true nominal values under pure data and the test becomes robust with respect to its level and power also.

Since the proposed DPDTS is consistent, we should examine its asymptotic power under the contiguous alternative $H_{1,n} : \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\Delta}$ with $\boldsymbol{\Delta} \in \mathbb{R}^p - \{0\}$. In this case we consider contamination over these alternatives. As argued in Hampel et al. (1986), we must consider contaminations such that their effect tends 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. So, we consider the contaminated distributions

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

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

$$\begin{aligned} LIF(\mathbf{t}; T^{(1)}_\gamma, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) &= \lim_{n\to\infty} \frac{\partial}{\partial\epsilon} P_{\underline{\mathbf{F}}^L_{n,\epsilon,\mathbf{t}}}(T_\gamma(\boldsymbol{\theta}^\tau_n, \boldsymbol{\theta}_0) > t^{\tau,\gamma}_\alpha)\big|_{\epsilon=0}, \\ PIF(\mathbf{t}; T^{(1)}_\gamma, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) &= \lim_{n\to\infty} \frac{\partial}{\partial\epsilon} P_{\underline{\mathbf{F}}^P_{n,\epsilon,\mathbf{t}}}(T_\gamma(\boldsymbol{\theta}^\tau_n, \boldsymbol{\theta}_0) > t^{\tau,\gamma}_\alpha)\big|_{\epsilon=0}. \end{aligned}$$

We first derive the asymptotic power under the contaminated distribution $\underline{\mathbf{F}}^P_{n,\epsilon,\mathbf{y}}$ and examine its special cases by substituting specific values of $\boldsymbol{\Delta}$ and ϵ.

Theorem 3.4. *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 $\underline{\mathbf{F}}^P_{n,\epsilon,\mathbf{t}}$ is the same as the distribution of the quadratic form $\boldsymbol{W}^T\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\boldsymbol{W}$, where $\boldsymbol{W} \sim N_p\left(\widetilde{\boldsymbol{\Delta}}, \boldsymbol{\Sigma}_\tau(\boldsymbol{\theta}_0)\right)$ with $\widetilde{\boldsymbol{\Delta}} = \left[\boldsymbol{\Delta} + \epsilon IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})\right]$ and $\boldsymbol{\Sigma}_\tau$ as defined in Theorem 3.1.

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

$$\begin{aligned} P_{\tau,\gamma}(\boldsymbol{\Delta}, \epsilon; \alpha) &= \lim_{n\to\infty} P_{\underline{\mathbf{F}}^L_{n,\epsilon,\mathbf{t}}}(T_\gamma(\boldsymbol{\theta}^\tau_n, \boldsymbol{\theta}_0) > t^{\tau,\gamma}_\alpha), \\ &= \sum_{v=0}^{\infty} C^{\gamma,\tau}_v(\boldsymbol{\theta}_0, \widetilde{\boldsymbol{\Delta}}) P\left(\chi^2_{r+2v} > \frac{t^{\tau,\gamma}_\alpha}{\zeta^{\gamma,\tau}_{(1)}(\boldsymbol{\theta}_0)}\right), \end{aligned} \tag{11}$$

where χ^2_p denotes a chi-square random variable with p degrees of freedom, $\zeta^{\gamma,\tau}_{(1)}(\boldsymbol{\theta}_0)$ is the minimum of $\zeta^{\gamma,\tau}_i(\boldsymbol{\theta}_0)$s for $i = 1, \ldots, r$ defined in Theorem 3.1 and

$$C^{\gamma,\tau}_v(\boldsymbol{\theta}_0, \widetilde{\boldsymbol{\Delta}}) = \frac{1}{v!}\left(\prod_{j=1}^{r} \frac{\zeta^{\gamma,\tau}_{(1)}(\boldsymbol{\theta}_0)}{\zeta^{\gamma,\tau}_j(\boldsymbol{\theta}_0)}\right)^{1/2} e^{-\frac{1}{2}\sum\limits_{j=1}^{r}\delta_j} E\left[\left(\widehat{Q}\right)^v\right],$$

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

for r independent standard normal random variables $Z_1, \ldots, Z_r$ and δ_is are as defined in Remark 3.1 below.

Corollary 3.5. *Putting $\epsilon = 0$ in the above theorem, we get the asymptotic power under the contiguous alternatives $H_{1,n} : \boldsymbol{\theta} = \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-\frac{1}{2}}\boldsymbol{\Delta}$ as*

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

Corollary 3.6. *Putting $\boldsymbol{\Delta} = \mathbf{0}$ in the above theorem, we get the asymptotic level under the probability distribution $\underline{\mathbf{F}}^L_{n,\epsilon,\mathbf{t}}$ as*

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

Remark 3.1. *The asymptotic distribution of $T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0)$ under $\underline{\mathbf{F}}^P_{n,\epsilon,\mathbf{t}}$, as derived in Part (i) of Theorem 3.4, is also the same as that of $\sum_{i=1}^{r} \zeta_i^{\gamma,\tau}(\boldsymbol{\theta}_0)\chi^2_{1,\delta_i}$, where $\zeta_i^{\gamma,\tau}(\boldsymbol{\theta}_0)$s are as in Theorem 3.1 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 3.2. *Note that the infinite series used in the expressions of asymptotic level and power under contiguous alternative with contamination can be approximated, in practice, by truncating it up to a finite number (N) of terms. The error incurred by such truncation can be made smaller than any pre-specific limit by choosing N suitably large. For any practical usage, we can use a finite truncated sum up to the N-th term as an approximation of the infinite series considered in (11). and then the error in this approximation can be bounded by*

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

See Kotz et al. (1967a,b) for more accurate error bounds for such approximations.

Starting with the expression of $P_{\tau,\gamma}(\boldsymbol{\Delta}, \epsilon; \alpha)$ as obtained in Theorem 3.4 and differentiating, we get the power influence function $PIF(\cdot)$ as given in the following theorem. The theorem shows that the PIF is bounded whenever the IF of the MDPDE is bounded. But such is the case for $\tau > 0$ in most statistical models, implying the power robustness of the proposed DPDTS with $\tau > 0$.

Theorem 3.7. *Assume that the Lehmann and Ghosh-Basu conditions hold for the model density and (C1)-(C2) hold at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Also, suppose that 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

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

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

4 Testing Composite Hypothesis under the I-NH Set-up

In this section, we consider the composite null hypothesis. Consider again the I-NH set-up with notations as in Section 1 and take a fixed (proper) subspace Θ_0 of Θ. Based on the observed data, we want to test the hypothesis

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad \text{against} \quad H_1 : \boldsymbol{\theta} \notin \Theta_0. \tag{12}$$

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, for some $\boldsymbol{\theta}_0 \in \Theta_0$. Then, 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. In place of $\widehat{\boldsymbol{\theta}}$, we take the MDPDE $\boldsymbol{\theta}_n^\tau$ of $\boldsymbol{\theta}$ with tuning parameter τ. And, in place of the $\widetilde{\boldsymbol{\theta}}$, we consider 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 this estimator $\widetilde{\boldsymbol{\theta}}_n^\tau$ as the restricted MDPDE (RMDPDE) and discuss its properties in Section 4.1. Thus, our test statistic (DPDTS$_C$) for the composite hypothesis given in (12) based on the DPD with parameter γ is defined as

$$S_\gamma(\boldsymbol{\theta}_n^\tau, \widetilde{\boldsymbol{\theta}}_n^\tau) = 2 \sum_{i=1}^{n} d_\gamma(f_i(.; \boldsymbol{\theta}_n^\tau), f_i(.; \widetilde{\boldsymbol{\theta}}_n^\tau)). \tag{13}$$

4.1 Properties of the RMDPDE under I-NH Set-up

The restricted minimum density power divergence estimators (RMDPDE) $\widetilde{\boldsymbol{\theta}}_n^\tau$ of $\boldsymbol{\theta}$ is defined as the minimizer of the DPD objective function $H_n(\boldsymbol{\theta})$ given by Equation (1) with tuning parameter τ subject to a set of r restrictions of the form

$$\boldsymbol{\upsilon}(\boldsymbol{\theta}) = \mathbf{0}, \tag{14}$$

where $\boldsymbol{\upsilon} : \mathbb{R}^p \mapsto \mathbb{R}^r$ is some vector valued function. For the null hypothesis in (12), such restrictions are given by the definition of the null parameter space Θ_0. Further, we assume that the $p \times r$ matrix $\boldsymbol{\Upsilon}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\upsilon}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ exists and it is continuous in $\boldsymbol{\theta}$ with rank r. Then, the RMDPDE has to satisfy

$$\left.\begin{array}{rcl} \nabla H_n(\boldsymbol{\theta}) + \boldsymbol{\Upsilon}(\boldsymbol{\theta})\boldsymbol{\lambda}_n & = & 0 \\ \boldsymbol{\upsilon}(\boldsymbol{\theta}) & = & 0 \end{array}\right\}, \tag{15}$$

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

Theorem 4.1. *Assume that the Ghosh-Basu conditions are satisfied with respect to Θ_0 (instead of Θ). Then the following results hold:*

(i) There exists a consistent sequence $\widetilde{\boldsymbol{\theta}}_n^\tau$ of roots to the restricted minimum density power divergence estimating equations (15).

