Estimation and adaptive-to-model testing for regressions with diverging number of predictors *

Falong Tan and Lixing Zhu

Department of Mathematics, Hong Kong Baptist University, Hong Kong

Abstract

The research described in this paper is motivated by model checking for parametric single-index models with diverging number of predictors. To construct a test statistic, we first study the asymptotic property of the estimators of involved parameters of interest under the null and alternative hypothesis when the dimension is divergent to infinity as the sample size goes to infinity. For the testing problem, we study an adaptive-to-model residual-marked empirical process as the basis for constructing a test statistic. By modifying the approach in the literature to suit the diverging dimension settings, we construct a martingale transformation. Under the null, local and global alternative hypothesis, the weak limits of the empirical process are derived and then the asymptotic properties of the test statistic are investigated. Simulation studies are carried out to examine the performance of the test.

Key words: Adaptive-to-model test; Empirical process; Martingale transformation; Parametric single-index models; Sufficient dimension reduction.

*Lixing Zhu is a Chair professor of Department of Mathematics at Hong Kong Baptist University, Hong Kong, China. He was supported by a grant from the University Grants Council of Hong Kong, Hong Kong, China.
1 Introduction

Regression modelling is a vital problem in regression analysis. One important step in regression modelling is to check the adequacy of a model that would be used in further analysis to prevent possible wrong conclusions. There are a number of proposals available in the literature, which will be reviewed later. However, there is an important issue that has not been well studied. We notice that in high dimensional data analysis, the dimension $p$ of the predictor vector is often large even though it is still small compared with the sample size $n$. In this case, we often regard $p$ as a diverging number as $n$ goes to infinity. A relevant reference is Huber (1973) who considered a problem where $p$ goes to infinity at the rate of order $O(n^{1/4})$.

In this paper, we focus on inference for parametric single-index models. Although they are in form generalized linear models, we do not use this name as generalized linear models have their own definitions in the literature. Let $Y$ be a response variable associated with a $p$-dimensional predictor vector $X \in \mathbb{R}^p$. If $Y$ is integrable, the regression function $g(x) = E(Y|X = x)$ is well-defined. Let $\mathcal{G} = \{g(\beta^\top \cdot, \theta) : \beta \in \mathbb{R}^p, \theta \in \mathbb{R}^d\}$ be a given parametric family of functions. The study herewith is motivated by checking whether $g(\cdot, \cdot)$ belongs to $\mathcal{G}$ or not. Thus the null hypothesis we want to test is that $(Y, X)$ follows a parametric single-index model as

$$Y = g(\beta_0^\top X, \theta_0) + \varepsilon \quad \text{for some } \beta_0 \in \mathbb{R}^p, \; \theta_0 \in \mathbb{R}^d,$$

where $\varepsilon = Y - E(Y|X)$ is the error term, $d$ is fixed, $p$ diverges as the sample size $n$ tends to infinity, and $\top$ denotes the transposition.

We now review existing methodologies in the literature. 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 et al. (2008) as examples. Globally smoothing tests construct test statistics based on averages of functionals of empirical processes and then avoid nonparametric estimation. They are called
globally smoothing tests as averaging is also a globally smoothing step. Examples include Bierens
(1982, 1990), Stute (1997), Stute, Thies, and Zhu (1998), Stute et al. (1998), Khmadladze and
Koul (2004).

All existing methods are limited to the fixed dimension settings. The extension to a diversing
dimension case is by no means trivial. When the dimension \( p \) is large, most existing tests,
especially locally smoothing tests, perform badly. 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 Stute and Zhu’s (2002) test is not omnibus, i.e., it fails to be consistent against all alternative
hypotheses and thus is a directional test. Escanciano (2006) gave some detailed comments on
this issue, and proposed, as well as Lavergne and Patilea (2008, 2012), tests that are based on
projected covariates. Guo et al. (2016) did it also 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 help of sufficient dimension reduction
techniques, Tan et al. (2017) generalized Stute and Zhu’s (2002) method and obtained an
omnibus test which is asymptotically distribution-free and inherits the dimension reduction
properties. It performs very well, but still requires the condition that \( p \) is fixed. In this paper,
we develop a consistent diagnostic test for checking the adequacy of a single-index model when
the dimension \( p \) of the predictor vector diverges to infinity as the sample size \( n \) tends to infinity.

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

\[
Y = G(B^\top X) + \varepsilon. \tag{1.2}
\]

where \( E(\varepsilon|X) = 0 \) and \( G(\cdot) \) is an unknown smooth function and \( B \) is a \( p \times q \) orthonormal matrix
with an unknown \( q \) with \( 1 \leq q \leq p \). Note that this is a more general model of (1.2) than the
nonparametric model \( Y = G(X) + \varepsilon \) as it is a special case when \( B \) is an \( p \times p \) orthonormal
matrix with $q = p$.

Similarly as Stute and Zhu (2002), we still use residual-marked empirical process and the martingale transformation to construct a test statistic when projected predictors vector is used. However, when the projected predictors vector under the null hypothesis is used to construct a test statistic as Stute and Zhu (2002) did, it cannot be an omnibus test. Stute et al (1998a) constructed a residual-marked empirical process by using the original predictors vector. When $p$ is divergent, the test severely suffers from the curse of dimensionality in theory. To alleviate these difficulties, we will adopt a model adaptation strategy as Tan et al (2017) did. It can adaptively uses projected predictors under the null and alternative hypothesis. Under the null, only one projected predictor is used like that in Stute and Zhu’s construction, while under the alternatives, it can automatically uses all projections on $q$-dimensional unit sphere to guarantee the omnibus property. Although this idea seems workable, the theoretical investigation, due to the dimensionality divergence, becomes very complicated. There are no no relevant results in the literature about the convergence of residual-marked empirical process with diverging $p$. Even when we can obtain its limiting Gaussian process, the shift term created by estimating the parameter of interest has no a simple formula so that we can easily motivate the martingale transformation construction proposed by Stute, Thies, and Zhu (1998) to make the test asymptotically distribution-free. This is a typical problem when $p$ is divergent, which does not happen when $p$ is fixed.

Therefore, the paper is then organized as follows. Section 2 contains the asymptotic properties of the ordinary least squares estimator in the diverging dimension setting. Based on this, we define an adaptive-to-model residual-marked empirical process as the basis of the proposed test statistic. Since sufficient dimension reduction theory plays a crucial role to achieve the adaptive-to-model property, we give a brief review in this section and give the study on the convergence rate of the relevant estimators. In Section 3, we present the limit of the adaptive-to-model empirical process under the null hypothesis and give the investigation for its asymptotics. Then
we use a modified approach to define a martingale transformation because the shift term has no close form in the diverging dimension settings. The asymptotic properties of the martingale transformation-based innovation process under both the null and alternatives are studied. We also show that when $p$ is fixed, this transformation is equivalent to the Stute and Zhu’s (2002) martingale transformation. In Section 4, we give the test statistic for practical use and then several simulation studies are conducted. A real data example is analysed in Section 5 for illustration. Section 6 contains a discussion. Technical proofs are deferred to Appendix.

2 Adaptive-to-model residual-marked empirical process

2.1 Preliminary

Let $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ be an i.i.d. sample with the same distribution as $(X, Y)$ and let $\varepsilon = Y - E(Y|X)$ be the unpredictable part of $Y$ given $X$. Recall that $G = \{g(\beta^T \cdot, \theta) : \beta \in \mathbb{R}^p, \theta \in \mathbb{R}^d\}$. We want to test whether or not

$$H_0 : Y = g(\beta_0^T x, \theta_0) + \varepsilon \text{ for some } \beta_0 \in \mathbb{R}^p, \theta_0 \in \mathbb{R}^d.$$ 

For estimating the unknown $(\beta_0, \theta_0)$, we in this paper restrict ourselves to the ordinary least squares method. Let

$$(\hat{\beta}_n, \hat{\theta}_n) = \arg\min_{\beta, \theta} \sum_{i=1}^n [Y_i - g(\beta^T X_i, \theta)]^2.$$ 

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

$$(\tilde{\beta}_0, \tilde{\theta}_0) = \arg\min_{\beta, \theta} E[Y - g(\beta^T X, \theta)]^2.$$ 

It is easy to see that if $g(\cdot, \cdot) \in G$, we have $(\tilde{\beta}_0, \tilde{\theta}_0) = (\beta_0, \theta_0)$. If $g \notin G$, $(\tilde{\beta}_0, \tilde{\theta}_0)$ typically depends on the distribution of $X$. Let $e = Y - g(\tilde{\beta}_0^T X, \tilde{\theta}_0)$. Then under the null hypothesis we have $e = \varepsilon$. 

5
To study the asymptotic properties of \( (\hat{\beta}_n, \hat{\theta}_n) \) as \( p \) is divergent, we first give some notations and the regularity conditions postpone to Appendix. Suppose that \( g(\beta^\top x, \theta) \) is third differentiable with respect to \( (\beta, \theta) \). Let

\[
g'(\beta, \theta, x) = \frac{\partial g(\beta^\top x, \theta)}{\partial (\beta, \theta)}, \quad g''(\beta, \theta, x) = \frac{\partial g'(\beta, \theta, x)}{\partial (\beta, \theta)}.
\]

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

\[
\Sigma_n = E[g'(\hat{\beta}_0, \hat{\theta}_0, X)g'(\hat{\beta}_0, \hat{\theta}_0, X)^\top] - E[eg''(\hat{\beta}_0, \hat{\theta}_0, X)] =: \Sigma_1n - \Sigma_2n.
\]

The next two results give the norm consistency of \( (\hat{\beta}_n, \hat{\theta}_n) \) with respective to \( (\hat{\beta}_0, \hat{\theta}_0) \) and the decomposition of \( (\hat{\beta}_n, \hat{\theta}_n - \beta_0, \theta_0) \) into independent and identically distributed summands. This decomposition generalizes the results of White (1981) to the case where the dimension \( p \) of the predictor vector diverges. For simplicity, we define hereafter \( \hat{\gamma}_n = (\hat{\beta}_n^\top, \hat{\theta}_n^\top)^\top, \hat{\gamma}_0 = (\hat{\beta}_0^\top, \hat{\theta}_0^\top)^\top \) and \( \gamma_0 = (\beta_0^\top, \theta_0^\top)^\top \).

**Proposition 1.** Suppose that conditions (A1)-(A6) in Appendix hold. If \( p^4/n \to 0 \), then \( \hat{\gamma}_n \) is a norm consistent estimator of \( \gamma_0 \) in the sense that \( \|\hat{\gamma}_n - \gamma_0\| = O_p(\sqrt{p/n}) \), where \( \| \cdot \| \) denotes the Frobenius norm.

The convergence rate of order \( \sqrt{p/n} \) is in line of the results of the M-estimator that was obtained by Huber (1973) and Portnoy (1984) when the number of parameters \( p \) diverges. For the asymptotic decomposition, we have the following result.

**Proposition 2.** If \( p^5/n \to 0 \) and conditions (A1)-(A6) in Appendix hold, we then have

\[
\hat{\gamma}_n - \gamma_0 = \Sigma_n^{-1} \frac{1}{n} \sum_{i=1}^{n} [Y_i - g(\hat{\beta}_0^\top X_i, \hat{\theta}_0)]g'(\hat{\beta}_0, \hat{\theta}_0, X_i) + o_p\left(\frac{1}{\sqrt{n}}\right).
\]

**Remark 1.** The rate \( p^4/n \to 0 \) or \( p^5/n \to 0 \) as \( n \to \infty \) seems slow. According to the arguments for proving Propositions 1 and 2 in Appendix, we can see that if \( g(\beta^\top X, \theta) = \beta^\top X \) follows a linear model, then \( g''(\beta, \theta, x) = 0 \) and \( g''(\beta, \theta, x) = 0 \). Thus we can obtain the norm consistency of \( \hat{\gamma}_n \) to \( \gamma_0 \) and the asymptotic decomposition of \( \hat{\gamma}_n - \gamma_0 \) under the conditions \( p^2/n \to 0 \) and...
\( p^3/n \to 0 \), respectively. This condition is the same as that of Huber (1973) who only considered the linear model therein. Portnoy (1984, 1985) obtained the norm consistency and the asymptotic normality under weaker conditions again for linear models. However, his conditions are hard to check in practice what kinds of models, other than linear models, can satisfy. Further, extending their results to handle the parametric single-index models as we consider here is, to the best of our knowledge, still an open question.

2.2 Basic test statistic construction

Recall the null hypothesis:

\[
H_0 : \mathbb{P}\{E(Y|X) = g(\beta_0^\top X, \theta_0)\} = 1 \quad \text{for some } \beta_0 \in \mathbb{R}^p, \theta_0 \in \mathbb{R}^d,
\]

against the alternative hypothesis:

\[
H_1 : \mathbb{P}\{E(Y|X) = G(B^\top X) = g(\beta^\top X, \theta)\} < 1 \quad \forall \beta \in \mathbb{R}^p, \theta \in \mathbb{R}^d
\]

where \( G(\cdot) \) is an unknown smooth function and the \( p \times q \) orthonormal matrix \( B \) is given in (1.2). We assume that \( \tilde{\beta}_0 \in S_{E(Y|X)} \) under both the null and alternative hypothesis where \( S_{E(Y|X)} \) is the central mean subspace such that \( S_{E(Y|X)} = \text{span}(B) \). Under the null hypothesis, this is obvious. Under the alternative hypothesis, \( \tilde{\beta}_0 \) would not necessarily parallel to \( \beta_0 \), but reasonably be a linear combination of all columns of the matrix \( B \). Thus the assumption is not restrictive.

Also recall \( \varepsilon = Y - E(Y|X) \) and \( e = Y - g(\tilde{\beta}_0^\top X, \tilde{\theta}_0) \). Under the null hypothesis, \( e = \varepsilon, q = 1 \) and \( B = \kappa \beta_0 \) with \( \kappa = \pm \frac{1}{\|\beta_0\|} \). Therefore, we obtain that \( E(e|B^\top X) = E(e|\beta_0^\top X) = 0 \). Under the alternative hypothesis, we have \( E(e|B^\top X) = G(B^\top X) - g(\tilde{\beta}_0^\top X, \tilde{\theta}_0) \neq 0 \). Then it follows that under the null hypothesis

\[
E[eI(B^\top X \leq u)] = E[eI(\kappa \beta_0^\top X \leq u)] = 0.
\]

While under the alternative, by Lemma 1 of Escanciaco (2006), there exists an \( \alpha \in S_q^+ \) such that \( E(e|\alpha^\top B^\top X) \neq 0 \), where \( S_q^+ = \{\alpha = (a_1, \cdots, a_q)^\top \in \mathbb{R}^q : \|\alpha\| = 1 \text{ and } a_1 \geq 0\} \). Then it
follows that
\[ E[eI(\alpha^\top B^\top X \leq u)] \neq 0 \] (2.3)

Note that under the null we have \( q = 1 \) and \( S_q^+ = \{1\} \). Thus the quantity \( E[eI(\alpha^\top B^\top X \leq u)] \) 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

\[ V_n(\hat{\alpha}, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(\hat{\beta}_n^\top X_i, \hat{\theta}_n)] I(\hat{\alpha}^\top \hat{B}_n^\top X_i \leq u), \] (2.4)

\[ V_n(u) = \sup_{\hat{\alpha} \in S_q^+} |V_n(\hat{\alpha}, u)| \] (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 \( B \) with an estimated structural dimension \( \hat{q} \) of \( q \), which will be specified later. For \( V_n(u) \), one can also use the integral over \( S_q^+ \) 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 \( q \) and the matrix \( B \), when \( p \) diverges to infinity. We give a brief review below on this topic.

### 2.3 Adaptive-to-model approach

In this methodology, we need to identify the dimension \( q \) and the matrix \( B \). This can be done by using the methods in sufficient dimension reduction. We then give a brief description. Recall under the alternative hypothesis the model is as

\[ Y = G(B^\top X) + \varepsilon, \] (2.6)

where \( E(\varepsilon|X) = 0 \) and \( G(\cdot) \) is an unknown smooth function and \( B \) is a \( p \times q \) orthonormal matrix with \( 1 \leq q \leq p \). We can see that under both the null and alternative hypothesis, the conditional independence holds respectively:

\[ Y \perp \!\!\!\!\perp E(Y|X)|\beta_0^\top X, \quad \text{and} \quad Y \perp \!\!\!\!\perp E(Y|X)|B^\top X, \]
where \( \bot \) means statistical independence. Define \( S_{E(Y|X)} \) as the central mean subspace of \( Y \) with respect to \( X \) (see, Cook and Li 2002) that is the intersection of all subspaces spanned by the columns of \( A \) \( \text{span}(A) \) such that \( Y \bot E(Y|X)A^\top X \). The dimension of \( S_{E(Y|X)} \) is called the structural dimension, denoted as \( d_{E(Y|X)} \). Under mild conditions, such a subspace \( S_{E(Y|X)} \) always exists (see Cook and Li, 2002). If \( S_{E(Y|X)} = \text{span}(A) \), then \( E(Y|X) = E(Y|A^\top X) \). Under the null hypothesis (1.1), \( d_{E(Y|X)} = 1 \) and \( S_{E(Y|X)} = \text{span}(\beta_0/\|\beta_0\|) \). Under the alternative (1.2), \( d_{E(Y|X)} = q \) and \( S_{E(Y|X)} = \text{span}(B) \). For simplicity, we assume throughout this paper that \( S_{E(Y|X)} = S_{Y|X} \). Here \( S_{Y|X} \) is the central subspace of \( Y \) with respect to \( X \) (see, Cook 1998).

There are several estimation proposals available in the literature. For instance, sliced inverse regression (SIR, Li (1991)), sliced average variance estimation (SAVE, Cook and Weisberg (1991)), minimum average variance estimation (MAVE, Xia et al. (2002)), directional regression (DR, Li and Wang, (2007)), discretization-expectation estimation (DEE, Zhu, et al. (2010a)). All these methods assumed that \( p \) is fixed. Zhu, Miao, and Peng (2006) first discussed the asymptotic properties of SIR when \( p \) diverges to infinity. In this paper, we adapt cumulative slicing estimation (CSE, Zhu, Zhu, and Feng (2010b)) to identify the central subspace, which is similar to discretization-expectation estimation (DEE, Zhu, et al. (2010a)). This is because both of them are very easily implemented and easy to be extended to handle the case where the dimension \( p \) grows to infinity.

The procedure of CSE is as follows. For simplicity, we assume \( E(X) = 0, Var(X) = I_p \) for a moment. If the linearity condition (see Li, 1991) holds, it is easy to see that \( E[Xh(Y)] \in S_{Y|X} \) for any function \( h(\cdot) \). Theoretically, we obtain infinity amount of vectors in \( S_{Y|X} \). Zhu et.al. (2010b) suggested a determining class of indicator functions to replace \( h(\cdot) \). Let \( h_t(Y) = I(Y \leq t) \). It follows that

\[
Y \bot X|B^\top X \iff h_t(Y) \bot X|B^\top X, \quad \forall \ t \in \mathbb{R}.
\]
Define the target matrix

\[ M = \int E[X h_t(Y)] E[X^\top h_t(Y)] dF_Y(t), \quad (2.7) \]

where \( F_Y \) denotes the cumulative distribution function of \( Y \). If the rank of \( M \) is \( q \), then \( \text{span}(M) = \mathcal{S}_{Y|X} \). Based on this, it is easy to obtain the sample version of \( M \). Let \( Z_i \) be the standardized \( X_i \) and \( \hat{\alpha}_t = \frac{1}{n} \sum_{i=1}^{n} Z_i I(Y_i \leq t) \). The estimator of \( M \) is given by

\[ \hat{M}_n = \frac{1}{n} \sum_{j=1}^{n} \hat{\alpha}_Y j \hat{\alpha}_Y^\top. \quad (2.8) \]

If the structural dimension \( q \) is given, an estimator \( \hat{B}_n(q) \) of \( B \) consists of the eigenvectors corresponding to the largest \( q \) eigenvalues of \( \hat{M}_n \). Throughout this paper, we assume that \( q \) is fixed.

Yet we need a consistent estimator \( \hat{q} \) of \( q \) as \( q \) is usually unknown under the alternative hypothesis. Later we will see that even when \( q \) is given, we still want a consistent estimator because we wish the test to have 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 \( q \). Let \( \hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \) and \( \lambda_1 \geq \cdots \geq \lambda_p \) be the eigenvalues of the matrix \( \hat{M}_n \) and \( M \) respectively. Since \( \text{rank}(M) = q \), it follows that

\[ \lambda_1 \geq \cdots \geq \lambda_q > \lambda_{q+1} = \cdots = \lambda_p = 0. \]

Hence we estimate the structural dimension \( q \) by

\[ \hat{q} = \arg \min_{1 \leq i \leq p} \left\{ i : \frac{\hat{\lambda}_{i+1}^2 + c}{\hat{\lambda}_i^2 + c} \right\}. \quad (2.9) \]

Here \( \hat{\lambda}_{p+1} \) is defined as 0 and the ridge \( c \) is a positive constant. The following result shows that the consistency of MRER is adaptive to the underlying models, when \( c \) equals to some appropriate constant. Its proof will be given in Appendix.

