

# A robust adaptive-to-model enhancement test for parametric single-index models

Cuizhen Niu<sup>1</sup> · Lixing Zhu<sup>1,2</sup>

Received: 17 November 2016 / Revised: 20 July 2017 / Published online: 2 November 2017  
© The Institute of Statistical Mathematics, Tokyo 2017

**Abstract** This paper is devoted to test the parametric single-index structure of the underlying model when there are outliers in observations. First, a test that is robust against outliers is suggested. The Hampel’s second-order influence function of the test statistic is proved to be bounded. Second, the test fully uses the dimension reduction structure of the hypothetical model and automatically adapts to alternative models when the null hypothesis is false. Thus, the test can greatly overcome the dimensionality problem and is still omnibus against general alternative models. The performance of the test is demonstrated by both Monte Carlo simulation studies and an application to a real dataset.

**Keywords** Bounded influence function · Dimension reduction · Model checking · Omnibus property · Robust adaptive-to-model test

## 1 Introduction

Testing the validity of a specified model up to some unknown parameters is an important issue in statistical inference. A long-standing focus for this model checking problem is their sensitivity to outlying observations or heavy-tailed distributions. This is because these observations may have destructive effects even though with small

---

✉ Lixing Zhu  
lzhu@hkbue.edu.hk

Cuizhen Niu  
nczlb\_890@sina.cn

<sup>1</sup> School of Statistics, Beijing Normal University, No. 19, Xijiekouwai Street, Haidian District, Beijing 100875, China

<sup>2</sup> Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China

violation in usual observations. However, in the past decades, researchers have made more efforts on robust estimation, but paid less attention to robust hypothesis testing. Therefore, it is critical to develop a robust test that can be against outlier contamination.

Sample containing outliers or contaminated data is not uncommon and an ubiquitous problem in many disciplines, for example clinical trials, medical research, longitudinal studies, and so forth. When there exist outliers in the data, robust statistical inference procedures can improve the accuracy and reliability of statistical analyses. The purpose of robust hypothesis testing is twofold, just as stated in [Heritier and Ronchetti \(1994\)](#): One is that under small and arbitrary departures from the null hypothesis, the level of a test should be stable, which is called the robustness of validity; the other is that the test can still make a good power performance under small and arbitrary departures from specified alternatives, that is called the robustness of efficiency. [Wang and Qu \(2007\)](#) suggested a robust version of [Zheng's \(1996\)](#) test. Their numerical studies also showed the necessity of using a robust testing procedure: the effect of outliers on [Zheng's](#) original test is dramatic and destructive so that it cannot maintain the significance level.

Many efforts have been devoted to the development of robust testing procedures. For linear regression models, [Schrader and Hettmansperger \(1980\)](#) proposed the  $\rho_c$  test based on Huber's  $M$  estimators; [Markatou and Hettmansperger \(1990\)](#) introduced an aligned generalized  $M$  test for testing subhypotheses in general linear models, which is a robustification of the well-known  $F$  test and can be viewed as a generalization of [Sen's \(1982\)](#)  $M$  test for linear models. Afterward, [Heritier and Ronchetti \(1994\)](#) and [Markatou and Manos \(1996\)](#) presented robust versions of the Wald, score and drop-in-dispersion tests for general parametric models and nonlinear regression models, respectively. [Wang and Qu \(2007\)](#) developed a robust approach for testing the parametric form of a regression function versus an omnibus alternative, which can be viewed as a robustification of a smoothing-based conditional moment test. [Feng et al. \(2015\)](#) recommended a robust testing procedure to make comparison of two regression curves through combining a Wilcoxon-type artificial likelihood function with generalized likelihood ratio test.

There are a number of proposals available in the literature on testing consistently the correct specification of a parametric regression model. Most of existing test procedures can be classified into two categories: global smoothing tests and local smoothing tests. As mentioned in a comprehensive review paper of [González-Manteiga and Crujeiras \(2013\)](#), the global smoothing tests mainly involve empirical process, which can avoid subjective selection of the smoothing parameter, such as bandwidth; the local smoothing tests are based on nonparametric smoothing techniques such as Nadaraya–Watson kernel estimation ([Nadaraya 1964](#); [Watson 1964](#)), smoothing spline estimation or other local smoothing techniques. Examples of global smoothing tests include [Bierens \(1990\)](#), [Bierens and Ploberger \(1997\)](#), [Stute \(1997\)](#), [Stute et al. \(1998b\)](#), [Whang \(2000\)](#), [Escanciano \(2006a\)](#), among many others. This class of methods enjoys fast convergence rate of order  $O(n^{-1/2})$  (see [Stute et al. \(1998a\)](#)). However, this type of tests is often less sensitive to high-frequency models. To certain extent, this shortcoming makes their practical limitation since it is not uncommon to have high-order or high-frequency models. Also, in higher than 2-dimensional scenarios, the limiting null distribution is usually intractable which requires the assistance from resampling

approximation to determine critical values. Further, the power performance in high-dimensional cases is not very encouraging.

As for local smoothing tests, examples include the tests suggested by [Härdle and Mammen \(1993\)](#), [Zheng \(1996\)](#), [Fan et al. \(2001\)](#), [Horowitz and Spokoiny \(2001\)](#), [Koul and Ni \(2004\)](#), and [Van Keilegom et al. \(2008\)](#). Since they must involve multivariate nonparametric function estimation procedures and thus inevitably and severely suffer from the curse of dimensionality when the number of covariates is large, even moderate. This typical problem is a big obstacle for local smoothing tests to well maintain the significance level and to sense the alternative models. Because of the data sparseness in multi-dimensional space, the behavior of nonparametric smooth estimators very quickly deteriorates as the dimension increases, see [Stone \(1980\)](#). Further, even when there are no outliers, local smoothing tests have the typical slow convergence rate of order  $O(n^{-1/2}h^{-p/4})$  to their limits when  $p$  is large where  $p$  is the number of covariates. Besides, a suitable choice of smooth parameter is difficult but necessary for these tests. Although existing empirical studies show that the effect of bandwidth selection is not too profound in small  $p$  situations, how to make a practically suitable bandwidth choice is still a long-standing problem when the dimension  $p$  is relatively large.

The above observations suggest that a common difficulty both global and local smoothing test procedures suffer from is the data sparseness in high-dimensional space even for large sample sizes (see [Escanciano \(2007\)](#)). To attack this challenge, a representative method documented as projection-pursuit technique was proposed and experimented. The significant feature of this method is to employ the projections of original data onto one-dimensional subspaces: first projecting the original high-dimensional covariates to one-dimensional space to form a linear combination and a test can be obtained as an average of tests based on these selected combinations, see [Huber \(1985\)](#) for detail. [Escanciano \(2006b\)](#) proposed a consistent test for the goodness of fit of parametric regression models, which applied a residual marked empirical process. Some earlier references include [Zhu and Li \(1998\)](#) who suggested a test that is based on an unweighted integral of expectations with respect to all one-dimensional directions. [Zhu and An \(1992\)](#) could be an earlier reference that had already used this idea to deal with a relevant testing problem. [Zhu \(2003\)](#) constructed a lack-of-fit test via seeking for a good projection direction for plotting to achieve the dimension reduction aim. [Lavergne and Patilea \(2008\)](#) introduced the projection-pursuit technique to local smoothing-based tests to avoid the effect of dimension. Afterward, [Lavergne and Patilea \(2012\)](#) suggested a smooth integrated conditional moment (ICM) test. This test is an omnibus test based on the kernel estimation that performs against a sequence of directional nonparametric alternatives as if there were only one covariate whatever the number of covariates is. All of these tests either have no tractable sampling and limiting null distributions or have tractable limiting null distributions that are not useful for the significance level maintenance if critical values are determined by them. Thus, resampling/bootstrap approximations are commonly resorted to determine critical values. However, the algorithm of approximation is very computationally intensive when all projections are involved. [Stute and Zhu \(2002\)](#) simply used a one-dimensional projected covariate that is a linear combination of all covariates to alleviate this typical dimensionality problem. But the disadvantage is also

obvious: it is a directional rather than an omnibus test which cannot detect general alternatives. Another relevant reference is [Stute et al. \(2008\)](#) who also suggested a dimension reduction test that is based on the residual empirical process marked by a set of functions of covariates. This test relies solely on selecting proper functions for the significance level maintenance and power enhancement. A data-driven selection procedure would be in demand.

Recently, a dimension reduction model–adaptive test is proposed by [Guo et al. \(2016\)](#), which is an omnibus test against global alternative models. This local smoothing test procedure introduces a novel model–adaptation concept in model checking for parametric regression models. The test statistic under the hypothetical model can converge to its limit at the rate of order  $O(n^{-1/2}h^{-1/4})$  and detect local alternatives distinct from the null model at this rate, which is not affected by the number of covariates. Their test behaves like a local smoothing test as if the number of covariates were one. Another superiority is that it owns tractable limiting null distribution and can work very well even with moderate sample sizes without the assistance of resampling approximation to determine critical values.

All of the above tests can avoid the curse of dimensionality to some extent; they are, however, not robust against outliers, and their efficiency is adversely affected by outlying observations. Our subsequent numerical analysis suggests that the test proposed by [Guo et al. \(2016\)](#) fails to work when there exist outliers because a linear local average of the responses is not robust, as elaborated in [Härdle \(1992\)](#). To tickle this problem, we incorporate the idea in [Guo et al. \(2016\)](#) into a robust model–adaptive smoothing-based conditional moment test. It can possess the robustness property and simultaneously solve the dimensionality problem.

The hypothetical model is the following with a dimension reduction structure:

$$Y = g(\beta^\top X, \theta) + \varepsilon, \quad (1)$$

where  $Y$  is the response with the covariate vector  $X \in \mathbf{R}^p$ . The random error term  $\varepsilon$  and  $X$  are independent.  $\beta$  and  $\theta$  are unknown parameter vectors of dimensions  $p$  and  $d$ , respectively. In addition,  $g(\cdot)$  is a known function and the superscript  $\top$  denotes transposition. As we often have no much information in advance on model structure when this parametric structure assumption is false, a general alternative model is considered as follows:

$$Y = m(B^\top X) + \varepsilon, \quad (2)$$

where  $m(\cdot)$  is an unknown smooth function and  $\varepsilon$  denotes the error. Here,  $\varepsilon$  is independent of  $X$ .  $B$  is an unknown  $p \times q$  orthonormal matrix for an unknown integer  $q$  with  $1 \leq q \leq p$ . This model treats the nonparametric regression  $Y = m(X) + \varepsilon$  as a special case in which the matrix  $B = I_p$  with  $q = p$ .

In this paper, we construct a robust dimension reduction adaptive-to-model test (R-DREAM). It sufficiently invokes the information in both the null and alternative models to get rid of curse of dimensionality and employs centered asymptotic rank transformation technique to achieve the goal of robustness. We further study the local robustness via influence function analysis, which indicates that our R-DREAM has first-order influence function of zero and second-order influence function bounded in

the response direction. Therefore, it is argued that our test can have more stable and robust performance when there are outliers in responses than existing local smoothing tests. It is worthwhile to point out that our test is mainly robust against the error’s distribution and does not need to impose strong conditions on it. However, as the test is for the regression mean function, the second moments of  $X$  is required to derive the asymptotic properties. Thus, like Wang and Qu (2007) and Feng et al. (2015), it cannot handle models without moments of covariates. This deserves a further investigation.

The rest of this article is organized as follows. In Sect. 2, the test is constructed. The approaches to estimate the matrix  $B$  and to identify its structure dimension  $q$  are also stated in this section. Section 3 presents the large sample properties under the null, global and local alternative hypothesis. The local robustness properties through the Hampel influence function analysis are also discussed in this section. Numerical studies including simulation studies and a real data analysis are, respectively, reported in Sects. 4 and 5. All of the proofs are relegated to “Appendix”.

## 2 A robust dimension reduction adaptive-to-model test

As discussed before, the hypotheses of interest are:

$$\begin{aligned}
 H_0 &: Y = g(\beta^\top X, \theta) + \varepsilon \text{ for some } \beta \in R^p, \theta \in R^d; \\
 H_1 &: Y = m(B^\top X) + \varepsilon \text{ and } m(B^\top X) \neq g(\beta^\top X, \theta) \text{ for any } \beta \in R^p, \theta \in R^d, \quad (3)
 \end{aligned}$$

where  $g(\cdot)$  is a known link function and  $m(\cdot)$  is an unknown link function.  $\beta$  and  $\theta$  are, respectively, the parameter vectors of  $p$  and  $d$  dimensions.  $\varepsilon$  denotes the error, which is independent of  $X$ .  $B$  is a  $p \times q$  orthonormal matrix where  $B^\top B = I_q$  and  $1 \leq q \leq p$ .

### 2.1 Test statistic construction

Denote  $e = Y - g(\beta^\top X, \theta)$  and let  $e^* = H(e) - 1/2$ , where  $H(\cdot)$  is the cumulative distribution function of  $e$  and  $H(e)$  follows uniformly distributed on  $(0, 1)$ . Under the null hypothesis  $H_0$ ,

$$E(e^*|X) = E \left\{ H(e) - \frac{1}{2} | X \right\} = 0. \quad (4)$$

In this case, the model (1) is with  $q = 1$ , and

$$E(e^*|X) = 0 \Rightarrow E(e^*|\beta^\top X) = E(e^*|B^\top X) = 0.$$

Further, the following formula

$$E\{e^* E(e^*|B^\top X) f(B^\top X)\} = E \left\{ \left( E(e^*|B^\top X) \right)^2 f(B^\top X) \right\} = 0 \quad (5)$$

holds under  $H_0$ , where  $f(\cdot)$  is the probability density function of  $B^\top X$ .

