Abhik Ghosh[1] / Nirian Martin[2] / Ayanendranath Basu[1] / Leandro Pardo[3]

A New Class of Robust Two-Sample Wald-Type Tests

[1] Kolkata Interdisciplinary Statistical Research Unit 203, Indian Statistical Institute, B. T. Road, Kolkata - 700108, India, E-mail: ayanbasu@isical.ac.in
[2] Departamento de Estadistica e I.O., Complutense University of Madrid, II Avenida de Islas Filipinas 3, Madrid 28003, Spain
[3] Departamento de Estadistica e I.O., Complutense University of Madrid, Plaza de Ciencias 3, Madrid 28040, Spain

Abstract:
Parametric hypothesis testing associated with two independent samples arises frequently in several applications in biology, medical sciences, epidemiology, reliability and many more. In this paper, we propose robust Wald-type tests for testing such two sample problems using the minimum density power divergence estimators of the underlying parameters. In particular, we consider the simple two-sample hypothesis concerning the full parametric homogeneity as well as the general two-sample (composite) hypotheses involving some nuisance parameters. The asymptotic and theoretical robustness properties of the proposed Wald-type tests have been developed for both the simple and general composite hypotheses. Some particular cases of testing against one-sided alternatives are discussed with specific attention to testing the effectiveness of a treatment in clinical trials. Performances of the proposed tests have also been illustrated numerically through appropriate real data examples.

Keywords: robust hypothesis testing, two-sample problems, minimum density power divergence estimator, influence function, clinical trial

DOI: 10.1515/ijb-2017-0023

Received: March 14, 2017; **Revised:** January 6, 2018; **Accepted:** June 25, 2018

1 Introduction

Testing of parametric hypothesis is an important paradigm of statistical inference. In many real life applications like medical sciences, biology, epidemiology, sociology, reliability etc., we need to compare data from two independent samples through appropriate two-sample tests of hypotheses. Examples include, but are not limited to, comparing the means of any biomarker or success of any treatment between control and treatment groups, comparing lifetime of two populations in reliability, etc.

Mathematically, let $(\mathcal{X}, \beta_{\mathcal{X}}, P_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}$ be the statistical space associated with the random variable X, where $\beta_{\mathcal{X}}$ is the σ-field of Borel subsets $A \subset \mathcal{X}$ and $\{P_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}$ is a family of probability distributions defined on the measurable space $(\mathcal{X}, \beta_{\mathcal{X}})$ where $\boldsymbol{\Theta}$ is an open subset of $\mathbb{R}^p$, with $p \geq 1$. Probability measures $P_{\boldsymbol{\theta}}$ are assumed to possess the densities $f_{\boldsymbol{\theta}}(x) = dP_{\boldsymbol{\theta}}/d\mu(x)$, where μ is a σ-finite measure on $(\mathcal{X}, \beta_{\mathcal{X}})$. We shall denote by $\mathcal{F} = \{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\}$ a set of parametric model densities.

On the basis of two independent random samples $X_1, ..., X_n$ and $Y_1, ..., Y_m$ of sizes n and m, respectively, from two densities $f_{\boldsymbol{\theta}_1}(x)$ and $f_{\boldsymbol{\theta}_2}(x)$ belonging to $\mathcal{F}$, we can solve the problem of complete homogeneity by testing

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \text{ versus } H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2. \tag{1}$$

The classical test statistics for testing eq. (1) are the likelihood ratio test, Wald test and Rao test, where the unknown parameters are estimated by the maximum likelihood estimators (MLEs). Some alternative test statistics have also been presented in the literature based on divergence measures; see, for instance, Basu et al. [1] and Pardo [2]. It is well-known that the MLE is a BAN estimator, i.e., asymptotically efficient, but at the same time it has serious lack of robustness against data contamination and model misspecification. In order to avoid the robustness problem, appropriate testing procedures have been developed in the statistical literature based on suitable robust estimators. For example, Basu et al. [3] have introduced a family of test statistics for testing eq. (1) based on the density power divergence (DPD) measure between $f_{\boldsymbol{\theta}_1}$ and $f_{\boldsymbol{\theta}_2}$ when the parameters are estimated by the minimum density power divergence estimator (MDPDE) of Basu et al. [4]; see Section 1.1 for more details about the MDPDE.

Ayanendranath Basu is the corresponding author.

Note that, if the problem considered in eq. (1) has been solved, we will be able to apply it for the particular (and most common) normal populations with $f_{\boldsymbol{\theta}_1} \equiv N(\mu_1, \sigma_1)$ and $f_{\boldsymbol{\theta}_2} \equiv N(\mu_2, \sigma_2)$ to test the following problem of complete homogeneity

$$H_0 : (\mu_1, \sigma_1) = (\mu_2, \sigma_2) \text{ versus } H_1 : (\mu_1, \sigma_1) \neq (\mu_2, \sigma_2) .$$

But there are other interesting problems of testing for partial homogeneity, for instance, to test

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2$$

when the variances are the same but unknown; this is a particular case of the general composite hypotheses involving two samples. This particular problem with normal population has been considered in [5] on the basis of a family of test statistics based on the DPD measure and by estimating the unknown parameters using the MDPDE. The results presented in their paper have been excellent in relation to the robustness and efficiency trade-off; for some suggested members of their proposed test family the loss in efficiency based on the size and the power under pure data was not really significant but the improvement in terms of robustness under contaminated data was highly significant. Although their approach can theoretically be extended beyond the simple case of normal populations, from a practical point of view, it is often not very easy to compute the density power divergence measure between $f_{\boldsymbol{\theta}_1}$ and $f_{\boldsymbol{\theta}_2}$. In this paper we present a new family of test statistics which are easy to calculate based on only the MDPDEs for any general two-sample problem (involving nuisance parameters as well) and with any parametric distribution. These test statistics are Wald-type test statistics and their usefulness have been illustrated in the literature of one sample testing problems by Basu et al. [6] and Ghosh et al. [7]. In the present paper, not only will we present the asymptotic distribution of the proposed Wald-type test statistics for the two-sample problems but will also provide a theoretical study of their robustness properties along with suitable examples and numerical illustrations.

The rest of the paper is organized as follows: In Section 1.1 we present some important background results and definitions in relation to the MDPDE that will be necessary for the rest of the paper. Section 2 is devoted to developing the family of Wald-type tests for solving the problem of complete homogeneity given by eq. (1). We study its asymptotic distribution as well as the theoretical robustness properties with examples in the same section. In Section 3, we present a family of Wald-type tests for the more general composite hypotheses in the two sample context. We again derive their asymptotic distributions and robustness properties. Illustrations are provided for the special case of testing partial homogeneity in presence of nuisance parameters like, for example, testing equality of two normal means with unknown (nuisance) variances. In Section 4, we briefly describe the extensions for testing the two-sample hypotheses against one-sided alternatives. Section 5 presents several real life applications of our proposal with interesting data from applied sciences like medical science, biology, reliability etc. Appropriate simulation studies with some comments on the choice of the robustness tuning parameters are presented in Section 6. The paper ends with a short concluding remark in Section 7. For brevity in presentation, the proofs of all the results have been moved to Appendix 8.

1.1 The minimum density power divergence estimator: Asymptotic properties and robustness

Given any two densities $f_{\boldsymbol{\theta}_1}$ and $f_{\boldsymbol{\theta}_2}$ from $\mathcal{F}$, the density power divergence with a nonnegative tuning parameter β, is defined as [4]

$$d_\beta(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2}) = \begin{cases} \int \left\{ f_{\boldsymbol{\theta}_2}^{1+\beta}(x) - \left(1 + \frac{1}{\beta}\right) f_{\boldsymbol{\theta}_2}^{\beta}(x) f_{\boldsymbol{\theta}_1}(\boldsymbol{x}) + \frac{1}{\beta} f_{\boldsymbol{\theta}_1}^{1+\beta}(x) \right\} dx, & \text{for } \beta > 0, \\ \int f_{\boldsymbol{\theta}_1}(x) \ln \left(\frac{f_{\boldsymbol{\theta}_1}(x)}{f_{\boldsymbol{\theta}_2}(x)} \right) dx, & \text{for } \beta = 0. \end{cases} \tag{2}$$

The divergence corresponding to $\beta = 0$ may be derived from the general case by taking the continuous limit as $\beta \to 0^+$, and the resulting $d_0(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2})$ turns out to be the Kullback-Leibler divergence.

Let G represent the distribution function corresponding to the underlying true density g that generates the data and we want to model it by the parametric model density $f_{\boldsymbol{\theta}} \in \mathcal{F}$. The corresponding minimum DPD functional at G with tuning parameter β, denoted by $\boldsymbol{U}_\beta(G)$, is defined through the relation $d_\beta(g, f_{\boldsymbol{U}_\beta(G)}) = \min_{\boldsymbol{\theta} \in \Theta} d_\beta(g, f_{\boldsymbol{\theta}})$. Therefore the MDPDE of $\boldsymbol{\theta}$ with tuning parameter β is given by $\hat{\boldsymbol{\theta}}_\beta = \boldsymbol{U}_\beta(G_n)$,

where G_n is the empirical distribution function associated with the observed random sample $X_1, \ldots, X_n$ from the population having density g. As the last term of eq. (2) does not depend on $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_\beta$ is indeed given by

$$\hat{\boldsymbol{\theta}}_\beta = \arg\min_{\boldsymbol{\theta} \in \Theta} \left\{ \int f_{\boldsymbol{\theta}}^{1+\beta}(x) dx - \left(1 + \frac{1}{\beta}\right) \frac{1}{n} \sum_{i=1}^{n} f_{\boldsymbol{\theta}}^{\beta}(X_i) \right\}, \quad \text{if } \beta > 0, \tag{3}$$

$$\text{and} \qquad \hat{\boldsymbol{\theta}}_\beta = \arg\min_{\boldsymbol{\theta}\in\Theta}\left\{-\tfrac{1}{n}\sum_{i=1}^{n}\ln f_{\boldsymbol{\theta}}(X_i)\right\}, \qquad \text{if } \beta = 0. \tag{4}$$

Notice that $\hat{\boldsymbol{\theta}}_\beta$ for $\beta = 0$ coincides with the maximum likelihood estimator (MLE). Denoting

$$V_{\boldsymbol{\theta}}(x) = \int f_{\boldsymbol{\theta}}^{1+\beta}(\boldsymbol{x})d\boldsymbol{x} - \left(1+\frac{1}{\beta}\right)f_{\boldsymbol{\theta}}^{\beta}(x),$$

the expression in eq. (3) can be written as $\hat{\boldsymbol{\theta}}_\beta = \arg\min_{\boldsymbol{\theta}\in\Theta}\frac{1}{n}\sum_{i=1}^{n}V_{\boldsymbol{\theta}}(X_i)$. It shows that the MDPDE is an M-estimator.

The functional $\boldsymbol{U}_\beta(G)$ is Fisher consistent; it takes the value $\boldsymbol{\theta}_0$, the true value of the parameter, when the true density is a member of the model with $g = f_{\boldsymbol{\theta}_0}$. Let us assume $g = f_{\boldsymbol{\theta}_0}$ and define the quantities

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \int \boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x})\boldsymbol{u}_{\boldsymbol{\theta}}^T(\boldsymbol{x})f_{\boldsymbol{\theta}}^{1+\beta}(\boldsymbol{x})d\boldsymbol{x}, \quad \boldsymbol{K}_\beta(\boldsymbol{\theta}) = \int \boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x})\boldsymbol{u}_{\boldsymbol{\theta}}^T(\boldsymbol{x})f_{\boldsymbol{\theta}}^{1+2\beta}(\boldsymbol{x})d\boldsymbol{x} - \boldsymbol{\xi}_\beta(\boldsymbol{\theta})\boldsymbol{\xi}_\beta^T(\boldsymbol{\theta}), \tag{5}$$

where $\boldsymbol{\xi}_\beta(\boldsymbol{\theta}) = \int \boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x})f_{\boldsymbol{\theta}}^{1+\beta}(\boldsymbol{x})d\boldsymbol{x}$ and $\boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x}) = \frac{\partial}{\partial\boldsymbol{\theta}}\ln f_{\boldsymbol{\theta}}(\boldsymbol{x})$. Then, following [1, 4], it can be shown that, under Assumptions (D1)–(D5) of Basu et al. [1][p. 304] to be referred as "Basu et al. conditions" in the rest of the paper,

$$n^{1/2}(\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0) \xrightarrow[n\to\infty]{\mathcal{L}} N(\boldsymbol{0}_p, \boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)), \tag{6}$$

where $\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}) = \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta})\boldsymbol{K}_\beta(\boldsymbol{\theta})\boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta})$. It is a simple exercise to see that for $\beta = 0$, $\boldsymbol{J}_{\beta=0}(\boldsymbol{\theta}) = \boldsymbol{K}_{\beta=0}(\boldsymbol{\theta}) = \boldsymbol{I}_F(\boldsymbol{\theta})$, the Fisher information matrix associated to the model under consideration. Therefore we obtain the classical well known result,

$$n^{1/2}(\hat{\boldsymbol{\theta}}_{\beta=0} - \boldsymbol{\theta}_0) \xrightarrow[n\to\infty]{\mathcal{L}} N(\boldsymbol{0}_p, \boldsymbol{I}_F^{-1}(\boldsymbol{\theta}_0)).$$

Next, the influence function (IF) can be used to study the robustness of the MDPDE. Note that, if the influence function is bounded, the corresponding estimator or test statistic is said to have local robustness against infinitesimal contamination. More simply, the influence function $\mathcal{IF}\left(x, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}_0}\right)$ is the first derivative of an estimator or statistic viewed as a functional $\boldsymbol{U}_\beta$, which describes the normalized influence on the estimate or statistic of an infinitesimal contamination at a distant point x in the sample space. In [4] it was established that the influence function (IF) of the minimum DPD functional is given by

$$\mathcal{IF}\left(x, \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}_0}\right) = \lim_{\varepsilon\to 0}\frac{\boldsymbol{U}_\beta(F_\varepsilon) - \boldsymbol{U}_\beta\left(F_{\boldsymbol{\theta}_0}\right)}{\varepsilon} = \boldsymbol{J}_\beta^{-1}(\boldsymbol{\theta}_0)\left(\boldsymbol{u}_{\boldsymbol{\theta}}(x)f_{\boldsymbol{\theta}_0}^{\beta}(x) - \boldsymbol{\xi}(\boldsymbol{\theta}_0)\right), \tag{7}$$

where $F_\varepsilon = (1-\varepsilon)F_{\boldsymbol{\theta}_0} + \varepsilon\wedge_x$ is the ε-contaminated distribution of $F_{\boldsymbol{\theta}_0}$, the distribution function corresponding to $f_{\boldsymbol{\theta}}$, with respect to the point mass distribution $\wedge_x$ at x. If we assume that $\boldsymbol{J}_\beta(\boldsymbol{\theta}_0)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_0)$ are finite, the IF is a bounded function of x whenever $\boldsymbol{u}_{\boldsymbol{\theta}}(x)f_{\boldsymbol{\theta}_0}^{\beta}(x)$ is bounded. And this is the case for most common parametric models at $\beta > 0$ implying the robustness of MDPDEs with $\beta > 0$.

2 A simple two-sample problem

Let $X_1, ..., X_n$ and $Y_1, ..., Y_m$ be two samples of sizes n and m respectively from two populations having densities belonging to $\mathcal{F}$ with parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. The most common problem under this setup is to test the complete homogeneity of the two populations given by eq. (1). But some component of the parameters can also be just nuisance in many applications, for example, as in eq. (2).

In general notation, let us assume that

$$\boldsymbol{\theta}_1 = \left(\theta_{1,1}, ..., \theta_{1,r}, \theta_{1,r+1}, ..., \theta_{1,p}\right)^T = \left({}^*\boldsymbol{\theta}_1^T, {}^0\boldsymbol{\theta}_1^T\right)^T \text{ and } \boldsymbol{\theta}_2 = \left(\theta_{21,1}, ..., \theta_{2,r}, \theta_{2,r+1}, ..., \theta_{2,p}\right)^T = \left({}^*\boldsymbol{\theta}_2^T, {}^0\boldsymbol{\theta}_2^T\right)^T$$

with ${}^0\boldsymbol{\theta}_1$ and ${}^0\boldsymbol{\theta}_2$ known $(p-r)$-vectors. Based on $X_1, ..., X_n$ we can get the MLE, ${}^*\hat{\boldsymbol{\theta}}_1$, of ${}^*\boldsymbol{\theta}_1$ and based on $Y_1, ..., Y_m$ the MLE, ${}^*\hat{\boldsymbol{\theta}}_2$, of ${}^*\boldsymbol{\theta}_2$. Assuming ${}^*\boldsymbol{\theta}_1 = {}^*\boldsymbol{\theta}_2$ we can obtain an estimator, ${}^*_{(o)}\hat{\boldsymbol{\theta}}_1$ of the common value ${}^*\boldsymbol{\theta}_1$ by using the two random samples $X_1, ..., X_n$ and $Y_1, ..., Y_m$ together. It is well-known that, under ${}^*\boldsymbol{\theta}_1 = {}^*\boldsymbol{\theta}_2$,

$$\sqrt{\frac{mn}{m+n}}\left({}^*\hat{\boldsymbol{\theta}}_1 - {}^*\hat{\boldsymbol{\theta}}_2\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(\boldsymbol{0}_r, \omega\boldsymbol{I}_F^{-1}({}^*\boldsymbol{\theta}_1, {}^0\boldsymbol{\theta}_1) + (1-\omega)\boldsymbol{I}_F^{-1}({}^*\boldsymbol{\theta}_1, {}^0\boldsymbol{\theta}_2)) \tag{8}$$

with

$$\omega = \lim_{m,n\to\infty} \frac{m}{m+n}.$$

Based on eq. (8), the classical Wald test statistic for testing

$$H_0 :^* \boldsymbol{\theta}_1 =^* \boldsymbol{\theta}_2 \text{ versus } H_1 :^* \boldsymbol{\theta}_1 \neq^* \boldsymbol{\theta}_2, \tag{9}$$

is given by $W_{m,n} = \frac{mn}{m+n}(*\hat{}_1 - *\hat{}_2)^T \left(\frac{m\mathbf{I}_F^{-1}(*\hat{}_1, {}^0{}_1)}{m+n} + \frac{n\mathbf{I}_F^{-1}(*\hat{}_2, {}^0{}_2)}{m+n} \right)^{-1} (*\hat{}_1 - *\hat{}_2)$

$$= mn\,(*\hat{}_1 - *\hat{}_2)^T \left(m\mathbf{I}_F^{-1}(*\hat{}_1, {}^0{}_1) + n\mathbf{I}_F^{-1}(*\hat{}_2, {}^0{}_2) \right)^{-1} (*\hat{}_1 - *\hat{}_2).$$

We can observe that, when $r = p$ we have $\boldsymbol{I}_F^{-1}(*\boldsymbol{\theta}_1, {}^0\boldsymbol{\theta}_1) = \boldsymbol{I}_F^{-1}(*\boldsymbol{\theta}_1, {}^0\boldsymbol{\theta}_2) = \boldsymbol{I}_F^{-1}(\boldsymbol{\theta}_0)$, with $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}_0$ and the Wald test statistic becomes

$$W_{m,n} = \frac{mn}{m+n} \left(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2 \right)^T \boldsymbol{I}_F({}^{(0)}\hat{\boldsymbol{\theta}}) \left(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2 \right), \tag{10}$$

where ${}^{(0)}\hat{\boldsymbol{\theta}}$ denotes the MLE of $\boldsymbol{\theta}_0$ based on the pooled sample.