**Proposition 3.** Suppose that the regularity conditions of Theorem 3 in Zhu et al. (2010b) hold. Let \( \hat{B}_n(q) \) be a matrix whose columns are the eigenvectors that are associated with the largest \( q \)
eigenvalues of $\hat{M}_n$. If $c = \log n/n$, then

1. under $H_0$, we have $P(\hat{q} = 1) \to 1$ and $\|\hat{B}_n(1) - \kappa\beta_0\| = O_p(\sqrt{p/n})$;
2. under $H_1$, we have $P(\hat{q} = q) \to 1$ and $\|\hat{B}_n(q) - B\| = O_p(\sqrt{p/n})$.

3 Main results

3.1 Basic properties of the process

First, we discuss the asymptotic properties of the process $V_n(\hat{\alpha}, u)$ under the null hypothesis. Since the distributional limit theory becomes much simpler if we replace the estimators by their true values, we define the following process

$$V_n^0(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(\beta_0^\top X_i, \theta_0)] I(\kappa\beta_0^\top X_i \leq u).$$

Put

$$\sigma_n^2(v) = \text{Var}(Y|\kappa\beta_0^\top X = v) \quad \psi_n(u) = E[\text{Var}(Y|\kappa\beta_0^\top X) I(\kappa\beta_0^\top X \leq u)].$$

Then we have $\sigma_n^2(v) = E(\varepsilon^2|\kappa\beta_0^\top X = v)$ and $\psi_n(u) = \int_{-\infty}^{u} \sigma_n^2(v) F_{\kappa\beta_0}(dv)$ where $F_{\kappa\beta_0}$ is the cumulate distribution function of $\kappa\beta_0^\top X$. Obviously, $\psi_n(u)$ is a nondecreasing and nonnegative function. Since $V_n^0(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i I(\kappa\beta_0^\top X_i \leq u)$ is a centered residual cusum process, it is readily seen that

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

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

$$V_n^0(u) \longrightarrow V_{\infty}(u) \quad \text{in distribution}, \quad (3.1)$$
in the space $\ell^\infty(\mathbb{R})$, where $V_\infty(u)$ is a centred Gaussian process with the covariance function $\psi(s \wedge t)$. 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.

For composite model checks, the unknown parameters in $V_0^n(u)$ should be replaced by their estimators, so we go back to $V_n(\hat{\alpha}, u)$ as defined in (2.4). By Proposition 3, $\mathbb{P}(\hat{q} = 1) \to 1$ under the null hypothesis. Thus we only need to work on the event $\{\hat{q} = 1\}$. Consequently, $S_q^+ = \{1\}$ and $V_n(\hat{\alpha}, u)$ can be rewritten as

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

Under some regularity conditions stated in Appendix and on the event $\{\hat{q} = 1\}$, we can show that under the null hypothesis

$$V_n(\hat{\alpha}, u) = V_0^n(u) - \sqrt{n}(\hat{\gamma}_n - \gamma_0)^\top M_n(u) + o_p(1) \tag{3.2}$$

uniformly in $u$, where $M_n(u) = E[g'(\beta_0, \theta_0, X)I(\kappa\beta_0^\top X \leq u)]$. A proof of (3.2) will be given in Appendix. Combined (3.2) with Proposition 2 and some elementary calculations, we have

$$V_n(\hat{\alpha}, u) = V_0^n(u) - \frac{1}{\sqrt{n}} M_n(u)^\top \Sigma_n^{-1} \sum_{i=1}^{n} g'(\beta_0, \theta_0, X_i) \varepsilon_i + o_p(1) \tag{3.3}$$

uniformly in $u$. It is easy to see that the second term of the right hand side of (3.3) is also asymptotically tight. Altogether we then obtain the following result.

**Theorem 3.1.** Suppose that the regularity conditions in Appendix hold. when $p^5/n \to 0$, then under the null hypothesis, we have in distribution

$$V_n(u) \to |V_1^\infty(u)|,$$

where $V_1^\infty(u)$ is a zero mean Gaussian process with a covariance function $K(s, t)$ that is the pointwise limit of $K_n(s, t)$ as

$$K_n(s, t) = E[\varepsilon^2 I(\kappa\beta_0^\top X \leq s \wedge t)] - M_n(s)^\top \Sigma_n^{-1} E[\varepsilon^2 g'(\beta_0, \theta_0, X)I(\kappa\beta_0^\top X \leq t)]$$
$$- M_n(t)^\top \Sigma_n^{-1} E[\varepsilon^2 g'(\beta_0, \theta_0, X)I(\kappa\beta_0^\top X \leq s)]$$
$$+ M_n(s)^\top \Sigma_n^{-1} E[\varepsilon^2 g'(\beta_0, \theta_0, X)g'(\beta_0, \theta_0, X)^\top] \Sigma_n^{-1} M_n(t).$$
3.2 Martingale transformation

If \( p \) is fixed, \( V_{\infty}^1(u) \) can be rewritten as \( V_{\infty}^1(u) = V_{\infty}(u) + M(u)^{\top}V \) in distribution and its covariance function can be specified. The shift term \( M(u)^{\top}V \) is brought out from the second term in \( 3.3 \). Stute, Thies, and Zhu (1998) first proposed a martingale transformation to eliminate \( M(u)^{\top}V \) in \( V_{\infty}^1(u) \) and then 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 convenient 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 Khmadladze and Koul (2004) and Stute, Xu and Zhu (2008). However, when \( p \) diverges as \( n \) goes to infinity, the form of the shift term that would be a limit of \( M(u)^{\top}V \) can not be given specifically, as stated in the above theorem. The martingale transformation cannot directly target \( M(u)^{\top}V \). 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 \( 3.2 \). This is because in the case with the fixed \( p \), \( M(u)^{\top}V \) is just its weak limit. Thus, we then target that term directly at the sample level.

Following Stute, Thies, and Zhu (1998) or Stute and Zhu (2002), recall that \( M_n(u) = E[g'(\beta_0, \theta_0, X)I(\kappa\beta_0^{\top}X \leq u)] \) and \( \psi_n(u) = \int_{-\infty}^{u} \sigma_n^2(v) F_{\kappa\beta_0}(dv) \). Let \( a_n(u) = \frac{\partial M_n(u)}{\partial \psi_n(u)} \) be the Radon-Nikodym derivative of \( M_n(u) \) with respect to \( \psi_n(u) \). Next, define a \((p+d) \times (p+d)\) matrix

\[
A_n(u) = \int_{u}^{\infty} a_n(z) M_n^{\top}(dz) = \int_{u}^{\infty} a_n(z) a_n(z)^{\top} \sigma_n^2(z) F_{\kappa\beta_0}(dz).
\]

It can also be written as

\[
A_n(u) = E[a_n(\kappa\beta_0^{\top}X)g'(\beta_0, \theta_0, X)^{\top}I(\kappa\beta_0^{\top}X \geq u)].
\]
Mimicking the martingale transformation in Stute and Zhu (2002) at the sample level, we have

\[(T_n f_n)(u) = f_n(u) - \int_{-\infty}^{u} a_n(z)^\top A_n^{-1}(z) \left( \int_z^{\infty} a_n(v) f_n(dv) \right) \psi_n(dz). \quad (3.4)\]

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

Some elementary computation concludes that \((T_n \sqrt{n}(\hat{\gamma}_n - \gamma_0)^\top M_n) = 0\). Next, we discuss the approximation properties of \(T_n V_n^0\). Note that

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

and

\[\int_z^{\infty} a_n(v) V_n^0(dv) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_n(\kappa \beta_0^\top X_i) I(\kappa \beta_0^\top X_i \geq z) \varepsilon_i.\]

Combining these two formulas, we obtain that

\[T_n V_n^0(u) = V_n^0(u) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{-\infty}^{u} a_n(z)^\top A_n^{-1}(z) a_n(\kappa \beta_0^\top X_i) I(\kappa \beta_0^\top X_i \geq z) \psi_n(dz) \varepsilon_i.\]

Therefore, \(T_n V_n^0\) is also an i.i.d. centered residual cusum process with a covariance function

\[Cov[T_n V_n^0(s), T_n V_n^0(t)] = Cov[V_n^0(s), V_n^0(t)] = \psi_n(s \wedge t). \quad (3.5)\]

This means that \(T_n V_n^0(u)\) admits the same limiting distribution as that of \(V_n^0(u)\), i.e.,

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

Consequently, we get rid of the annoying shift term \(\sqrt{n}(\hat{\gamma}_n - \gamma_0)^\top M_n\) and obtain the process \(V_\infty(u)\) whose supremum over all \(u\) has a tractable limiting distribution. The assertions (3.5) and (3.6) will be justified in Appendix (Lemma 1).

The transformation \(T_n\) obviously contains some unknown quantities and therefore needs to be substituted by their empirical analogues. For this, let

\[g_1'(t, \theta) = \frac{\partial g(t, \theta)}{\partial t} \quad \text{and} \quad g_2'(t, \theta) = \frac{\partial g(t, \theta)}{\partial \theta}.\]

It follows that

\[g'(\beta_0, \theta_0, X) = \left( g_1'(\beta_0^\top X, \theta_0) X^\top, g_2'(\beta_0^\top X, \theta_0)^\top \right)^\top.\]
Consequently, we have
\[ M_n(u) = \left( \int_{-\infty}^{u} g'_1(z/\kappa, \theta_0) r_n(z)^T F_{\kappa \beta_0}(dz), \int_{-\infty}^{u} g'_2(z/\kappa, \theta_0)^T F_{\kappa \beta_0}(dz) \right)^T \]
where \( r_n(v) = E(X|\kappa \beta_0^T X = v) \). Conclude that
\[ a_n(u) = \left( \frac{g'_1(u/\kappa, \theta_0)r_n(u)^T}{\sigma_n^2(u)}, \frac{g'_2(u/\kappa, \theta_0)^T}{\sigma_n^2(u)} \right)^T. \]
Since \( a_n(u) \) depends on \( r_n(u) \) and \( \sigma_n^2(u) \) on which we do not make any assumption rather than smoothness, they need to be estimated in a nonparametric way. For instance, we may adopt a standard Nadaraya-Watson estimator for \( r_n(v) \):
\[ \hat{r}_n(v) = \frac{\sum_{i=1}^{n} X_i K\left(\frac{v - \hat{\beta}^\top B_n^\top X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{v - \hat{\beta}^\top B_n^\top X_i}{h}\right)} \]
where \( K(\cdot) \) is an univariate kernel function and \( h \) is a bandwidth. Similarly for \( \sigma_n^2(u) \). Thus we obtain the empirical estimators \( \hat{a}_n(u) \) and \( \hat{A}_n(u) \) of \( a_n(u) \) and \( A_n(u) \) respectively:
\[ \hat{a}_n(u) = \left( \frac{g'_1(u/\hat{\kappa}_n, \hat{\theta}_n)r_n(u)^T}{\hat{\sigma}_n^2(u)}, \frac{g'_2(u/\hat{\kappa}_n, \hat{\theta}_n)^T}{\hat{\sigma}_n^2(u)} \right)^T, \]
\[ \hat{A}_n(u) = \frac{1}{n} \sum_{i=1}^{n} \hat{a}_n(\hat{\alpha}^\top B_n^\top X_i)g'(\hat{\beta}_n, \hat{\theta}_n, X_i)^\top I(\hat{\alpha}^\top B_n^\top X_i \geq u). \]
Finally, we can give an estimator \( \hat{T}_n \) of \( T_n \):
\[ \hat{T}_n V_n(\hat{\alpha}, u) = V_n(\hat{\alpha}, u) - \int_{-\infty}^{u} \hat{a}_n(z)^\top \hat{A}^{-1}_n(z) \left( \int_{z}^{\infty} \hat{a}_n(v)V_n(\hat{\alpha}, dv) \right) \hat{\sigma}_n^2(z)F_{\hat{\alpha}}(dz) \]
\[ = \frac{1}{n^{1/2}} \sum_{i=1}^{n} [Y_i - g(\hat{\beta}_n^\top X_i, \hat{\theta}_n)] I(\hat{\alpha}^\top B_n^\top X_i \leq u) - \]
\[ \frac{1}{n^{3/2}} \sum_{i,j=1}^{n} I(\hat{\alpha}^\top B_n^\top X_j \geq \hat{\alpha}^\top B_n^\top X_i)[Y_j - g(\hat{\beta}_n^\top X_j, \hat{\theta}_n)]\hat{\sigma}_n^2(\hat{\alpha}^\top B_n^\top X_i) \]
where \( \hat{\kappa}_n \) is the estimator of \( \kappa \) and \( F_{\hat{\alpha}} \) is the empirical distribution function of \( \hat{\alpha}^\top B_n^\top X_i, 1 \leq i \leq n \). Making sure the columns of \( B_n \) have the same direction as \( \hat{\beta}_n \), we can assume \( \kappa = 1/\|\beta_0\| \) 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 $u$. If $p^5/n \to 0$, under the null hypothesis $H_0$ and the regularity conditions in Appendix, we have
\[
\sup_{\hat{\alpha} \in \mathcal{S}^+} |\hat{T}_n V_n(\hat{\alpha}, u)| \longrightarrow |V_\infty(u)|
\]
in distribution in the space $\ell^\infty([-\infty,x_0])$ for any $x_0 \in \mathbb{R}$.

Note that we use $\hat{A}_n(u)$ in the process $\hat{T}_n V_n(\hat{\alpha}, u)$. In concrete data analysis, these matrices may be unbounded for large $u$ and thus the distributional behavior of the underlying process may become very unstable in the extreme right tails. These may severely damage the approximation accuracy of the test statistic based on all $\hat{T}_n V_n$. Therefore, we restrict $\hat{T}_n V_n$ to compact intervals $[-\infty,u_0]$ and obtain the convergence of $\sup_{\hat{\alpha} \in \mathcal{S}^+} |\hat{T}_n V_n(\hat{\alpha}, u)|$ in the space $\ell^\infty([-\infty,x_0])$.

In a special case where the predictor $X$ follows a spherically contoured distribution or its extension, the elliptically contoured distribution, we can show that the calculations of the martingale transformation will become much simpler. The idea is similar to Stute and Zhu (2002). Without loss of generality, we only consider spherically contoured distributions. Here we shall assume the regression function $g$ does not depend on $\theta$. Let $g'(t)$ be the derivative of $g(t)$ with respective to $t$. It follows that
\[
M_n(u) = E[g'(\beta_0^\top X)XI(\kappa\beta_0^\top X \leq u)] = \Gamma^\top E[g'(\beta_0^\top X)\Gamma XI(\kappa\beta_0^\top X \leq u)],
\]
where $\Gamma$ is an $p \times p$ orthonormal matrix with the first row $\kappa\beta_0^\top$ (or $\beta_0^\top / \|\beta_0\|$). Since the conditional expectation of the other components of $\Gamma X$ given the first is zero, it follows that
\[
M_n(u) = \frac{\beta_0}{\|\beta_0\|^2} E[g'(\beta_0^\top X)\beta_0^\top XI(\kappa\beta_0^\top X \leq u)] = \frac{\beta_0}{\|\beta_0\|^2} \int_{-\infty}^{u} g'(z/\kappa)z/\kappa F_{\kappa\beta_0}(dz),
\]
whence,
\[
a_n(u) = \frac{\beta_0}{\|\beta_0\|^2} \frac{g'(u/\kappa)u/\kappa}{\sigma_n^2(u)},
\]
\[
A_n(u) = \frac{\beta_0\beta_0^\top}{\|\beta_0\|^4} \int_{u}^{\infty} \frac{[g'(z/\kappa)z/\kappa]^2}{\sigma_n^2(z)} F_{\kappa\beta_0}(dz).
\]
Note that $A_n(z)$ is a matrix with rank 1 and is singular when $p > 1$. Thus the martingale transformation cannot apply directly. However, if we go back to (3.2) and set

$$\tilde{M}_n(u) = E[g'(\beta_0^\top X)\beta_0^\top X I(\kappa\beta_0^\top X \leq u)],$$

then (3.2) can be rewritten as

$$V_n(\hat{\alpha}, u) = V_n^0(u) - \sqrt{n} (\hat{\gamma}_n - \gamma_0)^\top \beta_0 / \|\beta_0\|^2 \tilde{M}_n(u) + o_p(1)$$

(3.7)

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

$$a_n(u) = \frac{\partial \tilde{M}_n(u)}{\partial \psi_n(u)} = \frac{g'(u/\kappa)u/\kappa}{\sigma_n^2(u)} \quad \text{and} \quad A_n(u) = \int_u^\infty \frac{[g'(z/\kappa)z/\kappa]^2}{\sigma_n^2(z)} F_{\kappa,\beta_0}(dz).$$

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

Hall and Li (1993) shown that, if $p \to \infty$ as $n \to \infty$, expectation over a large number of random variables behaves more or less like expectation over the multivariate normal distribution. Note that $M_n(u) = E[g'(\beta_0^\top X)X I(\kappa\beta_0^\top X \leq u)]$ and multivariate normal distribution is elliptically-contoured. Consequently, even when $X$ is not multivariate normal distributed, $M_n(u)$ can be viewed as expectation on multivariate normal distribution and then the martingale transformation $T_n$ can apply to the real-valued $a_n(u)$ and $A_n(u)$ in practice for large $p$.

### 3.3 The properties under the alternative hypothesis

Now we discuss the asymptotic properties of $\sup_{\hat{\alpha} \in S_+^1} |\hat{T}_n V_n(\hat{\alpha}, u)|$ under a sequence of local alternatives converging to the null hypothesis at a parametric rate $1/\sqrt{n}$. Consider

$$H_{1n} : Y = g(\beta_0^\top X, \theta_0) + \frac{1}{\sqrt{n}} G(X) + \varepsilon,$$

(3.8)

where $E(\varepsilon|X) = 0$, $G(X)$ is a random variable with zero mean and satisfies $\mathbb{P}\{G(X) = 0\} < 1$. To derive the asymptotic distribution of $\hat{T}_n V_n(\hat{\alpha}, u)$ under $H_{1n}$, we need the asymptotic properties of $\hat{q}$ and $\hat{\gamma}_n$, when $p$ diverges to infinity.
Proposition 4. Assume the regularity conditions of Theorem 3 in Zhu et al. (2010b) hold. Let \( \hat{B}_n(1) \) be an eigenvector associating with the largest eigenvalues of \( \hat{M}_n \), then we have, under \( H_{1n} \), \( \mathbb{P}(\hat{q} = 1) \to 1 \) and \( \|\hat{B}_n(1) - \kappa\beta_0\| = O_p(\sqrt{p/n}) \).

Next, we derive the norm consistency of \( \hat{\gamma}_n \) with respective to \( \gamma_0 \) and a asymptotical decomposition of \( \hat{\gamma}_n - \gamma_0 \) under \( H_{1n} \). Here \( \hat{\gamma}_n = (\hat{\beta}_n^\top, \hat{\theta}_n^\top) \) and \( \gamma_0 = (\beta_0^\top, \theta_0^\top) \) as mentioned before.

Proposition 5. Suppose the regularity conditions in Appendix and (3.8) hold. If \( p^4/n \to 0 \), then \( \hat{\gamma}_n \) is a norm consistent estimator for \( \gamma_0 \) with \( \|\hat{\gamma}_n - \gamma_0\| = O_p(\sqrt{p/n}) \). Moreover, if \( p^5/n \to 0 \), we have

\[
\sqrt{n}(\hat{\gamma}_n - \gamma_0) = \Sigma_n^{-1/2} \sum_{i=1}^n \varepsilon_i g'(\beta_0, \theta_0, X_i) + \Sigma_n^{-1} E[G(X)g'(\beta_0, \theta_0, X)] + o_p(1).
\] (3.9)

The following theorem states the asymptotic results under various alternatives.

