Robust Bounded Influence Tests for Independent Non-Homogeneous Observations

Abhik Ghosh and Ayanendranath Basu

Indian Statistical Institute, Kolkata, India

Supplementary Material

S1 Required Assumptions

Ghosh-Basu Conditions: Assumptions (A1)–(A7) of Ghosh and Basu (2013):

Consider the general I-NH set-up as described in Section 1 of the main paper.

(A1) The support $\chi = \{y | f_i(y;\boldsymbol{\theta}) > 0\}$ is independent of i and $\boldsymbol{\theta}$; the true distributions G_i are also supported on χ for all i.

(A2) There is an open subset of ω of the parameter space Θ, containing the best fitting parameter $\boldsymbol{\theta}^g$ such that for almost all $y \in \chi$, and all $\boldsymbol{\theta} \in \Theta$, all $i = 1, 2, \ldots$, the density $f_i(y;\boldsymbol{\theta})$ is thrice differentiable with respect to $\boldsymbol{\theta}$ and the third partial derivatives are continuous with respect to $\boldsymbol{\theta}$.

(A3) For $i = 1, 2, \ldots$, the integrals $\int f_i(y;\boldsymbol{\theta})^{1+\tau} dy$ and $\int f_i(y;\boldsymbol{\theta})^{\tau} g_i(y) dy$ can be differentiated thrice with respect to $\boldsymbol{\theta}$, and the derivatives can be taken under the integral sign.

(A4) For each $i = 1, 2, \ldots$, the matrix $\boldsymbol{J}^{(i)}$, defined in Section 2 of the main paper, is positive definite. Also the matrix $\boldsymbol{\Psi}_n^\tau$, defined in Equation (6) of the main paper, satisfies

$$\lambda_0 = \inf_n \quad [\text{min eigenvalue of } \boldsymbol{\Psi}_n^\tau] > 0$$

(A5) For each $i = 1, 2, \ldots$, there exists a function $M_{jkl}^{(i)}(y)$ such that

$|\nabla_{jkl} V_i(y;\boldsymbol{\theta})| \leq M_{jkl}^{(i)}(y)$ for all $\boldsymbol{\theta} \in \Theta$, and $\dfrac{1}{n}\displaystyle\sum_{i=1}^n E_{g_i}\left[M_{jkl}^{(i)}(Y)\right] = O(1)$, for all j, k, l,

where $V_i(y;\boldsymbol{\theta})$ is as defined in Section 2 of the main paper.

(A6) For all j, k, we have

$$\lim_{N\to\infty}\sup_{n>1}\left\{\frac{1}{n}\sum_{i=1}^n E_{g_i}\left[|\nabla_j V_i(Y;\boldsymbol{\theta})| I(|\nabla_j V_i(Y;\boldsymbol{\theta})| > N)\right]\right\} = 0, \tag{S1.1}$$

and

$$\lim_{N\to\infty}\sup_{n>1}\left\{\frac{1}{n}\sum_{i=1}^n E_{g_i}\left[|\nabla_{jk} V_i(Y;\boldsymbol{\theta}) - E_{g_i}(\nabla_{jk} V_i(Y;\boldsymbol{\theta}))| \right.\right.$$
$$\left.\left. \times I(|\nabla_{jk} V_i(Y;\boldsymbol{\theta}) - E_{g_i}(\nabla_{jk} V_i(Y;\boldsymbol{\theta}))| > N)\right]\right\} = 0, \tag{S1.2}$$

where $I(B)$ denotes the indicator variable of the event B.

(A7) For all $\epsilon > 0$, we have

$$\lim_{n\to\infty}\left\{\frac{1}{n}\sum_{i=1}^n E_{g_i}\left[||\boldsymbol{\Omega}_n^\tau(\boldsymbol{\theta})^{-1/2}\nabla V_i(Y;\boldsymbol{\theta})||^2 I\left(||\boldsymbol{\Omega}_n^\tau(\boldsymbol{\theta})^{-1/2}\nabla V_i(Y;\boldsymbol{\theta})|| > \epsilon\sqrt{n}\right)\right]\right\} = 0, \tag{S1.3}$$

where $\boldsymbol{\Omega}_n^\tau(\boldsymbol{\theta})$ is as defined in Equation (6) of the main paper.

Assumptions (R1)–(R2) of Ghosh and Basu (2013):

Consider the set-up of normal linear regression model with fixed design as described in Section 5 of the main paper.

(R1) The values of $\boldsymbol{x}_i$'s are such that for all j, k, and l

$$\sup_{n>1} \max_{1\leq i\leq n} |x_{ij}| = O(1), \qquad \sup_{n>1} \max_{1\leq i\leq n} |x_{ij}x_{ik}| = O(1), \qquad \frac{1}{n}\sum_{i=1}^{n} |x_{ij}x_{ik}x_{il}| = O(1). \tag{S1.4}$$

(R2) The matrix $\boldsymbol{X}^T = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)_{p\times n}$ satisfies

$$\inf_n \ [\text{min eigenvalue of } \frac{(\boldsymbol{X}^T\boldsymbol{X})}{n}] > 0, \tag{S1.5}$$

which also implies that the matrix $\boldsymbol{X}$ has full column rank, and

$$n \max_{1\leq i\leq n} [\boldsymbol{x}_i{}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i] = O(1). \tag{S1.6}$$

Lehmann conditions (Lehmann, 1983):

(A) There is an open subset of ω of the parameter space Θ, containing the true parameter value $\boldsymbol{\theta}_0$ such that for almost all $x \in \mathcal{X}$, and all $\boldsymbol{\theta} \in \omega$, the density $f_i(x, \boldsymbol{\theta})$ is three times differentiable with respect to $\boldsymbol{\theta}$ for each i.

(B) For each i, the first and second logarithmic derivatives of $f_i(\cdot, \boldsymbol{\theta})$ satisfy the equations

$$E_{\boldsymbol{\theta}} \left[\nabla \log f_i(X, \boldsymbol{\theta})\right] = 0,$$

and

$$\boldsymbol{I}_i(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \left[(\nabla \log f_i(X, \boldsymbol{\theta}))(\nabla \log f_i(X, \boldsymbol{\theta}))^T\right] = E_{\boldsymbol{\theta}} \left[-\nabla^2 \log f_i(X, \boldsymbol{\theta})\right].$$

(C) The matrix $\boldsymbol{I}_i(\boldsymbol{\theta})$ is positive definite with all entries finite for all i and all $\boldsymbol{\theta} \in \omega$, and hence the components $(\nabla \log f_i(X, \boldsymbol{\theta}))$ are affinely independent with probability one.

(D) For all j, k, l, there exists functions M_{jkl} with finite expectation (under true distribution) such that for each i

$$|\nabla_{jkl} \log f_i(X, \boldsymbol{\theta})| \leq M_{jkl}(x), \qquad \text{for all } \boldsymbol{\theta} \in \omega.$$

S2 Background and Some Details on the Linear Regression Model with Fixed Covariates

S2.1 The MDPDE (Ghosh and Basu, 2013)

Ghosh and Basu (2013) extensively studied and formally established the properties of the minimum density power divergence estimators for normal regression model.

As this work is of critical importance to the present research, we describe, briefly, the main findings of that work.

Consider the normal linear regression model set up with given values of the explanatory variables as described in Section 5 of the main paper. The estimating equation for the MDPDE of the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2)^T$ in this regression model, corresponding to the DPD with tuning parameter τ, is given by

$$\sum_{i=1}^{n} x_{ij}(y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta})e^{-\frac{\tau(y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta})^2}{2\sigma^2}} = 0 \quad \forall j = 1, \ldots, p \tag{S2.7}$$

$$\sum_{i=1}^{n} \left[1 - \frac{(y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta})^2}{\sigma^2}\right] e^{-\frac{\tau(y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta})^2}{2\sigma^2}} = \frac{\tau}{(1+\tau)^{\frac{3}{2}}}. \tag{S2.8}$$

We can then obtain the MDPDE $\widehat{\boldsymbol{\theta}}^T = (\widehat{\boldsymbol{\beta}}^T, \widehat{\sigma}^2)$ of regression parameters by solving these estimating equations numerically over the parameter space. Ghosh and Basu (2013) proved the asymptotic properties of this MDPDE under some suitable assumptions [Assumptions (R1) and (R2) of their paper, also presented above in Section 1] on the given values of explanatory variables, which basically imply the boundedness of the $\boldsymbol{x}_i$s in large samples and that the spectrum of $(\boldsymbol{X}^T\boldsymbol{X})$ remains bounded away from zero.

With these Conditions (R1) and (R2), let us also assume that the true density generating the observed sample belongs to the assumed model family. Then, we have

(i) The minimum DPD estimating equations (S2.7) and (S2.8) have a consistent sequence of roots $\widehat{\boldsymbol{\theta}}^T = (\widehat{\boldsymbol{\beta}}^T, \widehat{\sigma}^2)$.

(ii) The estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$ are asymptotically independent.

(iii) $(\boldsymbol{X}^T\boldsymbol{X})^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ asymptotically follows a p-dimensional normal with (vector) mean zero and covariance matrix $\upsilon_\tau^{\boldsymbol{\beta}}\boldsymbol{I}_p$ and $\sqrt{n}(\widehat{\sigma}^2-\sigma^2)$ follows asymptotically a normal distribution with mean 0 and variance υ_τ^e, where $\upsilon_\tau^{\boldsymbol{\beta}}$ and υ_τ^e are as defined in Theorem 5.1 in the main paper.

This asymptotic results follows from the general theory of Ghosh and Basu (2013) discussed in Section 2 of the main paper. Since the MDPDE corresponding to $\tau = 0$ coincides with the MLEs, substituting $\tau = 0$ in the above expressions, we get back exactly the corresponding results for their MLE.

The robustness of the MDPDE of regression parameters can be observed from the boundedness of their influence functions. In fact, the influence function the MDPDE $\boldsymbol{U}_\tau^{\boldsymbol{\beta}}$ of the regression coefficient $\boldsymbol{\beta}$ at the model distribution $\underline{\mathbf{F}}_{\boldsymbol{\theta}}$ for contamination only in i_0-th data-point is given by

$$IF_{i_0}(t_{i_0},\boldsymbol{U}_\tau^{\boldsymbol{\beta}},\underline{\mathbf{F}}_{\boldsymbol{\theta}}) = (1+\tau)^{\frac{3}{2}}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_{i_0}(t_{i_0}-{\boldsymbol{x}_{i_0}}^T\boldsymbol{\beta})e^{-\frac{\tau(t_{i_0}-{\boldsymbol{x}_{i_0}}^T\boldsymbol{\beta})^2}{2\sigma^2}}, \tag{S2.9}$$

and the same for the MDPDE $\boldsymbol{U}_\tau^{\sigma}$ of the error variance σ^2 is given by

$$IF_{i_0}(t_{i_0},\boldsymbol{U}_\tau^{\sigma},\underline{\mathbf{F}}_{\boldsymbol{\theta}}) = \frac{2(1+\tau)^{\frac{5}{2}}}{n(2+\tau^2)}\left\{(t_{i_0}-{\boldsymbol{x}_{i_0}}^T\boldsymbol{\beta})^2-\sigma^2\right\}e^{-\frac{\tau(t_{i_0}-{\boldsymbol{x}_{i_0}}^T\boldsymbol{\beta})^2}{2\sigma^2}} + \frac{2\tau(1+\tau)^2}{n(2+\tau^2)}. \tag{S2.10}$$

Clearly, both the influence functions in (S2.9) and (S2.10) are bounded in t_{i_0} for all $\tau > 0$ and for any i_0 implying the robustness of the MDPDEs with $\tau > 0$. However the influence functions for the non-robust MLE (MDPDE with $\tau = 0$)

are clearly unbounded. See Ghosh and Basu (2013) for the influence function with contamination in all directions.