(ii) Asymptotically, $\boldsymbol{\Omega}_n^\tau(\widetilde{\boldsymbol{\theta}}^g)^{-\frac{1}{2}}\boldsymbol{P}_n^\tau(\widetilde{\boldsymbol{\theta}}^g)^{-1}[\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g)] \sim N_p(\mathbf{0}, \boldsymbol{I}_p)$ 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 corollary, we will further assume that

(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 4.2. *Along with the assumptions of the above theorem, let us also assume that (C1) and (C4) hold at $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}^g$. Then, asymptotically, $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g) \sim N_p\left(\mathbf{0}, \boldsymbol{P}_\tau(\widetilde{\boldsymbol{\theta}}^g)\boldsymbol{V}_\tau(\widetilde{\boldsymbol{\theta}}^g)\boldsymbol{P}_\tau(\widetilde{\boldsymbol{\theta}}^g)\right)$*

Next, we explore the robustness properties of the RMDPDEs in terms of their influence function. However, in the present case of I-NH data, the contamination can be in any one or more (or all) directions i ($i = 1, \cdots, n$) so that the corresponding IF depends on the sample size n as in the unrestricted case (Ghosh and Basu, 2013). Let us first consider the contamination in only one (i_0-th) direction as in Section 3.2.1. Also, 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 $\boldsymbol{\theta} = \widetilde{\boldsymbol{U}}_\tau(\underline{\mathbf{G}}_{i0,\epsilon})$ 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, $IF_{i_0}(t_{i_0}; \widetilde{\boldsymbol{U}}_\tau; \underline{\mathbf{G}}) = \frac{\partial}{\partial \epsilon}\widetilde{\boldsymbol{U}}_\tau(\underline{\mathbf{G}}_{i0,\epsilon})\big|_{\epsilon=0}$ as a solution of

$$\boldsymbol{\Psi}_n^{(0)}(\widetilde{\boldsymbol{\theta}}^g) IF_{i_0}(t_{i_0}, \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{G}}) - \frac{1}{n}\boldsymbol{D}_{\tau,i_0}^{(0)}(t_{i_0}; \widetilde{\boldsymbol{\theta}}^g) = 0, \tag{16}$$

where $\boldsymbol{D}_{\tau,i}^{(0)}(t;\boldsymbol{\theta}) = \left[f_i(t;\boldsymbol{\theta})^\tau \boldsymbol{u}_i{}^{(0)}(t;\boldsymbol{\theta}) - \boldsymbol{\xi}_i^{(0)}(\boldsymbol{\theta})\right]$ 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, but under the additional restriction $\boldsymbol{\upsilon}(\boldsymbol{\theta}) = 0$. Also, $\widetilde{\boldsymbol{U}}_\tau(\underline{\mathbf{G}}_{i0,\epsilon})$ must satisfy (14), from which we get

$$\boldsymbol{\Upsilon}(\widetilde{\boldsymbol{\theta}}^g)^T IF_{i_0}(t_{i_0}, \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{G}}) = \mathbf{0}. \tag{17}$$

Solving Equations (16) and (17) (as done for the i.i.d. case in Ghosh (2015)), we get a general expression for the IF of the RMDPDE given by

$$IF_{i_0}(t_{i_0}, \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{G}}) = \frac{1}{n}\boldsymbol{Q}(\widetilde{\boldsymbol{\theta}}^g)^{-1}\boldsymbol{\Psi}_n^{(0)}(\widetilde{\boldsymbol{\theta}}^g)^T\boldsymbol{D}_{\tau,i_0}^{(0)}(t_{i_0}; \widetilde{\boldsymbol{\theta}}^g), \tag{18}$$

where $\boldsymbol{Q}(\boldsymbol{\theta}) = \left[\boldsymbol{\Psi}_n^{(0)}(\boldsymbol{\theta})^T\boldsymbol{\Psi}_n^{(0)}(\boldsymbol{\theta}) + \boldsymbol{\Upsilon}(\boldsymbol{\theta})\boldsymbol{\Upsilon}(\boldsymbol{\theta})^T\right]$. 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.

Similarly, if we consider the contamination 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

Let us assume that Θ_0 is a proper subset of the parameter space Θ which can be defined in terms of r restrictions $\boldsymbol{\upsilon}(\boldsymbol{\theta}) = 0$ such that the $p \times r$ matrix $\boldsymbol{\Upsilon}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\upsilon}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ exists and it is a continuous function of $\boldsymbol{\theta}$ with rank r. Then, assuming the notation and conditions of the previous sections, we have the following theorem.

Theorem 4.3. *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 $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Define $\widetilde{\boldsymbol{\Sigma}}_\tau(\boldsymbol{\theta}_0) = [\boldsymbol{J}_\tau{}^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)]\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)[\boldsymbol{J}_\tau{}^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)]$. Then the asymptotic null distribution of the DPDTS$_C$ $S_\gamma(\boldsymbol{\theta}_n^\tau, \widetilde{\boldsymbol{\theta}}_n^\tau)$ coincides with 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 normal variables and $\widetilde{\zeta_1^{\gamma,\tau}}(\boldsymbol{\theta}_0)$, $\ldots$, $\widetilde{\zeta_r^{\gamma,\tau}}(\boldsymbol{\theta}_0)$ are the nonzero eigenvalues of $\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\widetilde{\boldsymbol{\Sigma}}_\tau(\boldsymbol{\theta}_0)$.*

Note that, we can find approximate critical values of the above asymptotic null distribution from the discussions in Basu et al. (2013a,b). In the next theorem, 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 explained 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

$$\sqrt{n}\begin{pmatrix} \boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}^* \\ \widetilde{\boldsymbol{\theta}}_n^\tau - \boldsymbol{\theta}_0 \end{pmatrix} \rightarrow N_{2p}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_\tau(\boldsymbol{\theta}^*) & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{12}^T & \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)\boldsymbol{P}_\tau(\boldsymbol{\theta}_0) \end{bmatrix}\right),$$

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

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

We then have the next theorem.

Theorem 4.4. *Suppose the model density satisfies the Lehmann and Ghosh-Basu conditions and take any $\boldsymbol{\theta}^* \notin \Theta_0$ for which (C1), (C4) and (C5) hold. Then, an approximation to the power function of the DPDTS$_C$ for testing (12) 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 4.5. *For any $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$, the probability of rejecting H_0 in (12) at level $\alpha > 0$ based on the DPDTS$_C$ tends to 1 as $n \rightarrow \infty$, provided $\frac{1}{n}\sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}^*), f_i(.;\boldsymbol{\theta}_0)) = O(1)$. So the proposed test is consistent.*

4.3 Robustness Properties of the Test

We again start with the IF of the DPDTS$_C$ to study its robustness properties. Using the functional form of $\boldsymbol{\theta}_n^\tau$ and $\widetilde{\boldsymbol{\theta}}_n^\tau$ and ignoring the multiplier 2 in our test statistic, we define the functional corresponding to the DPDTS$_C$ as

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

Clearly, the test functional depends on the sample size n implying the same dependency in its IF. Consider the contaminated distribution $G_{i,\epsilon}$ as defined in Section 3.2.1 and assume the contamination to be only in one fixed direction-i_0. Then the first order IF of $S^{(1)}_{\gamma,\tau}(\underline{\mathbf{G}})$ under this set-up is given by

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

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 as in Section 4.1. 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 $\boldsymbol{M}^{(i)}_{j,\gamma}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) = \mathbf{0}$ 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 $S^{(1)}_{\gamma,\tau}$ is given by

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

where $\boldsymbol{D}_{\tau,i_0}(t_{i_0}, \boldsymbol{\theta}_0) = \left[IF_{i_0}(t_{i_0}, \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) - IF_{i_0}(t_{i_0}, \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})\right]$. 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 the 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

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

where $\boldsymbol{D}_{\tau,o}(\mathbf{t}, \boldsymbol{\theta}_0) = \left[IF_o(\mathbf{t}, \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) - IF_o(\mathbf{t}, \widetilde{\boldsymbol{U}}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})\right]$. Again this IF behaves similarly as in the previous case implying the robustness for $\tau > 0$.

