*The Annals of Statistics*
2019, Vol. 47, No. 4, 1960–1994
https://doi.org/10.1214/18-AOS1735


# ADAPTIVE-TO-MODEL CHECKING FOR REGRESSIONS WITH DIVERGING NUMBER OF PREDICTORS[1]

BY FALONG TAN* AND LIXING ZHU†,‡

*Hunan University**, *Beijing Normal University*† *and Hong Kong Baptist University*‡

In this paper, we construct an adaptive-to-model residual-marked empirical process as the base of constructing a goodness-of-fit test for parametric single-index models with diverging number of predictors. To study the relevant asymptotic properties, we first investigate, under the null and alternative hypothesis, the estimation consistency and asymptotically linear representation of the nonlinear least squares estimator for the parameters of interest and then the convergence of the empirical process to a Gaussian process. We prove that under the null hypothesis the convergence of the process holds when the number of predictors diverges to infinity at a certain rate that can be of order, in some cases, $o(n^{1/3}/\log n)$ where $n$ is the sample size. The convergence is also studied under the local and global alternative hypothesis. These results are readily applied to other model checking problems. Further, by modifying the approach in the literature to suit the diverging dimension settings, we construct a martingale transformation and then the asymptotic properties of the test statistic are investigated. Numerical studies are conducted to examine the performance of the test.

**1. Introduction.** Parametric regression models have been widely used in practice. It is however necessary to check the model adequacy to prevent possible wrong conclusions in any further analysis. This issue has been well studied when the dimension $p$ of the predictor vector is fixed. Yet, for the cases with large dimension that may be regarded as a diverging number as the sample size goes to infinity, there are no tests for parametric models available in the literature. We now specify this problem.

Let $Y_n$ be a response variable associated with a $p_n$-dimensional predictor vector $X_n \in \mathbb{R}^{p_n}$ where $p_n$ diverges as the sample size $n$ tends to infinity. The regression function $m_n(x) = E(Y_n|X_n = x)$ is the conditional expectation of $Y_n$ given $X_n$. Let $\mathcal{G}_n = \{g(\beta^\top \cdot, \theta) : \beta \in \mathbb{R}^{p_n}, \theta \in \mathbb{R}^d\}$ be a given parametric family of functions where $g(\cdot, \cdot)$ is a given function. The study herewith is motivated by checking

Received January 2018; revised May 2018.

[1]Supported by the Fundamental Research Funds for the Central Universities, a grant from the University Grants Council of Hong Kong and National Natural Science Foundation of China Grant #NSFC11671042.

*MSC2010 subject classifications.* Primary 62G10; secondary 62M07.

*Key words and phrases.* Adaptive-to-model test, diverging number of predictors, empirical process, martingale transformation, parametric single-index models, sufficient dimension reduction.

whether $m_n(\cdot)$ belongs to $\mathcal{G}_n$ or not. Thus the null hypothesis we want to test is that $(Y_n, X_n)$ follows a parametric single-index model as

$$Y_n = g\big(\beta_{n0}^\top X_n, \theta_0\big) + \varepsilon_n \quad \text{for some } \beta_{n0} \in \mathbb{R}^{p_n}, \theta_0 \in \mathbb{R}^d. \tag{1.1}$$

Here $\varepsilon_n = Y_n - E(Y_n|X_n)$ is the error term with $\varepsilon_n = V_1(\beta_{n0}^\top X_n)\tilde{\varepsilon}_n$ where $\tilde{\varepsilon}_n$ is independent of $X_n$ and has zero mean, $V_1$ is a squared integrable nonparametric function, $d$ is fixed and $\top$ denotes the transposition. In this paper, we call this model the parametric single-index model. Although they are in form a generalized linear model, we do not use this name as generalized linear models have their own definitions in the literature.

To make full use of the model structures under both the null and the alternative hypothesis, we consider the following alternative model

$$Y_n = G\big(B_n^\top X_n\big) + \varepsilon_n. \tag{1.2}$$

Here $\varepsilon_n = Y_n - E(Y_n|X_n)$ is the error term with $\varepsilon_n = V_2(B_n^\top X_n)\tilde{\varepsilon}_n$ where $\tilde{\varepsilon}_n$ is independent of $X_n$ and has zero mean, $V_2$ is also a squared integrable nonparametric function, $G(\cdot)$ is an unknown smooth function and $B_n$ is a $p_n \times q_n$ orthonormal matrix with an unknown $q_n$ with $1 \le q_n \le p_n$. Note that this is a more general model than the nonparametric model $Y_n = G(X_n) + \varepsilon_n$ that is a special case when $B_n$ is an $p_n \times p_n$ identity matrix with $q_n = p_n$.

We now review the existing methodologies in the literature when the dimension $p_n = p$ is fixed. Two major classes of tests are: locally smoothing tests and globally smoothing tests. Locally smoothing tests use nonparametric smoothing estimators to construct test statistics; see Härdle and Mammen (1993), Zheng (1996), Fan and Li (1996), Dette (1999), Fan and Huang (2001), Koul and Ni (2004) and Van Keilegom, González Manteiga and Sánchez Sellero (2008) as examples. Globally smoothing tests construct test statistics based on averages of functionals of empirical processes and thus avoid nonparametric estimation. They are called globally smoothing tests as averaging is a globally smoothing step. Examples include Bierens (1982, 1990), Stute (1997), Stute, Thies, and Zhu (1998), Stute, González Manteiga and Presedo Quindimil (1998) and Khmaladze and Koul (2004). González-Manteiga and Crujeiras (2013) is a nice review paper.

When the dimension $p$ is fixed but large (even moderate), most existing tests, especially locally smoothing tests, perform badly due to the use of nonparametric estimations. Globally smoothing tests also suffer from the data sparseness in high-dimensional space. Several efforts have been devoted to solving this problem. Stute and Zhu (2002) can be regarded as a dimension reduction-based test. A martingale transformation leads it to be asymptotically distribution-free. This test has been proved to be powerful in many cases, even when $p$ is large. But it is not omnibus while is a directional test. Escanciano (2006) gave some detailed comments on this issue, and proposed a test that is based on projected predictors. Lavergne and Patilea (2008, 2012) also suggested projection-based tests. An early relevant reference is Zhu and An (1992). Guo, Wang and Zhu (2016) also commented on this issue and put forward to a model adaptation notion in hypothesis testing. This innovative notion provides a deep insight into model checking for regressions and the adaptive-to-model approach can fully use the model structures under both the null and alternative hypothesis. Recently, with the assistance of sufficient dimension reduction techniques, Tan, Zhu and Zhu (2018) generalized Stute and Zhu's (2002) method and obtained an omnibus test which is asymptotically distribution-free and inherits the dimension reduction properties.

However, extending the existing methods to diverging dimension cases is by no means trivial. In this paper, we devote the effort on this issue to construct an adaptive-to-model residual-marked empirical process as the base of constructing a test. Under the null hypothesis, the process is similar to that of Stute and Zhu (2002). However, it is of importance to investigate at which rate of $p_n$ to infinity, the convergence of the empirical process to a Gaussian process can be achieved. As there are no relevant results in the literature about this, investigating this issue is one of the main focuses in this paper. We find that the leading rate $n^{1/3}$ of $p_n$ for ensuring the convergence of the empirical processes would not be easy to improve as the technical proof shows this, although we cannot give a definitive answer right now. A brief comment will be given in Section 3. The results are of particular interest as the theoretical results can be applied to other model checking problems when any residual-marked empirical process is used to construct test statistic in diverging dimension settings.

This study also relates to parameter estimation when $p_n$ is divergent. For linear models, this issue has been paid much attention in the literature. Huber (1973) was a pioneer work that provided norm consistency and asymptotic normality of the least squares estimator for linear models when $p_n$ goes to infinity at the rate of order $o(n^{1/3})$ where $n$ is the sample size. Portnoy (1984, 1985) refined the results under some more conditions. There are many developments afterwards. For instance, Zou and Zhang (2009) greatly improved the diverging rate of $p_n$ also for linear models. All these methods assume fixed designs that are different from the case in the present paper. When the model is nonlinear, there are few estimation results available in the literature. Thus, we also give a study on this.

Another interesting issue is that even its limiting Gaussian process can be derived, the shift term created by estimating the parameters of interest has no close form. Thus, we cannot directly follow the martingale transformation as Stute and Zhu (2002) did. This is a typical problem when $p_n$ is divergent, which does not happen when $p_n = p$ is fixed. Then a modified approach is suggested to define a martingale transformation in this scenario. The asymptotic properties of the martingale transformation-based innovation process under both the null and alternatives are also studied. We show that when $p_n = p$ is fixed, this transformation is equivalent to that in Stute and Zhu (2002). The test based on the constructed process can be consistent against all global alternatives as well as the local alternatives distinct from the null at the rate slower than $1/\sqrt{n}$. We also prove that it can detect the local alternatives converging to the null at the rate of $1/\sqrt{n}$, a fastest possible rate in hypothesis testing. These are interesting results, that the sensitivity of the test to the local alternatives can still be at a parametric rate even when $p_n$ is divergent. Further, the test statistic construction is also under the model adaptation framework such that the test is still omnibus even when the one-dimensional dimension reduction structure under the null has been fully used like Stute and Zhu (2002). This inherits the properties of Tan, Zhu and Zhu's (2018) test.

The paper is organized as follows. Section 2 contains the asymptotic properties of the ordinary least squares estimator with diverging dimension. Based on the estimator, we define an adaptive-to-model residual-marked empirical process. Since sufficient dimension reduction theory plays a crucial role to achieve the model adaptation property, we give a brief review in this section and study the convergence rate of the relevant estimators. In Section 3, we present the result that the adaptive-to-model empirical process converges weakly to a Gaussian process under the null hypothesis and the asymptotic properties under the local and global alternative hypothesis. We also give the test statistic for practical use. In Section 4, several simulation studies are conducted to examine the performance of the test and a real data example is analysed for illustration. Section 5 contains some discussions and topics in the future study. As Theorem 3.1 is an important result to show the convergence of the empirical process when the dimension $p_n$ is divergent, we give the proof in the Appendix. The regularity conditions and the proofs for the other theorems and propositions are contained in the Supplementary Material (Tan and Zhu (2019)) for saving space.

## 2. An adaptive-to-model residual-marked empirical process.

2.1. *Parameter estimation.* Let $\{(X_{ni}, Y_{ni}), i = 1, \dots, n\}$ be an i.i.d. sample with the same distribution as $(X_n, Y_n)$ and let $\varepsilon_n = Y_n - E(Y_n|X_n)$ be the unpredictable part of $Y_n$ given $X_n$. Recall that $\mathcal{G}_n = \{g(\beta^\top \cdot, \theta) : \beta \in \mathbb{R}^{p_n}, \theta \in \mathbb{R}^d\}$. We want to test whether or not

$$H_0 : Y_n = g(\beta_{n0}^\top X_n, \theta_0) + \varepsilon_n \quad \text{for some } \beta_{n0} \in \mathbb{R}^{p_n}, \theta_0 \in \mathbb{R}^d.$$

For estimating the unknown $(\beta_{n0}, \theta_0)$, in this paper we restrict ourselves to the ordinary least squares method. Let

$$(\hat{\beta}_n, \hat{\theta}_n) = \underset{\beta,\theta}{\operatorname{argmin}} \sum_{i=1}^{n} [Y_{ni} - g(\beta^\top X_{ni}, \theta)]^2.$$

To analyze the asymptotic property of $(\hat{\beta}_n, \hat{\theta}_n)$, define

$$(\tilde{\beta}_{n0}, \tilde{\theta}_0) = \underset{\beta,\theta}{\operatorname{argmin}} E[Y_n - g(\beta^\top X_n, \theta)]^2.$$

It is easy to see that if $m_n(\cdot) \in \mathcal{G}_n$, we have $(\tilde{\beta}_{n0}, \tilde{\theta}_0) = (\beta_{n0}, \theta_0)$. If $m_n(\cdot) \notin \mathcal{G}_n$, $(\tilde{\beta}_{n0}, \tilde{\theta}_0)$ typically depends on the distribution of $X_n$. Let $e_n = Y_n - g(\tilde{\beta}_{n0}^\top X_n, \tilde{\theta}_0)$. Then under the null hypothesis we have $e_n = \varepsilon_n$.

To study the asymptotic properties of $(\hat{\beta}_n, \hat{\theta}_n)$ as $p_n$ is divergent, we first give some notations. The regularity conditions are postponed to the Supplementary Material (Tan and Zhu (2019)). Suppose that $g(\beta^\top x, \theta)$ is third differentiable with respect to $(\beta, \theta)$. Let

$$g'(\beta, \theta, x) = \frac{\partial g(\beta^\top x, \theta)}{\partial(\beta, \theta)} \quad \text{and} \quad g''(\beta, \theta, x) = \frac{\partial g'(\beta, \theta, x)}{\partial(\beta, \theta)}.$$

The matrix $g''(\beta, \theta, x)$ is used within the following matrix $\Sigma_n$ which will play a crucial role in deriving the asymptotic properties of $(\hat{\beta}_n, \hat{\theta}_n)$:

$$\Sigma_n = E\big[g'(\tilde{\beta}_{n0}, \tilde{\theta}_0, X_n) g'(\tilde{\beta}_{n0}, \tilde{\theta}_0, X_n)^\top\big] - E\big[e_n g''(\tilde{\beta}_{n0}, \tilde{\theta}_0, X_n)\big] =: \Sigma_{n1} - \Sigma_{n2}.$$

The next two results give the norm consistency of $(\hat{\beta}_n, \hat{\theta}_n)$ to $(\tilde{\beta}_{n0}, \tilde{\theta}_0)$ and the asymptotically linear representation of $\binom{\hat{\beta}_n - \tilde{\beta}_{n0}}{\hat{\theta}_n - \tilde{\theta}_0}$. The representation generalizes the results of White (1981) to the diverging dimension settings. For simplicity, we define hereafter $\hat{\gamma}_n = (\hat{\beta}_n^\top, \hat{\theta}_n^\top)^\top$, $\tilde{\gamma}_{n0} = (\tilde{\beta}_{n0}^\top, \tilde{\theta}_0^\top)^\top$ and $\gamma_{n0} = (\beta_{n0}^\top, \theta_0^\top)^\top$.

PROPOSITION 1. *Assume that conditions* (A1)–(A6) *in the Supplementary Material hold. If $p_n^2/n \to 0$ and $g''(\beta, \theta, x) \equiv 0$, or $p_n^4/n \to 0$, then $\hat{\gamma}_n$ is a norm consistent estimator of $\tilde{\gamma}_{n0}$ in the sense that $\|\hat{\gamma}_n - \tilde{\gamma}_{n0}\| = O_p(\sqrt{p_n/n})$, where $\|\cdot\|$ denotes the Frobenius norm.*

The convergence rate of order $\sqrt{p_n/n}$ is in line with the results in Huber (1973) and Portnoy (1984) when $p_n$ diverges. For the asymptotically linear representation, we have the following result.

