Biometrika (2022), **109**, 1, *pp.* 103–122
Printed in Great Britain
doi: 10.1093/biomet/asab009
Advance Access publication 31 August 2021

Integrated conditional moment test and beyond: when the number of covariates is divergent

BY FALONG TAN

Department of Statistics, Hunan University, Changsha 410082, China
falongtan@hnu.edu.cn

AND LIXING ZHU

Center for Statistics and Data Science, Beijing Normal University at Zhuhai, Zhuhai, China
lzhu@hkbu.edu.hk

SUMMARY

The classical integrated conditional moment test is a promising method for model checking and its basic idea has been applied to develop several variants. However, in diverging-dimension scenarios, the integrated conditional moment test may break down and has completely different limiting properties from the fixed-dimension case. Furthermore, the related wild bootstrap approximation can also be invalid. To extend this classical test to diverging dimension settings, we propose a projected adaptive-to-model version of the integrated conditional moment test. We study the asymptotic properties of the new test under both the null and alternative hypotheses to examine if it maintains significance level, and its sensitivity to the global and local alternatives that are distinct from the null at the rate $n^{-1/2}$. The corresponding wild bootstrap approximation can still work for the new test in diverging-dimension scenarios. We also derive the consistency and asymptotically linear representation of the least squares estimator when the parameter diverges at the fastest possible known rate in the literature. Numerical studies show that the new test can greatly enhance the performance of the integrated conditional moment test in high-dimensional cases. We also apply the test to a real dataset for illustration.

Some key words: Adaptive-to-model test; Dimension reduction; Integrated conditional moment test; Least squares estimation; Model checking; Wild bootstrap.

1. INTRODUCTION

Testing for model specification is one of the most important issues in regression analysis, as using a misspecified regression model may induce misleading statistical inference. There are a number of proposals available in the literature when the dimension p of the predictor vector is fixed. However, for high-dimensional paradigms where the dimension p is treated as a divergent number when the sample size n tends to infinity, such model-checking problems have not yet been systematically studied. Thus, the research described in this paper is motivated by developing a goodness-of-fit test for parametric multiple-index regression models in diverging-dimension settings.

Let Y be a real-valued response variable and $X \in \mathbb{R}^p$ be its associated random-design predictor vector. The null hypothesis we want to test is that (Y, X) follows a parametric multiple-index model as

$$Y = g(\beta_{01}^{\mathrm{T}} X, \ldots, \beta_{0d}^{\mathrm{T}} X, \vartheta_0) + \varepsilon, \tag{1}$$

where g is a given smooth function, $\vartheta_0 \in \mathbb{R}^l$ and $\beta_0 = (\beta_{01}, \ldots, \beta_{0d}) \in \mathbb{R}^{p \times d}$ are the unknown matrix of regression parameters, $\varepsilon = Y - E(Y \mid X)$ is the unpredictable part of Y given X, and the notation $^{\mathrm{T}}$ denotes the transpose. Without loss of generality, we assume that $\beta_{01}, \ldots, \beta_{0d}$ are orthogonal. To make full use of the model structures under both the null and the alternative hypothesis, consider the general alternative model

$$Y = G(B^{\mathrm{T}} X) + \varepsilon, \tag{2}$$

where $E(\varepsilon \mid X) = 0$, $G(\cdot)$ is an unknown smooth function and B is a $p \times q$ orthonormal matrix such that $\mathcal{S}_{E(Y|X)} = \operatorname{span}(B)$. Here, $\mathcal{S}_{E(Y|X)}$ is the central mean space of Y with respect to X, which will be specified in § 4. Throughout this paper, we assume that q is fixed.

When the dimension p of the predictor vector is fixed, local smoothing tests by using nonparametric estimation methods is one of the primary class of tests for model checking; see Härdle & Mammen (1993), Zheng (1996), Dette (1999), Fan & Huang (2001), Horowitz & Spokoiny (2001), Koul & Ni (2004), Van Keilegom et al. (2008) and Lavergne & Patilea (2008, 2012), amongst others. The use of nonparametric regression estimators usually causes them to suffer severely from the curse of dimensionality in high-dimensional cases; Guo et al. (2016) gave some detailed comments on this issue. Another class of tests for model checking is the class of global smoothing tests which are constructed in terms of converting the constraint on the conditional expectation $E(\varepsilon \mid X) = 0$ to infinite and parametric unconditional orthogonality restrictions, that is,

$$E(\varepsilon \mid X) = 0 \quad \Leftrightarrow \quad E\{\varepsilon w(X, t)\} = 0 \qquad \forall\, t \in \Gamma, \tag{3}$$

where Γ is some proper space. There exist several parametric families $w(X, t)$ such that the equivalence (3) holds; see Bierens & Ploberger (1997) and Escanciano (2006b) for more details on the primitive conditions for $w(\cdot, t)$ to satisfy this equivalence. The indicator function $w(X, t) = I(X \leqslant t)$ is commonly used as a weight function in the literature; see, e.g., Stute (1997), Stute et al. (1998a,b), Zhu (2003), Khmaladze & Koul (2004) and Stute et al. (2008), among many others. Recently, Shah & Bühlmann (2018) and Janková et al. (2020) proposed goodness-of-fit tests for high-dimensional linear models and generalized linear models with fixed design, respectively. The setting of these two works is different from the present paper as we consider regression models with random design here.

In a seminal paper, Bierens (1982) used the characteristic function $w(X, t) = \exp(\mathrm{i} t^{\mathrm{T}} X)$, where $i = \sqrt{(-1)}$ denotes the imaginary unit, as the weight function and constructed an integrated conditional moment, ICM, test. The ICM test statistic is

$$ICM_n = \int_\Gamma \left| n^{-1/2} \sum_{j=1}^{n} \hat{\eta}_j \exp\{\mathrm{i} t^{\mathrm{T}} \Phi(X_j)\} \right|^2 \mathrm{d}\mu(t), \tag{4}$$

where $\hat{\eta}_j$ is the residual, $\Gamma = \prod_{i=1}^{p} [-\epsilon_i, \epsilon_i]$ with ϵ_i being arbitrarily positive numbers, μ is the Lebesgue measure on Γ, and $\Phi(\cdot)$ is a one-to-one bounded smoothing function from $\mathbb{R}^p$ to $\mathbb{R}^p$. Since high-dimensional numerical integrations are extremely hard to handle in practice, the computation of the integral in (4) or its approximation becomes difficult and time-consuming for large dimensions p. Thus, it is usually suggested in the literature to use the standard normal measure in (4), and then Bierens' test statistic has an analytic formula:

$$ICM_n = \int_{\mathbb{R}^p} \left| n^{-1/2} \sum_{j=1}^{n} \hat{\eta}_j \exp(\mathrm{i} t^{\mathrm{T}} X_j) \right|^2 \phi(t)\, \mathrm{d}t$$
$$= n^{-1} \sum_{j,k=1}^{n} \hat{\eta}_j \hat{\eta}_k \exp\left(-\frac{1}{2} \|X_j - X_k\|^2 \right), \quad (5)$$

where $\phi(t)$ denotes the standard normal density on $\mathbb{R}^p$ (see Escanciano, 2006a; Lavergne & Patilea, 2008, 2012). Since the ICM test statistic is not asymptotically distribution-free, it needs to resort to resampling methods such as the wild bootstrap to approximate the null distribution. When p is fixed, Stute et al. (1998a) showed that the distribution of the residual marked empirical process-based test statistic can be approximated by the wild bootstrap. Dominguez (2005) further proved the validity of the wild bootstrap for the ICM test in the fixed-dimension setting. Some relevant approaches are summarized in Zhu (2005). However, the ICM test suffers from the dimensionality problem. When the dimension $p \to \infty$ as the sample size $n \to \infty$, we found that the test statistic ICM_n has completely different limiting properties from those in fixed-dimension cases, and this also causes the related wild bootstrap approximation to be invalid. Under mild conditions, we will show that the limiting distributions of ICM_n under the null and alternative hypotheses, degenerate to finite fixed values, and the wild bootstrap version of ICM_n has the exact same limits as ICM_n, rather than giving an approximation to the limiting null distribution. In other words, the related wild bootstrap approximation would fail to work for the ICM test when the dimension p is large. Therefore, this unique feature causes the ICM test either power loss or inflation of the significance level in diverging scenarios. These results are presented in § 3. The simulation studies in § 6 also confirm this phenomenon. That is, the ICM test based on the wild bootstrap works very well when the dimension p is relatively small. However, in high-dimensional cases its power may drop very quickly, especially when the components of covariates are uncorrelated.

The primary purpose of this research is to extend the ICM test to the diverging-dimension setting under the multiple-index model structure. To this end, we adopt projection and sufficient dimension reduction techniques to develop a projected adaptive-to-model version of the ICM test. The asymptotic properties of the proposed test statistic are investigated under both the null and alternative hypotheses when p is divergent. We show that the proposed test has a nondegenerate limiting null distribution, is consistent against any alternative hypothesis in (2), and can detect the local alternatives converging to the null at the parametric rate $n^{-1/2}$. Although the ICM test based on the wild bootstrap fails to work in diverging-dimension settings, the proposed test can still be approximated by the wild bootstrap provided $p = o(n/\log n)^{1/3}$.

The innovative model adaptation methodology was first proposed by Guo et al. (2016) in the literature for testing the specification of parametric single-index models in fixed-dimension cases. Tan & Zhu (2019) also used this methodology to attack the same testing problem as Guo et al. (2016) with diverging dimension p, and proposed the martingale transformation to obtain an asymptotically distribution-free test. To make sure that these two tests are omnibus, a restrictive assumption was imposed on the central mean subspace $\mathcal{S}_{E(Y|X)}$ which may not hold for nonlinear regression models. In this paper we modify the model adaptation procedure to remove this restrictive assumption and extend it to test more general multiple-index models. Further, as the martingale transformation proposed in Tan & Zhu (2019) relies on the single-index structure under the null hypothesis, it fails to work for multiple-index null models in general. Thus, the test of Tan & Zhu (2019) is not extendable to handle parametric multiple-index models.

To study the asymptotic properties of the test statistic, the first step is to derive the asymptotic properties of the estimator $(\hat{\beta}, \hat{\vartheta})$ of (β_0, ϑ_0) in divergent-dimension cases. This problem has been well investigated for linear regression models, see Huber (1973), Portnoy (1984, 1985) and Zou & Zhang (2009). Nevertheless, it has rarely been studied for nonlinear models. Tan & Zhu (2019) obtained the estimation consistency and the asymptotically linear representation of $(\hat{\beta}, \hat{\vartheta})$ at the rates $p = o(n^{1/4})$ and $p = o(n^{1/5})$ of divergence, respectively. In this paper, with the help of some high-dimensional empirical processes techniques, we will greatly improve these rates of divergence to $p = o(n^{1/2})$ and $p = o(n/\log n)^{1/3}$, and obtain the uniformly asymptotically linear representation of $(\hat{\beta}, \hat{\vartheta})$. Under the conditions we assume, the rates are, to our knowledge, the best in the relevant research in the literature. Shi et al. (2019) obtained similar results for generalized linear models with fixed design in the divergent-dimension setting, but, as we actually deal with general nonlinear models with random designs here, our approach is significantly different from theirs.