We can also derive the level and power influence functions of the proposed test for the composite hypothesis also as done in Section 3.2.2 for the simple null. Following a referee's comment its details have been deferred to the online supplement (Section 2). In summary, both the LIF and the PIF are seen to be bounded whenever the IFs of the MDPDE under the null and overall parameter space are bounded. But this is the case for most statistical models at $\tau > 0$, implying the size and power robustness of the corresponding DPDTS$_C$.

5 Application (I): Normal Linear Regression

Possibly the simplest (but extremely important) area of application for the proposed theory is the linear regression model with normally distributed error, and fixed covariates. Consider the linear regression model

$$y_i = \boldsymbol{x}_i{}^T\boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n, \tag{19}$$

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 y_i's are independent but not identically distributed. The MDPDEs of $\boldsymbol{\beta}$ and σ^2 and their properties are described in Section 3.1 of the online supplement.

5.1 Testing for the regression coefficients with known σ

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

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0, \qquad \text{against} \qquad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0, \tag{20}$$

for some pre-specified $\boldsymbol{\beta}_0(= \boldsymbol{\theta}_0)$. The treatment of unknown σ case with generalized linear hypotheses has been presented in Section 3.3 of the Online Supplement for brevity.

Here we refer to Section 3 and consider the test statistics $T_\gamma(\boldsymbol{\beta}_n^\tau, \boldsymbol{\beta}_0)$ for testing (20), where $\boldsymbol{\beta}_n^\tau$ is the MDPDE of $\boldsymbol{\beta}$ with tuning parameter τ and known $\sigma = \sigma_0$. Using the form of the normal density, we get

$$\begin{aligned} T_\gamma(\boldsymbol{\beta}_n^\tau, \boldsymbol{\beta}_0) &= \frac{2\sqrt{1+\gamma}}{\gamma(\sqrt{2\pi}\sigma_0)^\gamma}\left[n - \sum_{i=1}^n e^{-\frac{\gamma(\boldsymbol{\beta}_n^\tau - \boldsymbol{\beta}_0)^T(\boldsymbol{x}_i\boldsymbol{x}_i{}^T)(\boldsymbol{\beta}_n^\tau - \boldsymbol{\beta}_0)}{2(\gamma(\sigma_n^\tau)^2 + \sigma_0^2)}}\right], \text{ if } \gamma > 0, \\ \text{and } T_0(\boldsymbol{\beta}_n^\tau, \boldsymbol{\beta}_0) &= \frac{(\boldsymbol{\beta}_n^\tau - \boldsymbol{\beta}_0)^T(\boldsymbol{X}^T\boldsymbol{X})(\boldsymbol{\beta}_n^\tau - \boldsymbol{\beta}_0)}{\sigma_0^2}, \end{aligned}$$

where $\boldsymbol{X} = [\boldsymbol{x}_1 \cdots \boldsymbol{x}_n]^T$. Note that the estimator $\boldsymbol{\beta}_n^{(0)}$, the MDPDE with $\tau = 0$, is indeed the MLE of $\boldsymbol{\beta}$. Also the usual LRT statistics for this problem is defined by $-2\log\left[\frac{\prod_{i=1}^n \phi(y_i, \boldsymbol{x}_i{}^T\boldsymbol{\beta}_0, \sigma_0)}{\prod_{i=1}^n \phi(y_i; \boldsymbol{x}_i{}^T\boldsymbol{\beta}_n^{(0)}, \sigma_0)}\right]$; after simplification, this statistics turns out to be exactly 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

Assume Conditions (R1)–(R2) of Ghosh and Basu (2013), also presented in Section 1 of the online supplement, hold true and also assume

(C6) $\frac{1}{n}(\boldsymbol{X}^T\boldsymbol{X})$ 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 $\zeta_\tau = (2\pi)^{-\frac{\tau}{2}}\sigma^{-(\tau+2)}(1+\tau)^{-\frac{3}{2}}$.

Now, Theorem 3.1 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 (chi-square distribution with p degrees of freedom) with the multiplier being $\zeta_1^{\gamma,\tau} = (\sqrt{2\pi}\sigma_0)^{-\gamma}(1+\gamma)^{-\frac{1}{2}}\left(1+\frac{\tau^2}{1+2\tau}\right)^{\frac{3}{2}}$. So, the critical region for testing (20) at the significance level α is given by

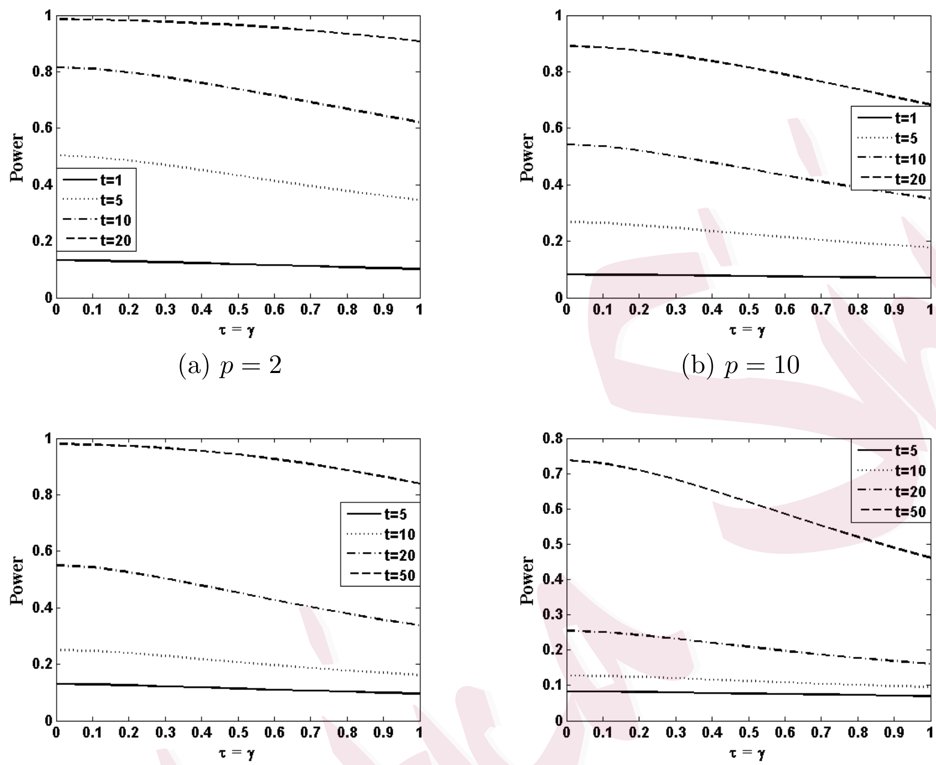
$$\left\{T_\gamma(\boldsymbol{\beta}_n^\tau, \boldsymbol{\beta}_0) > \zeta_1^{\gamma,\tau}\chi_{p,\alpha}^2\right\},$$

where $\chi_{p,\alpha}^2$ is the $(1-\alpha)$-th quantile of the χ_p^2 distribution. Further, at $\gamma = \tau = 0$, we have $\zeta_1^{0,0} = 1$ so that $T_0(\boldsymbol{\theta}_n^{(0)}, \boldsymbol{\theta}_0)$ follows asymptotically a χ_p^2 distribution under H_0, as expected from its relation to the LRT.