As an example, in the case of two normal populations, with known variances σ_1^2 and σ_2^2, we can test $H_0 : \mu_1 = \mu_2$ by the Wald test statistic

$$W_{m,n} = mn \frac{(\hat{\mu}_1 - \hat{\mu}_2)^2}{m\sigma_1^2 + n\sigma_2^2} = \frac{(\hat{\mu}_1 - \hat{\mu}_2)^2}{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}.$$

Although it has several nice optimum properties, it is highly non-robust in presence of outliers even in any one sample. Here, we will generalize this classical Wald test to make it robust through replacing the non-robust MLEs by the corresponding robust MDPDEs. In the following we will present the results for $r = p$, i.e., to test for the hypothesis in eq. (1). The case $r = p$ can be studied in a similar way.

Let us assume ${}^{(1)}\hat{\boldsymbol{\theta}}_\beta$ and ${}^{(2)}\hat{\boldsymbol{\theta}}_\beta$ denote the MDPDEs of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ respectively, obtained by minimizing the DPD with tuning parameter β for each of the two samples separately. Further, under the null hypothesis $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}_0$ in eq. (1), we can consider the two samples pooled together as one i.i.d. sample of size $m+n$ from a population having density function $f_{\boldsymbol{\theta}_0}$; let ${}^{(0)}\hat{\boldsymbol{\theta}}_\beta$ denote the corresponding MDPDE of $\boldsymbol{\theta}_0$ with tuning parameter β based on the pooled sample. Note that, all the three estimators ${}^{(1)}\hat{\boldsymbol{\theta}}_\beta$, ${}^{(2)}\hat{\boldsymbol{\theta}}_\beta$ and ${}^{(0)}\hat{\boldsymbol{\theta}}_\beta$ should coincide with $\boldsymbol{\theta}_0$ asymptotically under H_0 with probability tending to one. Assuming identifiability of the model family, the difference between the two estimators ${}^{(1)}\hat{\boldsymbol{\theta}}_\beta$ and ${}^{(2)}\hat{\boldsymbol{\theta}}_\beta$ gives us an idea of the distinction between the two samples and hence indicate any departure from the null hypothesis. So, we define a generalized Wald-type test statistic by

$$T_{m,n}^{(\beta)} = \frac{nm}{n+m} \left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - {}^{(2)}\hat{\boldsymbol{\theta}}_\beta \right)^T \boldsymbol{\Sigma}_\beta \left({}^{(0)}\hat{\boldsymbol{\theta}}_\beta \right)^{-1} \left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - {}^{(2)}\hat{\boldsymbol{\theta}}_\beta \right). \tag{11}$$

Note that, at $\beta = 0$, all the MDPDEs used coincide with corresponding MLEs and hence the generalized Wald-type test statistic $T_{m,n}^{(\beta)}$ coincides with the classical Wald test statistic $W_{m,n}$ given in eq. (10).

2.1 Asymptotic properties

In order to perform any statistical test, we first need to derive the asymptotic distribution of the test statistics under H_0. Using the asymptotic properties of the MDPDEs presented in Section 1.1, we can easily obtain the asymptotic null distribution of the proposed test statistics $T_{m,n}^{(\beta)}$ which is presented in the following theorem. Throughout the rest of the paper, we will assume Conditions (A)–(D) of Lehmann [8][p. 429] about the assumed model family which we will refer as "Lehmann conditions". Also, we consider the following assumption.

Assumption (A):

1. $\frac{m}{m+n} \to \omega \in (0,1)$ as $m, n \to \infty$
2. The asymptotic variance-covariance matrix $\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta})$ of the MDPDE with tuning parameter β is continuous in $\boldsymbol{\theta}$.

Theorem 2.1

Suppose the model density satisfies the Lehmann and Basu et al. conditions, and Assumption (A) holds. Then the asymptotic distribution of $T_{m,n}^{(\beta)}$ under the null hypothesis in eq. (9) is χ_p^2, the chi-square distribution with p degrees of freedom.

The asymptotic null distribution of the test in [3] is a linear combination of chi-square distribution and hence it is somewhat difficult to obtain the critical values of their test in practice. On the contrary, our proposed tests have a simple chi-square limit under the null hypothesis and hence are much easier to perform. Our proposal provides, in this sense, an advantageous procedure for testing.

However, when the null hypothesis is not correct, i.e., $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, then the pooled estimator ${}^{(0)}\hat{\boldsymbol{\theta}}_\beta$ no longer converges to $\boldsymbol{\theta}_1$ or $\boldsymbol{\theta}_2$; rather it will then converge in probability to a new value $\boldsymbol{\theta}_3$, say, which is a function of $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and ω. For example, if the estimators are additive in sample data, e.g. sample mean, then $\boldsymbol{\theta}_3 = (1-\omega)\boldsymbol{\theta}_1 + \omega\boldsymbol{\theta}_2$. Define $l^*_{\boldsymbol{\theta}_3,\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = (\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_3)^{-1}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)$. Then we have the following result.

Theorem 2.2

Suppose the model density satisfies the Lehmann and Basu et al. conditions, and Assumption (A) holds. Then, as $m,n \to \infty$, we have for any $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$

$$\sqrt{\frac{mn}{m+n}}\left[l^*_{{}^{(0)}\hat{\boldsymbol{\theta}}_\beta,\beta}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta) - l^*_{\boldsymbol{\theta}_3,\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\right] \xrightarrow[m,n\to\infty]{\mathscr{L}} N\left(0, 4\sigma^2_{\boldsymbol{\theta}_3,\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\right),$$

where $\sigma^2_{\boldsymbol{\theta}_3,\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = (\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_3)^{-1}\left[\omega\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_1) + (1-\omega)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_2)\right]\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_3)^{-1}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)$.

This theorem leads to an approximation to the power function $\pi^{(\beta)}_{m,n,\alpha}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = P\left(T^{(\beta)}_{m,n} > \chi^2_{p,\alpha}\right)$ of the proposed Wald-type tests for testing eq. (9) at the significance level α, where $\chi^2_{p,\alpha}$ denotes the $(1-\alpha)$-th quantile of the χ^2_p distribution.

Corollary 2.3

Under the assumption of Theorem 2.2, we have

$$\pi^{(\beta)}_{m,n,\alpha}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = 1 - \Phi_n\left(\frac{\sqrt{\frac{n+m}{nm}}}{2\sigma_{\boldsymbol{\theta}_3,\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}\left[\chi^2_{p,\alpha} - \frac{nm}{n+m}l^*_{\boldsymbol{\theta}_3,\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\right]\right), \qquad \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2,$$

for a sequence of distributions $\Phi_n(\cdot)$ tending uniformly to the standard normal distribution $\Phi(\cdot)$.

The corollary also helps us to determine the sample size requirement for our proposed test to achieve any pre-specified power level. Further, we have $\pi^{(\beta)}_{m,n,\alpha}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \to 1$ for any $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ as $m,n \to \infty$. Hence the proposed test with rejection rule $\left\{T^{(\beta)}_{m,n} > \chi^2_{p,\alpha}\right\}$ is consistent.

Corollary 2.4

Under the assumption of Theorem Theorem 2.2, the proposed Wald-type test is consistent in the Fraser's sense.

Next, we look at the performance of the proposed test under contiguous alternatives. Now, in case of two sample problem, we can have different types of contiguous alternatives. For example, we can assume $\boldsymbol{\theta}_2$ to be fixed and $\boldsymbol{\theta}_1$ converging to $\boldsymbol{\theta}_2$ so that $H'_{1,n} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1,n} = \boldsymbol{\theta}_2 + n^{-\frac{1}{2}}\boldsymbol{\Delta}_1$ for some p-vector $\boldsymbol{\Delta}_1$ of non-zero reals such that $\boldsymbol{\theta}_2 + n^{-\frac{1}{2}}\boldsymbol{\Delta}_1 \in \Theta$. Conversely, we can have $\boldsymbol{\theta}_1$ to be fixed and $H''_{1,m} : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{2,m} = \boldsymbol{\theta}_1 + m^{-\frac{1}{2}}\boldsymbol{\Delta}_2$ for some $\boldsymbol{\Delta}_2 \in \mathbb{R}^p - \{\mathbf{0}\}$ with $\boldsymbol{\theta}_1 + m^{-\frac{1}{2}}\boldsymbol{\Delta}_2 \in \Theta$. Here, we consider a general form of the contiguous alternative given by

$$H_{1,n,m} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1,n} = \boldsymbol{\theta}_0 + n^{-\frac{1}{2}}\boldsymbol{\Delta}_1, \;\; \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{2,m} = \boldsymbol{\theta}_0 + m^{-\frac{1}{2}}\boldsymbol{\Delta}_2, \qquad (\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2) \in \mathbb{R}^p \times \mathbb{R}^p - \{(\mathbf{0}_p,\mathbf{0}_p)\}, \tag{12}$$

for some fixed $\boldsymbol{\theta}_0 \in \Theta$. Note that, putting $\boldsymbol{\Delta}_2 = \mathbf{0}$ in eq. (12) we get $H'_{1,n}$ back from $H_{1,n,m}$, whereas $\boldsymbol{\Delta}_1 = \mathbf{0}$ yields $H''_{1,m}$. The following theorem gives the asymptotic distribution of the proposed test statistics $T^{(\beta)}_{m,n}$ under this general contiguous alternatives $H_{1,m,n}$.

Theorem 2.5

Suppose the model density satisfies the Lehmann and Basu et al. conditions and the assumption (A) holds. Then the asymptotic distribution of $T^{(\beta)}_{m,n}$ under the contiguous alternative $H_{1,n,m}$ given by eq. (12) is $\chi^2_p(\delta_\beta)$, the non-central chi-square distribution with p degrees of freedom and non-centrality parameter $\delta_\beta = \boldsymbol{W}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)^T\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)^{-1}\boldsymbol{W}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)$ with $\boldsymbol{W}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2) = \left[\sqrt{\omega}\boldsymbol{\Delta}_1 - \sqrt{1-\omega}\boldsymbol{\Delta}_2\right]$.

We can easily obtain the asymptotic power $\pi_\beta(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)$ under the contiguous alternatives $H_{1,n,m}$ from the above theorem. In particular, denoting the distribution function of a random variable Z by F_Z, we have

$$\pi_\beta(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2) = 1 - F_{\chi^2_p(\delta_\beta)}(\chi^2_{p,\alpha}). \tag{13}$$

Example 2.1 (Testing equality of two Normal means with known equal variances)

We first present the simplest possible case of testing two normal means with known equal variance σ^2. Here the model family is $\mathcal{F} = \{N(\theta, \sigma^2) : \theta \in \mathbb{R}\}$ with σ being known. In this case, the asymptotic variance $\Sigma_\beta(\theta)$ of the MDPDE with tuning parameter β is given by $\Sigma_\beta(\theta) = \left(1 + \frac{\beta^2}{1+2\beta}\right)^{3/2} \sigma^2$. Hence, our generalized Wald-type test statistics has a much simpler form in this case given by

$$T_{m,n}^{(\beta)} = \frac{mn}{m+n}\left(1 + \frac{\beta^2}{1+2\beta}\right)^{-3/2}\left(\frac{^{(1)}\hat{\theta}_\beta - ^{(2)}\hat{\theta}_\beta}{\sigma}\right)^2,$$

and it has χ_1^2 asymptotic distribution under H_0. Note that, at $\beta = 0$, this test statistic coincides with the classical Wald-test statistic $W_{m,n} = \frac{mn}{m+n}\left(\frac{^{(1)}\hat{\theta}_0 - ^{(2)}\hat{\theta}_0}{\sigma}\right)^2 = \frac{mn}{m+n}\left(\frac{\bar{X}-\bar{Y}}{\sigma}\right)^2$, where $\bar{X}$ and $\bar{Y}$ are the sample means of $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ respectively.

Clearly, these tests are consistent for any $\beta \geq 0$ by Corollary 2.4. Further, the asymptotic power of the proposed test under contiguous alternatives $H_{1,m,n}$ can be easily obtained as

$$\pi_\beta(\Delta_1, \Delta_2) = 1 - F_{\chi_1^2(\delta_\beta)}(\chi_{1,\alpha}^2),$$

with $\delta_\beta = \left(1 + \frac{\beta^2}{1+2\beta}\right)^{-3/2} \sigma^{-2} W(\Delta_1, \Delta_2)^2$.Table 1 presents the values of $\pi_\beta(\Delta_1, \Delta_2)$ over $\beta \in [0, 1]$ for different values of $W(\Delta_1, \Delta_2)$. Note that, whenever $W(\Delta_1, \Delta_2) = 0$, the alternative coincides with null and hence we get back the level of the test and as $W(\Delta_1, \Delta_2)$ increases the power also increases as expected. Clearly, this asymptotic power decreases as β increases but this loss is not significant at small positive values of β. This fact is quite intuitive as the classical Wald-test at $\beta = 0$ is asymptotically most powerful under pure model. But, as we will see in the next two subsections, we can gain much higher robustness with respect to the outliers at the cost of this small loss in asymptotic power.

Table 1: Asymptotic contiguous power of the proposed Wald-type test at 95% level for testing equality of two normal means as in Example 2.1 with known common $\sigma^2 = 1$.

	β						
W_{Δ^1,Δ^2}	**0**	**0.1**	**0.3**	**0.5**	**0.7**	**0.9**	**1**
0	0.050	0.050	0.050	0.050	0.050	0.050	0.050
1	0.170	0.169	0.160	0.150	0.140	0.131	0.127
2	0.516	0.511	0.484	0.449	0.413	0.380	0.364
3	0.851	0.847	0.821	0.784	0.742	0.698	0.677
5	0.999	0.999	0.998	0.996	0.992	0.985	0.981

2.2 Influence function of the wald-type test statistics

The robustness of any two sample test is relatively complicated compared to the one sample case because, here, one may have contamination in either of the two sample or even in both the samples. Let us first derive the Hampel's influence function (IF) of the two sample Wald-type test statistics to study the robustness of the proposed test. Consider the set-up of previous subsection and denote $G_1 = F_{\boldsymbol{\theta}_1}$ and $G_2 = F_{\boldsymbol{\theta}_2}$. Then, ignoring the multiplier $\frac{nm}{n+m}$, we can define the statistical functional corresponding to the proposed Wald-type test statistics $T_{m,n}^{(\beta)}$ as

$$T_\beta(G_1, G_2) = \left(\boldsymbol{U}_\beta(G_1) - \boldsymbol{U}_\beta(G_2)\right)^T \boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\theta}_0)\left(\boldsymbol{U}_\beta(G_1) - \boldsymbol{U}_\beta(G_2)\right),$$

where $\boldsymbol{U}_\beta$ is the MDPDE functional defined in Section 1.1.

Now consider the contaminated distributions $G_{1,\varepsilon} = (1-\varepsilon)G_1 + \varepsilon\wedge_x$ and $G_{2,\varepsilon} = (1-\varepsilon)G_2 + \varepsilon\wedge_y$ where ε is the contaminated proportion and x, y are the point of contamination in the two samples respectively. Then the Hampel's first-order influence function of our test functional, when the contamination is only in the first sample, is given by

$$IF^{(1)}(x; T_\beta, G_1, G_2) = \frac{\partial}{\partial\varepsilon} T_\beta(G_{1,\varepsilon}, G_2)\Big|_{\varepsilon=0} = 2(\boldsymbol{U}_\beta(G_1) - \boldsymbol{U}_\beta(G_2))^T \boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\theta}_0) \mathcal{IF}(x; \boldsymbol{U}_\beta, G_1).$$

Similarly, if there is contamination only in the second sample, then the corresponding IF is given by

$$IF^{(2)}(y;T_\beta,G_1,G_2) = \frac{\partial}{\partial\varepsilon}T_\beta(G_1,G_{2,\varepsilon})\Big|_{\varepsilon=0} = -2(\boldsymbol{U}_\beta(G_1)-\boldsymbol{U}_\beta(G_2))^T\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\theta}_0)\mathcal{IF}(y;\boldsymbol{U}_\beta,G_1).$$

Finally, if we assume that the contamination is in both the samples, Hampel's IF turns out to be

$$IF(x,y;T_\beta,G_1,G_2) = \frac{\partial}{\partial\varepsilon}T_\beta(G_{1,\varepsilon},G_{2,\varepsilon})\Big|_{\varepsilon=0} = 2(\boldsymbol{U}_\beta(G_1)-\boldsymbol{U}_\beta(G_2))^T\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{D}_\beta(x,y),$$

where $\boldsymbol{D}_\beta(x,y) = \left[\mathcal{IF}(x;\boldsymbol{U}_\beta,G_1)-\mathcal{IF}(y;\boldsymbol{U}_\beta,G_2)\right]$. Now, in particular, if we assume the null hypothesis to be true with $G_1 = G_2 = F_{\boldsymbol{\theta}_1}$, then $\boldsymbol{U}_\beta(G_1) = \boldsymbol{U}_\beta(G_2) = \boldsymbol{\theta}_1$. Therefore, all the above three types of influence function will be zero at the null hypothesis in eq. (9), which implies that the Wald-type tests are not robust for all $\beta \geq 0$. This is clearly not informative about the robustness of the tests as we all know the non-robust nature of $T^{(0)}_{m,n}$ (which is the classical Wald test statistic $W_{m,n}$).

Therefore, we need to consider the second order influence function for this case of two sample problem. When there is contamination only in the first sample, the corresponding second order IF is given by

$$\begin{aligned} IF_2^{(1)}(x;T_\beta,G_1,G_2) &= \frac{\partial^2}{\partial^2\varepsilon}T_\beta(G_{1,\varepsilon},G_2)|_{\varepsilon=0} \\ &= 2(\boldsymbol{U}_\beta(G_1)-\boldsymbol{U}_\beta(G_2))^T\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\theta}_0)\mathcal{IF}_2(x;\boldsymbol{U}_\beta,G_1) \\ &\quad + 2\mathcal{IF}(x;\boldsymbol{U}_\beta,G_1)^T\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\theta}_0)\mathcal{IF}(x;\boldsymbol{U}_\beta,G_1). \end{aligned}$$

For the particular case of null distribution $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, it simplifies to

$$IF_2^{(1)}(x;T_\beta,F_{\boldsymbol{\theta}_1},F_{\boldsymbol{\theta}_1}) = 2\mathcal{IF}(x;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_1})^T\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\theta}_0)\mathcal{IF}(x;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_1}).$$

Similarly, if the contamination is in the second sample only, then the second order IF simplifies to

$$IF_2^{(2)}(y;T_\beta,F_{\boldsymbol{\theta}_1},F_{\boldsymbol{\theta}_1}) = 2\mathcal{IF}(y;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_1})^T\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\theta}_0)\mathcal{IF}(y;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_1}).$$

Note that these two IFs are bounded with respect to the contamination points x or y if and only if the IF of the corresponding MDPDE used is bounded; but it is the case for all $\beta > 0$ under most common parametric models. Hence for any $\beta > 0$, the proposed test gives robust inference with respect to contamination in any one of the samples. However, at $\beta = 0$ the MDPDE becomes the non-robust MLE having unbounded influence function and so using that estimator makes the classical Wald test statistic to be highly non-robust also.