PROPOSITION 2. *Assume that conditions* (A1)–(A6) *in the Supplementary Material* (*Tan and Zhu* (2019)) *hold. If $p_n^3/n \to 0$ and $g''(\beta, \theta, x) \equiv 0$, or $p_n^5/n \to 0$, we then have the following asymptotically linear representation*:

$$\hat{\gamma}_n - \tilde{\gamma}_{n0} = \Sigma_n^{-1} \frac{1}{n} \sum_{i=1}^{n} [Y_{ni} - g(\tilde{\beta}_{n0}^\top X_{ni}, \tilde{\theta}_0)] g'(\tilde{\beta}_{n0}, \tilde{\theta}_0, X_{ni}) + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{2.1}$$

*where the remaining term $o_p(\frac{1}{\sqrt{n}})$ is in the sense of norm consistency.*

REMARK 1. Both the rate $p_n^4/n \to 0$ and $p_n^5/n \to 0$ in these two propositions as $n \to \infty$ seem slow. When the condition $g''(\beta, \theta, \cdot) \equiv 0$ holds, the rates can be improved to $p_n^2/n \to 0$ and $p_n^3/n \to 0$, respectively as Huber (1973) obtained, where he only considered linear models. Portnoy (1984, 1985) also obtained the norm consistency and the normal approximation under a weaker condition where he still considered linear models. Note that if $g(\beta^\top X, \theta) = \beta^\top X$, $g''(\beta, \theta, \cdot) \equiv 0$. Thus, our result is more general, though slightly. Further, for the least squares estimator for linear models, when fixed design is considered, the rates can be improved to be $p_n = o(n)$ and $p_n = O(n^\alpha)$ for $0 \le \alpha < 1$ for the norm consistency and asymptotically linear representation when some regularity conditions are assumed. See Zou and Zhang (2009). However, for the random design in this paper, it is still unknown whether these faster rates can achieve or not. More specifically, if without other extra conditions, the fastest passible convergence rate is $p_n^2/n$. This is because we need the convergence rate of $\|\hat{\Sigma}_n - \Sigma_n\|$ in proving Propositions 1 and 2 where $\hat{\Sigma}_n$ is the estimator of $\Sigma_n$. Since each element in $\hat{\Sigma}_n - \Sigma_n$ has an optimal rate $1/\sqrt{n}$ and $\|\hat{\Sigma}_n - \Sigma_n\|$ is a square root of the sum of $p_n^2$ elements, it is easy to see that $\|\hat{\Sigma}_n - \Sigma_n\| = O_p(\sqrt{p_n^2/n})$ as presented in Propositions 1 and 2. Moreover, for nonlinear models without $g''(\beta, \theta, \cdot) \equiv 0$, the remainders after the Taylor expansion involve $g''(\beta, \theta, \cdot)$ and $g'''(\beta, \theta, \cdot)$. Thus, the rate gets slower and the optimal rate is still unknown.

2.2. *Empirical process construction.* Recall the null hypothesis:

$$H_0: \quad \mathbb{P}\{E(Y_n|X_n) = g(\beta_{n0}^\top X_n, \theta_0)\} = 1 \quad \text{for some } \beta_{n0} \in \mathbb{R}^{p_n}, \theta_0 \in \mathbb{R}^d,$$

against the alternative hypothesis:

$$H_1: \quad \mathbb{P}\{E(Y_n|X_n) = G(B_n^\top X_n) \neq g(\beta^\top X_n, \theta)\} < 1 \quad \forall \beta \in \mathbb{R}^{p_n}, \theta \in \mathbb{R}^d,$$

where $G(\cdot)$ is an unknown smooth function and the $p_n \times q_n$ orthonormal matrix $B_n$ is given in (1.2). We assume that $\tilde{\beta}_{n0} \in \mathcal{S}_{E(Y_n|X_n)}$ under both the null and alternative hypothesis where $\mathcal{S}_{E(Y_n|X_n)}$ is the central mean subspace such that $\mathcal{S}_{E(Y_n|X_n)} = \text{span}(B_n)$. Under the null hypothesis, this is obvious. Under the alternative hypothesis, if $g(\beta^\top X_n, \theta) = \beta^\top X_n$ is a linear model, we have $\tilde{\beta}_{n0} = [E(X_n X_n^\top)]^{-1} E(X_n Y_n) \in \mathcal{S}_{E(Y_n|X_n)}$. For other models, $\tilde{\beta}_{n0}$ may not be necessarily in $\mathcal{S}_{E(Y_n|X_n)}$. If $\tilde{\beta}_{n0} \notin \mathcal{S}_{E(Y_n|X_n)}$ under the alternative hypothesis, in Section 5 we will give a detailed discussion and provide a partial solution to relax the assumption we impose.

Also recall $\varepsilon_n = Y_n - E(Y_n|X_n)$ and $e_n = Y_n - g(\tilde{\beta}_{n0}^\top X_n, \tilde{\theta}_0)$. Under the null hypothesis, $e_n = \varepsilon_n$, $q = 1$ and $B_n = \kappa_n \beta_{n0}$ with $\kappa_n = \pm \frac{1}{\|\beta_{n0}\|}$. Therefore, we obtain that $E(e_n|B_n^\top X_n) = E(e_n|\beta_{n0}^\top X_n) = 0$. Under the alternative hypothesis, we have $E(e_n|B_n^\top X_n) = G(B_n^\top X_n) - g(\tilde{\beta}_{n0}^\top X_n, \tilde{\theta}_0) \neq 0$. Then it follows that under the null hypothesis

$$E[e_n I(B_n^\top X_n \le u)] = E[e_n I(\kappa_n \beta_{n0}^\top X_n \le u)] = 0. \tag{2.2}$$

While under the alternatives, by Lemma 1 of Escanciano (2006), there exists an $\alpha_n \in \mathcal{S}_{q_n}^+$ such that $E(e_n|\alpha_n^\top B_n^\top X_n) \neq 0$, where $\mathcal{S}_{q_n}^+ = \{\alpha_n = (a_1, \dots, a_{q_n})^\top \in \mathbb{R}^{q_n} : \|\alpha_n\| = 1 \text{ and } a_1 \ge 0\}$. Then it follows that

$$E[e_n I(\alpha_n^\top B_n^\top X_n \le u)] \neq 0. \tag{2.3}$$

Note that under the null we have $q_n = 1$ and $\mathcal{S}_{q_n}^{+} = \{1\}$. Thus the quantity $E[e_n I(\alpha_n^{\top} B_n^{\top} X_n \le u)]$ actually has the same form in both (2.2) and (2.3). Define an adaptive-to-model residual marked empirical process $V_n(u)$ in the diverging dimension setting as below

$$V_n(\hat{\alpha}_n, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_{ni} - g(\hat{\beta}_n^{\top} X_{ni}, \hat{\theta}_n)] I(\hat{\alpha}_n^{\top} \hat{B}_n^{\top} X_{ni} \le u), \tag{2.4}$$

$$V_n(u) = \sup_{\hat{\alpha}_n \in \mathcal{S}_{\hat{q}_n}^{+}} |V_n(\hat{\alpha}_n, u)|, \tag{2.5}$$

where $\hat{\beta}_n$ and $\hat{\theta}_n$ are defined as before and $\hat{B}_n$ is the sufficient dimension reduction estimator of $B_n$ with an estimated structural dimension $\hat{q}_n$ of $q_n$, which will be specified later. For $V_n(u)$, one can also use the integral over $\mathcal{S}_{\hat{q}_n}^{+}$ to define a test statistic.

To achieve the model adaptation property of the process, we need sufficient dimension reduction (SDR) techniques to identify the structural dimension $q_n$ and the matrix $B_n$. We give a brief review below on this topic and extend the results to diverging dimension settings.

2.3. *The estimation of the matrix $B_n$ and its structural dimension.* Recall under the alternative hypothesis the model is as

$$Y_n = G(B_n^{\top} X_n) + \varepsilon_n, \tag{2.6}$$

where the error term satisfies that $E(\varepsilon_n | B_n^{\top} X_n) = 0$, $G(\cdot)$ is an unknown smooth function and $B_n$ is a $p_n \times q_n$ orthonormal matrix with $1 \le q_n \le p_n$. We can see that under both the null and alternative hypothesis, the conditional independence holds respectively:

$$Y_n \perp\!\!\!\perp E(Y_n|X_n) | \beta_{n0}^{\top} X_n, \quad \text{and} \quad Y_n \perp\!\!\!\perp E(Y_n|X_n) | B_n^{\top} X_n,$$

where $\perp\!\!\!\perp$ means statistical independence. Define $\mathcal{S}_{E(Y_n|X_n)}$ as the central mean subspace of $Y_n$ with respect to $X_n$ (see Cook and Li (2002)) that is, the intersection of all subspaces $\mathrm{span}(B_n)$ spanned by the columns of $B_n$ such that $Y_n \perp\!\!\!\perp E(Y_n|X_n)|B_n^{\top} X_n$. The dimension of $\mathcal{S}_{E(Y_n|X_n)}$ is called the structural dimension, denoted as $d_{E(Y_n|X_n)}$. Under mild conditions, such a subspace $\mathcal{S}_{E(Y_n|X_n)}$ always exists (see Cook and Li (2002)). If $\mathcal{S}_{E(Y_n|X_n)} = \mathrm{span}(B_n)$, then $E(Y_n|X_n) = E(Y_n|B_n^{\top} X_n)$. Under the null hypothesis (1.1), $d_{E(Y_n|X_n)} = 1$ and $\mathcal{S}_{E(Y_n|X_n)} = \mathrm{span}(\beta_{n0}/\|\beta_{n0}\|)$. Under the alternative (1.2), $d_{E(Y_n|X_n)} = q_n$ and $\mathcal{S}_{E(Y_n|X_n)} = \mathrm{span}(B_n)$. There are some methods to estimate the central mean subspace such that a matrix $B_n C_n$ can be identified where $C_n$ is a $q \times q$ orthonormal matrix. Principal Hessian directions (pHd, Li (1992)) is a popularly used method for this purpose when $p$ is fixed. However, when $p$ is divergent, there are no corresponding asymptotic results about pHd and we guess that the convergence rates of the corresponding estimated eigenvalues and matrix $B_n$ would be very slow. Thus, in this paper we consider a method that is for identifying the central subspace defined below.

The central subspace is related to the conditional distribution of $Y_n|X_n$ (see Cook (1998)), denoted by $\mathcal{S}_{Y_n|X_n}$. This space is the intersection of all subspaces span$(B_n)$ such that $Y_n \perp\!\!\!\perp X_n|B_n^\top X_n$. Obviously, $\mathcal{S}_{E(Y_n|X_n)} \subset \mathcal{S}_{Y_n|X_n}$. Under the conditions on the error term under the null and alternative hypothesis, $\mathcal{S}_{E(Y_n|X_n)} = \mathcal{S}_{Y_n|X_n}$. That is, our conditions on the error term can make sure that $\mathcal{S}_{E(Y_n|X_n)} = \mathcal{S}_{Y_n|X_n}$ as $\varepsilon_n = V_1(\beta_{n0}^\top X_n)\tilde{\varepsilon}_n$ under the null and $\varepsilon_n = V_2(B_n^\top X_n)\tilde{\varepsilon}_n$ under the alternative where $\tilde{\varepsilon}_n$ is independent of $X_n$. In Section 5, we will discuss how to relax these conditions as well.

There are also several estimation proposals available in the literature. For instance, sliced inverse regression (SIR, Li (1991)), sliced average variance estimation (SAVE, Cook and Weisberg (1991)), directional regression (DR, Li and Wang (2007)) and discretization-expectation estimation (DEE, Zhu et al. (2010)). All these methods assumed that $p$ is fixed. Zhu, Miao and Peng (2006) first discussed the asymptotic properties of SIR when $p_n$ diverges to infinity. In this paper, we adapt cumulative slicing estimation (CSE, Zhu, Zhu and Feng (2010)) to identify the central subspace. This is because it is very easily implemented and can handle the case where the dimension $p_n$ grows to infinity. Note that CSE requires the linearity condition. This is satisfied if the predictors $X_n$ are elliptically symmetric. Hall and Li (1993) showed that when $p_n \to \infty$ as $n \to \infty$, the linear combinations of the covariates are approximately normally distributed. Thus the linearity condition for CSE is approximately satisfied when the dimension $p_n$ is large.

The procedure of CSE is as follows. For simplicity, we assume $E(X_n) = 0$, $\mathrm{Var}(X_n) = I_{p_n}$ for a moment. Thus it is easy to see that $E[X_n h(Y_n)] \in \mathcal{S}_{Y_n|X_n}$ for any function $h(\cdot)$. Theoretically, we obtain infinity amount of vectors in $\mathcal{S}_{Y_n|X_n}$. Zhu et al. (2010) suggested a determining class of indicator functions to replace $h(\cdot)$. Let $h_t(Y_n) = I(Y_n \le t)$. It follows that

$$Y_n \perp\!\!\!\perp X_n|B_n^\top X_n \quad \Longleftrightarrow \quad h_t(Y_n) \perp\!\!\!\perp X_n|B_n^\top X_n, \quad \forall t \in \mathbb{R}.$$

Define the target matrix

$$M_n = \int E\big[X_n h_t(Y_n)\big]E\big[X_n^\top h_t(Y_n)\big]\,dF_{Y_n}(t), \tag{2.7}$$

where $F_{Y_n}$ denotes the cumulative distribution function of $Y_n$. If the rank of $M_n$ is $q_n$, then span$(M_n) = \mathcal{S}_{Y_n|X_n}$. Based on this, it is easy to obtain the sample version of $M_n$. Let $Z_{ni}$ be the standardized $X_{ni}$ and $\hat{\xi}_t = \frac{1}{n}\sum_{i=1}^n Z_{ni} I(Y_{ni} \le t)$. The estimator of $M_n$ is given by

$$\hat{M}_n = \frac{1}{n}\sum_{j=1}^n \hat{\xi}_{Y_{nj}}\hat{\xi}_{Y_{nj}}^\top. \tag{2.8}$$

If the structural dimension $q_n$ is given, an estimator $\hat{B}_n(q)$ of $B_n$ consists of the eigenvectors corresponding to the largest $q_n$ eigenvalues of $\hat{M}_n$.

Yet we need a consistent estimator $\hat{q}_n$ of $q_n$ that is usually unknown under the alternative hypothesis. Later we will see that even when $q_n$ is given, we still want a consistent estimator because we wish the test to have the model adaptation property to fully use the dimension reduction structure under the null hypothesis. Inspired by Xia et al. (2015), we suggest a minimum ridge-type eigenvalue ratio estimator (MRER) to determine $q_n$. Let $\hat{\lambda}_{n1} \geq \cdots \geq \hat{\lambda}_{np_n}$ and $\lambda_{n1} \geq \cdots \geq \lambda_{np_n}$ be the eigenvalues of the matrix $\hat{M}_n$ and $M_n$ respectively. Since $\operatorname{rank}(M_n) = q_n$, it follows that

$$\lambda_{n1} \geq \cdots \geq \lambda_{nq_n} > \lambda_{n,q_n+1} = \cdots = \lambda_{n,p_n} = 0.$$

Hence we estimate the structural dimension $q_n$ by

$$\hat{q}_n = \arg \min_{1 \leq i \leq p_n} \left\{ i : \frac{\hat{\lambda}_{n,i+1}^2 + c_n}{\hat{\lambda}_{ni}^2 + c_n} \right\}. \tag{2.9}$$

Here $\hat{\lambda}_{n,p_n+1} = 0$ and the ridge $c_n$ is a positive constant depending on $n$. The following result shows that the consistency of MRER is adaptive to the underlying models. Its proof is given in the Supplementary Material (Tan and Zhu (2019)).