Under  $H_1$ , we have  $e = Y - g(\beta^\top X, \theta) = m(B^\top X) - g(\beta^\top X, \theta) + \varepsilon$ . The conditional expectation  $E(e^*|X)$  is

$$E(e^*|X) = E \left\{ H \left( m(B^\top X) - g(\beta^\top X, \theta) + \varepsilon \right) - \frac{1}{2} \middle| X \right\}. \tag{6}$$

When  $P(m(B^\top X) - g(\beta^\top X, \theta) = C) = 1$ , the conditional expectation  $E(e^*|X)$  reduces to the unconditional expectation  $E(e^*) = E(H(e) - 1/2) = 0$ . Here, the event that  $E(e^*|X)$  equals to zero occurs only when  $P(m(B^\top X) - g(\beta^\top X, \theta) = C) = 1$  for some constant  $C$ , which only holds under the null hypothesis  $H_0$ . Since we can enlarge the null class of models by including some location shifts, if  $g(\beta^\top X, \theta)$  belongs to the null class of models, then so does  $g(\beta^\top X, \theta) + C$ . In other words, it is reasonable to assume that the null class of models is sufficiently general to contain all location shifts in the  $Y$  direction. However, under  $H_1$ ,  $P(m(B^\top X) - g(\beta^\top X, \theta) = C) \neq 1$  holds. Hence, the conditional expectation  $E(e^*|X)$  is not equal to zero but relates to  $X$ . Different values of  $X$  would make different values of the conditional expectation  $E(e^*|X)$ . Therefore, when  $H_1$  holds,  $E(e^*|X)$  is not equal to zero; further,

$$E(e^*|X) \neq 0 \Rightarrow E(e^*|B^\top X) \neq 0.$$

Thus, we have

$$E\{e^*E(e^*|B^\top X)f(B^\top X)\} = E \left\{ \left( E(e^*|B^\top X) \right)^2 f(B^\top X) \right\} > 0. \tag{7}$$

Based on the different performances of  $E\{e^*E(e^*|B^\top X)f(B^\top X)\}$  under  $H_0$  and  $H_1$  in (5) and (7), respectively, its empirical version can be used as a test statistic. The null hypothesis  $H_0$  is rejected for large values of the test statistic.

Given a random sample  $\{(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)\}$ , define the asymptotic rank transform of  $\hat{e}_i = y_i - g(\hat{\beta}^\top x_i, \hat{\theta})$  as  $n^{-1} \sum_{l=1}^n I(\hat{e}_l \leq \hat{e}_i)$  where  $\sum_{l=1}^n I(\hat{e}_l \leq \hat{e}_i)$  is the rank of  $\hat{e}_i$  among all of the  $n$  residuals. Here,  $\hat{\beta}$  and  $\hat{\theta}$  come from robust estimates of  $\beta$  and  $\theta$ . Further, the corresponding centered asymptotic rank transform of residuals is as follows

$$\hat{e}_i^* = \frac{1}{n} \sum_{l=1}^n I(\hat{e}_l \leq \hat{e}_i) - \frac{n+1}{2n}, \quad l = 1, \dots, n, \tag{8}$$

where  $I(\cdot)$  is the indicator function.

Once an estimator of  $\hat{B}(\hat{q})$  is available, a kernel estimator of the regression function  $E(e_i^*|B^\top X_i)$  can be estimated as follows:

$$\hat{E}(e_i^*|\hat{B}(\hat{q})^\top x_i) = \frac{\frac{1}{n-1} \sum_{j \neq i} \hat{e}_j^* \mathcal{K}_h \{ \hat{B}(\hat{q})^\top (x_i - x_j) \}}{\hat{f}(\hat{B}(\hat{q})^\top x_i)},$$

where  $\hat{e}_j^*$  has been defined in (8).  $\hat{B}(\hat{q})$  is an sufficient dimension reduction (SDR) estimate of the matrix  $B$  with an estimated structural dimension  $\hat{q}$  of  $q$  and the estimates will be specified later. Besides,  $\mathcal{K}_h(\cdot) = \mathcal{K}(\cdot/h)/h^{\hat{q}}$ , where  $\mathcal{K}(\cdot)$  is a  $\hat{q}$ -dimensional kernel function and  $h$  is the bandwidth, and  $\hat{f}(\hat{B}(\hat{q})^\top x_i)$  is a kernel estimator of the density function of  $f(B^\top x_i)$ ,

$$\hat{f}(\hat{B}(\hat{q})^\top x_i) = \frac{1}{n-1} \sum_{j \neq i}^n \mathcal{K}_h\{\hat{B}(\hat{q})^\top(x_i - x_j)\}.$$

Further, a robust dimension reduction adaptive-to-model test (R-DREAM) can be constructed as follows:

$$V_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}_h\{\hat{B}(\hat{q})^\top(x_i - x_j)\} \hat{e}_i^* \hat{e}_j^*. \tag{9}$$

*Remark 1* From the above construction of  $V_n$  in (9), it seems that except for the estimates of the matrix  $B$  and structural dimension  $q$ , the test statistic makes no difference with the test proposed by Wang and Qu (2007) as follows:

$$\tilde{V}_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \tilde{\mathcal{K}}_h(x_i - x_j) \hat{e}_i^* \hat{e}_j^*, \tag{10}$$

where  $\tilde{\mathcal{K}}_h(\cdot) = \tilde{\mathcal{K}}(\cdot/h)/h^p$  with  $\tilde{\mathcal{K}}(\cdot)$  being a  $p$ -dimensional kernel function. Comparing the test statistic in (9) with that in (10), we note that  $\hat{q}$ -dimensional kernel function ( $\hat{q} \leq p$ ) is required in  $V_n$ . The result in Sect. 3 shows that under the null hypothesis  $H_0$ ,  $\hat{q} = 1$  with a probability going to one, which can avoid the curse of dimensionality greatly. Another superiority of the new test is the model-adaptive property, that is, through estimating the matrix  $B$ , the test can automatically adapt the hypothetical and alternative model such that it can have better performance in the significance level maintenance and power enhancement. To be specific, under  $H_0$ ,  $\hat{B}(\hat{q}) \rightarrow c\beta$  for a constant  $c$ , and under  $H_1$   $\hat{q} = q \geq 1$  with a probability going to one,  $\hat{B}(\hat{q}) \rightarrow BC$  for a  $q \times q$  orthogonal matrix  $C$ , adaptive to the alternative model (2).

*Remark 2* Another related test is the dimension reduction model-adaptive test  $T_n^{\text{GWZ}}$  proposed by Guo et al. (2016):

$$T_n^{\text{GWZ}} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}_h\{\hat{B}(\hat{q})^\top(x_i - x_j)\} \hat{e}_i \hat{e}_j. \tag{11}$$

In Sect. 3.3, we will show through Hampel influence function analysis that R-DREAM has more stable and robust behavior than  $T_n^{\text{GWZ}}$  when the response is contaminated.

## 2.2 Identification and estimation of $B$

As the estimates of the matrix  $B$  and structural dimension  $q$  are crucial for our R-DREAM, we first specify the estimate for the matrix  $B$  under given  $q$  and then study how to select  $q$  consistently. The method to estimate  $B$  is inspired by sufficient dimension reduction technique. In fact,  $B$  is not identifiable since for any  $q \times q$  orthogonal matrix  $C$ ,  $m(B^\top X)$  can always be rewritten as  $\tilde{m}(C^\top B^\top X)$ . Therefore, what we can identify is the space spanned by  $B$  via sufficient dimension reduction technique, or in other words, we can identify  $q$  base vectors of the space spanned by  $B$ . In view of easy operation, good performance and robustness of DEE, we consider to employ it to estimate  $B$ . One can read [Zhu et al. \(2010\)](#) to get the details on SIR-based DEE method and [Guo et al. \(2016\)](#) also gives a simple review of it.

Compared with SIR, DEE can avoid selecting the number of slices as there is no optimal solution. Define the new response variable  $\mathcal{U}(t) = I(Y \leq t)$  for any  $t$ , where the indicator function  $I(Y \leq t)$  takes the value 1 if  $Y \leq t$  and 0, otherwise. When SIR is applied, the original related matrix  $\mathcal{M}(t)$  based on SIR is a  $p \times p$  positive semi-definite matrix such that  $\text{span}\{\mathcal{M}(t)\} = \mathcal{S}_{\mathcal{U}(t)|X}$ . Here,  $\mathcal{S}_{\mathcal{U}(t)|X}$  is the central subspace of  $\mathcal{U}(t)|X$ . Given  $\mathcal{M} = E\{\mathcal{M}(T)\}$ , according to Theorem 1 in [Zhu et al. \(2010\)](#),  $\text{span}\{\mathcal{M}\} = \mathcal{S}_{Y|X}$ . In our paper, since  $\varepsilon \perp X$  is assumed and thus we get  $\mathcal{S}_{Y|X} = \mathcal{S}_{E(Y|X)}$ , where the later is the central mean subspace of  $Y$  given  $X$ .

Based on the above description, estimating  $\mathcal{S}_{E(Y|X)}$  amounts to estimating  $\mathcal{M}$ . Given the sample  $\{(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)\}$ , we define the dichotomized responses as  $u_i(y_j) = I(y_i \leq y_j)$ ,  $i, j = 1, \dots, n$ . Thus, for each fixed  $y_j$ , we can obtain a new sample  $\{(u_1(y_j), x_1), (u_2(y_j), x_2), \dots, (u_n(y_j), x_n)\}$  and the estimate  $\mathcal{M}_n(y_j)$  of  $\mathcal{M}(y_j)$  can be gained with SIR. Thus,  $\mathcal{M}$  can be estimated as

$$\mathcal{M}_{n,n} = n^{-1} \sum_{j=1}^n \mathcal{M}_n(y_j), \quad (12)$$

which has been proved to be root- $n$  consistent to  $\mathcal{M}$  by [Zhu et al. \(2010\)](#). The  $q$  eigenvectors of  $\mathcal{M}_{n,n}$  corresponding to its  $q$  largest eigenvalues are applied to estimate  $B$ . For this method, a mild linearity condition is assumed:  $E(X|B^\top X = z)$  is linear in  $z$  ([Li 1991](#)).

## 2.3 Identification of $B$ and Estimation of the structural dimension $q$

In order to get R-DREAM in (9), the estimate of structural dimension  $q$  is necessary for the above method of identifying  $B$ . Here, a ridge-type ratio estimate (RRE) method, which is inspired by [Xia et al. \(2015\)](#), is suggested to determine  $q$  for DEE. It is based on the ratios of the eigenvalues with an artificially added ridge value  $c$ . Denote  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$  to be the eigenvalues of  $\mathcal{M}_{n,n}$  in (12).  $\hat{q}$  can be obtained as

$$\hat{q} = \arg \min_{k=1, \dots, p-1} \frac{\lambda_{k+1} + c}{\lambda_k + c},$$



where the constant  $c = 1/\sqrt{nh}$  is recommended. The consistencies of  $\hat{q}$  under the null hypothesis (1) and global alternative hypothesis (2) are shown in the following lemma.

**Lemma 1** *Assume that the DEE-based matrix  $\mathcal{M}_{n,n}$  is root- $n$  consistent to  $\mathcal{M}$ . Then, the corresponding estimate  $\hat{q} = q$  (under  $H_0$ ,  $q = 1$ ) with a probability going to one as  $n \rightarrow \infty$ . Therefore, for a  $q \times q$  orthogonal matrix  $C$ ,  $\hat{B}(\hat{q})$  is a root- $n$  consistent estimate of  $BC^\top$ .*

### 3 Asymptotic properties

In this section, the large sample properties of the R-DREAM test statistic  $V_n$  in (9) are investigated via its asymptotic distributions under the null hypothesis, global alternative hypothesis and local alternative hypothesis. The robustness properties of our proposed test are also studied through Hampel influence function analysis.

#### 3.1 Limiting null distribution

The asymptotic normality discussed in the following requires that the regression parameter is root- $n$  consistently estimated under  $H_0$  and the residuals must come from a robust fit. Let  $Z = B^\top X$  and

$$Var = \frac{1}{72} \int \mathcal{K}^2(u)du \int p^2(z)dz,$$

where  $p(z)$  is the probability density function of  $Z$ . Moreover,  $Var$  can be consistently estimated by:

$$\widehat{Var} = \frac{1}{72n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K}^2 \left\{ \frac{\hat{B}(\hat{q})^\top (x_i - x_j)}{h} \right\}.$$

We first state the asymptotic property of the R-DREAM test statistic in (9) under the null hypothesis  $H_0$  as follows:

**Theorem 1** *Suppose that conditions (C1)–(C8) in Appendix hold. Under  $H_0$ , we have*

$$nh^{1/2}V_n \Rightarrow N(0, Var).$$

Plugging in a consistent estimator of  $Var$ , a standardized version of  $V_n$  can be defined as

$$S_n = \frac{n-1}{n} \frac{nh^{1/2}V_n}{\sqrt{\widehat{Var}}}. \tag{13}$$

The following corollary can be easily obtained.

**Corollary 1** Under  $H_0$  and Conditions (C1)–(C8) in Appendix, we have

$$S_n^2 \Rightarrow \chi_1^2,$$

where  $\chi_1^2$  is the Chi-square distribution with one degree of freedom.