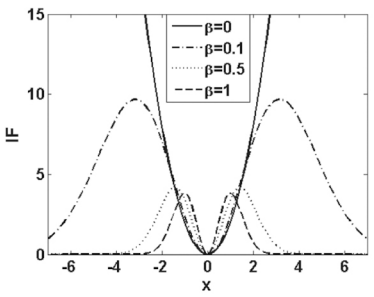
Finally for the case of contamination in both samples, the corresponding second order IF is given by

$$\begin{aligned} IF_2(x,y;T_\beta,G_1,G_2) &= \frac{\partial^2}{\partial^2\varepsilon}T_\beta(G_{1,\varepsilon},G_{2,\varepsilon})\Big|_{\varepsilon=0} \\ &= 2(\boldsymbol{U}_\beta(G_1)-\boldsymbol{U}_\beta(G_2))^T\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\theta}_0)\left[\mathcal{IF}_2(x;\boldsymbol{U}_\beta,G_1)-\mathcal{IF}_2(y;\boldsymbol{U}_\beta,G_2)\right] \\ &\quad + 2\boldsymbol{D}_\beta(x,y)^T\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{D}_\beta(x,y). \end{aligned}$$

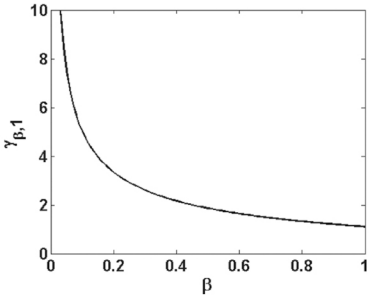
In particular, at the null hypothesis $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, we have

$$IF_2(x,y;T_\beta,F_{\boldsymbol{\theta}_1},F_{\boldsymbol{\theta}_1}) = 2\boldsymbol{D}_\beta(x,y)^T\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{D}_\beta(x,y).$$

Note that if $x = y$ then $D_\beta(x,y) = 0$ and hence this second order influence function is zero implying the robustness of the proposed test with any values of the parameter; this is expected intuitively as the same contamination in both the samples nullifies each other for testing the equivalence of the two samples as in eq. (9). However, if $x \neq y$, then the influence function of our test is bounded if and only if the difference $D_\beta(x,y)$ between the influence functions of the MDPDEs used is bounded. This happens whenever the IF of the MDPDE is bounded, i.e., at $\beta > 0$.



(a) Influence function



(b) Gross error sensitivity

Figure 1: Second order influence function of the proposed Wald-type test statistics and corresponding gross error sensitivity $\gamma_{\beta,1}$ under contamination only in first sample for testing equality of two normal means as in Example 2.1 with known common $\sigma^2 = 1$.

Example 2.2 (Continuation of Example 2.1)

Let us again consider the previous example on testing two normal means as in Example 2.1. We have seen that the proposed Wald-type tests are consistent for all $\beta \geq 0$ but their power against contiguous alternatives decreases slightly as β increases. Now let us verify the claimed robustness of these tests.

Clearly, the first order IFs of the test statistics will always be zero. For contamination only in the first sample, the second order IF of the test statistic T_β at the null hypothesis in eq. (9) has a simpler form given by

$$IF_2^{(1)}(x; T_\beta, F_{\theta_1}, F_{\theta_1}) = \frac{2}{\sigma^2}\left(1+2\beta\right)^{3/2}(x-\theta_1)^2 e^{-\frac{\beta(x-\theta_1)^2}{\sigma^2}}.$$

Figure 1a presents the plot of this second order IF for different values of $\beta \in [0, 1]$. It is evident from the figure that the second order IF is unbounded at $\beta = 0$ implying the non-robustness of the classical Wald test statistic; but it is bounded for all $\beta > 0$ implying the robustness of our proposals. Further, Figure 1b presents the plot of the maximum possible influence of infinitesimal contamination on the test statistics, known as the "gross error sensitivity", computed as

$$\gamma_{\beta,1} = \sup_x \left\| IF_2^{(1)}(x; T_\beta, F_{\theta_1}, F_{\theta_1}) \right\| = \frac{2}{\sigma^3\sqrt{\beta}}\left(1+\frac{\beta}{1+\beta}\right)e^{-\frac{\sqrt{\beta}}{\sigma}}.$$

It clearly shows that the robustness of our proposed test statistics increases as β increases (since $\gamma_{\beta,1}$ decreases). Thus, just like the trade-off between efficiency and robustness of MDPDE, the parameter β again controls the trade-off between asymptotic contiguous power and robustness for the proposed MDPDE based test statistics.

Similar inferences can also be drawn for contamination only in the second sample.

Next consider the case when there is contamination in both the samples. In this case, the second order IF is given by

$$IF_2(x, y; T_\beta, F_{\theta_1}, F_{\theta_1}) = \frac{2}{\sigma^2}\left(1+2\beta\right)^{3/2}\left[(x-\theta_1)e^{-\frac{\beta(x-\theta_1)^2}{2\sigma^2}} - (y-\theta_1)e^{-\frac{\beta(y-\theta_1)^2}{2\sigma^2}}\right]^2.$$

The plot of $IF_2(x, y; T_\beta, F_{\theta_1}, F_{\theta_1})$ have been presented in Figure 2, which clearly show the robust nature of our proposals at $\beta > 0$ and the non-robust nature of the classical Wald test (at $\beta = 0$) unless $x = y$. By looking at the maximum possible influence in this case, we can again see that, even under contamination in both the samples, the robustness of our proposed Wald-type test statistics increases as β increases.

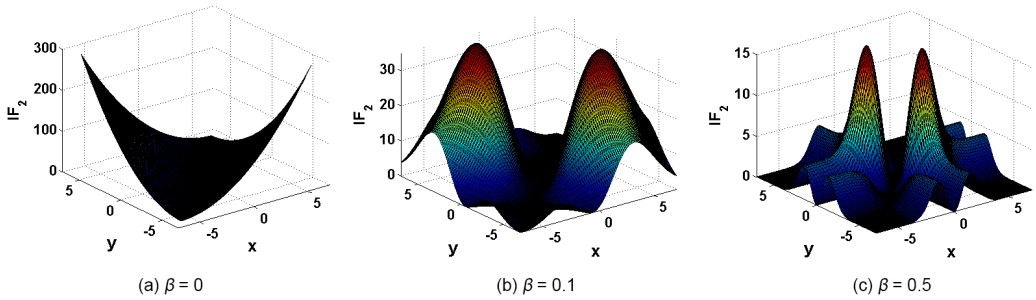


Figure 2: Second order influence function of the proposed Wald-type test statistics under contamination in both the samples for testing equality of two normal means as in Example 2.1 with known common $\sigma^2 = 1$.

2.3 Power and level influence functions

The robustness of a test statistic, although necessary, may not be sufficient in all the cases since the performance of any test is finally measured through its level and power. In this section, we consider the effect of contamination on the asymptotic power and level of the proposed Wald-type tests. Due to consistency, the asymptotic power against any fixed alternative will be one. So, we again consider the contiguous alternatives $H_{1,m,n}$ given by eq. (12) along with contamination over these alternatives. Following Hampel et al. [9], the effect of contaminations should tend to zero, as the alternatives tend to the null (i.e., $\boldsymbol{\theta}_{1,n} \to \boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_{2,m} \to \boldsymbol{\theta}_0$ as $m, n \to \infty$) at the same rate to avoid confusion between the neighborhoods of the two hypotheses (also see [7, 10–13] for

some one sample applications). Further, in case of the present two sample problem, the contamination can be in any one sample or in both the samples. When the contamination is only in the first sample, we consider the corresponding contamination distribution for the first population as

$$F^L_{1,n,\varepsilon,x} = \left(1 - \frac{\varepsilon}{\sqrt{n}}\right) F_{\boldsymbol{\theta}_0} + \frac{\varepsilon}{\sqrt{n}} \wedge_x \quad F^P_{1,n,\varepsilon,x} = \left(1 - \frac{\varepsilon}{\sqrt{n}}\right) F_{\boldsymbol{\theta}_{1,n}} + \frac{\varepsilon}{\sqrt{n}} \wedge_x,$$

for the level and power calculations respectively along with the usual uncontaminated distributions for the second population. Then the corresponding level influence function (LIF) and the power influence function (PIF) at the null $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}_0$ are given by

$$LIF^{(1)}(x; T_\beta, F_{\boldsymbol{\theta}_0}) = \lim_{m,n\to\infty} \frac{\partial}{\partial\varepsilon} P_{(F^L_{1,n,\varepsilon,x}, F_{\boldsymbol{\theta}_0})}(T^{(\beta)}_{m,n} > \chi^2_{p,\alpha})|_{\varepsilon=0},$$

$$PIF^{(1)}(x; T_\beta, F_{\boldsymbol{\theta}_0}) = \lim_{m,n\to\infty} \frac{\partial}{\partial\varepsilon} P_{(F^P_{1,n,\varepsilon,x}, F_{\boldsymbol{\theta}_{2,m}})}(T^{(\beta)}_{m,n} > \chi^2_{p,\alpha})|_{\varepsilon=0}.$$

Similarly, when contamination is assumed to be only in the second sample, then we take the uncontaminated distributions for the first population and the contaminated distribution for the second population as

$$F^L_{2,m,\varepsilon,y} = \left(1 - \frac{\varepsilon}{\sqrt{m}}\right) F_{\boldsymbol{\theta}_0} + \frac{\varepsilon}{\sqrt{m}} \wedge_y \quad F^P_{2,m,\varepsilon,y} = \left(1 - \frac{\varepsilon}{\sqrt{m}}\right) F_{\boldsymbol{\theta}_{2,m}} + \frac{\varepsilon}{\sqrt{m}} \wedge_y,$$

for the level and power calculations respectively. Corresponding LIF and PIF at the null $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}_0$ are given by

$$LIF^{(2)}(y; T_\beta, F_{\boldsymbol{\theta}_0}) = \lim_{m,n\to\infty} \frac{\partial}{\partial\varepsilon} P_{(F_{\boldsymbol{\theta}_0}, F^L_{2,m,\varepsilon,y})}(T^{(\beta)}_{m,n} > \chi^2_{p,\alpha})|_{\varepsilon=0},$$

$$PIF^{(2)}(y; T_\beta, F_{\boldsymbol{\theta}_0}) = \lim_{m,n\to\infty} \frac{\partial}{\partial\varepsilon} P_{(F_{\boldsymbol{\theta}_{1,n}}, F^P_{2,m,\varepsilon,y})}(T^{(\beta)}_{m,n} > \chi^2_{p,\alpha})|_{\varepsilon=0}.$$

Finally, while considering contamination in both the samples with above contaminated distributions, we define the corresponding LIF and PIF as

$$LIF(x, y; T_\beta, F_{\boldsymbol{\theta}_0}) = \lim_{m,n\to\infty} \frac{\partial}{\partial\varepsilon} P_{(F^L_{1,n,\varepsilon,x}, F^L_{2,m,\varepsilon,y})}(T^{(\beta)}_{m,n} > \chi^2_{p,\alpha})|_{\varepsilon=0},$$

$$PIF(x, y; T_\beta, F_{\boldsymbol{\theta}_0}) = \lim_{m,n\to\infty} \frac{\partial}{\partial\varepsilon} P_{(F^P_{1,n,\varepsilon,x}, F^P_{2,m,\varepsilon,y})}(T^{(\beta)}_{m,n} > \chi^2_{p,\alpha})|_{\varepsilon=0}.$$

First let us derive the asymptotic distribution of the proposed Wald-type test statistics $T^{(\beta)}_{m,n}$ under the contaminated distributions. Let us define $\widetilde{\boldsymbol{\Delta}}_i = \boldsymbol{\Delta}_i + \varepsilon \mathcal{IF}(x_i; \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}_0})$ for $i = 1,2$ with $x_1 = x$ and $x_2 = y$. Then we have the following theorem.

Theorem 2.6

Suppose the model density satisfies the Lehmann and Basu et al. conditions and Assumption (A) holds. Then the asymptotic distribution of $T^{(\beta)}_{m,n}$ under any contaminated contiguous alternative distributions (D_1, D_2) is $\chi^2_p(\lambda)$ where λ is the parameter of non-centrality given by $\lambda = \widetilde{\boldsymbol{W}}^T_\varepsilon \boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)^{-1} \widetilde{\boldsymbol{W}}_\varepsilon$, where

$$\begin{aligned} \widetilde{\boldsymbol{W}}_\varepsilon &= \boldsymbol{W}(\widetilde{\boldsymbol{\Delta}}_1, \boldsymbol{\Delta}_2), \quad \text{if } (D_1, D_2) = (F^P_{1,n,\varepsilon,x}, F_{\boldsymbol{\theta}_{2,m}}), \\ &= \boldsymbol{W}(\boldsymbol{\Delta}_1, \widetilde{\boldsymbol{\Delta}}_2), \quad \text{if } (D_1, D_2) = (F_{\boldsymbol{\theta}_{1,n}}, F^P_{2,m,\varepsilon,y}), \\ &= \boldsymbol{W}(\widetilde{\boldsymbol{\Delta}}_1, \widetilde{\boldsymbol{\Delta}}_2), \quad \text{if } (D_1, D_2) = (F^P_{1,n,\varepsilon,x}, F^P_{2,m,\varepsilon,y}). \end{aligned} \tag{14}$$

From the above theorem, we get the asymptotic power of the proposed Wald-type tests under the contaminated contiguous alternatives as

$$\pi_\beta(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2; \varepsilon) = P_{(D_1,D_2)}\left(T^{(\beta)}_{m,n} > \chi^2_{p,\alpha}\right) = 1 - F_{\chi^2_p\left(\widetilde{\boldsymbol{W}}^T_\varepsilon \boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)^{-1} \widetilde{\boldsymbol{W}}_\varepsilon\right)}(\chi^2_{p,\alpha}).$$

Using infinite series expansion of a non-central chi-square distribution function [14], we get

$$\pi_\beta(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2;\varepsilon) = \sum_{v=0}^{\infty} C_v\left(\widetilde{\boldsymbol{W}}_\varepsilon,\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)^{-1}\right) P\left(\chi^2_{p+2v} > \chi^2_{p,\alpha}\right),$$

where $$C_v(\boldsymbol{t},\boldsymbol{A}) = \frac{(\boldsymbol{t}^T\boldsymbol{A}\boldsymbol{t})^v}{v!\,2^v} e^{-\frac{1}{2}\boldsymbol{t}^T\boldsymbol{A}\boldsymbol{t}}.$$

In particular, substituting $\varepsilon = 0$ in the above theorem, we get back Theorem Theorem 2.5 on the asymptotic contiguous power of our tests and hence expression eq. (13) can be written as

$$\pi_\beta(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2) = \pi_\beta(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2;\boldsymbol{0}_p) = \sum_{v=0}^{\infty} C_v\left(\boldsymbol{W}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2),\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)^{-1}\right) P\left(\chi^2_{p+2v} > \chi^2_{p,\alpha}\right).$$

Further, substituting $\boldsymbol{\Delta}_1 = \boldsymbol{\Delta}_2 = \boldsymbol{0}_p$, we get the asymptotic level of our Wald-type tests under the contamination as $\alpha_\varepsilon = \pi_\beta(\boldsymbol{0},\boldsymbol{0};\varepsilon)$.

Now we can define the power influence functions of our proposed tests which is nothing but $\frac{\partial}{\partial\varepsilon}\pi_\beta(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2;\varepsilon)\big|_{\varepsilon=0}$ under standard regularity conditions. Using the infinite series expression of a non-central chi-square distribution function, we can derive an explicit form of the PIFs as presented in the following theorem.

Theorem 2.7

Suppose the model density satisfies the Lehmann and Basu et al. conditions, and Assumption (A) holds. Then the power influence functions of our proposed Wald-type tests are given by

$$\begin{aligned}
PIF^{(1)}(x;T_\beta,F_{\boldsymbol{\theta}_0}) &= \sqrt{\omega}K_p^*\left(\delta_\beta\right)\boldsymbol{W}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)^T\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)^{-1}\mathcal{IF}(x;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_0}),\\
PIF^{(2)}(y;T_\beta,F_{\boldsymbol{\theta}_0}) &= \sqrt{1-\omega}K_p^*\left(\delta_\beta\right)\boldsymbol{W}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)^T\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)^{-1}\mathcal{IF}(y;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_0}),\\
PIF(x,y;T_\beta,F_{\boldsymbol{\theta}_0}) &= K_p^*\left(\delta_\beta\right)\boldsymbol{W}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)^T\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)^{-1}\boldsymbol{W}\left(\mathcal{IF}(x;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_0}),\mathcal{IF}(x;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_0})\right),
\end{aligned}$$

where δ_β and $\boldsymbol{W}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)$ are as defined in Theorem Theorem 2.5 and

$$K_p^*(s) = e^{-\frac{s}{2}}\sum_{v=0}^{\infty}\frac{s^{v-1}}{v!\,2^v}(2v-s)\,P\left(\chi^2_{p+2v} > \chi^2_{p,\alpha}\right).$$

Note that the PIFs are also a function of the influence function of the MDPDE used and hence they are bounded whenever $\beta > 0$. Thus the proposed tests will be robust for all $\beta > 0$. However, at $\beta = 0$, these PIFs will be unbounded (unless there is contamination at the same points $x = y$ in both the samples) which proves the non-robust nature of the classical Wald test.

Note that, although there is no direct relationship between the IF of test statistics with the corresponding PIF in general, in this present case they are seen to be related indirectly via the IF of the MDPDE. So, using a robust MDPDE with $\beta > 0$ in the proposed Wald-type tests will make both the test statistics and its asymptotic power robust under infinitesimal contamination.