Theorem 3.3. Suppose the regularity conditions in Appendix hold. If \( p^5/n \to 0 \),

1. under the global alternative \( H_1 \), we have in probability

\[
\frac{1}{\sqrt{n}} \sup_{\hat{\alpha} \in S_q^+} |\hat{T}_n V_n(\hat{\alpha}, u)| \to |L(u)|,
\]

where \( L(u) \) is some nonzero function;

2. under the local alternative \( H_{1n} \), we have in distribution

\[
\sup_{\hat{\alpha} \in S_q^+} |\hat{T}_n V_n(\hat{\alpha}, u)| \to |V_\infty(u) + G_1(u) - G_2(u)|,
\]

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 limit of \( G_{1n}(u) \) and \( G_{2n}(u) \), respectively which are as follows

\[
G_{1n}(u) = E[G(X)I(\kappa\beta_0^\top X \leq u)],
G_{2n}(u) = E\{G(X) \int_{-\infty}^u a_n(z) A_n^{-1}(z) a_n(\kappa\beta_0^\top X) I(\kappa\beta_0^\top X \geq z) \psi_n(dz)\}.
\]

These results show that under the global alternative, the process diverges to infinity at the rate of order \( 1/\sqrt{n} \) and under the local alternatives distinct from the null at the rate of order \( 1/\sqrt{n} \), the process converges to a stochastic process. Thus, the test that is based on this process can detect such alternatives.
4 Numerical studies

4.1 Test statistics in practical use

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

\[ CM_n^2 = \int_{-\infty}^{u_0} \sup_{\hat{\alpha} \in S^+_\beta} |\hat{T}_n V_n(\hat{\alpha}, u)|^2 F_n(du), \]  

(4.1)

where \( F_n \) is the empirical distribution function of \( \beta_0^T X_i/\|\beta_0\|, 1 \leq i \leq n \). According to Theorem 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,

\[ CM_n^2 \rightarrow \int_{-\infty}^{u_0} \frac{B^2(\psi(u))}{\sigma^2(u)} \psi(du) \text{ in distribution}, \]

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

\[ \int_{-\infty}^{u_0} B^2(\psi(u)) \psi(du) = \psi^2(u_0) \int_{0}^{1} B(t)^2 dt \text{ in distribution}. \]

Consequently, we consider

\[ ACM_n^2 = \frac{1}{\psi_n(u_0)^2} \int_{-\infty}^{u_0} \sup_{\hat{\alpha} \in S^+_\beta} |\hat{T}_n V_n(\hat{\alpha}, u)|^2 \hat{\sigma}_n^2(u) F_n(du). \]  

(4.2)

Here we use \( \hat{\psi}_n(u_0) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - g(\hat{\beta}_i^T X_i, \hat{\theta}_n))^2 I(\hat{\alpha}^T \hat{B}_n^T X_i \leq u_0) \) as an estimator of \( \psi(u) \). Therefore, we obtain

\[ ACM_n^2 \rightarrow \int_{0}^{1} B^2(u) du \text{ in distribution}. \]

In the homoscedastic models, \( \sigma_n^2(u) \) is free of \( u \) and thus we can estimate it by

\[ \hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} [Y_i - g(\hat{\beta}_n^T X_i, \hat{\theta}_n)]^2. \]
Now we also have $\psi_n(u_0) = \sigma_n^2 F_{\kappa_0}(u_0)$ and thus it can be estimated by $\hat{\sigma}_n^2 F_n(u_0)$. Consequently, $ACM_n^2$ becomes

$$ACM_n^2 = \frac{1}{\hat{\sigma}_n^2 F_n(u_0)^2} \int_{-\infty}^{u_0} \sup_{\hat{\alpha} \in S_i^+} |\hat{T}_n V_n(\hat{\alpha}, u)|^2 F_n(du).$$

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

### 4.2 Numerical studies

In this subsection we conduct some simulation studies to examine the performance of the proposed test in this paper. From the results, we set $p = \lfloor 4n^{1/4} \rfloor - 5$ with $n = 100, 200, 400$ and 800, as used in Fan and Peng (2004). As there are no relevant tests dealing with the case with divergent dimension, we give comparisons with some existing tests that were developed with fixed dimension as for practical use, they would be workable.

1. Stute and Zhu’s (2002) test is given by

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

where

$$R_n^1(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(\hat{\beta}_n^T X_i, \hat{\theta}_n)] I(\hat{\beta}_n^T X_i \leq u);$$

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

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

2. Bierens (1982) proposed an integrated conditional moment (ICM) test which is based on the following statistic:

$$ICM_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{e}_i \hat{e}_j \exp\left(-\frac{1}{2} |X_i - X_j| \right)$$

where $\hat{e}_i = Y_i - g(\hat{\beta}_n^T X_i, \hat{\theta}_n)$. 

20
3. Escanciano’s (2006) test statistic is defined as

\[ PCvM_n = \frac{1}{n^2} \sum_{i,j,r=1}^{n} \hat{e}_i \hat{e}_j \int_{S_p} I(\beta^\top X_i \leq \beta^\top X_r) I(\beta^\top X_j \leq \beta^\top X_r) d\beta \]

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

4. Zheng (1996) proposed a locally smoothing test whose statistic is given by

\[ T_{ZH}^n = \sum_{i \neq j} K((X_i - X_j)/h) \hat{e}_i \hat{e}_j \left\{ \sum_{i \neq j} 2K^2((X_i - X_j)/h) \hat{e}_i^2 \hat{e}_j^2 \right\}^{1/2} \]

5. An adaptive-to-model test defined in Guo et. al. (2016) with the test statistic:

\[ T_{GWZ}^n = h^{1/2} \sum_{i \neq j} \hat{e}_i \hat{e}_j K(\hat{B}_n^\top (X_i - X_j)/h) \]

\[ \left\{ 2 \sum_{i \neq j} \hat{e}_i^2 \hat{e}_j^2 K^2(\hat{B}_n^\top (X_i - X_j)/h) \right\}^{1/2} \]

Here we use the kernel function \( K(u) = (15/16)(1 - u^2)^2 \mathbb{I}(|u| \leq 1) \) and the bandwidth \( h = 1.5n^{1/(4+\hat{q})} \) as in Guo et. al. (2016) and \( \hat{B}_n \) is a sufficient dimension estimate of \( B \) with an estimated structural dimension \( \hat{q} \) of \( q \).

The significance levels are set to be \( \alpha = 0.1, 0.05, \) and \( 0.01 \). 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:

\[ H_{11} : Y = \beta_0^\top X + a \exp(- (\beta_0^\top X)^2) + \varepsilon; \]
\[ H_{12} : Y = \beta_0^\top X + a \cos(0.6\pi \beta_0^\top X) + \varepsilon; \]
\[ H_{13} : Y = \beta_1^\top X + a (\beta_2^\top X)^2 + \varepsilon; \]
\[ H_{14} : Y = \beta_1^\top X + a \exp(\beta_2^\top X) + \varepsilon; \]

where \( \beta_0 = (1, \cdots, 1)^\top / \sqrt{p}, \beta_1 = (1, \cdots, 1, 0, \ldots, 0)^\top / \sqrt{p_{11}} \) and \( \beta_2 = (0, \ldots, 0, 1, \ldots, 1)^\top / \sqrt{p_{11}} \) with \( p_1 = [p/2] \). The predictors \( \{X_i, 1 \leq i \leq n\} \) are i.i.d. from \( N(0, I_p) \) and \( \varepsilon \) is 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.

The simulation results are reported in Tables 1 to 4. We can see that both $ACM_n$ and $T_n^{SZ}$ maintain the significance levels very well. The empirical sizes of $PCvM_n$ are also very close to the significance levels, but slightly more unstable in some cases. $T_n^{GWZ}$ can only maintain the significance level when it is $\alpha = 0.05$. $T_n^{ZH}$ can maintain the significance levels occasionally, but generally, it is conservative with smaller sizes. $ICM_n$ is the worst among these tests in both the significance level maintenance and power performance. According to our experience, when $p$ is smaller than 5, $ICM_n$ could work well. The powers of $ACM_n^2$, $T_n^{SZ}$, $PCvM_n$ and $T_n^{GWZ}$ are all very high for models $H_{11}$, $H_{13}$ and $H_{14}$. But $T_n^{GWZ}$’s power grows slightly slower than the other three, while, for model $H_{12}$, $T_n^{GWZ}$ beats the other competitors. These may validate again the empirical experience in this area that locally smoothing tests perform better for high frequency/oscillating models, while globally smoothing tests work better for low frequency models. Nevertheless, $T_n^{ZH}$, a representative of locally smoothing tests, has very low power for model $H_{12}$. This is because $T_n^{ZH}$ severely suffers from the dimensionality problem, while $T_n^{GWZ}$ uses a dimension reduction technique to greatly alleviate the curse of dimensionality.

Tables 1 – 4 about here

The null 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

$$H_{21} : Y = (\beta_1^\top X)^3 + a(\beta_2^\top X)^2 + \varepsilon;$$

$$H_{22} : Y = \exp(\beta_1^\top X) + a(\beta_2^\top X) + \varepsilon,$$

where $\beta_1 = (1, \ldots, 1, 0, \ldots, 0)^\top / \sqrt{p_1}$ and $\beta_2 = (0, \ldots, 0, 1, \ldots, 1)^\top / \sqrt{p_1}$ with $p_1 = [p/2]$, $\varepsilon$ is
\( N(0,1) \), and \( X \) is \( N(0, I_p) \) independent of \( \varepsilon \).

We report the empirical sizes and powers in Tables 5 and 6. For model \( H_{21} \), the conclusions are very similar to those in Study 1. For model \( H_{22} \), we can see that the empirical sizes of \( ACM_n^2, T_n^{SZ} \) and \( PCvM_n \) are very close to the significance levels, while \( T_n^{ZH} \) and \( T_n^{GWZ} \) can only control the level of \( \alpha = 0.1 \). \( ICM_n \) is still the worst one. The empirical powers of \( T_n^{GWZ} \) and \( ACM_n^2 \) are higher than the other competitors, while \( T_n^{SZ} \)'s empirical powers grow very slow in this case. This would confirm the theoretical result that \( T_n^{SZ} \) is not an omnibus test.

Tables 5 – 6 about here

Therefore, overall, the proposed test in this paper performs well and can detect different alternatives. Further, the dimension of predictors has less negative impact on its performance.

4.3 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](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, and \( X_{12} \): errors; and \( X_{13} \): 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 \( X = (X_1, \ldots, X_{16})^\top \), we first test for a linear regression model by the proposed test because the dimension \( 16 \approx (337)^{0.476} \) and in the simplest case with linear model, the proposed test can theoretically handle \( p = O(n^{1/2}) \). The value of the test statistic is \( ACM_n^2 = 1.3651 \) 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 dataset is $q = 1$. This means that $Y$ may be conditionally independent of $X$ given the projected covariate $\hat{\beta}_1^\top X$ where

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

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 $(Y, X)$ is not reasonable.

To further exhaust possible projected covariates, we consider the second projected covariate $\hat{\beta}_2^\top X$ obtained by CSE. The scatter 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 regression information of $(Y, X)$. 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 mode:

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

Figure 1(b) adds the fitted curve on the scatter plot. The value of the test statistic $ACM_n^2 = 0.1038$ 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 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 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 cases with divergent dimension. Third, the model adaptation property shows its advantage in maintaining the significance level and enhancing power performance. The research also leaves some unsolved topics. An important topic is about how to relax the condition on the diverging rate of the dimension. In this paper, we cannot have faster rate than \( p = o(n^{1/4}) \) in general although for some special regression models such as linear models, it can achieve \( p = o(n^{1/2}) \). This is mainly because of technical difficulties in estimation. Thus, to attack this problem, we need to improve the asymptotic properties of involved estimators. This is beyond the scope of this paper and deserves further studies.

6 Appendix

6.1 Regularity Conditions

In this subsection we present some regularity conditions for the theoretical results. Although these conditions may not be the weakest possible, they make technical arguments easy to understand. In the following, \( C \) always stands for a constant which may be different in different
cases.

First, we give some regularity conditions for the norm consistency of \((\hat{\beta}_n, \hat{\theta}_n)\) to \((\tilde{\beta}_0, \tilde{\theta}_0)\) and the decomposition of \(\left( \hat{\beta}_n - \tilde{\beta}_0 \; \hat{\theta}_n - \tilde{\theta}_0 \right)\).

(A1) The matrix \(\Sigma_n\) is positive definite and satisfies the following condition

\[0 < \Delta \leq \lambda_{\min}(\Sigma_n) \leq \lambda_{\max}(\Sigma_n) \leq \bar{\lambda} < \infty\]

for all \(n\), where \(\lambda_{\min}(\Sigma_n)\) and \(\lambda_{\max}(\Sigma_n)\) are the smallest and largest eigenvalues of \(\Sigma\), respectively.

The first to third derivatives of the regression function \(g(\cdot)\) satisfy the conditions:

(A2) \(|E|Y|^4 \leq C, \; E|e|^8 \leq C; \; E|g'_j(\tilde{\beta}_0, \tilde{\theta}_0, X)|^8 \leq C;\)

(A3) \(|g'(\beta, \theta, x)| \leq F(x)\) with \(EF(X)^4 \leq C\) for all \((\beta, \theta)\);

(A4) \(|g'_j(\beta, \theta, x)| \leq F_j(x)\) with \(EF_j(X)^4 \leq C\) for all \(j\) and \((\beta, \theta)\);

(A5) \(|g''_{jk}(\beta, \theta, x)| \leq F_{jk}(x)\) with \(EF_{jk}(X)^4 \leq C\) for all \(j, k\), and \((\beta, \theta)\);

(A6) \(|g'''_{jkl}(\beta, \theta, x)| \leq F_{jkl}(x)\) with \(EF_{jkl}(X)^4 \leq C\) for all \(j, k, l\), and \((\beta, \theta)\);

where \(g'_j(\beta, \theta, x)\) is the \(j\)-th component of \(g'(\beta, \theta, x)\), \(g''_{jk}(\beta, \theta, x)\) is the \((j, k)\)-element of \(g''(\beta, \theta, x)\), and \(g'''_{jkl}(\beta, \theta, x)\) is the \((j, k, l)\)-element of \(g'''(\beta, \theta, x)\).

Condition (A1) is similar to the regularity condition on the Fisher information matrix \(I_n\) proposed by Fan and Peng (2004), where the Fisher information matrix \(I_n\) plays the same role in deriving the asymptotic theory as the matrix \(\Sigma_n\) does here. Conditions (A2)-(A6) are standard for nonlinear least squares estimation, see, e.g., Jennrich (1969) and White (1981).

Next, we present some regularity condition for the convergence of the adaptive-to-model residual marked empirical process.

(B1) There exists a constant \(C\) such that if \(\|\beta - \kappa \beta_0\| \leq C \sqrt{p \log n/n}\), then

\[P(\{\beta^\top X \leq u\} \triangle \{\kappa \beta_0^\top X \leq u\}) \leq \sqrt{2p \log n/n},\]

where \(\triangle\) denotes the symmetric difference of two sets. This condition is given by Zhu (1993) who showed the existence of distributions satisfying this condition.
(B2) If \( M_n(u) = E[g'(\beta_0, \theta_0, X)I(\kappa \beta_0^\top X \leq u)] \), \( \|M_n(u)\| = O(1) \) uniformly in \( u \).

(B3) For any unit non-random vector \( \gamma \in \mathbb{R}^p \), there exist \( F_Y \)-integrable functions \( h_i(t) \) such that

\[
E[\gamma^\top XI(Y \leq t)] \rightarrow h_1(t), \\
E[\gamma^\top X f_{Y|X}(t)G(X)] \rightarrow h_2(t),
\]

where \( G(X) \) is given by (3.8) and \( f_{Y|X} \) is the conditional density of \( Y \) given \( X \).

6.2 Lemmas

In this subsection we present some Lemmas that will be needed in proving the propositions and theorems. Since we consider the empirical process with diverging dimension, there are no relevant results available in the literature. Thus, in the following Lemmas, we give the results about the convergence rate of the involved empirical process, which are different from the classical ones with fixed dimension in the literature.

**Lemma 1.** Suppose \( A_n(u) \) is nonsingular for all \( u \), then we have

\[
\text{Cov}[T_n V^0_n(s), T_n V^0_n(t)] = \text{Cov}[V^0_n(s), V^0_n(t)] = \psi_n(s \wedge t),
\]

that is, (3.5) and thus (3.6) hold.

**Proof.** Assume \( t \leq s \). By the definition of \( T_n V^0_n \) and the Fubini Theorem, the left-hand side of (3.5) equals

\[
\begin{align*}
\text{Cov}[V^0_n(s), V^0_n(t)] & = -\int_{-\infty}^{t} a_n(z)^\top A_n(z)^{-1} \int_{z}^{t} a_n(v) \sigma_n^2(v) F_{\kappa \beta_0}(dv) \psi_n(dz) \\
& - \int_{-\infty}^{t} a_n(z)^\top A_n(z)^{-1} \int_{z}^{s} a_n(v) \sigma_n^2(v) F_{\kappa \beta_0}(dv) \psi_n(dz) \\
& + \int_{-\infty}^{t} \int_{-\infty}^{s} a_n(z_1)^\top A_n(z_1)^{-1} A_n(z_1 \vee z_2) A_n(z_2)^{-1} a_n(z_2) \psi_n(dz_1) \psi_n(dz_2).
\end{align*}
\]

It is easy to see that the sum of the last three terms is equal to zero. Thus we complete the proof. \( \square \)
Next we consider the convergence rate of the involved empirical processes in the diverging dimension. Let $F(x)$ be a fixed function and $\mathcal{F}_n$ be a VC-class of functions with a VC-index $V(\mathcal{F}_n)$ which may depend on $n$. Let $N_1(\epsilon, \mathcal{F}_n, L_1(Q))$ be the covering number of $\mathcal{F}_n$ with respective to the seminorm $L_1(Q)$. See e.g. Pollard (1984) for details. Suppose $\sup_{\mathcal{F}_n} |f(x)| \leq 1$ for any $n$ and $x \in \mathbb{R}^p$ and

$$\sup_{Q} N_2(\epsilon, \mathcal{F}_n, L_2(Q)) \leq A_n \epsilon^{-W_n} \quad \text{for all} \quad 0 < \epsilon < 1.$$ 

Set $\tilde{\mathcal{F}}_n = \{F(x)f(x) : f(x) \in \mathcal{F}_n\}$. By some elementary calculations, we have

$$N_1(\sqrt{QF^2} \epsilon, \tilde{\mathcal{F}}_n, L_1(Q)) \leq N_2(\epsilon, \mathcal{F}_n, L_2(Q)),$$

whence

$$\sup_{Q} N_1(\sqrt{QF^2} \epsilon, \tilde{\mathcal{F}}_n, L_1(Q)) \leq A_n \epsilon^{-W_n} \quad \text{for all} \quad 0 < \epsilon < 1.$$

**Lemma 2.** Let $\tau_n$ and $\epsilon_n$ be positive sequences. If $E|F|^4 < \infty$ and $\text{Var}(P_nFf)/(4\epsilon_n^2) \leq 1/2$ for $n$ large enough, then

$$\mathbb{P}\{\sup_{\mathcal{F}_n} |P_nFf - PFf| > 8\epsilon_n\} \leq 8A_n \tau_n W_n \epsilon_n^{-W_n} \exp\left(-\frac{1}{2}n\epsilon_n^2/\tau_n^2\right) + 4\mathbb{P}\{P_nF^2 > \tau_n^2\}$$

where $A_n$ and $W_n$ are constants which may depend on $n$.

**Proof.** The proof is similar to Theorem 37 in Chapter 2 of Pollard (1984) and Theorem 3.1 in Zhu (1993). Since $\text{Var}(P_nFf)/(4\epsilon_n^2) \leq 1/2$ for $n$ large enough, by the formula (30) in Chapter 2 of Pollard (1984), we have

$$\mathbb{P}\{\sup_{\mathcal{F}_n} |P_nFf - PFf| > 8\epsilon_n\} \leq 4\mathbb{P}\{\sup_{\mathcal{F}_n} |P_n^oFf| > 2\epsilon_n\},$$

Conditionally on $\{X_1, \cdots, X_n\}$. Using the same argument as that for proving the inequality (31) in Chapter 2 of Pollard (1984), it follows that

$$\mathbb{P}\left\{\sup_{\mathcal{F}_n} |P_n^oFf| > 2\epsilon_n|X_1, \cdots, X_n\right\} \leq 2N_1(\epsilon_n, \tilde{\mathcal{F}}_n, L_1(P_n)) \exp\left(-\frac{1}{2}n\epsilon_n^2/(\max_j P_ng_j^2)\right) \wedge 1$$

$$\leq 2A_n(P_nF^2)^{W_n/2} \epsilon_n^{-W_n} \exp\left(-\frac{1}{2}n\epsilon_n^2/P_nF^2\right) \wedge 1.$$
Taking expectation, we obtain that

\[ \mathbb{P}\{\sup_{F_n} |P_n^o F f| > 2\epsilon_n\} \leq 2A_n^W n^{n} \epsilon_n^W \exp\left(-\frac{1}{2} n\epsilon_n^2 / \tau_n^2\right) + \mathbb{P}\{P_n^o \geq \tau_n^2\}. \]

Consequently,

\[ \mathbb{P}\{\sup_{F_n} |P_n F f - \bar{F} f| > 8\epsilon_n\} \leq 8A_n^W n^{n} \epsilon_n^W \exp\left(-\frac{1}{2} n\epsilon_n^2 / \tau_n^2\right) + 4\mathbb{P}\{P_n^2 \geq \tau_n^2\}. \]

Therefore, we complete the proof. \( \square \)

**Lemma 3.** If \( |g_j'(\beta_0, \theta_0, x)| \leq F(x) \) and \( E[F(X)]^4 < \infty \), then we have

\[ \sup_u \left\| \frac{1}{n} \sum_{i=1}^n g'(\beta_0, \theta_0, X_i)I(\kappa \beta_0^T X_i \leq u) - M_n(u) \right\| = o_p\left( \sqrt{\frac{p}{n}} p^{1/4} \log n \right), \]

where \( M_n(u) = E[g'(\beta_0, \theta_0, X)I(\kappa \beta_0^T X \leq u)] \).