Next we study the performance of the proposed test under pure data through its asymptotic power. However, its asymptotic power against any fixed alternative will be one due to its consistency. So, we derive its asymptotic power under the contiguous alternatives $H_{1,n}$ using Corollary 3.5. Note that 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 $\delta = \frac{1}{\upsilon_\tau^{\boldsymbol{\beta}}}\boldsymbol{\Delta}^T\boldsymbol{\Sigma}_x\boldsymbol{\Delta}$, where $\upsilon_\tau^{\boldsymbol{\beta}} = \sigma^2\left(1+\frac{\tau^2}{1+2\tau}\right)^{\frac{3}{2}}$. Thus its asymptotic contiguous power turns out to be

$$P_{\tau,\gamma}(\boldsymbol{\Delta}, 0; \alpha) = P\left(\zeta_1^{\gamma,\tau}W_{p,\delta} > \zeta_1^{\gamma,\tau}\chi_{p,\alpha}^2\right) = 1 - G_{p,\delta}(\chi_{p,\alpha}^2),$$

where $G_{p,\delta}$ denote the distribution function of $\chi_{p,\delta}^2$. Figure 1 shows the nature of this asymptotic power over the tuning parameters $\gamma = \tau$ for different values of $\boldsymbol{\Delta}^T\boldsymbol{\Sigma}_x\boldsymbol{\Delta}$ ($= t$, say). Note that this asymptotic contiguous power does not depend on the tuning parameter γ and so we have taken $\gamma = \tau$ for illustrations here. Also, this contiguous power is clearly seen to depend on the distance ($\boldsymbol{\Delta}$) of the contiguous alternatives from null and the limiting second order moments ($\boldsymbol{\Sigma}_x$) of the covariates through $t = \boldsymbol{\Delta}^T\boldsymbol{\Sigma}_x\boldsymbol{\Delta}$; for any fixed $\tau = \gamma$ it increases as the value of t increases. Further this asymptotic power also depends on the number (p) of explanatory variables used in the regression. In Figure 1, we have shown the case of small values of p such as 2 and 10 as well as its large values $p = 50, 200$. Again for any fixed values of $\gamma = \tau$ and $\boldsymbol{\Delta}_1^T\boldsymbol{\Sigma}_x\boldsymbol{\Delta}_1$, the power decreases as p increases; this is expected as more the number of explanatory variables more is the number of component of $\boldsymbol{\beta}$ to be tested and so it becomes more difficult to test any fixed departure from null. Finally the asymptotic power against any contiguous alternative and any model is seen to decrease slightly with increasing values of τ; however the extent of this loss is not significant at moderate values of τ. Notice that this dependence of asymptotic power on τ is exactly similar to that of the asymptotic relative efficiency of the MDPDE $\boldsymbol{\beta}_n^\tau$ used in the test statistics. This is because the asymptotic power depends on τ through the non-centrality parameter δ and hence trough the asymptotic variance $\upsilon_\tau^{\boldsymbol{\beta}}$ of each element of $(\boldsymbol{X}^T\boldsymbol{X})^{1/2}\boldsymbol{\beta}_n^\tau$. As τ increases, the variance $\upsilon_\tau^{\boldsymbol{\beta}}$ increases slightly and hence both the efficiency of the MDPDE and the asymptotic contiguous power of the DPDTS decreases slightly. This is the price we pay for gaining additional robustness of the procedures as τ increases and is common in robustness literature.

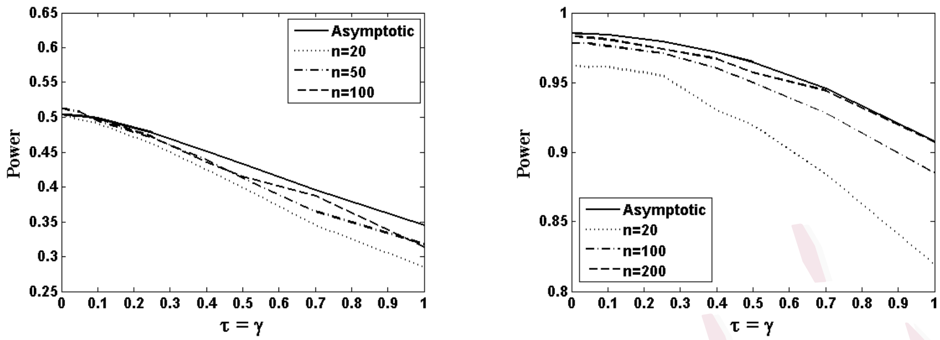


(a) $p = 2$ (b) $p = 10$

(c) $p = 50$ (d) $p = 200$

Figure 1: Asymptotic contiguous power of simple DPD based test of $\boldsymbol{\beta}$ for different values of $t = \boldsymbol{\Delta}^T \boldsymbol{\Sigma}_x \boldsymbol{\Delta}$ and p, the number of explanatory variables

Further, in order to examine how close are these asymptotic power functions to the corresponding empirical power curves in finite-sample situations, we have performed suitable simulations for different sample sizes n and different choices of $t = \boldsymbol{\Delta}^T \boldsymbol{\Sigma}_x \boldsymbol{\Delta}$ and p. Note that, this convergence of the finite-sample power to asymptotic value depends on the convergence rate of $\frac{1}{n}(\boldsymbol{X}^T \boldsymbol{X})$ in Condition (C6); here we have taken X following a p-variate normal distribution with mean 0 and covariance matrix $\sigma_x^2 \boldsymbol{I}_p$ so that (C6) holds with $\boldsymbol{\Sigma}_x = \sigma_x^2 \boldsymbol{I}_p$. For brevity, results from two such simulations with $p = 2$ and $\sigma_x^2 = 5$ are presented in Figure 2; other cases have the similar pattern. Clearly the finite-sample powers are quite close to the asymptotic power in moderate sample sizes like $n = 100$ and the convergence rate becomes little slower for larger values of $t = \boldsymbol{\Delta}^T \boldsymbol{\Sigma}_x \boldsymbol{\Delta}$.



(a) $\mathbf{\Delta}^T\mathbf{\Sigma}_x\mathbf{\Delta} = 5$ (b) $\mathbf{\Delta}^T\mathbf{\Sigma}_x\mathbf{\Delta} = 20$

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

5.1.2 Robustness Results

We study the robustness of the proposed tests under contamination through the influence function analysis as developed in Section 3.2. Since the first order IF of DPDTS $T_\gamma(\boldsymbol{\beta}_n^\tau, \boldsymbol{\beta}_0)$ is zero at any simple null hypothesis, we measure its stability by the second order IF. In particular, considering contamination in only one direction (i_0^{th} direction), the second order IF at the null hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ simplifies to

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

Clearly, the IF depends on the outliers and the leverage points through $(t_{i_0} - \boldsymbol{x}_{i_0}{}^T\boldsymbol{\beta}_0)$ and $[\boldsymbol{x}_{i_0}{}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_{i_0}]$, as expected from our intuition.

Also, the PIF of the proposed DPDTS under contiguous alternatives can be obtained from Theorem 3.7 and is presented in Section 3.2 of the online supplement. One can check that this PIF depends on the contamination points t_i only through $(t_i - \boldsymbol{x}_i^T\boldsymbol{\beta}_0)$.

Therefore, 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. But, both of them are unbounded at t_{i_0} for the proposed test with $\gamma = \tau = 0$, which is also the LRT statistic, indicating the non-robustness of the LRT. Further, interestingly the LIF of this test turns out to be identically zero for all $\tau, \gamma \geq 0$ (see Section 3.2 of the online supplement) implying no asymptotic influence of contiguous contamination on its size.

6 Application (II): Generalized Linear Model

Generalized linear models (GLMs) are generalizations of the normal linear regression model where the response variables Y_i are independent and assumed to follow a general exponential family distribution having density

$$f(y_i;\theta_i,\phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right\}; \tag{21}$$

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 $\eta_i = \boldsymbol{x}_i{}^T\boldsymbol{\beta}$. This general structure has a wide scope of application and includes normal linear regression, Poisson regression and logistic regression as special cases.

Clearly, the GLMs with fixed predictors consist one major subclass of the general I-NH set-up. The properties of the MDPDEs of $\boldsymbol{\theta} = (\boldsymbol{\beta},\phi)$ in the GLM was derived in Ghosh and Basu (2016) and is also presented in Section 4 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 we want to test for the hypothesis

$$H_0 : \boldsymbol{L}^T\boldsymbol{\beta} = \boldsymbol{l}_0 \qquad \text{against} \qquad H_1 : \boldsymbol{L}^T\boldsymbol{\beta} \neq \boldsymbol{l}_0, \tag{22}$$

where $\boldsymbol{L}$ is a $p\times r$ known matrix with $p \geq r$ and $\boldsymbol{l}_0$ is an r-vector of reals. Here, we assume that the nuisance parameter ϕ is unknown to us; the case of known ϕ can be derived easily from the general case.

The DPD based test statistics (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.