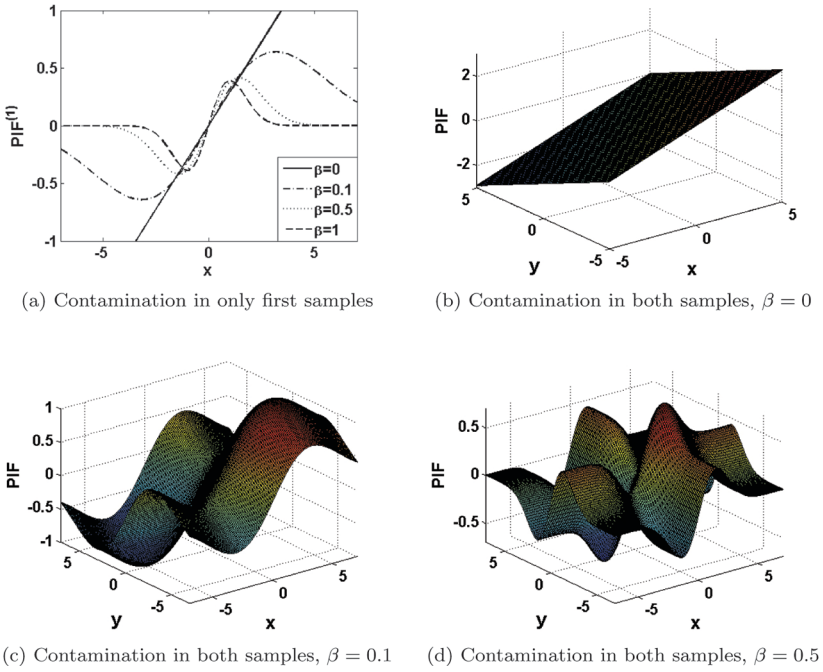
Finally, we can find the level influence function of the proposed Wald-type tests either starting from α_ε and following the same steps as in the case of PIFs or just by substituting $\boldsymbol{\Delta}_1 = \boldsymbol{\Delta}_2 = \boldsymbol{0}$ in the expression of the PIFs given in Theorem 2.7. In either case, since $\boldsymbol{W}(\boldsymbol{0},\boldsymbol{0}) = \boldsymbol{0}$, it turns out that

$$LIF^{(1)}(x;T_\beta,F_{\boldsymbol{\theta}_0}) = 0,\quad LIF^{(2)}(y;T_\beta,F_{\boldsymbol{\theta}_0}) = 0,\quad LIF(x,y;T_\beta,F_{\boldsymbol{\theta}_0}) = 0,$$

provided the corresponding IF of $\boldsymbol{U}_\beta$ is bounded, which is true at $\beta > 0$. Hence the asymptotic level of our Wald-type tests is always stable with respect infinitesimal contamination. This fact was also expected as we are using the asymptotic critical values for testing.

Example 2.3 (Continuation of Examples 2.1 and 2.1)

Let us again consider the problem of testing for normal means as in Examples 2.1 and 2.6. As seen above, the level influence function is always zero implying the level robustness of our proposed Wald-type test for all $\beta > 0$. Next, to study the power robustness, we compute the functions $PIF^{(1)}(x;T_\beta,F_{\theta_0})$ and $PIF(x,y;T_\beta,F_{\theta_0})$ numerically for different values of β with $\theta_0 = 0$ and plot them over the contamination points x and y in Figure 3. $PIF^{(2)}(y;T_\beta,F_{\theta_0})$ has the same nature as $PIF^{(1)}(x;T_\beta,F_{\theta_0})$. The figures clearly show the robustness of the proposed Wald-type tests with $\beta > 0$, where the robustness increases (i.e., maximum possible PIF decreases) as β increases. Further, all the PIFs at $\beta = 0$ are unbounded implying the non-robust nature of the classical Wald test.



(a) Contamination in only first samples (b) Contamination in both samples, $\beta = 0$

(c) Contamination in both samples, $\beta = 0.1$ (d) Contamination in both samples, $\beta = 0.5$

Figure 3: Power influence functions of the proposed Wald-type test statistics at 95% level for testing equality of two normal means as in Example Example 2.2 (Continuation of Example 2.1) with known common $\sigma^2 = 1$, $W(\Delta_1, \Delta_2) = 2$ and $\omega = 0.5$ ($n = m$).

3 General composite hypotheses with two samples

In the previous section, we have considered the simplest two sample problem which tests for equality of all the model parameters. However, in practice, we need to test many different complicated hypotheses which cannot be solved just by considering the Wald-type test statistic $T^{(\beta)}_{m,n}$ defined in the previous section. For example, in many real life problems, we are only interested in a proper subset of the parameters ignoring the rest as nuisance parameters; example includes popular mean test taking variance parameter unknown and nuisance. Further, in case of testing for multiplicative heteroscedasticity of two samples, we have to test if the ratio of variance parameters equals a pre-specified limit with means being unknown and nuisance. Neither of them belongs to the problem considered in the previous section.

In this section, we will consider a general class of hypotheses involving two independent samples, which would include most of the above real life testing problems. Suppose $\boldsymbol{\psi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ denote a general function from $\mathbb{R}^p \times \mathbb{R}^p$ to $\mathbb{R}^r$. Then, considering the set-up of the previous section, we want to develop a family of robust tests for the general class of hypothesis given by

$$H_0 : \boldsymbol{\psi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{0}_r \quad \text{against} \quad H_1 : \boldsymbol{\psi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \neq \mathbf{0}_r. \tag{15}$$

In particular, the problem of testing normal mean with unknown variance can be seen as a particular case of the above general set-up with $\boldsymbol{\psi}((\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2)) = \mu_1 - \mu_2$. Further, to test for multiplicative heteroscedasticity, we can take $\boldsymbol{\psi}((\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2)) = \frac{\sigma_1^2}{\sigma_2^2} - C_0$ for some known constant C_0 and apply the above general set-up. It is interesting to note that, this general class of hypotheses in eq. (15) also contains the simple hypothesis in eq. (9) as its special case with $\boldsymbol{\psi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$.

Now, to define a robust Wald-type test statistics for this general set-up, we again consider the MDPDEs of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ with tuning parameter β as given by ${}^{(1)}\hat{\boldsymbol{\theta}}_\beta$ and ${}^{(2)}\hat{\boldsymbol{\theta}}_\beta$ based on the individual samples separately. Note that, whenever H_0 is true, we should have $\boldsymbol{\psi}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta) \approx \mathbf{0}_r$ in large sample and so its observed value provide the indication of any departure from the null hypothesis. Using its asymptotic variance-covariance matrix as a normalizing factor, we define the corresponding Wald-type test statistic as

$$\widetilde{T^{(\beta)}_{m,n}} = \frac{nm}{n+m}\, \boldsymbol{\psi}\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta\right)^T \widetilde{\boldsymbol{\Sigma}}_\beta({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta)^{-1} \boldsymbol{\psi}\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta\right), \tag{16}$$

where $\widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \omega\boldsymbol{\Psi}_1(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_1)\boldsymbol{\Psi}_1(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) + (1-\omega)\boldsymbol{\Psi}_2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_2)\boldsymbol{\Psi}_2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ with

$$\boldsymbol{\Psi}_i(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \frac{\partial}{\partial\boldsymbol{\theta}_i}\boldsymbol{\psi}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T, \quad i=1,2.$$

Note that, at $\beta = 0$, the Wald-type test statistics $\widetilde{T^{(0)}_{m,n}}$ is again nothing but the classical Wald test statistics for the general hypothesis eq. (15) and hence our proposal is indeed a generalization of the classical Wald test.

Interestingly, although the general hypothesis contains the hypothesis eq. (9) as its special case, the Wald-type test statistics $\widetilde{T^{(\beta)}_{m,n}}$ with $\boldsymbol{\psi}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$ is not the same as the Wald-type test statistics $T^{(\beta)}_{m,n}$ considered in the previous section. However, whenever $\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta})$ is linear in the parameters, these two Wald-type test statistics coincide asymptotically with probability tending to one. In this section, we present the properties of the statistics $\widetilde{T^{(\beta)}_{m,n}}$ with general $\boldsymbol{\psi}$-function satisfying the following assumption.

Assumption (B):

– $\boldsymbol{\Psi}_i(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$, $i=1,2$, exist, have rank r and are continuous with respect to its arguments.

3.1 Asymptotic properties

We again start with the asymptotic null distribution of the proposed Wald-type test statistics $\widetilde{T^{(\beta)}_{m,n}}$ in order to obtain the required critical values for the test.

Theorem 3.1

Suppose the model density satisfies the Lehmann and Basu et al. conditions and Assumptions (A) and (B) hold. Then, under the null hypothesis in eq. (15), $\widetilde{T^{(\beta)}_{m,n}}$ asymptotically follows a χ^2_r distribution.

Therefore, the level-α critical region for the proposed test based on $\widetilde{T^{(\beta)}_{m,n}}$ for testing eq. (15) is given by

$$\widetilde{T^{(\beta)}_{m,n}} > \chi^2_{r,\alpha}.$$

Next, in order to consider an approximation to the asymptotic power for this general test based on $\widetilde{T^{(\beta)}_{m,n}}$, we are going to use the following function

$$\widetilde{l^*}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \boldsymbol{\psi}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T\widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^{-1}\boldsymbol{\psi}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2).$$

Theorem 3.2

Suppose the model density satisfies the Lehmann and Basu et al. conditions and Assumptions (A)-(B) hold. Then, whenever $\boldsymbol{\psi}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \neq \mathbf{0}_r$, we have

$$\sqrt{\frac{mn}{m+n}}\left[\widetilde{l^*}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta) - \widetilde{l^*}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\right] \underset{m,n\to\infty}{\mathscr{L}} N\left(0, 4\widetilde{l^*}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\right), \text{ as } m,n\to\infty.$$

Note that, from the above theorem, we can easily obtain an approximation to the power function of the proposed level-α Wald-type tests based on $\widetilde{T^{(\beta)}_{m,n}}$ as

$$\widetilde{\pi_{m,n,\alpha}}^{(\beta)}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = P\left(\widetilde{T^{(\beta)}_{m,n}} > \chi^2_{r,\alpha}\right) = 1 - \Phi_n\left(\frac{\sqrt{\frac{n+m}{nm}}}{2\sqrt{\widetilde{l^*}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}}\left[\chi^2_{r,\alpha} - \frac{nm}{n+m}\widetilde{l^*}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\right]\right),$$

for a sequence of distributions $\Phi_n(\cdot)$ tending uniformly to the standard normal distribution $\Phi(\cdot)$, whenever $\boldsymbol{\psi}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \neq \mathbf{0}_r$.

In such cases, it can be easily checked that $\widetilde{\pi_{m,n,\alpha}}^{(\beta)}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \to 1$ as $m,n\to\infty$. This proves the consistency of our proposed tests.

Corollary 3.3

Under the assumptions of Theorem Theorem 3.2, the proposed Wald-type tests based on $\widetilde{T^{(\beta)}_{m,n}}$ are consistent.

Now, let us study the performance of the proposed general two-sample Wald-type tests under the contiguous alternative hypotheses. As discussed in the previous section, there could be different choices for the contiguous alternative hypotheses for any general null hypothesis. Here, following the similar idea as in the alternatives in eq. (12), we consider the general form of the contiguous alternatives given by

$$H_{1,n,m}: \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1,n} = \boldsymbol{\theta}_{10} + n^{-\frac{1}{2}}\boldsymbol{\Delta}_1, \;\; \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{2,m} = \boldsymbol{\theta}_{20} + m^{-\frac{1}{2}}\boldsymbol{\Delta}_2, \qquad (\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2) \in \mathbb{R}^p \times \mathbb{R}^p - \{(\mathbf{0}_p, \mathbf{0}_p)\}, \tag{17}$$

for some fixed $(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20}) \in \Theta_0 = \{(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \Theta \times \Theta : \boldsymbol{\psi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = 0\}$. The asymptotic distribution of $\widetilde{T^{(\beta)}_{m,n}}$ under these alternatives $H_{1,m,n}$ has been presented in the following theorem.

Theorem 3.4

Suppose the model density satisfies the Lehmann and Basu et al. conditions and Assumptions (A)-(B) hold. Then the asymptotic distribution of $\widetilde{T^{(\beta)}_{m,n}}$ under $H_{1,n,m}$ in eq. (17) is $\chi^2_r(\widetilde{\delta_\beta})$, where

$$\widetilde{\delta_\beta} = \boldsymbol{W}_{\boldsymbol{\psi}}(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2)^T \widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^{-1} \boldsymbol{W}_{\boldsymbol{\psi}}(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2)$$

with $\boldsymbol{W}_{\boldsymbol{\psi}}(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2) = \left[\sqrt{\omega}\boldsymbol{\Psi}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^T\boldsymbol{\Delta}_1 + \sqrt{1-\omega}\boldsymbol{\Psi}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^T\boldsymbol{\Delta}_2\right]$.

The above theorem directly helps us to obtain the asymptotic power $\widetilde{\pi}_\beta(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2)$ of our general Wald-type tests based on $\widetilde{T^{(\beta)}_{m,n}}$ under the contiguous alternatives $H_{1,n,m}$ in eq. (17) as

$$\widetilde{\pi}_\beta(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2) = 1 - F_{\chi^2_r(\widetilde{\delta_\beta})}(\chi^2_{r,\alpha}).$$

3.2 Robustness properties

Let us now study the robustness properties of the proposed general two-sample Wald-type tests based on $\widetilde{T^{(\beta)}_{m,n}}$. We first consider the influence function of the Wald-type test statistics. Define the statistical functional corresponding to $\widetilde{T^{(\beta)}_{m,n}}$ ignoring the multiplier $\frac{nm}{n+m}$ as

$$\widetilde{T_\beta}(G_1, G_2) = \boldsymbol{\psi}\left(\boldsymbol{U}_\beta(G_1), \boldsymbol{U}_\beta(G_2)\right)^T \widetilde{\boldsymbol{\Sigma}_\beta}^{-1}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\boldsymbol{\psi}\left(\boldsymbol{U}_\beta(G_1), \boldsymbol{U}_\beta(G_2)\right),$$

where $\boldsymbol{U}_\beta$ is the corresponding MDPDE functional. Then, we can derive the first and second order influence functions of the Wald-type test statistics following the derivations similar to that of Section 2.2. So, here we will skip those derivations for brevity and present only the final results in the following theorem.

Theorem 3.5

Consider the notations of Section 2.2. Under the null hypothesis in eq. (15) with $G_1 = F_{\boldsymbol{\theta}_{10}}$, $G_2 = F_{\boldsymbol{\theta}_{20}}$ and $\boldsymbol{\psi}(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20}) = 0$, the first and second order influence functions of our general two-sample Wald-type test statistics are given as follows:

For contamination only in the i-th sample (i = 1,2) at the point x_i ($x_1 = x, x_2 = y$)

$$\begin{aligned} IF^{(i)}(x_i; \widetilde{T_\beta}, F_{\boldsymbol{\theta}_{10}}, F_{\boldsymbol{\theta}_{20}}) &= 0, \\ IF_2^{(i)}(x_i; \widetilde{T_\beta}, F_{\boldsymbol{\theta}_{i0}}, F_{\boldsymbol{\theta}_{20}}) &= 2\mathcal{JF}(x_i; \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}_{10}})^T \boldsymbol{\Psi}_i(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})^T \widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})^{-1} \\ &\quad \boldsymbol{\Psi}_i(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})\mathcal{JF}(x_i; \boldsymbol{U}_\beta, \boldsymbol{\theta}_{i0}). \end{aligned}$$

For contamination in both the samples

$$\begin{aligned} IF(x, y; \widetilde{T_\beta}, F_{\boldsymbol{\theta}_{10}}, F_{\boldsymbol{\theta}_{20}}) &= 0 \\ IF_2(x, y; \widetilde{T_\beta}, F_{\boldsymbol{\theta}_{10}}, F_{\boldsymbol{\theta}_{20}}) &= 2\boldsymbol{Q}_\beta(x, y)^T \widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})^{-1} \boldsymbol{Q}_\beta(x, y). \end{aligned}$$

with $\boldsymbol{Q}_\beta(x, y) = \boldsymbol{\Psi}_1(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})^T \mathcal{JF}(x; \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}_{10}}) + \boldsymbol{\Psi}_2(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})^T \mathcal{JF}(y; \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}_{20}})$.

Clearly, as in the previous case of simple two sample problem in Section 2.2, here also the first order IF of the test statistics are always zero and hence non-informative about their robustness. However, their second order IFs are clearly bounded whenever the IF of the corresponding MDPDE is bounded which holds for all $\beta > 0$. Thus, the proposed general two sample Wald-type tests with any $\beta > 0$ yield robust solution under contamination in either of the samples or in both. Further, in case of contamination in both the samples, if the IF of the MDPDE is not bounded (at $\beta = 0$), then also the corresponding second order IF can be bounded generating robust inference provided the term $\boldsymbol{Q}_\beta(x,y)$ is bounded. One example of such situation arises in case of the simpler problem of Section 2 under the choice $x = y$, because in that case $\boldsymbol{\Psi}_1(\boldsymbol{\theta}_{10},\boldsymbol{\theta}_{20}) = -\boldsymbol{\Psi}_2(\boldsymbol{\theta}_{10},\boldsymbol{\theta}_{20}) = \boldsymbol{I}_p$, the identity matrix of oder p, and hence $\boldsymbol{Q}_\beta(x,y)$ becomes identically zero.

Next, we consider the effect of contamination on the asymptotic power and level of the proposed general Wald-type tests based on $\widetilde{T^{(\beta)}_{m,n}}$. For this general case, we consider the contiguous alternatives $H_{1,m,n}$ as defined in eq. (17) but now with the null baseline parameter values as $\boldsymbol{\theta}_{10}$ and $\boldsymbol{\theta}_{20}$ for the two samples respectively instead of the common $\boldsymbol{\theta}_0$ and define the level and power influence functions using the corresponding contaminated distributions as in Section 2.3. Following theorem presents the asymptotic distribution of the test statistics under the contiguous and contaminated distributions, where $\widetilde{\boldsymbol{\Delta}}_i$s ($i = 1,2$) are as defined in Section 2.3.

Theorem 3.6

Suppose the model density satisfies the Lehmann and Basu et al. conditions and Assumptions (A)-(B) hold. Then, the asymptotic distribution of the general Wald-type test statistics $\widetilde{T^{(\beta)}_{m,n}}$ under any contaminated contiguous alternative distributions (D_1, D_2) is non-central chi-square with r degrees of freedom and non-centrality parameter $\widetilde{\boldsymbol{W}^*_\varepsilon}^T \widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^{-1}\widetilde{\boldsymbol{W}^*_\varepsilon}$, where

$$\begin{aligned}
\widetilde{\boldsymbol{W}^*_\varepsilon} &= \boldsymbol{W}_{\boldsymbol{\psi}}(\widetilde{\boldsymbol{\Delta}}_1,\boldsymbol{\Delta}_2), \quad \text{if } (D_1,D_2) = (F^P_{1,n,\varepsilon,x}, F_{\boldsymbol{\theta}_{2,m}}),\\
&= \boldsymbol{W}_{\boldsymbol{\psi}}(\boldsymbol{\Delta}_1,\widetilde{\boldsymbol{\Delta}}_2), \quad \text{if } (D_1,D_2) = (F_{\boldsymbol{\theta}_{1,n}}, F^P_{2,m,\varepsilon,y}),\\
&= \boldsymbol{W}_{\boldsymbol{\psi}}(\widetilde{\boldsymbol{\Delta}}_2,\widetilde{\boldsymbol{\Delta}}_2), \quad \text{if } (D_1,D_2) = (F^P_{1,n,\varepsilon,x}, F^P_{2,m,\varepsilon,y}).
\end{aligned}$$

The above theorem can be used to get the asymptotic power of the proposed general two-sample Wald-type tests under the contiguous contaminated alternatives in terms of an infinite series following Section 2.3 (arguments after 2.6). This can be also simplified by substituting $\varepsilon = 0$ or $\boldsymbol{\Delta}_1 = \boldsymbol{\Delta}_2 = \boldsymbol{0}_p$ to get asymptotic power under contiguous alternatives or the asymptotic level under contiguous contamination respectively. Further, the resulting infinite series expressions can now be used to obtain the power and level influence functions for this general case. Since the derivations are the same as that of Theorem 2.7, for brevity, we will only present the resulting expressions skipping the details in the following Theorem.