**Proof.** Fix \( \epsilon > 0 \) and let \( \epsilon_n^2 = \epsilon^2 (\log n)^2 \sqrt{p}/n \). We have

\[
\begin{align*}
\mathbb{P}& \left( \sup_u \left\| \frac{1}{n} \sum_{i=1}^n g'(\beta_0, \theta_0, X_i)I(\kappa \beta_0^T X_i \leq u) - M_n(u) \right\| > 8\sqrt{p}\epsilon_n \right) \\
& \leq \mathbb{P} \left( \sup_u \left\| \frac{1}{n} \sum_{i=1}^n g'(\beta_0, \theta_0, X_i)I(\kappa \beta_0^T X_i \leq u) - M_n(u) \right\|^2 > 64\epsilon_n^2 \right) \\
& = \mathbb{P} \left( \sup_u \sum_{j=1}^p \left\| \frac{1}{n} \sum_{i=1}^n g_j'(\beta_0, \theta_0, X_i)I(\kappa \beta_0^T X_i \leq u) - M_{nj}(u) \right\|^2 > 64\epsilon_n^2 \right) \\
& \leq \sum_{j=1}^p \mathbb{P} \left( \sup_u \left\| \frac{1}{n} \sum_{i=1}^n g_j'(\beta_0, \theta_0, X_i)I(\kappa \beta_0^T X_i \leq u) - M_{nj}(u) \right\| > 8\epsilon_n \right). 
\end{align*}
\]

For every term in the last sum, we use Lemma 2. Let

\[
\mathcal{F}_{1n} = \{f_u(x) = I(\kappa \beta_0^T x \leq u) : u \in \mathbb{R}\}
\]

\[
\mathcal{F}_{1n} = \{g_j'(\beta_0, \theta_0, x) f_u(x) : f_u(x) \in \mathcal{F}_{1n}\}.
\]

29
It is easy to see that $\mathcal{F}_{1n}$ is a VC-class with VC-index $V(\mathcal{F}_{1n}) = 2$. By Theorem 2.6.7 in Van Der Vaart and Wellner (1996), we obtain that

$$\sup_Q N_2(\epsilon, \mathcal{F}_{1n}, L_2(Q)) \leq 2K(16e)^2 \epsilon^{-2},$$

$$\sup_Q N_1(\epsilon(Qg_j^2)^{1/2}, \mathcal{F}_{1n}, L_1(Q)) \leq 2K(16e)^2 \epsilon^{-2},$$

where $K$ is a universal constant. Set $A = 2K(16e)^2$ and $\tau_n^2 = \sqrt{p \log n}$. Lemma 2 leads to

$$\mathbb{P} \left( \sup_{\mathcal{F}_n} |P_n g_j^f f_{\beta,u} - P g_j^f f_{\beta,u}| > 8\epsilon_n \right) \leq 8A \tau_n^2 \epsilon_n^{-2} \exp \left( -\frac{n \epsilon_n^2}{2 \tau_n^2} \right) + 4\mathbb{P} \{ P_n F^2 \geq \tau_n^2 \},$$

whence

$$\mathbb{P} \left( \sup_{\mathcal{F}_n} \frac{1}{n} \sum_{i=1}^{n} g'((\beta_0, \theta_0, X_i) I(\kappa_{\beta_0}^T X_i \leq u) - M_n(u) \| > 8\sqrt{p} \epsilon_n \right) \leq 8A \tau_n^2 \epsilon_n^{-2} \exp \left( -\frac{n \epsilon_n^2}{2 \tau_n^2} \right) + 4\mathbb{P} \{ P_n F^2 \geq \tau_n^2 \} \leq 8A \frac{p}{\epsilon^2 (\log n)^{3/2}} \exp \left( -\frac{1}{2} \epsilon \sqrt{\log n \log n} \right) + 4pF^4/\log n = o(1).$$

Therefore, we obtain the result.  

**Lemma 4.** Let $\mathcal{F}$ be a permissible class of functions with $|f| \leq 1$ and $P|f| \leq \delta$ for all $f \in \mathcal{F}$. Then

$$\mathbb{P}(\sup_{\mathcal{F}} |P_n f| > 8\delta) \leq 4P[N_1(\delta, \mathcal{F}, L_1(P_n)) \exp(-n\delta^2) \land 1].$$

For the definition of “a permissible class of functions”, one can refer to Chapter 2 of Pollard (1984) for details.

**Proof.** This Lemma is a slightly modified version of Lemma 33 in Chapter 2 of Pollard (1984) as we need the result with diverging $p$. But the proof can be very similar and thus is omitted here.  

**Lemma 5.** Let $\delta_n$ and $\alpha_n$ be positive real valued sequences. Suppose $P|f| \leq \delta_n$ for all $f(x) \in \mathcal{F}_n$ and $\text{Var}(P_n F f)/(4\epsilon_n)^2 \leq 1/2$ for $n$ large enough. If $E|F|^8 < \infty$, then

$$\mathbb{P}(\sup_{\mathcal{F}_n} |P_n F f - PF f| > 8\epsilon_n) \leq 8A_n \alpha_n^W \epsilon_n^{-W_n} \exp \left( -\frac{1}{2} n \epsilon_n^2 /[\alpha_n^2 (8\delta_n)^{\frac{3}{2}}] \right) + 4\mathbb{P}(P_n F^8 > \alpha_n^8) + 16A_n \delta_n^{-W_n} \exp(-n\delta_n^2).$$
where $A_n$ and $W_n$ are constants which may depend on $n$.

**Proof.** The proof is similar to Theorem 37 in Chapter 2 of Pollard (1984) and Theorem 3.1 of Zhu (1993). Since $\text{Var}(P_n F f)/(4\epsilon_n)^2 \leq 1/2$, similar to the proof for Lemma 2, we have

$$\mathbb{P}\{\sup_{\mathcal{F}_n} |P_n F f - P f f| > 8\epsilon_n\} \leq 4\mathbb{P}\{\sup_{\mathcal{F}_n} |P_n^0 F f| > 2\epsilon_n\}.$$  

Conditionally on $\{X_1, \ldots, X_n\}$, we obtain

$$\mathbb{P}\left\{\sup_{\mathcal{F}_n} |P_n^0 F f| > 2\epsilon_n |X_1, \ldots, X_n\right\} \leq 2N_1(\epsilon_n, \tilde{F}_n, L_1(P_n)) \exp\left(-\frac{1}{2}n\epsilon_n^2 / (\max_j P_n F^2 f_j^2)\right) + 1$$

$$\leq 2A_n(P_n F^2)^{W_n/2} \epsilon_n^{-W_n} \exp\left(-\frac{1}{2}n\epsilon_n^2 / (\max_j [P_n F^8]^{1/4} [P_n |f_j|]^{2})\right) + 1$$

Take expectation to obtain

$$\mathbb{P}\{\sup_{\mathcal{F}_n} |P_n^0 F f| > 2\epsilon_n\} \leq 2A_n \alpha_n^{W_n} \epsilon_n^{-W_n} \exp\left(-\frac{1}{2}n\epsilon_n^2 / [\alpha_n^2 (8\delta_n)^2]\right) + \mathbb{P}(P_n F^8 > \alpha_n^8) + \mathbb{P}\{\sup_{\mathcal{F}_n} |f| \geq 8\delta_n\}$$

$$\leq 2A_n \alpha_n^{W_n} \epsilon_n^{-W_n} \exp\left(-\frac{1}{2}n\epsilon_n^2 / [\alpha_n^2 (8\delta_n)^2]\right) + \mathbb{P}(P_n F^8 > \alpha_n^8) + 4A_n \delta_n \exp(-n\delta_n^2).$$

The last inequality is due to Lemma 4. Altogether we complete the proof. \qed

**Lemma 6.** Suppose $H_0$ and condition (B1) hold. If $E\epsilon^8 < \infty$ and $p^4/n \to 0$, then we have

$$\sup_{u} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i [I(\hat{\beta}_n^\top X_i \leq u) - I(\kappa \beta_0^\top X_i \leq u)] \right\} = o_p(1)$$

**Proof.** Fix $\epsilon > 0$ and set $H_C = \{ \beta : \beta \in \mathbb{R}^p, \|\beta\| \leq 1, \|\beta - \kappa \beta_0\| \leq C \sqrt{p \log n/n} \}$. Since $\|\hat{\beta}_n - \kappa \beta_0\| = O_p(\sqrt{p/n})$, by condition (B1), it suffices to prove

$$\mathbb{P}\left\{ \sup_{\beta \in H_C} \sup_{u} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i [I(\beta^\top X_i \leq u) - I(\kappa \beta_0^\top X_i \leq u)] \right\} > \frac{8}{\sqrt{n}} \epsilon \right\} \to 0,$$
Let
\[
\mathcal{F}_{2n} = \{ f_{\beta,u}(x) = I(\beta^\top x \leq u) - I(\kappa\beta_0^\top x \leq u) : \beta \in H_C, u \in \mathbb{R} \},
\]
\[
\mathcal{F}_{2n}^1 = \{ f_{\beta,u}(x) = I(\beta^\top x \leq u) : \beta \in H_C, u \in \mathbb{R} \},
\]
\[
\mathcal{F}_{2n}^2 = \{ f_u(x) = I(\kappa\beta_0^\top x \leq u) : u \in \mathbb{R} \}.
\]
Then it is easy to see that
\[
N_2(2\epsilon, \mathcal{F}_{2n}, L_2(Q)) \leq N_2(\epsilon, \mathcal{F}_{2n}^1, L_2(Q)) \cdot N_2(\epsilon, \mathcal{F}_{2n}^2, L_2(Q)).
\]
Since \(\mathcal{F}_{2n}^1\) and \(\mathcal{F}_{2n}^2\) are both VC-classes with the VC-index \(p + 2\) and \(2\) respectively, by Theorem 2.6.7 in Van Der Vaart and Wellner (1996), we obtain that
\[
\sup_Q N_2(\epsilon, \mathcal{F}_{2n}^1, L_2(Q)) \leq K(p + 2)(16\epsilon)^{p+2} \epsilon^{-2(p+1)},
\]
\[
\sup_Q N_2(\epsilon, \mathcal{F}_{2n}^2, L_2(Q)) \leq 2(16\epsilon)^2 \epsilon^{-2},
\]
whence
\[
\sup_Q N_2(\epsilon, \mathcal{F}_{2n}, L_2(Q)) \leq K(p + 2)(64\epsilon)^{p+2} \epsilon^{-2(p+2)}.
\]
Let \(\tilde{\mathcal{F}}_{2n} = \{ \epsilon f_{\beta,u}(x) : f_{\beta,u} \in \mathcal{F}_{2n} \}\). It follows that
\[
\sup_Q N_1(\epsilon \sqrt{Q\epsilon^2}, \tilde{\mathcal{F}}_{2n}, L_1(Q)) \leq K(p + 2)(64\epsilon)^{p+2} \epsilon^{-2(p+2)}.
\]
Let \(\delta_n = \sqrt{2p \log n/n}\), \(\alpha_n^8 = \log n\) and \(\epsilon_n = \epsilon/\sqrt{n}\). Since
\[
P|f_{\beta,u}| = \mathbb{P}(\{\beta^\top X \leq u\} \triangle \{\kappa\beta_0^\top X \leq u\}) \leq \delta_n,
\]
\[
\frac{\text{Var}(P_n\epsilon f_{\beta,u})}{(4\epsilon_n^2)^{1/2} Pf_{\beta,u}} \leq \frac{1}{2} \quad \text{for } n \text{ large enough},
\]
by Lemma 5, we have
\[
P\left\{ \sup_{\mathcal{F}_{2n}} |P_n \epsilon f_{\beta,u}| > \frac{8}{\sqrt{n}} \right\} \leq 8K(p + 2)(64\epsilon)^{p+2} \alpha_n^2(\epsilon_n^2)(\epsilon_n^2)^{-2(p+2)} \exp\left( -\frac{n\epsilon_n^2}{2\alpha_n^2(8\delta_n)^2} \right) + 4P(P_n \epsilon^8 > \alpha_n^8) + 16K(p + 2)(64\epsilon)^{p+2} \delta_n^{-2(p+2)} \exp(-n\delta_n^2).
\]
Then we have (A2) and (B1) hold. If by the same argument for proving Lemma 3, we obtain Lemma 7. Let \( \tilde{M}_n(\beta, u) = E\{g'(\beta_0, \theta_0, X)|I(\beta^T X_i \leq u) - I(\kappa \beta_0^T X_i \leq u)|\} \). Suppose conditions (A2) and (B1) hold. If \( p^4/n \to 0 \), then
\[
\sup_u \frac{1}{n} \sum_{i=1}^{n} g'(\beta_0, \theta_0, X_i)[I(\hat{\beta}_n^T X_i \leq u) - I(\kappa\beta_0^T X_i \leq u)] - \tilde{M}_n(\beta, u) = o_p(\sqrt{\frac{p^{3/2} \log n}{n}}).
\]

**Proof.** Fix \( \epsilon > 0 \) and set \( \epsilon_n = \epsilon \sqrt{\frac{p^{1/2} \log n}{n}}, \delta_n = 2p \log n/n \), and \( \alpha_n = p \log n \). Similar to the proof for Lemma 6, it suffices to prove
\[
P \left\{ \sup_{\beta \in H_C} \sup_u \frac{1}{n} \sum_{i=1}^{n} g'(\beta_0, \theta_0, X_i)[I(\beta^T X_i \leq u) - I(\kappa \beta_0^T X_i \leq u)] - \tilde{M}_n(\beta, u) > 8p \epsilon_n \right\} \to 0.
\]
By the same argument for proving Lemma 3, we obtain
\[
P \left\{ \sup_{\beta \in H_C} \sup_u \frac{1}{n} \sum_{i=1}^{n} g'(\beta_0, \theta_0, X_i)[I(\beta^T X_i \leq u) - I(\kappa \beta_0^T X_i \leq u)] - \tilde{M}_n(\beta, u) > 8p \epsilon_n \right\}
\leq \sum_{j=1}^{p+d} P \left\{ \sup_{\beta \in H_C} \sup_u \frac{1}{n} \sum_{i=1}^{n} g_j'(\beta_0, \theta_0, X_i)[I(\beta^T X_i \leq u) - I(\kappa \beta_0^T X_i \leq u)] - \tilde{M}_{nj}(\beta, u) > 8 \epsilon_n \right\}.
\]
For every term in the last sum, we use Lemma 5 to derive the result. Let
\[
\mathcal{F}_{3n} = \{g_j'(\beta_0, \theta_0, x)f_{\beta,u}(x) : f_{\beta,u}(x) \in \mathcal{F}_{3n}\},
\]
\[
\mathcal{F}_{3n} = \{f_{\beta,u}(x) = I(\beta^T x \leq u) - I(\kappa \beta_0^T x \leq u) : \beta \in H_C, u \in \mathbb{R}\}.
\]
Then we have
\[
\sup_Q N_2(\epsilon, \mathcal{F}_{3n}, L_2(Q)) \leq K(p + 2)(64 \epsilon)^{p+2} \epsilon^{-2(p+2)},
\]
\[
\sup_Q N_1(\epsilon \sqrt{Q g_j''}, \mathcal{F}_{3n}, L_1(Q)) \leq K(p + 2)(64 \epsilon)^{p+2} \epsilon^{-2(p+2)}.
\]

33
where $K$ is a universal constant free of $n$.

Recall $\epsilon_n = \epsilon \sqrt{p^{1/2} \log n}/n$ and $\delta_n = \sqrt{2p \log n}/n$. By conditions (A2) and (B1), we have

$$P|f_{\beta,u}| = \mathbb{P}(\{\beta^\top X \leq u\} \Delta \{\kappa \beta_0^\top X \leq u\}) \leq \delta_n;$$

$$\frac{\text{Var}(P_n g_j f_{\beta,u})}{(4\epsilon_n)^2} \leq \frac{Pg_j^2}{16\epsilon^2 p \log n} < \frac{1}{2} \text{ for } n \text{ large enough.}$$

By Lemma 5, we obtain that

$$\mathbb{P}\left\{ \sup_{\beta \in H_{C}} \sup_{u} \left| \frac{1}{n} \sum_{i=1}^{n} g_j'(\beta_0, \theta_0, X)[I(\beta^\top X_i \leq u) - I(\kappa \beta_0^\top X_i \leq u)] - \tilde{M}_n(\beta, u) \right| > 8\epsilon_n \right\}$$

$$\leq 8K(p + 2)(64\epsilon)^{p+2}\alpha_n^{2(p+2)}\epsilon^{-2(p+2)} \exp\left(-\frac{n\epsilon_n^2}{2\alpha_n^2(8\delta_n)^4}\right) + 4\mathbb{P}(P_n g_j^8 > \alpha_n^8)$$

$$+ 16K(p + 2)(64\epsilon)^{p+2}\delta_n^{-2(p+2)} \exp(-n\delta_n^2),$$

whence

$$\mathbb{P}\left\{ \sup_{\beta \in H_{C}} \sup_{u} \left| \frac{1}{n} \sum_{i=1}^{n} g_j'(\beta_0, \theta_0, X)[I(\beta^\top X_i \leq u) - I(\kappa \beta_0^\top X_i \leq u)] - \tilde{M}_n(\beta, u) \right| > 8\sqrt{p}\epsilon_n \right\}$$

$$\leq 8K(p + d)(p + 2)(64\epsilon)^{p+2}\alpha_n^{2(p+2)}\epsilon^{-2(p+2)} \exp\left(-\frac{n\epsilon_n^2}{2\alpha_n^2(8\delta_n)^4}\right) + \sum_{j=1}^{p+d} 4\mathbb{P}(P_n g_j^8 > \alpha_n^8)$$

$$+ 16K(p + d)(p + 2)(64\epsilon)^{p+2}\delta_n^{-2(p+2)} \exp(-n\delta_n^2),$$

Since $p^4/n \to 0$, it follows that the right-hand side of the above inequality tends to zero. Hence we complete the proof. \hfill \square

In the next lemma, we give the convergence rate of the kernel regression function estimator $\hat{r}_n(y)$. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample from $(X, Y)$, $f(y)$ be the density function of $Y$ with a support $C$ and $m_n(y) = r_n(y)f(y) = \{E(X_{11}|Y = y)f(y), \ldots, E(X_{1p}|Y = y)f(y)\}^T$.

Suppose that

$$0 < \inf_{y \in C} f(y) \leq \sup_{y \in C} f(y) < \infty,$$

$$E(X_{1i}^2|Y = y) \leq C \text{ for all } y, 1 \leq i \leq p.$$
It follows that \( m_n(y) = O_p(\sqrt{p}) \) uniformly in \( y \). Set

\[
\hat{m}_n(y) = \frac{1}{nh} \sum_{j=1}^{n} X_j K\left(\frac{y - Y_j}{h}\right) \quad \text{and} \quad \hat{f}(y) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{y - Y_j}{h}\right).
\]

Then \( \hat{r}_n(y) = \hat{m}_n(y)/\hat{f}(y) \). Here \( K(\cdot) \) is the kernel function and \( h \) is a bandwidth.

**Lemma 8.** Suppose the above conditions hold. If \( ph^3 \log n \to 0 \), then we have

\[
\sup_y \| \hat{m}_n(y) - m_n(y) \| = O_p(\sqrt{\frac{p \log n}{nh^2}}) + O_p(\sqrt{ph^4}),
\]

\[
\sup_y \| \hat{r}_n(y) - r_n(y) \| = O_p(\sqrt{\frac{p \log n}{nh^2}}) + O_p(\sqrt{ph^4}).
\]

**Proof.** Let \( r_m(y), \hat{r}_m(y), m_m(y), \) and \( \hat{m}_m(y) \) be the \( i \)-th component of \( r_n(y), \hat{r}_n(y), m_n(y), \) and \( \hat{m}_n(y) \) respectively. For fixed \( \epsilon \), set \( \epsilon_i = \frac{\log n}{n} \epsilon^2, \delta_n = h, \) and \( \alpha_8 = p \log n \). Then

\[
\mathbb{P}(\sup_y \| \hat{m}_n(y) - E\hat{m}_n(y) \| > 8\sqrt{\frac{p \log n}{nh^2}} \epsilon) \leq \sum_{i=1}^{p} \mathbb{P}(\sup_{y} |h\hat{m}_n(y) - hE\hat{m}_n(y)| > 8\epsilon_n)
\]

Define

\[
\mathcal{F}_{4n} = \left\{ f_{y,h}(u) : f_{y,h}(u) = K\left(\frac{y - u}{h}\right), y \in \mathcal{C} \right\}.
\]

Without loss of generality, assume \( |K(x)| \leq 1 \) and \( f(y) \leq 1 \). By the arguments in Example 38 of Chapter 2 of Pollard (1984), we obtain that

\[
\sup_{Q} N_2(\epsilon, \mathcal{F}_n, L_2(Q)) \leq A \epsilon^{-W} \quad \text{for all} \quad 0 < \epsilon < 1,
\]

where \( A \) and \( W \) are free of \( n \). Let \( \tilde{\mathcal{F}}_{4n} = \{ zf_{y,h}(u) : f_{y,h}(u) \in \mathcal{F}_{4n} \} \). Then

\[
\sup_{Q} N_1(\sqrt{Qz^2} \epsilon, \tilde{\mathcal{F}}_{4n}, L_1(Q)) \leq A \epsilon^{-W}.
\]

Since

\[
P|f_{y,h}| = \int |K\left(\frac{y - u}{h}\right)| f(u)du = h \int |K(u)| f(y - uh)du \leq h = \delta_n
\]

and

\[
\frac{Var(P_{y,h} K(\frac{y - u}{h}))}{16\epsilon_n^2} \leq \frac{EX_1^2 K^2(\frac{y - y}{h})}{16\epsilon_n^2} \leq \frac{hC \int K^2(u)du}{16n\epsilon_n^2} = \frac{Ch \int K^2(u)du}{\epsilon \log n} < \frac{1}{2}
\]
for $n$ large enough, Lemma 5 yields that

$$P(\sup_y |\hat{m}_{ni}(y) - E\hat{m}_{ni}(y)| > 8\epsilon_n)$$

$$\leq 8A(p \log n)^{W/8} \epsilon^{-W/(n \log n)^{W/2}} \exp\left\{ -\frac{\epsilon^2 \log n}{2(p \log n)^{1/4}(8h)^{3/4}} \right\} + 4P\left( P_n X_{i1}^8 > p \log n \right)$$

$$+ 16Ah^{-W} \exp(-nh^2),$$

whence

$$P(\sup_y \|\hat{m}_n(y) - E\hat{m}_n(y)\| > 8\sqrt{\frac{p \log n}{nh^2}} \epsilon)$$

$$\leq 8Ap(p \log n)^{W/8} \epsilon^{-W/(n \log n)^{W/2}} \exp\left\{ -\frac{\epsilon^2 \log n}{2(p \log n)^{1/4}(8h)^{3/4}} \right\} + \sum_{i=1}^p 4P\left( P_n X_{i1}^8 > p \log n \right)$$

$$+ 16Ap\epsilon^{-W} \exp(-nh^2).$$

Since $ph^3 \log n \to 0$, it is easy to see that the right-hand side of the inequality tends to zero. Thus $\sup_y \|\hat{m}_n(y) - E\hat{m}_n(y)\| = o_p(\sqrt{p \log n/(nh^2)})$. By the arguments for proving Lemma 3.3 of Zhu and Fang (1996), we obtain that

$$\sup_y |E\hat{m}_{ni}(y) - m_{ni}(y)| \leq Ch^4 \int |K(u)|u^4 du.$$ 

Consequently,

$$\sup_y \|E\hat{m}_n(y) - m_n(y)\| \leq C\sqrt{ph^4} \int |K(u)|u^4 du.$$ 

Thus we obtain the first result. For the second, note that

$$\|\hat{r}_n(y) - r_n(y)\| \leq \frac{\|\hat{m}_n(y) - E\hat{m}_n(y)\|}{|f(y)|} + \|E\hat{m}_n(y)\| \left| \frac{1}{\hat{f}(y)} - \frac{1}{f(y)} \right| + \|\frac{E\hat{m}_n(y) - m_n(y)}{f(y)}\|$$

and

$$\sup_y |\hat{f}(y) - f(y)| = O_p(h^4) + O_p(\sqrt{\log n/(nh^2)}).$$

Combining these with the uniformly boundedness of $f(y)$, the proof is concluded. $\square$
6.3 Proofs of The Propositions and Theorems

For simplicity of notations, we consider a parametric family of functions $G = \{ g(\beta, \cdot) : \beta \in \Theta \subset \mathbb{R}^p \}$. Let $\hat{\beta}_n = \text{argmin}_\beta \sum_{i=1}^n [Y_i - g(\beta, X_i)]^2$ and $\tilde{\beta}_0 = \text{argmin}_\beta E[Y - g(\beta, X)]^2$.