2. PARAMETER ESTIMATION

To estimate the unknown parameters (β_0, ϑ_0) in the null model (1), we here restrict ourselves to the least squares method. For notational simplicity, define $\theta = \{\mathrm{vec}(\beta)^{\mathrm{T}}, \vartheta^{\mathrm{T}}\}^{\mathrm{T}}$ where $\mathrm{vec}(\beta) = (\beta_1^{\mathrm{T}}, \ldots, \beta_d^{\mathrm{T}})^{\mathrm{T}}$. Then, the null model (1) can be restated as $Y = g(\theta_0, X) + \varepsilon$, where $\theta_0 \in \Theta \subset \mathbb{R}^m$ with $m = pd + l$ and $E(\varepsilon \mid X) = 0$, Θ being a compact subset of $\mathbb{R}^m$. Let

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \{Y_i - g(\theta, X_i)\}^2.$$

To analyse the asymptotic properties of $\hat{\theta}$, set

$$\tilde{\theta}_0 = \arg\min_{\theta \in \Theta} E\{Y - g(\theta, X)\}^2 = \arg\min_{\theta \in \Theta} E\{m(X) - g(\theta, X)\}^2, \tag{6}$$

where $m(x) = E(Y \mid X = x)$ denotes the regression function. Under the regularity condition (A4) specified in the Supplementary Material, $\tilde{\theta}_0 = \theta_0$ under the null hypothesis. Under the alternatives, $\tilde{\theta}_0$ typically depends on the cumulative distribution of X. Under the regularity conditions (A1)–(A5) in the Supplementary Material, we can derive the asymptotic properties of $\hat{\theta}$ when the dimension m is divergent. Although we only need the asymptotic properties of the parameters for multiple-index models in this paper, the asymptotic properties of $\tilde{\theta}_0$ also hold for general parametric regression models. Putting $g'(\theta, x) = \frac{\partial g(\theta, x)}{\partial \theta}$ and $g''(\theta, x) = \frac{\partial g'(\theta, x)}{\partial \theta^{\mathrm{T}}}$, let $\Sigma = E\{g'(\tilde{\theta}_0, X) g'(\tilde{\theta}_0, X)^{\mathrm{T}}\} - E[\{m(X) - g(\tilde{\theta}_0, X)\} g''(\tilde{\theta}_0, X)]$, and let $\| \cdot \|$ denote the Frobenius norm.

THEOREM 1. *Suppose that (A1)–(A4) in the Supplementary Material hold. If $m^2/n \to 0$, then it follows that $\hat{\theta}$ is a norm-consistent estimator of $\tilde{\theta}_0$ in the sense that $\|\hat{\theta} - \tilde{\theta}_0\| = O_p(m/n)^{1/2}$. Furthermore, if $(m^3 \log n)/n \to 0$, then under (A1)–(A5) in the Supplementary Material,*

$$n^{1/2} \alpha^{\mathrm{T}} (\hat{\theta} - \tilde{\theta}_0) = n^{-1/2} \sum_{i=1}^{n} \{Y_i - g(\tilde{\theta}_0, X_i)\} \alpha^{\mathrm{T}} \Sigma^{-1} g'(\tilde{\theta}_0, X_i) + o_p(1),$$

where the remaining term $o_p(1)$ is uniform over $\alpha \in \mathcal{S}^{m-1}$, i.e., the unit sphere in $\mathbb{R}^m$.

The convergence rate $O_p(m/n)^{1/2}$ in Theorem 1 is in line with the results of the M-estimator, for linear regression models $g(\theta, X) = \theta^{\mathrm{T}} X$, studied by Huber (1973) and Portnoy (1984) when the number of parameters is divergent. The uniformly asymptotically linear representation of $\hat{\theta} - \tilde{\theta}$ in Theorem 1 is much stronger than that in Portnoy (1985), who obtained the asymptotically linear representation for a given $\alpha \in \mathcal{S}^{m-1}$ for linear regression models. Further, in the linear regression setting and sub-Gaussian cases, we can respectively strengthen the results to faster rates $p = o(n)$ and $p = o(n^{1/2})$. Here we say a random variable $Z \in \mathbb{R}$ is sub-Gaussian with variance proxy σ^2 if $E[\exp\{u(Z - EZ)\}] \leqslant \exp(\sigma^2 u^2/2)$ for all $u \in \mathbb{R}$. A random vector $X \in \mathbb{R}^p$ is sub-Gaussian with variance proxy σ^2 if $v^{\mathrm{T}} X$ is sub-Gaussian with variance proxy σ^2 for all $v \in \mathcal{S}^{p-1}$.

THEOREM 2. *Suppose that (A1)–(A4) in the Supplementary Material hold for linear regression models, and X is sub-Gaussian with the variance proxy σ^2. If $p/n \to 0$, then we have $\|\hat{\theta} - \tilde{\theta}_0\| = O_p(p/n)^{1/2}$. Moreover, if $p^2/n \to 0$, it follows that*

$$n^{1/2}\alpha^{\mathrm{T}}(\hat{\theta} - \tilde{\theta}_0) = n^{-1/2}\sum_{i=1}^{n}(Y_i - \tilde{\theta}_0^{\mathrm{T}} X_i)\alpha^{\mathrm{T}}\Sigma^{-1}X_i + o_p(1),$$

where $\Sigma = E(XX^{\mathrm{T}})$ and the term $o_p(1)$ is uniform over $\alpha \in \mathcal{S}^{p-1}$.

3. THE ICM TEST AND THE WILD BOOTSTRAP IN DIVERGING DIMENSION SETTINGS

When the dimension p is fixed, the limiting properties of the ICM test, or more generalized Cramér–von Mises tests, have been well investigated in the literature, see, for instance, Bierens & Ploberger (1997) and Escanciano (2009). Specifically, when the dimension p is fixed, it can be shown that, under the null,

$$ICM_n \longrightarrow \sum_{j=1}^{\infty}\lambda_j Z_j^2 \quad \text{in distribution,}$$

where $\{\lambda_j\}_{j=1}^{\infty}$ depend on the joint distribution of (Y, X), and $\{Z_j\}_{j=1}^{\infty}$ are independent and identically distributed $N(0,1)$ random variables. Under the alternative, ICM_n tends to infinity in probability. The ICM test is not asymptotically distribution-free as its limiting null distribution relies on the unknown data-generating process. Thus, the wild bootstrap is often a promising approach in the literature to approximate the null distribution (Stute et al., 1998a; Dominguez, 2005). However, if p diverges as n tends to infinity, we will see that the classic limiting results of ICM_n no longer hold and the related wild bootstrap approximation is invalid. To see this, recall that ICM_n can be rewritten as

$$\begin{aligned} ICM_n &= \frac{1}{n}\sum_{j,k=1}^{n}\hat{\eta}_j\hat{\eta}_k \exp\left(-\frac{1}{2}\|X_j - X_k\|^2\right) \\ &= \frac{1}{n}\sum_{j=1}^{n}\hat{\eta}_j^2 + \frac{2}{n}\sum_{j<k}\hat{\eta}_j\hat{\eta}_k \exp\left(-\frac{1}{2}\|X_j - X_k\|^2\right), \end{aligned}$$

where $\hat{\eta}_j = Y_j - g(\hat{\theta}, X_j)$. Assuming that (A1)–(A5) in the Supplementary Material hold, we can show that, under the null hypothesis,

$$ICM_n = E(\varepsilon^2) + O_p(p^2/n)^{1/2} + p^2[E\{\exp(-2\|X_1 - X_2\|^2)\}]^{1/4} O_p(1), \tag{7}$$

while under the alternative hypothesis,

$$ICM_n = E\{\varepsilon + m(X) - g(\tilde{\theta}_0, X)\}^2 + O_p(p^2/n)^{1/2} + n[E\{\exp(-\|X_1 - X_2\|^2)\}]^{1/2} O_p(1). \quad (8)$$

The proofs for (7) and (8) are given in the Supplementary Material. To obtain the exact limiting distribution of ICM_n in the diverging-dimension setting, we take a special case for illustration. Suppose that the components $\{X_{1k} : k = 1, \ldots, p\}$ of X_1 are independent and identically distributed. It is readily seen that $E\{\exp(-2\|X_1 - X_2\|^2)\} = [E\{\exp(-2|X_{11} - X_{21}|^2)\}]^p$. This means that $E\{\exp(-2\|X_1 - X_2\|^2)\}$ converges to zero at an exponential rate. If $p^2/n \to 0$ and $(\log n)/p \to 0$, it is easy to see that ICM_n converges to finite points in probability under both the null and alternative hypotheses.

Now we can show that the wild bootstrap used for critical value determination would be invalid when p is divergent. The procedure of the wild bootstrap is as follows. Set $X_j^* = X_j$, $Y_j^* = g(\hat{\theta}, X_j) + \hat{\eta}_j V_j^*$ $(j = 1, \ldots, n)$, where $\hat{\eta}_j$ is the residual and $\{V_j^*\}_{j=1}^n$ are independent and identically distributed bounded random variables with zero mean and unit variance, independent of the original sample $\{(X_i, Y_i)\}_{i=1}^n$. A popularly used sequence of $\{V_j^*\}_{j=1}^n$ is the independent and identically distributed Bernoulli variates with

$$\mathbb{P}\left(V_j^* = \frac{1-\sqrt{5}}{2}\right) = \frac{1+\sqrt{5}}{2\sqrt{5}}, \qquad \mathbb{P}\left(V_j^* = \frac{1+\sqrt{5}}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}}.$$

For other examples of $\{V_j^*\}_{j=1}^n$, one can refer to Mammen (1993). Let $\hat{\theta}^*$ be the bootstrap estimator obtained by the least squares method based on the bootstrap sample $\{(X_j, Y_j^*)\}_{j=1}^n$. Then the bootstrap version of the ICM test statistic is

$$ICM_n^* = \int_{\mathbb{R}^p} \left| n^{-1/2} \sum_{j=1}^n \hat{\eta}_j^* \exp(\mathrm{i}t^{\mathrm{T}} X_j) \right|^2 \phi(t)\,\mathrm{d}t = \frac{1}{n} \sum_{j,k=1}^n \hat{\eta}_j^* \hat{\eta}_k^* \exp\left(-\frac{1}{2}\|X_j - X_k\|^2\right),$$

where $\hat{\eta}_j^* = Y_j^* - g(\hat{\theta}^*, X_j)$ and $\phi(t)$ is the standard normal density in $\mathbb{R}^p$.