PROPOSITION 3. *Suppose that the regularity conditions of Theorem 3 in Zhu et al. (2010) hold. Let $\hat{B}_n(q_n)$ be a matrix whose columns are the eigenvectors that are associated with the largest $q_n$ eigenvalues of $\hat{M}_n$. Assume further that $0 < c_0 \leq \lambda_{nq_n} \leq \cdots \leq \lambda_{n1} \leq C_0 < \infty$ and $c_n = (\log n)/n$. Then:*

(1) *under $H_0$, we have $\mathbb{P}(\hat{q}_n = 1) \to 1$ and $\|\hat{B}_n(1) - \kappa_n \beta_{n0}\| = O_p(\sqrt{p_n/n})$;*
(2) *under $H_1$, we have $\mathbb{P}(\hat{q}_n = q_n) \to 1$ and $\|\hat{B}_n(q_n) - B_n\| = O_p(\sqrt{p_n q_n/n})$.*

## 3. Main results.

3.1. *Asymptotic properties of the process.* First, we discuss the asymptotic properties of the process $V_n(\hat{\alpha}_n, u)$ under the null hypothesis. To facilitate the study, we define the following process:

$$V_n^0(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_{ni} - g(\beta_{n0}^\top X_{ni}, \theta_0)] I(\kappa_n \beta_{n0}^\top X_{ni} \leq u).$$

Put

$$\sigma_n^2(v) = \operatorname{Var}(Y_n | \kappa_n \beta_{n0}^\top X_n = v),$$
$$\psi_n(u) = E[\operatorname{Var}(Y_n | \kappa_n \beta_{n0}^\top X_n) I(\kappa_n \beta_{n0}^\top X_n \leq u)].$$

Then we have $\sigma_n^2(v) = E(\varepsilon_n^2|\kappa_n\beta_{n0}^\top X_n = v)$ and $\psi_n(u) = \int_{-\infty}^u \sigma_n^2(v)F_{\kappa_n\beta_{n0}}(dv)$ where $F_{\kappa_n\beta_{n0}}$ is the cumulative distribution function of $\kappa_n\beta_{n0}^\top X_n$. Obviously, $\psi_n(u)$ is a nondecreasing and nonnegative function. Since

$$V_n^0(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_{ni} I(\kappa_n\beta_{n0}^\top X_{ni} \le u)$$

is a centered residual cusum process, it is readily seen that

$$\mathrm{Cov}[V_n^0(s), V_n^0(t)] = \psi_n(s \wedge t).$$

By Theorem 2.11.22 in van der Vaart and Wellner (1996), we obtain that $V_n^0(u)$ is asymptotically tight. If $\psi_n(u) \to \psi(u)$ pointwisely in $u$, it follows that

$$V_n^0(u) \longrightarrow V_\infty(u) \quad \text{in distribution,} \tag{3.1}$$

in the space $\ell^\infty(\overline{R})$, where $V_\infty(u)$ is a centred Gaussian process with the covariance function $\psi(s \wedge t)$ and $\ell^\infty(\overline{R})$ is the set of all real bounded functions on $\overline{R}$ (see Section 1.5 in van der Vaart and Wellner (1996)). Since $\psi(u)$ is also nondecreasing and nonnegative, it follows that $V_\infty(u) = B(\psi(u))$ in distribution, where $B(u)$ is a standard Brownian motion.

We now study $V_n(\hat{\alpha}_n, u)$ defined in (2.4). By Proposition 3, $\mathbb{P}(\hat{q}_n = 1) \to 1$ under the null hypothesis. Thus we only need to work on the event $\{\hat{q}_n = 1\}$. Consequently, $\mathcal{S}_{\hat{q}_n}^+ = \{1\}$ and $V_n(\hat{\alpha}_n, u)$ can be rewritten as

$$V_n(\hat{\alpha}_n, u) = \frac{1}{\sqrt{n}}\sum_{i=1}^n [Y_{ni} - g(\hat{\beta}_n^\top X_{ni}, \hat{\theta}_n)] I(\hat{B}_n^\top X_{ni} \le u).$$

Under some regularity conditions stated in the Supplementary Material (Tan and Zhu (2019))and on the event $\{\hat{q}_n = 1\}$, we can show that under the null hypothesis

$$V_n(\hat{\alpha}_n, u) = V_n^0(u) - \sqrt{n}(\hat{\gamma}_n - \gamma_{n0})^\top R_n(u) + o_p(1) \tag{3.2}$$

uniformly in $u$, where $R_n(u) = E[g'(\beta_{n0}, \theta_0, X_n) I(\kappa_n\beta_{n0}^\top X_n \le u)]$. A proof of (3.2) is given in the Appendix. Combining (3.2) with Proposition 2, some elementary calculations yield

$$V_n(\hat{\alpha}_n, u) = V_n^0(u) - R_n(u)^\top \Sigma_n^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n g'(\beta_{n0}, \theta_0, X_{ni})\varepsilon_{ni} + o_p(1) \tag{3.3}$$

uniformly in $u$. Altogether we obtain the following result.

THEOREM 3.1. *Suppose that the regularity conditions in the Supplementary Material hold. When* $(p_n \log n)^3/n \to 0$ *and* $g''(\beta, \theta, x) \equiv 0$, *or* $p_n^5/n \to 0$, *then under the null hypothesis, we have in distribution*

$$V_n(u) \longrightarrow |V_\infty^1(u)|,$$

*where $V_\infty^1(u)$ is a zero mean Gaussian process with a covariance function $K(s,t)$ that is the pointwise limit of $K_n(s,t)$ as*

$$\begin{aligned}K_n(s,t) = E\bigl[\varepsilon_n^2 I\bigl(\kappa_n\beta_{n0}^\top X_n \le s\wedge t\bigr)\bigr] \\ - R_n(s)^\top\Sigma_n^{-1}E\bigl[\varepsilon_n^2 g'(\beta_{n0},\theta_0,X_n)I\bigl(\kappa_n\beta_{n0}^\top X_n\le t\bigr)\bigr] \\ - R_n(t)^\top\Sigma_n^{-1}E\bigl[\varepsilon_n^2 g'(\beta_{n0},\theta_0,X_n)I\bigl(\kappa_n\beta_{n0}^\top X_n\le s\bigr)\bigr] \\ + R_n(s)^\top\Sigma_n^{-1}E\bigl[\varepsilon_n^2 g'(\beta_{n0},\theta_0,X_n)g'(\beta_{n0},\theta_0,X_n)^\top\bigr]\Sigma_n^{-1}R_n(t).\end{aligned}$$

REMARK 2. Note that the dimension $p_n$ is required to have divergence rate slower than $n^{1/3}/(\log n)$ under some conditions. From the lemmas in the Supplementary Material (Tan and Zhu (2019)), we can see that the leading term $n^{1/3}$ would be close to optimal. This is because when $p_n$ is divergent, the covering number of index functions in the empirical process diverges at an exponential rate and the equicontinuity of the process requires such a rate. This conjecture is based on a similar case for the projection pursuit-type Kolmogonov–Smirnov test investigated by Zhu and Cheng (1994) who gave the same rate for the lower and upper bound of the tail probability. Of course, when the underlying model has a sparse structure and a lower-dimensional model can be selected, the rate of $p_n$ can be faster. This is beyond the scope of this paper.

3.2. *Martingale transformation.* If $p$ is fixed, $V_\infty^1(u)$ can be rewritten as $V_\infty^1(u) = V_\infty(u) + R(u)^\top V$ in distribution and its covariance function can be specified. The shift term $R(u)^\top V$ is brought out from the second term in (3.3). Stute, Thies and Zhu (1998) first proposed a martingale transformation to eliminate $R(u)^\top V$ in $V_\infty^1(u)$ and then to obtain a tractable limiting distribution of a functional of $V_\infty(u)$. This has become one of the basic methodologies in the area of model checking to derive asymptotically distribution-free tests. It was motivated by the Khmaladze martingale transformation in constructing goodness of fit tests for hypothetical distribution functions (Khmaladze (1982)). There are a number of follow-up studies in the literature to extend this methodology to various high-dimensional models such as Khmaladze and Koul (2004) and Stute, Xu and Zhu (2008). However, when $p_n$ diverges as $n$ goes to infinity, the shift term does not have such a close form as that in the fixed dimension case. The martingale transformation cannot directly target $R(u)^\top V$. We then bypass this difficulty by checking its shift term at the sample level. Note that the shift term comes from the second term in (3.2) and in the case with the fixed $p$, $R(u)^\top V$ is just its weak limit.

Recall that $R_n(u) = E[g'(\beta_{n0},\theta_0,X_n)I(\kappa_n\beta_{n0}^\top X_n \le u)]$ and $\psi_n(u) = \int_{-\infty}^u \sigma_n^2(v)F_{\kappa_n\beta_{n0}}(dv)$. Let

$$a_n(u) = \frac{\partial R_n(u)}{\partial\psi_n(u)}$$

be the Radon–Nikodym derivative of $R_n(u)$ with respect to $\psi_n(u)$. Next, define a $(p_n + d) \times (p_n + d)$ matrix

$$A_n(u) = \int_u^\infty a_n(z) R_n^\top(dz) = \int_u^\infty a_n(z) a_n(z)^\top \sigma_n^2(z) F_{\kappa_n \beta_{n0}}(dz).$$

It can also be written as

$$A_n(u) = E\big[a_n(\kappa_n \beta_{n0}^\top X_n) g'(\beta_{n0}, \theta_0, X_n)^\top I(\kappa_n \beta_{n0}^\top X_n \geq u)\big].$$

Mimicking the martingale transformation in Stute and Zhu (2002) at the sample level, we have

$$(T_n f_n)(u) = f_n(u) - \int_{-\infty}^u a_n(z)^\top A_n^{-1}(z) \bigg( \int_z^\infty a_n(v) f_n(dv) \bigg) \psi_n(dz). \tag{3.4}$$

Here we should assume that $A_n(u)$ is nonsingular and the process $f_n(u)$ should be either of bounded variation or a Brownian motion.

Some elementary computations conclude that $T_n(\sqrt{n}(\hat{\gamma}_n - \gamma_{n0})^\top R_n) = 0$. We now discuss the asymptotic properties of $T_n V_n^0$. Note that

$$(T_n V_n^0)(u) = V_n^0(u) - \int_{-\infty}^u a_n(z)^\top A_n^{-1}(z) \bigg( \int_z^\infty a_n(v) V_n^0(dv) \bigg) \psi_n(dz)$$

and

$$\int_z^\infty a_n(v) V_n^0(dv) = \frac{1}{\sqrt{n}} \sum_{i=1}^n a_n(\kappa_n \beta_{n0}^\top X_{ni}) I(\kappa_n \beta_{n0}^\top X_{ni} \geq z) \varepsilon_{ni}.$$

Combining these two formulas, we obtain that

$$\begin{aligned} T_n V_n^0(u) = V_n^0(u) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{-\infty}^u & a_n(z)^\top A_n^{-1}(z) a_n(\kappa_n \beta_{n0}^\top X_{ni}) \\ & \times I(\kappa_n \beta_{n0}^\top X_{ni} \geq z) \psi_n(dz) \varepsilon_{ni}. \end{aligned}$$

Therefore, $T_n V_n^0$ is also a cusum process of i.i.d. centered residuals with the covariance function

$$\mathrm{Cov}\big[T_n V_n^0(s), T_n V_n^0(t)\big] = \mathrm{Cov}\big[V_n^0(s), V_n^0(t)\big] = \psi_n(s \wedge t). \tag{3.5}$$

This means that $T_n V_n^0(u)$ admits the same limiting distribution as that of $V_n^0(u)$, that is,

$$T_n V_n^0(u) \longrightarrow V_\infty(u) \quad \text{in distribution.} \tag{3.6}$$

Consequently, we get rid of the annoying shift term $\sqrt{n}(\hat{\gamma}_n - \gamma_{n0})^\top R_n$ and obtain the process $V_\infty(u)$ whose supremum over all $u$ has a tractable distribution. The assertions (3.5) and (3.6) will be justified in the Supplementary Material (Tan and Zhu (2019); see Lemma 1).

The transformation $T_n$ obviously contains some unknown quantities and therefore needs to be substituted by their empirical analogues. For this, let $g_1'(t,\theta) = \frac{\partial g(t,\theta)}{\partial t}$ and $g_2'(t,\theta) = \frac{\partial g(t,\theta)}{\partial \theta}$. It follows that

$$g'(\beta_{n0}, \theta_0, X_n) = \big(g_1'(\beta_{n0}^\top X_n, \theta_0)X_n^\top, g_2'(\beta_{n0}^\top X_n, \theta_0)^\top\big)^\top.$$

Consequently, we have

$$R_n(u) = \left(\int_{-\infty}^u g_1'(z/\kappa_n, \theta_0) r_n(z)^\top F_{\kappa_n\beta_{n0}}(dz), \int_{-\infty}^u g_2'(z/\kappa_n, \theta_0)^\top F_{\kappa_n\beta_{n0}}(dz)\right)^\top,$$

where $r_n(v) = E(X_n|\kappa_n\beta_{n0}^\top X_n = v)$. Conclude that

$$a_n(u) = \left(\frac{g_1'(u/\kappa_n, \theta_0) r_n(u)^\top}{\sigma_n^2(u)}, \frac{g_2'(u/\kappa_n, \theta_0)^\top}{\sigma_n^2(u)}\right)^\top.$$

Since $a_n(u)$ is related to $r_n(u)$ and $\sigma_n^2(u)$ on which we do not make any assumption rather than smoothness, we then need to estimate them in a nonparametric way. Thus, a standard Nadaraya–Watson estimator for $r_n(v)$ is defined by

$$\hat{r}_n(v) = \frac{\sum_{i=1}^n X_{ni} K(\frac{v-\hat{\alpha}_n^\top \hat{B}_n^\top X_{ni}}{h})}{\sum_{i=1}^n K(\frac{v-\hat{\alpha}_n^\top \hat{B}_n^\top X_{ni}}{h})},$$

where $K(\cdot)$ is an univariate kernel function and $h$ is a bandwidth. Similarly for $\sigma_n^2(u)$. Thus we obtain two estimators $\hat{a}_n(u)$ and $\hat{A}_n(u)$ of $a_n(u)$ and $A_n(u)$ respectively:

$$\hat{a}_n(u) = \left(\frac{g_1'(u/\hat{\kappa}_n, \hat{\theta}_n)\hat{r}_n(u)^\top}{\hat{\sigma}_n^2(u)}, \frac{g_2'(u/\hat{\kappa}_n, \hat{\theta}_n)^\top}{\hat{\sigma}_n^2(u)}\right)^\top,$$

$$\hat{A}_n(u) = \frac{1}{n}\sum_{i=1}^n \hat{a}_n(\hat{\alpha}_n^\top \hat{B}_n^\top X_{ni}) g'(\hat{\beta}_n, \hat{\theta}_n, X_{ni})^\top I(\hat{\alpha}_n^\top \hat{B}_n^\top X_{ni} \ge u).$$

Finally, we can give an estimator $\hat{T}_n$ of $T_n$:

$$\begin{aligned}
&\hat{T}_n V_n(\hat{\alpha}_n, u)\\
&\quad = V_n(\hat{\alpha}_n, u) - \int_{-\infty}^u \hat{a}_n(z)^\top \hat{A}_n^{-1}(z)\left(\int_z^\infty \hat{a}_n(v) V_n(\hat{\alpha}_n, dv)\right)\hat{\sigma}_n^2(z) F_{\hat{\alpha}_n}(dz)\\
&\quad = \frac{1}{n^{1/2}}\sum_{i=1}^n [Y_{ni} - g(\hat{\beta}_n^\top X_{ni}, \hat{\theta}_n)] I(\hat{\alpha}_n^\top \hat{B}_n^\top X_{ni} \le u)\\
&\qquad - \frac{1}{n^{3/2}}\sum_{i,j=1}^n I(\hat{\alpha}_n^\top \hat{B}_n^\top X_{ni} \le u)\hat{a}_n(\hat{\alpha}_n^\top \hat{B}_n^\top X_{ni})^\top\\
&\qquad \times \hat{A}_n^{-1}(\hat{\alpha}_n^\top \hat{B}_n^\top X_{ni})\hat{a}_n(\hat{\alpha}_n^\top \hat{B}_n^\top X_{nj})\\
&\qquad \times I(\hat{\alpha}_n^\top \hat{B}_n^\top X_{nj} \ge \hat{\alpha}_n^\top \hat{B}_n^\top X_{ni})[Y_{nj} - g(\hat{\beta}_n^\top X_{nj}, \hat{\theta}_n)]\hat{\sigma}_n^2(\hat{\alpha}_n^\top \hat{B}_n^\top X_{ni}),
\end{aligned}$$

where $\hat{\kappa}_n$ is an estimator of $\kappa_n$ and $F_{\hat{\alpha}_n}$ is the empirical distribution function of $\hat{\alpha}_n^\top \hat{B}_n^\top X_{ni}$, $1 \le i \le n$. Making sure the columns of $\hat{B}_n$ have the same direction as $\hat{\beta}_n$, we can assume $\kappa_n = 1/\|\beta_{n0}\|$ and $\hat{\kappa}_n = 1/\|\hat{\beta}_n\|$.