S2.2 LIF and PIF for testing for the regression coefficient with known σ

Consider the set-up and notations of Section 5.1 of the main paper. For the corresponding DPDTS, the PIF under contiguous alternatives can be obtained from Theorem 3.7 of the main paper. Some calculations simplify it to have the form

$$\begin{aligned} PIF(\mathbf{t}; T_{\gamma,\lambda}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) &= K_\tau^*\left(\boldsymbol{\Delta}^T\boldsymbol{\Sigma}_x\boldsymbol{\Delta}, p\right)\sum_{i=1}^n(\boldsymbol{\Delta}^T\boldsymbol{x}_i)(t_i - {\boldsymbol{x}_i}^T\boldsymbol{\beta}_0)e^{-\frac{\tau(t_i - {\boldsymbol{x}_i}^T\boldsymbol{\beta}_0)^2}{2\sigma_0^2}}. \\ \text{where} \quad K_\tau^*(s,p) &= (1+\tau)^{3/2}e^{-\frac{s}{2\upsilon_\tau^{\boldsymbol{\beta}}}}\sum_{k=0}^\infty\frac{(2k-s)\, s^{k-1}}{k!(2\upsilon_\tau^{\boldsymbol{\beta}})^k}P\left(Z_{p+2k} > \chi_{p,\alpha}^2\right). \end{aligned}$$

Note that this PIF depends on the contamination points t_is only through $(t_i - {\boldsymbol{x}_i}^T\boldsymbol{\beta}_0)$ and is bounded whenever $\tau > 0$ implying the power stability of the DPDTS. But, for $\gamma = \tau = 0$ the PIF simplifies to a linear function of t_is which is clearly unbounded, implying the non-robust nature of the LRT.

Further, substituting $\boldsymbol{\Delta} = \mathbf{0}$ in the PIF derived above, we get the LIF of the proposed DPDTS, which is identically zero for all $\gamma, \tau > 0$. This implies no influence of contiguous contamination on its asymptotic level.

S3 The MDPDE under Generalized Linear Model (GLM) with Fixed Covariates (Ghosh and Basu, 2016)

The minimum density power divergence estimator in case of the generalized linear regression model was explained in detail in Ghosh and Basu (2016). Suppose $(y_i; \boldsymbol{x}_i)$, $i = 1, \cdots, n$ denotes n independent observations from a generalized linear model with density given by Equation (24) of the main paper and mean μ_i depending on the given fixed values of $\boldsymbol{x}_i$'s. Then, following the approach of Section 2 of the main paper, the estimating equation of the parameters are given by

$$\sum_{i=1}^{n} \boldsymbol{x}_i \left[\int K_{1i}(y; (\boldsymbol{\beta}, \phi)) f_i(y; (\boldsymbol{\beta}, \phi))^{1+\tau} dy - K_{1i}(y_i; (\boldsymbol{\beta}, \phi)) f_i(y_i; (\boldsymbol{\beta}, \phi))^{\tau} \right] = 0, \tag{S3.11}$$

$$\sum_{i=1}^{n} \left[\int K_{2i}(y; (\boldsymbol{\beta}, \phi)) f_i(y; (\boldsymbol{\beta}, \phi))^{1+\tau} dy - K_{2i}(y_i; (\boldsymbol{\beta}, \phi)) f_i(y_i; (\boldsymbol{\beta}, \phi))^{\tau} \right] = 0, \tag{S3.12}$$

where $K_{1i}(y_i; (\boldsymbol{\beta}, \phi)) = \frac{(y_i - \mu_i)}{Var(y_i) g'(\mu_i)}$ and $K_{2i}(y_i; (\boldsymbol{\beta}, \phi)) = -\frac{(y_i\theta_i - b(\theta_i))}{a^2(\phi)} a'(\phi) + \frac{\partial}{\partial \phi} c(y_i, \phi)$. In particular, assuming ϕ to be known and taking $\tau = 0$, the estimating equations for $\boldsymbol{\beta}$ turn out to be the same as that of the maximum likelihood estimator (MLE) as well as that of the ordinary least squares (OLS) estimator given by $\sum_{i=1}^{n} \frac{(Y_i - \mu_i)}{Var(Y_i) g'(\mu_i)} \boldsymbol{x}_i = 0$.

Further, assume that the true data generating distribution also belongs to the model density with parameters $(\boldsymbol{\beta}^g, \phi^g)$. Then, it follows from Ghosh and Basu (2016) that, under suitable assumptions, there exists a consistent sequence $(\widehat{\boldsymbol{\beta}}_n, \widehat{\phi}_n)$ of roots to the minimum DPD estimating equations (S3.11) and (S3.12) and

$$\boldsymbol{\Omega}_n^\tau(\boldsymbol{\beta}^g,\phi^g)^{-\frac{1}{2}}\boldsymbol{\Psi}_n^\tau(\boldsymbol{\beta}^g,\phi^g)[\sqrt{n}((\widehat{\boldsymbol{\beta}}_n,\widehat{\phi}_n)-(\boldsymbol{\beta}^g,\phi^g))]$$

follows asymptotically a $(p+1)$-dimensional normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{I}_{p+1}$, the identity matrix of dimension $p+1$. Here

$$\boldsymbol{\Psi}_n^\tau(\boldsymbol{\beta},\phi)=\frac{1}{n}\begin{pmatrix}\boldsymbol{X}^T\boldsymbol{\Gamma}_{11}^{(\tau)}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{\Gamma}_{12}^{(\tau)}\mathbf{1}\\ \mathbf{1}^T\boldsymbol{\Gamma}_{12}^{(\tau)}X & \mathbf{1}^T\boldsymbol{\Gamma}_{22}^{(\tau)}\mathbf{1}\end{pmatrix},\quad \boldsymbol{\Omega}_n^\tau(\boldsymbol{\beta},\phi)=\frac{1}{n}\begin{pmatrix}\boldsymbol{X}^T\widetilde{\boldsymbol{\Gamma}_{11}}^{(\tau)}\boldsymbol{X} & \boldsymbol{X}^T\widetilde{\boldsymbol{\Gamma}_{12}}^{(\tau)}\mathbf{1}\\ \mathbf{1}^T\widetilde{\boldsymbol{\Gamma}_{21}}^{(\tau)}X & \mathbf{1}^T\widetilde{\boldsymbol{\Gamma}_{22}}^{(\tau)}\mathbf{1}\end{pmatrix},$$

with $\boldsymbol{\Gamma}_j^{(\tau)}=Diag(\gamma_{ji})_{i=1,\cdots,n}$, $\boldsymbol{\Gamma}_{jk}^{(\tau)}=Diag(\gamma_{jki})_{i=1,\cdots,n}$ and $\widetilde{\boldsymbol{\Gamma}_{jk}}^{(\tau)}=\left[\boldsymbol{\Gamma}_{jk}^{(2\tau)}-\boldsymbol{\Gamma}_j^{(\tau)T}\boldsymbol{\Gamma}_k^{(\tau)}\right]$ for $j,k=1,2$, $\boldsymbol{X}^T=[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n]$ and for $i=1,\ldots,n$, we define

$$\gamma_{ji}=\gamma_{ji}^{1+\tau}(\boldsymbol{\beta},\phi)=\int K_{ji}(y;(\boldsymbol{\beta},\phi))f_i(y;(\boldsymbol{\beta},\phi))^{1+\tau}dy,\quad j=1,2,$$

$$\gamma_{jki}=\gamma_{jki}^{1+\tau}(\boldsymbol{\beta},\phi)=\int K_{ji}(y;(\boldsymbol{\beta},\phi))K_{ki}(y;(\boldsymbol{\beta},\phi))f_i(y;(\boldsymbol{\beta},\phi))^{1+\tau}dy,\quad j,k=1,2.$$

It is interesting to note that the estimators $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\phi}_n$ are not asymptotically independent for any general GLM; a set of sufficient conditions for this to happen is $\gamma_{12i}^{1+2\tau}=0$ and $\gamma_{1i}^{1+\tau}\gamma_{2i}^{1+\tau}=0$ for all i which holds for only some particular GLM including the normal regression case.

The robustness of the MDPDE in the GLM can be described in terms of its influence function following Section 2.4 of Ghosh and Basu (2016). In particular, the influence function of the minimum DPD functional $\boldsymbol{U}_\tau$ for contamination only in the

i_0^{th} direction is given by

$$IF_{i_0}(t_{i_0}, \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}}) \;=\; (\boldsymbol{\Psi}_n^\tau)^{-1}\, \frac{1}{n} \begin{pmatrix} [f_{i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi))^\tau K_{1i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi)) - \gamma_{1i_0}]\, \boldsymbol{x}_i \\ f_{i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi))^\tau K_{2i_0}(t_{i_0}; (\boldsymbol{\beta}, \phi)) - \gamma_{2i_0} \end{pmatrix}.$$

Clearly, the boundedness of this influence function for any fixed sample size n and any given (finite) values of $\boldsymbol{x}_i$'s depends on the boundedness of the terms $f_i(t_i; (\boldsymbol{\beta}, \phi))^\tau K_{ji}(t_i; (\boldsymbol{\beta}, \phi))$ for all i and $j = 1, 2$; the terms $\boldsymbol{\Psi}_n$ and γ_{ji_0} are generally bounded by the conditions assumed for asymptotic normality of the corresponding estimators. Further, whenever $\gamma_{12i}^{1+2\tau} = 0$ and $\gamma_{1i}^{1+\tau}\gamma_{2i}^{1+\tau} = 0$ for all i one can separate out the influence function for the MDPDE of $\boldsymbol{\beta}$ and ϕ and they will be independent of each other.

S4 Proofs of the Results

S4.1 Proof of Theorem 3.1 of the main paper

Fix any $i = 1, \ldots, n$. We consider the second order Taylor series expansion of $d_\gamma(f_i(\cdot; \boldsymbol{\theta}), f_i(\cdot; \boldsymbol{\theta}_0))$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_n^\tau$ as

$$\begin{aligned} d_\gamma(f_i(\cdot; \boldsymbol{\theta}_n^\tau), f_i(\cdot; \boldsymbol{\theta}_0)) \;=\;& d_\gamma(f_i(\cdot; \boldsymbol{\theta}_0), f_i(\cdot; \boldsymbol{\theta}_0)) \\ &+ \sum_{j=1}^p \nabla_j d_\gamma(f_i(.; \boldsymbol{\theta}), f_i(.; \boldsymbol{\theta}_0))|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}((\theta_n^\tau)^j - \theta_0^j) \\ &+ \frac{1}{2} \sum_{j,k} \nabla_{jk}^2 d_\gamma(f_i(.; \boldsymbol{\theta}), f_i(.; \boldsymbol{\theta}_0))|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}((\theta_n^\tau)^j - \theta_0^j)((\theta_n^\tau)^k - \theta_0^k) \\ &+ o(||\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0||^2), \end{aligned}$$

where the superscripts denote the corresponding components. Now we have

$$d_\gamma(f_i(.;\boldsymbol{\theta}_0), f_i(.;\boldsymbol{\theta}_0)) = 0 \quad \text{and} \quad \nabla_j d_\gamma(f_i(.;\boldsymbol{\theta}), f_i(.;\boldsymbol{\theta}_0))|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0.$$

Also, we have

$$\begin{aligned}\nabla^2_{jk} d_\gamma(f_i(.;\boldsymbol{\theta}), f_i(.;\boldsymbol{\theta}_0))|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= (1+\gamma)\int f_i(\cdot;\boldsymbol{\theta}_0)^{\gamma-1}\frac{\partial f_i(\cdot;\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_j}\frac{\partial f_i(\cdot;\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_k} \\ &= (j,k) - \text{th element of } \boldsymbol{A}_\gamma^{(i)}(\boldsymbol{\theta}_0).\end{aligned}$$

Now from the above Taylor series expansion it is clear that the random variables

$$T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0) = 2\sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}_n^\tau), f_i(.;\boldsymbol{\theta}_0)) \quad \text{and} \quad \sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0)^T \boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0)$$

have the same asymptotic distribution because $n \times o(||\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0||^2) = o_P(1)$. However, since $\boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_0) \to \boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)$ element-wise as $n \to \infty$, by Slutsky's Theorem the asymptotic distribution of the test statistics $T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0)$ is in turn the same as that of $\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0)^T \boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0)$. Also it follows from Ghosh and Basu (2013) that the asymptotic distribution of $\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0)$ is normal with mean zero and variance $\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0)\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0)$.