Let P^* be the probability measure induced by the wild bootstrap resampling conditional on the original sample $\{(X_i, Y_i) : i = 1, \ldots, n\}$. Then, under (A1)–(A5) in the Supplementary Material, we can show that

$$\begin{aligned} ICM_n^* = {} & E\{\varepsilon + m(X) - g(\tilde{\theta}_0, X)\}^2 + O_{P^*}(p^2/n)^{1/2} \\ & + p^2[E\{\exp(-2\|X_1 - X_2\|^2)\}]^{1/4} O_{P^*}(1). \end{aligned} \quad (9)$$

The proof of (9) is also given in the Supplementary Material.

When $p^2/n \to 0$ and $(\log n)/p \to 0$, the limiting distributions of ICM_n under the null and alternative hypotheses, respectively, degenerate to finite fixed values, and its wild bootstrap version ICM_n^* has the exact same limits as ICM_n under both the null and alternatives. Thus, ICM_n^* cannot approximate well the limiting null distribution of ICM_n under the alternatives. This causes Bierens' ICM test, when using the wild bootstrap, to not maintain the significance level well and have reasonable power in large-dimension scenarios. The simulation studies in § 6 also validate these phenomena.

Remark 1. It is often helpful to analyse the power performance of goodness-of-fit tests in regression based on local alternative hypotheses. It is well known that most local smoothing tests can only detect alternative hypotheses distinct from the null hypothesis at the rate of order $r_n = n^{-1/2} h^{-p/4}$, see, for example, Zheng (1996). Here, h is the bandwidth in the nonparametric estimation. When p is large, this convergence rate can be very slow, and thus local smoothing tests would be less powerful for detecting alternative models. However, this cannot explain why global smoothing tests still lose power quickly when the dimension p becomes large, as it can be shown that global smoothing tests can detect the alternatives that converge to the null at the rate $n^{-1/2}$ and are not related to the dimension p, see for instance Bierens & Ploberger (1997) and Stute (1997). From this observation, global smoothing tests seem to be powerful in large-dimension cases. But this is not the case in practice. Most global smoothing tests still suffer from the curse of dimensionality. In this paper, we give a new explanation as to why the ICM test, as a representative of global smoothing tests, loses power in large-dimension scenarios. Furthermore, our method can be readily extended to analyse the power performance of other global smoothing tests for large dimension p.

4. Adaptive-to-model version of the ICM test

4.1. *Test statistic construction*

When the dimension p diverges as the sample size n tends to infinity, the main reason that the ICM test statistic ICM_n and its related wild bootstrap fail to work is that the weight function $\exp(-\|X_j - X_k\|^2/2)$ in (5) degenerates to zero at an exponential rate for all $j \neq k$. To extend the ICM test to diverging-dimension settings under the multiple-index model structure, we adopt projection and sufficient dimension reduction techniques to reduce the original dimension p to a much lower dimension and then propose a projected adaptive-to-model version of the ICM test. First, we restate the hypotheses as

$$H_0: \ E(Y \mid X) = g(\beta_0^{\mathrm{T}} X, \vartheta_0) \quad \text{for some } \beta_0 \in \mathbb{R}^{p\times d} \text{ and } \vartheta_0 \in \mathbb{R}^l,$$
$$H_1: \ E(Y \mid X) = G(B^{\mathrm{T}} X) \neq g(\beta^{\mathrm{T}} X, \vartheta) \quad \text{for all } \beta \in \mathbb{R}^{p\times d} \text{ and } \vartheta \in \mathbb{R}^l,$$

where G is an unknown smoothing function and B is a $p \times q$ orthonormal matrix satisfying $\mathcal{S}_{E(Y|X)} = \mathrm{span}(B)$. Here, $\mathcal{S}_{E(Y|X)}$ is the central mean space of Y with respect to X, which is defined as the intersection of all subspaces $\mathrm{span}(A)$ such that $Y \perp\!\!\!\perp E(Y \mid X) \mid A^{\mathrm{T}} X$, where $\perp\!\!\!\perp$ means statistical independence and $\mathrm{span}(A)$ is the subspace spanned by the columns of A. Under mild conditions, such a subspace $\mathcal{S}_{E(Y|X)}$ always exists (Cook & Li, 2002). If $\mathcal{S}_{E(Y|X)} = \mathrm{span}(B)$, it follows that $E(Y \mid X) = E(Y \mid B^{\mathrm{T}} X)$. Thus, it is reasonable to use the regression format $G(B^{\mathrm{T}} X)$ instead of the fully nonparametric function $G(X)$ in the alternative hypothesis H_1. The dimension of $\mathcal{S}_{E(Y|X)}$ is called the structural dimension, which is q under the alternatives. Similarly, under the null we have $\mathcal{S}_{E(Y|X)} = \mathrm{span}(B) = \mathrm{span}(\beta_0)$ with a structural dimension d.

Recall that $\varepsilon = Y - E(Y \mid X)$ and

$$(\tilde\beta_0, \tilde\vartheta_0) = \mathop{\arg\min}_{\beta \in \mathbb{R}^{p\times d}, \vartheta \in \mathbb{R}^l} E\{Y - g(\beta^{\mathrm{T}} X, \vartheta)\}^2.$$

Let $\eta = Y - g(\tilde\beta_0^{\mathrm{T}} X, \tilde\vartheta_0)$ and $\tilde{\mathfrak{B}} = (\tilde\beta_0, B)$. Under H_0, we have $\tilde\beta_0 = \beta_0$ and $\mathrm{span}(B) = \mathrm{span}(\beta_0)$. Consequently,

$$E(\eta \mid \tilde{\mathfrak{B}}^{\mathrm{T}} X) = E(Y \mid \tilde\beta_0^{\mathrm{T}} X, B^{\mathrm{T}} X) - g(\tilde\beta_0^{\mathrm{T}} X, \tilde\vartheta_0) = E(Y \mid \beta_0^{\mathrm{T}} X) - g(\beta_0^{\mathrm{T}} X, \vartheta_0) = 0.$$

Under H_1, we have $\eta = G(B^{\mathrm{T}} X) - g(\tilde\beta_0^{\mathrm{T}} X, \tilde\vartheta_0) + \varepsilon$, and then $E(\eta \mid \tilde{\mathfrak{B}}^{\mathrm{T}} X) = G(B^{\mathrm{T}} X) - g(\tilde\beta_0^{\mathrm{T}} X, \tilde\vartheta_0) \neq 0$.

By Theorem 1 of Bierens (1982), under the null we have $E\{\eta \exp(\mathrm{i}t^{\mathrm{T}}\tilde{\mathfrak{B}}^{\mathrm{T}}X)\} = 0$ for all t. Consequently,

$$\int_{\mathbb{R}^{2d}} |E\{\eta \exp(\mathrm{i}t^{\mathrm{T}}\tilde{\mathfrak{B}}^{\mathrm{T}}X)\}|^2 \varphi(t)\, \mathrm{d}t = 0. \tag{10}$$

Under the alternatives we have $E\{\eta \exp(\mathrm{i}t_0^{\mathrm{T}}\tilde{\mathfrak{B}}^{\mathrm{T}}X)\} \neq 0$ for some $t_0 \in \mathbb{R}^{d+q}$. Therefore,

$$\int_{\mathbb{R}^{d+q}} |E\{\eta \exp(\mathrm{i}t^{\mathrm{T}}\tilde{\mathfrak{B}}^{\mathrm{T}}X)\}|^2 \varphi(t)\, \mathrm{d}t > 0, \tag{11}$$

where $\varphi(t)$ denotes a positive weight function that will be specified later. Thereby, we reject the null hypothesis for large values of the empirical version of the left-hand side of (10) and (11). Let $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ be a random sample from the distribution of (X, Y). We then propose an adaptive-to-model integrated conditional moment test statistic as

$$\hat{V}_n(t) = n^{-1/2} \sum_{j=1}^{n} \{Y_j - g(\hat{\beta}^{\mathrm{T}}X_j, \hat{\vartheta}_0)\} \exp(\mathrm{i}t^{\mathrm{T}}\hat{\mathfrak{B}}^{\mathrm{T}}X_j),$$

$$AICM_n = \int_{\mathbb{R}^{d+\hat{q}}} |\hat{V}_n(t)|^2 \varphi(t)\, \mathrm{d}t,$$

where $\hat{\mathfrak{B}} = (\hat{\beta}, \hat{B})$, $\hat{B}$ is a sufficient dimension reduction estimator of B with an estimated structural dimension $\hat{q}$ of q, and $\hat{\beta}$ is a norm-consistent estimator of $\tilde{\beta}_0$. In this paper, we restrict ourselves to the least squares estimator of $\tilde{\beta}_0$.

Remark 2. It is worth mentioning that Guo et al. (2016) first used sufficient dimension reduction techniques to construct a goodness-of-fit test for parametric single-index models when the dimension p is fixed. However, they only used the matrix B rather than $(\tilde{\beta}_0, B)$ as we consider here. To make sure $E(\eta \mid B^{\mathrm{T}}X) \neq 0$ under the alternative hypotheses, they needed an extra condition that $\tilde{\beta}_0 \in \mathrm{span}(B)$. This restrictive condition does not always hold for nonlinear regression models. To avoid this extra condition, we use the matrix $\tilde{\mathfrak{B}}$ in the test construction to ensure the conditional expectation $E(\eta \mid \tilde{\mathfrak{B}}^{\mathrm{T}}X) \neq 0$ for any alternative hypothesis in H_1.