Theorem 1 and Corollary 1 characterize the asymptotic properties of the test statistic  $V_n$ . Based on Corollary 1,  $p$ -values of R-DREAM can be easily determined by the quantiles of the Chi-square distribution with one degree of freedom. The null hypothesis  $H_0$  is rejected when  $S_n \geq \chi_{1-\alpha}^2(1)$  where  $\chi_{1-\alpha}^2(1)$  is the  $1 - \alpha$  upper quantile of the Chi-square distribution.

### 3.2 Power study

We are now in the position to examine the power performance of our R-DREAM under alternative hypothesis. More specifically, the following sequence of alternative models is under consideration:

$$H_{1n} : Y = g(\tilde{\beta}^\top X, \tilde{\theta}) + C_n m(B^\top X) + \varepsilon, \tag{14}$$

where  $E[m^2(B^\top X)] < \infty$  and  $\{C_n\}$  is a constant sequence. When  $C_n = C$  for a nonzero constant  $C$ , the model is a global alternative model, while when  $C_n$  goes to zero, it is a sequence of local alternative models. In this sequence of models,  $\beta$  is one of the columns in  $B$ . Denote  $\tilde{\alpha} = (\tilde{\beta}^\top, \tilde{\theta}^\top)^\top$ . For a robust estimate  $\hat{\alpha}$ , we have  $\hat{\alpha} - \tilde{\alpha} = O_p(1/\sqrt{n})$ . We first discuss the consistency of  $\hat{q}$  under the local alternative hypothesis (14). When  $n \rightarrow \infty$ , the local alternative models converge to the null model,  $\hat{q}_s$  under the local alternative models are expected to converge to  $\hat{q}$  under the null model, which finally converge to the structural dimension  $q = 1$  under the null model.

**Lemma 2** Assume conditions (C1)–(C8) in Appendix hold, and under the local alternative hypothesis (14) with  $C_n = n^{-1/2}h^{-1/4}$ , we have  $\hat{q} = 1$  as  $n \rightarrow \infty$  with a probability going to one, where  $\hat{q}$  is the DEE-based estimate.

The asymptotic properties under global and local alternative hypotheses are concluded in the following theorem.

**Theorem 2** Under Conditions (C1)–(C8) in Appendix, we have:

(i) Under the global alternative of (2) or equivalently the above model with  $C_n = C > 0$ ,

$$S_n/(nh^{1/2}) \Rightarrow \text{Constant} > 0;$$

(ii) Under the local alternative hypothesis (14) with  $C_n = n^{-1/2}h^{-1/4}$ , we have

$$nh^{1/2}V_n \Rightarrow N(\mu, \text{Var}) \quad \text{and} \quad S_n^2 \Rightarrow \chi_1^2(\mu^2/\text{Var}),$$

where

$$\mu = E[h^2(\varepsilon)m^2(B^\top X)p(X)],$$

$h(\cdot)$  denotes the probability density function of  $\varepsilon$  and  $\chi_1^2(\mu^2/\text{Var})$  is a non-central Chi-squared random variable with one degree of freedom and the non-centrality parameter  $\mu^2/\text{Var}$ .

The above theorem indicates that under the global alternative hypothesis, the R-DREAM test is consistent with the asymptotic power 1, and it can also detect the local alternatives distinct from the null hypothesis at a nonparametric rate of order  $n^{-1/2}h^{-1/4}$ , which is the optimal rate with one-dimensional predictor for the test  $\tilde{V}_n$  in (10).

*Remark 3* From the above theorem, we can observe that under the null hypothesis, the use (automatically) of lower-order kernel function in  $V_n$  in (9) makes a very significant improvement than  $\tilde{V}_n$  in (10). Based on Theorem 1,  $V_n$  owns a much faster convergence rate of order  $nh^{1/2}$  and  $nh^{1/2}V_n$  is asymptotically normal under the null, whereas the rate of order  $nh^{p/2}$  is for  $\tilde{V}_n$ . Further, according to Theorem 2, the conclusion can be made that  $V_n$  is much more sensitive than Wang and Qu's test  $\tilde{V}_n$  in the sense that  $V_n$  can detect the local alternatives distinct from the null at the rate of  $n^{-1/2}h^{-1/4}$ , whereas  $\tilde{V}_n$  is only workable at the rate of order  $n^{-1/2}h^{-p/4}$ . Therefore, the power performance of the proposed test can be much enhanced.

*Remark 4* Our original simulation results based on  $S_n$  in (13) with DEE-based estimate suggest the conservative sizes of tests. Thus, the following size adjustment is needed for the test statistics with the DEE-based estimate:

$$\tilde{S}_n = (1 + 4n^{-4/5})S_n, \quad (15)$$

The size adjustment constant is chosen through intensive simulation with various different values, and this one is recommended. With such a size adjustment, our new test  $\tilde{S}_n$  can better control type I errors. It is worth noting that the size adjustment is asymptotically negligible when  $n \rightarrow \infty$  since  $\tilde{S}_n - S_n$  tends to zero at the rate of order  $n^{-4/5}$  as  $n \rightarrow \infty$ .

### 3.3 Robustness property

In this section, we investigate the local stability of R-DREAM under infinitesimal local contamination through Hampel influence function, which was introduced by Hampel (1974). The Hampel influence analysis reveals that R-DREAM has the desired local robustness property. In the following, we first give von Mises functional expansion of R-DREAM and further derive the Hampel influence function. Compared with Wang and Qu (2007), the key difference is between the used covariate  $X$  and  $\hat{B}(\hat{q})^\top X$  in the respective test statistics, and  $\hat{B}(\hat{q})^\top X$  is automatically 1- and  $q$ -dimensional under the null and alternative hypothesis. Thus, we only give some brief descriptions about the results.

The von Mises analysis (see Fernholz (1983)) can provide the basis of the Hampel influence function calculation. We first discuss the von Mises functional expansion. Let  $Z = B^\top X$ ,  $\hat{Z} = \hat{B}(\hat{q})^\top X$  and denote  $\hat{H}_n$  as the empirical distribution function

of  $Y_1 - g(\hat{\beta}^\top X_1, \hat{\theta}), \dots, Y_n - g(\hat{\beta}^\top X_n, \hat{\theta})$ . When  $H_0$  holds,  $B = c\beta$ , and  $\hat{B}(\hat{q})$  is an estimator of  $\beta$  up to a scalar constant  $c$ . The R-DREAM statistic  $V_n$  in (9) can be asymptotically expressed as

$$T(\hat{F}_h, F_n) =: \int \int \left[ \hat{H}_n(y - g((\beta^\top x)(F_n), \theta(F_n))) - \frac{1}{2} \right] \times \left( \int \left[ \hat{H}_n(y_1 - g((\beta^\top x)(F_n), \theta(F_n))) - \frac{1}{2} \right] \times \hat{f}_h(\hat{z}, y_1) dy_1 \right) dF_n(\hat{z}, y),$$

where  $F_n(\cdot)$  is the empirical distribution function of  $(\hat{Z}_i, Y_i)$ 's,  $((\beta^\top x)(F_n), \theta(F_n))^\top$  is an estimator of  $(\beta^\top x, \theta)^\top$ , which can be rewritten as a functional of the empirical distribution  $F_n$ , and  $\hat{f}_h(z_i, y_i)$  is a smoothing kernel estimation of the joint density function of  $(\hat{Z}, Y)$ , which has the following form:

$$\hat{f}_h(\hat{z}_i, y_i) = \frac{1}{n-1} \sum_{i \neq j}^n \mathcal{K}_{1,h}(y_i - y_j) \mathcal{K}_{2,h}(\hat{z}_i - \hat{z}_j),$$

where  $\mathcal{K}_1(\cdot), \mathcal{K}_2(\cdot)$  are two kernel functions and satisfy  $\mathcal{K}_{1,h}(\cdot) = \mathcal{K}_1(\cdot/h)/h$  and  $\mathcal{K}_{2,h}(\cdot) = \mathcal{K}_2(\cdot/h)/h^{\hat{q}}$ , respectively. An appropriate functional for R-DREAM is bivariate with the form

$$T(F, F) =: \int \int \left[ H(y - g((\beta^\top x)(F), \theta(F))) - \frac{1}{2} \right] \times \left( \int \left[ H(y_1 - g((\beta^\top x)(F), \theta(F))) - \frac{1}{2} \right] \times f(z, y_1) dy_1 \right) dF(z, y), \tag{16}$$

here,  $F(\cdot)$  is the distribution function of  $(Z, Y)$  and  $H(\cdot)$  is the distribution function of  $Y - g(\beta^\top X, \theta)$ :  $H(v) = \int \int_{y \leq g((\beta^\top x)(F), \theta(F)) + v} dF(z, y)$ .

Now we are in the position to analyze Hampel influence function (Hampel 1974). When the second-order influence function in the response direction is bounded, the regression functions is robustly estimated. The detailed derivations for the first-order and the second-order influence function of  $\tilde{V}_n$  in (10) are given in Wang and Qu (2007). Here, we only report the properties for influence functions of  $T(F, F)$  in (16) at the point  $(z_0, y_0)$ , and the detailed proofs are left into Appendix.

**Theorem 3** *Suppose the conditions (C1)–(C8) in Appendix hold, we have*

- (i) *R-DREAM has a degenerate first-order influence function.*
- (ii) *The second-order influence function of R-DREAM is bounded in the response direction.*

For the purpose of comparison, the first-order and second-order influence functions for the test  $T_n^{GWZ}$  in (11) are derived. The first-order influence function is also zero and the second-order influence function are given by (26) in Appendix. Based on this formulation, we can see that the second-order influence function in the  $y$ -direction is not bounded. The above influence function analysis indicates R-DREAM possesses

more stable and robust performance than the test  $T_n^{GWZ}$  when the response is under contamination.

*Remark 5* In order to have the above desired robustness property, a robust method must be used to implement the parameter estimation. In our paper, the popular Tukey bisquare M-estimate method (P.29, Maronna et al. 2006) is used to estimate the parameters. The robust estimate  $(\hat{\beta}^\top, \hat{\theta}^\top)^\top$  is obtained by solving the following minimization problem:

$$(\hat{\beta}^\top, \hat{\theta}^\top) = \arg \min_{\beta, \theta} \sum_{i=1}^n \rho(y_i - g(\beta^\top x_i, \theta)),$$

where

$$\rho(u) = \begin{cases} 1 - [1 - (u/k)^2]^3 & \text{if } |u| \leq k, \\ 1 & \text{if } |u| > k. \end{cases}$$

Here,  $k = 4.685\sigma_\varepsilon$  is the tuning constant and  $\sigma_\varepsilon$  is the standard deviation of the error  $\varepsilon$ . In this way, the  $\sqrt{n}$  consistent estimates of the parameters can be derived. Note that using  $k = 4.685\sigma_\varepsilon$  can achieve the good estimation efficiency. However, the  $\sqrt{n}$  consistency holds for any fixed value in the place of  $\sigma_\varepsilon$  if we do not consider the estimation efficiency. In our case, we can simply use  $k = 4.685$  to avoid estimating the unknown  $\sigma_\varepsilon$ . It is worth mentioning that the other robust estimation methods can also be applicable for estimating the parameters as long as the  $\sqrt{n}$  consistent estimates of the parameters can be obtained.

### 4 Simulation studies

In this section, two simulation studies are conducted to examine the theory and the finite-sample performance of the proposed R-DREAM. The purpose of the simulation studies is twofold: to check the usefulness of dimension reduction strategy to overcome the curse of dimensionality and to check how much it would lose in the cases without outliers when compared to existing non-robust tests.

*Study 1:* Consider the following models

$$\begin{aligned} H_{11} : Y &= \beta_1^\top X + 0.75a(\beta_2^\top X + 1.2)^3 + \varepsilon, \\ H_{12} : Y &= \beta_1^\top X + a \times \left\{ 3.8 \exp(1.2\beta_2^\top X) - 10 \right\} + \varepsilon, \\ H_{13} : Y &= \beta_1^\top X + 1.2a \times \exp(-1.3\beta_1^\top X) + \varepsilon, \end{aligned}$$

where  $\beta_1 = (1, \dots, 1)^\top / \sqrt{p}$ ,  $\beta_2 = (1, \dots, 1, 0, \dots, 0)^\top / \sqrt{p/2}$ ,  $p = 4, 2$  and  $n = 100$ . The covariates  $X = (X_1, \dots, X_p)^\top$  are generated from multivariate normal distributions  $N(0, \Sigma_k)$ ,  $k = 1, 2$  with  $\Sigma_1 = I_p$  and  $\Sigma_2 = \{0.2^{|i-j|}\}_{p \times p}$ . Two kinds of errors are considered:  $\varepsilon \sim N(0, 1)$  and  $\varepsilon \sim t(1)$ . For  $H_{11}$  and  $H_{12}$ , when  $a \neq 0$ , we have  $q = 2$  and  $B = (\beta_1, \beta_2)$ . However, for  $H_{13}$ ,  $q = 1$  and  $B = \beta_1$ . For the

first two alternative models  $H_{11}$  and  $H_{12}$ , 10% of the responses are randomly added by observations from the two following nonlinear models, respectively:

$$\text{Case 1: } Y = 8 \exp\{-|\beta_1^\top X|\} + 10 + \varepsilon,$$

$$\text{Case 2: } Y = 8 \exp\{-|\beta_1^\top X|\} + 48 + \varepsilon.$$

As to the last alternative model  $H_{13}$ , two outlier situations are considered: Case 3: 10% of the responses are randomly added by an outlying value of 5; Case 4: 10% of the responses are randomly added by an outlying value of 15. We set  $a = 0, 0.2, \dots, 1.0$  where  $a = 0$  corresponds to the null hypothesis and  $a \neq 0$  to the alternative hypothesis. We intend to apply these alternative models to examine the effect of dimensionality on the proposed test  $\tilde{S}_n$  and  $T_n^{WQ}$  of Wang and Qu (2007).