Further we know that for $\boldsymbol{X} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma})$, and a q-dimensional real symmetric matrix $\boldsymbol{A}$, the distribution of the quadratic form $\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}$ is the same as that of $\sum_{i=1}^r \zeta_i Z_i^2$, where $Z_1, \ldots, Z_r$ are independent standard normal variables, $r = rank(\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma})$, $r \geq 1$ and $\zeta_1, \ldots, \zeta_r$ are the nonzero eigenvalues of $\boldsymbol{A}\boldsymbol{\Sigma}$ (Dik and Gunst, 1985, Corollary 2.1). Applying this result with $\boldsymbol{X} = \sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0)$ we get the theorem with

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

Finally, from the Corollary 8.3.3 of Harville (2008), it follows that,

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

□

S4.2 Proof of Theorem 3.2 of the main paper

Fix any $i = 1, 2, \ldots, n$. Consider the first order Taylor series expansion of $d_\gamma(f_i(\cdot;\boldsymbol{\theta}), f_i(\cdot;\boldsymbol{\theta}_0))$ under $f_i(\cdot;\boldsymbol{\theta}^*)$ as

$$d_\gamma(f_i(\cdot;\boldsymbol{\theta}_n^\tau), f_i(\cdot;\boldsymbol{\theta}_0)) = d_\gamma(f_i(\cdot;\boldsymbol{\theta}^*), f_i(\cdot;\boldsymbol{\theta}_0)) + \boldsymbol{M}_\gamma^{(i)}(\boldsymbol{\theta}^*)^T(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}^*) + o(||\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}^*||).$$

Now we know that, under $\boldsymbol{\theta}^*$,

$$\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}^*) \xrightarrow{\mathcal{D}} N_p(\boldsymbol{0}, \boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}^*)\boldsymbol{V}_\tau(\boldsymbol{\theta}^*)\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}^*)) \quad \text{as} \quad n \to \infty,$$

and $\sqrt{n} \times o(||\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}^*||) = o_P(1)$. Thus we get that the random variables

$$\frac{1}{\sqrt{n}}\left[\sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}_n^\tau), f_i(.;\boldsymbol{\theta}_0)) - \sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}^*), f_i(.;\boldsymbol{\theta}_0))\right] \quad \text{and} \quad \boldsymbol{M}_n^\gamma(\boldsymbol{\theta}^*)^T\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}^*)$$

have the same asymptotic distribution. Using the convergence $\boldsymbol{M}_n^\gamma(\boldsymbol{\theta}^*) \to \boldsymbol{M}_\gamma(\boldsymbol{\theta}^*)$, we have

$$\frac{1}{\sqrt{n}}\left[\sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}_n^\tau), f_i(.;\boldsymbol{\theta}_0)) - \sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}^*), f_i(.;\boldsymbol{\theta}_0))\right] \xrightarrow{\mathcal{D}} N(0, \sigma_{\tau,\gamma}(\boldsymbol{\theta}^*)),$$

where $\sigma_{\tau,\gamma}(\boldsymbol{\theta}^*)$ is as defined in the theorem. The desired approximation to the power function follows from the above asymptotic distribution. □

S4.3 Proof of Theorem 3.4 of the main paper

Let us denote $\boldsymbol{\theta}_n^* = \boldsymbol{U}_\tau(\underline{\mathbf{F}}_{n,\epsilon,\mathbf{t}}^P)$ and fix any $i = 1, \ldots, n$. We consider the second order Taylor series expansion of $d_\gamma(f_i(\cdot;\boldsymbol{\theta}), f_i(\cdot;\boldsymbol{\theta}_0))$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_n^*$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_n^\tau$ as,

$$\begin{aligned} d_\gamma(f_i(\cdot;\boldsymbol{\theta}_n^\tau), f_i(\cdot;\boldsymbol{\theta}_0)) &= d_\gamma(f_i(\cdot;\boldsymbol{\theta}_n^*), f_i(\cdot;\boldsymbol{\theta}_0)) + \boldsymbol{M}_\gamma^{(i)}(\boldsymbol{\theta}_n^*)^T(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_n^*) \\ &\quad + \frac{1}{2}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_n^*)^T \boldsymbol{A}_\gamma^{(i)}(\boldsymbol{\theta}_n^*)(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_n^*) + o(||\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_n^*||^2). \end{aligned}$$

Now from Section 2 of the main paper and using the consistency of $\boldsymbol{\theta}_n^*$ we know that, under $\underline{\mathbf{F}}_{n,\epsilon,\mathbf{t}}^P$,

$$\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_n^*) \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \boldsymbol{\Sigma}_\tau(\boldsymbol{\theta}_0)).$$

Further using the Taylor series expansion of $\boldsymbol{M}_\gamma^{(i)}(\boldsymbol{\theta})$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_n^*$, we get

$$\boldsymbol{M}_\gamma^{(i)}(\boldsymbol{\theta}_n^*) - \boldsymbol{M}_\gamma^{(i)}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}\boldsymbol{A}_\gamma^{(i)}(\boldsymbol{\theta}_0)\boldsymbol{\Delta} + \frac{\epsilon}{\sqrt{n}} IF(\mathbf{t}; \boldsymbol{M}_\gamma^{(i)}(\boldsymbol{U}_\tau), \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) + o\left(\frac{1}{\sqrt{n}}\right).$$

But $\boldsymbol{M}_\gamma^{(i)}(\boldsymbol{\theta}_0) = 0$ and $IF(\mathbf{t}; \boldsymbol{M}_\gamma^{(i)}(\boldsymbol{U}_\tau), \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = \boldsymbol{A}_\gamma^{(i)}(\boldsymbol{\theta}_0) IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})$, so that we get

$$\begin{aligned} \sqrt{n}\boldsymbol{M}_\gamma^{(i)}(\boldsymbol{\theta}_n^*) &= \boldsymbol{A}_\gamma^{(i)}(\boldsymbol{\theta}_0)\boldsymbol{\Delta} + \epsilon \boldsymbol{A}_\gamma^{(i)}(\boldsymbol{\theta}_0) IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) + o(1) \\ &= \boldsymbol{A}_\gamma^{(i)}(\boldsymbol{\theta}_0)\widetilde{\boldsymbol{\Delta}} + o(1), \end{aligned}$$

where $\widetilde{\boldsymbol{\Delta}} = \left[\boldsymbol{\Delta} + \epsilon IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})\right]$. Again using the second order Taylor series expansion of $d_\gamma(f_i(\cdot;\boldsymbol{\theta}), f_i(\cdot;\boldsymbol{\theta}_0))$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_n^*$ and taking summation over all $i = 1, \ldots, n$, we get

$$\begin{aligned}\sum_{i=1}^n & \left[d_\gamma(f_i(\cdot;\boldsymbol{\theta}_n^*), f_i(\cdot;\boldsymbol{\theta}_0)) - d_\gamma(f_i(\cdot;\boldsymbol{\theta}_0), f_i(\cdot;\boldsymbol{\theta}_0))\right] \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\Delta}^T \boldsymbol{M}_\gamma^{(i)}(\boldsymbol{\theta}_0) + \epsilon\sqrt{n} IF(\mathbf{t}; T_{\gamma,\lambda}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) + \frac{1}{2}\boldsymbol{\Delta}^T \boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_0)\boldsymbol{\Delta} \\ &\quad + \frac{\epsilon^2}{2} IF^{(2)}(\mathbf{t}; T_{\gamma,\lambda}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) + \frac{\epsilon}{n}\sum_{i=1}^n \boldsymbol{\Delta}^T IF(\mathbf{t}; \boldsymbol{M}_\gamma^{(i)}(\boldsymbol{U}_\tau), \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) + o(1).\end{aligned}$$

But, $d_\gamma(f_i(\cdot;\boldsymbol{\theta}_0), f_i(\cdot;\boldsymbol{\theta}_0)) = 0$, $\boldsymbol{M}_\gamma^{(i)}(\boldsymbol{\theta}_0) = 0$ for all $i = 1, \ldots, n$, $IF(\mathbf{t}; T_{\gamma,\lambda}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = 0$ and

$$IF^{(2)}(\mathbf{t}; T_{\gamma,\lambda}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) = IF(\mathbf{t}; U_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})^T \boldsymbol{A}_\gamma(\boldsymbol{\theta}_0) IF(\mathbf{t}; U_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}).$$

Thus, taking summation over $i = 1, \ldots, n$, the above equation simplifies to

$$\begin{aligned}2\sum_{i=1}^n d_\gamma(f_i(\cdot;\boldsymbol{\theta}_n^*), f_i(\cdot;\boldsymbol{\theta}_0)) &= \boldsymbol{\Delta}^T \boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_0)\boldsymbol{\Delta} + \epsilon^2 IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})\boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_0)^T IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) \\ &\quad + 2\epsilon\boldsymbol{\Delta}^T \boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_0) IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) + o(1).\end{aligned}$$

Hence, noting that $n \times o(||\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_n^*||^2) = o_P(1)$, we get

$$\begin{aligned}2\sum_{i=1}^n d_\gamma(f_i(\cdot;\boldsymbol{\theta}_n^\tau), f_i(\cdot;\boldsymbol{\theta}_0)) &= \widetilde{\boldsymbol{\Delta}}^T \boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_0)\widetilde{\boldsymbol{\Delta}} + 2\widetilde{\boldsymbol{\Delta}}^T \boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_n^*) \\ &\quad + \sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_n^*)^T \boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_n^*)\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_n^*) + o_P(1) + o(1) \\ &= \left[\widetilde{\boldsymbol{\Delta}} + \sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_n^*)\right]^T \boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_0)\left[\widetilde{\boldsymbol{\Delta}} + \sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_n^*)\right] + o_P(1).\end{aligned}$$

Thus under the probability $\underline{\mathbf{F}}^P_{n,\epsilon,\mathbf{t}}$, asymptotic distribution of the proposed DPD based test statistics $T_\gamma(\boldsymbol{\theta}_n^\tau, \boldsymbol{\theta}_0) = 2\sum_{i=1}^n d_\gamma(f_i(\cdot;\boldsymbol{\theta}_n^\tau), f_i(\cdot;\boldsymbol{\theta}_0))$ is the same as the distribution of $(\widetilde{\boldsymbol{\Delta}} + \boldsymbol{W}_0)^T \boldsymbol{A}_n^\gamma(\boldsymbol{\theta}_0)(\widetilde{\boldsymbol{\Delta}} + \boldsymbol{W}_0)$, where $\boldsymbol{W}_0$ follows the asymptotic distribution of $\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_n^*)$. Hence Part (i) follows by taking $\boldsymbol{W} = (\widetilde{\boldsymbol{\Delta}} + \boldsymbol{W}_0)$.