Theorem 3.7

Suppose the model density satisfies the Lehmann and Basu et al. conditions, and Assumptions (A)–(B) hold. Then we have the following results for the proposed Wald-type test functional $\widetilde{T}_\beta$ for testing the general two-sample hypothesis in eq. (15).

The power influence functions are given by

$$\begin{aligned}
PIF^{(1)}(x;\widetilde{T_\beta},F_{\boldsymbol{\theta}_{10}},F_{\boldsymbol{\theta}_{20}}) &= \sqrt{\omega}K^*_r\left(\widetilde{\delta_\beta}\right)\boldsymbol{W}_{\boldsymbol{\psi}}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)^T\widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_0)^{-1}\\
&\quad \boldsymbol{\Psi}_1(\boldsymbol{\theta}_{10},\boldsymbol{\theta}_{20})^T\mathcal{IF}(x;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_{10}}),\\
PIF^{(2)}(y;\widetilde{T_\beta},F_{\boldsymbol{\theta}_{10}},F_{\boldsymbol{\theta}_{20}}) &= \sqrt{1-\omega}K^*_r\left(\widetilde{\delta_\beta}\right)\boldsymbol{W}_{\boldsymbol{\psi}}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)^T\widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_0)^{-1}\\
&\quad \boldsymbol{\Psi}_2(\boldsymbol{\theta}_{10},\boldsymbol{\theta}_{20})^T\mathcal{IF}(y;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_{20}}),\\
PIF(x,y;\widetilde{T_\beta},F_{\boldsymbol{\theta}_{10}},F_{\boldsymbol{\theta}_{20}}) &= K^*_r\left(\widetilde{\delta_\beta}\right)\boldsymbol{W}_{\boldsymbol{\psi}}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)^T\widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_0)^{-1}\\
&\quad \boldsymbol{W}_{\boldsymbol{\psi}}\left(\mathcal{IF}(x;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_{10}}),\mathcal{IF}(y;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_{20}})\right),
\end{aligned}$$

where $\widetilde{\delta_\beta}$ and $\boldsymbol{W}_{\boldsymbol{\psi}}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)$ are as defined in Theorem 3.4 and $K^*_r(s)$ is as defined in Theorem 2.7.

Provided the IF of the MDPDE $\boldsymbol{U}_\beta$ is bounded, the level influence functions are given by

$$LIF^{(1)}(x;\widetilde{T_\beta},F_{\boldsymbol{\theta}_{10}},F_{\boldsymbol{\theta}_{20}}) = 0, \quad LIF^{(2)}(y;\widetilde{T_\beta},F_{\boldsymbol{\theta}_{10}},F_{\boldsymbol{\theta}_{20}}) = 0, \quad LIF(x,y;\widetilde{T_\beta},F_{\boldsymbol{\theta}_{10}},F_{\boldsymbol{\theta}_{20}}) = 0.$$

Note that for the general two-sample hypothesis eq. (15) also, the LIFs and the PIFs of our proposed test are bounded whenever the influence function of the MDPDE used is bounded which holds for all $\beta > 0$. Thus, our proposal with $\beta > 0$ is robust also for testing any general two-sample problem.

3.3 Special case: Testing partial homogeneity with nuisance parameters

Let us consider a simplified and possibly the most common special case of the general hypothesis in eq. (15), where we test for partial homogeneity of the two samples assuming some parameters to be nuisance. Mathematically, let us consider the partition of the parameters $\boldsymbol{\theta}_1 = \left({}^*\boldsymbol{\theta}_1^T, {}^0\boldsymbol{\theta}_1^T\right)^T$ and $\boldsymbol{\theta}_2 = \left({}^*\boldsymbol{\theta}_2^T, {}^0\boldsymbol{\theta}_2^T\right)^T$ as in the beginning of Section 2, but now we assume both, ${}^0\boldsymbol{\theta}_1$ and ${}^0\boldsymbol{\theta}_2$, to be unknown and nuisance parameters. Under these notations, we consider the hypothesis of partial homogeneity as given by

$$H_0 : {}^*\boldsymbol{\theta}_1 = {}^*\boldsymbol{\theta}_2 \quad \text{against} \quad H_1 : {}^*\boldsymbol{\theta}_1 \neq {}^*\boldsymbol{\theta}_2, \tag{18}$$

with ${}^0\boldsymbol{\theta}_1$ and ${}^0\boldsymbol{\theta}_2$ being unknown under both hypotheses. Note that, this special case contains the problem of testing normal mean with unknown variances with ${}^*\boldsymbol{\theta}_i$ being the mean and ${}^0\boldsymbol{\theta}_i$ being the variance parameter for each $i = 1,2$. In practice we can either assume ${}^0\boldsymbol{\theta}_1 = {}^0\ \boldsymbol{\theta}_2$ (e.g., equal variances) or ${}^0\boldsymbol{\theta}_1 \neq {}^0\ \boldsymbol{\theta}_2$ (e.g., unequal variances). Here, we will consider the general case assuming ${}^0\boldsymbol{\theta}_1 \neq {}^0\ \boldsymbol{\theta}_2$; other case can also be dealt similarly.

Note that the hypothesis eq. (18) is indeed a special case of the general hypothesis in eq. (15) with $\boldsymbol{\psi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = {}^*\boldsymbol{\theta}_1 - {}^*\boldsymbol{\theta}_2$. Hence, the proposed MDPDE based Wald-type test statistics for testing eq. (18) is given by

$$\widetilde{T^{(\beta)}_{m,n}} = \frac{nm}{n+m}\left({}^{(1)*}\hat{\boldsymbol{\theta}}_\beta - {}^{(2)*}\hat{\boldsymbol{\theta}}_\beta\right)^T \left[\omega\boldsymbol{\Sigma}^{11}_\beta({}^{(1)}\hat{\boldsymbol{\theta}}_\beta) + (1-\omega)\boldsymbol{\Sigma}^{11}_\beta({}^{(2)}\hat{\boldsymbol{\theta}}_\beta)\right]^{-1}\left({}^{(1)*}\hat{\boldsymbol{\theta}}_\beta - {}^{(2)*}\hat{\boldsymbol{\theta}}_\beta\right), \tag{19}$$

where ${}^{(1)*}\hat{\boldsymbol{\theta}}_\beta$ and ${}^{(2)*}\hat{\boldsymbol{\theta}}_\beta$ are the first r-components of the MDPDEs ${}^{(1)}\hat{\boldsymbol{\theta}}_\beta = ({}^{(1)*}\hat{\boldsymbol{\theta}}_\beta^T, {}^{(1)0}\hat{\boldsymbol{\theta}}_\beta^T)^T$ and ${}^{(2)}\hat{\boldsymbol{\theta}}_\beta = ({}^{(2)*}\hat{\boldsymbol{\theta}}_\beta^T, {}^{(2)0}\hat{\boldsymbol{\theta}}_\beta^T)^T$ of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ respectively and $\boldsymbol{\Sigma}^{11}_\beta(\boldsymbol{\theta})$ denotes the $r\times r$ principle minor of the asymptotic variance-covariance matrix $\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}^{11}_\beta(\boldsymbol{\theta}) & \boldsymbol{\Sigma}^{12}_\beta(\boldsymbol{\theta}) \\ \boldsymbol{\Sigma}^{12}_\beta(\boldsymbol{\theta})^T & \boldsymbol{\Sigma}^{22}_\beta(\boldsymbol{\theta}) \end{pmatrix}$. Also note that Assumption (B) always holds for the hypothesis eq. (18). Following Theorem Theorem 3.1, the asymptotic distribution of $\widetilde{T^{(\beta)}_{m,n}}$ in eq. (19) under the null hypothesis in eq. (18) is χ^2_r and the test is consistent against any fixed alternatives by Corollary Corollary 3.3. To study the asymptotic contiguous power in this case, we consider the contiguous alternatives

$$H'_{1,n,m} : {}^*\boldsymbol{\theta}_1 = {}^*\boldsymbol{\theta}_0 + n^{-\frac{1}{2}}\boldsymbol{\Delta}_1,\ \ {}^*\boldsymbol{\theta}_2 = {}^*\boldsymbol{\theta}_0 + m^{-\frac{1}{2}}\boldsymbol{\Delta}_2, \qquad (\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2) \in \mathbb{R}^r \times \mathbb{R}^r - \{(\mathbf{0}_r, \mathbf{0}_r)\}, \tag{20}$$

for some fixed ${}^*\boldsymbol{\theta}_0 \in \Theta$. Then, by Theorem 3.4, the asymptotic distribution of the Wald-type test statistics $\widetilde{T^{(\beta)}_{m,n}}$ in eq. (19) under $H'_{1,n,m}$ in eq. (20) is a non-central chi-square distribution with r degrees of freedom and non-centrality parameter ${}^*\widetilde{\delta_\beta} = \boldsymbol{W}(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2)^T \left[\omega\boldsymbol{\Sigma}^{11}_\beta({}^{(1)}\hat{\boldsymbol{\theta}}_\beta) + (1-\omega)\boldsymbol{\Sigma}^{11}_\beta({}^{(2)}\hat{\boldsymbol{\theta}}_\beta)\right]^{-1}\boldsymbol{W}(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2)$ from which the power can be calculated easily.

Next, for examining robustness properties, we define the corresponding test functional following Section 3.2 as given by

$$\widetilde{T_\beta}(G_1, G_2) = \left({}^*\boldsymbol{U}_\beta(G_1) - {}^*\boldsymbol{U}_\beta(G_2)\right)^T \left[\omega\boldsymbol{\Sigma}^{11}_\beta({}^{(1)}\hat{\boldsymbol{\theta}}_\beta) + (1-\omega)\boldsymbol{\Sigma}^{11}_\beta({}^{(2)}\hat{\boldsymbol{\theta}}_\beta)\right]^{-1}\left({}^*\boldsymbol{U}_\beta(G_1) - {}^*\boldsymbol{U}_\beta(G_2)\right),$$

where ${}^*\boldsymbol{U}_\beta$ denotes first r-components of the minimum DPD functional $\boldsymbol{U}_\beta$. Then, we can get the IF for this test statistics from Theorem Theorem 3.5. In particular, the first order influence function is identically zero for any kind of contamination and hence non-informative. And its second order influence function for contamination in i-th sample at the point x_i ($i = 1,2$) is given by

$$IF^{(i)}_2(x_i; \widetilde{T_\beta}, F_{\boldsymbol{\theta}_{i0}}, F_{\boldsymbol{\theta}_{20}}) = 2\mathcal{IF}(x_i; {}^*\boldsymbol{U}_\beta, F_{\boldsymbol{\theta}_{10}})^T \left[\omega\boldsymbol{\Sigma}^{11}_\beta({}^{(1)}\hat{\boldsymbol{\theta}}_\beta) + (1-\omega)\boldsymbol{\Sigma}^{11}_\beta({}^{(2)}\hat{\boldsymbol{\theta}}_\beta)\right]^{-1}\mathcal{IF}(x_i; {}^*\boldsymbol{U}_\beta, \boldsymbol{\theta}_{i0}).$$

and the same for contamination in both samples is given by

$$IF_2(x, y; \widetilde{T_\beta}, F_{\boldsymbol{\theta}_{10}}, F_{\boldsymbol{\theta}_{20}}) = 2\ {}^*\boldsymbol{Q}_\beta(x,y)^T \left[\omega\boldsymbol{\Sigma}^{11}_\beta({}^{(1)}\hat{\boldsymbol{\theta}}_\beta) + (1-\omega)\boldsymbol{\Sigma}^{11}_\beta({}^{(2)}\hat{\boldsymbol{\theta}}_\beta)\right]^{-1}{}^*\boldsymbol{Q}_\beta(x,y),$$

with ${}^*\boldsymbol{Q}_\beta(x,y) = \mathcal{IF}(x; {}^*\boldsymbol{U}_\beta, F_{\boldsymbol{\theta}_{10}}) - \mathcal{IF}(y; {}^*\boldsymbol{U}_\beta, F_{\boldsymbol{\theta}_{20}})$. Similarly, following Theorem 3.7, the level influence functions are always zero and the power influence functions under contiguous contamination in each sample

separately or in both the samples are respectively given by

$$\begin{aligned}
PIF^{(1)}(x;\widetilde{T_\beta},F_{\boldsymbol{\theta}_{10}},F_{\boldsymbol{\theta}_{20}}) &= \sqrt{\omega}K_r^*\left({}^*\widetilde{\delta_\beta}\right)\boldsymbol{W}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)^T\left[\omega\boldsymbol{\Sigma}_\beta^{11}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta)+(1-\omega)\boldsymbol{\Sigma}_\beta^{11}({}^{(2)}\hat{\boldsymbol{\theta}}_\beta)\right]^{-1}\\
&\quad \mathcal{IF}(x;{}^*\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_0}),\\
PIF^{(2)}(y;\widetilde{T_\beta},F_{\boldsymbol{\theta}_{10}},F_{\boldsymbol{\theta}_{20}}) &= \sqrt{1-\omega}K_r^*\left({}^*\widetilde{\delta_\beta}\right)\boldsymbol{W}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)^T\left[\omega\boldsymbol{\Sigma}_\beta^{11}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta)+(1-\omega)\boldsymbol{\Sigma}_\beta^{11}({}^{(2)}\hat{\boldsymbol{\theta}}_\beta)\right]^{-1}\\
&\quad \mathcal{IF}(y;{}^*\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_0}),\\
PIF(x,y;\widetilde{T_\beta},F_{\boldsymbol{\theta}_{10}},F_{\boldsymbol{\theta}_{20}}) &= K_r^*\left({}^*\widetilde{\delta_\beta}\right)\boldsymbol{W}(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)^T\left[\omega\boldsymbol{\Sigma}_\beta^{11}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta)+(1-\omega)\boldsymbol{\Sigma}_\beta^{11}({}^{(2)}\hat{\boldsymbol{\theta}}_\beta)\right]^{-1}\\
&\quad \times\boldsymbol{W}\left(\mathcal{IF}(x;{}^*\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_0}),\mathcal{IF}(x;{}^*\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_0})\right),
\end{aligned}$$

where ${}^*\widetilde{\delta_\beta}$ and ${}^*\boldsymbol{U}_\beta$ are as defined previously in this subsection. The nature of these PIFs are exactly the same as in the previous cases and indicates robustness of our proposals with $\beta>0$.

Example 3.1 (Testing equality of two Normal means with unknown and unequal variances)
We again consider the example of comparing two normal means (say μ_1 and μ_2), but now with unknown and unequal variances (say σ_1^2 and σ_2^2) for the two populations. Hence the model family is $\mathcal{F}=\{N(\mu,\sigma^2):\boldsymbol{\theta}=(\mu,\sigma)^T\in\mathbb{R}\times[0,\infty)\}$ and we want to test for the hypothesis

$$H_0:\mu_1=\mu_2 \quad \text{against} \quad H_1:\mu_1\neq\mu_2, \tag{21}$$

with σ_1^2 and σ_2^2 being unknown under both hypotheses. Let us denote the MDPDEs based on the i-th sample ($i=1,2$) as ${}^{(i)}\hat{\boldsymbol{\theta}}_\beta={}^{(i)}\hat{\mu}_\beta,{}^{(i)}\hat{\sigma}_\beta)^T$ and its asymptotic variance matrix $\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta})$ is given by

$$\boldsymbol{\Sigma}_\beta(\mu,\sigma)=\begin{pmatrix}\left(1+\frac{\beta^2}{1+2\beta}\right)^{3/2}\sigma^2 & 0\\ 0 & \frac{(1+\beta)^2}{(2+\beta^2)^2}\left\{\frac{2\zeta_\beta}{(1+2\beta)^{5/2}}-\beta^2\right\}\end{pmatrix},$$

with $\zeta_\beta=1+3\beta+5\beta^2+7\beta^3+6\beta^4+2\beta^5$. Then, noting that the hypothesis eq. (21) is of the form eq. (18), our proposed generalized Wald-type test statistics eq. (19) simplifies to

$$\widetilde{T_{m,n}^{(\beta)}}=\frac{mn}{m+n}\left(1+\frac{\beta^2}{1+2\beta}\right)^{-3/2}\frac{\left({}^{(1)}\hat{\mu}_\beta-{}^{(2)}\hat{\mu}_\beta\right)^2}{\left(\omega^{(1)}\widehat{\sigma}_\beta^2+(1-\omega)^{(2)}\widehat{\sigma}_\beta^2\right)}, \tag{22}$$

whose null asymptotic distribution is χ_1^2 from Theorem Theorem 3.1. In the particular case of $\beta=0$, we have

$$\widetilde{T_{m,n}^{(0)}}=\frac{mn}{m+n}\frac{\left({}^{(1)}\hat{\mu}_0-{}^{(2)}\hat{\mu}_0\right)^2}{\left(\omega^{(1)}\widehat{\sigma}_0^2+(1-\omega)^{(2)}\widehat{\sigma}_0^2\right)}=\frac{mn}{m+n}\frac{\left(\bar{X}-\bar{Y}\right)^2}{\left(\omega s_X^2+(1-\omega)s_Y^2\right)},$$

where $\bar{X}$ and $\bar{Y}$ are the sample means and s_X^2 and s_Y^2 are the sample variances of $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_m$ respectively, and this is nothing but the classical MLE based Wald test statistic.

We can now study the asymptotic and robustness properties of these proposed Wald-type tests following the theoretical results derived in this section. However, due to the asymptotic independence of the MDPDEs of μ and σ under normal model, all the properties of the Wald-type test statistics in eq. (22) turn out to be similar in nature to those of the proposed Wald-type test with known σ as discussed in Examples 2.1, 2.6 and Example 2.2 (Continuation of Example 2.1) with the common variance σ^2 there replaced by $[\omega\sigma_1^2+(1-\omega)\sigma_2^2]$ in the present case. This fact can also be observed intuitively by noting that the Wald-type test statistics in eq. (22) have a similar form as the corresponding Wald-type test statistics for known common σ^2 case (in Example 2.1) with the known value there being replaced by $\left[\omega^{(1)}\widehat{\sigma}_\beta^2+(1-\omega)^{(2)}\widehat{\sigma}_\beta^2\right]$. So, we will skip these details for the present general case for brevity. However, examining them, one can easily verify that, in this case of unknown and unequal variances also, the asymptotic contiguous power of the proposed Wald-type test decreases only slightly as β increases (exactly in the same rate as in Table 1) but the robustness increases significantly having bounded (second order) influence functions of the Wald-type test statistics and bounded power and level influence functions for all $\beta>0$.