Throughout these simulations, unless otherwise specified, the kernel function is taken to be  $\mathcal{K}(u) = 15/16(1 - u^2)^2$  if  $|u| \leq 1$  and 0, otherwise. Our experience in the simulations suggests that  $\tilde{S}_n$  is not very sensitive to the choice of kernel function. The bandwidth is recommended as  $h = 0.5n^{-1/(\hat{q}+4)}$  through intensive numerical computation. The significance level is set to be  $\alpha = 0.05$  and the sample size  $n = 100$  is considered. Every simulation result is the average of 4000 replications.

The simulation results under models  $H_{11}$  and  $H_{12}$  are reported in Table 1 for  $a = 0, 0.2, \dots, 1$  at the significance level  $\alpha = 0.05$ . From this table, we can observe that in all the cases we conduct, the two tests have empirical sizes close to the significance level. Also, the simulated powers of the tests become higher with increase in the parameter  $a$ . It can be seen clearly that the test  $\tilde{S}_n$  works very well in power performance and is not significantly affected by the dimension of  $X$ . However,  $T_n^{WQ}$  severely suffers from the curse of dimensionality. When the dimension of  $X$  gets larger,  $T_n^{WQ}$  completely fails to detect the alternatives. In addition, our proposed test  $\tilde{S}_n$  shows robust performance to the correlated covariates and the heavy tail error distribution  $t(1)$ . Figure 1 reports the simulated power curves under  $H_{13}$  for different values of  $a$  and sample sizes  $n = 100$ . The similar conclusion can be arrived.

*Study 2:* We generate the data from the following model:

$$H_{21} : Y = \beta^\top X + 0.5a(\beta^\top X + 1)^3 + \varepsilon,$$

where  $\beta = (1, \dots, 1)^\top / \sqrt{p}$  and  $p = 8, 12$ . Here,  $X \sim N(0, I_p)$  and  $\varepsilon \sim N(0, 1)$ . In this study, we investigate two issues: One is whether the test  $T_n^{\text{GWZ}}$  can maintain the significance level when there exist some outlier values in the responses and the other is how much  $\tilde{S}_n$  loses in power when there are not outliers.  $\varrho$  is the percentage of which the responses are randomly replaced by the observations from a nonlinear model  $Y = (3\beta^\top X + 5)^2 + \varepsilon$ . Here, we only consider  $T_n^{\text{GWZ}}$  with SIR-based DEE estimate.

Figure 2 presents the empirical size curves or “significance trace” of the three tests for different values of the ratio  $\varrho$  and sample sizes  $n = 100$ . In this case, the ratio  $\varrho = 0, 0.02, \dots, 0.1$ . From this figure, we can see that when  $\varrho = 0$ , all of the three tests can control empirical sizes very well which are all close to the pre-specified significance level 0.05. However, with the increase in the percentage  $\varrho$ , our robust test

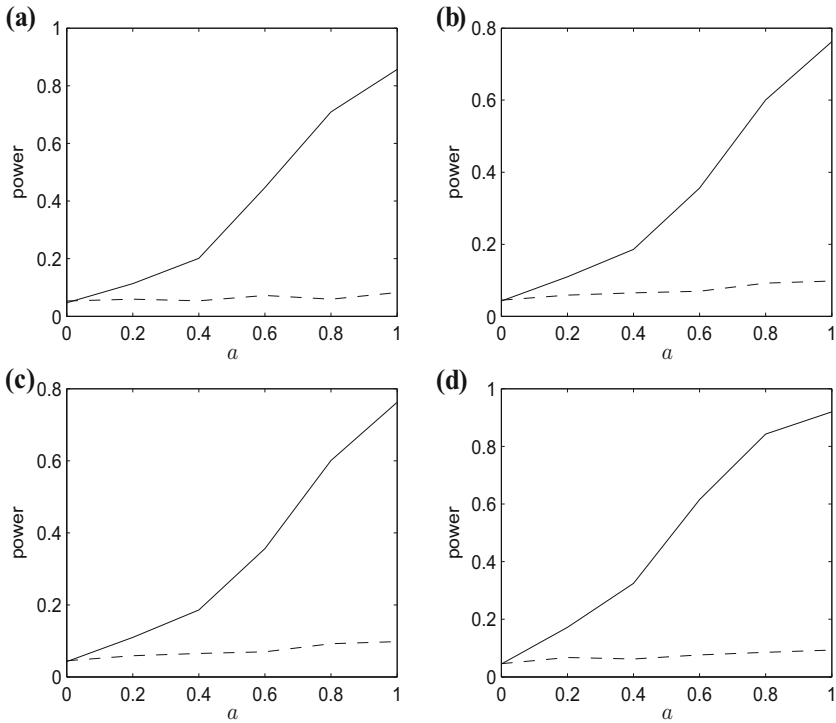
**Table 1** Empirical sizes and powers of  $\tilde{S}_n$  and  $T_n^{WQ}$  for  $H_{10}$  versus  $H_{11}$  and  $H_{12}$  at the significance level  $\alpha = 0.05$  with  $n = 100$

$X, \varepsilon$	$a$	$p = 4$				$p = 2$				
		Case 1		Case 2		Case 1		Case 2		
		$\tilde{S}_n$	$T_n^{WQ}$	$\tilde{S}_n$	$T_n^{WQ}$	$\tilde{S}_n$	$T_n^{WQ}$	$\tilde{S}_n$	$T_n^{WQ}$	
$H_{11}$ $X \sim MVN(0, \Sigma_1)$ $\varepsilon \sim N(0, 1)$	0	0.048	0.055	0.051	0.053	0.048	0.046	0.050	0.048	
	0.2	0.202	0.074	0.194	0.063	0.216	0.142	0.221	0.145	
	0.4	0.581	0.080	0.575	0.078	0.645	0.346	0.654	0.323	
	0.6	0.842	0.093	0.813	0.088	0.889	0.495	0.867	0.503	
	0.8	0.935	0.113	0.908	0.101	0.954	0.641	0.941	0.659	
	1.0	0.962	0.115	0.947	0.106	0.989	0.748	0.976	0.717	
	$X \sim MVN(0, \Sigma_2)$ $\varepsilon \sim t(1)$	0	0.053	0.055	0.055	0.051	0.055	0.044	0.056	0.045
		0.2	0.110	0.057	0.096	0.053	0.116	0.076	0.104	0.065
		0.4	0.255	0.059	0.252	0.055	0.256	0.118	0.252	0.109
		0.6	0.483	0.078	0.442	0.067	0.489	0.194	0.457	0.187
0.8		0.656	0.079	0.574	0.074	0.614	0.276	0.594	0.261	
$H_{12}$ $X \sim MVN(0, \Sigma_1)$ $\varepsilon \sim N(0, 1)$	0	0.049	0.055	0.045	0.055	0.047	0.054	0.045	0.046	
	0.2	0.365	0.072	0.347	0.068	0.415	0.204	0.402	0.193	
	0.4	0.817	0.099	0.785	0.090	0.878	0.466	0.849	0.485	
	0.6	0.941	0.113	0.917	0.094	0.971	0.648	0.966	0.650	
	0.8	0.975	0.110	0.970	0.106	0.994	0.757	0.985	0.742	
	1.0	0.994	0.120	0.983	0.120	0.999	0.813	1.000	0.821	
	$X \sim MVN(0, \Sigma_2)$ $\varepsilon \sim t(1)$	0	0.054	0.055	0.053	0.049	0.047	0.046	0.052	0.046
		0.2	0.154	0.057	0.151	0.055	0.178	0.094	0.177	0.078
		0.4	0.430	0.069	0.396	0.061	0.455	0.177	0.408	0.163
		0.6	0.648	0.073	0.609	0.070	0.678	0.270	0.609	0.265
0.8		0.795	0.075	0.762	0.073	0.828	0.363	0.774	0.349	
1.0	0.889	0.081	0.854	0.077	0.915	0.442	0.855	0.415		

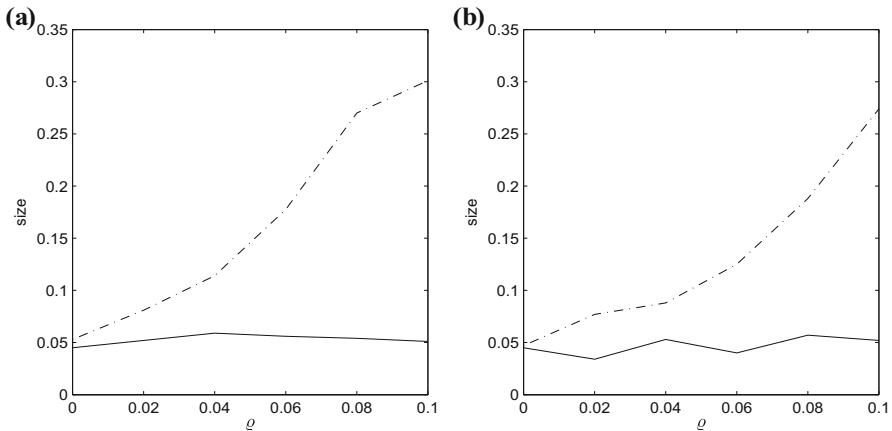
$\tilde{S}_n$  outperforms the test  $T_n^{GWZ}$ . The simulation results indicate that  $\tilde{S}_n$  is not affected by the outlier values and it is more robust, whereas  $T_n^{GWZ}$  fails to work when outliers exist.

We display the simulated power curves for different values of  $a$  in Fig. 3. Here, we consider  $\varrho = 0$  and  $n = 100$ . In other words, there are no outliers in the data. The parameter  $a$  is set to be  $0, 0.2, \dots, 1$ . From the figure, we can see that compared with  $T_n^{GWZ}$ , it is anticipated that  $T_n^{GWZ}$  has higher powers since their test employs more value information, whereas the robust tests only utilize the rank information of the responses. However, the powers of  $\tilde{S}_n$  are still acceptable.

The sensitivity analysis in this study is also considered under the same settings of Study 1 without outliers. Here,  $p = 8, X \sim N(0, \Sigma_2)$  with  $\Sigma_2 = \{0.2^{|i-j|}\}_{p \times p}$  and

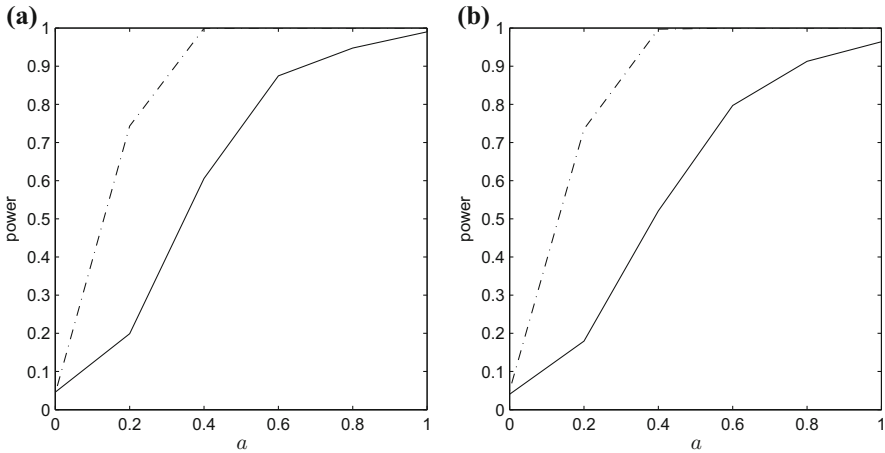


**Fig. 1** Empirical sizes and powers of  $\tilde{S}_n$  and  $T_n^{WQ}$  for  $H_{10}$  versus  $H_{13}$  at the significance level  $\alpha = 0.05$  with  $n = 100$ ,  $p = 4$  and  $\varepsilon \sim N(0, 1)$ . In four plots, the solid line and the dashed line are for  $\tilde{S}_n$  and  $T_n^{WQ}$ , respectively. **a** Powers for Case 3 under  $H_{13}$  with  $X \sim N(0, \Sigma_1)$ , **b** powers for Case 4 under  $H_{13}$  with  $X \sim N(0, \Sigma_2)$ , **c** powers for Case 4 under  $H_{13}$  with  $X \sim N(0, \Sigma_1)$ , **d** powers for Case 4 under  $H_{13}$  with  $X \sim N(0, \Sigma_2)$

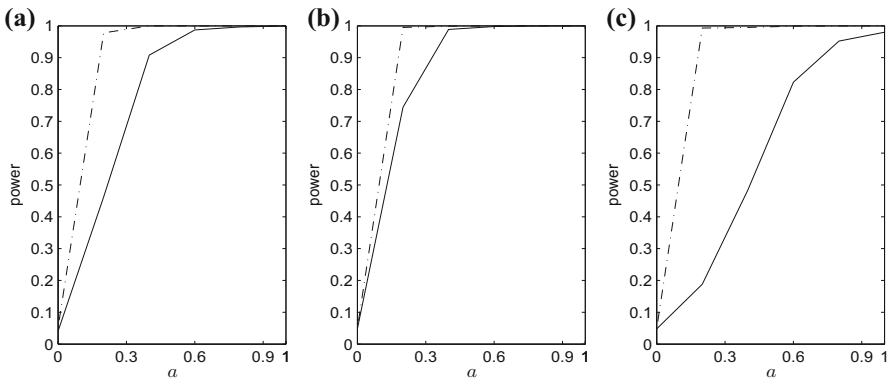


**Fig. 2** Empirical sizes of  $\tilde{S}_n$  and  $T_n^{GWZ}$  for  $H_{20}$  versus  $H_{21}$  at the significance level  $\alpha = 0.05$  with  $X \sim N(0, I_p)$ ,  $\varepsilon \sim N(0, 1)$  and different values of  $\rho$ . In two plots, the solid line and the dash-dotted line are for  $\tilde{S}_n$  and  $T_n^{GWZ}$ , respectively. **a** Empirical sizes for  $H_{21}$  with  $p = 8$ , **b** empirical sizes for  $H_{21}$  with  $p = 12$