**Proof of Proposition 1.** Let $\beta = \tilde{\beta}_0 + \alpha$ and $F(\alpha) = \sum_{i=1}^n [Y_i - g(\tilde{\beta}_0 + \alpha, X_i)]g'(\tilde{\beta}_0 + \alpha, X_i)$.

Then it suffices to show that there is a root $\alpha_n$ of $F(\alpha)$ such that $\|\alpha_n\|^2 = O_p(p/n)$. Applying the results in (6.3.4) of Ortega and Rheinboldt (1970), it in turn needs to show that $\alpha^T F(\alpha) < 0$ for $\|\alpha\|^2 = Cp/n$ where $C$ is some large enough constant.

Let $\alpha = \sqrt{p/n}U$ with $\|U\| = C$, and $e_i = Y_i - g(\tilde{\beta}_0, X_i)$. Using Taylor’s expansion we obtain

$$
\alpha^T F(\alpha) = \sum_{i=1}^n \alpha^T g'(\tilde{\beta}_0 + \alpha, X_i)[Y_i - g(\tilde{\beta}_0 + \alpha, X_i)] = \sum_{i=1}^n \alpha^T g'(\tilde{\beta}_0, X_i)e_i - \sum_{i=1}^n \alpha^T g'(\tilde{\beta}_0 + \alpha, X_i)[g(\tilde{\beta}_0 + \alpha, X_i) - g(\tilde{\beta}_0, X_i)]
$$

$$
= \sum_{i=1}^n \alpha^T g'(\tilde{\beta}_0, X_i)e_i + \alpha^T \sum_{i=1}^n e_i g''(\tilde{\beta}_0, X_i)\alpha + \frac{1}{2} \alpha^T \sum_{i=1}^n \alpha^T e_i g'''(\beta_{1n}, X_i)\alpha - \sum_{i=1}^n \alpha^T g''(\beta_{2n}, X_i)\alpha g'(\tilde{\beta}_0, X_i)^\top \alpha - \sum_{i=1}^n \alpha^T g''(\beta_{3n}, X_i)\alpha g'(\beta_{3n}, X_i)^\top \alpha
$$

$$
=: A_1 - A_2 + A_3,
$$

where $\beta_{1n}, \beta_{2n}, \beta_{3n}$ lie between $\tilde{\beta}_0$ and $\tilde{\beta}_0 + \alpha$ and

$$
A_1 = \sum_{i=1}^n \alpha^T g'(\tilde{\beta}_0, X_i)e_i,
$$

$$
A_2 = \alpha^T \sum_{i=1}^n [g'(\tilde{\beta}_0, X_i)g'(\tilde{\beta}_0, X_i)^\top - e_i g''(\tilde{\beta}_0, X_i)]\alpha,
$$

$$
A_3 = \frac{1}{2} \alpha^T \sum_{i=1}^n \alpha^T e_i g'''(\beta_{1n}, X_i)\alpha - \sum_{i=1}^n \alpha^T g''(\beta_{2n}, X_i)\alpha g'(\tilde{\beta}_0, X_i)^\top \alpha - \sum_{i=1}^n \alpha^T g''(\beta_{3n}, X_i)\alpha g'(\beta_{3n}, X_i)^\top \alpha.
$$
Then we have \( |A_1| \leq \sqrt{p/n}||U||\| \sum_{i=1}^n g' \beta_0, X_i e_i \|. \) Since \( E[g' \beta_0, X_i e_i] = 0 \), it follows that
\[
E\| \sum_{i=1}^n g' \beta_0, X_i e_i \|^2 = nE\{ g' \beta_0, X_i \|^2 = n \sum_{j=1}^p E[e_1 g'_j \beta_0, X_i]^2 \leq npC.
\]
Thus \( A_1 = p ||U|| O_p(1) \). Recall that \( \Sigma_n = E[g' \beta_0, X]g' \beta_0, X^\top \), \( \Sigma_{n2} = E[e g'' \beta_0, X] \), and \( \Sigma_n = \Sigma_n1 - \Sigma_{n2} \). Then we decompose the term \( A_2 \) as follows
\[
A_2 = p U^\top \Sigma_n U + p U^\top \{ \frac{1}{n} \sum_{i=1}^n [g' \beta_0, X_i]g' \beta_0, X_i^\top - e_i g'' \beta_0, X_i] - \Sigma_n \} U.
\]
By condition (A2), we obtain that
\[
E\| \frac{1}{n} \sum_{i=1}^n (g' \beta_0, X_i)g' \beta_0, X_i^\top - \Sigma_n1 \|^2
= \frac{1}{n^2} \sum_{j,k=1}^p E\{ \sum_{i=1}^n [g'_j \beta_0, X_i]g'_k \beta_0, X_i - \Sigma_{n1jk} \} \|^2
= \frac{1}{n^2} \sum_{j,k=1}^p \sum_{i=1}^n E[g'_j \beta_0, X_i]g'_k \beta_0, X_i - \Sigma_{n1jk} \] \] \] \[
\leq \frac{1}{n} \sum_{j,k=1}^p E[g'_j \beta_0, X_i]g'_k \beta_0, X_i \] \] \[
\leq \frac{p^2}{n} C.
\]
It follows that \( \frac{1}{n} \sum_{i=1}^n [g' \beta_0, X_i]g' \beta_0, X_i^\top - \Sigma_n1 = \frac{p}{\sqrt{n}} O_p(1) \). By the same argument, we have
\[
1 \sum_{i=1}^n [e_i g'' \beta_0, X_i] - \Sigma_{n2} = \frac{p}{\sqrt{n}} O_p(1).
\]
Therefore \( A_2 = p U^\top \Sigma_n U + \frac{p^2}{\sqrt{n}} ||U||^2 O_p(1) = p U^\top \Sigma_n U + p ||U||^2 O_p(1) \). For the first term of \( A_3 \), by the triangle inequality and condition (A6), we have
\[
E[\alpha^\top \sum_{i=1}^n \alpha^\top e_i g''' \beta_1n, X_i \alpha] \leq ||\alpha||^3 E\| \sum_{i=1}^n e_i g''' \beta_1n, X_i \| = ||\alpha||^3 n E\| e_i g''' \beta_1n, X_i \|
\leq n ||\alpha||^3 (E\| e_i g''' \beta_1n, X_i \|^2)^{\frac{1}{2}} \leq n ||\alpha||^3 (\sum_{j,k,l=1}^p E[e_1 g''' \beta_1n, X_i]^2)^{\frac{1}{2}}
\leq \frac{p^3}{\sqrt{n}} ||U||^3 C.
\]
For the second term of $A_3$, we have
\[ E \left| \sum_{i=1}^{n} \alpha^\top g''(\beta_{2n}, X_i) \alpha g'(\hat{\beta}_0, X_i)^\top \alpha \right| \leq \sum_{i=1}^{n} E \left| \alpha^\top g''(\beta_{2n}, X_i) \alpha g'(\hat{\beta}_0, X_i)^\top \alpha \right| \]
\[ \leq n \{ E[\alpha^\top g''(\beta_{2n}, X_1)\alpha]^2 \}^{1/2} \{ E[\alpha^\top g'(\hat{\beta}_0, X_1)^\top \alpha]^2 \}^{1/2} \]
\[ \leq n \| \alpha \|^3 \{ E \| g''(\beta_{2n}, X_1) \|^2 \}^{1/2} \{ E \| g'(\hat{\beta}_0, X_1) \|^2 \}^{1/2} \]
\[ \leq \frac{p^3}{\sqrt{n}} \| U \|^3 C. \]

By the same argument for the third and forth term of $A_3$, we obtain that $A_3 = \frac{p^3}{\sqrt{n}} \| U \|^3 O_p(1) + \frac{p^4}{n} \| U \|^4 O_p(1)$. Therefore
\[ \alpha^\top F(\alpha) = p \| U \| O_p(1) - p U^\top \Sigma_n U + p \| U \|^2 O_p(1) \]
\[ \leq p \| U \| O_p(1) - p \lambda_{\min}(\Sigma_n) \| U \|^2 + p \| U \|^2 O_p(1) \]
\[ = p \| U \| \{ O_p(1) - \lambda_{\min}(\Sigma_n) \| U \|^2 \} + \| U \| O_p(1) \}. \]

If $\| U \| = C$ be large enough, for any $\epsilon > 0$, we have
\[ \mathbb{P}(\alpha^\top F(\alpha) < 0) \geq \mathbb{P}\{ O_p(1) - \lambda_{\min}(\Sigma_n) \| U \|^2 + \| U \| O_p(1) < 0 \} \geq 1 - \epsilon. \]

Thus our result follows from (6.3.4) of Ortega and Rheinboldt (1970). \qed

If $g(\beta, X) = \beta^\top X$ follows a linear regression model, then $g''(\beta, x) = 0$ and $g'''(\beta, x) = 0$. According to the proof of Proposition 1, we can obtain the norm consistency of $\hat{\beta}_n$ under the weaker condition $p^2/n \rightarrow 0$.

**Proof of Proposition 2.** We use the same notations as those in the proof of Proposition 1. Let $\Psi_n(\beta) = \sum_{i=1}^{n} [Y_i - g(\beta, X_i)] g'(\beta, X_i)$. Then $\Psi_n(\hat{\beta}_n) = 0$. Applying Taylor’s expansion around $\hat{\beta}_0$, we obtain
\[ 0 = \Psi_n(\hat{\beta}_n) = \Psi_n(\hat{\beta}_0) + \Psi'_n(\hat{\beta}_0) (\hat{\beta}_n - \hat{\beta}_0) + \frac{1}{2} (\hat{\beta}_n - \hat{\beta}_0)^\top \Psi''_n(\beta_{4n}) (\hat{\beta}_n - \hat{\beta}_0) \]
where $\beta_{4n}$ lies between $\hat{\beta}_n$ and $\hat{\beta}_0$. Therefore
\[ \Sigma_n(\hat{\beta}_n - \hat{\beta}_0) = \frac{1}{n} \Psi_n(\hat{\beta}_0) + [\Sigma_n + \frac{1}{n} \Psi'_n(\hat{\beta}_0)] (\hat{\beta}_n - \hat{\beta}_0) + \frac{1}{2n} (\hat{\beta}_n - \hat{\beta}_0)^\top \Psi''_n(\beta_{4n}) (\hat{\beta}_n - \hat{\beta}_0). \]
Note that
\[
\Sigma_n + \frac{1}{n} \Psi'_n(\tilde{\beta}_0) = \frac{1}{n} \sum_{i=1}^{n} [\Sigma_{n1} - g'(\tilde{\beta}_0, X_i)g'(\tilde{\beta}_0, X_i)^\top] - \frac{1}{n} \sum_{i=1}^{n} [\Sigma_{n2} - [Y_i - g(\tilde{\beta}_0, X_i)]g''(\tilde{\beta}_0, X_i)].
\]

Following the same arguments in Proposition 1, we obtain that \( \Sigma_n + \frac{1}{n} \Psi'_n(\tilde{\beta}_0) = \frac{p}{\sqrt{n}} O_p(1) \) and \( \Psi''_n(\beta_{4n}) = n \sqrt{p^5} O_p(1) \). Since \( \|\tilde{\beta}_n - \tilde{\beta}_0\| = O_p(\sqrt{p/n}) \), it follows that
\[
\Sigma_n(\tilde{\beta}_n - \tilde{\beta}_0) = \frac{1}{n} \Psi_n(\tilde{\beta}_0) + \frac{\sqrt{p^3}}{n} O_p(1) + \frac{\sqrt{p^5}}{n} O_p(1).
\]

Because \( \Sigma_n^{-1} O_p(1) = O_p(1) \), the result follows. Indeed, \( \|\Sigma_n^{-1} O_p(1)\|^2 = O_p(1)^\top \Sigma_n^{-2} O_p(1) \leq \lambda_{\text{max}}(\Sigma_n^{-2}) \|O_p(1)\|^2 = O_p(1). \)
\[
\square
\]

If \( g(X, \beta) = \beta^\top X \), it is easy to see that \( \Psi''_n(\beta_{4n}) = 0 \). Consequently,
\[
\Sigma_n(\tilde{\beta}_n - \tilde{\beta}_0) = \frac{1}{n} \Psi_n(\tilde{\beta}_0) + \frac{\sqrt{p^3}}{n} O_p(1).
\]

Therefore only the convergence rate \( p^3/n \to 0 \) is needed to obtain the result in Proposition 2.

**Proof of Proposition 3.** (1) Suppose that \( M_n \beta_i = \lambda_i \beta_i \) and \( \hat{M}_n \beta_i = \hat{\lambda}_i \hat{\beta}_i \) for \( 1 \leq i \leq p \).

Similar to the arguments of Theorem 2.2 in Zhu and Fang (1996), we have
\[
\sqrt{n}(\hat{\lambda}_i - \lambda_i) = \sqrt{n} \beta_i^\top (\hat{M}_n - M_n) \beta_i + o_p(1).
\]

By Theorem 3 in Zhu et al. (2010b), we obtain that \( \sqrt{n} \beta_i^\top (\hat{M}_n - M_n) \beta_i \) is asymptotically normal. Thus \( \hat{\lambda}_i - \lambda_i = O_p(1/\sqrt{n}) \). Following the arguments of Lemma 1 in Tan et al. (2017), we obtain \( P(\hat{q} = 1) \to 1 \). Again, by Theorem 2.2 in Zhu and Fang (1996), we obtain
\[
\sqrt{n}(\hat{\beta}_1 - \beta_1) = \sqrt{n} \sum_{i=2}^{p} \frac{\beta_i \beta_i^\top (\hat{M}_n - M_n) \beta_1}{\lambda_1 - \lambda_i} + o_p(1).
\]

Note that \( \hat{B}_n(1) = \hat{\beta}_1 \) and \( \beta_1 = \kappa \beta_0 \) under \( H_0 \). Then we have
\[
\sqrt{n}(\hat{B}_n(1) - \kappa \beta_0) = \sqrt{n} \beta_1^\top (\hat{M}_n - M_n) \beta_1 \sum_{i=2}^{p} \frac{\beta_i}{\lambda_1 - \lambda_i} + o_p(1).
\]

Since \( \sqrt{n} \beta_1^\top (\hat{M}_n - M_n) \beta_1 = O_p(1) \) and \( \|\sum_{i=2}^{p} \frac{\beta_i}{\lambda_1 - \lambda_i}\| = O(p) \), it follows that \( \|\sqrt{n}(\hat{B}_n(q) - \kappa \beta_0)\| = O_p(\sqrt{p}) \).
\[
40
\]
Note that $q$ is free of $n$ under $H_1$. The proof is concluded from the argument for proving (1). \hfill \Box

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

$$V_n(\hat{\alpha}, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(\hat{\beta}_n^T X_i, \theta_n)] I(\hat{B}_n^T X_i \leq u)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(\hat{\beta}_n^T X_i, \theta_n)] I(\kappa \beta_0^T X_i \leq u) +$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(\hat{\beta}_n^T X_i, \theta_n)] [I(\hat{B}_n^T X_i \leq u) - I(\kappa \beta_0^T X_i \leq u)]$$

$$= V_{n1} + V_{n2}.$$

Let $\gamma = (\beta^T, \theta^T)^T$. Then we obtain that

$$V_{n1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i I(\kappa \beta_0^T X_i \leq u) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [g(\hat{\beta}_n^T X_i, \theta_n) - g(\beta_0^T X_i, \theta_0)] I(\kappa \beta_0^T X_i \leq u)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i I(\kappa \beta_0^T X_i \leq u) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\gamma}_n - \gamma_0)^T g'(\beta_0, \theta_0, X_i) I(\kappa \beta_0^T X_i \leq u) -$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\gamma}_n - \gamma_0)^T g''(\beta_1, \theta_1, X_i)(\hat{\gamma}_n - \gamma_0) I(\kappa \beta_0^T X_i \leq u)$$

$$= V_{n11} - V_{n12} - V_{n13}$$

where $(\beta_1, \theta_1)$ lies between $(\hat{\beta}, \hat{\theta})$ and $(\beta_0, \theta_0)$. For the third term $V_{n13}$ in $V_{n1}$, note that

$$E \sup_u \left\| \sum_{i=1}^{n} g''(\beta_1, \theta_1, X_i) I(\kappa \beta_0^T X_i \leq u) \right\| \leq \sum_{i=1}^{n} E \sup_u \left\| g''(\beta_1, \theta_1, X_i) I(\kappa \beta_0^T X_i \leq u) \right\|$$

$$\leq \sum_{i=1}^{n} \left[ E \sup_u \left\| g''(\beta_1, \theta_1, X_i) I(\kappa \beta_0^T X_i \leq u) \right\|^2 \right]^{1/2}$$

$$\leq \sum_{i=1}^{n} \left( \sum_{j,k=1}^{p+d} E g''_{jk}(\beta_1, \theta_1, X_i)^2 \right)^{1/2}$$

$$\leq Cn(p + d).$$

41
Therefore \( V_{n13} = \frac{1}{\sqrt{n}} \sum_n (p + d) o_p(1) = o_p(1) \) uniformly in \( u \). For \( V_{n12} \), recall that \( M_n(u) = E[g'(\beta_0, \theta_0, X) I(\kappa \beta_0^\top X \leq u)] \). Then we decompose \( V_{n12} \) as follows

\[
V_{n12} = \sqrt{n} (\hat{\gamma}_n - \gamma_0)^\top M_n(u) + \sqrt{n} (\hat{\gamma}_n - \gamma_0)^\top \left( \frac{1}{n} \sum_{i=1}^{n} g'(\beta_0, \theta_0, X_i) I(\kappa \beta_0^\top X_i \leq u) - M_n(u) \right).
\]

For the second term in \( V_{n12} \), by Lemma 3, we have

\[
\sup_u \| \frac{1}{n} \sum_{i=1}^{n} g'(\beta_0, \theta_0, X_i) I(\kappa \beta_0^\top X_i \leq u) - M_n(u) \| = o_p(\sqrt{\frac{p^{3/2} \log n}{n}}).
\]

Conclude that

\[
\sqrt{n} (\hat{\gamma}_n - \gamma_0)^\top \left( \frac{1}{n} \sum_{i=1}^{n} g'(\beta_0, \theta_0, X_i) I(\kappa \beta_0^\top X_i \leq u) - M_n(u) \right) = \sqrt{p} \sqrt{\frac{p^{3/2} \log n}{n}} o_p(1) = o_p(1).
\]

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

\[
V_{n12} = M_n(u)^\top \Sigma_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(\beta_0^\top X_i, \theta_0)] g'(\beta_0, \theta_0, X_i) + o_p(1).
\]

Therefore, we obtain that

\[
V_{n1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i I(\kappa \beta_0^\top X_i \leq u) - \frac{1}{\sqrt{n}} M_n(u)^\top \Sigma_n^{-1} \sum_{i=1}^{n} \varepsilon_i g'(\beta_0, \theta_0, X_i) + o_p(1). \tag{6.1}
\]