4 The cases of one-sided alternatives

As we have mentioned in the introduction (Section 1), majority of common practical applications of the two-sample problems are in comparing the treatment and control groups in any experimental or clinical trials or any observational studies among two such groups of population. However, in most of such cases, researchers want to test weather there is any improvement in the treatment group over the control groups due to the treatment effects. For example, one might be interested to test if the success rate of cure (modeled by binomial probability model) is reduced, or if the number of attacks of a disease (modeled by Poisson model) decreases in the treat group, or some continuous biomarkers like blood pressure etc. (modeled by normal model) changes in the targeted direction from control to treatment group. All of them lead to the one-sided alternatives in contrast to the omnibus two-sided alternatives considered so far in this paper. Although the case of general one-sided alternatives with vector parameters are much difficult to define and dealt with and hence need more targeted future research, our proposal of robust Wald-type tests in this paper can be easily extended for comparing any scalar parameters with one-sided alternatives. Noting that all the above motivating practical scenarios indeed deal with scalar parameter comparison, in this section we extend our proposal to these particular one sample problems.

In general, we consider the class of one-sided version of eq. (15) with $r = 1$. So, $\psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is a real function of the parameters and we develop the robust test for the one-sided hypothesis given by

$$H_0 : \psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = 0 \quad \text{against} \quad H_1 : \psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) > 0. \tag{23}$$

Note that the one sided version of the simple two-sided hypothesis in eq. (9) with scalar parameters ($p = 1$), that contains the motivating examples for Poisson and binomial models and normal model with known variances, belong to this general class eq. (23). Also, this general class of hypotheses contains many more useful cases like testing for increase (or decrease) in normal means with unknown variances.

For testing the one sided hypothesis eq. (23), we define the corresponding robust Wald-type test statistics by taking a signed square-root of our two-sided Wald-type test statistics $\widetilde{T^{(\beta)}_{m,n}}$ in eq. (16)

$$\widetilde{T^{(\beta)P}_{m,n}} = sgn\left(\psi\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta\right)\right)\sqrt{\widetilde{T^{(\beta)}_{m,n}}} = \sqrt{\frac{nm}{n+m}}\frac{\psi\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta\right)}{\sqrt{\widetilde{\Sigma}_\beta({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta)}}, \tag{24}$$

where $sgn(\cdot)$ denotes the sign function and note that $\widetilde{\Sigma}_\beta(\boldsymbol{\theta}_\beta, \boldsymbol{\theta}_\beta)$ is a scalar for $r = 1$. Then, we have the following null asymptotic distribution.

Theorem 4.1

Under the assumptions of Theorem Theorem 3.1, the asymptotic null distribution of the one-sided test statistics $\widetilde{T^{(\beta)P}_{m,n}}$ for testing eq. (23) is standard normal.

Following the above theorem, the level-α critical region for testing the one-sided hypothesis in eq. (23) is given by $\left\{\widetilde{T^{(\beta)P}_{m,n}} > z_{1-\alpha}\right\}$, where $z_{1-\alpha}$ denotes the $(1-\alpha)$-th quantile of the standard normal distribution.

Further, as in the case of two-side alternatives, we can also derive an power approximation of these proposed Wald-type tests at any fixed alternative $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ satisfying $\psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) > 0$ as follows:

$$\begin{aligned}
\widetilde{\pi_{m,n,\alpha}}^{(\beta)P}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= P\left(\widetilde{T^{(\beta)P}_{m,n}} > z_{1-\alpha}\right) \\
&= P\left(\sqrt{\frac{nm}{n+m}}\frac{\left[\psi\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta\right) - \psi\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right)\right]}{\sqrt{\widetilde{\Sigma}_\beta({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta)}} > z_{1-\alpha} - \sqrt{\frac{nm}{n+m}}\frac{\psi\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right)}{\sqrt{\widetilde{\Sigma}_\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}}\right) \\
&= 1 - \Phi_n\left(z_{1-\alpha} - \sqrt{\frac{nm}{n+m}}\frac{\psi\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right)}{\sqrt{\widetilde{\Sigma}_\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}}\right),
\end{aligned}$$

for a sequence of distributions $\Phi_n(\cdot)$ tending uniformly to the standard normal distribution $\Phi(\cdot)$, since under the alternative parameter values $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$

$$\sqrt{\frac{nm}{n+m}}\frac{\left[\psi\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta\right) - \psi\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right)\right]}{\sqrt{\widetilde{\Sigma}_\beta({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta)}} \underset{m,n\to\infty}{\mathscr{L}} N(0,1).$$

Now, since $\psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) > 0$ under the alternatives in eq. (23), we have $\widetilde{\pi_{m,n,\alpha}}^{(\beta)P}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \to 1$ as $m, n \to \infty$ and hence the proposed Wald-type tests are consistent for the one-sided alternatives also.

Next to study the contiguous power of the proposed Wald-type tests, we can consider the class of contiguous alternatives in eq. (17) but now with $(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2)$ being such that $\psi(\boldsymbol{\theta}_{1,n}, \boldsymbol{\theta}_{2,m}) > 0$ for all m,n. This can be equivalently (asymptotic) expressed in terms of the sequence of alternatives

$$H^P_{1,m,n} \; : \; \psi(\boldsymbol{\theta}_{1,n}, \boldsymbol{\theta}_{2,m}) = \sqrt{\frac{m+n}{mn}}d, \tag{25}$$

with $d = W_\psi(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2) > 0$. The following theorem then gives the asymptotic distribution of our Wald-type test statistics under the contiguous alternatives in eq. (25) and the corresponding asymptotic power.

Theorem 4.2

Under the assumptions of Theorem 3.4, the asymptotic distribution of $\widetilde{T^{(\beta)P}_{m,n}}$ in eq. (24) under the sequence of contiguous alternatives in eq. (25) is normal with mean $d/\sqrt{\widetilde{\Sigma_\beta}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}$ and variance 1. Hence, the corresponding asymptotic contiguous power of the proposed Wald-type tests is given by

$$\widetilde{\pi}^P_\beta(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2) = \widetilde{\pi}^P_\beta(d) = 1 - \Phi\left(z_{1-\alpha} - d/\sqrt{\widetilde{\Sigma_\beta}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}\right).$$

Now we can also derive the robustness properties of the proposed Wald-type tests against one-sided alternatives by defining the corresponding statistical function as

$$\widetilde{T}^P_\beta(G_1, G_2) = \psi\left(\boldsymbol{U}_\beta(G_1), \boldsymbol{U}_\beta(G_2)\right) \Big/ \sqrt{\widetilde{\Sigma_\beta}(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})}.$$

Then, under the assumptions of Theorem Theorem 3.5 with contamination in only i-th sample at the point x_i ($i = 1,2$), the first order influence function of the proposed Wald-type test statistics at the null hypothesis in eq. (23) is given by

$$IF^{(i)}(x_i; \widetilde{T}^P_\beta, F_{\boldsymbol{\theta}_{10}}, F_{\boldsymbol{\theta}_{20}}) \;\; = \;\; \boldsymbol{\Psi}_i(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})^T \mathcal{J}\mathcal{F}(x_i; \boldsymbol{U}_\beta, \boldsymbol{\theta}_{i0}) \Big/ \sqrt{\widetilde{\Sigma_\beta}(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})},$$

and the same for contamination in both the samples is given by

$$IF(x_1, x_2; \widetilde{T}^P_\beta, F_{\boldsymbol{\theta}_{10}}, F_{\boldsymbol{\theta}_{20}}) \;\; = \;\; Q_\beta(x_1, x_2) \Big/ \sqrt{\widetilde{\Sigma_\beta}(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})},$$

with $Q_\beta(\cdot,\cdot)$ being as defined in Theorem Theorem 3.5 (but is a scalar now). Note that, unlike the two-sided hypotheses, here the first order influence function of the proposed Wald-type test statistics is non-zero. Further, it is bounded whenever te IF of the corresponding MDPDE is bounded, i.e., only for $\beta > 0$ and unbounded at $\beta = 0$ implying the robustness of our proposal with $\beta > 0$.

In order to derive the corresponding level and power influence functions, we consider the same set of hypothesis as in Section 3.2 but now with the restriction $\psi(\boldsymbol{\theta}_{1,n}, \boldsymbol{\theta}_{2,m}) > 0$ for all m,n under the alternative sequence, which is ensured by assuming $W_\psi(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2) > 0$. Then, the following theorem gives the asymptotic distribution of the one-sided test statistics $\widetilde{T^{(\beta)P}_{m,n}}$ under the contiguous contaminated distributions.

Theorem 4.3

Under the assumptions of Theorem 3.2, the asymptotic distribution of $\widetilde{T^{(\beta)P}_{m,n}}$ under any contaminated contiguous alternative distributions (D_1, D_2) is normal with mean $\widetilde{\boldsymbol{W}}^*_\varepsilon / \sqrt{\widetilde{\Sigma_\beta}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}$ and variance 1, where $\widetilde{\boldsymbol{W}}^*_\varepsilon$ is as defined in Theorem 3.2 for different (D_1, D_2).

Using above theorem and following the arguments similar to those for the two-sided alternatives in Section 3.2, we can get the power influence functions for this case of one-sided alternatives also, which is presented in the next theorem.

Theorem 4.4

Under the assumptions of Theorem 3.7, the power influence functions of our proposed Wald-type test functional $\widetilde{T}^P_\beta$ for testing the one-sided hypothesis in eq. (23) are given by

$$
\begin{aligned}
PIF^{(1)}(x;\widetilde{T_\beta},F_{\boldsymbol{\theta}_{10}},F_{\boldsymbol{\theta}_{20}}) &= \frac{\sqrt{\omega}}{\sqrt{\widetilde{\Sigma_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}}\phi\left(z_{1-\alpha}-\frac{W_\psi(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)}{\sqrt{\widetilde{\Sigma_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}}\right)\boldsymbol{\Psi}_1(\boldsymbol{\theta}_{10},\boldsymbol{\theta}_{20})^T\mathcal{IF}(x;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_{10}}),\\
PIF^{(2)}(y;\widetilde{T_\beta},F_{\boldsymbol{\theta}_{10}},F_{\boldsymbol{\theta}_{20}}) &= \frac{\sqrt{1-\omega}}{\sqrt{\widetilde{\Sigma_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}}\phi\left(z_{1-\alpha}-\frac{W_\psi(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)}{\sqrt{\widetilde{\Sigma_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}}\right)\boldsymbol{\Psi}_2(\boldsymbol{\theta}_{10},\boldsymbol{\theta}_{20})^T\mathcal{IF}(y;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_{20}}),\\
PIF(x,y;\widetilde{T_\beta},F_{\boldsymbol{\theta}_{10}},F_{\boldsymbol{\theta}_{20}}) &= \frac{\sqrt{1}}{\sqrt{\widetilde{\Sigma_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}}\phi\left(z_{1-\alpha}-\frac{W_\psi(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)}{\sqrt{\widetilde{\Sigma_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}}\right)W_\psi\left(\mathcal{IF}(x;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_{10}}),\right.\\
&\qquad \left.\mathcal{IF}(y;\boldsymbol{U}_\beta,F_{\boldsymbol{\theta}_{20}})\right).
\end{aligned}
$$

Note that, the nature of these PIFs with respect to the contamination points x and y are exactly same as those in the case of two-sided alternatives except for a multiplicative constant. In particular, they are bounded whenever the influence function of the MDPDE used is bounded, i.e., at $\beta > 0$, implying robustness of our proposal.

Finally, we can get the level influence functions from the above theorem by substituting $\boldsymbol{\Delta}_1 = \boldsymbol{\Delta}_2 = \mathbf{0}$ in the expressions of PIFs. Note that, in this case of one-sided hypothesis testing, the LIFs are not identically zero, but they are bounded only for $\beta > 0$ implying again the level stability of our proposed Wald-type tests.

For illustration, we will again present the case of normal model with one-sided alternatives in the following example. Other motivating models with relevant data examples will be provided in the next section.

Example 4.1 (Comparing two Normal means against one-sided alternatives)

Let us again consider the two-sample problem under normal model with unknown and unequal variances as in Example 3.1, but now with the one-sided alternatives so that our target hypothesis is

$$H_0 : \mu_1 = \mu_2 \quad \text{against} \quad H_1 : \mu_1 > \mu_2, \tag{26}$$

with the variance parameters σ_1 and σ_2 being unknown for both hypotheses. Considering the notations of Example 3.1, our proposed test statistics $\widetilde{T^{(\beta)P}_{m,n}}$ is then given by

$$\widetilde{T^{(\beta)P}_{m,n}} = \sqrt{\frac{mn}{m+n}}\left(1+\frac{\beta^2}{1+2\beta}\right)^{-3/4}\frac{\left({}^{(1)}\hat{\mu}_\beta - {}^{(2)}\hat{\mu}_\beta\right)}{\sqrt{\omega^{(1)}\widehat{\sigma}^2_\beta + (1-\omega)^{(2)}\widehat{\sigma}^2_\beta}}, \tag{27}$$

which has standard normal asymptotic distribution under the null. Clearly this statistic also coincides with the corresponding classical Wald test statistic at $\beta = 0$. Since the test is consistent at any fixed alternatives, we consider the contiguous alternatives $H^P_{1,m,n} : \psi(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \mu_1 - \mu_2 = \sqrt{\frac{m+n}{mn}}d$ with $d \,\&\, 0$, under which the test statistics has asymptotic distribution as normal with mean $\left(1+\frac{\beta^2}{1+2\beta}\right)^{-3/4} d\,[\omega\sigma_1^2 + (1-\omega)\sigma_2^2]^{-\frac{1}{2}}$ and variance 1. Corresponding asymptotic contiguous power at different values of d and β with $\sigma_1^2 = \sigma_2^2 = 1$ and $\omega = 0.5$ ($n = m$) is presented in Table 2. Note that, as expected this power decreases only slightly as β increases (note the similarity with Table 1).

Table 2: Asymptotic contiguous power of the proposed Wald-type tests at 95% level for testing equality of two normal means against one-sided alternatives as in Example 4.1.

	β						
d	**0**	**0.1**	**0.3**	**0.5**	**0.7**	**0.9**	**1**
0	0.050	0.050	0.050	0.050	0.050	0.050	0.050
1	0.260	0.258	0.247	0.233	0.219	0.207	0.201
2	0.639	0.634	0.608	0.574	0.538	0.503	0.487
3	0.912	0.909	0.891	0.865	0.833	0.798	0.780
5	1.000	1.000	0.999	0.998	0.997	0.994	0.991

Further, the influence function of the proposed Wald-type test statistics in this case of one-sided alternatives simplifies to

$$IF_2^{(i)}(x_i; \widetilde{T}_\beta^P, F_{\boldsymbol{\theta}_{10}}, F_{\boldsymbol{\theta}_{20}}) = \left[\omega\sigma_{10}^2 + (1-\omega)\sigma_{20}^2\right]^{-\frac{1}{2}} (1+2\beta)^{3/4} (x_i - \mu_{i0}) e^{-\frac{\beta(x_i-\mu_{i0})^2}{2\sigma_{i0}^2}},$$

and

$$IF_2(x_1, x_2; \widetilde{T}_\beta^P, F_{\boldsymbol{\theta}_{10}}, F_{\boldsymbol{\theta}_{20}}) = \frac{(1+2\beta)^{3/4}}{\sqrt{\omega\sigma_{10}^2 + (1-\omega)\sigma_{20}^2}} \left[(x_1 - \mu_{10}) e^{-\frac{\beta(x_1-\mu_{10})^2}{2\sigma_{10}^2}} - (x_2 - \mu_{20}) e^{-\frac{\beta(x_2-\mu_{20})^2}{2\sigma_{20}^2}} \right].$$

Note that these influence functions are square roots of the corresponding influence functions under two-sided alternatives in Example Example 2.2 (Continuation of Example 2.1) except for a multiplicative constant. Further, by the general theory developed above, the corresponding PIFs and LIFs in this case can be shown to be also a constant multiplication of the corresponding PIFs in the two-sided case presented in Example Example 2.2 (Continuation of Example 2.1). Therefore, the boundedness nature of all these influence functions for the one-sided alternative will be similar to those presented in Figure 1a and Figure 3, i.e., bounded at $\beta > 0$ and unbounded at $\beta = 0$. These again imply the robustness of our proposal with $\beta > 0$ over the classical Wald test at $\beta = 0$.

5 Real life applications

5.1 Poisson model for clinical trial: Adverse events data

In our first example we will consider the application of the proposed Wald-type tests with Poisson model to the adverse event data in an Asthma clinical trial conducted by Kerstjens et al. [15][Table 3]. In this two phase randomized controlled trials, 912 patients having asthma and receiving inhaled glucocorticoids and LABAs had been divided into treatment and control groups of the two trials and were randomly assigned a total dose of 5 g tiotropium (treatment group) or suitable placebo (control group) once daily for 48 weeks. Then, Kerstjens et al. [15] investigated the effect of this combined treatment on patient's lung function and exacerbations.

Table 3: No of Different adverse events reported in Trial 2 of the Kerstjens et al. [15] clinical trail study.

Treatment	91	49	19	12	12	3	13	10	6	3	3	7	6	5	4	4	3	2	0
Control	109	58	20	13	10	10	6	4	5	7	5	1	2	4	4	5	2	2	1

Here we will consider the data on 19 reported adverse effect on the patients in trail 2 of this study, presented in Table 3, that can be modeled by a Poisson distribution with mean θ. Note that the first two entry for both the groups (corresponding to the events of Asthma and Decreased rate of peak expiratory flow) clearly stands out as outliers from the remaining observations. Hence, in presence of these two observations the MLE of the Poisson parameters θ_1 and θ_2 in treatment and control groups (15 and 18.47 respectively) turns out to be drastically different from the MLEs without them (8.82 and 9.65 respectively). However the robust MDPDEs with larger β remains stable (see Table 4). Clearly, the number of average adverse effect decreases from control to treatment group; but to check how significant this change is, one might be interested in testing the one-side hypothesis

$$H_0 : \theta_2 = \theta_1 \quad \text{against} \quad H_1 : \theta_2 > \theta_1. \tag{28}$$

We have applied our proposed Wald-type tests for this problem, as developed in Section 4, to both the full dataset and after deleting the first two outliers from both the groups; the resulting p-values are presented in Figure 4a. Clearly, the classical Wald test results in completely different inference due to the inclusion of these outlying observations – it's p-value becomes significant from non-significant inference without them (at 95% level). On the other hand, proposed MDPDE based robust Wald-type tests with $\beta > 0$ gives stable results (accept the null hypothesis) even in presence of outlying observations.

Table 4: MDPDEs of Poisson parameter θ for the Adverse Events Data in Table 3.