(ii) This part follows from Part (i) using the series expansion of the distribution function of a linear combination of independent non-central chi-squares in terms of central chi-square distribution functions as derived in Kotz et al. (1967). □

S4.4 Proof of Theorem 3.7 of the main paper

Starting with the expression of $P_{\tau,\gamma}(\boldsymbol{\Delta},\epsilon;\alpha)$ as obtained in Theorem 3.4 of the main paper, we get the power influence function $PIF(\cdot)$ as

$$\begin{aligned} PIF(\mathbf{t}; T_{\gamma,\tau}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) &= \frac{\partial}{\partial\epsilon} P_{\tau,\gamma}(\boldsymbol{\Delta},\epsilon;\alpha)\Big|_{\epsilon=0} \\ &= \sum_{v=0}^{\infty} \frac{\partial}{\partial\epsilon} C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \widetilde{\boldsymbol{\Delta}})\Big|_{\epsilon=0} P\left(\chi^2_{r+2v} > \frac{t_\alpha^{\tau,\gamma}}{\zeta_{(1)}^{\gamma,\tau}(\boldsymbol{\theta}_0)}\right). \end{aligned} \tag{S4.13}$$

Now, note that for each $v \geq 0$, the quantities $C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \widetilde{\boldsymbol{\Delta}})$ depends on ϵ only through its second argument $\widetilde{\boldsymbol{\Delta}} = \left[\boldsymbol{\Delta} + \epsilon IF(\mathbf{t}; U_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})\right]$ and at $\epsilon = 0$ we have $\widetilde{\boldsymbol{\Delta}} = \boldsymbol{\Delta}$. Consider a Taylor series expansion of $C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \boldsymbol{d})$ with respect to $\boldsymbol{d}$ around $\boldsymbol{d} = \boldsymbol{\Delta}$ and evaluate it at $\boldsymbol{d} = \widetilde{\boldsymbol{\Delta}}$ to get

$$\begin{aligned} C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \widetilde{\boldsymbol{\Delta}}) &= C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \boldsymbol{\Delta}) + (\widetilde{\boldsymbol{\Delta}} - \boldsymbol{\Delta})^T \left[\frac{\partial}{\partial \boldsymbol{d}} C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \boldsymbol{d})^T\Big|_{\boldsymbol{d}=\boldsymbol{\Delta}}\right] + o(||\widetilde{\boldsymbol{\Delta}} - \boldsymbol{\Delta}||) \\ &= C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \boldsymbol{\Delta}) + \epsilon IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})^T \cdot \left[\frac{\partial}{\partial \boldsymbol{d}} C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \boldsymbol{d})\Big|_{\boldsymbol{d}=\boldsymbol{\Delta}}\right] + o(\epsilon||IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})||) \end{aligned} \tag{S4.14}$$

Now differentiating it with respect to ϵ and evaluating at $\epsilon = 0$, we get

$$\frac{\partial}{\partial \epsilon} C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \widetilde{\boldsymbol{\Delta}})\Big|_{\epsilon=0} = IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})^T \left[\frac{\partial}{\partial \boldsymbol{d}} C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \boldsymbol{d})\Big|_{\boldsymbol{d}=\boldsymbol{\Delta}}\right],$$

provided the influence function $IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})$ is finite. Combining it with Equation (S4.13), we finally get the required power influence function as

$$\begin{aligned}
PIF(\mathbf{t}; T_{\gamma,\lambda}^{(1)}, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0}) &= \frac{\partial}{\partial \epsilon} P_{\tau,\gamma}(\boldsymbol{\Delta}, \epsilon; \alpha)\big|_{\epsilon=0} \\
&= \sum_{v=0}^{\infty} \frac{\partial}{\partial \epsilon} C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \widetilde{\boldsymbol{\Delta}})\Big|_{\epsilon=0} P\left(\chi^2_{r+2v} > \frac{t_\alpha^{\tau,\gamma}}{\zeta_{(1)}^{\gamma,\tau}(\boldsymbol{\theta}_0)}\right) \\
&= IF(\mathbf{t}; \boldsymbol{U}_\tau, \underline{\mathbf{F}}_{\boldsymbol{\theta}_0})^T \left(\sum_{v=0}^{\infty} \left[\frac{\partial}{\partial \boldsymbol{d}} C_v^{\gamma,\tau}(\boldsymbol{\theta}_0, \boldsymbol{d})\Big|_{\boldsymbol{d}=\boldsymbol{\Delta}}\right] P\left(\chi^2_{r+2v} > \frac{t_\alpha^{\tau,\gamma}}{\zeta_{(1)}^{\gamma,\tau}(\boldsymbol{\theta}_0)}\right)\right).
\end{aligned}$$

□

S4.5 Proof of Theorem 4.1 of the main paper

Note that the consistency of the RMDPDE follows from the exactly same proof of Theorem 3.1 of Ghosh and Basu (2013), because the conditions (A1)–(A7) of their paper hold with respect to Θ_0. So, here we will only prove the asymptotic normality of the RMDPDE.

First note that, $\widetilde{\boldsymbol{\theta}}^g$ is the true RMDPDE in the sense that it minimizes the DPD subject to the given constraints. So, one can prove that (see proof of Theorem 3.1 of Ghosh and Basu (2013))

$$\frac{1}{1+\tau} \boldsymbol{\Omega}_n^{-\frac{1}{2}} \left[-\sqrt{n}\ \nabla H_n(\widetilde{\boldsymbol{\theta}}^g)\right] \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \boldsymbol{I}_p). \tag{S4.15}$$

Now, using a Taylor series expansion, we have

$$\sqrt{n}\ \nabla H_n(\widetilde{\boldsymbol{\theta}}_n^\tau) = \sqrt{n}\ \nabla H_n(\widetilde{\boldsymbol{\theta}}^g) + \nabla^2 H_n(\widetilde{\boldsymbol{\theta}}^g)\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g) + o\left(\sqrt{n}||\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g||^2\right).$$

But, by the consistency of $\widetilde{\boldsymbol{\theta}}_n^\tau$, we have $o\left(\sqrt{n}||\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g||^2\right) = o_P(1)$. Hence, we get that

$$\sqrt{n}\ \nabla H_n(\widetilde{\boldsymbol{\theta}}_n^\tau) = \sqrt{n}\ \nabla H_n(\widetilde{\boldsymbol{\theta}}^g) + \nabla^2 H_n(\widetilde{\boldsymbol{\theta}}^g)\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g) + o_P(\mathbf{1}). \tag{S4.16}$$

Similarly, we also have,

$$\sqrt{n}\ \boldsymbol{\upsilon}(\widetilde{\boldsymbol{\theta}}_n^\tau) = \boldsymbol{\Upsilon}^T(\widetilde{\boldsymbol{\theta}}^g)\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g) + o_P(\mathbf{1}), \tag{S4.17}$$

since $\boldsymbol{\upsilon}(\widetilde{\boldsymbol{\theta}}^g) = \mathbf{0}$. Now, the RMDPDE $\widetilde{\boldsymbol{\theta}}_n^\tau$ must satisfy the Equations (15) of the main paper. Using them, along with Equations (S4.16) and (S4.17), we get

$$\begin{aligned}
\sqrt{n}\ \nabla H_n(\widetilde{\boldsymbol{\theta}}^g) + \nabla^2 H_n(\widetilde{\boldsymbol{\theta}}^g)\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g) + \sqrt{n}\ \boldsymbol{\Upsilon}(\widetilde{\boldsymbol{\theta}}_n^\tau)\boldsymbol{\lambda}_n + o_P(\mathbf{1}) &= \mathbf{0},\\
\boldsymbol{\Upsilon}^T(\widetilde{\boldsymbol{\theta}}^g)\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g) + o_P(\mathbf{1}) &= \mathbf{0}.
\end{aligned}$$

Writing it in the matrix form, we get

$$\begin{bmatrix} \nabla^2 H_n(\widetilde{\boldsymbol{\theta}}^g) & \boldsymbol{\Upsilon}(\widetilde{\boldsymbol{\theta}}^g) \\ \boldsymbol{\Upsilon}^T(\widetilde{\boldsymbol{\theta}}^g) & \boldsymbol{O}_r \end{bmatrix} \cdot \begin{bmatrix} \sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g) \\ \sqrt{n}\ \boldsymbol{\lambda}_n \end{bmatrix} = \begin{bmatrix} -\sqrt{n}\ \nabla H_n(\widetilde{\boldsymbol{\theta}}^g) \\ \mathbf{0}_r \end{bmatrix} + o_P(\mathbf{1}),$$

where $\boldsymbol{O}_r$ denote the square null matrix of order r and $\mathbf{0}_r$ denote zero vector (column) of length r. Taking the inverse of the first matrix in the LHS of above equations, and simplifying, we get

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g) = -\frac{1}{1+\tau}\boldsymbol{P}_n^\tau(\widetilde{\boldsymbol{\theta}}^g)\sqrt{n}\ \nabla H_n(\widetilde{\boldsymbol{\theta}}^g) + o_p(1),$$

where $\boldsymbol{P}_n^\tau(\boldsymbol{\theta})$ is as defined in the theorem. Therefore,

$$\boldsymbol{\Omega}_n^\tau(\widetilde{\boldsymbol{\theta}}^g)^{-\frac{1}{2}}\boldsymbol{P}_n^\tau(\widetilde{\boldsymbol{\theta}}^g)^{-1}\left[\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \widetilde{\boldsymbol{\theta}}^g)\right] = \frac{1}{1+\tau}\boldsymbol{\Omega}_n^\tau(\widetilde{\boldsymbol{\theta}}^g)^{-\frac{1}{2}}\left[-\sqrt{n}\ \nabla H_n(\widetilde{\boldsymbol{\theta}}^g)\right] + o_p(\boldsymbol{1}).$$

Now, the theorem follows using Equation (S4.15). □

S4.6 Proof of Theorem 4.3 of the main paper

Fix any $i = 1, \ldots, n$. We consider the second order Taylor series expansion of $d_\gamma(f_i(\cdot;\boldsymbol{\theta}), f_i(\cdot;\widetilde{\boldsymbol{\theta}}_n^\tau))$ around $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}_n^\tau$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_n^\tau$ as,

$$\begin{aligned} d_\gamma(f_i(\cdot;\boldsymbol{\theta}_n^\tau), f_i(\cdot;\widetilde{\boldsymbol{\theta}}_n^\tau)) &= d_\gamma(f_i(\cdot;\widetilde{\boldsymbol{\theta}}_n^\tau), f_i(\cdot;\widetilde{\boldsymbol{\theta}}_n^\tau)) \\ &+ \sum_{j=1}^{p} \nabla_j d_\gamma(f_i(.;\boldsymbol{\theta}), f_i(.;\widetilde{\boldsymbol{\theta}}_n^\tau))|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}_n^\tau}((\theta_n^\tau)^j - (\widetilde{\theta}_n^\tau)^j) \\ &+ \frac{1}{2}\sum_{j,k} \nabla_{jk} d_\gamma(f_i(.;\boldsymbol{\theta}), f_i(.;\widetilde{\boldsymbol{\theta}}_n^\tau))|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}_n^\tau}((\theta_n^\tau)^j - (\widetilde{\theta}_n^\tau)^j)((\theta_n^\tau)^k - (\widetilde{\theta}_n^\tau)^k) \\ &+ o(||\boldsymbol{\theta}_n^\tau - \widetilde{\boldsymbol{\theta}}_n^\tau||^2), \end{aligned}$$

where the superscripts denote the corresponding components. But,

$$d_\gamma(f_i(.;\widetilde{\boldsymbol{\theta}}_n^\tau), f_i(.;\widetilde{\boldsymbol{\theta}}_n^\tau)) = 0, \qquad \nabla_j d_\gamma(f_i(.;\boldsymbol{\theta}), f_i(.;\widetilde{\boldsymbol{\theta}}_n^\tau))|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}_n^\tau} = 0$$

and

$$\begin{aligned} \nabla_{jk}^2 d_\gamma(f_i(.;\boldsymbol{\theta}), f_i(.;\widetilde{\boldsymbol{\theta}}_n^\tau))|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}_n^\tau} &= (1+\gamma)\int f_i(\cdot;\widetilde{\boldsymbol{\theta}}_n^\tau)^{\gamma-1}\frac{\partial f_i(\cdot;\widetilde{\boldsymbol{\theta}}_n^\tau)}{\partial \boldsymbol{\theta}_j}\frac{\partial f_i(\cdot;\widetilde{\boldsymbol{\theta}}_n^\tau)}{\partial \boldsymbol{\theta}_k} \\ &= (j,k)\text{-th element of } \boldsymbol{A}_\gamma^{(i)}(\widetilde{\boldsymbol{\theta}}_n^\tau). \end{aligned}$$