4.2. *The choice of* $\varphi(t)$

The choice of weight functions $\varphi(t)$ is flexible. Recall that our test statistic $AICM_n = \int_{\mathbb{R}^{d+\hat{q}}} |\hat{V}_n(t)|^2 \varphi(t)\, \mathrm{d}t$. For any even positive function $\varphi(t)$, some elementary algebra leads to

$$AICM_n = \frac{1}{n} \sum_{j,k=1}^{n} \hat{\eta}_j \hat{\eta}_k \int_{\mathbb{R}^{d+\hat{q}}} \cos\{t^{\mathrm{T}}(\hat{\mathfrak{B}}^{\mathrm{T}}X_j - \hat{\mathfrak{B}}^{\mathrm{T}}X_k)\} \varphi(t)\, \mathrm{d}t,$$

where $\hat{\eta}_j = Y_j - g(\hat{\beta}^{\mathrm{T}}X_j, \hat{\vartheta})$. Let $K_\varphi(x) = \int \cos(t^{\mathrm{T}}x)\varphi(t)\, \mathrm{d}t$; then, it follows that

$$AICM_n = \frac{1}{n} \sum_{j,k=1}^{n} \hat{\eta}_j \hat{\eta}_k K_\varphi(\hat{\mathfrak{B}}^{\mathrm{T}}X_j - \hat{\mathfrak{B}}^{\mathrm{T}}X_k).$$

Yet the calculation of $K_\varphi(x)$ is still complex in high-dimension settings, even when we use sufficient dimension reduction techniques. To facilitate the calculation of the test statistic, a closed form of the function $K_{\varphi}(x)$ is preferred. There is a large class of weight functions $\varphi(t)$ available for this purpose. Consider the sub-Gaussian symmetric a-stable distribution as an example. By Proposition 2.5.2 of Samorodnitsky & Taqqu (1994), we have

$$K_{\varphi_a}(x) = \int_{\mathbb{R}^p} \cos(t^{\mathrm{T}}x)\varphi_a(t)\,\mathrm{d}t = \exp\left(-\left|\frac{1}{2}\sum_{i,j=1}^{p} x_i x_j R_{ij}\right|^{a/2}\right),$$

where $\varphi_a(t)$ is the density of a sub-Gaussian symmetric a-stable distribution with $0 < a \leqslant 2$, and $R_{ij} = E(Z_i Z_j)$ are the covariances of the underlying zero-mean Gaussian random vector $Z = (Z_1, \ldots, Z_p)^{\mathrm{T}}$. The family of a-stable distributions includes many frequently used distributions such as multivariate Gaussian distributions with $a = 2$ and Cauchy distributions with $a = 1$. More details about a-stable distributions can be found in Chapter 2 of Samorodnitsky & Taqqu (1994). For a sub-Gaussian symmetric a-stable distribution, if the covariance matrix of Z is the identity matrix I_p, then we have $K_{\varphi_a}(x) = \exp\{-(2^{-1/2}\|x\|)^a\}$. Substituting this weight function into $AICM_n$, we obtain

$$AICM_n = \frac{1}{n}\sum_{j,k=1}^{n} \hat{\eta}_j \hat{\eta}_k \exp\{-(2^{-1/2}\|\hat{\mathfrak{B}}^{\mathrm{T}}X_j - \hat{\mathfrak{B}}^{\mathrm{T}}X_k\|)^a\}.$$

It would be interesting to investigate how the weight parameter a affects the performance of the test statistic, and how to determine the optimal value for a in the context of goodness-of-fit testing. This is still an open problem in the literature. As was pointed out by Hlávka et al. (2017), this problem is highly nontrivial even in the strict parametric context of independent and identically distributed testing for univariate normality. In the simulation studies, we choose the density $\phi(t)$ of a standard Gaussian distribution, i.e., $a = 2$, as a weight function. Then the test statistic can be stated as

$$AICM_n = \frac{1}{n}\sum_{j,k=1}^{n} \hat{\eta}_j \hat{\eta}_k \exp\left\{-\frac{1}{2}(\|\hat{B}^{\mathrm{T}}X_j - \hat{B}^{\mathrm{T}}X_k\|^2 + \|\hat{\beta}^{\mathrm{T}}X_j - \hat{\beta}^{\mathrm{T}}X_k\|^2)\right\}. \tag{12}$$

Remark 3. Unlike Bierens' (1982) test statistic ICM_n in (5), here we utilize the sufficient dimension reduction and projection techniques to reduce the original dimension p to $d+\hat{q}$. Furthermore, if the eigenvalues of $E(XX^{\mathrm{T}})$ are bounded away from infinity, then we have $E|\alpha^{\mathrm{T}}X|^2 = O(1)$ for any $\alpha \in \mathcal{S}^{p-1}$. Thus, it can be readily seen that $|\hat{B}_i^{\mathrm{T}}X_j|^2 = O_p(1)$ and $|\hat{\beta}_i^{\mathrm{T}}X_j|^2 = O_p(1)$, where $\{\hat{B}_i : 1 \leqslant i \leqslant \hat{q}\}$ and $\{\hat{\beta}_i : 1 \leqslant i \leqslant d\}$ are the columns of $\hat{B}$ and $\hat{\beta}$, respectively. Thus, no matter how large the dimension p, the exponential weight in (12) only relies on the number $d+q$. Furthermore, if $d+q$ is fixed, we will show that our test statistic $AICM_n$ would not degenerate to finite fixed values and the related wild bootstrap still works for our new test. In practice, when $d+q$ is much smaller than p, the dimensionality difficulty will be largely alleviated. The simulation results in § 6 validate our claims. However, if the structural dimension q of the underlying model is too large, which violates the purpose of dimension reduction for multiple-index model structure, the dimensionality issue comes back. Thus, we assume that the structural dimension q is fixed in this paper. We are working on a study for the large-q paradigm, although this is a challenging problem.

4.3. *Model adaptation property*

To achieve model adaptation, we need to identify the structural dimension q and the central mean subspace $\mathcal{S}_{E(Y|X)}$. When p is fixed, there are several methods available in the literature, such as principal Hessian directions (Li, 1992). However, when p is divergent, there are no corresponding asymptotic results. To overcome this difficulty, we consider the central subspace $\mathcal{S}_{Y|X}$ instead of the central mean subspace $\mathcal{S}_{E(Y|X)}$. The central subspace $\mathcal{S}_{Y|X}$ is defined as the intersection of all subspaces span(A) such that $Y \perp\!\!\!\perp X \mid A^{\mathrm{T}}X$. It is easy to see that $\mathcal{S}_{E(Y|X)} \subset \mathcal{S}_{Y|X}$. Thus, we further assume that $\mathcal{S}_{E(Y|X)} = \mathcal{S}_{Y|X}$. This can be achieved if the error term in the regression model has the same dimension reduction structure as the regression function. For the null model (1), if the error term satisfies $\varepsilon = \sigma_1(\beta_0^{\mathrm{T}}X)\tilde{\varepsilon}$ and $\tilde{\varepsilon} \perp\!\!\!\perp X$, then it can be readily seen that $\mathcal{S}_{E(Y|X)} = \mathcal{S}_{Y|X} = \mathrm{span}(\beta_0)$. Similarly, if $\varepsilon = \sigma_2(B^{\mathrm{T}}X)\tilde{\varepsilon}$ and $\tilde{\varepsilon} \perp\!\!\!\perp X$ in the alternative model (2), then it follows that $\mathcal{S}_{E(Y|X)} = \mathcal{S}_{Y|X} = \mathrm{span}(B)$.

To identify the central subspace $\mathcal{S}_{Y|X}$ in the diverging-dimension setting, we adopt cumulative slicing estimation (Zhu et al., 2010) in this paper, as it allows the divergence rate of p to be $p^2/n \to 0$ and is very easy to implement in practice. Cumulative slicing estimation uses the determining class of indicator functions to identify the central subspace. For simplicity, we first assume $E(X) = 0$ and $\mathrm{var}(X) = I_p$. Under the linearity condition (Li 1991), it can be shown that $E\{X \mid (Y \leqslant t)\} \in \mathcal{S}_{Y|X}$ for any $t \in \mathbb{R}$. This means we can obtain an infinite number of vectors in $\mathcal{S}_{Y|X}$. We then define the target matrix

$$M = \int_{-\infty}^{\infty} E\{X \mid (Y \leqslant t)\} E\{X^T I(Y \leqslant t)\}\, \mathrm{d}F_Y(t),$$

where $F_Y(t)$ is the cumulative distribution function of Y. Assuming the rank of the nonnegative matrix M is q, Zhu et al. (2010) showed that $\mathrm{span}(M) = \mathcal{S}_{Y|X}$. Let $\hat{\alpha}_t = n^{-1}\sum_{i=1}^n X_i I(Y_i \leqslant t)$; then, the sample version of M is

$$\hat{M} = n^{-1}\sum_{j=1}^n \hat{\alpha}_{Y_j}\hat{\alpha}_{Y_j}^{\mathrm{T}}.$$

If the structural dimension q is already known, then the estimator $\hat{B}(q)$ of B consists of the orthonormal eigenvectors corresponding to the largest q eigenvalues of $\hat{M}$. Here, B is the orthonormal matrix such that $\mathcal{S}_{Y|X} = \mathrm{span}(B)$.

Yet we need to estimate the unknown structural dimension q. For this, we suggest a minimum ridge-type eigenvalue ratio estimator to identify q that was first proposed by Zhu et al. (2017). Let $\{\hat{\lambda}_j, 1 \leqslant j \leqslant p\}$ and $\{\lambda_j, 1 \leqslant j \leqslant p\}$ be the eigenvalues of the matrix $\hat{M}$ and M, respectively. Assume that $\hat{\lambda}_{j+1} \leqslant \hat{\lambda}_j$ and $\lambda_{j+1} \leqslant \lambda_j$. As the rank of M is q, it follows that $\lambda_p = \cdots = \lambda_{q+1} = 0 < \lambda_q \leqslant \cdots \leqslant \lambda_1$. Thus, we can estimate the structural dimension q by

$$\hat{q} = \arg\min_{1\leqslant j\leqslant p}\left\{ j : \frac{\hat{\lambda}_{j+1}^2 + c_n}{\hat{\lambda}_j^2 + c_n} \right\}.$$

Here, $\hat{\lambda}_{p+1}^2 = 0$ and c_n is a sequence of positive numbers that will be specified later. The following result is a slight extension of Proposition 3 in Tan & Zhu (2019) that shows the consistency of the minimum ridge-type eigenvalue ratio estimator and the model adaptation to the underlying models.

PROPOSITION 1. *Assume that the regularity conditions of Theorem 3 in Zhu et al. (2010) hold. If* $c_n = \log n/n$ *and* $0 < c^{-1} \leqslant \lambda_q \leqslant \lambda_1 \leqslant c < \infty$ *for some fixed c, then: (i) under* H_0, $\mathbb{P}(\hat{q} = d) \to 1$ *and* $\|\hat{B}(d) - B\| = O_p(p/n)^{1/2}$; *(ii) under* H_1, $\mathbb{P}(\hat{q} = q) \to 1$ *and* $\|\hat{B}(q) - B\| = O_p(p/n)^{1/2}$. *Here, matrix B under the null hypothesis satisfies* $\mathrm{span}(B) = \mathrm{span}(\beta_0)$.