THEOREM 3.2. *Suppose that $A_n(u)$ is nonsingular and $\sigma_n^2(u)$ is bounded away from zero for all $n$ and $u$. If $(p_n \log n)^3/n \to 0$ and $g''(\beta, \theta, x) \equiv 0$, or $p_n^5/n \to 0$, under the null hypothesis $H_0$ and the regularity conditions in the Supplementary Material* (Tan and Zhu (2019)), *we have in distribution*

$$\sup_{\hat{\alpha}_n \in \mathcal{S}_{\hat{q}}^+} \left|\hat{T}_n V_n(\hat{\alpha}_n, u)\right| \longrightarrow \left|V_\infty(u)\right|$$

*in the space $\ell^\infty([-\infty, u_0])$ for any $u_0 \in \mathbb{R}$.*

Here $\ell^\infty([-\infty, u_0])$ is the space of all real bounded functions on $[-\infty, u_0]$. Note that we consider the convergence of $\sup_{\hat{\alpha}_n \in \mathcal{S}_{\hat{q}n}^+} |\hat{T}_n V_n(\hat{\alpha}_n, u)|$ in the space $\ell^\infty([-\infty, u_0])$. This is because $\hat{A}_n^{-1}(u)$ in the process $\hat{T}_n V_n(\hat{\alpha}_n, u)$ may be unbounded for large $u$ and thus the distributional behavior of the underlying process may become very unstable in the extreme right tails. Therefore, we restrict $\hat{T}_n V_n$ in the interval $[-\infty, u_0]$.

In a special case where the predictor $X_n$ follows a spherically contoured distribution or its extension, the elliptically contoured distribution, we can show that the calculations of the martingale transformation are much simpler. The idea is similar to Stute and Zhu (2002). Without loss of generality, we only consider spherically contoured distributions. Here we assume that the regression function $g$ does not depend on $\theta$. Let $g'(t)$ be the derivative of $g(t)$ with respect to $t$. It follows that

$$\begin{aligned} R_n(u) &= E\big[g'(\beta_{n0}^\top X_n) X_n I(\kappa_n \beta_{n0}^\top X_n \le u)\big] \\ &= \Gamma_n^\top E\big[g'(\beta_{n0}^\top X_n)\Gamma_n X_n I(\kappa_n \beta_{n0}^\top X_n \le u)\big], \end{aligned}$$

where $\Gamma_n$ is an $p_n \times p_n$ orthonormal matrix with the first row $\kappa_n \beta_{n0}^\top$ (or $\beta_{n0}^\top/\|\beta_{n0}\|$). Since the conditional expectation of the other components of $\Gamma_n X_n$ given the first is zero, it follows that

$$\begin{aligned} R_n(u) &= \frac{\beta_{n0}}{\|\beta_{n0}\|^2} E\big[g'(\beta_{n0}^\top X_n)\beta_{n0}^\top X_n I(\kappa_n \beta_{n0}^\top X_n \le u)\big] \\ &= \frac{\beta_{n0}}{\|\beta_{n0}\|^2} \int_{-\infty}^{u} g'(z/\kappa_n)(z/\kappa_n) F_{\kappa_n \beta_{n0}}(dz), \end{aligned}$$

whence,

$$\begin{aligned} a_n(u) &= \frac{\beta_{n0}}{\|\beta_{n0}\|^2} \frac{g'(u/\kappa_n) u/\kappa_n}{\sigma_n^2(u)}, \\ A_n(u) &= \frac{\beta_{n0}\beta_{n0}^\top}{\|\beta_{n0}\|^4} \int_u^\infty \frac{[g'(z/\kappa_n) z/\kappa_n]^2}{\sigma_n^2(z)} F_{\kappa_n \beta_{n0}}(dz). \end{aligned}$$

Note that $A_n(z)$ is a matrix with rank 1 and is singular when $p > 1$. Thus the martingale transformation can not apply directly. However, if we go back to (3.2) and set

$$\tilde{R}_n(u) = E[g'(\beta_{n0}^\top X_n)\beta_{n0}^\top X_n I(\kappa_n \beta_{n0}^\top X_n \le u)],$$

then (3.2) can be rewritten as

$$V_n(\hat{\alpha}_n, u) = V_n^0(u) - \sqrt{n}(\hat{\gamma}_n - \gamma_{n0})^\top \frac{\beta_{n0}}{\|\beta_{n0}\|^2} \tilde{R}_n(u) + o_p(1). \tag{3.7}$$

Conclude that the new $a_n(u)$ and $A_n(u)$ become real-valued with the formulas as

$$a_n(u) = \frac{\partial \tilde{R}_n(u)}{\partial \psi_n(u)} = \frac{g'(u/\kappa_n)u/\kappa_n}{\sigma_n^2(u)} \quad \text{and}$$

$$A_n(u) = \int_u^\infty \frac{[g'(z/\kappa_n)(z/\kappa_n)]^2}{\sigma_n^2(z)} F_{\kappa_n \beta_{n0}}(dz).$$

Clearly, Theorem 3.2 can be applied to these new functions.

Hall and Li (1993) showed that, if $p_n \to \infty$ as $n \to \infty$, the expectation over a few linear combinations of the components of $X_n$ behaves like the expectation over multivariate normal distributions. Note that $R_n(u) = E[g'(\beta_{n0}^\top X_n)X_n \times I(\kappa_n \beta_{n0}^\top X_n \le u)]$ and a multivariate normal distribution is elliptically-contoured. Consequently, in large dimension cases, even when $X_n$ does not follow a multivariate normal distribution, $R_n(u)$ can be viewed as an expectation on multivariate normal distributions and then the martingale transformation $T_n$ could be applied with the real-valued $a_n(u)$ and $A_n(u)$ in practice.

3.3. *The asymptotic properties under the alternative hypotheses.* Now we discuss the asymptotic properties of $\sup_{\hat{\alpha}_n \in \mathcal{S}_{\hat{q}_n}^+} |\hat{T}_n V_n(\hat{\alpha}_n, u)|$ under a sequence of alternatives. To see how sensitive the test to the alternative hypothesis is, we consider the alternatives which converge to the null hypothesis at the rate of $C_n = 1/\sqrt{sn}$:

$$H_{1n}: Y_n = g(\beta_{n0}^\top X_n, \theta_0) + C_n G(B_n^\top X_n) + \varepsilon_n, \tag{3.8}$$

where $0 < s \le 1$, $G(B_n^\top X_n)$ is a random variable with zero mean and satisfies $\mathbb{P}\{G(B_n^\top X_n) = 0\} < 1$. The constant $s$ can be fixed or dependent on $n$ tending to zero. To derive the limiting distribution of $\hat{T}_n V_n(\hat{\alpha}_n, u)$ under $H_{1n}$, we need the asymptotic properties of $\hat{q}_n$ and $\hat{\gamma}_n$, as $n \to \infty$ and $p_n \to \infty$.

PROPOSITION 4. *Assume the regularity conditions of Theorem* 3 *in Zhu et al.* (2010) *hold. Let $\hat{B}_n(1)$ be an eigenvector associating with the largest eigenvalues of $\hat{M}_n$. If $p_n C_n \to 0$ and $c_n = C_n^2 \log C_n^{-2}$, then under $H_{1n}$, we have $\mathbb{P}(\hat{q}_n = 1) \to 1$ and $\|\hat{B}_n(1) - \kappa_n \beta_{n0}\| = O_p(\sqrt{p_n} C_n)$.*

A special case is that when $C_n = 1/\sqrt{n}$ and $c_n = (\log n)/n$, we have $\mathbb{P}(\hat{q} = 1) \to 1$ and $\|\hat{B}_n(1) - \kappa\beta_{n0}\| = O_p(\sqrt{p/n})$. Next, we derive the norm consistency of $\hat{\gamma}_n$ to $\gamma_{n0}$ and an asymptotic decomposition of $\hat{\gamma}_n - \gamma_{n0}$ under $H_{1n}$. Here $\hat{\gamma}_n = (\hat{\beta}_n^\top, \hat{\theta}_n^\top)^\top$ and $\gamma_{n0} = (\beta_{n0}^\top, \theta_0^\top)^\top$ as defined before.

PROPOSITION 5. *Suppose the regularity conditions in the Supplementary Material and* (3.8) *hold. If* $np_nC_n^3 \to 0$ *and* $g''(\beta, \theta, x) \equiv 0$, *or* $np_n^2C_n^3 \to 0$, *then* $\hat{\gamma}_n$ *is a norm consistent estimator of* $\gamma_{n0}$ *with* $\|\hat{\gamma}_n - \gamma_{n0}\| = O_p(\sqrt{p_n}C_n)$. *Moreover, if* $np_n^{3/2}C_n^3 \to 0$ *and* $g''(\beta, \theta, x) \equiv 0$, *or* $np_n^{5/2}C_n^3 \to 0$, *we have*

$$
\begin{aligned}
\sqrt{n}(\hat{\gamma}_n - \gamma_{n0}) = {} & \Sigma_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_{ni} g'(\beta_{n0}, \theta_0, X_{ni}) \\
& + \sqrt{n} C_n \Sigma_n^{-1} E\big[G(X_n) g'(\beta_{n0}, \theta_0, X_n)\big] + o_p(1).
\end{aligned}
\tag{3.9}
$$

If $p_n^2/n \to 0$ and $g''(\beta, \theta, x) \equiv 0$, or $p_n^4/n \to 0$, under $H_{1n}$ with $C_n = 1/\sqrt{n}$, we have $\hat{\gamma}_n$ is a norm consistent estimator of $\gamma_{n0}$ with $\|\hat{\gamma}_n - \gamma_{n0}\| = O_p(\sqrt{p_n/n})$. Moreover, if $p_n^3/n \to 0$ and $g''(\beta, \theta, x) \equiv 0$, or $p_n^5/n \to 0$, then we have

$$
\begin{aligned}
\sqrt{n}(\hat{\gamma}_n - \gamma_{n0}) = {} & \Sigma_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_{ni} g'(\beta_{n0}, \theta_0, X_{ni}) \\
& + \Sigma_n^{-1} E\big[G(X_n) g'(\beta_{n0}, \theta_0, X_n)\big] + o_p(1).
\end{aligned}
\tag{3.10}
$$

The following theorem states the asymptotic results under various alternatives.

THEOREM 3.3. *Suppose the regularity conditions in the the Supplementary Material hold.*

(1) *If* $(p_n \log n)^3/n \to 0$ *and* $g''(\beta, \theta, x) \equiv 0$, *or* $p_n^5/n \to 0$, *under the global alternative* $H_1$, *we have in probability*

$$
\frac{1}{\sqrt{n}} \sup_{\hat{\alpha}_n \in \mathcal{S}^+_{\hat{q}n}} \big|\hat{T}_n V_n(\hat{\alpha}_n, u)\big| \longrightarrow \big|L_1(u)\big|,
$$

*where* $L_1(u)$ *is some nonzero function.*

(2) *If* $n(p_n \log n)^{\frac{3}{2}} C_n^3 \to 0$ *and* $g''(\beta, \theta, x) \equiv 0$, *or* $np_n^{5/2}C_n^3 \to 0$, *then under the local alternative* $H_{1n}$ *with* $C_n = 1/\sqrt{sn}$ *and* $s \to 0$, *we have in probability*

$$
\sqrt{s} \sup_{\hat{\alpha}_n \in \mathcal{S}^+_{\hat{q}n}} \big|\hat{T}_n V_n(\hat{\alpha}_n, u)\big| \longrightarrow \big|L_2(u)\big|.
$$

*where* $L_2(u)$ *is some nonzero function.*

(3) *If* $(p_n \log n)^3/n \to 0$ *and* $g''(\beta, \theta, x) \equiv 0$, *or* $p_n^5/n \to 0$, *then under the local alternative* $H_{1n}$ *with* $C_n = 1/\sqrt{n}$, *we have in distribution*

$$\sup_{\hat{\alpha}_n \in \mathcal{S}^+_{\hat{q}_n}} \left|\hat{T}_n V_n(\hat{\alpha}_n, u)\right| \longrightarrow \left|V_\infty(u) + G_1(u) - G_2(u)\right|,$$

*where* $V_\infty(u)$ *is a zero-mean Gaussian process given by* (3.1) *and* $G_1(u)$ *and* $G_2(u)$ *are the uniform limits of* $G_{1n}(u)$ *and* $G_{2n}(u)$ *respectively, which are as follows*:

$$G_{1n}(u) = E\left[G(X_n) I\left(\kappa_n \beta_{n0}^\top X_n \le u\right)\right],$$

$$G_{2n}(u) = E\left\{G(X_n) \int_{-\infty}^{u} a_n(z)^\top A_n^{-1}(z) a_n\left(\kappa_n \beta_{n0}^\top X_n\right) I\left(\kappa_n \beta_{n0}^\top X_n \ge z\right) \psi_n(dz)\right\}.$$

REMARK 3. The results in Theorem 3.3 show that the test is consistent against all global alternatives with fixed $C_n$ as well as the local alternatives distinct from the null at the rate of $C_n$ slower than $1/\sqrt{n}$. That is, the process diverges to infinity at the rate $\sqrt{n} C_n$. It can also detect the local alternatives converging to the null at the rate of order $1/\sqrt{n}$ in the sense that the process has a shift term from the one under the null hypothesis. These results include the fixed $p$ cases as special cases. These results indicate that although $p_n$ is divergent, the sensitivity of the test is identical to that when $p_n$ is fixed in the asymptotic sense.

3.4. *Test statistic*. In this subsection, we use the Cramér–von Mises (CM) functional to construct the test statistic. Consider

$$\mathrm{CM}_n^2 = \int_{-\infty}^{u_0} \sup_{\hat{\alpha}_n \in \mathcal{S}^+_{\hat{q}_n}} \left|\hat{T}_n V_n(\hat{\alpha}_n, u)\right|^2 F_n(du), \tag{3.11}$$

where $F_n$ is the empirical distribution function of $\beta_{n0}^\top X_{ni}/\|\beta_{n0}\|$, $1 \le i \le n$. According to Theroem 3.2 and the Extended Continuous Mapping Theorem (see Theorem 1.11.1 in van der Vaart and Wellner (1996)), we obtain, under the null,

$$\mathrm{CM}_n^2 \longrightarrow \int_{-\infty}^{u_0} \frac{B^2(\psi(u))}{\sigma^2(u)} \psi(du) \quad \text{in distribution},$$

where $B(t)$ is a standard Brownian motion and $\sigma^2(u)$ is the pointwise limit of $\sigma_n^2(u)$. Since $B(t\psi(u_0))/\sqrt{\psi(u_0)} = B(t)$ in distribution, it follows that

$$\int_{-\infty}^{u_0} B^2\left(\psi(u)\right) \psi(du) = \psi^2(u_0) \int_0^1 B(t)^2 \, dt \quad \text{in distribution}.$$

Consequently, the resulting test statistic is

$$\mathrm{ACM}_n^2 = \frac{1}{\hat{\psi}_n(u_0)^2} \int_{-\infty}^{u_0} \sup_{\hat{\alpha}_n \in \mathcal{S}^+_{\hat{q}_n}} \left|\hat{T}_n V_n(\hat{\alpha}_n, u)\right|^2 \hat{\sigma}_n^2(u) F_n(du). \tag{3.12}$$

Here we use $\hat{\psi}_n(u_0) = \frac{1}{n}\sum_{i=1}^n [Y_{ni} - g(\hat{\beta}_n^\top X_{ni}, \hat{\theta}_n)]^2 I(\hat{\alpha}_n^\top \hat{B}_n^\top X_{ni} \le u_0)$ as an estimator of $\psi(u)$. Therefore, we obtain

$$\mathrm{ACM}_n^2 \longrightarrow \int_0^1 B^2(u)\,du \quad \text{in distribution.}$$

In homoscedastic models, $\sigma_n^2(u)$ is free of $u$ and thus can be estimated by

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n [Y_{ni} - g(\hat{\beta}_n^\top X_{ni}, \hat{\theta}_n)]^2.$$

We then also have $\psi_n(u_0) = \sigma_n^2 F_{\kappa_n \beta_{n0}}(u_0)$ that can be estimated by $\hat{\sigma}_n^2 F_n(u_0)$. $\mathrm{ACM}_n^2$ becomes

$$\mathrm{ACM}_n^2 = \frac{1}{\hat{\sigma}_n^2 F_n(u_0)^2}\int_{-\infty}^{u_0} \sup_{\hat{\alpha}_n \in \mathcal{S}_{\hat{q}_n}^+} |\hat{T}_n V_n(\hat{\alpha}_n, u)|^2 F_n(du).$$

For $u_0$, as suggested by Stute and Zhu (2002), we take the 99% quantile of $F_n$ in practical use.