Now we consider the term \( V_{n2} \). It can be decomposed as follow

\[
V_{n2} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i [I(\hat{B}_n^\top X_i \leq u) - I(\kappa \beta_0^\top X_i \leq u)] - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [g(\hat{\beta}_n^\top X_i, \hat{\theta}_n) - g(\beta_0^\top X_i, \theta_0)] [I(\hat{B}_n^\top X_i \leq u) - I(\kappa \beta_0^\top X_i \leq u)]
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i [I(\hat{B}_n^\top X_i \leq u) - I(\kappa \beta_0^\top X_i \leq u)] - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\gamma}_n - \gamma_0)^\top g'(\beta_0, \theta_0, X_i) [I(\hat{B}_n^\top X_i \leq u) - I(\kappa \beta_0^\top X_i \leq u)] - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\gamma}_n - \gamma_0)^\top g''(\beta_0, \theta_0, X_i) (\hat{\gamma}_n - \gamma_0) [I(\hat{B}_n^\top X_i \leq u) - I(\kappa \beta_0^\top X_i \leq u)]
\]

\[
= V_{n21} - V_{n22} - V_{n23}.
\]
By Lemma 6, we obtain that $V_{n21} = o_p(1)$ uniformly in $u$. For the second term $V_{n22}$, let

$$\tilde{M}_n(\beta, u) = E\{g'(\beta_0, \theta_0, X)[I(\beta^\top X_i \leq u) - I(\kappa\beta_0 X_i \leq u)]\}. $$

By Lemma 7, we have

$$\sup_u \left\| \frac{1}{n} \sum_{i=1}^{n} g'(\beta_0, \theta_0, X_i) [I(\tilde{B}_n^\top X_i \leq u) - I(\kappa\beta_0 X_i \leq u)] - \tilde{M}_n(\hat{B}_n, u) \right\| = o_p(1).$$

Therefore, we derive that

$$V_{n22} = \sqrt{n}(\hat{\gamma}_n - \gamma_0)^\top \tilde{M}_n(\hat{B}_n, u) + \sqrt{\frac{p^{5/2} \log n}{n}} o_p(1).$$

Let $M_n(\beta, u) = E[g'(\beta, \theta_0, X)I(\beta^\top X_i \leq u)]$. By condition (B1), it is easy to see that

$$\tilde{M}_n(\hat{B}_n, u) = M_n(\hat{B}_n, u) - M_n(\kappa\beta_0, u) = \frac{p^{7/8} \log n}{n^{3/8}} o_p(1).$$

Consequently,

$$V_{n22} = \frac{p^{11/8} \log n}{n^{3/8}} o_p(1) + \sqrt{\frac{p^{5/2} \log n}{n}} o_p(1).$$

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

Similar to the term $V_{n13}$, we obtain that $V_{n23} = o_p(1)$ uniformly in $u$. Combining these with (6.1), we obtain that

$$V_n(\hat{\alpha}, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i I(\kappa\beta_0^\top X_i \leq u) - \frac{1}{\sqrt{n}} M_n(u)^\top \Sigma_n^{-1} \sum_{i=1}^{n} \varepsilon_i g'(\beta_0, \theta_0, X_i) + o_p(1). \quad (6.2)$$

It is easy to see that the first and second terms of the right-hand side of (6.2) 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

$$Y_{ni}(u) = \frac{1}{\sqrt{n}} \varepsilon_i [I(\kappa\beta_0^\top X_i \leq u) - M_n(u)^\top \Sigma_n^{-1} g'(\beta_0, \theta_0, X_i)].$$
For any \( \delta > 0 \), we have

\[
\sum_{i=1}^{n} E\|Y_{ni}\|^2 I(\|Y_{ni}\| > \delta) = nE\{|Y_{ni}|^2 I(\|Y_{ni}\| > \delta)\} \leq n\{E\|Y_{ni}\|^4\}^{1/2}\{P(\|Y_{ni}\| > \delta)\}^{1/2}.
\]

Since

\[
P(\|Y_{ni}\| > \delta) = P(Y_{n1}(u_1)^2 + \ldots + Y_{n1}(u_m)^2 > \delta^2) \leq \sum_{j=1}^{m} P(Y_{n1}(u_j)^2 > \frac{\delta^2}{m}),
\]

and

\[
P(Y_{n1}(u)^2 > \frac{\delta^2}{m}) = P(\varepsilon_1^2[I(\kappa\beta_0^\top X_1 \leq u) - M_n(u)\Sigma^{-1}_n g'(\beta_0, \theta_0, X_1)]^2 > \frac{n\delta^2}{m}) \leq \frac{2mE\varepsilon_1^2}{n\delta^2} + \frac{2m\lambda^2_{\max}(\Sigma^{-1})\|M_n(u)\|E\{\varepsilon_1^2\|g'(\beta_0, \theta_0, X_1)\|\}^2}{n\delta^2},
\]

it follows that \( P(\|Y_{ni}\| > \delta) = O(p/n) \). For \( E\|Y_{ni}\|^4 \), it is easy to see that

\[
E\|Y_{ni}\|^4 \leq m[EY_{n1}(u_1)^4 + \ldots + EY_{n1}(u_m)^4].
\]

Since

\[
EY_{n1}(u)^4 = \frac{1}{n^2} E\{\varepsilon_1^4[I(\kappa\beta_0^\top X_1 \leq u) - M_n(u)\Sigma^{-1}_n g'(\beta_0, \theta_0, X_1)]^4\}
\]

\[
\leq \frac{8}{n^2} \{E[\varepsilon_1^4 I(\kappa\beta_0^\top X_1 \leq u)] + E[\varepsilon_1 M_n(u)\Sigma^{-1}_n g'(\beta_0, \theta_0, X_1)]^4\}
\]

\[
\leq \frac{8}{n^2} \{E[\varepsilon_1^4 I(\kappa\beta_0^\top X_1 \leq u)] + \lambda^4_{\max}(\Sigma^{-1})\|M_n(u)\|^4 E[\varepsilon_1^4 \|g'(\beta_0, \theta_0, X_1)\|]\}
\]

\[
\leq \frac{8}{n^2} \{E\varepsilon_1^4 + \lambda^4_{\max}(\Sigma^{-1})\|M_n(u)\|^4 \sum_{j,k=1}^{p+d} E[\varepsilon_1^4 g_j'(\beta_0, \theta_0, X_1)^2 g_k'(\beta_0, \theta_0, X_1)^2]\},
\]

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

For the covariance matrix \( \sum_{i=1}^{n} Cov(Y_{ni}) \), we only need to consider \( \sum_{i=1}^{n} Cov\{Y_{ni}(s), Y_{ni}(t)\} \).
It is easy to see that
\[ \sum_{i=1}^{n} \text{Cov}\{Y_{ni}(s), Y_{ni}(t)\} \]
\[ = E[\epsilon_1^2 I(\kappa \beta_0^\top X_1 \leq s \land t)] - M_n(s)^\top \Sigma_n^{-1} E[\epsilon_1^2 g'(\beta_0, \theta_0, X_1) I(\kappa \beta_0^\top X_1 \leq s)] \]
\[ - M_n(t)^\top \Sigma_n^{-1} E[\epsilon_1^2 g'(\beta_0, \theta_0, X_1) I(\kappa \beta_0^\top X_1 \leq s)] \]
\[ + M_n(s)^\top \Sigma_n^{-1} E[\epsilon_1^2 g'(\beta_0, \theta_0, X_1) g'(\beta_0, \theta_0, X_1)^\top] \Sigma_n^{-1} M_n(t). \]

Thus \( \sum_{i=1}^{n} \text{Cov}\{Y_{ni}(s), Y_{ni}(t)\} = K_n(s, t). \) Since \( K_n(s, t) \to K(s, t) \), it follows that \( Y_{ni} \) satisfies the conditions of Lindeberg-Feller Central limit theorem. Hence convergence of the finite-dimensional distributions holds. All together we have
\[ V_n(u) \to |V_1^1(u)| \]
where \( V_1^1(u) \) is a zero mean Gaussian process with covariance function \( K(s, t) \). Hence we complete the proof. \( \square \)

**Proof of Theorem 3.2.** Similar to the proof for Theorem 3.1, we only need to work on the event \( \{\hat{q} = 1\} \). Let
\[ V_n^1(\hat{\alpha}, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(\beta_0^\top X_i, \theta_0)] I(\hat{\alpha}^\top \hat{B}_n^\top X_i \leq u). \]
On the event \( \{\hat{q} = 1\} \), we have \( S_{\hat{q}}^+ = \{1\} \) and then \( \hat{\alpha} = 1 \). Consequently, \( V_n^1(\hat{\alpha}, u) \) can be rewritten as
\[ V_n^1(\hat{\alpha}, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(\beta_0^\top X_i, \theta_0)] I(\hat{B}_n^\top X_i \leq u). \]

Next we divide the whole proof of Theorem 3.2 into three parts.

(I) First, to prove that \( \hat{T}_n V_n(\hat{\alpha}, u) - \hat{T}_n V_n^1(\hat{\alpha}, u) = o_p(1) \) uniformly in \( u \). Recall that
\[ \hat{T}_n V_n(\hat{\alpha}, u) = V_n(\hat{\alpha}, u) - \int_{-\infty}^{u} \hat{a}_n(z)^\top \hat{A}_{n}^{-1}(z) \left( \int_{z}^{\infty} \hat{a}_n(v) V_n(\hat{\alpha}, dv) \right) \hat{\sigma}^2_n(z) F_{\hat{\alpha}}(dz) \]
\[ \hat{T}_n V_n^1(\hat{\alpha}, u) = V_n^1(\hat{\alpha}, u) - \int_{-\infty}^{u} \hat{a}_n(z)^\top \hat{A}_{n}^{-1}(z) \left( \int_{z}^{\infty} \hat{a}_n(v) V_n^1(\hat{\alpha}, dv) \right) \hat{\sigma}^2_n(z) F_{\hat{\alpha}}(dz). \]
Since

$$V_n(\hat{\alpha}, u) - V_n^1(\hat{\alpha}, u) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [g(\hat{\beta}_n^T X_i, \hat{\theta}_n) - g(\beta_i^T X_i, \theta_0)] I(\hat{B}_n^T X_i \leq u),$$

by the same arguments in the proof of Theorem 3.3, we obtain that

$$V_n(\hat{\alpha}, u) - V_n^1(\hat{\alpha}, u) = -\sqrt{n}(\hat{\gamma}_n - \gamma_0)^T M_n(u) + o_p(1)$$

uniformly in $u$. The two integrals in $\hat{T}_n V_n(\hat{\alpha}, u)$ and $\hat{T}_n V_n^1(\hat{\alpha}, u)$ differ by

$$\int_{-\infty}^{u} \hat{a}_n(z)^T \hat{A}_n^{-1}(z) \left( \int_{z}^{\infty} \hat{a}_n(v)(V_n^1(\hat{\alpha}, dv) - V_n(\hat{\alpha}, dv)) \right) \hat{\sigma}_n^2(z) F_{\hat{\alpha}}(dz).$$

It equals

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{-\infty}^{u} \hat{a}_n(z)^T \hat{A}_n^{-1}(z) \hat{\sigma}_n^2(z) I(\hat{B}_n^T X_i \geq z) \hat{a}_n(\hat{B}_n^T X_i) [g(\hat{\beta}_n^T X_i, \hat{\theta}_n) - g(\beta_0^T X_i, \theta_0)] F_{\hat{\alpha}}(dz)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{-\infty}^{u} \hat{a}_n(z)^T \hat{A}_n^{-1}(z) \hat{\sigma}_n^2(z) I(\hat{B}_n^T X_i \geq z) \hat{a}_n(\hat{B}_n^T X_i) [g' (\hat{\beta}_n, \hat{\theta}_n, X_i) (\hat{\gamma}_n - \gamma_0) +$$

$$(\hat{\gamma}_n - \gamma_0) \hat{\beta}_n^T X_i + (\hat{\gamma}_n - \gamma_0) \hat{\beta}_n^T X_i + (\hat{\gamma}_n - \gamma_0) \hat{\beta}_n^T X_i + (\hat{\gamma}_n - \gamma_0) \hat{\beta}_n^T X_i + (\hat{\gamma}_n - \gamma_0) \hat{\beta}_n^T X_i] F_{\hat{\alpha}}(dz)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{-\infty}^{u} \hat{a}_n(z)^T \hat{A}_n^{-1}(z) \hat{\sigma}_n^2(z) I(\hat{B}_n^T X_i \geq z) \hat{a}_n(\hat{B}_n^T X_i) [g' (\hat{\beta}_n, \hat{\theta}_n, X_i) (\hat{\gamma}_n - \gamma_0) F_{\hat{\alpha}}(dz) + o_p(1),$$

where $(\hat{\beta}_1, \hat{\theta}_1)$ and $(\hat{\beta}_2, \hat{\theta}_2)$ both lie between $(\hat{\beta}_n, \hat{\theta}_n)$ and $(\beta_0, \theta_0)$. Recall that

$$\hat{A}_n(z) = \frac{1}{n} \sum_{i=1}^{n} \hat{a}_n(\hat{B}_n^T X_i) g' (\hat{\beta}_n, \hat{\theta}_n, X_i) I(\hat{B}_n^T X_i \geq z).$$

Then the two integrals differ by

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0)^T \int_{-\infty}^{u} \hat{a}_n(z) \hat{\sigma}_n^2(z) F_{\hat{\alpha}}(dz) + o_p(1)$$

$$= \sqrt{n}(\hat{\gamma}_n - \gamma_0)^T \int_{-\infty}^{u} a_n(z) \sigma_n^2(z) F_{\kappa \beta_0}(dz) + o_p(1).$$

Since $\int_{-\infty}^{u} a_n(z) \sigma_n^2(z) F_{\kappa \beta_0}(dz) = M_n(u)$, it follows that $\hat{T}_n V_n(\hat{\alpha}, u) - \hat{T}_n V_n^1(\hat{\alpha}, u) = o_p(1)$ uniformly in $u$. 46
(II) Second, to prove $T_n V_n^1(\hat{\alpha}, u) - \hat{T}_n V_n^1(\hat{\alpha}, u) = o_p(1)$ uniformly in $u$. Indeed,

$$ T_n V_n^1(\hat{\alpha}, u) - \hat{T}_n V_n^1(\hat{\alpha}, u) $$

$$ = \int_{-\infty}^{\infty} \hat{a}(z)^T A^{-1}(z) \left( \int_{z}^{\infty} \hat{a}(v) V_n^1(\hat{\alpha}, dv) \right) \sigma_n^2(z) F_\alpha(dz) - $$

$$ \int_{-\infty}^{\infty} a_n(z)^T A_n^{-1}(z) \left( \int_{z}^{\infty} a_n(v) V_n^1(\hat{\alpha}, dv) \right) \sigma_n^2(z) F_\alpha(dz) $$

$$ = \int_{-\infty}^{\infty} \hat{a}(z)^T A^{-1}(z) \left( \int_{z}^{\infty} \hat{a}(v) V_n^1(\hat{\alpha}, dv) \right) \sigma_n^2(z) F_\alpha(dz) - $$

$$ \int_{-\infty}^{\infty} a_n(z)^T A_n^{-1}(z) \left( \int_{z}^{\infty} a_n(v) V_n^1(\hat{\alpha}, dv) \right) \sigma_n^2(z) F_\alpha(dz) + $$

$$ \int_{-\infty}^{\infty} a_n(z)^T A_n^{-1}(z) \left( \int_{z}^{\infty} a_n(v) V_n^1(\hat{\alpha}, dv) \right) \sigma_n^2(z) \{F_\alpha(dz) - F_n(\hat{\beta}_0)(dz)\} $$

$$ =: T_{n1} - T_{n2} + T_{n3}. $$

Putting

$$ h_n(z) = a_n(z)^T A_n^{-1}(z) \left( \int_{z}^{\infty} a_n(v) V_n^1(\hat{\alpha}, dv) \right) \sigma_n^2(z), $$

it follows that

$$ h_n(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i a_n(z)^T A_n^{-1}(z) a_n(\kappa \beta_0^T X_i) I(\kappa \beta_0^T X_i \geq z) \sigma_n^2(z) + o_p(1). $$

By the uniformly boundedness of $\sigma_n^2(z)$, we have the sequence $\{h_n(z)\}$ is asymptotically tight. According to Lemma 3.4 in Stute, Thies, and Zhu (1998) and the arguments thereafter, we obtain that $T_{n3} = o_p(1)$ uniformly in $u \in [-\infty, u_0]$. For $T_{n1} - T_{n2}$, since both $a_n(z)$ and $A_n(z)$ depend on $(\beta_0, \theta_0)$, we rewrite $a_n(z)$ and $A_n(z)$ as $a_n(\beta_0, \theta_0, z)$ and $A_n(\beta_0, \theta_0, z)$ respectively and define

$$ l_n(\beta, \theta, u) = \int_{-\infty}^{\infty} a_n(\beta, \theta, z)^T A_n^{-1}(\beta, \theta, z) \left( \int_{z}^{\infty} a_n(\beta, \theta, v) V_n^1(\hat{\alpha}, dv) \right) \sigma_n^2(z) F_\alpha(dz). $$

By the boundedness of $\sigma_n(u)$ and Condition (B1), we obtain that $l_n(\hat{\beta}_n, \hat{\theta}_n, u) - l_n(\beta_0, \theta_0, u) = o_p(1)$. By Lemma 8, we show that

$$ \sup_v \|\hat{r}_n(v) - r_n(v)\| = O_p\left(\sqrt{p \log n/n^{4/5}}\right). $$
Combining this with the uniformly boundedness of $\hat{\sigma}_n^2$, we obtain $T_{n1} - T_{n2}$ tends to zero in probability.

(III) Finally, to prove $T_nV_n^1(\hat{\alpha}, u) - T_nV_n^0(u) = o_p(1)$ uniformly in $u$.

$$
T_nV_n^1(\hat{\alpha}, u) - T_nV_n^0(u) = V_n^1(\hat{\alpha}, u) - \int_{-\infty}^u a_n(z)^T A_n^{-1}(z) \left( \int_{\infty}^\infty a_n(v) V_n^1(\hat{\alpha}, dv) \right) \psi_n(dz) - \left\{ V_n^0(u) - \int_{-\infty}^u a_n(z)^T A_n^{-1}(z) \left( \int_{\infty}^\infty a_n(v) V_n^0(dv) \right) \psi_n(dz) \right\}.
$$

Since

$$
V_n^1(\hat{\alpha}, u) - V_n^0(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i [I(\hat{B}_n^\top X_i \leq u) - I(\kappa\beta_0^\top X_i \leq u)],
$$

by the same argument in Theorem 3.1 we obtain that $V_n^1(\hat{\alpha}, u) - V_n^0(u) = o_p(1)$ uniformly in $u$. For the integrals in $T_nV_n^1(\hat{\alpha}, u) - T_nV_n^0(u)$, note that the two integrals differ by

$$
\int_{-\infty}^u a_n(z)^T A_n^{-1}(z) \left\{ \int_{\infty}^\infty a_n(v) [V_n^1(\hat{\alpha}, dv) - V_n^0(dv)] \right\} \psi_n(dz)
$$

$$
= \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i \int_{-\infty}^u a_n(z)^T A_n^{-1}(z) a_n(\hat{B}_n^\top X_i) I(\hat{B}_n^\top X_i \geq z) \psi_n(dz) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i \int_{-\infty}^u a_n(z)^T A_n^{-1}(z) a_n(\kappa\beta_0^\top X_i) I(\kappa\beta_0^\top X_i \geq z) \psi_n(dz).
$$

Since $\|\hat{B}_n - \kappa\beta_0\| = O_p(\sqrt{p/n})$, similar to the arguments in Lemma 6, the difference between the two integrals in $T_nV_n^1(\hat{\alpha}, u) - T_nV_n^0(u)$ tends to zero. Hence $T_nV_n^1(\hat{\alpha}, u) - T_nV_n^0(u) = o_p(1)$ uniformly in $u$. All together we conclude that

$$
\sup_{\hat{\alpha} \in \hat{\mathcal{H}}^+} |\hat{T}_n V_n(\hat{\alpha}, u)| \to |V_\infty(u)|
$$
in distribution. \qed

**Proof of Proposition 4.** Let $Y = g(\beta_0^\top X, \theta_0) + \varepsilon$, $\alpha_t = E[XY(Y \leq t)]$, $\hat{\alpha}_t = E[XI(Y_n \leq t)]$, $M_n = \int \alpha_t a_t^\top F_Y(dt)$, and $\hat{M}_n = \int \hat{\alpha}_t \hat{a}_t^\top F_{Y_n}(dt)$. Then the space $\text{span}(M_n) \in \mathcal{S}_{1/X}$ and the space $\text{span}(\hat{M}_n) \in \mathcal{S}_{1/X}$. If we show that $\sqrt{n}\gamma^\top (\hat{M}_n - M_n)\gamma$ is asymptotically normal for any unit
vector $\gamma$, the result of this proposition follows from the exact arguments for proving Proposition 3.