5. Asymptotic properties of the test statistic

5.1. *Limiting null distribution*

We first consider the asymptotic property of the test statistic $AICM_n$ under the null hypothesis. Recall that

$$AICM_n = \int_{\mathbb{R}^{d+\hat{q}}} \left| n^{-1/2} \sum_{j=1}^{n} \hat{\eta}_j \exp(\mathrm{i}t^{\mathrm{T}}\hat{\mathfrak{B}}^{\mathrm{T}}X_j) \right|^2 \varphi(t)\,\mathrm{d}t,$$

where $\hat{\eta}_j = Y_j - g(\hat{\beta}^{\mathrm{T}}X_j, \hat{\vartheta}_j)$. To facilitate the derivation of the asymptotic properties, we define the following empirical process:

$$\hat{V}_n^1(t) = n^{-1/2} \sum_{j=1}^{n} \hat{\eta}_j \{\cos(t^{\mathrm{T}}\hat{\mathfrak{B}}^{\mathrm{T}}X_j) + \sin(t^{\mathrm{T}}\hat{\mathfrak{B}}^{\mathrm{T}}X_j)\}.$$

If $\varphi(t) = \varphi(-t)$, then it follows that

$$AICM_n = \int_{\mathbb{R}^{d+\hat{q}}} |\hat{V}_n^1(t)|^2 \varphi(t)\,\mathrm{d}t.$$

By Proposition 1, $\mathbb{P}(\hat{q} = d) \to 1$ under the null hypothesis. Thus, we only need to work on the event $\{\hat{q} = d\}$. Consequently, $\hat{\mathfrak{B}} = \{\hat{\beta}_0, \hat{B}(d)\}$ and $AICM_n$ becomes

$$AICM_n = \int_{\mathbb{R}^{2d}} |\hat{V}_n^1(t)|^2 \varphi(t)\,\mathrm{d}t.$$

By Theorem 1 and the regularity conditions (A1)–(A5) and (B1)–(B3) in the Supplementary Material, we can show that, under the null hypothesis,

$$\begin{aligned} \hat{V}_n^1(t) &= n^{-1/2} \sum_{j=1}^{n} \varepsilon_j \{\cos(t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X_j) + \sin(t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X_j) - M(t)^{\mathrm{T}}\Sigma^{-1}g'(\theta_0, X_j)\} + R_n(t) \\ &=: V_n^1(t) + R_n(t), \end{aligned} \tag{13}$$

where $\mathfrak{B}_0 = (\beta_0, B)$, $M(t) = E[g'(\theta_0, X)\{\cos(t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X) + \sin(t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X)\}]$ and $R_n(t)$ is a remainder satisfying $\int_{\mathbb{R}^{2d}} |R_n(t)|^2\varphi(t)\,\mathrm{d}t = o_p(1)$. The proof of (13) will be given in the Supplementary Material. Altogether, we can obtain the following result.

Theorem 3. *Assume that the conditions (A1)–(A5) and (B1)–(B3) in the Supplementary Material hold. If $(p^3 \log n)/n \to 0$, then, under H_0, we have, in distribution,*

$$AICM_n \longrightarrow \int_{R^{2d}} |V_\infty^1(t)|^2 \varphi(t)\,\mathrm{d}t, \tag{14}$$

where $V_\infty^1(t)$ is a zero-mean Gaussian process with a covariance function $K(s,t)$ which is the pointwise limit of $K_n(s,t)$. Here, $K_n(s,t)$ is the covariance function of $V_n^1(t)$, that is,

$$\begin{aligned} K_n(s,t) &= \mathrm{cov}\{V_n^1(s), V_n^2(t)\} \\ &= E[\varepsilon^2\{\cos((s-t)^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X) + \sin((s+t)^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X)\}] \\ &\quad - M(t)^{\mathrm{T}}\Sigma^{-1}E[\varepsilon^2\{\cos(s^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X) + \sin(s^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X)\}g'(\theta_0, X)] \end{aligned}$$

$$
\begin{aligned}
&- M(s)^{\mathrm{T}}\Sigma^{-1}E[\varepsilon^2\{\cos(t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X) + \sin(t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X)\}g'(\theta_0, X)] \\
&+ M(s)^{\mathrm{T}}\Sigma^{-1}E\{\varepsilon^2 g'(\theta_0, X)XX^{\mathrm{T}}\}\Sigma^{-1}M(t).
\end{aligned}
$$

For single-index null models, Tan & Zhu (2019) showed that the residual marked empirical process involved in their test statistic converges to a Gaussian process under the rate $p = o(n^{1/5})$. This rate can be improved to $p = o(n^{1/3}/\log n)$ for linear models. In the current paper, we improve this divergence rate to $p = o(n/\log n)^{1/3}$ for more general multiple-index models. Tan & Zhu (2019) conjectured that the leading term $n^{1/3}$ would be close to optimal (Tan & Zhu, 2019, Remark 2). Chen & Lockhart (2001) showed that for linear regression models, the residual empirical process admits a uniform asymptotic linear representation, and thus converges to a zero-mean Gaussian process provided $p = o(n^{1/3}/\log^{2/3} n)$. They also gave an example to show that this rate cannot be improved, in general. This makes the conjecture more reasonable, although, for the residual marked empirical process, we still do not know whether this example can work in our setting.

5.2. *Limiting distribution under the alternative hypotheses*

Now we discuss the asymptotic property of $AICM_n$ under the alternative hypotheses. Consider the following sequence of alternative hypotheses:

$$
H_{1n}: \ Y_n = g(\beta_0^{\mathrm{T}}X, \vartheta_0) + r_n G(B^{\mathrm{T}}X) + \varepsilon,
$$

where $E(\varepsilon \mid X) = 0$, $G(B^{\mathrm{T}}X)$ is a random variable satisfying $E\{G(B^{\mathrm{T}}X)\} = 0$ and $\mathbb{P}\{G(B^{\mathrm{T}}X) = 0\} < 1$. The convergence rate r_n satisfies $r_n = n^{-1/2}$ or $r_n n^{1/2} \to \infty$. To obtain the asymptotic distribution of $AICM_n$ under the alternatives H_{1n}, we first derive the asymptotic properties of the estimators $\hat{q}$ and $\hat{\beta}$ when the dimension p diverges to infinity.

PROPOSITION 2. *Suppose that the regularity conditions of Theorem 3 in Zhu et al. (2010) hold. If $pr_n \to 0$, $c_n = \log n/n$ and $0 < c^{-1} \leqslant \lambda_q \leqslant \lambda_1 \leqslant c < \infty$ for some fixed c, then, under H_{1n}, we have $\mathbb{P}(\hat{q} = d) \to 1$ and $\|\hat{B}(d) - B_L\| = O_p(p^{1/2}r_n)$. Here, B_L is a $p \times d$ orthonormal matrix satisfying* $\mathrm{span}(B_L) = \mathrm{span}(\beta_0)$.

It is worth mentioning that under the local alternatives H_{1n} with $r_n = o(1/p)$, the estimated structural dimension $\hat{q}$ is not equal to the true structural dimension, but to d asymptotically. This means $\hat{q}$ is not a consistent estimator of the true structural dimension in this case. A special case is that if $r_n = n^{-1/2}$, it follows that $\mathbb{P}(\hat{q} = d) \to 1$ and $\|\hat{B}(d) - B_L\| = O_p(p/n)^{1/2}$. Yet we need to derive the asymptotic properties of $(\hat{\beta}, \hat{\vartheta})$ with respect to (β_0, ϑ_0). Recall that $\hat{\theta} = \{\mathrm{vec}(\hat{\beta})^{\mathrm{T}}, \hat{\vartheta}^{\mathrm{T}}\}^{\mathrm{T}}$ and $\theta_0 = \{\mathrm{vec}(\beta_0)^{\mathrm{T}}, \vartheta_0^{\mathrm{T}}\}^{\mathrm{T}}$, where $\mathrm{vec}(\beta) = \{\beta_1^{\mathrm{T}}, \ldots, \beta_d^{\mathrm{T}}\}^{\mathrm{T}}$.

THEOREM 4. *Suppose that H_{1n} and the conditions (A1)–(A4) in the Supplementary Material hold. If $pr_n \to 0$, then $\hat{\theta}$ is a norm-consistent estimator for θ_0 with $\|\hat{\theta} - \theta_0\| = O_p(p^{1/2}r_n)$. Moreover, if the conditions (A1)–(A5) in the Supplementary Material hold and $np^3r_n^4 \log n \to 0$, then under the alternatives H_{1n}, we have*

$$
\begin{aligned}
n^{1/2}\alpha^{\mathrm{T}}(\hat{\theta} - \theta_0) = n^{-1/2}\sum_{j=1}^{n} \varepsilon_j \alpha^{\mathrm{T}}\Sigma^{-1}g'(\theta_0, X_j) \\
+ n^{1/2}r_n\alpha^{\mathrm{T}}\Sigma^{-1}E\{g'(\theta_0, X)G(B^{\mathrm{T}}X)\} + o_p(1),
\end{aligned}
$$

where $\Sigma = E\{g'(\theta_0, X)g'(\theta_0, X)^{\mathrm{T}}\}$ and the term $o_p(1)$ is uniform over $\alpha \in \mathcal{S}^{p-1}$.

Under the alternatives H_{1n} with $r_n = n^{-1/2}$, if $p^2/n \to 0$, then we have that $\hat{\theta}$ is a norm-consistent estimator for θ_0 with $\|\hat{\theta} - \theta_0\| = O_p(p/n)^{1/2}$. Moreover, if $(p^3 \log n)/n \to 0$, it follows that

$$n^{1/2}\alpha^{\mathrm{T}}(\hat{\theta} - \theta_0) = n^{-1/2}\sum_{j=1}^{n}\varepsilon_j\alpha^{\mathrm{T}}\Sigma^{-1}g'(\theta_0, X_j) + \alpha^{\mathrm{T}}\Sigma^{-1}E\{g'(\theta_0, X)G(B^{\mathrm{T}}X)\} + o_p(1).$$

THEOREM 5. *Assume that the conditions (A1)–(A5) and (B1)–(B3) in the Supplementary Material hold.*

(i) If $(p^3 \log n)/n \to 0$*, under the global alternative* H_1 *we have, in probability,*

$$\frac{1}{n}\int_{\mathbb{R}^{d+\hat{q}}} |\hat{V}_n(t)|^2\varphi(t)\,\mathrm{d}t \longrightarrow C_1,$$

where C_1 *is a positive constant.*

(ii) If $np^3 r_n^4 \log n \to 0$*, under the local alternatives* H_{1n} *with* $n^{1/2}r_n \to \infty$ *we have, in probability,*

$$\frac{1}{nr_n^2}\int_{\mathbb{R}^{d+\hat{q}}} |\hat{V}_n(t)|^2\varphi(t)\,\mathrm{d}t \longrightarrow C_2,$$

where C_2 *is a positive constant.*

(iii) If $(p^3 \log n)/n \to 0$*, under the local alternatives* H_{1n} *with* $r_n = n^{-1/2}$ *we have, in distribution,*

$$\int_{\mathbb{R}^{d+\hat{q}}} |\hat{V}_n(t)|^2\varphi(t)\,\mathrm{d}t \longrightarrow \int_{\mathbb{R}^{2d}} |V_\infty^1(t) + L_1(t) - L_2(t)|^2\varphi(t)\,\mathrm{d}t,$$

where V_∞^1 *is a zero-mean Gaussian process given by (14), and* $L_1(t)$ *and* $L_2(t)$ *are the uniform limits of* $L_{n1}(t)$ *and* $L_{n2}(t)$*, respectively. The functions* $L_{n1}(t)$ *and* $L_{n2}(t)$ *are*

$$L_{n1}(t) = M^{\mathrm{T}}(t)\Sigma^{-1}E\{g'(\theta_0, X)G(B^{\mathrm{T}}X)\},$$
$$L_{n2}(t) = E[G(B^{\mathrm{T}}X)\{\cos(t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X) + \sin(t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X)\}].$$