In order to derive the asymptotic distribution of the RMDPDE $(\widetilde{\boldsymbol{\beta}_n}^\tau, \widetilde{\phi_n}^\tau)$ of $(\boldsymbol{\beta},\phi)$ from Theorem 4.1, some simple matrix algebra leads us to

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

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 $\boldsymbol{\Gamma}_{ij}^{(\tau)}$ as defined in Section 4 of the online Supplement and $\boldsymbol{M}_{11} = (\boldsymbol{X}^T\boldsymbol{\Gamma}_{11}^{(\tau)}\boldsymbol{X})^{-1}$.

Corollary 6.1. *Suppose the Ghosh-Basu conditions hold with respect to Θ_0. Then, the RMDPDE $(\widetilde{\boldsymbol{\beta}_n}, \widetilde{\phi_n})$ exists and are consistent for $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}^g, \phi^g)$, true parameter value under Θ_0. Also, the asymptotic distribution of $\boldsymbol{\Omega}_n^{-\frac{1}{2}}\boldsymbol{P}_n\left[\sqrt{n}\left((\widetilde{\boldsymbol{\beta}_n}, \widetilde{\phi_n}) - (\widetilde{\boldsymbol{\beta}^g}, \widetilde{\phi^g})\right)\right]$ 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 $\boldsymbol{\Omega}_n(\boldsymbol{\beta},\phi)$ as defined in Section 4 of the online supplement.*

As in the case of unrestricted MDPDE, the restricted MDPDE of $\boldsymbol{\beta}$ and ϕ are also not always asymptotically independent. They will be independent if $\gamma_{12i}^{1+2\tau} = 0$ and $\gamma_{1i}^{1+\tau}\gamma_{2i}^{1+\tau} = 0$ for all i; the same conditions as in the unrestricted MDPDE (see Section 4 of the online supplement) and hold true for the normal regression model.

Next, to derive asymptotic distribution of the DPDTS$_C$ we assume the fixed covariates $\boldsymbol{x}_i$s to be such that the matrices $\boldsymbol{\Psi}_n^\tau(\widetilde{\boldsymbol{\theta}}^g)$ and $\boldsymbol{\Omega}_n^\tau(\widetilde{\boldsymbol{\theta}}^g)$, as defined in Section 4 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 their partition as

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

where $\boldsymbol{J}_{11,\tau}$ and $\boldsymbol{V}_{11,\tau}$ are of order $p \times p$. We will generally suppress τ in above notations 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 4.3 provided the Ghosh-Basu conditions hold for the model under H_0.

Corollary 6.2. *Consider the above mentioned set-up of GLM and 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, $\zeta_1^{\gamma,\tau}(\boldsymbol{\theta}_0), \cdots, \zeta_r^{\gamma,\tau}(\boldsymbol{\theta}_0)$ 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 null distribution helps us to obtain the critical values of the proposed DPD based test. All the other asymptotic results regarding power and robustness of the test can be derived by direct application of the general theory developed in Section 4; we will not report them again for brevity. We just report one robustness measure of the test, namely the second order IF of the test statistics at the null hypothesis, when there is contamination in only one fixed direction-i_0, as given by

$$IF_{i_0}^{(2)}(t_{i_0}, S_{\gamma,\tau}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = n(1+\gamma) \cdot \boldsymbol{W}^T\boldsymbol{\Psi}_n^\gamma\boldsymbol{W}, \tag{23}$$

$$\text{where} \qquad \boldsymbol{W} = \boldsymbol{\Psi}_n^{-1}\frac{1}{n}\begin{pmatrix} [f_{i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi))^\tau K_{1i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi)) - \gamma_{1i_0}]\,\boldsymbol{x}_i \\ f_{i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi))^\tau K_{2i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi)) - \gamma_{2i_0} \end{pmatrix}$$
$$- \boldsymbol{Q}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Psi}_n^{(0)}(\boldsymbol{\theta}_0)^T\frac{1}{n}\begin{pmatrix} f_{i_0}(t_{i_0}; \boldsymbol{\theta}_0)^\tau \boldsymbol{u}_{1i_0}^{(0)}(t_{i_0}; \boldsymbol{\theta}_0) - \boldsymbol{\gamma}_{1i_0}^{(0)} \\ f_{i_0}(t_{i_0}; \boldsymbol{\theta}_0)^\tau \boldsymbol{u}_{2i_0}^{(0)}(t_{i_0}; \boldsymbol{\theta}_0) - \boldsymbol{\gamma}_{2i_0}^{(0)} \end{pmatrix},$$

with $K_{ji_0}(t_{i_0}; (\boldsymbol{\beta}, \phi))$ being as defined in Section 4 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 6.1 (*Testing for the first r components of $\boldsymbol{\beta}$*)**.** Consider the hypothesis to test for the first r components ($r \leq p$) of the regression coefficient $\boldsymbol{\beta}$ at a pre-fixed value $\boldsymbol{\beta}_0^{(1)}$. In the particular case $r = 1$, it reduces to the problem of testing significance of individual components of $\boldsymbol{\beta}$. Here the null hypothesis to be tested is given by (22) with $\boldsymbol{L} = \begin{bmatrix} \boldsymbol{I}_r \\ \boldsymbol{O}_{(p-r)\times r} \end{bmatrix}$.

Let us 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 $\boldsymbol{X}$. Also, consider

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

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

In this particular case, the asymptotic distribution of the DPD based test statistics $S_\gamma(\boldsymbol{\theta}_n^\tau, \widetilde{\boldsymbol{\theta}}_n^\tau)$ under the null is given by 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, $\zeta_1^{\gamma,\tau}(\boldsymbol{\theta}_0), \cdots, \zeta_r^{\gamma,\tau}(\boldsymbol{\theta}_0)$ are r nonzero eigenvalues of the matrix $(1+\gamma)\boldsymbol{J}_{11,\gamma}^{11} \boldsymbol{J}_{11.2}^{-11} \boldsymbol{V}_{11}^{11} \boldsymbol{J}_{11.2}^{-11}$.

Further the second order IF of the DPDTS$_C$ can be obtained by using

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

Clearly, there is no influence of contamination on the first r components of the restricted MDPDE; this is expected as those r components are pre-fixed under null. Then, the second order IF of the DPDTS$_C$ follows from expression (23) with the simple form of $\boldsymbol{W}$ as above. □

7 Numerical Illustrations

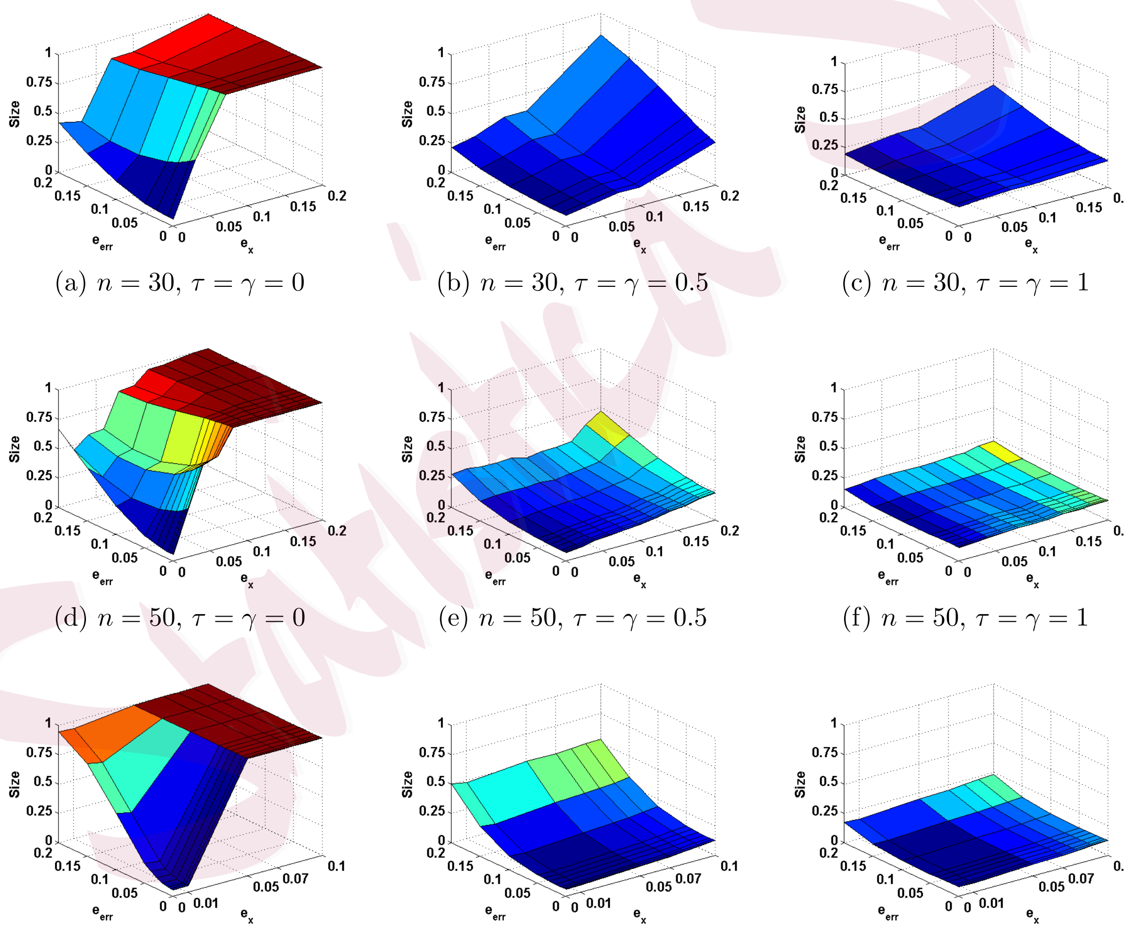
To examine the performance of the proposed tests in small or moderate samples, we have performed several simulation studies and applied them to analyze several interesting real data sets. For brevity, only one simulation result and one interesting real example for the linear regression model are presented here; some more simulation results and real data examples (including one each from Poisson and logistic regression set-up) are presented in the online supplement.