We now prove the above asymptotic normality. Under $H_{1n}$, we have

$$\hat{\alpha}_t = E\{XF_{Y|X}(t - \frac{1}{\sqrt{n}}G(X))\},$$

where $F_{Y|X}$ is the conditional distribution of $Y$ given $X$. By Taylor’s expansion, we derive

$$\hat{\alpha}_t = \alpha_t - \frac{1}{\sqrt{n}} E\{Xf_{Y|X}(t)G(X)\} + \frac{1}{2n} E\{Xf'_{Y|X}(\xi_t(X))G(X)^2\}.$$ 

Here $\xi_t(X)$ lies between $t - \frac{1}{\sqrt{n}} G(X)$ and $t$ and $f_{Y|X}$ is the conditional density function of $Y$ given $X$. Therefore,

$$\hat{M}_n = \int \alpha_t \hat{\alpha}_t^{\top} F_{Y_n}(dt) - \frac{1}{\sqrt{n}} \int \alpha_t E\{X^{\top} f_{Y|X}(t)G(X)\} F_{Y_n}(dt) - \frac{1}{\sqrt{n}} \int \{E[Xf_{Y|X}(t)G(X)]\} \alpha_t^{\top} F_{Y_n}(dt) + O_p(\frac{P}{n}).$$

Note that $F_{Y_n}(t) = F_Y(t) - \frac{1}{\sqrt{n}} E[G(X)f_{Y|X}(t)] + \frac{1}{2n} E[f'_{Y|X}(\xi_t(X))G(X)^2]$. Consequently,

$$\hat{M}_n = M_n - \frac{1}{\sqrt{n}} \int \{E[X^{\top} f_{Y|X}(t)G(X)] + E[Xf_{Y|X}(t)G(X)]\alpha_t^{\top}\} F_Y(dt) + O_p(\frac{P}{n}).$$

By Theorem 3 in Zhu et al. (2010b), we have $\sqrt{n}\gamma^{\top}(\hat{M}_n - M_n)\gamma$ is asymptotically normal. By condition (B3) in Appendix, $\sqrt{n}\gamma^{\top}(\hat{M}_n - M_n)\gamma$ is also asymptotically normal.

**Proof of Proposition 5.** The proof is similar to that for proving Propositions 1 and 2 with $e_i = \varepsilon_i$ and $\Sigma_n = E[g'(\beta_0, \theta_0, X)g'(\beta_0, \theta_0, X)^\top]$. 

**Proof of Theorem 3.3.** (1) Under $H_1$, Proposition 1 asserts that $P(\hat{q} = q) \to 1$. Thus we only need work on the event $\{\hat{q} = q\}$. It follows that $\sup_{\hat{\alpha} \in S_q^+} |\hat{T}_n V_n(\hat{\alpha}, u)| = \sup_{\alpha \in S_q^+} |\hat{T}_n V_n(\alpha, u)|$. 

Putting

$$\hat{V}_n^1(\alpha, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [Y_i - g(\hat{\beta}_0^{\top} X_i, \hat{\theta}_0)] I(\alpha^{\top} \hat{B}_n^{\top} X_i \leq u),$$

$$\hat{V}_n^0(\alpha, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [Y_i - g(\hat{\beta}_0^{\top} X_i, \hat{\theta}_0)] I(\alpha^{\top} B^{\top} X_i \leq u)$$

49
and

\[ r_n(\alpha, z) = E(X|\alpha^T B^TX = z); \]
\[ \sigma_n^2(\alpha, z) = E((Y - g(\tilde{\beta}_0, \tilde{\theta}_0))^2|\alpha^T B^TX = z); \]
\[ a_n(\alpha, z) = \{g_1'(z/\|\tilde{\beta}_0\|, \tilde{\theta}_0)r_n(\alpha, z)^T / \sigma_n^2(\alpha, z), g_2'(z/\|\tilde{\beta}_0\|, \tilde{\theta}_0)^T / \sigma_n^2(\alpha, z)\}; \]
\[ A_n(\alpha, z) = E\{a_n(\alpha, \alpha^T B^TX)g'((\tilde{\beta}_0, \theta_0, X)^T I(\alpha^T B^TX \geq z))\}. \]

Following the arguments in Theorem 3.2, we obtain that

\[ \hat{T}_n V_n(\alpha, u) - T_n \tilde{V}_n^1(\alpha, u) = o_p(1), \]

where

\[ T_n \tilde{V}_n^1(\alpha, u) = \tilde{V}_n^1(\alpha, u) - \int_{-\infty}^{u} a_n(\alpha, z)^T A_n^{-1}(\alpha, z) \left( \int_{z}^{\infty} a_n(\alpha, v)\tilde{V}_n^1(\alpha, dv) \right) \sigma_n^2(\alpha, z) F_\alpha(dz) \]

and \( F_\alpha \) is the cumulative distribution function of \( \alpha^T B^TX \). Consider

\[ T_n \tilde{V}_n^1(\alpha, u) - T_n \tilde{V}_n^0(\alpha, u) \]
\[ = \tilde{V}_n^1(\alpha, u) - \int_{-\infty}^{u} a_n(\alpha, z)^T A_n^{-1}(\alpha, z) \left( \int_{z}^{\infty} a_n(\alpha, v)\tilde{V}_n^1(\alpha, dv) \right) \sigma_n^2(\alpha, z) F_\alpha(dz) \]
\[ \tilde{V}_n^0(\alpha, u) - \int_{-\infty}^{u} a_n(\alpha, z)^T A_n^{-1}(\alpha, z) \left( \int_{z}^{\infty} a_n(\alpha, v)\tilde{V}_n^0(\alpha, dv) \right) \sigma_n^2(\alpha, z) F_\alpha(dz). \]

Since

\[ \frac{1}{\sqrt{n}} (\tilde{V}_n^1(\alpha, u) - \tilde{V}_n^0(\alpha, u)) = \frac{1}{n} \sum_{i=1}^{n} [Y_i - g(\tilde{\beta}_0^TX_i, \tilde{\theta}_0)]\{I(\alpha^T \hat{B}_n^TX_i \leq u) - I(\alpha^T B^TX_i \leq u)\} \]
\[ = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \{I(\alpha^T \hat{B}_n^TX_i \leq u) - I(\alpha^T B^TX_i \leq u)\} + \]
\[ \frac{1}{n} \sum_{i=1}^{n} [G(X_i) - g(\tilde{\beta}_0^TX_i, \tilde{\theta}_0)]\{I(\alpha^T \hat{B}_n^TX_i \leq u) - I(\alpha^T B^TX_i \leq u)\}, \]

it follows that \( \frac{1}{\sqrt{n}} (\tilde{V}_n^1(\alpha, u) - \tilde{V}_n^0(\alpha, u)) = o_p(1) \). For the two integrals in \( T_n \tilde{V}_n^1(\alpha, u) - T_n \tilde{V}_n^0(\alpha, u) \),
we have
\[
\frac{1}{\sqrt{n}} \int_{-\infty}^{u} a_n(\alpha, z) A_n^{-1}(\alpha, z) \left( \int_{z}^{\infty} a_n(\alpha, v) (\tilde{V}_n^1(\alpha, dv) - \hat{V}_n^1(\alpha, dv)) \right) \sigma_n^2(\alpha, z) F_\alpha(dz)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \int_{-\infty}^{u} a_n(\alpha, z) A_n^{-1}(\alpha, z)a_n(\alpha, \alpha^T \hat{B}_n^T X_i) I(\alpha^T \hat{B}_n^T X_i \geq z) \sigma_n^2(\alpha, z) F_\alpha(dz) -
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \int_{-\infty}^{u} a_n(\alpha, z) A_n^{-1}(\alpha, z)a_n(\alpha, \alpha^T B^T X_i) I(\alpha^T B^T X_i \geq z) \sigma_n^2(\alpha, z) F_\alpha(dz) +
\]

\[
\frac{1}{n} \sum_{i=1}^{n} [G(X_i) - g(\beta_0^T X_i, \tilde{\theta}_0)] \int_{-\infty}^{u} a_n(\alpha, z) A_n^{-1}(\alpha, z)a_n(\alpha, \alpha^T \hat{B}_n^T X_i) I(\alpha^T \hat{B}_n^T X_i \geq z) \sigma_n^2(\alpha, z) F_\alpha(dz) -
\]

\[
\frac{1}{n} \sum_{i=1}^{n} [G(X_i) - g(\beta_0^T X_i, \tilde{\theta}_0)] \int_{-\infty}^{u} a_n(\alpha, z) A_n^{-1}(\alpha, z)a_n(\alpha, \alpha^T B^T X_i) I(\alpha^T B^T X_i \geq z) \sigma_n^2(\alpha, z) F_\alpha(dz)
\]

\[
= o_p(1).
\]

Therefore, we obtain that
\[
\frac{1}{\sqrt{n}} (\hat{T}_n V_n(\alpha, u) - T_n \hat{V}_n^0(\alpha, u)) = o_p(1).
\]

Note that
\[
\frac{1}{\sqrt{n}} T_n \hat{V}_n^0(\alpha, u) = \frac{1}{n} \sum_{i=1}^{n} [Y_i - g(\beta_0^T X_i, \tilde{\theta}_0)] I(\alpha^T B^T X_i \leq u) -
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \int_{-\infty}^{u} a_n(\alpha, z) A_n^{-1}(\alpha, z)a_n(\alpha, \alpha^T B^T X_i) I(\alpha^T B^T X_i \geq z) \sigma_n^2(\alpha, z) F_\alpha(dz) -
\]

\[
\frac{1}{n} \sum_{i=1}^{n} [G(X_i) - g(\beta_0^T X_i, \tilde{\theta}_0)] \int_{-\infty}^{u} a_n(\alpha, z) A_n^{-1}(\alpha, z)a_n(\alpha, \alpha^T B^T X_i) \times
\]

\[
I(\alpha^T B^T X_i \geq z) \sigma_n^2(\alpha, z) F_\alpha(dz).
\]

It follows that
\[
\frac{1}{\sqrt{n}} T_n \hat{V}_n^0(\alpha, u) = \hat{G}_{1n}(\alpha, u) - \hat{G}_{2n}(\alpha, u) + o_p(1) \longrightarrow \hat{G}_1(\alpha, u) - \hat{G}_2(\alpha, u)
\]

where
\[
\hat{G}_{1n}(\alpha, u) = E\{[G(X) - g(\beta_0^T X, \tilde{\theta}_0)] I(\alpha^T B^T X \leq u)\}
\]

\[
\hat{G}_{2n}(\alpha, u) = E\{[G(X) - g(\beta_0^T X, \tilde{\theta}_0)] \int_{-\infty}^{u} a_n(\alpha, z) A_n^{-1}(\alpha, z)a_n(\alpha, \alpha^T B^T X) \times
\]

\[
I(\alpha^T B^T X \geq z) \sigma_n^2(\alpha, z) F_\alpha(dz)\}.
\]

51
Therefore, we obtain that
\[
\frac{1}{\sqrt{n}} \sup_{\hat{\alpha} \in \mathcal{S}^+_q} |\hat{T}_n V_n(\hat{\alpha}, u)| \rightarrow |L(u)|
\]
where \(L(u)\) is an nonzero function.

(2) We use the same notations as in the arguments of Theorem \(3.2\). Under the local alternatives \((3.3)\), by Proposition 3, we have \(P\{\hat{q} = 1\} \rightarrow 1\). Thus we just work on this event \(\{\hat{q} = 1\}\). Hence \(\mathcal{S}^+_q = \{1\}\) and \(\sup_{\hat{\alpha} \in \mathcal{S}^+_q} |\hat{T}_n V_n(\hat{\alpha}, u)| = |\hat{T}_n V_n(\hat{\alpha}, u)|\).

Following the same arguments for Theorem \(3.2\) we obtain that
\[
\hat{T}_n V_n(\hat{\alpha}, u) - T_n V^1_n(\hat{\alpha}, u) = o_p(1)
\]
Next, we consider \(T_n V^1_n(\hat{\alpha}, u) - T_n V^0_n(u)\). Recall that
\[
V^0_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(\beta_0^T X_i, \theta_0)] I(\kappa \beta_0^T X_i \leq u),
\]
\[
V^1_n(\hat{\alpha}, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(\hat{\beta}_0^T X_i, \theta_0)] I(\hat{\beta}_n^T X_i \leq u).
\]
Under \(H_{1n}\), we have
\[
V^1_n(\hat{\alpha}, u) - V^0_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i [I(\hat{\beta}_n^T X_i \leq u) - I(\kappa \beta_0^T X_i \leq u)] + \frac{1}{n} \sum_{i=1}^{n} G(X_i) [I(\hat{\beta}_n^T X_i \leq u) - I(\kappa \beta_0^T X_i \leq u)].
\]
Then \( V_n^1(\hat{\alpha}, u) - V_n^0(u) = o_p(1) \). For the integrals in \( T_nV_n^1(\hat{\alpha}, u) - T_nV_n^0(u) \), since

\[
\int_{-\infty}^{u} a_n(z)^T A_n^{-1}(z) \left\{ \int_{z}^{\infty} a_n(v)[V_n^1(\hat{\alpha}, dv) - V_n^0(dv)] \right\} \psi_n(dz) = 1
\]

by the same arguments for Theorem 3.2, we have

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \int_{-\infty}^{u} a_n(z)^T A_n^{-1}(z)a_n(\hat{\beta}_0^T X_i)I(\hat{\beta}_0^T X_i \geq z)\psi_n(dz) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \int_{-\infty}^{u} a_n(z)^T A_n^{-1}(z)a_n(\kappa \beta_0^T X_i)I(\kappa \beta_0^T X_i \geq z)\psi_n(dz) + \frac{1}{n} \sum_{i=1}^{n} G(X_i) \int_{-\infty}^{u} a_n(z)^T A_n^{-1}(z)a_n(\hat{\beta}_0^T X_i)I(\hat{\beta}_0^T X_i \geq z)\psi_n(dz) - \frac{1}{n} \sum_{i=1}^{n} G(X_i) \int_{-\infty}^{u} a_n(z)^T A_n^{-1}(z)a_n(\kappa \beta_0^T X_i)I(\kappa \beta_0^T X_i \geq z)\psi_n(dz),
\]

Hence we obtain that \( T_nV_n^1(\hat{\alpha}, u) - T_nV_n^0(u) = o_p(1) \).

To complete the proof, it remains to derive the asymptotic distribution of \( T_nV_n^0(u) \). Under the alternatives, note that

\[
V_n^0(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i I(\kappa \beta_0^T X_i \leq u) + \frac{1}{n} \sum_{i=1}^{n} G(X_i)I(\kappa \beta_0^T X_i \leq u).
\]

It follows that

\[
T_nV_n^0(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i I(\kappa \beta_0^T X_i \leq u) + \frac{1}{n} \sum_{i=1}^{n} G(X_i)I(\kappa \beta_0^T X_i \leq u) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \int_{-\infty}^{u} a_n(z)^T A_n^{-1}(z)a_n(\kappa \beta_0^T X_i)I(\kappa \beta_0^T X_i \geq z)\psi_n(dz) - \frac{1}{n} \sum_{i=1}^{n} G(X_i) \int_{-\infty}^{u} a_n(z)^T A_n^{-1}(z)a_n(\kappa \beta_0^T X_i)I(\kappa \beta_0^T X_i \geq z)\psi_n(dz).
\]
By Glivenko-Cantelli Theorem, we have

\[ \frac{1}{n} \sum_{i=1}^{n} G(X_i) I(\kappa_0 \beta^T X_i \leq u) = E[G(X) I(\kappa_0 \beta^T X \leq u)] + o_p(1), \]

\[ \frac{1}{n} \sum_{i=1}^{n} G(X_i) \int_{-\infty}^{u} a_n(z)^T A_n^{-1}(z) a_n(\kappa \beta_0^T X_i) I(\kappa \beta_0^T X_i \geq z) \psi_n(dz) \]

\[ = E \left( G(X) \int_{-\infty}^{u} a_n(z)^T A_n^{-1}(z) a_n(\kappa \beta_0^T X) I(\kappa \beta_0^T X \geq z) \psi_n(dz) \right) + o_p(1). \]

Since \( E[G(X) I(\kappa_0 \beta^T X \leq u)] \to G_1(u) \) and

\[ E \left( G(X) \int_{-\infty}^{u} a_n(z)^T A_n^{-1}(z) a_n(\kappa \beta_0^T X) I(\kappa \beta_0^T X \geq z) \psi_n(dz) \right) \to G_2(u), \]

we conclude that

\[ T_n V_0^n(u) \longrightarrow V_\infty(u) + G_1(u) - G_2(u) \quad \text{in distribution}, \]

where \( V_\infty(u) \) is a zero-mean Gaussian process given by (3.6). \( \square \)
References

[1] Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, 20, 105-134.

[2] Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica*, 58, 1443-1458.

[3] Cook, R. D. (1998). *Regression Graphics: Ideas for Studying Regressions Through Graphics*. New York: Wiley.

[4] Cook, R. D. and Weisberg, S. (1991). Discussion of Sliced inverse regression for dimension reduction, by K. C. Li. *Journal of the American Statistical Association*, 86, 316-342.

[5] Cook, R. D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.*, 30, 455 – 474.

[6] Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimates. *The Annals of Statistics*. 27, 1012-1050.

[7] Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory*, 22, 1030-1051.

[8] Fan, J. Q. and Huang, L. S. (2001). Goodness-of-fit tests for parametric regression models, *Journal of the American Statistical Association*, 96, 640-652.

[9] Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*, 32, 928-961.

[10] Fan, Y. and Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica*, 64, 865-890.
[11] Guo, X., Wang, T. and Zhu, L. X. (2016). Model checking for generalized linear models: a dimension-reduction model-adaptive approach. *Journal of the Royal Statistical Society: Series B*, 78, 1013-1035.

[12] Hall, P. and Li, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. *The annals of Statistics*, 867-889.

[13] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21, 1926-1947.

[14] Huber, P. J. (1973) Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 799-821.

[15] Khmaladze, E V. (1982). Martingale Approach in the Theory of Goodness-of-fit Tests. *Theory of Probability & Its Applications*, 26, 240-257.

[16] Khmaladze, E V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *The Annals of Statistics*, 32, 995-1034.

[17] Koul, H. L. and Ni, P. P. (2004). Minimum distance regression model checking. *Journal of Statistical Planning and Inference*, 119, 109-141.

[18] Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40, 633-643.

[19] Lavergne, P. and Patilea, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *Journal of Econometrics*, 143, 103-122.

[20] Lavergne, P. and Patilea, V. (2012). One for all and all for one: regression checks with many regressors. *Journal of Business & Economic Statistics*, 30, 41-52.

[21] Li, K. C. (1991). Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, 86, 316-327.
[22] Li, B. and Wang, S. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association*, **102**, 997-1008.

[23] Ortega, J. M., and Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in several variables*. Society for Industrial and Applied Mathematics.

[24] Pollard, D. (1984). *Convergence of stochastic processes*. Springer Science and Business Media.

[25] Portnoy S. (1984). Asymptotic behavior of M-estimators of p regression parameters when \( p^2/n \) is large. I. Consistency. *The Annals of Statistics*, 1298-1309.

[26] Portnoy, S. (1985). Asymptotic behavior of M estimators of p regression parameters when \( p^2/n \) is large; II. Normal approximation. *The Annals of Statistics*, 1403-1417.

[27] Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*. **25**, 613-641.

[28] Stute, W., González-Manteiga, W. and Presedo-Quindimil, M. (1998). Bootstrap approximation in model checks for regression. *Journal of the American Statistical Association*. **93**, 141-149.

[29] Stute, W., Thies, S. and Zhu, L. X. (1998). Model checks for regression: An innovation process approach. *The Annals of Statistics*. **26**, 1916-1934.

[30] Stute, W., Xu, W. L. and Zhu, L. X. (2008). Model diagnosis for parametric regression in high dimensional spaces. *Biometrika*. **95**, 1-17.

[31] Stute, W. and Zhu, L. X. (2002). Model checks for generalized linear models. *Scandinavian Journal of Statistics*. **29**, 535-545.

[32] Tan, F. L., Zhu, X. H, and Zhu, L. X. (2017). A projection-based adaptive-to-model test for regressions. *Statistica Sinica*, accepted.
[33] Van Keilegom, I., Gonz´ ales-Manteiga, W. and S´ anchez Sellero, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST*, **17**, 401-415.

[34] Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer New York.

[35] White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association, 76*, 419-433.

[36] Xia, Y. C., Tong, H., Li, W. K. and Zhu, L. X. (2002). An adaptive estimation of dimension reduction space (with discussion). *Journal of the Royal Statistical Society: Series B, 64*, 363-410.

[37] Xia, Q., Xu, W. L. and Zhu, L. X. (2015). Consistently determining the number of factors in multivariate volatility modelling. *Statistica Sinica, 25*, 1025-1044.

[38] Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics, 75*, 263-289.

[39] Zhu, L. X. (1993). Convergence rates of empirical processes indexed by the classes of functions and its application. *Journal of System Science and Mathematical Science Chinese Series, 13*, 33-41.

[40] Zhu, L. X. and Fang, K. T. (1996). Asymptotics for the kernel estimates of sliced inverse regression. *Annals of Statistics, 24*, 1053-1067.

[41] Zhu, L. X., Miao, B. Q. and Peng, H. (2006). On Sliced Inverse Regression with High Dimensional Covariates. *Journal of the American Statistical Association, 101*, 630-643.