So from the above Taylor series expansion,

$$\begin{aligned} S_\gamma(\boldsymbol{\theta}_n^\tau, \widetilde{\boldsymbol{\theta}}_n^\tau) &= 2\sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}_n^\tau), f_i(.;\widetilde{\boldsymbol{\theta}}_n^\tau)) \\ &= \sqrt{n}(\boldsymbol{\theta}_n^\tau - \widetilde{\boldsymbol{\theta}}_n^\tau)^T \boldsymbol{A}_n^\gamma(\widetilde{\boldsymbol{\theta}}_n^\tau)\sqrt{n}(\boldsymbol{\theta}_n^\tau - \widetilde{\boldsymbol{\theta}}_n^\tau) + n \times o(||\boldsymbol{\theta}_n^\tau - \widetilde{\boldsymbol{\theta}}_n^\tau||^2). \end{aligned}$$

Now, since $\boldsymbol{\theta}_0 \in \Theta_0$ is the true value of the parameter, we have $\widetilde{\boldsymbol{\theta}}_n^\tau \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$ and so $\boldsymbol{A}_n^\gamma(\widetilde{\boldsymbol{\theta}}_n^\tau) \xrightarrow{\mathcal{P}} \boldsymbol{A}_\gamma(\widetilde{\boldsymbol{\theta}}_n^\tau)$ element-wise as $n \to \infty$. Let us define

$$\boldsymbol{D}_n(\boldsymbol{\theta}) = [\nabla^2 H_n(\boldsymbol{\theta})]^{-1}\boldsymbol{\Upsilon}(\boldsymbol{\theta})\left\{\boldsymbol{\Upsilon}(\boldsymbol{\theta})^T[\nabla^2 H_n(\boldsymbol{\theta})]^{-1}\boldsymbol{\Upsilon}(\boldsymbol{\theta})\right\}^{-1}\boldsymbol{\Upsilon}(\boldsymbol{\theta})^T.$$

Then noting that, at the model with parameter $\boldsymbol{\theta}_0$, $\boldsymbol{\Psi}_n^\tau(\boldsymbol{\theta}_0) = \frac{1}{1+\tau}[\nabla^2 H_n(\boldsymbol{\theta}_0)]$, we get $\boldsymbol{D}_n(\boldsymbol{\theta}_0) = \boldsymbol{I}_r - \boldsymbol{P}_n^\tau(\boldsymbol{\theta}_0)\boldsymbol{\Psi}_n^\tau(\boldsymbol{\theta}_0)$. Now, from the proof of Theorem 4.1, we have

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \boldsymbol{\theta}_0) = -\frac{1}{1+\tau}\boldsymbol{P}_n^\tau(\boldsymbol{\theta}_0)\sqrt{n}\ \nabla H_n(\boldsymbol{\theta}_0) + o_p(1).$$

Also, using a Taylor series expansion, we get

$$\sqrt{n}\ \nabla H_n(\boldsymbol{\theta}_0) = -\sqrt{n}\ \nabla^2 H_n(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0) + o_p(1).$$

Combining above two,

$$\begin{aligned} \sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^\tau - \boldsymbol{\theta}_0) &= \frac{1}{1+\tau}\boldsymbol{P}_n^\tau(\boldsymbol{\theta}_0)\sqrt{n}\ \nabla^2 H_n(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0) + o_p(1) \\ &= \boldsymbol{P}_n^\tau(\boldsymbol{\theta}_0)\boldsymbol{\Psi}_n(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0) + o_p(1) \\ &= \sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0) - \boldsymbol{D}_n(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0) + o_p(1). \end{aligned}$$

Therefore,

$$\sqrt{n}(\boldsymbol{\theta}_n^\tau - \widetilde{\boldsymbol{\theta}}_n^\tau) = \boldsymbol{D}_n(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0) + o_p(1).$$

But, by the assumptions of the Theorem, it follows that $\boldsymbol{D}_n(\boldsymbol{\theta}_0) \to [\boldsymbol{I}_r - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)\boldsymbol{J}_\tau(\boldsymbol{\theta}_0)]$ element-wise as $n \to \infty$ and from the asymptotic distribution of MDPDE in Section 2 of the main paper, we get that

$$\sqrt{n}(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}_0) \to N_p(\boldsymbol{0}, \boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0)\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0)).$$

Hence

$$\sqrt{n}(\boldsymbol{\theta}_n^\tau - \widetilde{\boldsymbol{\theta}}_n^\tau) \to N_p(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}_\tau(\boldsymbol{\theta}_0)),$$

where

$$\begin{aligned}\widetilde{\boldsymbol{\Sigma}}_\tau(\boldsymbol{\theta}_0) &= [\boldsymbol{I}_r - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)\boldsymbol{J}_\tau(\boldsymbol{\theta}_0)]\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0)\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0)[\boldsymbol{I}_r - \boldsymbol{J}_\tau(\boldsymbol{\theta}_0)\boldsymbol{P}_\tau(\boldsymbol{\theta}_0)] \\ &= [\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)]\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)[\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)].\end{aligned}$$

It also follows from above that $n \times o(||\boldsymbol{\theta}_n^\tau - \widetilde{\boldsymbol{\theta}}_n^\tau||^2) = o_P(1)$. Thus, the asymptotic distribution of the DPD test statistics $S_\gamma(\boldsymbol{\theta}_n^\tau, \widetilde{\boldsymbol{\theta}}_n^\tau)$ under $\boldsymbol{\theta}_0 \in \Theta_0$ coincides with that of $\sqrt{n}(\boldsymbol{\theta}_n^\tau - \widetilde{\boldsymbol{\theta}}_n^\tau)^T\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n^\tau - \widetilde{\boldsymbol{\theta}}_n^\tau)$, which is same as that of the random variable $\sum_{i=1}^r \zeta_i Z_i^2$, where $Z_1, \cdots, Z_r$ are independent standard normal variables, $\zeta_1^{\gamma,\tau}, \cdots, \zeta_r^{\gamma,\tau}$ are the nonzero eigenvalues of $\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\widetilde{\boldsymbol{\Sigma}}_\tau(\boldsymbol{\theta}_0)$ and $r = rank\left(\widetilde{\boldsymbol{\Sigma}}_\tau(\boldsymbol{\theta}_0)\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)\widetilde{\boldsymbol{\Sigma}}_\tau(\boldsymbol{\theta}_0)\right)$, by Corollary 2.1 of Dik and Gunst (1985) (see the last part of the proof of Theorem 3.1). Finally, from the Corollary 8.3.3 of Harville (2008), we get

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

□

S4.7 Proof of Theorem 4.4 of the main paper

Fix any $i = 1, 2, \ldots, n$. Considering the first order Taylor series expansion, we get

$$\begin{aligned} d_\gamma(f_i(\cdot;\boldsymbol{\theta}_n^\tau), f_i(\cdot;\widetilde{\boldsymbol{\theta}}_n^\tau)) &= d_\gamma(f_i(\cdot;\boldsymbol{\theta}^*), f_i(\cdot;\boldsymbol{\theta}_0)) + \boldsymbol{M}_{1,\gamma}^{(i)}(\boldsymbol{\theta}^*,\boldsymbol{\theta}_0)^T(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}^*) \\ &\quad + \boldsymbol{M}_{2,\gamma}^{(i)}(\boldsymbol{\theta}^*,\boldsymbol{\theta}_0)^T(\widetilde{\boldsymbol{\theta}}_n^\tau - \boldsymbol{\theta}^*) + o(||\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}^*|| + ||\widetilde{\boldsymbol{\theta}}_n^\tau - \boldsymbol{\theta}^*||). \end{aligned}$$

Averaging over i and multiplying by $\sqrt{n}$, we get

$$\begin{aligned} &\frac{1}{\sqrt{n}}\left[\sum_{i=1}^n d_\gamma(f_i(\cdot;\boldsymbol{\theta}_n^\tau), f_i(\cdot;\widetilde{\boldsymbol{\theta}}_n^\tau)) - \sum_{i=1}^n d_\gamma(f_i(.;\boldsymbol{\theta}^*), f_i(.;\boldsymbol{\theta}_0))\right] \\ &= \boldsymbol{M}_n^{1,\gamma}(\boldsymbol{\theta}^*,\boldsymbol{\theta}_0)^T(\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}^*) + \boldsymbol{M}_n^{2,\gamma}(\boldsymbol{\theta}^*,\boldsymbol{\theta}_0)^T(\widetilde{\boldsymbol{\theta}}_n^\tau - \boldsymbol{\theta}^*) \\ &\qquad + \sqrt{n} \times o(||\boldsymbol{\theta}_n^\tau - \boldsymbol{\theta}^*|| + ||\widetilde{\boldsymbol{\theta}}_n^\tau - \boldsymbol{\theta}^*||), \end{aligned}$$

from which the Theorem follows in a straightforward manner. □

S4.8 Proof of Corollary 6.2 of the main paper

The proof follows from the Theorem 4.3 of the main paper by noting that

$$\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)[\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)]\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)[\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)] = \begin{bmatrix} \boldsymbol{E} & \boldsymbol{0}_p \\ \boldsymbol{0}_p^T & 0 \end{bmatrix}$$

and

$$\begin{aligned} & rank(\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)[\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)]\boldsymbol{A}_\gamma(\boldsymbol{\theta}_0)[\boldsymbol{J}_\tau^{-1}(\boldsymbol{\theta}_0) - \boldsymbol{P}_\tau(\boldsymbol{\theta}_0)]\boldsymbol{V}_\tau(\boldsymbol{\theta}_0)) \\ = \; & rank((\boldsymbol{L}^T\boldsymbol{J}_{11.2}^{-1}\boldsymbol{L})^{-1}(\boldsymbol{L}^T\boldsymbol{J}_{11.2}^{-1}L)) = rank(\boldsymbol{I}_r) = r. \end{aligned}$$

Here the last two relations follow from some straightforward matrix algebra calculations and the fact that $\boldsymbol{J}_\tau$ and $\boldsymbol{V}_\tau$ are invertible. □

S5 Simulation Results for Simple Linear Regression

We consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, 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. In this case, outliers may affect the size or power of the test. We compute the empirical size and power of the proposed DPD based test for testing $H_0 : (\beta_0,\ \beta_1) = (\beta_0^g,\ \beta_1^g)$. assuming the error variance σ^2 to be both known and unknown separately. For our simulation, we will take $\boldsymbol{\beta}^g = (\beta_0^g,\ \beta_1^g) = (3,\ 2)$, the true value of σ^2 to be 3 and the explanatory variables are given from a normal distribution with mean 10 and variance 5. Then, we will check the stability of the size and power of this test under several types of contaminations based on 10000 Monte-Carlo iterations and for three different sample sizes $n = 30, 50$, and 100. We simulate ϵ_is from some specific distributions and contaminate the values of x_is to create different scenarios of large residuals and leverage points; however, due to the different objective and effects, simulation scheme has to be different for studying the stability of size and power of the test, which we take as follows:

1. For comparing size stability, we choose the error ϵ randomly from a $(1-e_{err})N(0,3)+e_{err}N(10,3)$ distribution so that $100e_{err}\%$ (approximately) of the residuals are large compared to the pure data. Further, we will contaminate $100e_x\%$ proportion of the explanatory variable X by changing its ne_x values to be given from a $N(16,5)$ distribution, when the sample size is n. This will produce $100e_x\%$ leverage points in the sample data. Then, for each simulation, we take the true values of the parameters to be $(\beta_0, \beta_1, \sigma) = (3, 2, \sqrt{3})$, and the empirical size of the test are calculated.

2. Note that, the proposed DPD based test is consistent and so we need to compute its power only against some contiguous alternatives; we will take here these alternatives to be $H_{1,n} : (\beta_0,\ \beta_1) = (\beta_0^g,\ \beta_1^g) + \Delta_n$ with $\Delta_n = \frac{1}{\sqrt{2n}}$ for sample size n. Then, to study power stability of the proposed DPD based test, we again choose the error ϵ randomly from a $(1-e_{err})N(0,3)+e_{err}N(10,3)$ distribution generating $100e_{err}\%$ (approximately) large residuals compared to the pure data. But, we now contaminate $100e_x\%$ proportion of X in a different ways; by changing its ne_x values by $\left[x_i(\frac{2-\Delta}{2})^2 - \Delta_n\right]$ for sample size n. This will produce $100e_x\%$ leverage points with the specific characteristics that they pull the data towards null from the contiguous alternative under consideration. Then, for each simulation, we take the true values of the parameters to be at the contiguous alternative, $(\beta_0, \beta_1, \sigma) = (3 + \Delta_n, 2 + \Delta_n, \sqrt{3})$, and compute the

empirical power of the above mentioned test at the alternative $H_{1,n}$.