It follows from Theorem 5 that under the global alternative H_1 and the local alternatives H_{1n} with $n^{1/2}r_n \to \infty$, our test statistic $AICM_n$ diverges to infinity at the rate of n and nr_n^2, respectively. For the local alternative with $r_n = n^{-1/2}$, it is readily seen that if $L_1(t) - L_2(t) \not\equiv 0$, the proposed test $AICM_n$ is still sensitive to the local alternatives distinct from the null at the rate of $n^{-1/2}$. Recall that $M(t) = E[g'(\theta_0, X)\{\cos(t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X) + \sin(t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X)\}]$; then we have

$$L_{n1}(t) = \xi_G^{\mathrm{T}}E[g'(\theta_0, X)\{\cos(t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X) + \sin(t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X)\}]$$

with $\xi_G = \Sigma^{-1}E\{g'(\theta_0, X)G(B^{\mathrm{T}}X)\}$. Consequently,

$$\int_{\mathbb{R}^{2d}} |L_{n1}(t) - L_{n2}(t)|^2\varphi(t)\,\mathrm{d}t = \int_{\mathbb{R}^{2d}} |E[\{G(B^{\mathrm{T}}X) - \xi_G^{\mathrm{T}}g'(\theta_0, X)\}\exp(\mathrm{i}t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X)]|^2\varphi(t)\,\mathrm{d}t.$$

By the Fourier reversal formula, we have

$$\begin{aligned} L_{n1}(t) - L_{n2}(t) = o(1) &\iff E[\{G(B^{\mathrm{T}}X) - \xi_G^{\mathrm{T}}g'(\theta_0, X)\}\exp(\mathrm{i}t^{\mathrm{T}}\mathfrak{B}_0^{\mathrm{T}}X)] = o(1) \\ &\iff E\{G(B^{\mathrm{T}}X) - \xi_G^{\mathrm{T}}g'(\theta_0, X) \mid \mathfrak{B}_0^{\mathrm{T}}X\} = o(1) \quad \text{almost surely.} \end{aligned}$$

Since $L_1(t) - L_2(t)$ is the uniform limit of $L_{n1}(t) - L_{n2}(t)$, the proposed test $AICM_n$ is able to asymptotically detect any local alternative converging to the null with a parametric convergence rate $n^{-1/2}$ if $E\{G(B^{\mathrm{T}}X) \mid \mathfrak{B}_0^{\mathrm{T}}X\}$ is not parallel to $E\{g'(\theta_0, X) \mid \mathfrak{B}_0^{\mathrm{T}}X\}$ upon an infinitesimal term.

5.3. *Bootstrap approximation*

The limiting null distribution of our test statistic $AICM_n$ depends on the unknown parameters β_0 and the matrix B, and thus is not tractable for critical value determination. Although we have shown that the wild bootstrap is not valid for Bierens' original ICM test when p is divergent, we now show that the wild bootstrap still works for the proposed test in diverging-dimension settings. Recall that the bootstrap sample $\{(X_j, Y_j^*)\}_{j=1}^n$ is

$$X_j^* = X_j, \qquad Y_j^* = g(\hat{\theta}, X_j) + \hat{\eta}_j V_j^* \quad (j = 1, \dots, n),$$

where $\hat{\eta}_j = Y_j - g(\hat{\theta}, X_j)$ and $\{V_j^*\}_{j=1}^n$ are the independent and identically distributed Bernoulli variates specified in § 3. Let $\hat{\theta}^*$ be the bootstrap estimator obtained by least squares based on the bootstrap sample $\{(X_j, Y_j^*)\}_{j=1}^n$. Then, we approximate the limiting null distribution of $AICM_n$ by that of

$$AICM_n^* = \int_{\mathbb{R}^{d+\hat{q}}} |\hat{V}_n^*(t)|^2 \varphi(t)\, \mathrm{d}t,$$

where $\hat{V}_n^*(t) = n^{-1/2} \sum_{j=1}^n \{Y_j^* - g(\hat{\theta}^*, X_j)\} \exp(\mathrm{i}t^{\mathrm{T}}\hat{\mathfrak{B}}^{\mathrm{T}}X_j)$. To determine critical values in practice, repeat the resampling process a large number of times, say b times. For a given nominal level τ, the critical value is determined by the upper τ-quantile of the bootstrap distribution $\{\hat{V}_{nk}^* : k = 1, \dots, b\}$.

In the next theorem we establish the validity of the wild bootstrap for the new test.

THEOREM 6. *Suppose that the bootstrap sample is generated from the wild bootstrap, and conditions (A1)–(A5) and (B1)–(B3) in the Supplementary Material hold.*

(i) If $(p^3 \log n)/n \to 0$, under the null hypothesis H_0 or under the alternative hypothesis with $r_n = n^{-1/2}$ we have, in probability,

$$AICM_n^* \longrightarrow \int_{\mathbb{R}^{2d}} |V_\infty^{1*}(t)|^2 \varphi(t)\, \mathrm{d}t \qquad \text{in distribution,}$$

where V_∞^{1*} have the same distribution as the Gaussian process V_∞^1 given in Theorem 3.

(ii) If $np^3 r_n^4 \log n \to 0$, under the local alternatives H_{1n} with $\sqrt{n}r_n \to \infty$, the result in (i) continues to hold.

(iii) If $(p^3 \log n)/n \to 0$, under the global alternative H_1, the distribution of $AICM_n^*$ converges to a finite limiting distribution which may be different from the limiting null distribution.

6. NUMERICAL STUDIES

6.1. *Simulations*

In this subsection we conduct some numerical studies to examine the performance of the proposed test in finite-sample cases. From the theoretical results in this paper, we set $p = [3n^{1/3}] - 5$ with the sample sizes $n = 100, 200, 400$ and 600, and make a comparison in Studies 1–3 among our test $AICM_n$, Bierens' (1982) ICM_n test, Zheng's (1996) T_n^{ZH} test, Stute & Zhu's (2002) T_n^{SZ} test, Escanciano's (2006a) $PCvM_n$ test, the Guo et al. (2016) T_n^{GWZ} test, and Tan & Zhu's (2019)

ACM_n test, although most of these dealt with fixed dimensions. Since the tests ICM_n and $PCvM_n$ are not asymptotically distribution-free in fixed-dimension cases, we adopt the wild bootstrap to determine the critical values as suggested by Escanciano (2006a) and Lavergne & Patilea (2008, 2012). To give a relatively thorough comparison among these tests, we also consider the fixed-dimension cases with $p = 2, 4, 6$ and $n = 100$ in the first two studies, and then some larger dimensions in Study 3. In Study 4, we compare these tests with the more recent goodness-of-fit test GRP_n proposed by Janková et al. (2020) in the logistic regression setting.

In the simulations that follow, $a = 0$ corresponds to the null and $a \neq 0$ to the alternatives. The significance level is $\alpha = 0.05$. The simulation results are based on the average of 1000 replications and the bootstrap approximation of $b = 500$ replications.

Study 1. Consider the following regression models:

$$
\begin{aligned}
H_{11} &: Y = \beta_0^{\mathrm{T}} X + a \exp\{-(\beta_0^{\mathrm{T}} X)^2\} + \varepsilon,\\
H_{12} &: Y = \beta_0^{\mathrm{T}} X + a \cos(0.6\pi \beta_0^{\mathrm{T}} X) + \varepsilon,\\
H_{13} &: Y = \beta_1^{\mathrm{T}} X + a(\beta_2^{\mathrm{T}} X)^2 + \varepsilon,\\
H_{14} &: Y = \beta_1^{\mathrm{T}} X + a \exp(\beta_2^{\mathrm{T}} X) + \varepsilon,
\end{aligned}
$$

where $\beta_0 = (1, \ldots, 1)^{\mathrm{T}}/\sqrt{p}$, $\beta_1 = (\underbrace{1, \ldots, 1}_{p_1}, 0, \ldots, 0)^{\mathrm{T}}/\sqrt{p_1}$ and $\beta_2 = (0, \ldots, 0, \underbrace{1, \ldots, 1}_{p_1})/\sqrt{p_1}$ with $p_1 = [p/2]$. The covariate X is $N(0, \Sigma_1)$ or $N(0, \Sigma_2)$, independent of the standard Gaussian error term ε. Here, $\Sigma_1 = I_p$ and $\Sigma_2 = (1/2^{|i-j|})_{p\times p}$. H_{12} is a high-frequency model and the others are low-frequency models. The structure dimension $q = 1$ under both the null and alternative hypotheses in the models H_{11} and H_{12}, while the structure dimension $q = 2$ under the alternative hypothesis in models H_{13} and H_{14}.

The empirical sizes and powers of H_{11} are presented in Table 1. The remaining simulation results are relegated to the Supplementary Material. First, we can see that when the dimension p is small, Bierens' (1982) original ICM_n test performs very well in all models. However, it breaks down in the large-dimension scenarios. Furthermore, this phenomenon seems unaffected by the correlated structure of the predictor vector X. For the other tests, we observe that $AICM_n$, ACM_n, T_n^{SZ} and T_n^{GWZ} can control the empirical sizes very well in all models and dimension cases. The empirical sizes of $PCvM_n$ are also close to the significance level, but slightly unstable in some cases, while T_n^{ZH} cannot maintain the significance level in most cases and is generally conservative with smaller empirical sizes. For the empirical power, we can see that $AICM_n$, ACM_n, T_n^{SZ} and $PCvM_n$ all perform very well for the low-frequency models H_{11}, H_{13} and H_{14}, whereas T_n^{GWZ} behaves slightly worse for these three low-frequency models. For the high-frequency model H_{12}, the test T_n^{GWZ} beats all other competitors except our new test $AICM_n$. This is somewhat surprising as local smoothing tests such as T_n^{GWZ} usually perform better for high-frequency models, and global smoothing tests work better for low-frequency models. While our new test can be viewed as a global smoothing test, it also seems to work well for high-frequency models. In contrast, Zheng's (1996) T_n^{ZH} test, which is a typical local smoothing test, has very small empirical powers in most cases when the dimension is large. This validates the well-known results that the traditional local smoothing tests suffer severely from the curse of dimensionality.