## 4. Numerical studies.

4.1. *Simulation studies.* In this subsection we present the results of several simulation studies to examine the performance of the proposed test. From the theoretical results in this paper and similarly as a relevant setting in Fan and Peng (2004), we set $p_n = [4n^{1/4}] - 5$ with the sample sizes $n = 100, 200, 400$ and 800 in Study 1 and 2 and try some bigger dimensions in Study 3. As there are no relevant tests dealing with the diverging dimension case, we give comparisons with some existing tests that were developed with fixed dimension, as for practical use they would be workable. Our theoretical investigations also show that the process has similar properties as Stute and Zhu (2002) under the null hypothesis even when $p_n$ is divergent.

1. Stute and Zhu's (2002) test:

$$T_n^{\mathrm{SZ}} = \frac{1}{\hat{\psi}_n(x_0)}\int_{-\infty}^{x_0} |\hat{T}_n R_n^1|^2 \hat{\sigma}_n^2\, dF_n,$$

where

$$R_n^1(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^n [Y_{ni} - g(\hat{\beta}_n^\top X_{ni}, \hat{\theta}_n)] I(\hat{\beta}_n^\top X_{ni} \le u);$$

$$\hat{T}_n R_n^1(u) = R_n^1(u) - \int_{-\infty}^u \hat{a}_n(z)^\top \hat{A}_n^{-1}(z)\left(\int_z^\infty \hat{a}_n(v) R_n^1(dv)\right)\hat{\sigma}_n^2(z) F_n(dz).$$

For $\hat{\psi}_n(x_0)$, $\hat{\sigma}_n^2$, $\hat{a}_n(z)$, $\hat{A}_n^{-1}(z)$, one can refer to their paper for detail.

2. Bierens' (1982) integrated conditional moment (ICM) test:

$$\mathrm{ICM}_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{e}_{ni}\hat{e}_{nj}\exp\Bigl(-\frac{1}{2}\|X_{ni}-X_{nj}\|\Bigr),$$

where $\hat{e}_{ni} = Y_{ni} - g(\hat{\beta}_n^\top X_{ni}, \hat{\theta}_n)$.

3. Escanciano's (2006) test:

$$\mathrm{PCvM}_n = \frac{1}{n^2}\sum_{i,j,r=1}^{n}\hat{e}_{ni}\hat{e}_{nj}\int_{S^{p_n}} I(\beta^\top X_{ni} \le \beta^\top X_{nr}) I(\beta^\top X_{nj} \le \beta^\top X_{nr})\,d\beta$$

with the critical value determination by the wild bootstrap. More details can be found in Escanciano (2006).

4. Zheng's (1996) test:

$$T_n^{\mathrm{ZH}} = \frac{\sum_{i\neq j} K((X_{ni}-X_{nj})/h)\hat{e}_{ni}\hat{e}_{nj}}{\{\sum_{i\neq j} 2K^2((X_{ni}-X_{nj})/h)\hat{e}_{ni}^2\hat{e}_{nj}^2\}^{1/2}}.$$

Here we use the kernel function $K(u) = (15/16)(1-u^2)^2 I(|u| \le 1)$ and the bandwidth $h = 1.5n^{-1/(4+p_n)}$.

The significance level is set to be $\alpha = 0.05$. We also did some simulations at $\alpha = 0.1$ and $\alpha = 0.01$ that are not reported here for saving space as the conclusions from those simulations are very similar. The simulation results are based on the averages of 2000 replications. In the following simulation studies, $a = 0$ corresponds to the null while $a \neq 0$ to the alternatives.

STUDY 1. The data are generated from the following models:

$$\begin{aligned}
H_{11}: &\quad Y_n = \beta_{n0}^\top X_n + a\exp\bigl(-(\beta_{n0}^\top X_n)^2\bigr) + \varepsilon_n;\\
H_{12}: &\quad Y_n = \beta_{n0}^\top X_n + a\cos\bigl(0.6\pi\beta_{n0}^\top X_n\bigr) + \varepsilon_n;\\
H_{13}: &\quad Y_n = \beta_{n1}^\top X_n + a\bigl(\beta_{n2}^\top X_n\bigr)^2 + \varepsilon_n;\\
H_{14}: &\quad Y_n = \beta_{n1}^\top X_n + a\exp\bigl(\beta_{n2}^\top X_n\bigr) + \varepsilon_n;
\end{aligned}$$

where $\beta_{n0} = (1,\dots,1)^\top/\sqrt{p_n}$, $\beta_{n1} = (\underbrace{1,\dots,1}_{p_{n1}},0,\dots,0)^\top/\sqrt{p_{n1}}$ and $\beta_{n2} = (0,\dots,0,\underbrace{1,\dots,1}_{p_{n1}})/\sqrt{p_{n1}}$ with $p_{n1} = [p_n/2]$. The predictor $X_n$ is from $N(0, I_{p_n})$ and $\varepsilon_n$ is a Gaussian white noise with variance 1. $H_{12}$ is a high-frequency/oscilating model and the other three are low-frequency models. In $H_{11}$ and $H_{12}$, the structural dimension equals 1 under both the null and the alternative, while, in $H_{13}$ and $H_{14}$, the structural dimension is 2 under the alternatives.

TABLE 1
*Empirical sizes and powers of* $\mathrm{ACM}_n^2$, $T_n^{\mathrm{SZ}}$, $\mathrm{PCvM}_n$, $\mathrm{ICM}_n$ *and* $T_n^{\mathrm{ZH}}$ *for* $H_{11}$ *in Study* 1

| | $a$ | $n=100$ $p=7$ | $n=200$ $p=10$ | $n=400$ $p=12$ | $n=800$ $p=16$ |
|---|---|---|---|---|---|
| $\mathrm{ACM}_n^2, \alpha=0.05$ | 0.0 | 0.0500 | 0.0530 | 0.0500 | 0.0505 |
| | 0.5 | 0.7770 | 0.9810 | 1.0000 | 1.0000 |
| $T_n^{\mathrm{SZ}}, \alpha=0.05$ | 0.0 | 0.0510 | 0.0470 | 0.0420 | 0.0495 |
| | 0.5 | 0.7825 | 0.9795 | 1.0000 | 1.0000 |
| $\mathrm{PCvM}_n, \alpha=0.05$ | 0.0 | 0.0480 | 0.0590 | 0.0650 | 0.0490 |
| | 0.5 | 0.8110 | 0.9860 | 1.0000 | 1.0000 |
| $\mathrm{ICM}_n, \alpha=0.05$ | 0.0 | 0.0070 | 0.0000 | 0.0000 | 0.0000 |
| | 0.5 | 0.3900 | 0.0910 | 0.0180 | 0.0000 |
| $T_n^{\mathrm{ZH}}, \alpha=0.05$ | 0.0 | 0.0305 | 0.0300 | 0.0330 | 0.0310 |
| | 0.5 | 0.1460 | 0.1285 | 0.1445 | 0.0980 |

The simulation results are reported in Tables 1 to 4. We can see that both $\mathrm{ACM}_n^2$ and $T_n^{\mathrm{SZ}}$ maintain the significance level very well, and reasonably have similar power performance. The empirical sizes of $\mathrm{PCvM}_n$ are also very close to the significance level, but slightly more unstable from model to model. $T_n^{\mathrm{ZH}}$ can maintain the significance level occasionally, but generally, it is conservative with smaller sizes. $\mathrm{ICM}_n$ is the worst among these tests in both the significance level maintenance and power performance. According to our experience, when $p_n$ is smaller than 5, $\mathrm{ICM}_n$ could work well. The powers of $\mathrm{ACM}_n^2$, $T_n^{\mathrm{SZ}}$ and $\mathrm{PCvM}_n$ are all very high for low frequency models $H_{11}$, $H_{13}$ and $H_{14}$. In contract, $T_n^{\mathrm{ZH}}$ has a

TABLE 2
*Empirical sizes and powers of* $\mathrm{ACM}_n^2$, $T_n^{\mathrm{SZ}}$, $\mathrm{PCvM}_n$, $\mathrm{ICM}_n$ *and* $T_n^{\mathrm{ZH}}$ *for* $H_{12}$ *in Study* 1

| | $a$ | $n=100$ $p=7$ | $n=200$ $p=10$ | $n=400$ $p=12$ | $n=800$ $p=16$ |
|---|---|---|---|---|---|
| $\mathrm{ACM}_n^2, \alpha=0.05$ | 0.0 | 0.0520 | 0.0465 | 0.0445 | 0.0515 |
| | 0.5 | 0.1445 | 0.3225 | 0.7550 | 1.0000 |
| $T_n^{\mathrm{SZ}}, \alpha=0.05$ | 0.0 | 0.0530 | 0.0480 | 0.0515 | 0.0495 |
| | 0.5 | 0.1760 | 0.3235 | 0.7350 | 0.9970 |
| $\mathrm{PCvM}_n, \alpha=0.05$ | 0.0 | 0.0530 | 0.0590 | 0.0440 | 0.0700 |
| | 0.5 | 0.1470 | 0.2320 | 0.4080 | 0.6250 |
| $\mathrm{ICM}_n, \alpha=0.05$ | 0.0 | 0.0110 | 0.0000 | 0.0000 | 0.0000 |
| | 0.5 | 0.0790 | 0.0020 | 0.0000 | 0.0000 |
| $T_n^{\mathrm{ZH}}, \alpha=0.05$ | 0.0 | 0.0325 | 0.0350 | 0.0320 | 0.0330 |
| | 0.5 | 0.0755 | 0.0775 | 0.0940 | 0.0615 |

TABLE 3
*Empirical sizes and powers of* $\mathrm{ACM}_n^2$, $T_n^{\mathrm{SZ}}$, $\mathrm{PCvM}_n$, $\mathrm{ICM}_n$ *and* $T_n^{\mathrm{ZH}}$ *for* $H_{13}$ *in Study* 1

| | $a$ | $n=100$, $p=7$ | $n=200$, $p=10$ | $n=400$, $p=12$ | $n=800$, $p=16$ |
|---|---|---|---|---|---|
| $\mathrm{ACM}_n^2, \alpha=0.05$ | 0.00 | 0.0500 | 0.0455 | 0.0435 | 0.0450 |
| | 0.25 | 0.5970 | 0.8945 | 0.9980 | 1.0000 |
| $T_n^{\mathrm{SZ}}, \alpha=0.05$ | 0.00 | 0.0505 | 0.0420 | 0.0470 | 0.0495 |
| | 0.25 | 0.5940 | 0.8980 | 0.9945 | 1.0000 |
| $\mathrm{PCvM}_n, \alpha=0.05$ | 0.00 | 0.0580 | 0.0600 | 0.0440 | 0.0570 |
| | 0.25 | 0.6160 | 0.8980 | 0.9970 | 1.0000 |
| $\mathrm{ICM}_n, \alpha=0.05$ | 0.00 | 0.0110 | 0.0000 | 0.0000 | 0.0000 |
| | 0.25 | 0.0590 | 0.0010 | 0.0000 | 0.0000 |
| $T_n^{\mathrm{ZH}}, \alpha=0.05$ | 0.00 | 0.0275 | 0.0310 | 0.0315 | 0.0340 |
| | 0.25 | 0.0730 | 0.0485 | 0.0745 | 0.0625 |

much low power for these three models. For models $H_{12}$, $\mathrm{ACM}_n^2$ and $T_n^{\mathrm{SZ}}$ have much better power performance than $\mathrm{PCvM}_n$ and $T_n^{\mathrm{ZH}}$ when $p_n$ is large. Note that $H_{12}$ is a high frequency/oscillating model. The empirical experience in this area shows that locally smoothing tests could perform better for such models in many cases. However, in our setting with relatively large dimension $p$, $T_n^{\mathrm{ZH}}$ that is a representative of locally smoothing tests, has very low power for model $H_{12}$. This is because $T_n^{\mathrm{ZH}}$ severely suffers from the dimensionality problem, which further shows the negative impact from dimensionality for nondimension reduction-type tests.

TABLE 4
*Empirical sizes and powers of* $\mathrm{ACM}_n^2$, $T_n^{\mathrm{SZ}}$, $\mathrm{PCvM}_n$, $\mathrm{ICM}_n$ *and* $T_n^{\mathrm{ZH}}$ *for* $H_{14}$ *in Study* 1

| | $a$ | $n=100$, $p=7$ | $n=200$, $p=10$ | $n=400$, $p=12$ | $n=800$, $p=16$ |
|---|---|---|---|---|---|
| $\mathrm{ACM}_n^2, \alpha=0.05$ | 0.00 | 0.0520 | 0.0460 | 0.0545 | 0.0490 |
| | 0.25 | 0.9525 | 1.0000 | 1.0000 | 1.0000 |
| $T_n^{\mathrm{SZ}}, \alpha=0.05$ | 0.00 | 0.0475 | 0.0490 | 0.0460 | 0.0555 |
| | 0.25 | 0.9605 | 0.9995 | 1.0000 | 1.0000 |
| $\mathrm{PCvM}_n, \alpha=0.05$ | 0.00 | 0.0580 | 0.0540 | 0.0570 | 0.0540 |
| | 0.25 | 0.9690 | 0.9990 | 1.0000 | 1.0000 |
| $\mathrm{ICM}_n, \alpha=0.05$ | 0.00 | 0.0050 | 0.0000 | 0.0000 | 0.0000 |
| | 0.25 | 0.3670 | 0.0740 | 0.0120 | 0.0000 |
| $T_n^{\mathrm{ZH}}, \alpha=0.05$ | 0.00 | 0.0320 | 0.0295 | 0.0325 | 0.0380 |
| | 0.25 | 0.1145 | 0.1195 | 0.1410 | 0.1145 |

The hypothetical models are all linear in Study 1. We then consider nonlinear hypothetical models in the next simulation study.

STUDY 2. The data are generated from the following models:

$$H_{21}: \quad Y_n = \exp(\beta_{n1}^\top X_n) + a(\beta_{n2}^\top X_n) + \varepsilon_n;$$

$$H_{22}: \quad Y_n = \exp(X_n^1) + a\{(X_n^2)^3 + \cos(\pi X_n^3) + X_n^4 \cdot X_n^5\} + \varepsilon_n;$$

$$H_{23}: \quad Y_n = \exp(X_n^1) + a\{X_n^2 + \cos(\pi X_n^3) + (X_n^4)^3 - (X_n^5)^2 - X_n^6 \cdot X_n^7\} + \varepsilon_n.$$

Here $\beta_{n1} = (\underbrace{1, \ldots, 1}_{p_{n1}}, 0, \ldots, 0)^\top / \sqrt{p_{n1}}$ and $\beta_{n2} = (0, \ldots, 0, \underbrace{1, \ldots, 1}_{p_{n1}})^\top / \sqrt{p_{n1}}$ with $p_{n1} = [p_n/2]$, $X_n^i$ is the $i$th component of $X_n$, $\varepsilon_n$ is $N(0, 1)$ and $X_n$ is $N(0, I_{p_n})$ independent of $\varepsilon_n$.