**Fig. 3** Simulated powers of  $\tilde{S}_n$  and  $T_n^{GWZ}$  for  $H_{20}$  versus  $H_{21}$  at the significance level  $\alpha = 0.05$  with  $X \sim N(0, I_p)$  and  $\varepsilon \sim N(0, 1)$ . In two plots, the solid line and the dash-dotted line are for  $\tilde{S}_n$  and  $T_n^{GWZ}$ , respectively. **a** Simulated powers for  $H_{21}$  with  $p = 8$ , **b** simulated powers for  $H_{21}$  with  $p = 12$



**Fig. 4** Simulated powers of  $\tilde{S}_n$  and  $T_n^{GWZ}$  for  $H_{10}$  versus  $H_{11}$ ,  $H_{12}$ ,  $H_{13}$  at the significance level  $\alpha = 0.05$  with  $X \sim N(0, \Sigma_2)$  and  $\varepsilon \sim N(0, 1)$ . In three plots, the solid line and the dash-dotted line are for  $\tilde{S}_n$  and  $T_n^{GWZ}$ , respectively. **a** Simulated powers for  $H_{11}$ , **b** simulated powers for  $H_{12}$ , **c** simulated powers for  $H_{13}$

$\varepsilon \sim N(0, 1)$  are considered. Figure 4 reports the simulated results of  $\tilde{S}_n$  and  $T_n^{GWZ}$ . From this figure, we can see that for the three alternatives  $H_{11}$ ,  $H_{12}$  and  $H_{13}$  without outliers, the simulated power curves for  $\tilde{S}_n$  and  $T_n^{GWZ}$  have the common sigmoidal shape. Although the powers for  $T_n^{GWZ}$  are slightly higher, the performance of our proposed test  $\tilde{S}_n$  remains accredited.

### 5 Real data analysis

In this section, we consider the CARS data to illustrate our test which was collected by Ernesto and David in the year of 1982. After that, this dataset was adopted by the

**Table 2** The values of test statistics and the corresponding  $p$ -values for the real data analysis

	Case 5		Case 6	
	Statistics	$p$ -values	Statistics	$p$ -values
$\tilde{S}_n$	4.627	0	5.599	0
$T_n^{\text{GWZ}}$	-0.711	0.762	-0.318	0.625

committee on Statistical Graphics of the American Statistical Association (ASA) in its Second Exposition of Statistical Graphics Technology in 1983 and can be found in <http://lib.stat.cmu.edu/datasets/cars.data>. Recently, Guo et al. (2016) analyzed this dataset with their proposed test and drew the conclusion that it is appropriate to describe this dataset by single-index model rather than simple linear regression model. This dataset includes 406 observations on the following 8 variables: miles per gallon ( $Y$ ), number of cylinders ( $X_1$ ), engine displacement (cu. inches,  $X_2$ ), horsepower ( $X_3$ ), vehicle weight (lbs.  $X_4$ ), time to accelerate from 0 to 60 mph (sec.  $X_5$ ), model year (modulo 100,  $X_6$ ) and origin of car (1. American, 2. European, 3. Japanese). Since the origin of car is classified into three categories, we refer to Xia (2007) to define two new indicator variables, that is, denote  $X_7 = 1$  if a car is from America and 0 otherwise and  $X_8 = 1$  if one is from Europe and 0 otherwise. Here, 392 complete observations are applied to implement our real data analysis. Before analysis, all the covariates are standardized separately and the responses are centered.

It is our interest to test whether the data  $(Y, X)$  can be fitted with linear regression models where  $X = (X_1, \dots, X_8)^\top$ , that is,

$$\begin{aligned} H_0 : Y &= \beta_0^\top X + \varepsilon \text{ for some } \beta_0 \in \mathbb{R}^8, \\ H_1 : Y &= m(B_1^\top X) + \varepsilon \text{ and } m(B_1^\top X) \neq \beta^\top X \text{ for any } \beta \in \mathbb{R}^8 \end{aligned} \quad (17)$$

The same kernel function and bandwidth are adopted as simulation section. For the original data, the value of the test statistic  $T_n^{\text{GWZ}}$  with SIR-based DEE estimate proposed by Guo et al. (2016) is 9.704 and the corresponding  $p$ -value is 0. With our proposed R-DREAM, we can obtain that our test statistic  $\tilde{S}_n$  is 8.2744 and the corresponding  $p$ -value is 0. Both of results indicate that the original real data cannot be fitted by linear regression models.

To investigate the influence of outlier values in the response space, the two following situations are considered: a). Case 5: we artificially add the first 5% and the last 5% responses of this dataset by an outlying value 50; b). Case 6: we artificially add the first 3% responses by the observations from the model:  $Y = \tilde{\beta}^\top X + \eta$ , where  $p = 8$ ,  $\tilde{\beta} = (1, 1, \dots, 1)^\top / \sqrt{p}$  and  $\eta \sim N(0, 1)$ . The values of test statistics and the corresponding  $p$ -values for  $T_n^{\text{GWZ}}$  and  $\tilde{S}_n$  are reported in Table 2. From this table, we can see that for both cases, our robust test  $\tilde{S}_n$  still suggests to reject the null hypothesis  $H_0$ , whereas  $T_n^{\text{GWZ}}$  cannot reject  $H_0$ . The similar conclusion with original data can be made with our test, on the contrary,  $T_n^{\text{GWZ}}$  tends to make a wrong conclusion, which indicates that our test method is more robust.

**Acknowledgements** The authors thank the Editor, the Associate Editor and referees for their constructive comments and suggestions, which led to the substantial improvement in an early manuscript. The research described herewith was supported by the Fundamental Research Funds for the Central Universities, China Postdoctoral Science Foundation (2016M600951), National Natural Science Foundation of China (11701034, 11671042) and a grant from the University Grants Council of Hong Kong.

## Appendix: Proofs of the theorems

The following conditions are required for proving the theorems in Sect. 3.

- (C1) The joint probability density function  $f(z, y)$  of  $(Z, Y)$  is bounded away from 0.
- (C2) The density function  $f(B^\top X)$  of  $B^\top X$  on support  $\mathcal{Z}$  exists and has two bounded derivatives and satisfies

$$0 < \inf f(z) < \sup f(z) < 1.$$

- (C3) The kernel function  $\mathcal{K}(\cdot)$  is a bounded and symmetric density function with compact support and a continuous derivative and all the moments of  $\mathcal{K}(\cdot)$  exist.
- (C4) The bandwidth satisfies  $nh^2 \rightarrow \infty$  under the null (1) and local alternative hypothesis (14);  $nh^q \rightarrow \infty$  under the global alternative hypothesis (2).
- (C5) There exists an estimator  $\hat{\alpha}$  such that under the null hypothesis,  $\sqrt{n}(\hat{\alpha} - \alpha) = O_p(1)$ , where  $\alpha = (\beta, \theta)$  and under the local alternative sequences,  $\sqrt{n}(\hat{\alpha} - \tilde{\alpha}) = O_p(1)$ , where  $\alpha$  and  $\tilde{\alpha}$  are both interior points of  $\Theta$ , a compact and convex set.
- (C6) Denote  $\alpha = (\beta, \theta)^\top$  and there exists a positive continuous function  $G(x)$  such that  $\forall \alpha_1, \alpha_2, |g(x, \alpha_1) - g(x, \alpha_2)| \leq G(x)|\alpha_1 - \alpha_2|$ .
- (C7)  $\mathcal{M}_n(s)$  has the following expansion:

$$\mathcal{M}_n(s) = \mathcal{M}(s) + E_n\{\psi(X, Y, s)\} + R_n(s),$$

where  $E_n(\cdot)$  denotes the average over all sample points,  $E\{\psi(X, Y, s)\} = 0$  and  $E\{\psi^2(X, Y, s)\} < \infty$ .

- (C8)  $\sup_s \|R_n(s)\|_{F=O_p(n^{-1/2})}$ , where  $\|\cdot\|_F$  denotes the Frobenius norm of a matrix.

*Remark 6* Conditions (C1),(C2),(C5) and (C6) are essentially the same as those in Wang and Qu (2007). These conditions are regular for ensuring the consistency and normality of the parameter estimator. Condition (C3) is standard for nonparametric regression estimation. Condition (C4) is the same as that in Guo et al. (2016). This condition on bandwidth is required to ensure the asymptotic normality of our test statistic. Conditions (C7) and (C8) are for the DEE estimation. Under the linearity condition, SIR-based DEE satisfies conditions (C7) and (C8). These conditions are quite mild and can be satisfied in many practical situations. It is worth mentioning that the second moment condition is not strong here because in DEE, we use the indicator function of  $Y$ , rather than the original  $Y$  and the target matrix only involves the square of the expectation of the product of  $X$  and the indicator of  $Y$ . Thus, these conditions only involve the second moment of  $X$ .

The following lemmas are used to prove the theorems in Sect. 3. We first give the proof of Lemma 1 in Sect. 2.

*Proof of Lemma 1* In the following, we give the proof of the DEE-based estimate and the same conditions in Theorem 4 of Zhu et al. (2010) are adopted.

Under the conditions designed by in Zhu et al. (2010), their Theorem 2 ensures that  $\mathcal{M}_{n,n} - \mathcal{M} = O_p(n^{-1/2})$ . Further, the root-n consistency of the eigenvalues of  $\mathcal{M}_{n,n}$  is retained, that is,  $\hat{\lambda}_i - \lambda_i = O_p(n^{-1/2})$ . Note that when  $l \leq q$ ,  $\lambda_l > 0$  and for  $l > q$ , we have  $\lambda_l = 0$ . As recommended,  $c = 1/\sqrt{nh}$  is used. when  $nh \rightarrow \infty, h \rightarrow 0$ , we have  $1/\sqrt{n} = o(c)$  and  $c = o(1)$ . For  $1 \leq l < q$ ,

$$\begin{aligned} \frac{\hat{\lambda}_{q+1} + c}{\hat{\lambda}_q + c} - \frac{\hat{\lambda}_{l+1} + c}{\hat{\lambda}_l + c} &= \frac{\lambda_{q+1} + c + O_p\left(\frac{1}{\sqrt{n}}\right)}{\lambda_q + c + O_p\left(\frac{1}{\sqrt{n}}\right)} - \frac{\lambda_{l+1} + c + O_p\left(\frac{1}{\sqrt{n}}\right)}{\lambda_l + c + O_p\left(\frac{1}{\sqrt{n}}\right)} \\ &= \frac{c + O_p\left(\frac{1}{\sqrt{n}}\right)}{\lambda_q + c + O_p\left(\frac{1}{\sqrt{n}}\right)} - \frac{\lambda_{l+1} + c + O_p\left(\frac{1}{\sqrt{n}}\right)}{\lambda_l + c + O_p\left(\frac{1}{\sqrt{n}}\right)} \Rightarrow -\frac{\lambda_{l+1}}{\lambda_l} < 0. \end{aligned}$$

When  $l > q$ ,

$$\begin{aligned} \frac{\hat{\lambda}_{q+1} + c}{\hat{\lambda}_q + c} - \frac{\hat{\lambda}_{l+1} + c}{\hat{\lambda}_l + c} &= \frac{\lambda_{q+1} + c + O_p\left(\frac{1}{\sqrt{n}}\right)}{\lambda_q + c + O_p\left(\frac{1}{\sqrt{n}}\right)} - \frac{\lambda_{l+1} + c + O_p\left(\frac{1}{\sqrt{n}}\right)}{\lambda_l + c + O_p\left(\frac{1}{\sqrt{n}}\right)} \\ &= \frac{c + O_p\left(\frac{1}{\sqrt{n}}\right)}{\lambda_q + c + O_p\left(\frac{1}{\sqrt{n}}\right)} - \frac{c + O_p\left(\frac{1}{\sqrt{n}}\right)}{c + O_p\left(\frac{1}{\sqrt{n}}\right)} \Rightarrow -1 < 0. \end{aligned}$$

Therefore, we can conclude that under the null hypothesis  $H_0$  and under the fixed alternative hypothesis (2),  $\hat{q} = q$  with a probability going to one as  $n \rightarrow \infty$ . The proof is concluded. □

The proof of Lemma 2 in Sect. 3 is given as follows.

*Proof of Lemma 2* From the proof of Theorem 2 in Guo et al. (2016), it is shown that under the local alternative hypothesis,  $\mathcal{M}_{n,n} - \mathcal{M} = O_p(C_n)$ . Further, we can get  $\hat{\lambda}_i - \lambda_i = O_p(C_n)$ .