The hypothetical models in Study 1 are all single-index models. Next we consider multiple-index models in the second simulation study. As the tests ACM_n, T_n^{SZ} and T_n^{GWZ} only dealt with parametric single-index models, we only compare our new test with $PCvM_n$, ICM_n and T_n^{ZH}.

Table 1. *Empirical sizes and powers of the tests for H_{11} in Study 1*

	a	$n = 100$, $p = 2$	$n = 100$, $p = 4$	$n = 100$, $p = 6$	$n = 100$, $p = 8$	$n = 200$, $p = 12$	$n = 400$, $p = 17$	$n = 600$, $p = 20$
$AICM_n$, Σ_1	0.00	0.0530	0.0510	0.0540	0.0580	0.0540	0.0410	0.0560
	0.25	0.3660	0.3440	0.3560	0.3670	0.6260	0.8710	0.9770
ACM_n, Σ_1	0.00	0.0600	0.0440	0.0505	0.0440	0.0495	0.0570	0.0510
	0.25	0.2985	0.2645	0.2675	0.2720	0.4960	0.8125	0.9335
$PCVM_n$, Σ_1	0.00	0.0540	0.0470	0.0620	0.0660	0.0530	0.0540	0.0650
	0.25	0.3170	0.3020	0.2910	0.3080	0.5140	0.8330	0.9350
T_n^{SZ}, Σ_1	0.00	0.0500	0.0435	0.0550	0.0475	0.0450	0.0450	0.0450
	0.25	0.2700	0.2880	0.2760	0.2665	0.4945	0.7970	0.9315
ICM_n, Σ_1	0.00	0.0610	0.0460	0.0200	0.0020	0.0000	0.0000	0.0000
	0.25	0.3220	0.2540	0.1500	0.0200	0.0000	0.0000	0.0000
T_n^{GWZ}, Σ_1	0.00	0.0505	0.0565	0.0505	0.0345	0.0580	0.0490	0.0605
	0.25	0.2800	0.2650	0.2515	0.2550	0.4435	0.7570	0.9100
T_n^{ZH}, Σ_1	0.00	0.0395	0.0370	0.0320	0.0305	0.0210	0.0250	0.0285
	0.25	0.1920	0.0855	0.0685	0.0490	0.0370	0.0225	0.0255
$AICM_n$, Σ_2	0.00	0.0530	0.0530	0.0460	0.0650	0.0510	0.0460	0.0550
	0.25	0.2950	0.2490	0.2400	0.2250	0.4310	0.6900	0.8570
ACM_n, Σ_2	0.00	0.0445	0.0395	0.0475	0.0455	0.0500	0.0520	0.0520
	0.25	0.2280	0.1805	0.1745	0.1800	0.2730	0.4935	0.6600
$PCVM_n$, Σ_2	0.00	0.0460	0.0600	0.0600	0.0740	0.0630	0.0510	0.0600
	0.25	0.2820	0.2090	0.2030	0.2070	0.2910	0.4940	0.6670
T_n^{SZ}, Σ_2	0.00	0.0545	0.0405	0.0425	0.0535	0.0555	0.0450	0.0505
	0.25	0.2350	0.2015	0.1805	0.1545	0.2965	0.4905	0.6780
ICM_n, Σ_2	0.00	0.0490	0.0500	0.0300	0.0030	0.0000	0.0000	0.0000
	0.25	0.2740	0.2130	0.1370	0.0360	0.0000	0.0000	0.0000
T_n^{GWZ}, Σ_2	0.00	0.0540	0.0595	0.0555	0.0490	0.0525	0.0500	0.0530
	0.25	0.2315	0.1755	0.1680	0.1690	0.2675	0.5045	0.6585
T_n^{ZH}, Σ_2	0.00	0.0385	0.0360	0.0300	0.0280	0.0305	0.0301	0.0245
	0.25	0.1850	0.0845	0.0560	0.0460	0.0525	0.0390	0.0375

Study 2. Generate data from the following models:

$$
\begin{aligned}
H_{21} &: Y = \beta_1^{\mathrm{T}} X + \exp(\beta_2^{\mathrm{T}} X) + a(\beta_2^{\mathrm{T}} X)^2 + \varepsilon, \\
H_{22} &: Y = \beta_1^{\mathrm{T}} X + \exp(\beta_2^{\mathrm{T}} X) + a\cos(0.6\pi\beta_2^{\mathrm{T}} X) + \varepsilon, \\
H_{23} &: Y = \beta_1^{\mathrm{T}} X + \exp(\beta_2^{\mathrm{T}} X) + a(\beta_1^{\mathrm{T}} X)(\beta_2^{\mathrm{T}} X) + \varepsilon,
\end{aligned}
$$

where β_1, β_2, ε and X are the same as in Study 1.

The simulation results are presented in the Supplementary Material. We can observe that our test $AICM_n$ and $PCvM_n$ perform much better than the other two, while Bierens' (1982) test ICM_n again does not work at all, and Zheng's (1996) test T_n^{ZH} cannot maintain the nominal level and has no empirical powers in both cases. For the empirical size, both our $AICM_n$ and $PCvM_n$ tests are slightly conservative with larger empirical sizes. This may be due to the inaccurate estimation of the related parameters when p is large. The empirical powers of $AICM_n$ and $PCvM_n$ both grow quickly under both the low-frequency model H_{21} and the high-frequency model H_{22}. In model H_{23}, our test has much better power performance than $PCvM_n$.

In practice, it is difficult to determine the asymptotic mechanisms between the dimension p and the sample size n. Thus, we conduct a further simulation study under the paradigm $p = 0.1n$

with $n = 50, 100, 200, 500, 1000$ to provide more information on when we may use the test in the large-dimensional cases.

Study 3. The data are generated from the following models:

$$H_{31} : Y = \beta_0^{\mathrm{T}} X + a \exp(\beta_0^{\mathrm{T}} X) + \varepsilon,$$
$$H_{32} : Y = \beta_1^{\mathrm{T}} X + \exp(\beta_2^{\mathrm{T}} X) + a \exp(-\beta_0^{\mathrm{T}} X) + \varepsilon,$$

where β_0, β_1, β_2, ε and X are the same as in Study 1.

The empirical sizes and powers are reported in the Supplementary Material. We can see that Bierens' (1982) test ICM_n and Zheng's (1996) test T_n^{ZH} can maintain the significance level occasionally when the dimension p is relatively small. When the dimension p becomes large, the empirical sizes of these two tests are completely out of control and the powers make no sense. For the model H_{31}, we can observe that ACM_n, T_n^{SZ} and T_n^{GWZ} still work very well even when $p = 100$. These three tests are dimension-reduction-based tests which are designed to test single-index models. In contrast, although $PCvM_n$ and the new test $AICM_n$ have high empirical powers, they cannot maintain the significance level when the dimension p increases. This phenomenon suggests that the new test may not be reliable under the paradigm $p = 0.1n$, and some new deep theories should be developed to handle this high-dimension regime.

Finally, we compare the performance of the tests mentioned above against the test GRP_n proposed by Janková et al. (2020). Although GRP_n is designed for testing for generalized linear models in the fixed design setting, it can also be applied in the random design setting.

Study 4. The simulated data are from the logistic regression model with $(n, p) = (300, 10)$ and $(300, 15)$ according to $Y \mid X \sim \mathrm{Ber}\{\mu(\beta_0^{\mathrm{T}} X + ag(X))\}$,

where $\mu(z) = 1/(1 + \exp(-z))$. Consider three different cases for the misspecification $g(\cdot)$:

$$H_{41} : \; g(X) = (\beta_0^{\mathrm{T}} X)^2,$$
$$H_{42} : \; g(X) = 2(\beta_1^{\mathrm{T}} X)^2,$$
$$H_{43} : \; g(X) = 2(\beta_1^{\mathrm{T}} X)(\beta_2^{\mathrm{T}} X).$$

Here, $\beta_0 = (1, 1, 1, 0, \ldots, 0)^{\mathrm{T}}$, $\beta_1 = (1, 0, \ldots, 0)^{\mathrm{T}}$ and $\beta_2 = (0, 1, 0, \ldots, 0)^{\mathrm{T}}$. The predictor X follows an $N(0, I_p)$ or $N(0, \Sigma_3)$ distribution, where $\Sigma_3 = (0.6^{|i-j|})_{p \times p}$. This simulation study in the cases (2) and (3) with $(n, p) = (300, 10)$ was also considered in § 4.1 in Janková et al. (2020), while $(n, p) = (300, 15)$ coincides with the setting with $p = [3n^{1/3}] - 5$.

The simulation results are presented in the Supplementary Material. We can observe that most of the tests perform very well when the dimension p is relatively small for all simulation models. When p is large, the empirical powers of the ICM test ICM_n and Zheng's (1996) test T_n^{ZH} drop very quickly, except for the model H_{41} with the covariance matrix Σ_3. We also find that GRP_n is slightly conservative with smaller empirical sizes. For the models H_{41} and H_{42}, the empirical powers of GRP_n grow slightly slower than our test, but for the model H_{43} GRP_n beats all other competitors.

6.2. *A real-data example*

In this subsection we apply the proposed test to the CSM dataset that was first analysed by Ahmed et al. (2015), which can be obtained through `https://archive.ics.uci.edu`. There are 217 observations in the original dataset and 30 observations with missing response and/or covariates. To apply our test, we first clean the data points of missing values, leaving 187 complete observations. The response value Y is Gross Income. There are 11 predictor variables:

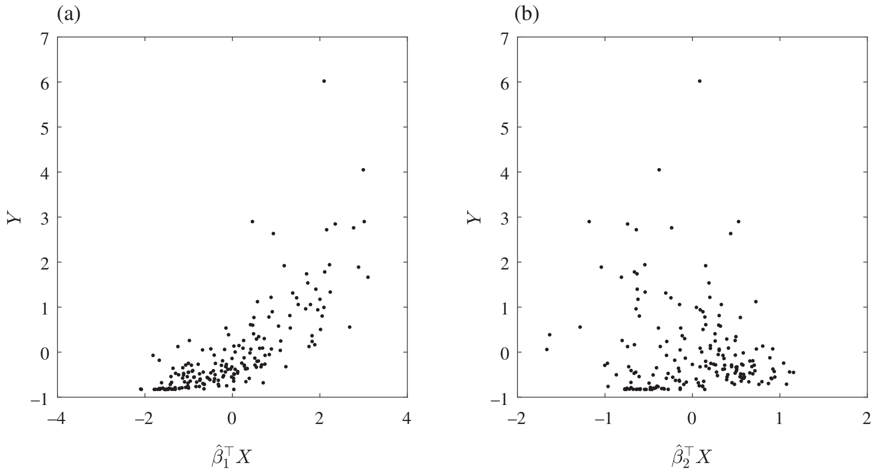

Fig. 1. (a) Scatterplot of Y versus the projected covariate $\hat{\beta}_1^{\mathrm{T}}X$. (b) Scatterplot of Y versus the projected covariate $\hat{\beta}_2^{\mathrm{T}}X$.