We report the empirical sizes and powers in Tables 5–7. From these tables, we can obviously see that the empirical sizes of $\mathrm{ACM}_n^2$, $T_n^{\mathrm{SZ}}$ and $\mathrm{PCvM}_n$ are again very close to the significance level, while $T_n^{\mathrm{ZH}}$ can only maintain the level sometimes. We also did some more simulations that are unreported here and found that $T_n^{\mathrm{ZH}}$ is in many cases conservative with even smaller empirical sizes. $\mathrm{ICM}_n$ is still the worst one. For these three models the empirical powers of $\mathrm{ACM}_n^2$ are higher than the other competitors, while $T_n^{\mathrm{SZ}}$'s empirical powers grow very slow in models $H_{21}$ and $H_{22}$. This would confirm the theoretical result that $T_n^{\mathrm{SZ}}$ is not an omnibus test.

As in practical use, the ratio between the dimension $p$ and the sample size $n$ is hard to be judged whether it has the rate of convergence in theory, we then consider the next simulation study.

TABLE 5
*Empirical sizes and powers of* $\mathrm{ACM}_n^2$, $T_n^{\mathrm{SZ}}$, $\mathrm{PCvM}_n$, $\mathrm{ICM}_n$ *and* $T_n^{\mathrm{ZH}}$ *for* $H_{21}$ *in Study* 2

| | $a$ | $n = 100$, $p = 7$ | $n = 200$, $p = 10$ | $n = 400$, $p = 12$ | $n = 800$, $p = 16$ |
|---|---|---|---|---|---|
| $\mathrm{ACM}_n^2, \alpha = 0.05$ | 0.0 | 0.0575 | 0.0550 | 0.0585 | 0.0530 |
| | 0.5 | 0.1295 | 0.1895 | 0.3030 | 0.5790 |
| $T_n^{\mathrm{SZ}}, \alpha = 0.05$ | 0.0 | 0.0650 | 0.0535 | 0.0550 | 0.0550 |
| | 0.5 | 0.0755 | 0.0970 | 0.0835 | 0.1195 |
| $\mathrm{PCvM}_n, \alpha = 0.05$ | 0.0 | 0.0470 | 0.0560 | 0.0690 | 0.0510 |
| | 0.5 | 0.1310 | 0.2000 | 0.2760 | 0.4450 |
| $\mathrm{ICM}_n, \alpha = 0.05$ | 0.0 | 0.0050 | 0.0000 | 0.0000 | 0.0000 |
| | 0.5 | 0.0210 | 0.0020 | 0.0000 | 0.0000 |
| $T_n^{\mathrm{ZH}}, \alpha = 0.05$ | 0.0 | 0.0445 | 0.0365 | 0.0410 | 0.0380 |
| | 0.5 | 0.0690 | 0.0765 | 0.0770 | 0.0545 |

TABLE 6
*Empirical sizes and powers of* $\mathrm{ACM}_n^2$, $T_n^{\mathrm{SZ}}$, $\mathrm{PCvM}_n$, $\mathrm{ICM}_n$ *and* $T_n^{\mathrm{ZH}}$ *for* $H_{22}$ *in Study* 2

| | $a$ | $n = 100$ $p = 7$ | $n = 200$ $p = 10$ | $n = 400$ $p = 12$ | $n = 800$ $p = 16$ |
|---|---|---|---|---|---|
| $\mathrm{ACM}_n^2, \alpha = 0.05$ | 0.00 | 0.0575 | 0.0540 | 0.0550 | 0.0505 |
| | 0.25 | 0.1840 | 0.2770 | 0.4725 | 0.7840 |
| $T_n^{\mathrm{SZ}}, \alpha = 0.05$ | 0.00 | 0.0615 | 0.0560 | 0.0555 | 0.0495 |
| | 0.25 | 0.1055 | 0.1350 | 0.1735 | 0.2670 |
| $\mathrm{PCvM}_n, \alpha = 0.05$ | 0.00 | 0.0510 | 0.0620 | 0.0430 | 0.0600 |
| | 0.25 | 0.0770 | 0.0870 | 0.0980 | 0.1620 |
| $\mathrm{ICM}_n, \alpha = 0.05$ | 0.00 | 0.0050 | 0.0000 | 0.0000 | 0.0000 |
| | 0.25 | 0.0150 | 0.0000 | 0.0000 | 0.0000 |
| $T_n^{\mathrm{ZH}}, \alpha = 0.05$ | 0.00 | 0.0310 | 0.0410 | 0.0310 | 0.0325 |
| | 0.25 | 0.0730 | 0.0695 | 0.0770 | 0.0615 |

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

$$H_{31} : Y_n = \beta_{n0}^{\top} X_n + a \exp(-\beta_{n0}^{\top} X_n) + \varepsilon_n;$$
$$H_{32} : Y_n = \exp(\beta_{n1}^{\top} X_n) + a(\beta_{n2}^{\top} X_n) + \varepsilon_n;$$

where $\beta_{n0}$, $\beta_{n1}$, $\beta_{n2}$, $X_n$ and $\varepsilon_n$ are the same as in Study 1. We set $p = 0.25n$ in model $H_{31}$ and $p = 0.1n$ in model $H_{32}$ with the sample size $n = 100, 200, 300, 400$.

TABLE 7
*Empirical sizes and powers of* $\mathrm{ACM}_n^2$, $T_n^{\mathrm{SZ}}$, $\mathrm{PCvM}_n$, $\mathrm{ICM}_n$ *and* $T_n^{\mathrm{ZH}}$ *for* $H_{23}$ *in Study* 2

| | $a$ | $n = 100$ $p = 7$ | $n = 200$ $p = 10$ | $n = 400$ $p = 12$ | $n = 800$ $p = 16$ |
|---|---|---|---|---|---|
| $\mathrm{ACM}_n^2, \alpha = 0.05$ | 0.0 | 0.0545 | 0.0620 | 0.0580 | 0.0520 |
| | 0.1 | 0.2310 | 0.4040 | 0.5855 | 0.8645 |
| $T_n^{\mathrm{SZ}}, \alpha = 0.05$ | 0.0 | 0.0660 | 0.0560 | 0.0500 | 0.0535 |
| | 0.1 | 0.1830 | 0.3020 | 0.4580 | 0.7355 |
| $\mathrm{PCvM}_n, \alpha = 0.05$ | 0.0 | 0.0580 | 0.0570 | 0.0580 | 0.0560 |
| | 0.1 | 0.1500 | 0.2150 | 0.3440 | 0.5970 |
| $\mathrm{ICM}_n, \alpha = 0.05$ | 0.0 | 0.0110 | 0.0000 | 0.0000 | 0.0000 |
| | 0.1 | 0.0210 | 0.0000 | 0.0000 | 0.0000 |
| $T_n^{\mathrm{ZH}}, \alpha = 0.05$ | 0.0 | 0.0430 | 0.0340 | 0.0415 | 0.0340 |
| | 0.1 | 0.0480 | 0.0500 | 0.0600 | 0.0525 |

TABLE 8
*Empirical sizes and powers of* $\mathrm{ACM}_n^2$, $T_n^{\mathrm{SZ}}$, $\mathrm{PCvM}_n$, $\mathrm{ICM}_n$ *and* $T_n^{\mathrm{ZH}}$ *for* $H_{31}$ *in Study* 3

| | $a$ | $n=100$ $p=25$ | $n=200$ $p=50$ | $n=300$ $p=75$ | $n=400$ $p=100$ |
|---|---|---|---|---|---|
| $\mathrm{ACM}_n^2, \alpha=0.05$ | 0.0 | 0.0455 | 0.0465 | 0.0525 | 0.0580 |
| | 0.1 | 0.3115 | 0.5695 | 0.7495 | 0.8565 |
| $T_n^{\mathrm{SZ}}, \alpha=0.05$ | 0.0 | 0.0525 | 0.0425 | 0.0585 | 0.0525 |
| | 0.1 | 0.2780 | 0.5095 | 0.7095 | 0.8210 |
| $\mathrm{PCvM}_n, \alpha=0.05$ | 0.0 | 0.1010 | 0.0830 | 0.0830 | 0.0920 |
| | 0.1 | 0.3910 | 0.6110 | 0.7640 | 0.8830 |
| $\mathrm{ICM}_n, \alpha=0.05$ | 0.0 | 0.3860 | 1.0000 | 1.0000 | 1.0000 |
| | 0.1 | 0.3400 | 0.9990 | 0.9990 | 0.9990 |
| $T_n^{\mathrm{ZH}}, \alpha=0.05$ | 0.0 | 0.1950 | 0.0095 | 0.0000 | 0.0000 |
| | 0.1 | 0.2115 | 0.0090 | 0.0000 | 0.0000 |

The simulation results are reported in Tables 8–9. For model $H_{31}$, we can see that $\mathrm{ACM}_n^2$ and $T_n^{\mathrm{SZ}}$ still perform very well even when the dimension $p$ is much larger than the cases in Study 1, while the empirical sizes of $\mathrm{PCvM}_n$ can not maintain the significance level. From our experience, if the hypothetical model is relatively simple such as a linear model, the proposed test can perform well even when $p < 0.4n$. For the nonlinear model $H_{32}$, although the dimension is smaller than the case in model $H_{31}$ with $p = 0.1n$, the empirical sizes of our test are higher than the significance level. The other unreported results also tell that the empirical sizes of our test are even higher when $p > 0.1n$. These massages indicate that the

TABLE 9
*Empirical sizes and powers of* $\mathrm{ACM}_n^2$, $T_n^{\mathrm{SZ}}$, $\mathrm{PCvM}_n$, $\mathrm{ICM}_n$ *and* $T_n^{\mathrm{ZH}}$ *for* $H_{32}$ *in Study* 3

| | $a$ | $n=100$ $p=10$ | $n=200$ $p=20$ | $n=300$ $p=30$ | $n=400$ $p=40$ |
|---|---|---|---|---|---|
| $\mathrm{ACM}_n^2, \alpha=0.05$ | 0.0 | 0.0650 | 0.0860 | 0.0980 | 0.1295 |
| | 0.5 | 0.1500 | 0.2510 | 0.3815 | 0.4960 |
| $T_n^{\mathrm{SZ}}, \alpha=0.05$ | 0.0 | 0.0650 | 0.0560 | 0.0535 | 0.0630 |
| | 0.5 | 0.0810 | 0.1035 | 0.1215 | 0.1475 |
| $\mathrm{PCvM}_n, \alpha=0.05$ | 0.0 | 0.0550 | 0.0640 | 0.0820 | 0.0670 |
| | 0.5 | 0.0550 | 0.0810 | 0.0980 | 0.0940 |
| $\mathrm{ICM}_n, \alpha=0.05$ | 0.0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 0.5 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $T_n^{\mathrm{ZH}}, \alpha=0.05$ | 0.0 | 0.0255 | 0.0135 | 0.3290 | 0.3245 |
| | 0.5 | 0.0430 | 0.0150 | 0.3585 | 0.3280 |

dimension cannot be too large unless some theory can be developed to handle the higher-dimensional scenarios. As expected, the competitors $T_n^{SZ}$ and $\mathrm{PCvM}_n$ can control the empirical sizes in most cases, while the empirical powers of these two tests are much lower than our test. The other two tests $T_n^{ZH}$ and $\mathrm{ICM}_n$ are still the worst among these tests and their type I errors are completely out of control.

Therefore, overall, the proposed test in this paper performs well and can detect different alternatives and in the high-dimensional cases, it shows the advantage over the other competitors.

4.2. *A real data example.* In this subsection we analyze the baseball salary data set that can be obtain through the website http://www4.stat.ncsu.edu/~boos/var.select/baseball.html. This data set contains 337 Major League Baseball players on the salary $Y$ from the year 1992 and 16 performance measures from the year 1991. The performance measures are $X_1$: Batting average, $X_2$: On-base percentage, $X_3$: runs, $X_4$: hits, $X_5$: doubles, $X_6$: triples, $X_7$: home runs, $X_8$: runs batted in, $X_9$: walks, $X_{10}$: strike-outs, $X_{11}$: stolen bases, $X_{12}$: errors, $X_{13}$: Indicator of free agency eligibility, $X_{14}$: Indicators of free agent in 1991/2, $X_{15}$: Indicators of arbitration eligibility and $X_{16}$: Indicators of arbitration in 1991/2. The dummy variables $X_{13}$–$X_{16}$ measure the freedom of movement of a player to another team. For easy interpretation, we standardize all variables separately. To obtain the regression relationship between $Y$ and the performance measures $X = (X_1, \dots, X_{16})^\top$, we first test for a linear regression model by the proposed test. The value of the test statistic is $\mathrm{ACM}_n^2 = 1.3651$ with the $p$-value equal to 0.077. Since the $p$-value is small although it is larger than, say, 0.05, an often-used significance level, we may consider a more plausible model to better fit this dataset. Hence we apply the dimension reduction techniques. Recalling in Section 2.3, we claimed that to estimate the central subspace, the CSE method is used. The estimated structural dimension of this datset is $\hat{q}_n = 1$. This means that $Y$ may be conditionally independent of $X$ given the projected covariate $\hat{\beta}_1^\top X$ where

$$\begin{aligned}\hat{\beta}_1 = (&0.0463, -0.1078, 0.0383, 0.2447, -0.0322, -0.0436, 0.0545, 0.2229,\\ &0.1173, -0.1718, 0.0491, -0.0494, 0.7479, -0.0965, 0.5022, -0.0165)^\top,\end{aligned}$$

is the first direction obtained by CSE. The scatter plot of $Y$ against $\hat{\beta}_1^\top X$ is presented in Figure 1(a). It indicates that a linear regression model for $(Y, X)$ may not be reasonable. To further exhaust possible projected covariates, we consider the second projected covariate $\hat{\beta}_2^\top X$ obtained by CSE. The 3-D plot of $Y$ against $(\hat{\beta}_1^\top X, \hat{\beta}_2^\top X)$ is presented in Figure 2. This figure shows that the second projected covariate $\hat{\beta}_2^\top X$ has no information in predicting the response $Y$, as the plot along $\hat{\beta}_2^\top X$ is almost invariable. This means that the projection of the data onto the subspace $\hat{\beta}_1^\top X$ would already contain most of the regression information of $(Y, X)$.

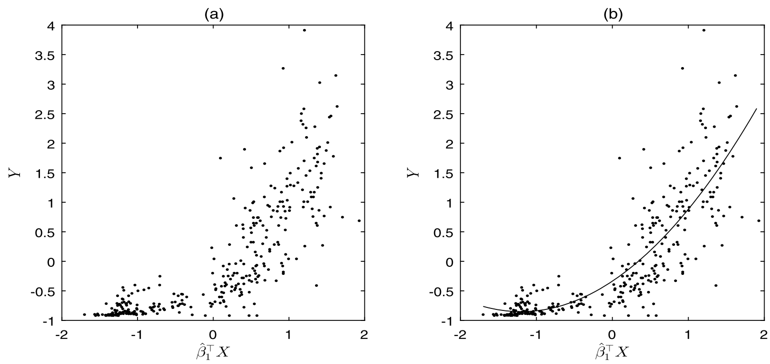


FIG. 1. *Scatter plots of the response $Y_n$ against the projected covariate $\hat{\beta}_1^\top X$ and the fitted quadratic polynomial curve where the direction $\hat{\beta}_1$ is obtained by CSE.*

Figure 1(a) seems to suggest a quadratic polynomial of $\hat{\beta}_1^\top X$ to fit the data. Hence we use the following regression model:

$$Y = \theta_1 + \theta_2(\beta^\top X) + \theta_3(\beta^\top X)^2 + \varepsilon.$$

Figure 1(b) adds the fitted curve on the scatter plot. The value of the test statistic $\mathrm{ACM}_n^2 = 0.1038$ and the $p$-value is about 0.83. Therefore the above regression model is plausible.