	(β)

	Group	0	0.1	0.3	0.5	0.7	0.9	1
With	Treatment	15.00	7.25	6.94	6.35	5.86	6.05	5.70
Outlier	Control	18.47	8.25	7.75	7.56	7.53	7.41	7.81
Without	Treatment	8.82	7.47	6.44	6.20	6.14	5.58	6.58
Outlier	Control	9.65	7.97	7.63	7.61	7.56	7.68	7.75

5.2 Poisson model for experimental trial: Drosophila data

We next consider another application to the Poisson model with data from an controlled experimental trial with Drosophila flies producing occasional spurious counts. The dataset contains two independent samples on the numbers of recessive lethal mutations observed among the daughters of male flies who are exposed either to a certain degree of chemical to be screened (treatment group) or to control conditions. This dataset has been previously analyzed by many statisticians including Woodru et al. [17], Simpson [16], Basu et al. [3] who have shown that the response data can be modeled by Poisson distribution, but there are two outlying observations in one sample that affects the likelihood based inference and so the classical Wald test. See Basu et al. [3][Table 7] for the dataset and the MDPDEs of the Poisson parameters.

Here, we will apply the proposed Wald-type tests for comparing the Poisson parameters for the two samples, say θ_1 and θ_2, through testing the one-sided hypothesis in eq. (28). The resulting p-values are presented in Figure 4b. Clearly, in presence of outliers, the classical rejects the null hypothesis indicating that the average number of mutation is significantly more for the second sample, which is the opposite of the true inference obtained after removing these outliers from the second sample. But, the proposed MDPDE based Wald-type tests with $\beta \geq 0.1$ produce robust results even in presence of outliers accepting the null hypothesis.

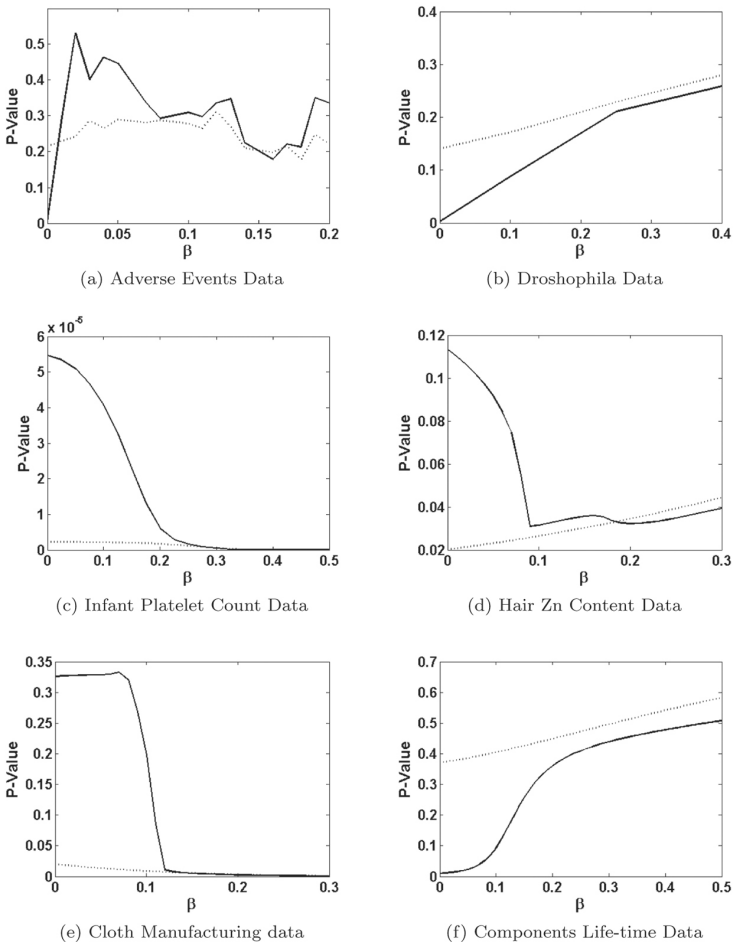


Figure 4: P-values of the proposed Wald-type tests under the real data examples with outliers (solid line) and without outliers (doted line).

5.3 Normal model for clinical trial: Infant platelet count data

We will now present another clinical trial example from Karpatkin et al. [18] to illustrate the applications under the normal model. This clinical trial was conducted to study if the infant platelet count can be increased by giving steroids to the mothers with autoimmune thrombocytopenia during pregnancy. The study consists of 19 mothers with 12 being given steroid (treatment group) and 7 not given steroid (control group) and the corresponding infant platelet counts (in thousands, per mm^3) after delivery are given in Table 5. These can be modeled by a normal model with means θ_1, θ_2 and the variances σ_1^2, σ_2^2 for the treatment and control groups respectively. Then, the primary research problem can be solved by testing the one-sided hypothesis in eq. (28) with σ_1^2 and σ_2^2 being unknown.

Table 5: Infant Platelet count after delivery (in thousands, per mm^3) in the Karpatkin et al. [18] clinical trail study.

Treatment	120	124	215	90	67	126	95	190	180	135	399	65
Control	12	20	112	32	60	40	18					

The p-values for this testing problem obtained by applying the proposed Wald-type tests, as described in Example 4.1, are presented in Figure 4c for different $\beta \geq 0$. One can easily observe that there is a large outlier value of 399 (thousands) in the treatment group that affects the classical Wald test (at $\beta = 0$). However, our MDPDE based proposal with $\beta > 0$ produces stable p-value ignoring the effect of the outlying observation.

5.4 Normal model for health study: Hair Zn content data

Two-sample test under the normal model has many possible applications from which we now present a health study to examine the impact of polluted urban environment over individual health in Sri Lanka. The dataset consist of the zinc (Zn) content of the hair of two independent samples taken from urban (polluted) and rural (unpolluted) Sri Lanka and our target is to check if the Zn content is more for polluted urban residents impacting their health conditions. The dataset was presented in Basu et al. [5][Table 6] and it has been shown their that each sample can be modeled by normal distributions with means θ_i and variance σ_i^2 ($i = 1,2$ for rural and urban groups respectively) except for two possible outliers. There is one outlier in each of the samples that affects the MLE based inference while testing for the targeted hypothesis eq. (28) of comparing θ_1 and θ_2 with unknown σ_1^2 and σ_2^2.

We have applied the proposed MDPDE based Wald-type test for this problem following Example 4.1 and the resulting p-values are presented in Figure 4d. Clearly, the significance increase of the zinc contents in urban residents cannot be identified by the classical Wald-test in presence of outliers, but our proposal with $\beta \geq 0.1$ gives stable and correct inference ignoring the effect of the outliers.

5.5 Normal model for quality control: Cloth manufacturing data

Our third and final example with normal model will be in the context of quality control based on the data from the Levi-Strauss clothing manufacturing plant. The dataset consists of 22 measurements on run-up (a percentage measure of wastage in cloth) for each of two particular mills supplying cloths to the plant [5][Table 1]. To control the quality of the cloths, the plant want to test for the consistency of the run-up measures from the two mills. Since the sample from each mill can be modeled by normal distribution with mean θ_i and variance σ_i^2 ($i = 1,2$), the objective is then to test for the both sided hypothesis

$$H_0 : \theta_1 = \theta_2 \quad \text{against} \quad H_1 : \theta_1 \neq \theta_2, \tag{29}$$

with σ_1^2 and σ_2^2 being unknown under both cases. However, as illustrated in [5], the dataset contains 3 potential outliers that make the MLE based inference highly non-robust. Hence the classical Wald test rejects the null hypothesis in presence of outliers whereas it accept the null after removing the outliers. When we apply the proposed MDPDE based Wald-type problem, following the description as in Example 3.1, the corresponding p-values (reported in Figure 4e) becomes highly stable for $\beta \geq 1.5$ rejecting the null hypothesis even in presence of the outliers.

5.6 Exponential model for reliability testing: Components life-time data

We will end this section with an example of exponential model used in reliability testing between two sets of products' lifetimes. We will use the (simulated) data from Perng [19] which consist of the lifetimes (in thousand of hours) of a particular electronic components produced by two different processes (see Table 6). Each sample can be then modeled by exponential distributions with mean θ_i ($i = 1,2$). Our objective in reliability testing of the manufacturing process is to test whether the lifetimes for both the process have the same distributions, i.e., if $\theta_1 = \theta_2$ against the both-sided alternatives as in the hypothesis eq. (29). It has been observed that there is no significant difference in the distributions of both the processes and so the null hypothesis should be accepted by any standard test.

Table 6: Lifetimes (in thousand of hours) of a particular electronic components produced by two different processes [19].

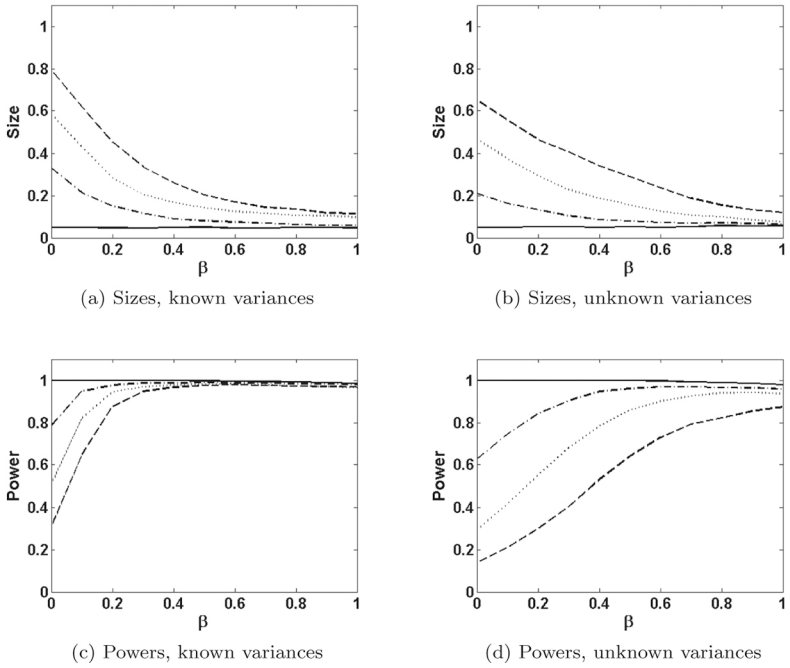
Process 1	0.044	0.134	0.142	0.158	0.216	0.625	0.649	0.658	1.062	1.140	1.159	1.238
Process 2	0.060	0.174	0.237	0.272	0.335	0.391	0.670	0.902	1.543	1.615	2.013	2.309

Since there is no outliers in this dataset, in order to study the robustness aspect of our proposal we add one outlying value of 20 (assuming a decimal is misplaced by one digit from 2.0) in the second sample. The resulting p-values obtained by the proposed Wald-type tests for both the pure data and with this artificial outlier are presented in Figure 4f for different β. Clearly, the classical Wald test changes drastically by rejecting null due to insertion of only one outlying observations, but our proposed Wald-type tests with $\beta \geq 0.1$ remains stable and still accept the null hypothesis robustly in presence of the outlier.

6 Simulation study and the choice of tuning parameter β

Finally, to examine the finite sample performances of our Wald-type tests, we have performed several simulation studies with all the models considered in the previous section for real datasets. However, noting the similarity of the results for different models, for brevity, here we will report the results from only one simulation study under normal model with two-sided alternatives.

We simulate 1000 pairs of samples, each of size $n = 50$, independently drawn from $N(\theta_i, 1)$ distributions ($i = 1,2$) and perform the proposed Wald-type tests for testing $H_0 : \theta_1 = \theta_2$ against the two-sided alternative $H_1 : \theta_1 \neq \theta_2$, first assuming the variances to be known (both equal to 1) and then for unknown and possibly unequal variances, following Examples 2.1 and 3.8, respectively. Then, we compute the empirical sizes and powers of the proposed test under such pure data over 1000 iterations, where for size calculation we have taken $\theta_1 = \theta_2 = 0$ and for power calculation $\theta_1 = 0$, $\theta_2 = 1$. Next, to study the robustness of these tests, we contaminate $100\varepsilon\%$ of the second sample in each iteration (for $\varepsilon = 0.1,0.15,0.2$) by observations from a $N(\theta_c, 1)$ distribution and repeat the above simulation to compute empirical sizes and powers under contamination. We have taken $\theta_c = 3$ and -3 for studying the robustness of size and power, respectively. Note that these contamination distributions are not very far from the corresponding true distributions and hence generate reasonably common practical situations. The resulting empirical sizes and powers are reported in Figure 5.



(a) Sizes, known variances (b) Sizes, unknown variances

(c) Powers, known variances (d) Powers, unknown variances

Figure 5: Empirical sizes and powers of the proposed Wald-type tests for testing equality of two normal means with both the known and unknown variance case at sample size $n = 50$ under pure data (solid line) and with contamination of 10% (dash-doted line), 15% (doted line) and 20% (dashed line).

It can be easily observed from Figure 5 that the size and power of the proposed Wald-type tests under pure data change (increase and decrease, respectively) only very slightly with increasing β, but their stabilities increase significantly. In particular, under contamination, both size and power of the tests near $\beta = 0$, the classical Wald test, are heavily affected. But larger positive values of β make these measures much more stable for both the known and unknown variance cases. However, for the cases of known (and correctly specified) variances we get highly stable results even for β as low as 0.3 or 0.4, whereas we need $\beta \approx 0.5, 0.6$ for similar stability in the case of unknown variances. This is intuitively expected since under the present contamination schemes the variance estimates also change and we need stronger downweighting to get overall stable inference with larger values of β.

To further illustrate the advantages of our proposed tests compared to the non-parametric Wilcoxon rank-sum test, we have repeated the above simulation exercise to derive the corresponding empirical sizes and powers of the Wilcoxon test. This Wilcoxon test is equivalent to the two-sample Mann-Whitney test and is the most commonly used default method for robust two-sample tests of hypotheses. The resulting values of its empirical sizes and powers are reported in Table 6 along with the same for the classical Wald test and the proposed MDPDE based Wald-type tests at some particular β assuming equal but unknown variances. It is evident from Table 6 that the non-parametric Wilcoxon test is slightly robust compared to the classical Wald test but it still has a high degree of non-robustness under higher contamination levels. Our proposed Wald-type tests with larger $\beta > 0$ perform much more robustly compared to both the Wald test and the Wilcoxon test under contaminated data and perform very competitively under pure data. These observations appear to indicate that, when the parametric model is even approximately correct, our proposed tests indeed serve as very useful and significantly improved simple robust alternatives to the existing likelihood based or non-parametric solutions for the two-sample problems arising frequently in biostatistics and many other disciplines.

Table 7: Empirical sizes and powers for the classical Wald test, the non-parametric Wilcoxon rank-sum test and the proposed MDPDE based Wald-type tests at different β under pure and contaminated data (assuming equal but unknown variances).

	Cont. Prop.	Wald Test	Wilcoxon Test	MDPDE based Wald-Type tests with β 0.1	0.3	0.5	0.7	1
Size	0%	0.049	0.047	0.049	0.053	0.052	0.052	0.058
	10%	0.209	0.116	0.163	0.104	0.079	0.069	0.064
	15%	0.466	0.248	0.374	0.228	0.155	0.106	0.075
	20%	0.652	0.395	0.556	0.408	0.289	0.187	0.119
Power	0%	1.000	1.000	1.000	1.000	0.999	0.993	0.979

10%	0.628	0.908	0.747	0.904	0.961	0.970	0.959
15%	0.292	0.728	0.416	0.681	0.859	0.926	0.937
20%	0.138	0.492	0.209	0.403	0.641	0.793	0.874

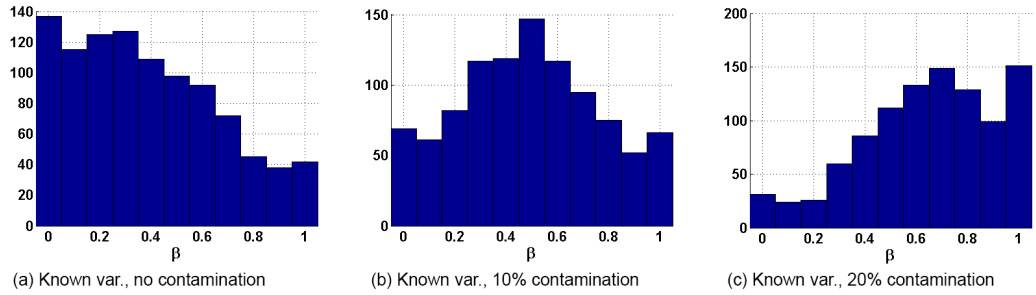
Throughout all our example and simulations above, we have notices that the tuning parameter β controls between robustness of the proposed Wald-type tests and its asymptotic contiguous power under pure data. So, we need to chose β properly for any practical applications. In particular we note that, in most of the example models, the loss in power is not significant enough at small positive β, whereas we get highly robust inferences for $\beta \geq 0.3$ (except for few cases with very high contaminations where we may need $\beta \approx 0.4, 0.5$). Therefore, an empirical suggestion for the choice of β in any application suspecting some contamination could be within the range $\beta \in [0.3, 0.5]$ for generating robust inference without significant loss in power.

Although this ad hoc empirical choice of β works well enough in most practical datasets suspectable to outliers, many practitioners will prefer a data-driven choice of β in case of no idea on the level of contamination in dataset that might produce a better trade-off. In this respect, we note that the performance of the proposed Wald-type tests directly depends on that of the MDPDE (with tuning parameter β) used in constructing the test statistics. In particular the asymptotic contiguous power of the proposed test has the same nature as the asymptotic efficiency of the corresponding MDPDE whereas all the robustness measures of our tests directly depend on the robustness of the MDPDE through its influence function. So, a suitable data-driven choice of β for our Wald-type test statistics also can be equivalently formed by adjusting the trade-off between efficiency and robustness of the MDPDE used. For this second problem, Warwick and Jones [20] proposed to minimize an estimator of MSE of the MDPDE to chose optimum β. Based on the first sample $X_1, \ldots, X_n$, they proposed to minimize the estimated MSE

$$\widehat{MSE}_n(\beta) = \left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_\beta^P\right)^T \left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_\beta^P\right) + \frac{1}{n} Trace\left(\hat{\boldsymbol{J}}_{\beta,n}^{-1} \widehat{\boldsymbol{K}}_{\beta,n} \hat{\boldsymbol{J}}_{\beta,n}^{-1}\right) \tag{30}$$

over β, where $\boldsymbol{\theta}_\beta^P$ is a pilot estimator of the target parameter and $\hat{\boldsymbol{J}}_{\beta,n}$ and $\widehat{\boldsymbol{K}}_{\beta,n}$ are estimators of the matrices $\boldsymbol{J}_\beta$ and $\boldsymbol{K}_\beta$ respectively, which can be easily obtained from their expressions by substituting $\boldsymbol{\theta}$ by the MDPDE and integrations by sample means.

Although there is no direct choice for $\boldsymbol{\theta}_\beta^P$, Warwick and Jones [20] suggested, based on an extensive simulation studies, that the MDPDE with $\beta = 1$ can serve the purpose well for the i.i.d. set-up and we will stick to that suggestion for the present case also (the non-i.i.d. cases have been studied in [21, 22]). However, the problem in the present two-sample case is that, the optimum β obtained by minimizing $\widehat{MSE}_n(\beta)$ based on the first sample may not be the same as that obtained for the second sample due to possible different level of contaminations. As a standard solution, we propose the minimization of the total estimated MSE, the sum of the MSE estimates based on two samples separately, over $\beta \in [0, 1]$ to obtain the optimum choice of the tuning parameter for the present two-sample testing problem.