Thus, note that  $\lambda_1 > 0$  and for any  $l > 1$ , we have  $\lambda_l = 0$ . Consequently, under the condition that  $C_n = o(c)$  and  $c = o(1)$ ,

$$\begin{aligned} \frac{\hat{\lambda}_2 + c}{\hat{\lambda}_1 + c} - \frac{\hat{\lambda}_{l+1} + c}{\hat{\lambda}_l + c} &= \frac{\lambda_2 + c + O_p(C_n)}{\lambda_1 + c + O_p(C_n)} - \frac{\lambda_{l+1} + c + O_p(C_n)}{\lambda_l + c + O_p(C_n)} \\ &= \frac{c + O_p(C_n)}{\lambda_1 + c + O_p(C_n)} - \frac{c + O_p(C_n)}{c + O_p(C_n)} \Rightarrow -1 < 0. \end{aligned}$$

Thus, under the local alternative (14), Lemma 1 holds. The proof of Lemma 2 is finished. □

**Lemma 3** *Given Conditions (C1)–(C8) in Appendix, we have*

$$nh^{1/2} (V_n - V_n^*) \xrightarrow{P} 0,$$

where  $\xrightarrow{P}$  represents convergence in probability and

$$V_n^* = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}_h \left\{ \hat{B}(\hat{q})^\top (x_i - x_j) \right\} \left[ H(e_i) - \frac{1}{2} \right] \left[ H(e_j) - \frac{1}{2} \right], \tag{18}$$

here,  $\mathcal{K}_h(\cdot) = \mathcal{K}(\cdot/h)/h^{\hat{q}}$ .

*Proof of Lemma 3* We first decompose  $V_n - V_n^*$  as

$$\begin{aligned} V_n - V_n^* &= \frac{1}{n^3(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l=1}^n \sum_{k=1}^n \mathcal{K}_h \left\{ \hat{B}(\hat{q})^\top (x_i - x_j) \right\} \\ &\quad \times \left[ I(\hat{e}_l \leq \hat{e}_i) - I(e_l \leq e_i) \right] \left[ I(\hat{e}_k \leq \hat{e}_j) - I(e_k \leq e_j) \right] \\ &\quad + \frac{2}{n^3(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l=1}^n \sum_{k=1}^n \mathcal{K}_h \left\{ \hat{B}(\hat{q})^\top (x_i - x_j) \right\} \\ &\quad \times \left[ I(\hat{e}_l \leq \hat{e}_i) - I(e_l \leq e_i) \right] \left[ I(e_k \leq e_j) - \frac{n+1}{2n} \right] \\ &\quad + \frac{1}{n^3(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l=1}^n \sum_{k=1}^n \mathcal{K}_h \left\{ \hat{B}(\hat{q})^\top (x_i - x_j) \right\} \\ &\quad \times \left\{ \left[ I(e_l \leq e_i) - \frac{n+1}{2n} \right] \left[ I(e_k \leq e_j) - \frac{n+1}{2n} \right] \right. \\ &\quad \left. - \left[ H(e_i) - \frac{1}{2} \right] \left[ H(e_j) - \frac{1}{2} \right] \right\} \\ &=: A_1 + A_2 + A_3. \end{aligned} \tag{19}$$

Since  $e_i = y_i - g(\beta^\top x_i, \theta)$ , we further have  $\hat{e}_i - e_i = g(\beta^\top x_i, \theta) - g(\hat{\beta}^\top x_i, \hat{\theta})$ . Let  $\alpha = (\beta, \theta)^\top$  and  $L(\hat{\alpha})$  as

$$L(\hat{\alpha}) = \max_{1 \leq i \leq n} \sum_{l=1, l \neq i}^n |I(\hat{e}_l \leq \hat{e}_i) - I(e_l \leq e_i)|.$$

Denote  $\Omega = \{\alpha^* : \sqrt{n}|\alpha^* - \alpha_0| \leq \delta\}$  for  $\delta = O(1)$  and  $t(x_i, x_l, \alpha, \alpha^*) = [g(\beta^\top x_i, \theta) - g(\beta^{*\top} x_i, \theta^*)] - [g(\beta^\top x_l, \theta) - g(\beta^{*\top} x_l, \theta^*)]$ , then

$$\begin{aligned}
 \sup_{\alpha^* \in \Omega} |L(\alpha^*)| &= \sup_{\alpha^* \in \Omega} \sum_{l=1, l \neq i}^n |I(e_l - e_i \leq t(x_i, x_l, \alpha, \alpha^*)) - I(e_l \leq e_i)| \\
 &\leq \sup_{\alpha^* \in \Omega} \sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq |t(x_i, x_l, \alpha, \alpha^*)|) \\
 &\leq \sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq Cn^{-1/2}), \tag{20}
 \end{aligned}$$

where  $C$  is a generic positive constant. From the above derivation, we can obtain that, conditional on  $e_i$ ,  $I(|e_l - e_i| \leq Cn^{-1/2}), l \neq i$  are iid Bernoulli random variables with  $O(n^{-1/2})$ -order success probability. Using Bernstein's inequality leads to  $P\{\sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq Cn^{-1/2}) \geq Cn^{1/2}|e_i\} \leq \exp(-Cn^{1/2})$ . Unconditionally, we still have that

$$P \left\{ \sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq Cn^{-1/2}) \geq Cn^{1/2} \right\} \leq \exp(-Cn^{1/2}).$$

Further,

$$P \left\{ \max_{1 \leq i \leq n} \sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq Cn^{-1/2}) \geq Cn^{1/2} \right\} \leq n \exp(-Cn^{1/2}).$$

When  $n \rightarrow \infty$ , for a positive constant  $C$ ,  $n \exp(-Cn^{1/2}) \rightarrow 0$ . Therefore,  $P\{\max_{1 \leq i \leq n} \sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq Cn^{-1/2}) < Cn^{1/2}\} = 1$ . Then,

$$\max_{1 \leq i \leq n} \sum_{l=1, l \neq i}^n I(|e_l - e_i| \leq Cn^{-1/2}) < Cn^{1/2} = O_p(n^{1/2}). \tag{21}$$

Combining (20) and (21), we can gain the following useful probability bound

$$L(\hat{\alpha}) = \max_{1 \leq i \leq n} \sum_{l=1, l \neq i}^n |I(\hat{e}_l \leq \hat{e}_i) - I(e_l \leq e_i)| = O_p(n^{1/2}).$$

An application of the formula (21) and for the term  $A_1$  in (19), we have

$$\begin{aligned}
 nh^{1/2}|A_1| &\leq \frac{h^{1/2}}{n^2(n-1)h^{\hat{q}}} \max_{1 \leq i \leq n} \sum_{l=1, l \neq i}^n |I(\hat{e}_l \leq \hat{e}_i) - I(e_l \leq e_i)| \\
 &\quad \times \max_{1 \leq j \leq n} \sum_{k=1, k \neq j}^n |I(\hat{e}_k \leq \hat{e}_j) - I(e_k \leq e_j)|
 \end{aligned}$$

$$\begin{aligned}
 & \times \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K} \left\{ \hat{B}(\hat{q})^\top (x_i - x_j) / h \right\} \\
 &= \frac{Ch^{1/2}}{n(n-1)h^{\hat{q}}} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K} \left\{ \hat{B}(\hat{q})^\top (x_i - x_j) / h \right\} \\
 &= \frac{Ch^{1/2}}{n(n-1)h^{\hat{q}}} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K} \left\{ B^\top (x_i - x_j) / h \right\} \\
 & \quad + \frac{Ch^{1/2}}{n(n-1)h^{\hat{q}}} \sum_{i=1}^n \sum_{j \neq i}^n \left[ \mathcal{K} \left\{ \hat{B}(\hat{q})^\top (x_i - x_j) / h \right\} - \mathcal{K} \left\{ B^\top (x_i - x_j) / h \right\} \right] \\
 &=: C_1(A_{11} + A_{12}), \tag{22}
 \end{aligned}$$

where  $C_1$  is a positive constant. Let  $Z = B^\top X$ . As to the term  $A_{11}$ , the term

$$\frac{1}{n(n-1)h^{\hat{q}}} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K} \left\{ B^\top (x_i - x_j) / h \right\}$$

is an U-statistic with the kernel as  $H_n(z_1, z_2) = h^{-\hat{q}} \mathcal{K}\{(z_1 - z_2)/h\}$ . In order to apply the theory for non-degenerate U-statistic (Serfling 1980),  $E[H_n(z_1, z_2)^2] = o(n)$  is needed. It can be verified that

$$\begin{aligned}
 E[H_n(z_1, z_2)^2] &= E \left\{ E \left[ H_n(z_1, z_2)^2 | z_1, z_2 \right] \right\} \\
 &= \int \frac{1}{h^{2\hat{q}}} \mathcal{K}^2 \left( \frac{z_1 - z_2}{h} \right) p(z_1) p(z_2) dz_1 dz_2 \\
 &= \int \frac{1}{h^{2\hat{q}}} \mathcal{K}^2(u) p(z_1) p(z_1 - hu) \left( -h^{\hat{q}} \right) dz_1 du \\
 &= -\frac{1}{h^{\hat{q}}} \int \mathcal{K}^2(u) p^2(z_1) dz_1 du + o(1) \\
 &= O \left( \frac{1}{h^{\hat{q}}} \right), \tag{23}
 \end{aligned}$$

where  $p(\cdot)$  is denoted as the probability density function. With the condition  $nh^{\hat{q}} \rightarrow \infty$ , we have  $E[H_n(z_1, z_2)^2] = O(1/h^{\hat{q}}) = o(n)$ . The condition of Lemma 3.1 of Zheng (1996) is satisfied, and we have  $A_{11} = h^{1/2} E[H_n(z_1, z_2)] + o_p(1)$ , where  $E[H_n(z_1, z_2)] = O(1)$ . Therefore, we can obtain that  $A_{11} = O_p(h^{1/2}) = o_p(1)$ . Denote

$$A_{12}^* = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n h^{1/2-\hat{q}} \mathcal{K}' \left\{ \tilde{B}^\top (x_i - x_j) / h \right\} (x_i - x_j)^\top \times \frac{\hat{B}(\hat{q}) - B}{h},$$

where  $\tilde{B}$  lies between  $B$  and  $\hat{B}$ . Then, for the term  $A_{12}$  in (22), we have

$$A_{12} = A_{12}^* + o_p(A_{12}^*).$$

Similar to  $A_{11}$ , the following term

$$\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n h^{-\hat{q}} \mathcal{K}' \left\{ \tilde{B}^\top (x_i - x_j) / h \right\} (x_i - x_j)^\top$$

can be regarded as an U-statistic. It can be similarly shown that the term is the order of  $O_p(h)$ . As  $\|\hat{B}(\hat{q}) - B\|_2 = O_p(1/\sqrt{n})$  and under the condition  $n \rightarrow \infty, h \rightarrow 0$ , we can obtain that  $A_{12} = o_p(1)$ . Toward (22), we have

$$nh^{1/2}|A_1| \leq o_p(1).$$

Similarly, we can derive that  $nh^{1/2}|A_i| = o_p(1)$  for  $i = 2, 3$ . Combining with the formula (19), it can be concluded that

$$nh^{1/2}(V_n - V_n^*) \xrightarrow{P} 0,$$

which completes the proof of Lemma 3. □

In the following, we give the proof of Theorem 1.

*Proof of Theorem 1* From Lemma 3, we know that the limiting distributions for  $nh^{1/2}V_n$  and  $nh^{1/2}V_n^*$  are the same. Thus, we just need to derive the asymptotic property of  $nh^{1/2}V_n^*$ . The term  $V_n^*$  in (18) can be decomposed as

$$\begin{aligned} V_n^* &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}_h \left\{ B^\top (x_i - x_j) \right\} \left[ H(e_i) - \frac{1}{2} \right] \left[ H(e_j) - \frac{1}{2} \right] \\ &\quad + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \left[ H(e_i) - \frac{1}{2} \right] \left[ H(e_j) - \frac{1}{2} \right] \\ &\quad \left[ \mathcal{K}_h \{ \hat{B}(\hat{q})^\top (x_i - x_j) \} - \mathcal{K}_h \{ B^\top (x_i - x_j) \} \right] \\ &=: V_{n1}^* + V_{n2}^*, \end{aligned}$$

where  $\mathcal{K}_h(\cdot) = \mathcal{K}(\cdot/h)/h^{\hat{q}}$ .

For the term  $V_{n1}^*$ , it is an U-statistic, since we always assume that the dimension of  $B^\top X$  is fixed in our paper. Under the null hypothesis,  $H(e_i), i = 1, \dots, n$  follows a uniform distribution on  $(0, 1), q = 1$  and  $\hat{q} \rightarrow 1$ . An application of Theorem 1 in Zheng (1996), it is not difficult to derive the asymptotic normality:  $nh^{1/2}V_{n1}^* \Rightarrow N(0, Var)$ , where

$$Var = \frac{1}{72} \int \mathcal{K}^2(u) du \int p^2(z) dz$$



with  $Z = B^\top X$ .