Below we will present the results for the case of known σ. The results for unknown σ case are similar and hence not reported here.

Size Stability for known σ

Let us assume that the error variance σ^2 is known to us and test for

$$H_0 : \beta_0 = 3, \beta_1 = 2 \tag{S5.18}$$

against the omnibus alternatives. We compute the empirical size of the DPD based test of this H_0 as per the above simulation scheme for sample sizes $n = 30$, 50 and 100 which are reported in Tables 2 to 4 respectively. The results are given for various degrees of contamination starting from no contamination to the heavy contamination of 20%. It is clear from the table that the likelihood ration test corresponding to the DPD based test with $\tau = 0$ is effected by the contamination most. Its sizes changes adversely in presence of very small proportion of contamination either in x or in error. However for the other DPD based test for larger values of τ the empirical size is highly stable in presence of small or moderate contamination proportion; it only changes a little bit for the heavy contamination of 15%-20% although the extend of change is very less that that of the likelihood ration test (LRT). Further, in case of no contamination also, the sizes of the DPD based test with different τ are very similar to each other; for larger τ the convergence to the asymptotic distribution used to obtain the critical values is slightly less than that of the LRT.

Power Stability for known σ

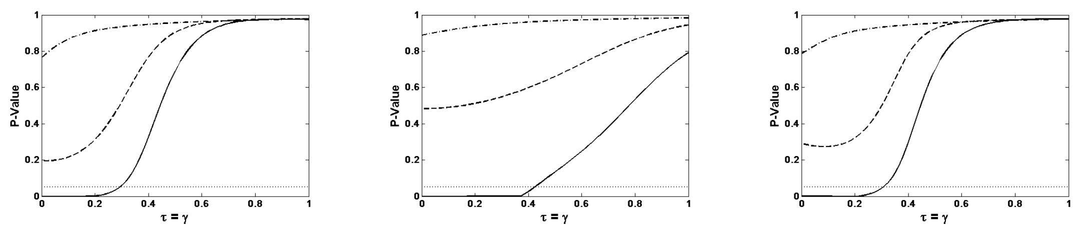
Now, let us see the power stability of the proposed DPD based test assuming the error variance σ^2 to be known and test for the hypothesis in (S5.18) against the contiguous alternatives $H_{1,n}$ as defined above. Then we calculate the empirical power of the DPD based test for various τ and sample sizes $n = 30$, 50 and 100 as per the above simulation scheme; the results obtained are reported in Tables 5 to 7 respectively. Again the likelihood ration test corresponding to the DPD based test with $\tau = 0$ is seen to be mostly effected by the insertion of contamination in data. It has very high power of 0.984 under pure data with small sample size of $n = 30$, which goes down to 0.421 for 20% contamination in error, or to 0.451 for 10% error and 16.7% leverage contamination for $n = 30$. But, under the same contamination scenarios and sample size, the DPD based test with $\tau = 0.5$ continues to provide high power of 0.755 and 0.688 respectively; also its power under no contamination is 0.955 for $n = 30$ which is very close to that of the LRT. Similar results are seen also for the larger sample sizes $n = 50$ and $n = 100$. Thus, the proposed DPD based test with large τ gives more power than the LRT in presence of contaminations, even under heavy contamination of 20%, and also has a comparable power under pure data. This proofs the usefulness of our proposed DPD based test with respect to the robustness of the process.

S6 Some Real Data Examples

S6.1 Hertzsprung-Russell Data: Normal linear regression

We consider the famous data of the Hertzsprung-Russell diagram of the star cluster CYG OB1 containing 47 stars in the direction of Cygnus (for details, see Rousseeuw and Leroy, 1987, Table 3, Chapter 2); the data are known to contain four huge outliers that affect the MLE and LRT drastically and is therefore useful in demonstrating the robust methodologies in simple linear regression. Here, the dependent variable is the logarithm of light intensity (L/L_0) of a star, which is to be modeled by a simple linear regression with only one predictor, the logarithm of its effective surface temperature (T_e). The corresponding parameter estimates based on the minimum DPD procedure were reported in Ghosh and Basu (2013); the robustness of the MDPDEs is quite clear from the results there.

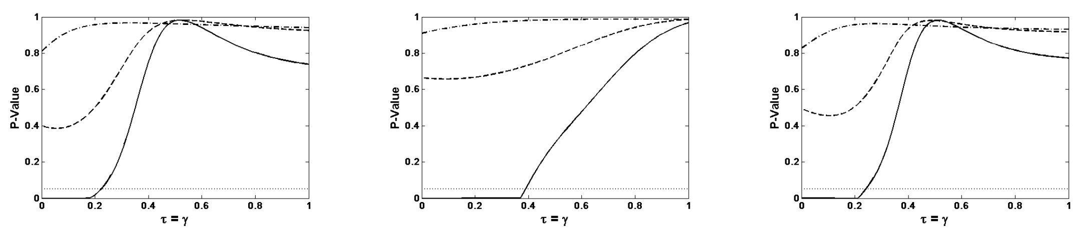
Here, we apply the proposed DPD based tests for testing the regression coefficients. Note that the estimates of regression coefficients $\boldsymbol{\beta}$ and error variance σ^2 differ greatly with τ due to the presence of outliers in the data set under consideration. So, we test for three different null hypotheses on $\boldsymbol{\beta}$ given by $H_0 : \boldsymbol{\beta} = (-8.03,\ 2.95)$, $H_0 : \boldsymbol{\beta} = (-7.22,\ 2.76)$ and $H_0 : \boldsymbol{\beta} = (6.7,\ -0.4)$ respectively; in the absence of the outliers, the LRT fails to reject either of the first two null hypotheses, but soundly rejects the third (this corresponds to the tests with moderately large values of the



(a) $H_0 : \boldsymbol{\beta} = (-8.03,\ 2.95)$ ($\sigma = 0.4$ known)

(b) $H_0 : \boldsymbol{\beta} = (-8.03,\ 2.95)$ ($\sigma = 0.6$ known)

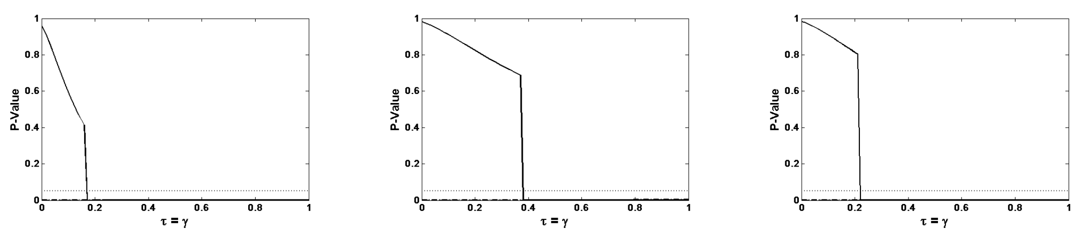
(c) $H_0 : \boldsymbol{\beta} = (-8.03,\ 2.95)$ (σ unknown)



(d) $H_0 : \boldsymbol{\beta} = (-7.22,\ 2.76)$ ($\sigma = 0.4$ known)

(e) $H_0 : \boldsymbol{\beta} = (-7.22,\ 2.76)$ ($\sigma = 0.6$ known)

(f) $H_0 : \boldsymbol{\beta} = (-7.22,\ 2.76)$ (σ unknown)



(g) $H_0 : \boldsymbol{\beta} = (6.7,\ -0.4)$ ($\sigma = 0.4$ known)

(h) $H_0 : \boldsymbol{\beta} = (6.7,\ -0.4)$ ($\sigma = 0.6$ known)

(i) $H_0 : \boldsymbol{\beta} = (6.7,\ -0.4)$ (σ unknown)

Figure 1: P-Values of the DPD based test for different H_0 on $\boldsymbol{\beta}$ with known and unknown σ^2 for the Hertzsprung-Russell data (solid line - for full data; dashed line - data after dropping 4 outliers; dashed-dotted line - data after dropping 5 outliers)

tuning parameter τ). For the full data, however, the LRT leads to the opposite conclusion in each case, whereas the conclusion for moderately large values of the tuning parameter remains unchanged." Figure 1 presents the p-values for testing these hypothesis based on the DPD based test over its tuning parameter $\tau = \gamma$ both with and without outliers; we considered the values of σ to be known at two values 0.4 (the robust estimate ignoring outlier) and 0.6 (the non-robust estimate in presence of outliers) and also repeat the test assuming σ to be unknown. It is clear from the figure that the DPD based test with $\tau \geq 0.3$ can ignore the outliers successfully leading to results which appropriately describes the homogeneous majority of the data. Even if the parameter σ is specified incorrectly the DPD based test gives appropriate robust results for larger values of $\tau \geq 0.5$. This example illustrates the strong robustness properties of the proposed DPD based tests with moderately large values of τ.

S6.2 Australia AIDS Data: Poisson Regression model

As an illustrative example of the generalized linear model, we will consider one of it's particular cases, namely the Poisson regression model applicable to count data. One can easily apply the proposed DPD based tests for testing under the Poisson regression model following the general theory developed in Section 6 of the main paper. Here, we will consider an interesting dataset on the counts of AIDS patients in Australia for successive quarters of 1984 to 1988; this dataset is obtained from Dobson (2002) and was previously analyzed by Ghosh and Basu (2016, Suppelentary material) to illustrate the robustness of the MDPDE under Poisson regression model.

The data on the AIDS count can be modeled by a Poisson regression model with the fixed covariate being the logarithm of time (slope) along with an intercept. Since there is no outlier in the original data, following Ghosh and Basu (2016, Suppelentary material) we also construct two artificial outliers by changing the observations at time 1 and time 20, from 1 to 10 and from 159 to 15 respectively. The corresponding MDPDEs under both the pure data and with outliers have been examined in Ghosh and Basu (2016, Suppelentary material); the work demonstrates strong robustness properties of these estimators with tuning parameter $\tau > 0$ over the classical MLE at $\tau = 0$ (see Table 1 for few examples).

Now, we will consider a simple linear hypothesis under this Poisson regression model given by

$$H_0 : \beta_1 = 3 \text{ against } H_1 : \beta_1 \neq 3, \tag{S6.19}$$

where β_1 is the slope coefficient corresponding to the covariate 'logarithmic time'. We apply the proposed DPD based test for this problem following Section 6 of the main paper, using the (unrestricted) MDPDEs of the slope and intercept coefficients and the restricted MDPDE of the intercept under H_0 (i.e., when the slope coefficient is 3). These estimates along with the resulting p-values for different values of tuning parameter $\tau = \gamma$ have been presented in Table 1 under both the original data without outlier and with two artificial outliers. Clearly, for the original data with no outliers, the null hypothesis in (S6.19) gets accepted for all values of tuning parameters $\tau = \gamma$.

However, in presence of two artificial outliers, the classical likelihood ratio test (at $\tau = \gamma = 0$) strongly rejects the null hypothesis but the proposed DPDTS with $\tau = \gamma > 0$ still accept the null hypothesis yielding robust inference.

Table 1: The unrestricted and restricted minimum density power divergence estimates and the resulting p-values for the AIDS Dataset without and with 2 artificial outliers

$\tau = \gamma$	0	0.1	0.2	0.5	1
Intercept (without outlier)	0.996	1.018	0.994	1.010	0.964
Slope (without outlier)	3.055	3.036	3.052	3.037	3.055
Restricted Intercept (without outlier)	1.057	1.057	1.055	1.048	1.038
Intercept (with 2 outliers)	1.730	1.278	1.227	1.022	1.155
Slope (with 2 outliers)	2.296	2.791	2.836	3.011	2.895
Restricted Intercept (with 2 outliers)	0.949	1.048	1.048	1.044	1.031
p-value (without outlier)	0.759	0.832	0.721	0.839	0.625
p-value (with 2 outliers)	0.000	0.321	0.431	0.761	1.000

S7 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 and Basu (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 in Section 7.2 of the main paper, 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 of the main paper. Similar phenomenon also hold for other datasets presented above in Section S6. 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., 2016) 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.