**5. Discussions.** In this paper we investigate model checking for regressions when the dimension of predictors diverges to infinity as the sample size tends to infinity. Three remarkable features are worthwhile to discuss. First, although the empirical process is similar to that in Stute and Zhu (2002), it involves much more

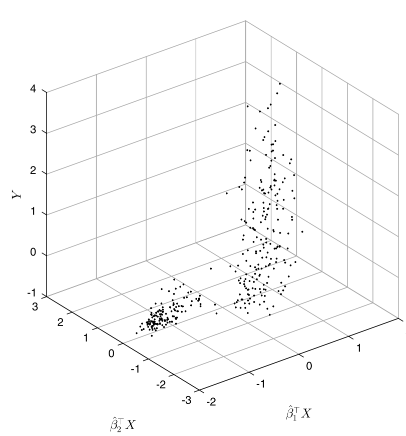


FIG. 2. *Scatter plot of the response $Y_n$ against the projected covariates $(\hat{\beta}_1^\top X, \hat{\beta}_2^\top X)$ where the directions $(\hat{\beta}_1, \hat{\beta}_2)$ are obtained by CSE.*

difficult estimation issues in the construction procedure of test statistics. Second, as the Khmaladze martingale transformation has become an important methodology for model checking as its asymptotically distribution-free property, we suggest another way to construct the transformation, rather than directly targeting the limit of the shift terms in the fixed dimension cases. The transformed process still has the same limiting Gaussian process as that with fixed dimension. This provides us an easy way to handle the diverging dimension cases. Third, the model adaptation property shows its advantage in maintaining the significance level and enhancing power performance.

The research also leaves some unsolved topics. They are beyond the scope of this paper and deserve further studies.

The first topic is about how to relax the condition on the diverging rate of the dimension. In this paper, we cannot have faster rate than $p_n = O(n^{1/3}/\log n)$ even when the hypothetical model is linear. From the lemmas, it seems not easy to improve. We guess that this rate would be close to optimal although we have not proved this result.

The second topic is about the assumption that $\tilde{\beta}_{n0} \in \mathcal{S}_{E(Y_n|X_n)}$ under the alternatives such that $E(e_n|B_n^\top X_n) = G(B_n^\top X_n) - g(\tilde{\beta}_{n0}^\top X_n, \tilde{\theta}_0) \neq 0$. If $g(\beta^\top X_n, \theta) = \beta^\top X_n$ follows a linear regression model, we have shown that $\tilde{\beta}_{n0} \in \mathcal{S}_{E(Y_n|X_n)}$ under the other conditions on the model we assume in the paper. For other models, this assumption may not hold. A simple solution is that we consider $\tilde{B}_n = (\tilde{\beta}_{n0}, B_n)$. It is easy to see that under the null, $B_n = \kappa_n \beta_{n0}$ and $\tilde{\beta}_{n0} = \beta_{n0}$. Then $E(e_n|\tilde{B}_n^\top X_n) = E(e_n|\beta_{n0}^\top X_n) = 0$. Under the alternatives, we can have that $E(e_n|\tilde{B}_n^\top X_n) = G(B_n^\top X_n) - g(\tilde{\beta}_{n0}^\top X_n, \tilde{\theta}_0) \neq 0$. Although the theoretical developments under the alternatives are similar as before, the limiting null distribution is no longer tractable. This is because the weak limit of the process under the null is $\sup_{\alpha \in \mathcal{S}_2^+} |B(\psi(\frac{u}{a_1+a_2}))|$ and $\mathcal{S}_2^+$ is not a single point $\{+1\}$ any more, its distribution is untractable. Here $\alpha = (a_1, a_2)$ and $B(t)$ is a Brownian motion. This destroys the advantage of the innovative process approach described in this paper.

We now provide a partial solution that is a refined adaptive-to-model method. The test statistic construction is based on the following fact. Define a new index set $\tilde{\mathcal{S}}_{q_n+1}^+ = \{\alpha_n = (I(q_n \neq 1)a_0, a_1, \dots, a_{q_n})^\top \in \mathbb{R}^{q_n+1} : \|\alpha_n\| = 1 \text{ and } a_1 \geq 0\}$. Under the null, $\tilde{\mathcal{S}}_{q_n+1}^+ = \{(0,1)^\top\}$. For any $\alpha_n \in \tilde{\mathcal{S}}_{q_n+1}^+$,

$$E\big[e_n I\big(\alpha_n^\top(\tilde{\beta}_{n0}, B_n)^\top X_n \leq u\big)\big] = E\big[e_n I\big(\kappa_n \beta_{n0}^\top X_n \leq u\big)\big] = 0.$$

Under the alternatives with $q_n \neq 1$, there exists an $\alpha_n \in \tilde{\mathcal{S}}_{q_n+1}^+$ such that

$$E\big[e_n I\big(\alpha_n^\top(\tilde{\beta}_{n0}, B_n)^\top X_n \leq u\big)\big] \neq 0.$$

Therefore, the revised test statistic becomes

$$V_n(\hat{\alpha}_n, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [Y_{ni} - g(\hat{\beta}_n^\top X_{ni}, \hat{\theta}_n)] I[\hat{\alpha}_n^\top(\hat{\beta}_n, \hat{B}_n)^\top X_{ni} \leq u],$$

$$V_n(u) = \sup_{\hat{\alpha}_n \in \tilde{\mathcal{S}}^+_{\hat{q}_n+1}} \left|V_n(\hat{\alpha}_n, u)\right|.$$

Then $V_n(u)$ can still have the desirable property described in the paper. That is, under the null, $\tilde{\mathcal{S}}^+_{\hat{q}_n+1} = \{(0, 1)^\top\}$ with a probability going to one. This is because $\hat{q}_n = 1$ in probability. Thus, the limiting null distribution of our test statistic is tractable. Under the alternatives, the theoretical results can be similar to those in the previous version.

As we mentioned above, this solution partly relaxes the condition. If the alternative model is a semiparametric single-index model with $q_n = 1$, it follows that $\tilde{\mathcal{S}}^+_{q_n+1} = \{(0, 1)^\top\}$. We still need to assume that $\tilde{\beta}_{n0}$ is proportional to $B_n$, otherwise under the alternatives $E(e_n|\alpha_n^\top \tilde{B}_n^\top X_n) = E(e_n|B_n^\top X_n)$ would not be necessarily equal to $G(B_n^\top X_n) - g(\tilde{\beta}_{n0}^\top X_n) \neq 0$. Thus, how to obtain a complete solution is still an interesting topic.

The third topic is about the assumption on the error term. We have assumed that the error term $\varepsilon_n$ has dimension reduction structures under the null and alternatives: $\varepsilon_n = V_1(\beta_{n0}^\top X_n)\tilde{\varepsilon}$ and $\varepsilon_n = V_2(B_n^\top X_n)\tilde{\varepsilon}_n$ respectively. In effect, if the methods for identifying the central mean subspace can be applied, such as pHd (Li (1992)), these conditions can be removed. However, as we mentioned in the main context, when $p_n$ is divergent, we have no relevant asymptotic results about pHd and guess that even if we can get some results, the convergence rate would be very slow. This is because it involves the square of Hessian matrix, not Hessian matrix itself and then the convergence rate of its estimator would have a rate of order $\sqrt{p_n^4/n}$ rather than $\sqrt{p_n^2/n}$. The theoretical development in our setting becomes difficult. Thus, we may discuss this issue in a further study.

## APPENDIX

In this section we only give the proof for Theorem 3.1. The regularity conditions and the proofs of the other theoretical results are put in the Supplementary Material (Tan and Zhu (2019)). This is because the main focus of this paper is to show at what diverging rate of $p_n$ the convergence of the empirical process can be derived. This theorem presents the relevant results.

PROOF OF THEOREM 3.1. Under the null hypothesis, we have $\mathbb{P}(\hat{q}_n = 1) \to 1$. Thus we need only to work on the event $\{\hat{q}_n = 1\}$. It follows that $\hat{\alpha}_n = 1$ and we can rewrite $V_n(\hat{\alpha}_n, u)$ as

$$\begin{aligned} V_n(\hat{\alpha}_n, u) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[Y_{ni} - g\left(\hat{\beta}_n^\top X_{ni}, \hat{\theta}_n\right)\right] I\left(\hat{B}_n^\top X_{ni} \le u\right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[Y_{ni} - g\left(\hat{\beta}_n^\top X_{ni}, \hat{\theta}_n\right)\right] I\left(\kappa_n \beta_{n0}^\top X_{ni} \le u\right) \end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[Y_{ni}-g\big(\hat{\beta}_n^{\top}X_{ni},\hat{\theta}_n\big)\big] \\
&\times\big[I\big(\hat{B}_n^{\top}X_{ni}\le u\big)-I\big(\kappa_n\beta_{n0}^{\top}X_{ni}\le u\big)\big] \\
=:{}& V_{n1}+V_{n2}.
\end{aligned}
$$

Recall $\gamma=(\beta^{\top},\theta^{\top})^{\top}$. Then we obtain that

$$
\begin{aligned}
V_{n1}={}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_{ni}I\big(\kappa_n\beta_{n0}^{\top}X_{ni}\le u\big) \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[g\big(\hat{\beta}_n^{\top}X_{ni},\hat{\theta}_n\big)-g\big(\beta_{n0}^{\top}X_{ni},\theta_0\big)\big]I\big(\kappa_n\beta_{n0}^{\top}X_{ni}\le u\big) \\
={}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_{ni}I\big(\kappa_n\beta_{n0}^{\top}X_{ni}\le u\big) \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\gamma}_n-\gamma_{n0})^{\top}g'(\beta_{n0},\theta_0,X_{ni})I\big(\kappa_n\beta_{n0}^{\top}X_{ni}\le u\big) \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\gamma}_n-\gamma_{n0})^{\top}g''(\beta_{1n},\theta_{1n},X_{ni})(\hat{\gamma}_n-\gamma_{n0})I\big(\kappa_n\beta_{n0}^{\top}X_{ni}\le u\big) \\
={}&V_{n11}-V_{n12}-V_{n13},
\end{aligned}
$$

where $(\beta_{1n},\theta_{1n})$ lies between $(\hat{\beta}_n,\hat{\theta}_n)$ and $(\beta_{n0},\theta_0)$. For the third term $V_{n13}$ in $V_{n1}$, note that

$$
\begin{aligned}
&E\sup_u\left\|\sum_{i=1}^{n}g''(\beta_{1n},\theta_{1n},X_{ni})I\big(\kappa_n\beta_{n0}^{\top}X_{ni}\le u\big)\right\| \\
&\quad\le\sum_{i=1}^{n}E\sup_u\big\|g''(\beta_{1n},\theta_{1n},X_{ni})I\big(\kappa_n\beta_{n0}^{\top}X_{ni}\le u\big)\big\| \\
&\quad\le\sum_{i=1}^{n}\Big[E\sup_u\big\|g''(\beta_{1n},\theta_{1n},X_{ni})I\big(\kappa_n\beta_{n0}^{\top}X_{ni}\le u\big)\big\|^2\Big]^{1/2} \\
&\quad\le\sum_{i=1}^{n}\left(\sum_{j,k=1}^{p+d}Eg''_{jk}(\beta_{1n},\theta_{1n},X_{ni})^2\right)^{1/2}\le Cn(p_n+d).
\end{aligned}
$$

Therefore $V_{n13}=\frac{1}{\sqrt{n}}\frac{p_n}{n}n(p_n+d)O_p(1)=o_p(1)$ uniformly in $u$. If $g''(\beta,\theta,x)\equiv 0$, then $V_{n13}=0$. For $V_{n12}$, recall that $R_n(u)=E[g'(\beta_{n0},\theta_0,X_n)I(\kappa_n\beta_{n0}^{\top}X\le u)]$.

Then we decompose $V_{n12}$ as

$$V_{n12} = \sqrt{n}(\hat{\gamma}_n - \gamma_{n0})^\top R_n(u) + \sqrt{n}(\hat{\gamma}_n - \gamma_{n0})^\top \left( \frac{1}{n} \sum_{i=1}^n g'(\beta_{n0}, \theta_0, X_{ni}) I(\kappa_n \beta_{n0}^\top X_{ni} \le u) - R_n(u) \right).$$

For the second term in $V_{n12}$, Lemma 3 in the Supplementary Material (Tan and Zhu (2019)) yields

$$\sup_u \left\| \frac{1}{n} \sum_{i=1}^n g'(\beta_{n0}, \theta_0, X_{ni}) I(\kappa_n \beta_{n0}^\top X_{ni} \le u) - R_n(u) \right\| = o_p\left( \sqrt{\frac{p_n^{3/2} \log n}{n}} \right).$$

Conclude that

$$\sqrt{n}(\hat{\gamma}_n - \gamma_{n0})^\top \left( \frac{1}{n} \sum_{i=1}^n g'(\beta_{n0}, \theta_0, X_{ni}) I(\kappa_n \beta_{n0}^\top X_{ni} \le u) - R_n(u) \right) = \sqrt{\frac{p_n^{5/2} \log n}{n}} o_p(1) = o_p(1).$$

Since $\|R_n(u)\| = O(1)$ uniformly in $u$, by Proposition 2, we have

$$V_{n12} = R_n(u)^\top \Sigma_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n [Y_{ni} - g(\beta_{n0}^\top X_{ni}, \theta_0)] g'(\beta_{n0}, \theta_0, X_{ni}) + o_p(1).$$

Therefore, we obtain that

$$V_{n1} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_{ni} I(\kappa_n \beta_{n0}^\top X_{ni} \le u) - \frac{1}{\sqrt{n}} R_n(u)^\top \Sigma_n^{-1} \sum_{i=1}^n \varepsilon_{ni} g'(\beta_{n0}, \theta_0, X_{ni}) + o_p(1). \tag{A.1}$$

Now we consider the term $V_{n2}$. It can be decomposed as

$$\begin{aligned} V_{n2} = {} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_{ni} [I(\hat{B}_n^\top X_{ni} \le u) - I(\kappa_n \beta_{n0}^\top X_{ni} \le u)] \\ & - \frac{1}{\sqrt{n}} \sum_{i=1}^n [g(\hat{\beta}_n^\top X_{ni}, \hat{\theta}_n) - g(\beta_{n0}^\top X_{ni}, \theta_0)] \\ & \times [I(\hat{B}_n^\top X_{ni} \le u) - I(\kappa_n \beta_{n0}^\top X_{ni} \le u)] \\ = {} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_{ni} [I(\hat{B}_n^\top X_{ni} \le u) - I(\kappa_n \beta_{n0}^\top X_{ni} \le u)] \end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\gamma}_n-\gamma_{n0})^\top g'(\beta_{n0},\theta_0,X_{ni})\big[I\big(\hat{B}_n^\top X_{ni}\le u\big)-I\big(\kappa_n\beta_{n0}^\top X_{ni}\le u\big)\big]\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\gamma}_n-\gamma_{n0})^\top g''(\beta_{1n},\theta_{1n},X_{ni})(\hat{\gamma}_n-\gamma_{n0})\\
&\times\big[I\big(\hat{B}_n^\top X_{ni}\le u\big)-I\big(\kappa_n\beta_{n0}^\top X_{ni}\le u\big)\big]\\
={}&V_{n21}-V_{n22}-V_{n23}.
\end{aligned}
$$

By Lemma 6 in the Supplementary Material (Tan and Zhu (2019)), we obtain that $V_{n21}=o_p(1)$ uniformly in $u$ when $(p_n\log n)^3/n\to 0$. For the second term $V_{n22}$, let

$$
W_n(\beta,u)=E\big\{g'(\beta_{n0},\theta_0,X_n)\big[I\big(\beta^\top X_{ni}\le u\big)-I(\kappa_n\beta_{n0}X_{ni}\le u)\big]\big\}.
$$

By Lemma 7 in the Supplementary Material, we have

$$
\begin{aligned}
&\sup_u\bigg\|\frac{1}{n}\sum_{i=1}^{n}g'(\beta_{n0},\theta_0,X_{ni})\big[I\big(\hat{B}_n^\top X_{ni}\le u\big)-I\big(\kappa_n\beta_{n0}^\top X_{ni}\le u\big)\big]-W_n(\hat{B}_n,u)\bigg\|\\
&=o_p\bigg(\sqrt{\frac{p_n^{3/2}\log n}{n}}\bigg).
\end{aligned}
$$

Therefore, we derive that

$$
V_{n22}=\sqrt{n}(\hat{\gamma}_n-\gamma_{n0})^\top W_n(\hat{B}_n,u)+\sqrt{\frac{p_n^{5/2}\log n}{n}}o_p(1).
$$

Let $R_n(\beta,u)=E[g'(\beta_{n0},\theta_0,X_n)I(\beta^\top X_{ni}\le u)]$. Then $W_n(\hat{B}_n,u)=R_n(\hat{B}_n,u)-R_n(\kappa_n\beta_{n0},u)$. By Taylor's expansion around $\kappa_n\beta_{n0}$ and condition (B2) in the Supplementary Material, we have

$$
\begin{aligned}
V_{n22}={}&\sqrt{n}(\hat{\gamma}_n-\gamma_{n0})^\top\bigg(\frac{\partial R_n}{\partial\beta}(\kappa_n\beta_{n0},u)(\hat{B}_n-\kappa_n\beta_{n0})+o_p\big(\|\hat{B}_n-\kappa_n\beta_{n0}\|\big)\bigg)\\
&+\sqrt{\frac{p_n^{5/2}\log n}{n}}o_p(1).
\end{aligned}
$$

It follows that $V_{n22}=o_p(1)$ uniformly in $u$.