7.1 A Simulation Study: Simple Linear Regression

We consider the simple linear regression model with only one predictor and intercept as $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where ϵ_is $(i = 1, \ldots, n)$ are i.i.d. normal with mean 0 and variance

σ^2. Note that, in case of hypothesis testing there can be two types of adverse effect of the outliers — one is to reject the null due to the contamination effect although it is correct under pure data; the second one is to accept the null through the influence of contamination although it would have been rejected in the absence of contamination. The first one affects the size of the test whereas second one affects its power.

Here, we will compute the empirical size and power of the proposed DPD based test for testing $H_0 : (\beta_0,\ \beta_1) = (\beta_0^g,\ \beta_1^g)$. In our simulation, we have created different contamination scenarios by introducing e_x% contamination in predictor values and e_{err}% contamination in the error component. However, due to the different objective and effects, the simulation scheme are decided differently for studying the stability of size and power of the test; details of the simulation scheme are presented in Section 6.1 of online supplement.



(a) $n = 30$, $\tau = \gamma = 0$ (b) $n = 30$, $\tau = \gamma = 0.5$ (c) $n = 30$, $\tau = \gamma = 1$

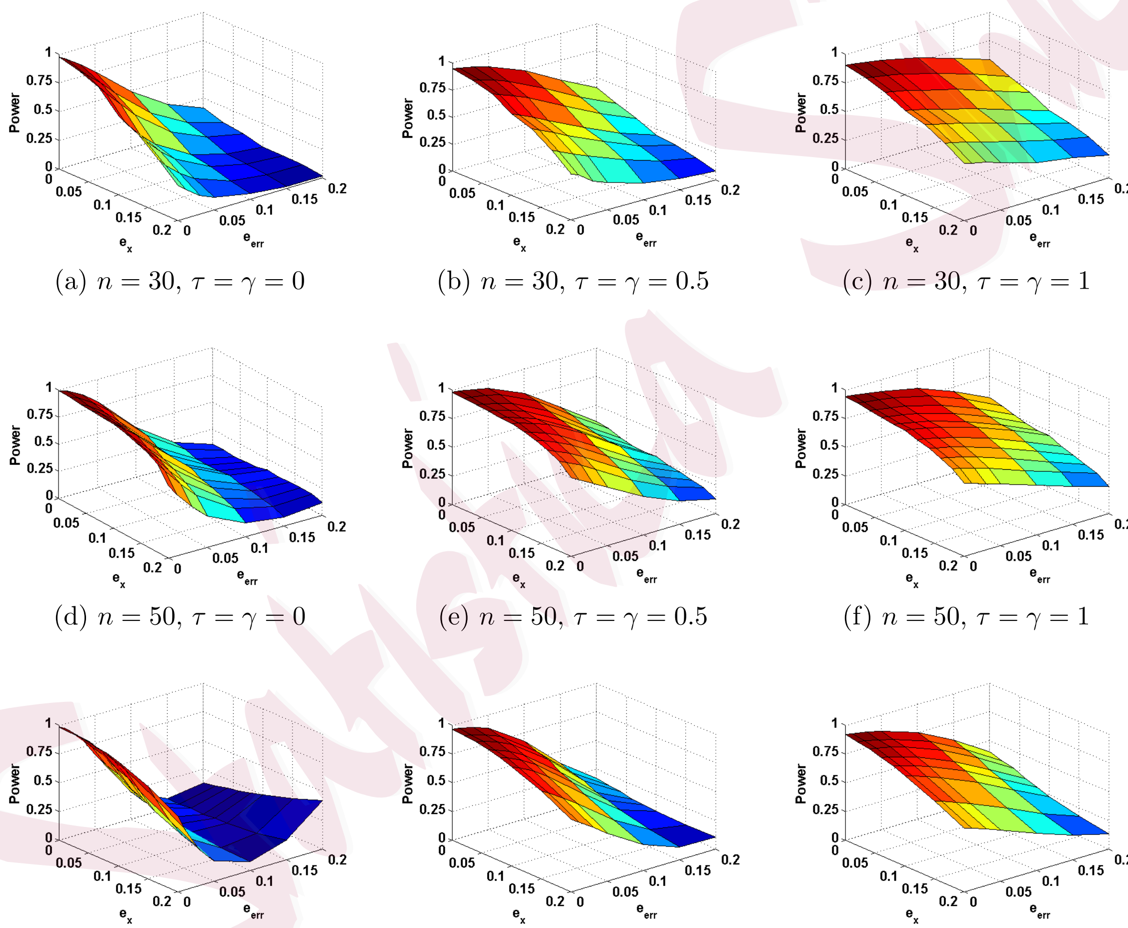
(d) $n = 50$, $\tau = \gamma = 0$ (e) $n = 50$, $\tau = \gamma = 0.5$ (f) $n = 50$, $\tau = \gamma = 1$

(g) $n = 100$, $\tau = \gamma = 0$ (h) $n = 100$, $\tau = \gamma = 0.5$ (i) $n = 100$, $\tau = \gamma = 1$

Figure 3: Empirical size of the DPD based test of $\boldsymbol{\beta}$ with unknown σ for different sample size n and different tuning parameter $\tau = \gamma$

The simulations are done for both the cases – known σ (simple hypothesis) and unknown σ (composite hypothesis). Detailed results for the case with known σ are presented in Section 6.2 of the online supplement. Here, for brevity, we present the surface plots of empirical sizes and powers over the contamination on both the predictor x and error ϵ in Figures 3 and 4 respectively for only three choices of the tuning parameter $\tau = \gamma = 0$ (equivalent to LRT), 0.5 and 1;

Clearly the LRT corresponding to $\tau = \gamma = 0$ is highly unstable with respect to both its size and power even in the presence of a fairly small contamination. However, the DPD based test with larger values of $\tau = \gamma$ are still robust again any kind of contamination in the data.



(a) $n = 30$, $\tau = \gamma = 0$ (b) $n = 30$, $\tau = \gamma = 0.5$ (c) $n = 30$, $\tau = \gamma = 1$

(d) $n = 50$, $\tau = \gamma = 0$ (e) $n = 50$, $\tau = \gamma = 0.5$ (f) $n = 50$, $\tau = \gamma = 1$

(g) $n = 100$, $\tau = \gamma = 0$ (h) $n = 100$, $\tau = \gamma = 0.5$ (i) $n = 100$, $\tau = \gamma = 1$

Figure 4: Empirical power of the DPD based test of $\boldsymbol{\beta}$ with unknown σ for different sample power n and different tuning parameter $\tau = \gamma$

Note that, the critical value (or critical region) of our proposed test has been obtained under the assumption of the data being free from contamination, since, for any practical application, we don't have any prior idea on the level of contamination in our data. However, as seen through simulations, use of these critical values inflates the sizes and reduces the powers of the proposed tests under serious contamination, as expected. Although these effects become less significant as the tuning parameters $\tau = \gamma$ increases, it would be interesting to study the theoretical level under some assumptions about data contamination such as heavy tail symmetric contamination etc., if known a priori, and correct the inflated type I and type II errors accordingly. A direct way to tackle this problem is through the asymptotic distribution of the test statistics under contaminated models. We hope to take it up in our future research work.