Table 2: The empirical size of the DPD based test with known σ for different contamination proportion and sample sizes $n = 30$.

e_x	e_{err}	$\tau = 0$	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 1$
0%	0%	0.050	0.048	0.046	0.045	0.044	0.044	0.045	0.044
0%	1%	0.067	0.061	0.057	0.052	0.050	0.049	0.050	0.050
0%	3%	0.112	0.097	0.087	0.070	0.064	0.061	0.061	0.059
0%	5%	0.176	0.147	0.127	0.093	0.080	0.077	0.072	0.068
0%	10%	0.378	0.322	0.273	0.185	0.145	0.130	0.113	0.099
0%	15%	0.593	0.526	0.463	0.321	0.244	0.213	0.175	0.147
0%	20%	0.767	0.707	0.643	0.476	0.364	0.314	0.256	0.206
3.3%	0%	0.866	0.446	0.211	0.092	0.078	0.074	0.073	0.075
3.3%	1%	0.851	0.443	0.220	0.104	0.086	0.080	0.080	0.082
3.3%	3%	0.826	0.442	0.246	0.129	0.106	0.098	0.094	0.090
3.3%	5%	0.810	0.454	0.280	0.159	0.130	0.119	0.111	0.104
3.3%	10%	0.801	0.527	0.394	0.261	0.206	0.183	0.162	0.143
3.3%	15%	0.828	0.637	0.543	0.390	0.311	0.272	0.232	0.199
3.3%	20%	0.876	0.749	0.688	0.534	0.429	0.380	0.315	0.261
10%	0%	1.000	0.969	0.538	0.160	0.129	0.122	0.116	0.112
10%	1%	1.000	0.964	0.543	0.172	0.139	0.131	0.124	0.119
10%	3%	1.000	0.955	0.557	0.200	0.162	0.150	0.142	0.133
10%	5%	1.000	0.947	0.575	0.236	0.188	0.173	0.162	0.151
10%	10%	0.999	0.935	0.638	0.338	0.264	0.244	0.216	0.196
10%	15%	0.999	0.937	0.727	0.464	0.362	0.331	0.291	0.251
10%	20%	0.998	0.945	0.810	0.598	0.482	0.436	0.374	0.318
20%	0%	1.000	1.000	0.997	0.507	0.312	0.269	0.234	0.215
20%	1%	1.000	1.000	0.997	0.518	0.324	0.280	0.245	0.226
20%	3%	1.000	1.000	0.996	0.541	0.354	0.308	0.268	0.245
20%	5%	1.000	1.000	0.995	0.569	0.386	0.337	0.294	0.265
20%	10%	1.000	1.000	0.993	0.640	0.466	0.413	0.355	0.315
20%	15%	1.000	1.000	0.992	0.725	0.561	0.505	0.439	0.381
20%	20%	1.000	1.000	0.993	0.799	0.653	0.593	0.516	0.450

Bibliography

Cantoni, E., and Ronchetti, E. (2001). Bounded Influence for Generalized Linear Models. *J. Amer. Statist. Assoc. 96*, 1022–1030.

Table 3: The empirical size of the DPD based test with known σ for different contamination proportion and sample sizes $n = 50$.

e_x	e_{err}	$\tau = 0$	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 1$
0%	0%	0.049	0.048	0.048	0.046	0.044	0.043	0.045	0.047
0%	1%	0.065	0.059	0.056	0.051	0.050	0.050	0.050	0.052
0%	3%	0.126	0.105	0.090	0.073	0.066	0.064	0.062	0.062
0%	5%	0.210	0.167	0.137	0.100	0.083	0.078	0.072	0.071
0%	10%	0.498	0.415	0.346	0.213	0.155	0.139	0.118	0.106
0%	15%	0.754	0.677	0.597	0.398	0.278	0.235	0.186	0.155
0%	20%	0.906	0.855	0.796	0.594	0.438	0.366	0.279	0.220
2%	0%	0.048	0.048	0.047	0.044	0.045	0.044	0.044	0.045
2%	1%	0.062	0.057	0.053	0.048	0.048	0.047	0.046	0.048
2%	3%	0.116	0.099	0.085	0.066	0.060	0.058	0.056	0.057
2%	5%	0.201	0.164	0.135	0.094	0.079	0.075	0.069	0.067
2%	10%	0.481	0.398	0.330	0.205	0.154	0.133	0.113	0.101
2%	15%	0.744	0.662	0.579	0.382	0.269	0.226	0.180	0.144
2%	20%	0.903	0.852	0.787	0.578	0.426	0.354	0.270	0.209
4%	0%	0.344	0.147	0.091	0.063	0.056	0.055	0.054	0.057
4%	1%	0.329	0.145	0.095	0.069	0.060	0.059	0.057	0.060
4%	3%	0.328	0.162	0.120	0.087	0.075	0.073	0.067	0.067
4%	5%	0.353	0.198	0.161	0.113	0.095	0.089	0.081	0.078
4%	10%	0.498	0.380	0.331	0.225	0.169	0.151	0.130	0.115
4%	15%	0.692	0.613	0.554	0.390	0.284	0.242	0.198	0.164
4%	20%	0.849	0.808	0.762	0.578	0.439	0.368	0.292	0.231
6%	0%	0.987	0.455	0.165	0.079	0.069	0.067	0.066	0.068
6%	1%	0.980	0.441	0.167	0.086	0.075	0.072	0.071	0.072
6%	3%	0.962	0.434	0.186	0.105	0.091	0.089	0.085	0.081
6%	5%	0.944	0.441	0.220	0.131	0.111	0.105	0.100	0.092
6%	10%	0.917	0.530	0.372	0.243	0.188	0.168	0.147	0.135
6%	15%	0.928	0.681	0.574	0.407	0.306	0.265	0.217	0.182
6%	20%	0.948	0.822	0.766	0.590	0.455	0.394	0.314	0.253
10%	0%	1.000	0.718	0.266	0.108	0.091	0.086	0.085	0.082
10%	1%	1.000	0.700	0.261	0.111	0.098	0.093	0.089	0.087
10%	3%	1.000	0.668	0.270	0.126	0.109	0.103	0.099	0.097
10%	5%	0.999	0.652	0.286	0.149	0.128	0.121	0.114	0.109
10%	10%	0.993	0.661	0.406	0.246	0.199	0.179	0.161	0.153
10%	15%	0.989	0.733	0.577	0.399	0.305	0.268	0.230	0.202
10%	20%	0.988	0.834	0.751	0.570	0.452	0.392	0.320	0.269
12%	0%	1.000	0.981	0.632	0.179	0.130	0.119	0.111	0.108

Table 4: The empirical size of the DPD based test with known σ for different contamination proportion and sample sizes $n = 100$.

e_x	e_{err}	$\tau = 0$	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 1$
0%	0%	0.050	0.049	0.049	0.050	0.049	0.049	0.048	0.048
0%	1%	0.071	0.065	0.061	0.059	0.055	0.054	0.053	0.052
0%	3%	0.160	0.132	0.110	0.081	0.070	0.068	0.062	0.059
0%	5%	0.302	0.241	0.192	0.118	0.092	0.084	0.075	0.071
0%	10%	0.713	0.612	0.515	0.298	0.195	0.160	0.129	0.110
0%	15%	0.934	0.884	0.811	0.560	0.375	0.295	0.213	0.165
0%	20%	0.993	0.979	0.956	0.788	0.584	0.475	0.336	0.240
1%	0%	0.365	0.097	0.067	0.056	0.055	0.055	0.054	0.053
1%	1%	0.347	0.102	0.075	0.065	0.062	0.061	0.059	0.057
1%	3%	0.357	0.149	0.119	0.089	0.079	0.073	0.068	0.065
1%	5%	0.421	0.239	0.197	0.125	0.099	0.091	0.083	0.077
1%	10%	0.688	0.576	0.508	0.307	0.207	0.171	0.138	0.119
1%	15%	0.907	0.855	0.802	0.567	0.382	0.306	0.225	0.176
1%	20%	0.983	0.971	0.949	0.792	0.591	0.488	0.345	0.253
5%	0%	1.000	0.729	0.275	0.123	0.110	0.107	0.103	0.103
5%	1%	1.000	0.707	0.273	0.132	0.117	0.115	0.110	0.108
5%	3%	1.000	0.682	0.286	0.155	0.138	0.133	0.125	0.119
5%	5%	1.000	0.678	0.328	0.192	0.163	0.154	0.142	0.131
5%	10%	1.000	0.767	0.539	0.356	0.269	0.237	0.202	0.179
5%	15%	1.000	0.894	0.785	0.589	0.433	0.366	0.290	0.239
5%	20%	1.000	0.969	0.934	0.794	0.619	0.530	0.404	0.315
7%	0%	1.000	0.975	0.512	0.156	0.127	0.122	0.117	0.115
7%	1%	1.000	0.969	0.503	0.162	0.135	0.130	0.124	0.119
7%	3%	1.000	0.956	0.496	0.187	0.156	0.150	0.140	0.132
7%	5%	1.000	0.945	0.509	0.222	0.180	0.171	0.158	0.147
7%	10%	1.000	0.938	0.632	0.374	0.286	0.254	0.220	0.196
7%	15%	1.000	0.962	0.807	0.594	0.449	0.382	0.305	0.257
7%	20%	1.000	0.983	0.935	0.787	0.626	0.539	0.422	0.337
8%	0%	1.000	0.994	0.568	0.173	0.142	0.136	0.132	0.128
9%	0%	1.000	0.997	0.602	0.179	0.149	0.145	0.138	0.136

Table 5: The empirical power of the DPD based test with known σ for different contamination proportion and sample sizes $n = 30$.