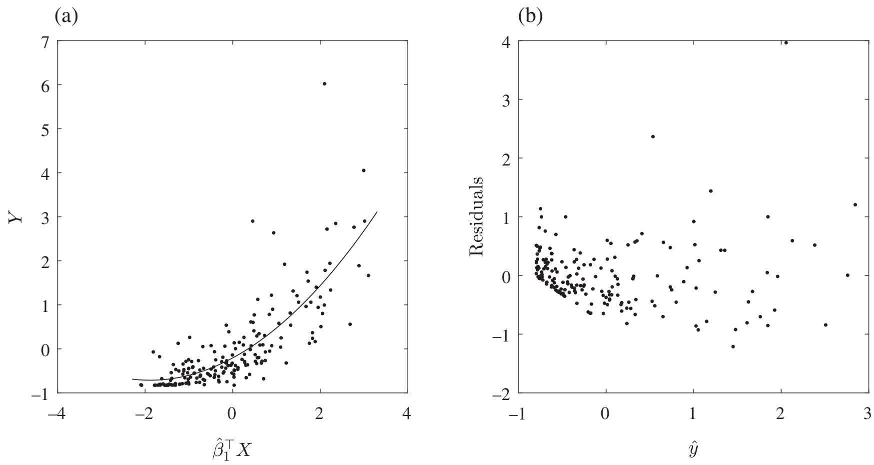

Fig. 2. (a) Scatterplot of Y versus the fitted quadratic polynomial curve. (b) Scatterplot of residuals from the model (15) versus the fitted values.

Rating X_1, Genre X_2, Budget X_3, Screens X_4, Sequel X_5, Sentiment X_6, Views X_7, Likes X_8, Dislikes X_9, Comments X_{10} and Aggregate Followers X_{11}. For easy explanation, we standardize all variables separately. To establish the relationship between the response Y and the predictor vector $X = (X_1, \ldots, X_{13})^{\mathrm{T}}$, we first check whether a linear regression model is adequate to fit this dataset. When the newly proposed test is applied, we find that the value of the test statistic is $AICM_n = 0.7435$ and the p-value is about 0.046. Thus, a linear regression model may not be adequate to predict the response. We may consider a more plausible model. To this end, we apply the sufficient dimension reduction techniques to the dataset. When cumulative slicing estimation is used, we find that the estimated structural dimension of this dataset is $\hat{q} = 1$, which indicates that Y may be conditionally independent of X given the projected predictor vector $\hat{\beta}_1^{\mathrm{T}}X$, where $\hat{\beta}_1 = (0.3186, 0.0750, 0.5351, 0.5384, 0.0735, -0.0546, -0.2497, 0.2082, 0.3813, -0.0532, 0.2333)^{\mathrm{T}}$ is the first direction obtained by cumulative slicing estimation. A scatterplot of Y against $\hat{\beta}_1^{\mathrm{T}}X$

is presented in Fig. 1(a), and suggests that a linear relationship between Y and X may not be reasonable.

To further explore the relationship between Y and X, we consider a more thorough search for the projected predictor vectors. Consider the second projected predictor vector $\hat{\beta}_2^{\mathrm{T}} X$ obtained by cumulative slicing estimation. A scatterplot of Y against $\hat{\beta}_2^{\mathrm{T}} X$ is presented in Fig. 1(b), from which it seems that there exists no obvious trend between the predictor Y and the second projected direction $\hat{\beta}_2^{\mathrm{T}} X$. This may indicate that the projection of the data onto the space $\hat{\beta}_1^{\mathrm{T}} X$ already contains almost all the regression information of the model structure. Further, from Fig. 1(a), it seems that there exists a quadratic relationship between Y and $\hat{\beta}_1^{\mathrm{T}} X$. Thus, we consider the following model to fit this dataset:

$$Y = \theta_1 + \theta_2(\beta^{\mathrm{T}} X) + \theta_3(\beta^{\mathrm{T}} X)^2 + \varepsilon. \tag{15}$$

When applying the proposed test to model (15), the value of the test statistic is $AICM_n = 0.4134$ and the p-value is about 0.646. This shows that the model (15) may be useful to fit this dataset. To further visualize this fit, Fig. 2(a) adds the fitted curve on the scatterplot. Figure 2(b) presents a scatterplot of residuals from the model (15) against the fitted values. We can see that there is no trend between the residuals and the fitted value. Thus, this model is plausible.

Acknowledgement

We thank the editor, associate editor and two referees for their constructive comments. Tan was supported by the National Natural Science Foundation of China (12071119) and the Natural Science Foundation of Hunan Province (2020JJ4162). Zhu was supported by the Research Grants Council of Hong Kong.

Supplementary material

Supplementary Material available at *Biometrika* online includes the regularity conditions, proofs of the theoretical results and additional simulation results.

References

Ahmed, M., Jahangir, M., Afzal, H., Majeed, A. & Siddiqi, I. (2015). Using crowd-source based features from social media and conventional features to predict the movies popularity. In *Proc. IEEE Int. Conf. Smart City/SocialCom/SustainCom (SmartCity)*, pp. 273–8.

Bierens, H. J. (1982). Consistent model specification tests. *J. Economet.* **20**, 105–34.

Bierens, H. J. & Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* **65**, 1129–51.

Chen, G. & Lockhart, R. A. (2001). Weak convergence of the empirical process of residuals in linear models with many parameters. *Ann. Statist.* **29**, 748–62.

Cook, R. D. & Li, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30**, 455–74.

Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* **27**, 1012–40.

Dominguez, M. A. (2005). On the power of bootstrapped specification tests. *Economet. Rev.* **23**, 215–28.

Escanciano, J. C. (2006a). A consistent diagnostic test for regression models using projections. *Economet. Theory* **22**, 1030–51.

Escanciano, J. C. (2006b). Goodness-of-fit tests for linear and nonlinear time series models. *J. Am. Statist. Assoc.* **101**, 531–41.

Escanciano, J. C. (2009). On the lack of power of omnibus specification tests. *Economet. Theory* **25**, 162–94.

Fan, J. & Huang, L. (2001). Goodness-of-fit tests for parametric regression models. *J. Am. Statist. Assoc.* **96**, 640–52.

Guo, X., Wang, T. & Zhu, L. (2016). Model checking for generalized linear models: a dimension-reduction model-adaptive approach. *J. R. Statist. Soc.* B **78**, 1013–35.

HÄRDLE, W. K. & MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926–47.

HLÁVKA, Z., HUSKOVA, M., KIRCH, C. & MEINTANIS, S. G. (2017). Fourier-type tests involving martingale difference processes. *Economet. Rev.* **36**, 468–92.

HOROWITZ, J. L. & SPOKOINY, V. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69**, 599–631.

HUBER, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1**, 799–821.

JANKOVÁ, J., SHAH, R. D., BÜHLMANN, P. & SAMWORTH, R. J. (2020). Goodness-of-fit testing in high dimensional generalized linear models. *J. R. Statist. Soc.* B **82**, 773–95.

KHMALADZE, E. V. & KOUL, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32**, 995–1034.

KOUL, H. L. & NI, P. (2004). Minimum distance regression model checking. *J. Statist. Plan. Infer.* **119**, 109–41.

LAVERGNE, P. & PATILEA, V. (2008). Breaking the curse of dimensionality in non-parametric testing. *J. Economet.* **143**, 103–22.

LAVERGNE, P. & PATILEA, V. (2012). One for all and all for one: regression checks with many regressors. *J. Bus. Econ. Statist.* **30**, 41–52.

LI, K. C. (1991). Sliced inverse regression for dimension reduction. *J. Am. Statist. Assoc.*, **86**, 316–27.

LI, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma. *J. Am. Statist. Assoc.* **87**, 1025–39.

MAMMEN, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.* **21**, 255–85.

PORTNOY, S. (1984). Asymptotic behavior of m-estimators of p regression parameters when p^2/n is large. I. Consistency. *Ann. Statist.* **12**, 1298–309.

PORTNOY, S. (1985). Asymptotic behavior of m estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Ann. Statist.* **13**, 1403–17.

SAMORODNITSKY, G. & TAQQU, M. (1994). *Stable Non-Gaussian Random Processes*. New York: Chapman and Hall.

SHAH, R. D. & BÜHLMANN, P. (2018). Goodness-of-fit tests for high dimensional linear models. *J. R. Statist. Soc.* B **80**, 113–35.

SHI, C., SONG, R., CHEN, Z. & LI, R. (2019). Linear hypothesis testing for high dimensional generalized linear models. *Ann. Statist.* **47**, 2671–703.

STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25**, 613–41.

STUTE, W., GONZÁLEZ MANTEIGA, W. & PRESEDO QUINDIMIL, M. (1998a). Bootstrap approximations in model checks for regression. *J. Am. Statist. Assoc.* **93**, 141–9.

STUTE, W., THIES, S. & ZHU, L. (1998b). Model checks for regression: an innovation process approach. *Ann. Statist.* **26**, 1916–34.

STUTE, W., XU, W. L. & ZHU, L. (2008). Model diagnosis for parametric regression in high-dimensional spaces. *Biometrika* **95**, 451–67.

STUTE, W. & ZHU. L. X. (2002). Model checks for generalized linear models. *Scand. J. Statist.* **29**, 535–45.

TAN, F. & ZHU, L. (2019). Adaptive-to-model checking for regressions with diverging number of predictors. *Ann. Statist.* **47**, 1960–94.

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**, 401–15.

ZHENG, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Economet.* **75**, 263–89.

ZHU, L. (2003). Model checking of dimension-reduction type for regression. *Statist. Sinica* **13**, 283–96.

ZHU, L. (2005). *Nonparametric Monte Carlo Tests and Their Applications*. New York: Springer.

ZHU, L., ZHU, L. & FENG, Z. (2010). Dimension reduction in regressions through cumulative slicing estimation. *J. Am. Statist. Assoc.* **105**, 1455–66.

ZHU, X., GUO, X. & ZHU, L. (2017). An adaptive-to-model test for partially parametric single-index models. *Statist. Comp.* **27**, 1193–1204.

ZOU, H. & ZHANG, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Ann. Statist.* **37**, 1733–51.

[*Received on* 18 *July* 2020. *Editorial decision on* 1 *February* 2021]