7.2 A Real Data Example: Salinity Data

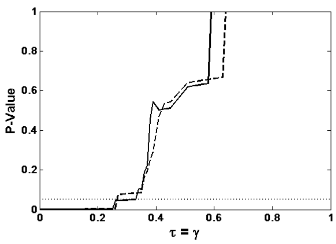
We consider an example of the multiple regression model through the popular "Salinity data" (Rousseeuw and Leroy, 1987, Table 5, Chapter 2), containing measurements of water's salt concentration and river discharge taken in North Carolina's Pamlico Sound. This dataset was originally discussed in Ruppert and Carroll (1980) and later examined by many authors including Rousseeuw and Leroy (1987) and Ghosh and Basu (2013). The analysis presented here shows, except two potential outliers, the data set can be modeled well by a multiple linear model, by 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 in the dataset contain are two outlying observations that corresponds to periods of very heavy discharge (Ruppert and Carroll, 1980).

The maximum likelihood estimator of the regression coefficient $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ and the error SD σ for the full data are $(9.6, 0.8, -0.03, -0.3)^T$ and 1.23. However, after deleting the 5th and 16th observations these estimators changes to $(23.39, 0.70, -0.25, -0.84)^T$ and 0.91 respectively, which indicates the dramatic effect of outliers. Ghosh and Basu (2013) illustrated that the MDPDE with $\tau \geq 0.25$ can successfully generate robust estimators even under presence of the two outlying 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.

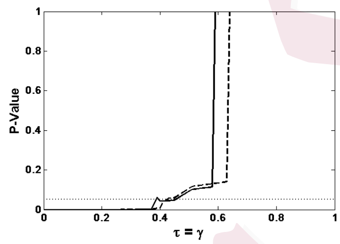
Here, we apply the proposed DPD based test using the full data and also after deleting the outlier from data. We test for several hypotheses on $\boldsymbol{\beta}$ as $H_0 : \boldsymbol{\beta} = (19.19,\ 0.71,\ -0.18,\ -0.66)^T$, $H_0 : \boldsymbol{\beta} = (18.4, 0.72, -0.2, -0.63)^T$ and $H_0 : \boldsymbol{\beta} = (9.6, 0.8, -0.03, -0.3)^T$. Note that these null hypotheses are 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 one. We have considered both the simple and composite hypotheses testing proposed in the paper by first assuming σ to be known and unknown respectively. Again, for the known σ case, we have assumed two distinct values of σ, namely 1.23 (a non-robust estimate, MLE) and 0.71 (a robust estimate, MDPDE at $\tau = 1$). The p-values obtained by applying the proposed DPD based tests for all these cases are presented in Figure 5.

Note that, whenever σ is assumed to be unknown, the DPD based tests with $\tau =$

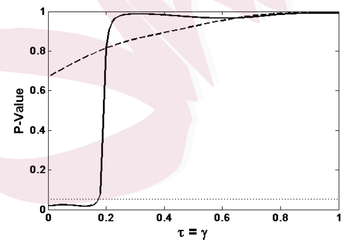
$\gamma \geq 0.2$ give robust results by accepting the first two hypotheses (Figures 5c, 5f) and rejecting the third one (Figure 5i) under full data. The performances of the LRT at $\tau = \gamma = 0$ is clearly non-robust under full data and hence the inferences change after deleting the outliers. When specifying σ by a robust estimator, under full the DPD based tests still accept the first two hypothesis at larger $\tau = \gamma \geq 0.5$ but LRT rejects them (Figures 5b, 5e). However, all the DPD based tests including the LRT successfully rejects the third hypotheses under full data for correctly specified robust σ (Figure 5h). Finally, when σ is incorrectly specified, the DPD based tests at $\tau = \gamma \geq 0.5$ still lead to the robust inferences while the LRT generates incorrect inferences for the first two hypotheses (Figures 5a, 5d); but the third hypotheses gets accepted by the DPD based tests at larger $\tau = \gamma$ due to the incorrect specification of σ (Figure 5g).



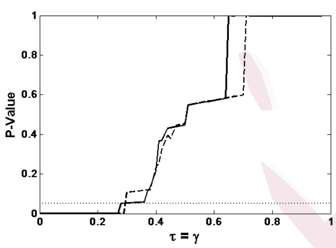
(a) $H_0 : \boldsymbol{\beta} = (19.19, 0.71, -0.18, -0.66)$ ($\sigma = 1.23$ known)



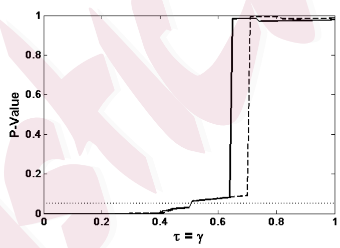
(b) $H_0 : \boldsymbol{\beta} = (19.19, 0.71, -0.18, -0.66)$ ($\sigma = 0.71$ known)



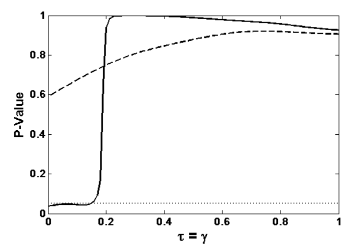
(c) $H_0 : \boldsymbol{\beta} = (19.19, 0.71, -0.18, -0.66)$ (σ unknown)



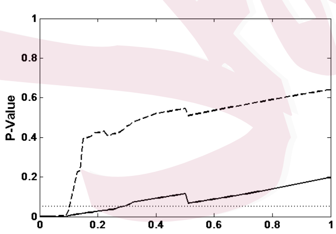
(d) $H_0 : \boldsymbol{\beta} = (18.4, 0.72, -0.2, -0.63)$ ($\sigma = 1.23$ known)



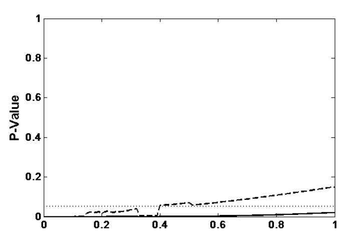
(e) $H_0 : \boldsymbol{\beta} = (18.4, 0.72, -0.2, -0.63)$ ($\sigma = 0.71$ known)



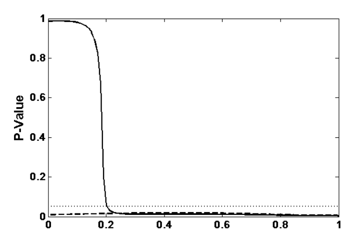
(f) $H_0 : \boldsymbol{\beta} = (18.4, 0.72, -0.2, -0.63)$ (σ unknown)



(g) $H_0 : \boldsymbol{\beta} = (9.6, 0.8, -0.03, -0.3)$ ($\sigma = 1.23$ known)



(h) $H_0 : \boldsymbol{\beta} = (9.6, 0.8, -0.03, -0.3)$ ($\sigma = 0.71$ known)



(i) $H_0 : \boldsymbol{\beta} = (9.6, 0.8, -0.03, -0.3)$ (σ unknown)

Figure 5: P-Values of the DPD based tests for different H_0 on $\boldsymbol{\beta}$ with known and unknown σ^2 for the Salinity data (Here, solid line - full data; dashed line - outlier deleted data)

8 On the Competitive Choice of Tuning parameters and Test Statistics

In defining the DPD test statistics we have tried to keep the method as general as possible in terms of the tuning parameters of the test statistics. As such the asymptotic distribution of the test statistics has been defined as a function of two independent tuning parameters τ and γ. In practice one could consider the totality of all tests obtained by varying the two different tuning parameters. In the theoretical sense we have done exactly that in this paper. Investigating all such tests numerically is, however, a huge task and for the present we have restricted our numerical investigations to the case where $\gamma = \tau$. We hope to further extend our numerical evaluation of this family of test statistics in the future by choosing distinct values for τ and γ. In particular, it may be of interest to observe the situation where $\gamma = 0$ and $\tau > 0$ so that we have an idea about the performance of the likelihood ratio statistics evaluated at a robust estimator.