e_x	e_{err}	$\tau = 0$	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 1$
0%	0%	0.984	0.984	0.982	0.976	0.964	0.955	0.931	0.885
0%	1%	0.972	0.974	0.974	0.969	0.958	0.949	0.925	0.880
0%	3%	0.940	0.947	0.953	0.954	0.947	0.937	0.914	0.871
0%	5%	0.894	0.909	0.923	0.937	0.933	0.926	0.905	0.862
0%	10%	0.734	0.773	0.808	0.865	0.880	0.881	0.868	0.832
0%	15%	0.558	0.613	0.663	0.763	0.808	0.822	0.826	0.800
0%	20%	0.421	0.469	0.522	0.649	0.728	0.755	0.776	0.761
3.3%	0%	0.970	0.968	0.966	0.959	0.947	0.936	0.907	0.855
3.3%	1%	0.953	0.954	0.954	0.950	0.938	0.929	0.901	0.851
3.3%	3%	0.909	0.921	0.926	0.930	0.922	0.913	0.888	0.838
3.3%	5%	0.851	0.873	0.886	0.905	0.904	0.898	0.876	0.827
3.3%	10%	0.675	0.720	0.756	0.821	0.842	0.845	0.836	0.798
3.3%	15%	0.507	0.558	0.612	0.714	0.767	0.782	0.787	0.764
3.3%	20%	0.393	0.431	0.481	0.606	0.686	0.716	0.742	0.733
6.7%	0%	0.946	0.947	0.946	0.936	0.920	0.908	0.877	0.820
6.7%	1%	0.922	0.927	0.930	0.924	0.911	0.898	0.869	0.814
6.7%	3%	0.868	0.883	0.894	0.900	0.891	0.880	0.854	0.802
6.7%	5%	0.799	0.825	0.843	0.872	0.870	0.863	0.841	0.794
6.7%	10%	0.619	0.668	0.705	0.780	0.804	0.807	0.799	0.764
6.7%	15%	0.465	0.512	0.559	0.672	0.731	0.745	0.754	0.732
6.7%	20%	0.383	0.413	0.450	0.573	0.650	0.683	0.710	0.702
10%	0%	0.892	0.896	0.898	0.894	0.875	0.861	0.826	0.764
10%	1%	0.859	0.869	0.876	0.878	0.863	0.850	0.819	0.759
10%	3%	0.784	0.807	0.826	0.843	0.836	0.828	0.802	0.748
10%	5%	0.704	0.739	0.765	0.810	0.815	0.810	0.786	0.735
10%	10%	0.509	0.560	0.607	0.697	0.733	0.743	0.740	0.707
10%	15%	0.377	0.419	0.466	0.588	0.657	0.680	0.694	0.677
10%	20%	0.333	0.354	0.383	0.503	0.588	0.623	0.651	0.651
16.7%	0%	0.778	0.787	0.789	0.787	0.771	0.754	0.721	0.662
16.7%	1%	0.732	0.749	0.758	0.768	0.756	0.743	0.713	0.657
16.7%	3%	0.643	0.673	0.692	0.726	0.728	0.720	0.697	0.646
16.7%	5%	0.562	0.601	0.629	0.682	0.696	0.695	0.678	0.635
16.7%	10%	0.402	0.443	0.479	0.572	0.618	0.631	0.633	0.605
16.7%	15%	0.336	0.354	0.385	0.486	0.556	0.581	0.600	0.589
16.7%	20%	0.355	0.349	0.358	0.438	0.516	0.547	0.579	0.577
20%	0%	0.691	0.703	0.709	0.716	0.700	0.687	0.652	0.595
20%	1%	0.645	0.664	0.676	0.694	0.686	0.677	0.647	0.589
20%	3%	0.557	0.585	0.609	0.648	0.656	0.651	0.628	0.581
20%	5%	0.483	0.519	0.548	0.606	0.625	0.626	0.614	0.571
20%	10%	0.355	0.382	0.417	0.503	0.551	0.567	0.573	0.549
20%	15%	0.329	0.331	0.351	0.437	0.498	0.528	0.549	0.539
20%	20%	0.382	0.358	0.357	0.410	0.476	0.506	0.539	0.537

Table 6: The empirical power of the DPD based test with known σ for different contamination proportion and sample sizes $n = 50$.

e_x	e_{err}	$\tau = 0$	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 1$
0%	0%	0.992	0.992	0.991	0.987	0.982	0.976	0.962	0.927
0%	1%	0.982	0.984	0.984	0.982	0.976	0.971	0.956	0.924
0%	3%	0.950	0.959	0.965	0.971	0.967	0.963	0.946	0.914
0%	5%	0.898	0.921	0.935	0.953	0.953	0.951	0.936	0.901
0%	10%	0.696	0.754	0.801	0.875	0.898	0.902	0.895	0.871
0%	15%	0.485	0.548	0.614	0.749	0.814	0.833	0.845	0.832
0%	20%	0.365	0.404	0.455	0.611	0.712	0.751	0.789	0.794
2%	0%	0.990	0.990	0.990	0.986	0.979	0.973	0.958	0.922
2%	1%	0.979	0.981	0.982	0.979	0.974	0.968	0.952	0.918
2%	3%	0.945	0.956	0.961	0.967	0.964	0.957	0.943	0.908
2%	5%	0.891	0.912	0.927	0.948	0.949	0.945	0.930	0.896
2%	10%	0.693	0.749	0.793	0.866	0.891	0.894	0.889	0.863
2%	15%	0.492	0.549	0.611	0.743	0.807	0.826	0.836	0.822
2%	20%	0.389	0.421	0.465	0.608	0.702	0.743	0.781	0.783
6%	0%	0.978	0.977	0.977	0.971	0.961	0.953	0.931	0.885
6%	1%	0.962	0.964	0.965	0.962	0.953	0.945	0.923	0.880
6%	3%	0.908	0.923	0.934	0.942	0.938	0.930	0.910	0.869
6%	5%	0.834	0.864	0.885	0.913	0.917	0.910	0.893	0.855
6%	10%	0.611	0.665	0.716	0.810	0.843	0.850	0.846	0.818
6%	15%	0.437	0.485	0.540	0.680	0.749	0.771	0.788	0.778
6%	20%	0.384	0.396	0.426	0.551	0.647	0.688	0.730	0.741
10%	0%	0.958	0.959	0.958	0.951	0.939	0.926	0.898	0.842
10%	1%	0.932	0.938	0.941	0.938	0.929	0.916	0.892	0.836
10%	3%	0.864	0.884	0.897	0.910	0.907	0.899	0.873	0.822
10%	5%	0.780	0.813	0.837	0.871	0.879	0.874	0.853	0.811
10%	10%	0.554	0.606	0.660	0.750	0.796	0.803	0.801	0.773
10%	15%	0.422	0.456	0.503	0.624	0.697	0.720	0.739	0.728
10%	20%	0.413	0.407	0.421	0.520	0.607	0.643	0.687	0.698
16%	0%	0.865	0.872	0.874	0.867	0.847	0.835	0.803	0.744
16%	1%	0.818	0.834	0.844	0.846	0.832	0.822	0.792	0.738
16%	3%	0.713	0.745	0.767	0.797	0.800	0.795	0.769	0.723
16%	5%	0.603	0.649	0.685	0.742	0.762	0.762	0.746	0.704
16%	10%	0.399	0.443	0.495	0.604	0.659	0.676	0.686	0.668
16%	15%	0.341	0.352	0.380	0.491	0.571	0.607	0.637	0.637
16%	20%	0.418	0.388	0.376	0.433	0.503	0.541	0.589	0.611
20%	0%	0.809	0.817	0.819	0.810	0.796	0.779	0.743	0.686
20%	1%	0.754	0.770	0.781	0.789	0.778	0.766	0.736	0.678
20%	3%	0.646	0.679	0.704	0.736	0.740	0.737	0.715	0.666
20%	5%	0.543	0.583	0.619	0.681	0.699	0.704	0.693	0.650
20%	10%	0.382	0.412	0.452	0.551	0.603	0.621	0.633	0.618
20%	15%	0.372	0.368	0.384	0.467	0.530	0.559	0.588	0.592
20%	20%	0.485	0.439	0.413	0.431	0.485	0.515	0.559	0.576

Table 7: The empirical power of the DPD based test with known σ for different contamination proportion and sample sizes $n = 100$.

e_x	e_{err}	$\tau = 0$	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 1$
0%	0%	0.987	0.986	0.985	0.982	0.975	0.967	0.951	0.916
0%	1%	0.970	0.974	0.976	0.976	0.969	0.962	0.944	0.910
0%	3%	0.909	0.928	0.939	0.954	0.952	0.946	0.930	0.897
0%	5%	0.800	0.844	0.875	0.920	0.927	0.923	0.911	0.879
0%	10%	0.446	0.529	0.607	0.766	0.829	0.845	0.854	0.834
0%	15%	0.298	0.316	0.364	0.556	0.682	0.726	0.771	0.777
0%	20%	0.457	0.378	0.340	0.396	0.529	0.598	0.676	0.716
2%	0%	0.980	0.980	0.980	0.975	0.965	0.958	0.939	0.897
2%	1%	0.960	0.965	0.968	0.967	0.958	0.951	0.932	0.889
2%	3%	0.885	0.908	0.921	0.939	0.938	0.934	0.916	0.876
2%	5%	0.761	0.812	0.845	0.896	0.908	0.907	0.893	0.857
2%	10%	0.409	0.488	0.563	0.729	0.798	0.817	0.826	0.811
2%	15%	0.295	0.299	0.339	0.520	0.646	0.695	0.741	0.750
2%	20%	0.481	0.397	0.349	0.375	0.500	0.569	0.647	0.688
6%	0%	0.960	0.959	0.959	0.953	0.940	0.929	0.901	0.850
6%	1%	0.926	0.932	0.936	0.937	0.927	0.917	0.890	0.842
6%	3%	0.813	0.846	0.868	0.894	0.894	0.889	0.870	0.824
6%	5%	0.667	0.727	0.770	0.838	0.857	0.857	0.845	0.807
6%	10%	0.349	0.410	0.478	0.645	0.727	0.752	0.768	0.751
6%	15%	0.310	0.289	0.301	0.447	0.570	0.624	0.675	0.688
6%	20%	0.544	0.446	0.383	0.349	0.442	0.505	0.584	0.627
10%	0%	0.924	0.925	0.922	0.912	0.894	0.878	0.845	0.786
10%	1%	0.866	0.878	0.885	0.886	0.876	0.861	0.831	0.778
10%	3%	0.722	0.762	0.791	0.827	0.832	0.827	0.805	0.758
10%	5%	0.563	0.624	0.675	0.757	0.783	0.786	0.776	0.735
10%	10%	0.295	0.337	0.396	0.544	0.637	0.666	0.687	0.679
10%	15%	0.348	0.298	0.284	0.377	0.486	0.541	0.599	0.618
10%	20%	0.617	0.507	0.423	0.335	0.393	0.444	0.521	0.566
15%	0%	0.820	0.825	0.825	0.813	0.794	0.775	0.738	0.671
15%	1%	0.741	0.762	0.772	0.779	0.771	0.757	0.723	0.662
15%	3%	0.572	0.615	0.651	0.701	0.714	0.712	0.690	0.638
15%	5%	0.411	0.469	0.519	0.620	0.652	0.660	0.651	0.615
15%	10%	0.230	0.241	0.277	0.409	0.500	0.534	0.565	0.563
15%	15%	0.384	0.300	0.263	0.288	0.375	0.424	0.487	0.519
15%	20%	0.701	0.583	0.472	0.312	0.331	0.367	0.427	0.481
20%	0%	0.711	0.715	0.716	0.707	0.686	0.669	0.626	0.566
20%	1%	0.626	0.643	0.655	0.671	0.660	0.647	0.613	0.558
20%	3%	0.446	0.490	0.524	0.584	0.601	0.597	0.576	0.539
20%	5%	0.315	0.359	0.404	0.501	0.540	0.549	0.543	0.519
20%	10%	0.242	0.228	0.238	0.324	0.402	0.435	0.470	0.474
20%	15%	0.477	0.373	0.304	0.265	0.317	0.357	0.409	0.440
20%	20%	0.780	0.681	0.567	0.355	0.324	0.337	0.376	0.417

Dik, J. J., and de Gunst, M. C. M. (1985). The distribution of general quadratic forms in normal variables. *Statist. Neerlandica 39*, 14–26.

Dobson, A. J. (2002). *An introduction to generalised linear models.* Second edition. Chapman & Hall/CRC Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, Florida.

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

Ghosh, A., and Basu, A. (2015). Robust Estimation for Non-Homogeneous Data and the Selection of the Optimal Tuning Parameter: The DPD Approach. *J. Appl. Stat. 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.

Harville, D. A. (2008). *Matrix Algebra from a statistician's perspective.* Springer, New York.

Hosseinian, S. (2009). *Robust inference for generalized linear models: binary and Poisson regression.* Ph. D. thesis, Ecole Polytechnique Federal de Lausanne.

Kotz, S., Johnson, N. L., and Boyd, D. W. (1967). Series representations of distributions of quadratic forms in normal variables. I. Non-central case. *Ann. Math. Stat. 38*, 838–848.

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

Markatou, M., and He, X. (1994). Bounded-Influence and High-Breakdown-Point Testing Procedures in Linear Models. *J. Amer. Statist. Assoc. 89*, 543–549.

Markatou, M., and Hettmansperger, T. P. (1990). Robust Bounded-Influence Tests in Linear Models. *J. Amer. Statist. Assoc. 85*, 187–190.

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.

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

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

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

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

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

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

Schrader, R. M., and Hettmansperger, T. P. (1980). Robust Analysis of Variance Based Upon a Likelihood Ratio Criterion. *Biometrika 67*, 93–101.

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

Wang, L., and Qu, A. (2007). Robust Tests in Regression Models With Omnibus Alternatives and Bounded Influence. *J. Amer. Statist. Assoc. 102*, 347–358.