Denote

$$\begin{aligned} \tilde{V}_{n2}^* &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K}' \left\{ \frac{\tilde{B}^\top (x_i - x_j)}{h} \right\} (x_i - x_j)^\top \\ &\quad \left[ H(e_i) - \frac{1}{2} \right] \left[ H(e_j) - \frac{1}{2} \right] \cdot \frac{\hat{B}(\hat{q}) - B}{h}, \end{aligned}$$

where  $\tilde{B}$  lies between  $\hat{B}$  and  $B$ . By an application of Taylor expansion yields

$$V_{n2}^* = \tilde{V}_{n2}^* + o_p \left( \tilde{V}_{n2}^* \right).$$

Because the kernel  $\mathcal{K}(\cdot)$  is spherical symmetric, the following term can be considered as an U-statistic:

$$\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K}' \left\{ \tilde{B}^\top (x_i - x_j) / h \right\} (x_i - x_j)^\top \left[ H(e_i) - \frac{1}{2} \right] \left[ H(e_j) - \frac{1}{2} \right].$$

Further note that

$$\begin{aligned} &E \left\{ \frac{1}{h^{\hat{q}}} \mathcal{K}' \{ \tilde{B}^\top (x_i - x_j) / h \} (x_i - x_j)^\top \left[ H(e_i) - \frac{1}{2} \right] \left[ H(e_j) - \frac{1}{2} \right] | x_i, y_i \right\} \\ &= E \left[ E \left\{ \frac{1}{h^{\hat{q}}} \mathcal{K}' \{ \tilde{B}^\top (x_i - x_j) / h \} (x_i - x_j)^\top \left[ H(e_i) - \frac{1}{2} \right] \left[ H(e_j) - \frac{1}{2} \right] | x_i, y_i, x_j \right\} | x_i, y_i \right] \\ &= E \left\{ \frac{1}{h^{\hat{q}}} \mathcal{K}' \{ \tilde{B}^\top (x_i - x_j) / h \} (x_i - x_j)^\top \left[ H(e_i) - \frac{1}{2} \right] \cdot E \left[ H(e_j) - \frac{1}{2} | x_j \right] | x_i, y_i \right\} = 0. \end{aligned}$$

Thus, the above U-statistic is degenerate. Similar as the derivation of  $V_{n1}^*$ , together with  $\|\hat{B}(\hat{q}) - B\|_2 = O_p(1/\sqrt{n})$  and  $1/nh^2 \rightarrow 0$ , we have  $nh^{1/2}V_{n2}^* = o_p(1)$ . Therefore, under the null hypothesis  $H_0$ , we can conclude that  $nh^{1/2}V_n^* \Rightarrow N(0, Var)$ . Based on Lemma 3, we have  $nh^{1/2}V_n \Rightarrow N(0, Var)$ .

An estimate of  $Var$  can be defined as

$$\widehat{Var} = \frac{1}{72n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K}^2 \left\{ \frac{\hat{B}(\hat{q})^\top (x_i - x_j)}{h} \right\}.$$

Since the proof is rather straightforward, we then only give a brief description. Using a similar argument as that for Lemma 3, we can get

$$\widehat{Var} = \frac{1}{72n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K}^2 \left\{ \frac{B^\top (x_i - x_j)}{h} \right\} + o_p(1).$$

The consistency can be derived through U-statistic theory. The proof for Theorem 1 is finished. □

The proof for Theorem 2 is given as follows.

*Proof of Theorem 2* Under the global alternative  $H_n$  in (2), we have  $e_i = m(B^\top x_i) + \varepsilon_i - g(\tilde{\beta}^\top x_i, \tilde{\theta})$ . Together with Lemma 3, it can be obtained that  $V_n = V_n^* + o_p(1)$ , where  $V_n^*$  can be rewritten as

$$\begin{aligned} V_n^* &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{\hat{B}(\hat{q})^\top (x_i - x_j)}{h} \right\} \left[ H(m(B^\top x_i) + \varepsilon_i - g(\tilde{\beta}^\top x_i, \tilde{\theta})) - \frac{1}{2} \right] \\ &\quad \left[ H(m(B^\top x_j) + \varepsilon_j - g(\tilde{\beta}^\top x_j, \tilde{\theta})) - \frac{1}{2} \right] \\ &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{B^\top (x_i - x_j)}{h} \right\} \left[ H(m(B^\top x_i) + \varepsilon_i - g(\tilde{\beta}^\top x_i, \tilde{\theta})) - \frac{1}{2} \right] \\ &\quad \left[ H(m(B^\top x_j) + \varepsilon_j - g(\tilde{\beta}^\top x_j, \tilde{\theta})) - \frac{1}{2} \right] + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \left[ H(m(B^\top x_i) + \varepsilon_i \right. \\ &\quad \left. - g(\tilde{\beta}^\top x_i, \tilde{\theta})) - \frac{1}{2} \right] \left[ H(m(B^\top x_j) + \varepsilon_j - g(\tilde{\beta}^\top x_j, \tilde{\theta})) - \frac{1}{2} \right] \left[ \mathcal{K}_h \{ \hat{B}(\hat{q})^\top (x_i - x_j) \} \right. \\ &\quad \left. - \mathcal{K}_h \{ B^\top (x_i - x_j) \} \right] \\ &= V_{n3}^* + V_{n4}^*, \end{aligned}$$

For the term  $V_{n3}^*$ , it is a standard U-statistic with

$$H_n(x_i, x_j) = \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{B^\top (x_i - x_j)}{h} \right\} l(x_i)l(x_j),$$

where  $l(x) = [H\{m(B^\top x) + \varepsilon - g(\tilde{\beta}^\top x, \tilde{\theta})\} - 1/2]$ . Similar to the proof of (23), when  $nh^{\hat{q}} \rightarrow \infty$ , we can derive that  $E[H^2(x_i, x_j)] = o(n)$  and the condition of Lemma 3.1 in Zheng (1996) can be shown to be satisfied. We further calculate

$$\begin{aligned} E[H_n(x_i, x_j)] &= E\{E[H_n(x_i, x_j)|x_i, x_j]\} \\ &= \frac{1}{h^{\hat{q}}} \int \mathcal{K} \left\{ \frac{z_i - z_j}{h} \right\} \tilde{l}(z_i)\tilde{l}(z_j)p(z_i)p(z_j)dz_i dz_j \\ &= \frac{1}{h^{\hat{q}}} \int \mathcal{K}(u)\tilde{l}(z_j + hu)\tilde{l}(z_j)p(z_j + hu)p(z_j) \times h^{\hat{q}} du dz_j \\ &= \int \tilde{l}^2(z_j)p^2(z_j)dz_j + o(1) \\ &= E[l^2(X)^2 p(X)] + o(1). \end{aligned}$$

where  $Z = B^\top X$ . Therefore,  $V_{n3}^* = E[l^2(X)^2 p(X)] + o_p(1) =: C_2$ , here,  $C_2$  is a positive constant.

As to the term  $V_{n4}^*$ , similarly as the term  $V_{n2}^*$ , we have

$$V_{n4}^* = \tilde{V}_{n4}^* + o_p(\tilde{V}_{n4}^*),$$

where,

$$\tilde{V}_{n4}^* = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K}' \left\{ \frac{\tilde{B}^\top(x_i - x_j)}{h} \right\} (x_i - x_j)^\top l(x_i) l(x_j) \cdot \frac{\hat{B}(\hat{q}) - B}{h},$$

here,  $\tilde{B}$  lies between  $B$  and  $\hat{B}$ . Similarly as the derivation of  $V_{n3}^*$ , together with  $\|\hat{B}(\hat{q}) - B\|_2 = O_p(1/\sqrt{n})$ , when  $nh^{\hat{q}} \rightarrow \infty$ , we have  $V_{n4}^* = O_p(h) \cdot O_p(1/\sqrt{n}) \cdot (1/h) = o_p(1)$ .

Based on the above analysis, we can derive that  $V_n = C_2 + o_p(1)$  and  $nh^{1/2}V_n \Rightarrow \infty$  in probability, which completes the proof of the global alternative situation.

We now consider the situation of local alternative  $H_{1n}$  in (14). Based on Lemma 3, we have  $V_n = V_n^* + o_p(1)$ . In this situation,  $e_i = C_n m(B^\top x_i) + \varepsilon_i$ . Therefore,  $V_n^*$  can be decomposed as

$$\begin{aligned} V_n^* &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{\hat{B}(\hat{q})^\top(x_i - x_j)}{h} \right\} \left[ H(C_n m(B^\top x_i) + \varepsilon_i) - \frac{1}{2} \right] \\ &\quad \left[ H(C_n m(B^\top x_j) + \varepsilon_j) - \frac{1}{2} \right] \\ &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{B^\top(x_i - x_j)}{h} \right\} \left[ H(C_n m(B^\top x_i) + \varepsilon_i) - \frac{1}{2} \right] \\ &\quad \left[ H(C_n m(B^\top x_j) + \varepsilon_j) - \frac{1}{2} \right] + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \left[ H(C_n m(B^\top x_i) + \varepsilon_i) \right. \\ &\quad \left. - \frac{1}{2} \right] \left[ H(C_n m(B^\top x_j) + \varepsilon_j) - \frac{1}{2} \right] \left[ \mathcal{K}_h\{\hat{B}(\hat{q})^\top(x_i - x_j)\} - \mathcal{K}_h\{B^\top(x_i - x_j)\} \right] \\ &= V_{n5}^* + V_{n6}^*. \end{aligned} \tag{24}$$

For the term  $V_{n5}^*$ , taking a Taylor expansion of  $H(C_n m(B^\top x_i) + \varepsilon_i)$  around  $C_n = 0$ , we have

$$H(C_n m(B^\top x_i) + \varepsilon_i) = H(\varepsilon_i) + C_n h(\varepsilon_i) m(B^\top x_i) + o_p(C_n^2).$$

Then, the term  $V_{n5}^*$  can be decomposed as

$$\begin{aligned} V_{n5}^* &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{B^\top(x_i - x_j)}{h} \right\} \left[ H(\varepsilon_i) - \frac{1}{2} \right] \left[ H(\varepsilon_j) - \frac{1}{2} \right] \\ &\quad + 2C_n \left\{ \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{B^\top(x_i - x_j)}{h} \right\} \left[ H(\varepsilon_i) - \frac{1}{2} \right] h(\varepsilon_j) m(B^\top x_j) \right\} \\ &\quad + C_n^2 \left\{ \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} \mathcal{K} \left\{ \frac{B^\top(x_i - x_j)}{h} \right\} h(\varepsilon_i) m(B^\top x_i) h(\varepsilon_j) m(B^\top x_j) \right\} \\ &\quad + o_p(C_n^2) \\ &= D_1 + 2C_n D_2 + C_n^2 D_3 + o_p(C_n^2). \end{aligned}$$

Under the local alternative hypothesis,  $\hat{q} \rightarrow 1$  can be obtained. For the term  $D_1$ , similarly to the proof of the term  $V_{n1}^*$  in Theorem 1, we can show that  $nh^{1/2}D_1 \Rightarrow N(0, Var)$ , where

$$Var = \frac{1}{72} \int \mathcal{K}^2(u) du \int p^2(z) dz$$

with  $Z = B^\top X$ . As to the term  $D_2$ , similarly as the proof of Lemma 3.3b in Zheng (1996), it can be obtained that  $D_2 = O_p(1/\sqrt{n})$ . When  $C_n = n^{-1/2}h^{-1/4}$ , we have  $nh^{1/2}C_n D_2 = O_p(h^{1/2})$ . Turn to the term  $D_3$ , similarly to the proof of  $V_{n3}^*$  in our Theorem 2, we have  $D_3 = E[h^2(\varepsilon)m^2(B^\top X)p(X)] + o_p(1)$ . Further,  $nh^{1/2}C_n^2 D_3 = E[h^2(\varepsilon)m^2(B^\top X)p(X)] + o_p(1)$ . Therefore,

$$nh^{1/2}V_{n5}^* \Rightarrow N(\mu, Var),$$

where  $\mu = E[h^2(\varepsilon)m^2(B^\top X)p(X)]$ .

As to the term  $V_{n6}^*$ , just similarly as the proof of the term  $V_{n2}^*$  in our Theorem 1, it can be gotten that  $nh^{1/2}V_{n6}^* = o_p(1)$ .

Combining Lemma 3 and the formula (24), under the local alternative, we have  $nh^{1/2}V_n \Rightarrow N(\mu, Var)$ .

The proof of Theorem 2 is finished. □

The proof of Theorem 3 is as follows.

*Proof of Theorem 3* We first calculate the first-order and second-order influence function of  $T(F, F)$  in (16) at the point  $(z_0, y_0)$ . When  $H_0$  holds, the Hampel’s first-order influence function of  $T(F, F)$  in (16) at the point  $(z_0, y_0)$  is defined as

$$IF^{(1)}(z_0, y_0; T) = \lim_{t \rightarrow 0} \frac{T(F_t, F_t) - T(F, F)}{t},$$

where  $T(F, F) = 0$  and  $F_t = (1 - t)F + t\Delta_{z_0, y_0}$ , here,  $\Delta_{z_0, y_0}$  is the point mass function at the point  $(z_0, y_0)$ . Denote  $L(t) = T(F_t, F_t) = T(F + tU, F + tU)$ ,

where  $U = \Delta_{z_0, y_0} - F$ . Thus, we have  $L(0) = 0$ . From the proof in Appendix, it is not difficult to obtain that  $\frac{dL(t)}{dt}|_{t=0} = 0$ . Therefore, R-DREAM has a degenerate first-order influence function.