Here we have examined the performances of the proposed DPD based test statistics at $\tau = \gamma$ through several theoretical results and numerical illustrations for the linear regression model and the GLMs. We have seen that the power of the proposed test against the contiguous alternative under pure data is asymptotically independent of γ and decreases slightly with increasing values of the parameters τ; but the loss in power is not significant even for $\tau = 0.5$. On the other hand the robustness of the proposed test under contamination, both in terms of its size and power, increase as $\tau = \gamma$ increases. So, we need to choose the tunning parameters $\tau = \gamma$ suitably to make a trade-off between these two.

In this respect, it is useful to note that the robustness properties of the proposed test depend mostly on the MDPDE of the parameter used through τ although the extent of robustness depends slightly on γ. However, we suggest to use $\gamma = \tau$ to make the test statistics compatible with the MDPDE used. So, it would be enough to choose the proper estimator with the optimal value of the parameter τ to be used in our test statistics. Ghosh and Basu (2015) has proposed one such approach of data-driven choice of the tunning parameter of the MDPDE in the context of I-NH set-up. The proposal had been successfully implemented in the case of linear regression and generalized linear models by Ghosh and Basu (2015) and Ghosh (2016) respectively. We have verified that the resulting choice of tunning parameter also provide us the desirable trade-off for the proposed testing procedures also. For example, the optimal choice of tunning parameter τ for the MDPDE under the Salinity Data-set had been seen to be $\tau = 0.5$ by Ghosh and Basu (2015). As we have seen above in Section 7.2, the choice of $\gamma = \tau = 0.5$ yields the robust inference for any kind of hypothesis for this data-set; also it has quite high power against the contiguous alternative under pure data which can be seen from Figure 1. Similar phenomenon also hold for other datasets presented in the online supplement. So, we suggest to choose the tunning parameters of the proposed testing procedures by means of the Ghosh and Basu (2015) proposal.

Further, as we have seen in case of linear regression and GLMs, the proposed DPD based test for positive γ and τ are computationally no more complicated than the popular LRT (corresponding to the DPD based test with $\gamma = \tau = 0$) but gives us the extra advantage of stability in presence of the outlying observations at the cost of only a small power loss under pure data. This very strong property of the proposed test will build its equity against the existing asymptotic tests for the present set-up.

For a brief comparison with the existing literature, it is to be noted that we have proposed a class of robust tests under a complete general set-up of I-NH set-up and as per the knowledge of the authors there is no such general approach available. However, there are some particular approaches for particular cases like linear regression (Ronchetti and Rousseeuw, 1980; Schrader and Hettmansperger, 1980; Ronchetti, 1982a,b, 1987; Sen, 1982; Markatou and Hettmansperger, 1990; Markatou and He, 1994; Salibian-Barrera et al., 2014) and some GLMs (Morgenthaler, 1992; Cantoni and Ronchetti, 2001; Maronna et al., 2006; Wang and Qu, 2007; Hosseinian, 2009); but most of them assume the covariates to be stochastic while we consider the case of fixed covariates. Even if we can apply a robust test procedure with stochastic covariate heuristically in case of regression models with given fixed covariates, their properties will directly depend on the robust estimations of the regression coefficient used in construction of the test statistics. And, as is extensively studied in Ghosh and Basu (2013, 2016), the MDPDE of the regression coefficients has several advantages over the existing robust estimators and we expect the same to hold in case of the proposed MDPDE based tests too. However, this surely needs a lot more research and considering the length of the present paper, we have decided to leave extensive comparisons for the future.

9 Conclusions

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

Among other possible extensions, we hope to study the corresponding two sample (or multi-sample) problem in the future which could be of obvious interest in real situations. When we have two independent non-homogeneous data systems, we may want to know whether the involved parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the same or whether they differ (including, perhaps, the direction of difference). In the simplest case this would be akin to testing for the equality of the slopes of two (or possibly several) regression lines, but this could be useful in many other more complicated scenarios as well.

Another natural extension could be to the case of heteroscedastic models. Obviously choosing completely independent error variances for each observations can make the problem infeasible. However, under certain assumptions on the form of these error variances this may be a reasonable problem; for example, this may be applicable in the scenarios where the error variances are functions of the covariates and the common σ (or, a few real parameters).

Acknowledgment: The authors wish to thank the editor, the associate editor and two anonymous referees for their constructive comments and suggestions.

References

Basu, A., I. R. Harris, N. L. Hjort, and M. C. Jones (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika 85*, 549–559.

Basu, A., Mandal, A., Martin, N., and Pardo, L. (2013). Testing statistical hypotheses based on the density power divergence. *Annals of the Institute of Statistical Mathematics 65*, 319–348.

Basu, A., Mandal, A., Martin, N., and Pardo, L. (2013). Density Power Divergence Tests for Composite Null Hypotheses. *ArXiv pre-print*, arXiv:1403.0330 [stat.ME].

Basu, A., Shioya, H. and Park, C. (2011). *Statistical Inference: The Minimum Distance Approach*. Chapman & Hall/CRC. Boca Raton, Florida.

Cantoni, E., and Ronchetti, E. (2001). Bounded Influence for Generalized Linear Models. *Journal of the American Statistical Association 96*, 1022–1030.

Ghosh, A. (2015). Influence Function of the Restricted Minimum Divergence Estimators : A General Form. *Electronic Journal of statistics 9*, 1017–1040.

Ghosh, A. (2016). Divergence based Robust Estimation of Tail Index with Exponential Regression Model. *Statistical Methods and Applications*, to appear.

Ghosh, A. and Basu, A. (2013). Robust Estimation for Independent Non-Homogeneous Observations using Density Power Divergence with Applications to Linear Regression. *Electronic Journal of statistics 7*, 2420–2456.

Ghosh, A. and Basu, A. (2015). Robust Estimation for Non-Homogeneous Data and the Selection of the Optimal Tuning Parameter: The DPD Approach. *Journal of Applied Statistics*, 42(9), 2056–2072.

Ghosh, A. and Basu, A. (2016). Robust Estimation in Generalised Linear Models : The Density Power Divergence Approach. *TEST*, 25(2), 269–290.

Ghosh, A., Harris, I. R., Maji, A., Basu, A., Pardo, L. (2016). A Generalized Divergence for Statistical Inference. *Bernoulli*, to appear.

Hampel, F. R., E. Ronchetti, P. J. Rousseeuw, and W. Stahel (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. *Journal of the American Statistical Association 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.

Huber-Carol, C. (1970). *Etude asymptotique de tests robustes.* Ph. D. thesis, ETH, Zurich.

Kotz, S., Johnson, N. L., and Boyd, D. W. (1967a). Series representations of distributions of quadratic forms in normal variables. I. Central case. *Annals of Mathematical Statistics 38*, 823–837.

Kotz, S., Johnson, N. L., and Boyd, D. W. (1967b). Series representations of distributions of quadratic forms in normal variables. I. Non-central case. *Annals of Mathematical Statistics 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. *Annals of Statistics 22*, 1081–1114.

Liu, R. C., Markatou, M., and Tsai, C. L. (2005). Robust Estimation and Testing in Nonlinear Regression Models. *International Journal of Pure and Applied Mathematics 21*, 525–552.

Markatou, M., and He, X. (1994). Bounded-Influence and High-Breakdown-Point Testing Procedures in Linear Models. *Journal of the American Statistical Association 89*, 543–549.

Markatou, M., and Hettmansperger, T. P. (1990). Robust Bounded-Influence Tests in Linear Models. *Journal of the American Statistical Association 85*, 187–190.

Markatou, M., and Manos, G. (1996). Robust Tests in Nonlinear Regression Models. *Journal of Statistical Planning and 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.

Morgenthaler, S. (1992). Least-absolute-deviations fits for generalized linear models. *Biometrika 79(4)*, 747–754.

Muller, C. (1998). Optimum robust testing in linear models. *Annals of Statistics 26(3)*, 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. *Statistica 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*, Brighton, England, 210–211.

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

Rousseeuw, P. J. and Ronchetti, E. (1979). The influence curve for tests. Research Report 21, Fachgruppe für Statistik, ETH, Zurich.

Rousseeuw, P. J. and Ronchetti, E. (1981). Influence curves for general statistics. *J. Comput. Appl. Math. 7*, 161–166.

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. (2014). Robust tests for linear regression models based on τ-estimates. *Computational Statistics and Data Analysis*, doi:10.1016/j.csda.2014.09.012.

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. *Journal of the American Statistical Association 84*, 107–113.

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

Wang, L. and Qu, A. (2007). Robust Tests in Regression Models With Omnibus Alternatives and Bounded Influence. *Journal of the American Statistical Association 102*, 347–358.