(a) Known var., no contamination (b) Known var., 10% contamination (c) Known var., 20% contamination

Figure 6: Histograms for optimally chosen tuning parameter β under normal models with different contamination levels.

We have implemented this proposal for the above simulation study with normal model to check its effectiveness. Figure 6 presents the histograms of the 1000 selected optimum β following this proposal for the normal model with known and equal variances under the simulation scheme used for studying size stability above (in Figure 5). Clearly, the mode of these optimum βs shift from 0 to 1 as the contamination proportion increases yielding the expected trade-off between the power and robustness based on the level of contaminations.

7 Concluding remarks

In this paper, we have considered the problem of testing with two independent samples of i.i.d. observations and proposed a class of robust Wald-type tests for both simple and composite hypothesis testing. These Wald-

type tests are constructed using the robust minimum density power divergence estimators of the underlying parameters in each sample. The asymptotic and robustness properties of the proposed Wald-type tests have been discussed along with their applications to several important real-life problems like clinical trial, medical experiment, reliability testing and many more.

Our focus in this paper has been on robust two sample tests. Nonparametric methods and robust methods share some common goals, yet robust methods are inherently different from nonparametric methods as they are essentially parametric, although they allow the parametric model to be only approximately true. It is well known that when a parametric model does hold, the parametric procedures are much more efficient compared to the nonparametric methods. However, when the parametric model holds only approximately, the robust methods are still often substantially more efficient in doing inference about the major component of the data generating distribution compared to nonparametric methods. This has been amply demonstrated by the simulations reported in Table 7. And while parametric models may never "exactly" fit the data, they often provide reasonable "approximate" fits to many practical data sets. So we expect that our method will have a better scope of application in real problems compared to classical parametric methods, and will have greater efficiency in many cases compared to nonparametric methods; in either case, our method will have better robustness properties.

Although we have discussed all possible types of general two-sample hypotheses, in this paper, we have restricted our attention to the cases where each of the two independent samples is identically distributed. The natural extension of this work will be to develop robust tests for hypotheses involving two independent samples from non-homogeneous populations; this also has many practical applications including comparing the regression lines between two groups of patients in a fixed design clinical trial. Also, one could further explore the possibility of robust hypothesis testing using the minimum density power divergence estimators for two paired samples or for more than two sample cases. we hope to pursue some of this possible extensions in our future research.

8 Proof of Results

8.1 Proof of Theorem 2.1

Using the asymptotic distribution of $\sqrt{n}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1)$ and $\sqrt{n}({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2)$, we have

$$\sqrt{\frac{mn}{m+n}}\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\right) \underset{m,n\to\infty}{\overset{\mathscr{L}}{\longrightarrow}} N(\mathbf{0}_p, \omega\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_1))$$

and

$$\sqrt{\frac{mn}{m+n}}\left({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\right) \underset{m,n\to\infty}{\overset{\mathscr{L}}{\longrightarrow}} N(\mathbf{0}_p, (1-\omega)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_2)).$$

Hence under $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}_0$, we get

$$\sqrt{\frac{mn}{m+n}}\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - {}^{(1)}\hat{\boldsymbol{\theta}}_\beta\right) \underset{m,n\to\infty}{\overset{\mathscr{L}}{\longrightarrow}} N(\mathbf{0}_p, \boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)).$$

Further, under H_0, ${}^{(0)}\hat{\boldsymbol{\theta}}_\beta \overset{\mathscr{P}}{\to} \boldsymbol{\theta}_0$ as $m+n \to \infty$. Then the theorem follows using the continuity of the matrix $\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta})$. □

8.2 Proof of Theorem 2.2

Note that, ${}^{(0)}\hat{\boldsymbol{\theta}}_\beta \xrightarrow[n,m\to\infty]{\mathscr{P}} \boldsymbol{\theta}_3$ and hence the asymptotic distribution of $l^*_{{}^{(0)}\hat{\boldsymbol{\theta}}_\beta,\beta}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta)$ is the same as that of $l^*_{\boldsymbol{\theta}_3,\beta}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta)$. Now, a suitable Taylor series expansion leads to

$$\begin{aligned} l^*_{\boldsymbol{\theta}_3,\beta}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta) - l^*_{\boldsymbol{\theta}_3,\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) &= \left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\right)^T \frac{\partial}{\partial\boldsymbol{\theta}_1} l^*_{\boldsymbol{\theta}_3,\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) + \left({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\right)^T \frac{\partial}{\partial\boldsymbol{\theta}_2} l^*_{\boldsymbol{\theta}_3,\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \\ &\quad + o_P\left(\|{}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\|^2\right) + o_P\left(\|{}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\|^2\right) \\ &= 2\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\right)^T \boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_3)^{-1}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) - 2\left({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\right)^T \\ &\quad \boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_3)^{-1}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \\ &\quad + o_P\left(\|{}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\|^2\right) + o_P\left(\|{}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\|^2\right) \\ &= 2\left[\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - {}^{(2)}\hat{\boldsymbol{\theta}}_\beta\right) - (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)\right]^T \boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_3)^{-1}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \\ &\quad + o_P\left(\|{}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\|^2\right) + o_P\left(\|{}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\|^2\right). \end{aligned}$$

Then, the theorem follows from the above expression by noting that

$$\sqrt{\frac{mn}{m+n}}\left[\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - {}^{(2)}\hat{\boldsymbol{\theta}}_\beta\right) - (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)\right] \xrightarrow[n,m\to\infty]{\mathscr{L}} \mathcal{N}\left(0, \left[\omega\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_1) + (1-\omega)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_2)\right]\right),$$

as $m, n \to \infty$ at any $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$. Here, the last convergence follows from the asymptotic distributions of the MDPDEs ${}^{(1)}\hat{\boldsymbol{\theta}}_\beta$ and ${}^{(2)}\hat{\boldsymbol{\theta}}_\beta$. □

8.3 Proof of Theorem 2.5

Using the asymptotic distribution of $\sqrt{n}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_{1,n})$ and $\sqrt{n}({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_{2,m})$ under $H_{1,n,m}$ and continuity of $\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)$, we have ${}^{(2)}\hat{\boldsymbol{\theta}}_\beta \xrightarrow[m\to\infty]{\mathscr{P}} \boldsymbol{\theta}_0$,

$$\sqrt{\frac{mn}{m+n}}\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0\right) \xrightarrow[m,n\to\infty]{\mathscr{L}} N(\sqrt{\omega}\boldsymbol{\Delta}_1, \omega\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0))$$

and

$$\sqrt{\frac{mn}{m+n}}\left({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0\right) \xrightarrow[m,n\to\infty]{\mathscr{L}} N(\sqrt{1-\omega}\boldsymbol{\Delta}_2, (1-\omega)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)).$$

Hence, under $H_{1,n,m}$, we get

$$\sqrt{\frac{mn}{m+n}}\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - {}^{(1)}\hat{\boldsymbol{\theta}}_\beta\right) \xrightarrow[m,n\to\infty]{\mathscr{L}} N(\sqrt{\omega}\boldsymbol{\Delta}_1 - \sqrt{1-\omega}\boldsymbol{\Delta}_2, \boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)),$$

from which the theorem follows immediately. □

8.4 Proof of 2.6

We will only prove the case $(D_1, D_2) = (F^P_{1,m,\varepsilon,x}, F^P_{2,n,\varepsilon,y})$. Other two cases will follow similarly.

Let us denote $\boldsymbol{\theta}^*_{1,n} = \boldsymbol{U}_\beta(F^P_{1,m,\varepsilon,x})$ and $\boldsymbol{\theta}^*_{2,m} = \boldsymbol{U}_\beta(F^P_{2,n,\varepsilon,y})$. Then using the continuity of $\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)$, we get under $(D_1, D_2) = (F^P_{1,m,\varepsilon,x}, F^P_{2,n,\varepsilon,y})$, the asymptotic distribution of $\sqrt{n}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^*_{1,n})$ and $\sqrt{n}({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}^*_{2,m})$ are both p-variate normal with mean zero and variance $\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)$. Further, a suitable Taylor series expansion yields

$$\begin{aligned} \boldsymbol{\theta}^*_{1,n} &= \boldsymbol{\theta}_{1,n} + \frac{\varepsilon}{\sqrt{n}}\mathcal{IF}(x; \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}_{1,n}}) + o(n^{-1/2}) \\ &= \boldsymbol{\theta}_0 + \frac{\boldsymbol{\Delta}_1}{\sqrt{n}} + \frac{\varepsilon}{\sqrt{n}}\mathcal{IF}(x; \boldsymbol{U}_\beta, F_{\boldsymbol{\theta}_{1,n}}) + o(n^{-1/2}) \\ &= \boldsymbol{\theta}_0 + \frac{\widetilde{\boldsymbol{\Delta}_1}}{\sqrt{n}} + o(n^{-1/2}). \end{aligned}$$

Similarly, we have

$$\boldsymbol{\theta}^*_{2,m} = \boldsymbol{\theta}_0 + \frac{\widetilde{\boldsymbol{\Delta}_2}}{\sqrt{n}} + o(n^{-1/2}).$$

Combining all these, we get

$$\sqrt{\frac{mn}{m+n}}\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0\right) \underset{m,n\to\infty}{\overset{\mathscr{L}}{\longrightarrow}} N(\sqrt{\omega}\widetilde{\boldsymbol{\Delta}}_1, \omega\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0))$$

and

$$\sqrt{\frac{mn}{m+n}}\left({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_0\right) \underset{m,n\to\infty}{\overset{\mathscr{L}}{\longrightarrow}} N(\sqrt{1-\omega}\widetilde{\boldsymbol{\Delta}}_2, (1-\omega)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)).$$

Hence, under $(D_1, D_2) = (F^P_{1,m,\varepsilon,x}, F^P_{2,n,\varepsilon,y})$, we get

$$\sqrt{\frac{mn}{m+n}}\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - {}^{(1)}\hat{\boldsymbol{\theta}}_\beta\right) \underset{m,n\to\infty}{\overset{\mathscr{L}}{\longrightarrow}} N(\sqrt{\omega}\widetilde{\boldsymbol{\Delta}}_1 - \sqrt{1-\omega}\widetilde{\boldsymbol{\Delta}}_2, \boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0))$$

and hence the theorem follows immediately. □

8.5 Proof of Theorem 3.1

Using suitable Taylor series expansion, we get

$$\begin{aligned}\boldsymbol{\psi}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta) &= \boldsymbol{\psi}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) + \boldsymbol{\Psi}_1(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1) + \boldsymbol{\Psi}_2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2)\\ &\quad + o_P\left(\|{}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\|\right) + o_P\left(\|{}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\|\right).\end{aligned}$$

Now, from the asymptotic distribution of $\sqrt{n}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1)$ and $\sqrt{n}({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2)$ it follows that

$$\sqrt{\frac{mn}{m+n}}\boldsymbol{\Psi}_1(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\right) \underset{m,n\to\infty}{\overset{\mathscr{L}}{\longrightarrow}} N(0, \omega\boldsymbol{\Psi}_1(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_1)\boldsymbol{\Psi}_1(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2))$$

and

$$\sqrt{\frac{mn}{m+n}}\boldsymbol{\Psi}_2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T\left({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\right) \underset{m,n\to\infty}{\overset{\mathscr{L}}{\longrightarrow}} N(0, (1-\omega)\boldsymbol{\Psi}_2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_2)\boldsymbol{\Psi}_2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)).$$

Hence under $H_0 : \boldsymbol{\psi}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \mathbf{0}_r$, we get

$$\sqrt{\frac{mn}{m+n}}\boldsymbol{\psi}\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta\right) \underset{m,n\to\infty}{\overset{\mathscr{L}}{\longrightarrow}} N(\mathbf{0}_r, \widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)).$$

Finally, by the consistency of the MDPDEs and the continuity of the matrices $\boldsymbol{\Psi}_1$, $\boldsymbol{\Psi}_2$ and $\boldsymbol{\Sigma}_\beta$, it follows that $\widetilde{\boldsymbol{\Sigma}_\beta}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta^{(2)}\hat{\boldsymbol{\theta}}_\beta) \overset{\mathscr{P}}{\to} \widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ as $m+n\to\infty$, from which the theorem follows immediately. □

8.6 Proof of Theorem 3.2

Using an appropriate Taylor series expansion, we get

$$\begin{aligned}&\widetilde{l^*}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta) - \widetilde{l^*}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2 = \left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\right)^T\frac{\partial}{\partial\boldsymbol{\theta}_1}\widetilde{l^*}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) + \left({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\right)^T\frac{\partial}{\partial\boldsymbol{\theta}_2}\widetilde{l^*}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\\ &\qquad + o_P\left(\|{}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\|^2\right) + o_P\left(\|{}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\|^2\right)\\ =\ & 2\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\right)^T\boldsymbol{\Psi}_1(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^{-1}\boldsymbol{\psi}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) + 2\left({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\right)^T\boldsymbol{\Psi}_2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^{-1}\boldsymbol{\psi}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\\ &\qquad + o_P\left(\|{}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\|^2\right) + o_P\left(\|{}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\|^2\right)\\ =\ & 2\left[\boldsymbol{\Psi}_1(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\right) + \boldsymbol{\Psi}_2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T\left({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\right)\right]^T\widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^{-1}\boldsymbol{\psi}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\\ &\qquad + o_P\left(\|{}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_1\|^2\right) + o_P\left(\|{}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_2\|^2\right).\end{aligned}$$

Then, the theorem follows from the asymptotic distributions of the MDPDEs ${}^{(1)}\hat{\boldsymbol{\theta}}_\beta$ and ${}^{(2)}\hat{\boldsymbol{\theta}}_\beta$. □

8.7 Proof of 3.4

Using the asymptotic distribution of $\sqrt{n}({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_{1,n})$ and $\sqrt{n}({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_{2,m})$ under $H_{1,n,m}$ and continuity of $\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_0)$, we have, as $m, n \to \infty$, ${}^{(2)}\hat{\boldsymbol{\theta}}_\beta \to^{\mathcal{P}} \boldsymbol{\theta}_0$,

$$\sqrt{\frac{mn}{m+n}}\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_{10}\right) \underset{m,n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} N(\sqrt{\omega}\boldsymbol{\Delta}_1, \omega\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_1))$$

and

$$\sqrt{\frac{mn}{m+n}}\left({}^{(2)}\hat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_{20}\right) \underset{m,n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} N(\sqrt{1-\omega}\boldsymbol{\Delta}_2, (1-\omega)\boldsymbol{\Sigma}_\beta(\boldsymbol{\theta}_2)).$$

Hence, following the proof of Theorem 2.5, we get under $H_{1,n,m}$

$$\sqrt{\frac{mn}{m+n}}\boldsymbol{\psi}\left({}^{(1)}\hat{\boldsymbol{\theta}}_\beta, {}^{(2)}\hat{\boldsymbol{\theta}}_\beta\right) \underset{m,n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} N\left(\left[\sqrt{\omega}\boldsymbol{\Psi}_1(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T\boldsymbol{\Delta}_1 + \sqrt{1-\omega}\boldsymbol{\Psi}_2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)^T\boldsymbol{\Delta}_2\right], \widetilde{\boldsymbol{\Sigma}_\beta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\right),$$

from which the theorem follows immediately. □

8.8 Proof of Theorems 3.2 and 3.7

These proofs are similar to that of Theorems 2.6 and 2.7 and hence omitted. □

References

[1] Basu A, Shioya H, Park C. Statistical inference: the minimum distance approach. Boca Raton, FL: Chapman & Hall/CRC, 2011.
[2] Pardo L. Statistical Inference based on Divergences. CRC/Chapman-Hall, 2006.
[3] Basu A, Mandal A, Martin N, Pardo L. Testing statistical hypotheses based on the density power divergence. Ann Inst Stat Math. 2013;65:319–48.
[4] Basu A, Harris IR, Hjort NL, Jones MC. Robust and efficient estimation by minimising a density power divergence. Biometrika. 1998;85:549–59.
[5] Basu A, Mandal A, Martin N, Pardo L. Robust tests for the equality of two normal means based on the density power divergence. Metrika. 2015;78:611–34.
[6] Basu A, Mandal A, Martin N, Pardo L. Generalized Wald-type tests based on minimum density power divergence estimators. Statistics. 2016;50:1–26.
[7] Ghosh A, Mandal A, Martin N, Pardo L. Influence analysis of robust Wald-type tests. J Multivariate Anal. 2016;147:102–26.
[8] Lehmann EL. Theory of point estimation. John Wiley & Sons, 1983.
[9] Hampel FR, Ronchetti E, Rousseeuw PJ, Stahel W. Robust statistics: the approach based on influence functions. New York, USA: John Wiley & Sons, 1986.
[10] Huber-Carol C. Etude asymptotique de tests robustes. Ph. D. thesis, ETH, Zurich, 1970.
[11] Heritier S, Ronchetti E. Robust bounded-influence tests in general parametric models. J Am Stat Assoc. 1994;89:897–904.
[12] Toma A, Broniatowski M. Dual divergence estimators and tests: robustness results. J Multivariate Anal. 2011;102:20–36
[13] Ghosh A, Basu A, Pardo L. On the robustness of a divergence based test of simple statistical hypotheses. J Stat Planning Inference. 2015;116:91–108.
[14] Kotz S, Johnson NL, Boyd DW. Series representations of distributions of quadratic forms in normal variables. I. Non-central case. Ann Math Stat. 1967;38:838–48.
[15] Kerstjens HAM, Engel M, Dahl R, Paggiaro P, Beck E, Vandewalker M, Sigmund R, Seibold W, Moroni-Zentgraf P, Bateman ED. Tiotropium in asthma poorly controlled with standard combination therapy New England J Med. 2012;367:1198–207.
[16] Simpson DG. Hellinger deviance test: efficiency, breakdown points, and examples. J Am Stat Assoc. 1989;84:107–13.
[17] Woodruff RC, Mason JM, Valencia R, Zimmering A. Chemical mutagenesis testing in drosophila – I: Comparison of positive and negative control data for sex-linked recessive lethal mutations and reciprocal translocations in three laboratories. Environ Mutagen. 1984;6:189–202.
[18] Karpatkin M, Porges RF, Karpatkin S. Platelet counts in infants of women with autoimmune thrombocytopenia: effect of steroid administration to the mother. New England J Med. 1981;305;936–9.
[19] Perng SK. A test for equality of two exponential distributions. Stat Neerlandica. 1978;32:93–102.
[20] Warwick J, Jones MC. Choosing a robustness tuning parameter. J Stat Comput Simul. 2005;75:581–8.
[21] Ghosh A, Basu A. Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. Electr J Stat. 2013;7:2420–56.
[22] Ghosh A, Basu A. Robust estimation for non-homogeneous data and the selection of the optimal tuning parameter: the density power divergence approach. J Appl Stat. 2015;42;2056–72.