To obtain the second-order influence function of R-DREAM, we first compute

$$\begin{aligned} & \frac{1}{2} \frac{d^2}{dt^2} L(t)|_{t=0} \\ &= \int \left( \int \dot{H}(y - g((\beta^\top x)(F), \theta(F))) f(z, y) dy \right)^2 dz \\ &+ \int \int \dot{H}(y - g((\beta^\top x)(F), \theta(F))) \times \int \left[ H(y_1 - g((\beta^\top x)(F), \theta(F))) - \frac{1}{2} \right] \\ &\times u(z, y_1) dy_1 dF(z, y) + \int \int \left[ H(y - g(\beta^\top x(F), \theta(F))) - \frac{1}{2} \right] \\ &\times \int \dot{H}(y_1 - g((\beta^\top x)(F), \theta(F))) \times f(z, y_1) dy_1 dU(z, y) \\ &+ \int \int \left[ H(y - g((\beta^\top x)(F), \theta(F))) - \frac{1}{2} \right] \times \left( \int \left[ H(y_1 - g((\beta^\top x)(F), \theta(F))) \right. \right. \\ &\left. \left. - \frac{1}{2} \right] u(z, y_1) dy_1 \right) dU(z, y), \end{aligned} \tag{25}$$

where  $\dot{H}(\cdot) =: \frac{dH_t(\cdot)}{dt}|_{t=0}$  and  $H_t(\cdot)$  represents  $H(\cdot)$  under contamination, that is,  $H_t(v) = \int \int_{y \leq g(\beta^\top x(F_t), \theta(F_t)) + v} dF_t(z, y)$ . Besides,  $u(z, y)$  is the probability density function of  $U(z, y)$ . Taking  $U = \Delta_{z_0, y_0} - F$  into the formula (25), it is shown that the four terms in (25) converge at the same rate. Further, based on Hampel’s definition, we can obtain the second-order influence function of R-DREAM at the point  $(z_0, y_0)$  as follows:

$$\begin{aligned} \text{IF}^{(2)}(z_0, y_0) &= \int \left( \int \dot{H}_{\Delta(z_0, y_0)}(y - g((\beta^\top x)(F), \theta(F))) f(z, y) dy \right)^2 dz \\ &+ \left[ H(y_0 - g((\beta^\top x_0)(F), \theta(F))) - \frac{1}{2} \right] \times \\ &\int \int \dot{H}_{\Delta(z_0, y_0)}(y - g((\beta^\top x)(F), \theta(F))) dF(z, y) \\ &+ \left[ H(y_0 - g((\beta^\top x)_0(F), \theta(F))) - \frac{1}{2} \right] \\ &\times \int \dot{H}_{\Delta(z_0, y_0)}(y - g(\beta^\top x_0(F), \theta(F))) \\ &\times f(z_0, y) dy + \left[ H(y_0 - g((\beta^\top x)_0(F), \theta(F))) - \frac{1}{2} \right]^2, \end{aligned}$$

where  $z_0 = \beta^\top x_0$  and  $\dot{H}_{\Delta(z_0, y_0)}(y - g((\beta^\top x)(F), \theta(F)))$  denotes  $\frac{d}{dt} H(y - g((\beta^\top x)(F_t), \theta(F_t)))|_{t=0, U=\Delta(z_0, y_0)-F}$ . The detailed expression of  $\dot{H}_{\Delta(z_0, y_0)}(y -$

$g((\beta^\top x)(F), \theta(F))$  can be written as

$$\begin{aligned} & \dot{H}_{\Delta(z_0, y_0)}(y - g((\beta^\top x)(F), \theta(F))) \\ &= \int h(y - g((\beta^\top x)(F), \theta(F))) \times \text{grad}_\alpha \{g((\beta^\top x)(F), \theta(F))\}^\top \times \frac{d\alpha}{dt} dF_X(x) \\ & \quad + I \left( y_0 \leq y + g((\beta^\top x)_0(F), \theta(F)) - g((\beta^\top x)(F), \theta(F)) \right) \\ & \quad - H(y - g((\beta^\top x)(F), \theta(F))), \end{aligned}$$

where  $\alpha = (\beta, \theta)^\top$  and  $\text{grad}_\alpha \{g((\beta^\top x)x(F), \theta(F))\}^\top$  represents the gradient of  $g((\beta^\top x)(F), \theta(F))$  with respect to  $\alpha$ . Since the parameter  $\alpha$  comes from a robust fit, we have that  $d\alpha/dt$  is bounded. Together with the conditions (C1) and (C5) in Appendix, it can be shown that  $\dot{H}_{\Delta(z_0, y_0)}(y - g((\beta^\top x)(F), \theta(F)))$  is also bounded. Further, the second-order influence function  $\text{IF}^{(2)}(z_0, y_0)$  of R-DREAM is bounded in the response direction.

For the purpose of comparison, the first-order and second-order influence functions for the test  $T_n^{\text{GWZ}}$  in (11) are derived. The first-order influence function is also zero, and the second-order influence function can be derived as

$$\begin{aligned} \text{IF}_{\text{GWZ}}^{(2)}(z_0, y_0) &= \int \left( \frac{d}{dt} g((\beta^\top x)(F), \theta(F)) f(x) \right)^2 dz + \left[ y_0 - g((\beta^\top x)_0(F), \theta(F)) \right]^2 \\ & \quad - \left[ y_0 - g((\beta^\top x)_0(F), \theta(F)) \right] \int \frac{d}{dt} g((\beta^\top x)(F), \theta(F)) dz \\ & \quad - \left[ y_0 - g((\beta^\top x)_0(F), \theta(F)) \right] \frac{d}{dt} g((\beta^\top x)_0(F), \theta(F)) f(z_0), \quad (26) \end{aligned}$$

where  $\frac{d}{dt} g((\beta^\top x)(F), \theta(F)) = \frac{d}{dt} g((\beta^\top x)(F_t), \theta(F_t))|_{t=0, U=\Delta(z_0, y_0)-F}$ . The second-order influence function  $\text{IF}_{\text{GWZ}}^{(2)}(z_0, y_0)$  in the  $y$ -direction is not bounded.  $\square$

The verification for the formula (25) is as follows.

*Verification of (25).* Let  $f_t$  and  $u$  be the probability density functions of  $F_t$  and  $U$ . Recall  $F_t = F + tU$ , then  $dF_t = dF + t dU$  and  $f_t = f + tu$ , we further have

$$\begin{aligned} L(t) &= \int \int \left[ H(y - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \\ & \quad \times \left( \int \left[ H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] f_t(z, y_1) dy_1 \right) dF_t(z, y), \\ &= \int \int \left[ H(y - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \\ & \quad \times \left( \int \left[ H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \right. \\ & \quad \left. \times f(z, y_1) dy_1 \right) dF(z, y) + t \int \int \left[ H(y - g(\beta^\top X(F_t), \theta(F_t))) - \frac{1}{2} \right] \end{aligned}$$

$$\begin{aligned}
 & \times \left( \int \left[ H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] u(z, y_1) dy_1 \right) dF(z, y) \\
 & + t \int \int \left[ H(y - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \\
 & \times \left( \int \left[ H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \right. \\
 & \times f(z, y_1) dy_1 \Big) dU(z, y) + t^2 \int \int \left[ H(y - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \\
 & \times \left( \int \left[ H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \times u(z, y_1) dy_1 \right) dU(z, y) \\
 & =: L_1(t) + L_2(t) + L_3(t) + L_4(t).
 \end{aligned}$$

For the first term  $L_1(t)$ , we have

$$\begin{aligned}
 \frac{dL_1(t)}{dt} &= \int \int \frac{d}{dt} H(y - g(\beta^\top x(F_t), \theta(F_t))) \\
 & \times \left( \int \left[ H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \right. \\
 & \times f(z, y_1) dy_1 \Big) dF(z, y) + \int \int \left[ H(y - g(\beta^\top x(F_t), \theta(F_t))) - \frac{1}{2} \right] \\
 & \times \left( \int \frac{d}{dt} H(y_1 - g(\beta^\top x(F_t), \theta(F_t))) \times f(z, y_1) dy_1 \right) dF(z, y).
 \end{aligned}$$

Since  $\int [H(y - g(\beta^\top x(F_t), \theta(F_t))) - 1/2] dF(y|z) = 0$ , we have that

$$\frac{dL_1(t)}{dt} \Big|_{t=0} = 0.$$

Further,

$$\begin{aligned}
 \frac{d^2 L_1(t)}{dt^2} \Big|_{t=0} &= 2 \int \int \dot{H}(y - g(\beta^\top x(F), \theta(F))) \times \left( \int \dot{H}(y_1 - g(\beta^\top x(F), \theta(F))) \right. \\
 & \times f(z, y_1) dy_1 \Big) dF(z, y).
 \end{aligned}$$

Similarly, it is not difficult to obtain that  $\frac{dL_i(t)}{dt} \Big|_{t=0} = 0, i = 2, 3, 4$  and  $\frac{1}{2} \frac{d^2 L_2(t)}{dt^2} \Big|_{t=0}, i = 2, 3, 4$  are equal to other three terms in the formula (25), which completes the proof. □

### References

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Ecomometrica*, 58, 1443–1458.

- Bierens, H. J., Ploberger, W. (1997). Asymptotic theory of integrated conditional moment test. *Econometrica*, 65, 1129–1151.
- Escanciano, J. C. (2006a). Goodness-of-fit tests for linear and nonlinear time series models. *Journal of the American Statistical Association*, 101, 531–541.
- Escanciano, J. C. (2006b). A consistent diagnostic test for regression models using projections. *Econometric Theory*, 22, 1030–1051.
- Escanciano, J. C. (2007). Model checks using residual marked empirical processes. *Statistica Sinica*, 17, 115–138.
- Fan, J. Q., Zhang, C., Zhang, J. (2001). Generalized likelihood ratio statistics and the wilks phenomenon. *The Annals of Statistics*, 29, 153–193.
- Feng, L., Zou, C. L., Wang, Z. J., Zhu, L. X. (2015). Robust comparison of regression curves. *Test*, 24, 185–204.
- Fernholz, L. T. (1983). *von Mises calculus for statistical functionals*. New York: Springer.
- González-Manteiga, W., Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *Test*, 22, 361–411.
- Guo, X., Wang, T., Zhu, L. X. (2016). Model checking for parametric single index models: A dimension-reduction model-adaptive approach. *Journal of the Royal Statistical Society: Series B*, 78, 1013–1035.
- Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69, 383–393.
- Härdle, W. (1992). *Applied nonparametric regression*. Cambridge: Cambridge University Press.
- Härdle, W., Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21, 1296–1947.
- Heritier, S., Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association*, 89, 897–904.
- Horowitz, J. L., Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric meanregression model against a nonparametric alternative. *Econometrica*, 69, 599–631.
- Huber, P. J. (1985). Projection pursuit. *The Annals of Statistics*, 13, 435–475.
- Koul, H. L., Ni, P. P. (2004). Minimum distance regression model checking. *Journal of Statistical Planning and Inference*, 119, 109–141.
- Lavergne, P., Patilea, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *Journal of Econometrics*, 143, 103–122.
- Lavergne, P., Patilea, V. (2012). One for all and all for one: Regression checks with many regressors. *Journal of Business & Economic Statistics*, 30, 41–52.
- Li, K. C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86, 316–327.
- Markatou, M., Hettmansperger, T. P. (1990). Robust bounded-influence tests in linear models. *Journal of American Statistical Association*, 85, 187–190.
- Markatou, M., Manos, G. (1996). Robust tests in nonlinear regression models. *Journal of Statistical Planning and Inference*, 55, 205–217.
- Maronna, R., Martín, D., Yohai, V. (2006). *Robust statistics: Theory and methods*. New York: Wiley.
- Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications*, 10, 186–196.
- Schrader, R. M., Hettmansperger, T. P. (1980). Robust analysis of variance based upon a likelihood ratio criterion. *Biometrika*, 67, 93–101.
- Sen, P. K. (1982). On  $M$ -tests in linear models. *Biometrika*, 69, 245–248.
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.
- Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Annals of Statistics*, 8, 1348–1360.
- Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics*, 25, 613–641.
- Stute, W., Zhu, L. X. (2002). Model checks for generalized linear models. *Schandinavian Journal of Statistics*, 29, 535–545.
- Stute, W., González-Manteiga, W., Presedo-Quindimil, M. (1998a). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, 93, 141–149.
- Stute, W., Thies, S., Zhu, L. X. (1998b). Model checks for regression: An innovation process approach. *Annals of Statistics*, 26, 1916–1934.
- Stute, W., Xu, W. L., Zhu, L. X. (2008). Model diagnosis for parametric regression in high dimensional spaces. *Biometrika*, 95, 1–17.



- Van Keilegom, I., González-Manteiga, W., Sánchez Sellero, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Test*, 17, 401–415.
- Wang, L., Qu, A. N. (2007). Robust tests in regression models with omnibus alternatives and bounded influence. *Journal of the American Statistical Association*, 102, 347–358.
- Watson, G. S. (1964). Smoothing regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26, 359–372.
- Whang, Y. J. (2000). Consistent bootstrap tests of parametric regression functions. *Journal of Econometrics*, 98, 27–46.
- Xia, Q., Xu, W. L., Zhu, L. X. (2015). Consistently determining the number of factors in multivariate volatility modelling. *Statistica Sinica*, 25, 1025–1044.
- Xia, Y. C. (2007). A constructive approach to the estimation of dimension reduction directions. *Annals of Statistics*, 35, 2654–2690.
- Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75, 263–289.
- Zhu, L. P., Wang, T., Zhu, L. X., Ferré, L. (2010). Sufficient dimension reduction through discretization-expectation estimation. *Biometrika*, 97, 295–304.
- Zhu, L. X. (2003). Model checking of dimension-reduction type for regression. *Statistica Sinica*, 13, 283–296.
- Zhu, L. X., An, H. Z. (1992). A test for nonlinearity in regression models. *Journal of Mathematics*, 4, 391–397. (Chinese).
- Zhu, L. X., Li, R. Z. (1998). Dimension-reduction type test for linearity of a stochastic model. *Acta Mathematicae Applicatae Sinica*, 14, 165–175.