We can obtain that $V_{n23}=o_p(1)$ uniformly in $u$ similarly as that for $V_{n13}$. Combining these with (A.1), we obtain that

$$
\begin{aligned}
V_n(\hat{\alpha}_n,u)={}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_{ni}I\big(\kappa_n\beta_{n0}^\top X_{ni}\le u\big)\\
&-\frac{1}{\sqrt{n}}R_n(u)^\top\Sigma_n^{-1}\sum_{i=1}^{n}\varepsilon_{ni}g'(\beta_{n0},\theta_0,X_{ni})+o_p(1).
\end{aligned}
\tag{A.2}
$$

By Theorem 2.11.22 in van der Vaart and Wellner (1996), we have the first two terms of the right-hand side of (A.2) are asymptotically tight.

Now we consider the convergence of finite-dimensional distributions. Let $Y_{ni} = (Y_{ni}(u_1), \ldots, Y_{ni}(u_m))^\top$ where

$$Y_{ni}(u) = \frac{1}{\sqrt{n}}\varepsilon_{ni}\bigl[I\bigl(\kappa_n\beta_{n0}^\top X_{ni} \le u\bigr) - R_n(u)^\top \Sigma_n^{-1} g'(\beta_{n0}, \theta_0, X_{ni})\bigr].$$

For any $\delta > 0$, we have

$$\begin{aligned}\sum_{i=1}^n E\|Y_{ni}\|^2 I\bigl(\|Y_{ni}\| > \delta\bigr) &= nE\bigl\{\|Y_{n1}\|^2 I\bigl(\|Y_{n1}\| > \delta\bigr)\bigr\}\\ &\le n\bigl\{E\|Y_{n1}\|^4\bigr\}^{1/2}\bigl\{\mathbb{P}\bigl(\|Y_{n1}\| > \delta\bigr)\bigr\}^{1/2}.\end{aligned}$$

Since

$$\begin{aligned}\mathbb{P}\bigl(\|Y_{n1}\| > \delta\bigr) &= \mathbb{P}\bigl(Y_{n1}(u_1)^2 + \cdots + Y_{n1}(u_m)^2 > \delta^2\bigr)\\ &\le \sum_{j=1}^m \mathbb{P}\biggl(Y_{n1}(u_j)^2 > \frac{\delta^2}{m}\biggr)\end{aligned}$$

and

$$\begin{aligned}&\mathbb{P}\biggl(Y_{n1}(u)^2 > \frac{\delta^2}{m}\biggr)\\ &\quad= \mathbb{P}\biggl(\varepsilon_{n1}^2\bigl[I\bigl(\kappa_n\beta_{n0}^\top X_{n1} \le u\bigr) - R_n(u)^\top \Sigma_n^{-1} g'(\beta_{n0}, \theta_0, X_{n1})\bigr]^2 > \frac{n\delta^2}{m}\biggr)\\ &\quad\le \frac{2mE\varepsilon_{n1}^2 + 2mE\{\varepsilon_{n1}^2[R_n(u)^\top \Sigma_n^{-1} g'(\beta_{n0}, \theta_0, X_{n1})]^2\}}{n\delta^2}\\ &\quad\le \frac{2mE\varepsilon_{n1}^2 + 2m\lambda_{\max}^2(\Sigma^{-1})\|R_n(u)\|^2 E\{\varepsilon_{n1}^2\|g'(\beta_{n0}, \theta_0, X_{n1})\|^2\}}{n\delta^2},\end{aligned}$$

it follows that $\mathbb{P}(\|Y_{n1}\| > \delta) = O(p_n/n)$. Further, it is easy to see that

$$E\|Y_{n1}\|^4 \le m\bigl[EY_{n1}(u_1)^4 + \cdots + EY_{n1}(u_m)^4\bigr].$$

Since

$$\begin{aligned}&EY_{n1}(u)^4\\ &\quad= \frac{1}{n^2}E\bigl\{\varepsilon_{n1}^4\bigl[I\bigl(\kappa_n\beta_{n0}^\top X_{n1} \le u\bigr) - R_n(u)^\top \Sigma_n^{-1} g'(\beta_{n0}, \theta_0, X_{n1})\bigr]^4\bigr\}\\ &\quad\le \frac{8}{n^2}\bigl\{E\bigl[\varepsilon_{n1}^4 I\bigl(\kappa_n\beta_{n0}^\top X_{n1} \le u\bigr)\bigr] + E\bigl[\varepsilon_{n1} R_n(u)^\top \Sigma_n^{-1} g'(\beta_{n0}, \theta_0, X_{n1})\bigr]^4\bigr\}\\ &\quad\le \frac{8}{n^2}\bigl\{E\bigl[\varepsilon_{n1}^4 I\bigl(\kappa_n\beta_{n0}^\top X_{n1} \le u\bigr)\bigr]\end{aligned}$$

$$
\begin{aligned}
&\quad + \lambda_{\max}^4(\Sigma_n^{-1}) \|R_n(u)\|^4 E[\varepsilon_{n1}^4 \|g'(\beta_{n0}, \theta_0, X_{n1})\|^4]\} \\
&\leq \frac{8}{n^2} \Biggl\{ E\varepsilon_{n1}^4 + \lambda_{\max}^4(\Sigma_n^{-1}) \|R_n(u)\|^4 \\
&\quad \times \sum_{j,k=1}^{p+d} E[\varepsilon_{n1}^4 g_j'(\beta_{n0}, \theta_0, X_{n1})^2 g_k'(\beta_{n0}, \theta_0, X_{n1})^2] \Biggr\},
\end{aligned}
$$

it follows that $EY_{n1}(u)^4 = O(p_n^2/n^2)$. Hence $\sum_{i=1}^n E\|Y_{ni}\|^2 I(\|Y_{ni}\| > \delta) = O(\sqrt{p_n^3/n}) = o(1)$.

For the covariance matrix $\sum_{i=1}^n \mathrm{Cov}(Y_{ni})$, we only need to consider $\sum_{i=1}^n \mathrm{Cov}\{Y_{ni}(s), Y_{ni}(t)\}$. It is easy to see that

$$
\begin{aligned}
&\sum_{i=1}^n \mathrm{Cov}\{Y_{ni}(s), Y_{ni}(t)\} \\
&\quad = E[\varepsilon_{n1}^2 I(\kappa_n \beta_{n0}^\top X_{n1} \leq s \wedge t)] \\
&\qquad - R_n(s)^\top \Sigma_n^{-1} E[\varepsilon_{n1}^2 g'(\beta_{n0}, \theta_0, X_{n1}) I(\kappa_n \beta_{n0}^\top X_{n1} \leq t)] \\
&\qquad - R_n(t)^\top \Sigma_n^{-1} E[\varepsilon_{n1}^2 g'(\beta_{n0}, \theta_0, X_{n1}) I(\kappa_n \beta_{n0}^\top X_{n1} \leq s)] \\
&\qquad + R_n(s)^\top \Sigma_n^{-1} E[\varepsilon_{n1}^2 g'(\beta_{n0}, \theta_0, X_{n1}) g'(\beta_{n0}, \theta_0, X_{n1})^\top] \Sigma_n^{-1} R_n(t).
\end{aligned}
$$

Thus $\sum_{i=1}^n \mathrm{Cov}\{Y_{ni}(s), Y_{ni}(t)\} = K_n(s,t)$. Since $K_n(s,t) \to K(s,t)$, it follows that $Y_{ni}$ satisfies the conditions of Lindeberg–Feller Central limit theorem. Hence the convergence of the finite-dimensional distributions holds. Altogether, we have

$$
V_n(u) \longrightarrow |V_\infty^1(u)|,
$$

where $V_\infty^1(u)$ is a zero mean Gaussian process with covariance function $K(s,t)$. Hence we complete the proof. □

## SUPPLEMENTARY MATERIAL

**Supplementary Material to "Adaptive-to-model checking for regressions with diverging number of predictors."** (DOI: 10.1214/18-AOS1735SUPP; .pdf). This Supplementary Material contains three parts with the regularity conditions, technical lemmas and proofs of the main results.

## REFERENCES

BIERENS, H. J. (1982). Consistent model specification tests. *J. Econometrics* **20** 105–134. MR0685673

BIERENS, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica* **58** 1443–1458. MR1080813

COOK, R. D. (1998). *Regression Graphics*: *Ideas for Studying Regressions Through Graphics*. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. Wiley, New York. MR1645673

COOK, R. D. and LI, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30** 455–474. MR1902895

COOK, R. D. and WEISBERG, S. (1991). Discussion of Sliced inverse regression for dimension reduction, by K. C. Li. *J. Amer. Statist. Assoc.* **86** 316–342.

DETTE, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* **27** 1012–1040. MR1724039

ESCANCIANO, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22** 1030–1051. MR2328527

FAN, J. and HUANG, L.-S. (2001). Goodness-of-fit tests for parametric regression models. *J. Amer. Statist. Assoc.* **96** 640–652. MR1946431

FAN, Y. and LI, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* **64** 865–890. MR1399221

FAN, J. and PENG, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32** 928–961. MR2065194

GONZÁLEZ MANTEIGA, W. and CRUJEIRAS, R. (2013). An updated review of Goodness-of-Fit tests for regression models. *TEST* **22** 361–411.

GUO, X., WANG, T. and ZHU, L. (2016). Model checking for parametric single-index models: A dimension reduction model-adaptive approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 1013–1035. MR3557187

HALL, P. and LI, K.-C. (1993). On almost linearity of low-dimensional projections from high-dimensional data. *Ann. Statist.* **21** 867–889. MR1232523

HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947. MR1245774

HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821. MR0356373

KHMALADZE, E. V. (1982). Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26** 240–257.

KHMALADZE, E. V. and KOUL, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32** 995–1034. MR2065196

KOUL, H. L. and NI, P. (2004). Minimum distance regression model checking. *J. Statist. Plann. Inference* **119** 109–141. MR2018453

LAVERGNE, P. and PATILEA, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *J. Econometrics* **143** 103–122. MR2384435

LAVERGNE, P. and PATILEA, V. (2012). One for all and all for one: Regression checks with many regressors. *J. Bus. Econom. Statist.* **30** 41–52. MR2899183

LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–342. MR1137117

LI, K.-C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. *J. Amer. Statist. Assoc.* **87** 1025–1039. MR1209564

LI, B. and WANG, S. (2007). On directional regression for dimension reduction. *J. Amer. Statist. Assoc.* **102** 997–1008. MR2354409

PORTNOY, S. (1984). Asymptotic behavior of $M$-estimators of $p$ regression parameters when $p^2/n$ is large. I. Consistency. *Ann. Statist.* **12** 1298–1309. MR0760690

PORTNOY, S. (1985). Asymptotic behavior of $M$ estimators of $p$ regression parameters when $p^2/n$ is large. II. Normal approximation. *Ann. Statist.* **13** 1403–1417. MR0811499

STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641. MR1439316

STUTE, W., GONZÁLEZ MANTEIGA, W. and PRESEDO QUINDIMIL, M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93** 141–149. MR1614600

STUTE, W., THIES, S. and ZHU, L.-X. (1998). Model checks for regression: An innovation process approach. *Ann. Statist.* **26** 1916–1934. MR1673284

STUTE, W., XU, W. L. and ZHU, L. X. (2008). Model diagnosis for parametric regression in high-dimensional spaces. *Biometrika* **95** 451–467. MR2521592

STUTE, W. and ZHU, L.-X. (2002). Model checks for generalized linear models. *Scand. J. Stat.* **29** 535–545. MR1925573

TAN, F. and ZHU, L. (2019). Supplement to "Adaptive-to-model checking for regressions with diverging number of predictors." DOI:10.1214/18-AOS1735SUPP.

TAN, F., ZHU, X. and ZHU, L. (2018). A projection-based adaptive-to-model test for regressions. *Statist. Sinica* **28** 157–188.

VAN KEILEGOM, I., GONZÁLEZ MANTEIGA, W. and SÁNCHEZ SELLERO, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST* **17** 401–415. MR2434335

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. Springer, New York. MR1385671

WHITE, H. (1981). Consequences and detection of misspecified nonlinear regression models. *J. Amer. Statist. Assoc.* **76** 419–433. MR0624344

XIA, Q., XU, W. and ZHU, L. (2015). Consistently determining the number of factors in multivariate volatility modelling. *Statist. Sinica* **25** 1025–1044. MR3409736

ZHENG, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75** 263–289. MR1413644

ZHU, L. X. and AN, H. Z. (1992). A method for testing nonlinearity in regression models. *J. Math. (Wuhan)* **12** 391–397. MR1254621

ZHU, L.-X. and CHENG, P. (1994). The optimal lower bound for the tail probability of the Kolmogorov distance in the worst direction. *Sankhyā Ser. A* **56** 265–293. MR1664915

ZHU, L., MIAO, B. and PENG, H. (2006). On sliced inverse regression with high-dimensional covariates. *J. Amer. Statist. Assoc.* **101** 630–643. MR2281245

ZHU, L.-P., ZHU, L.-X. and FENG, Z.-H. (2010). Dimension reduction in regressions through cumulative slicing estimation. *J. Amer. Statist. Assoc.* **105** 1455–1466. MR2796563

ZHU, L., WANG, T., ZHU, L. and FERRÉ, L. (2010). Sufficient dimension reduction through discretization-expectation estimation. *Biometrika* **97** 295–304. MR2650739

ZOU, H. and ZHANG, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Ann. Statist.* **37** 1733–1751. MR2533470

COLLEGE OF FINANCE AND STATISTICS
HUNAN UNIVERSITY
CHANGSHA 410006
CHINA
E-MAIL: falongtan@hnu.edu.cn

DEPARTMENT OF MATHEMATICS
HONG KONG BAPTIST UNIVERSITY
HONG KONG 999077
CHINA
E-MAIL: lzhu@hkbu.edu.hk