[42] Zhu, L. P., Zhu, L. X. , Ferré, L. and Wang, T. (2010a). Sufficient dimension reduction through discretization-expectation estimation. *Biometrika, 97*, 295-304.
[43] Zhu, L. P., Zhu, L. X. and Feng, Z. H. (2010b). Dimension reduction in regressions through cumulative slicing estimation. *Journal of the American Statistical Association*, 105, 1455-1466.
Table 1: Empirical sizes and powers of $ACM_n^2$, $T_n^{SZ}$, $PCvM_n$, $ICM_n$, $T_n^{ZH}$ and $T_n^{GWZ}$ for $H_0$ vs. $H_{11}$ in Study 1.

|             | n=100 | n=200 | n=400 | n=800 |
|-------------|-------|-------|-------|-------|
| $ACM_n^2$, $\alpha = 0.10$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0970| 0.0905| 0.0890| 0.1020|
| $p=10$      | 0.5   | 0.8650| 0.9915| 1.0000| 1.0000|
| $ACM_n^2$, $\alpha = 0.05$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0500| 0.0530| 0.0500| 0.0505|
| $p=10$      | 0.5   | 0.7770| 0.9810| 1.0000| 1.0000|
| $ACM_n^2$, $\alpha = 0.01$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0085| 0.0105| 0.0115| 0.0130|
| $p=10$      | 0.5   | 0.5620| 0.9095| 0.9975| 1.0000|
| $T_n^{SZ}$, $\alpha = 0.10$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0915| 0.0995| 0.1060| 0.0985|
| $p=10$      | 0.5   | 0.8675| 0.9865| 1.0000| 1.0000|
| $T_n^{SZ}$, $\alpha = 0.05$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0510| 0.0470| 0.0420| 0.0495|
| $p=10$      | 0.5   | 0.7825| 0.9795| 1.0000| 1.0000|
| $T_n^{SZ}$, $\alpha = 0.01$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0120| 0.0090| 0.0120| 0.0100|
| $p=10$      | 0.5   | 0.5290| 0.9065| 0.9990| 1.0000|
| $PCvM_n$, $\alpha = 0.10$ |       |       |       |       |
| $p=7$       | 0.0   | 0.1140| 0.1220| 0.0980| 0.1190|
| $p=10$      | 0.5   | 0.8850| 0.9880| 1.0000| 1.0000|
| $PCvM_n$, $\alpha = 0.05$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0480| 0.0590| 0.0650| 0.0490|
| $p=10$      | 0.5   | 0.8110| 0.9860| 1.0000| 1.0000|
| $PCvM_n$, $\alpha = 0.01$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0150| 0.0100| 0.0110| 0.0090|
| $p=10$      | 0.5   | 0.6190| 0.9310| 0.9970| 1.0000|
| $ICM_n$, $\alpha = 0.10$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0390| 0.0810| 0.0000| 0.0000|
| $p=10$      | 0.5   | 0.5490| 0.2910| 0.1760| 0.0000|
| $ICM_n$, $\alpha = 0.05$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0070| 0.0000| 0.0000| 0.0000|
| $p=10$      | 0.5   | 0.3900| 0.0910| 0.0180| 0.0000|
| $ICM_n$, $\alpha = 0.01$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0000| 0.0000| 0.0000| 0.0000|
| $p=10$      | 0.5   | 0.0000| 0.0000| 0.0000| 0.0000|
| $T_n^{ZH}$, $\alpha = 0.10$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0805| 0.0950| 0.1055| 0.1060|
| $p=10$      | 0.5   | 0.2240| 0.2205| 0.2420| 0.2430|
| $T_n^{ZH}$, $\alpha = 0.05$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0305| 0.0300| 0.0330| 0.0310|
| $p=10$      | 0.5   | 0.1460| 0.1285| 0.1445| 0.0980|
| $T_n^{ZH}$, $\alpha = 0.01$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0015| 0.0020| 0.0025| 0.0025|
| $p=10$      | 0.5   | 0.0420| 0.0210| 0.0225| 0.0150|
| $T_n^{GWZ}$, $\alpha = 0.10$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0710| 0.0755| 0.0850| 0.0830|
| $p=10$      | 0.5   | 0.8170| 0.9795| 1.0000| 1.0000|
| $T_n^{GWZ}$, $\alpha = 0.05$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0525| 0.0430| 0.0585| 0.0475|
| $p=10$      | 0.5   | 0.7690| 0.9690| 1.0000| 1.0000|
| $T_n^{GWZ}$, $\alpha = 0.01$ |       |       |       |       |
| $p=7$       | 0.0   | 0.0220| 0.0170| 0.0205| 0.0170|
| $p=10$      | 0.5   | 0.6510| 0.9455| 0.9995| 1.0000|
Table 2: Empirical sizes and powers of $ACM^2_n$, $T^{SZ}_n$, $PCvM_n$, $ICM_n$, $T^{ZH}_n$ and $T^{GWZ}_n$ for $H_0$ vs. $H_{12}$ in Study 1.

|                | $a$ | $n=100$ | $n=200$ | $n=400$ | $n=800$ |
|----------------|-----|---------|---------|---------|---------|
| $ACM^2_n$, $\alpha = 0.10$ | 0.0 | 0.1010  | 0.0925  | 0.1055  | 0.0900  |
|                | 0.5 | 0.2550  | 0.5135  | 0.9190  | 1.0000  |
| $ACM^2_n$, $\alpha = 0.05$ | 0.0 | 0.0520  | 0.0465  | 0.0445  | 0.0515  |
|                | 0.5 | 0.1445  | 0.3225  | 0.7550  | 1.0000  |
| $ACM^2_n$, $\alpha = 0.01$ | 0.0 | 0.0095  | 0.0090  | 0.0120  | 0.0070  |
|                | 0.5 | 0.0460  | 0.1060  | 0.3485  | 0.9140  |
| $T^{SZ}_n$, $\alpha = 0.10$ | 0.0 | 0.0980  | 0.0990  | 0.0865  | 0.0930  |
|                | 0.5 | 0.2630  | 0.5265  | 0.9240  | 1.0000  |
| $T^{SZ}_n$, $\alpha = 0.05$ | 0.0 | 0.0530  | 0.0480  | 0.0515  | 0.0495  |
|                | 0.5 | 0.1760  | 0.3235  | 0.7350  | 0.9970  |
| $T^{SZ}_n$, $\alpha = 0.01$ | 0.0 | 0.0100  | 0.0060  | 0.0085  | 0.0105  |
|                | 0.5 | 0.0470  | 0.1145  | 0.3580  | 0.9350  |
| $PCvM_n$, $\alpha = 0.10$ | 0.0 | 0.1080  | 0.1170  | 0.1230  | 0.1000  |
|                | 0.5 | 0.2560  | 0.3390  | 0.5160  | 0.7590  |
| $PCvM_n$, $\alpha = 0.05$ | 0.0 | 0.0530  | 0.0590  | 0.0440  | 0.0700  |
|                | 0.5 | 0.1470  | 0.2320  | 0.4080  | 0.6250  |
| $PCvM_n$, $\alpha = 0.01$ | 0.0 | 0.0130  | 0.0130  | 0.0080  | 0.0130  |
|                | 0.5 | 0.0450  | 0.1020  | 0.2010  | 0.4080  |
| $ICM_n$, $\alpha = 0.10$ | 0.0 | 0.0370  | 0.0000  | 0.0000  | 0.0000  |
|                | 0.5 | 0.1950  | 0.0330  | 0.0020  | 0.0000  |
| $ICM_n$, $\alpha = 0.05$ | 0.0 | 0.0110  | 0.0000  | 0.0000  | 0.0000  |
|                | 0.5 | 0.0790  | 0.0020  | 0.0000  | 0.0000  |
| $ICM_n$, $\alpha = 0.01$ | 0.0 | 0.0020  | 0.0000  | 0.0000  | 0.0000  |
| $T^{ZH}_n$, $\alpha = 0.10$ | 0.0 | 0.0805  | 0.0830  | 0.0800  | 0.1095  |
|                | 0.5 | 0.1630  | 0.1515  | 0.1825  | 0.1665  |
| $T^{ZH}_n$, $\alpha = 0.05$ | 0.0 | 0.0325  | 0.0350  | 0.0320  | 0.0330  |
|                | 0.5 | 0.0755  | 0.0775  | 0.0940  | 0.0615  |
| $T^{ZH}_n$, $\alpha = 0.01$ | 0.0 | 0.0045  | 0.0015  | 0.0035  | 0.0035  |
|                | 0.5 | 0.0155  | 0.0095  | 0.0125  | 0.0060  |
| $T^{GWZ}_n$, $\alpha = 0.10$ | 0.0 | 0.0820  | 0.0725  | 0.0810  | 0.0745  |
|                | 0.5 | 0.6765  | 0.9460  | 1.0000  | 1.0000  |
| $T^{GWZ}_n$, $\alpha = 0.05$ | 0.0 | 0.0495  | 0.0500  | 0.0500  | 0.0535  |
|                | 0.5 | 0.6035  | 0.9335  | 0.9995  | 1.0000  |
| $T^{GWZ}_n$, $\alpha = 0.01$ | 0.0 | 0.0190  | 0.0165  | 0.0180  | 0.0210  |
|                | 0.5 | 0.4660  | 0.8705  | 0.9980  | 1.0000  |
Table 3: Empirical sizes and powers of $ACM^2_n$, $T^SZ_n$, $PCvM_n$, $ICM_n$, $T^ZH_n$ and $T^GWZ_n$ for $H_0$ vs. $H_{13}$ in Study 1.

|                | a  | n=100 | n=200 | n=400 | n=800 |
|----------------|-----|-------|-------|-------|-------|
| $ACM^2_n$, $\alpha = 0.10$ |     |       |       |       |       |
|                | 0.00 | 0.9985 | 0.1050 | 0.1085 | 0.1090 |
|                | 0.25 | 0.7130 | 0.9410 | 0.9955 | 1.0000 |
| $ACM^2_n$, $\alpha = 0.05$ |     |       |       |       |       |
|                | 0.00 | 0.0500 | 0.0455 | 0.0435 | 0.0450 |
|                | 0.25 | 0.5970 | 0.8945 | 0.9980 | 1.0000 |
| $ACM^2_n$, $\alpha = 0.01$ |     |       |       |       |       |
|                | 0.00 | 0.0095 | 0.0090 | 0.0095 | 0.0090 |
|                | 0.25 | 0.3470 | 0.7225 | 0.9840 | 1.0000 |
| $T^SZ_n$, $\alpha = 0.10$ |     |       |       |       |       |
|                | 0.00 | 0.0960 | 0.1055 | 0.1060 | 0.0960 |
|                | 0.25 | 0.7190 | 0.9405 | 0.9975 | 1.0000 |
| $T^SZ_n$, $\alpha = 0.05$ |     |       |       |       |       |
|                | 0.00 | 0.0505 | 0.0420 | 0.0470 | 0.0495 |
|                | 0.25 | 0.5940 | 0.8980 | 0.9945 | 1.0000 |
| $T^SZ_n$, $\alpha = 0.01$ |     |       |       |       |       |
|                | 0.00 | 0.0080 | 0.0125 | 0.0095 | 0.0115 |
|                | 0.25 | 0.3310 | 0.7190 | 0.9705 | 0.9995 |
| $PCvM_n$, $\alpha = 0.10$ |     |       |       |       |       |
|                | 0.00 | 0.1030 | 0.0980 | 0.1140 | 0.1210 |
|                | 0.25 | 0.7180 | 0.9500 | 0.9970 | 1.0000 |
| $PCvM_n$, $\alpha = 0.05$ |     |       |       |       |       |
|                | 0.00 | 0.0580 | 0.0600 | 0.0440 | 0.0570 |
|                | 0.25 | 0.6160 | 0.8980 | 0.9970 | 1.0000 |
| $PCvM_n$, $\alpha = 0.01$ |     |       |       |       |       |
|                | 0.00 | 0.0060 | 0.0150 | 0.0080 | 0.0070 |
|                | 0.25 | 0.3870 | 0.7360 | 0.9780 | 1.0000 |
| $ICM_n$, $\alpha = 0.10$ |     |       |       |       |       |
|                | 0.00 | 0.0290 | 0.0010 | 0.0000 | 0.0000 |
|                | 0.25 | 0.1590 | 0.0190 | 0.0030 | 0.0000 |
| $ICM_n$, $\alpha = 0.05$ |     |       |       |       |       |
|                | 0.00 | 0.0110 | 0.0000 | 0.0000 | 0.0000 |
|                | 0.25 | 0.0590 | 0.0010 | 0.0000 | 0.0000 |
| $ICM_n$, $\alpha = 0.01$ |     |       |       |       |       |
|                | 0.00 | 0.0010 | 0.0000 | 0.0000 | 0.0000 |
|                | 0.25 | 0.0140 | 0.0000 | 0.0000 | 0.0000 |
| $T^ZH_n$, $\alpha = 0.10$ |     |       |       |       |       |
|                | 0.00 | 0.0765 | 0.0810 | 0.0940 | 0.0970 |
|                | 0.25 | 0.1135 | 0.1185 | 0.1400 | 0.1305 |
| $T^ZH_n$, $\alpha = 0.05$ |     |       |       |       |       |
|                | 0.00 | 0.0275 | 0.0310 | 0.0315 | 0.0340 |
|                | 0.25 | 0.0730 | 0.0485 | 0.0745 | 0.0625 |
| $T^ZH_n$, $\alpha = 0.01$ |     |       |       |       |       |
|                | 0.00 | 0.0030 | 0.0020 | 0.0030 | 0.0010 |
|                | 0.25 | 0.0055 | 0.0060 | 0.0080 | 0.0030 |
| $T^GWZ_n$, $\alpha = 0.10$ |     |       |       |       |       |
|                | 0.00 | 0.0800 | 0.0735 | 0.0710 | 0.0765 |
|                | 0.25 | 0.4580 | 0.7430 | 0.9795 | 0.9995 |
| $T^GWZ_n$, $\alpha = 0.05$ |     |       |       |       |       |
|                | 0.00 | 0.0510 | 0.0505 | 0.0540 | 0.0490 |
|                | 0.25 | 0.3840 | 0.6660 | 0.9465 | 1.0000 |
| $T^GWZ_n$, $\alpha = 0.01$ |     |       |       |       |       |
|                | 0.00 | 0.0290 | 0.0225 | 0.0235 | 0.0240 |
|                | 0.25 | 0.2590 | 0.5570 | 0.9040 | 0.9995 |
Table 4: Empirical sizes and powers of $ACM_n^2$, $T_n^{SZ}$, $PCvM_n$, $ICM_n$, $T_n^{ZH}$ and $T_n^{GWZ}$ for $H_0$ vs. $H_{14}$ in Study 1.

| Test Statistic | $\alpha$ | n=100 | n=200 | n=400 | n=800 |
|----------------|----------|-------|-------|-------|-------|
|                | p=7      | p=10  | p=12  | p=16  |       |
| $ACM_n^2$, $\alpha = 0.10$ | 0.00 | 0.1130 | 0.1000 | 0.0970 | 0.0955 |
|                | 0.25 | 0.9825 | 1.0000 | 1.0000 | 1.0000 |
| $ACM_n^2$, $\alpha = 0.05$ | 0.00 | 0.0520 | 0.0460 | 0.0545 | 0.0490 |
|                | 0.25 | 0.9525 | 1.0000 | 1.0000 | 1.0000 |
| $ACM_n^2$, $\alpha = 0.01$ | 0.00 | 0.0110 | 0.0090 | 0.0075 | 0.0105 |
|                | 0.25 | 0.8680 | 0.9950 | 1.0000 | 1.0000 |
| $T_n^{SZ}$, $\alpha = 0.10$ | 0.00 | 0.1090 | 0.0970 | 0.0910 | 0.1090 |
|                | 0.25 | 0.9805 | 0.9990 | 1.0000 | 1.0000 |
| $T_n^{SZ}$, $\alpha = 0.05$ | 0.00 | 0.0475 | 0.0490 | 0.0460 | 0.0555 |
|                | 0.25 | 0.9605 | 0.9995 | 1.0000 | 1.0000 |
| $T_n^{SZ}$, $\alpha = 0.01$ | 0.00 | 0.0095 | 0.0115 | 0.0075 | 0.0090 |
|                | 0.25 | 0.8700 | 0.9970 | 1.0000 | 1.0000 |
| $PCvM_n$, $\alpha = 0.10$ | 0.00 | 0.0950 | 0.1130 | 0.1110 | 0.1040 |
|                | 0.25 | 0.9960 | 1.0000 | 1.0000 | 1.0000 |
| $PCvM_n$, $\alpha = 0.05$ | 0.00 | 0.0580 | 0.0540 | 0.0570 | 0.0540 |
|                | 0.25 | 0.9690 | 0.9990 | 1.0000 | 1.0000 |
| $PCvM_n$, $\alpha = 0.01$ | 0.00 | 0.0140 | 0.0170 | 0.0080 | 0.0150 |
|                | 0.25 | 0.8730 | 0.9980 | 1.0000 | 1.0000 |
| $ICM_n$, $\alpha = 0.10$ | 0.00 | 0.0290 | 0.0010 | 0.0000 | 0.0000 |
|                | 0.25 | 0.5680 | 0.2420 | 0.1330 | 0.0000 |
| $ICM_n$, $\alpha = 0.05$ | 0.00 | 0.0050 | 0.0000 | 0.0000 | 0.0000 |
|                | 0.25 | 0.3670 | 0.0740 | 0.0120 | 0.0000 |
| $ICM_n$, $\alpha = 0.01$ | 0.00 | 0.0010 | 0.0000 | 0.0000 | 0.0000 |
|                | 0.25 | 0.1060 | 0.0040 | 0.0000 | 0.0000 |
| $T_n^{ZH}$, $\alpha = 0.10$ | 0.00 | 0.0700 | 0.0910 | 0.0875 | 0.0985 |
|                | 0.25 | 0.2420 | 0.2125 | 0.2680 | 0.2210 |
| $T_n^{ZH}$, $\alpha = 0.05$ | 0.00 | 0.0320 | 0.0295 | 0.0325 | 0.0380 |
|                | 0.25 | 0.1145 | 0.1195 | 0.1410 | 0.1145 |
| $T_n^{ZH}$, $\alpha = 0.01$ | 0.00 | 0.0015 | 0.0045 | 0.0050 | 0.0035 |
|                | 0.25 | 0.0335 | 0.0230 | 0.0220 | 0.0095 |
| $T_n^{GWZ}$, $\alpha = 0.10$ | 0.00 | 0.0780 | 0.0805 | 0.0815 | 0.0830 |
|                | 0.25 | 0.8645 | 0.9935 | 1.0000 | 1.0000 |
| $T_n^{GWZ}$, $\alpha = 0.05$ | 0.00 | 0.0455 | 0.0560 | 0.0540 | 0.0625 |
|                | 0.25 | 0.8405 | 0.9870 | 1.0000 | 1.0000 |
| $T_n^{GWZ}$, $\alpha = 0.01$ | 0.00 | 0.0210 | 0.0195 | 0.0225 | 0.0195 |
|                | 0.25 | 0.7285 | 0.9735 | 1.0000 | 1.0000 |
Table 5: Empirical sizes and powers of $ACM^2_n$, $T^S_{nZ}$, $PCvM_n$, $ICM_n$, $T^ZH_n$ and $T^GWZ_n$ for $H_0$ vs. $H_{21}$ in Study 2.

|                  | $a=100$ | $a=200$ | $a=400$ | $a=800$ |
|------------------|---------|---------|---------|---------|
|                  | $n=7$   | $n=10$  | $n=12$  | $n=16$  |
| $ACM^2_n$, $\alpha = 0.10$ | 0.00    | 0.1075  | 0.0965  | 0.0910  | 0.1035  |
|                  | 0.25    | 0.6185  | 0.8980  | 0.9955  | 1.0000  |
| $ACM^2_n$, $\alpha = 0.05$  | 0.00    | 0.0520  | 0.0490  | 0.0495  | 0.0570  |
|                  | 0.25    | 0.4895  | 0.8185  | 0.9925  | 1.0000  |
| $ACM^2_n$, $\alpha = 0.01$  | 0.00    | 0.0100  | 0.0085  | 0.0100  | 0.0115  |
|                  | 0.25    | 0.2505  | 0.5920  | 0.9450  | 0.9995  |
| $T^S_{nZ}$, $\alpha = 0.10$ | 0.00    | 0.0935  | 0.0935  | 0.1070  | 0.1055  |
|                  | 0.25    | 0.7005  | 0.9120  | 0.9965  | 1.0000  |
| $T^S_{nZ}$, $\alpha = 0.05$  | 0.00    | 0.0515  | 0.0425  | 0.0460  | 0.0445  |
|                  | 0.25    | 0.5600  | 0.8505  | 0.9940  | 1.0000  |
| $T^S_{nZ}$, $\alpha = 0.01$  | 0.00    | 0.0080  | 0.0100  | 0.0060  | 0.0100  |
|                  | 0.25    | 0.3180  | 0.6680  | 0.9665  | 1.0000  |
| $PCvM_n$, $\alpha = 0.10$  | 0.00    | 0.1150  | 0.0910  | 0.1090  | 0.1050  |
|                  | 0.25    | 0.7080  | 0.9320  | 0.9990  | 1.0000  |
| $PCvM_n$, $\alpha = 0.05$  | 0.00    | 0.0569  | 0.0480  | 0.0570  | 0.0430  |
|                  | 0.25    | 0.6230  | 0.9080  | 0.9960  | 1.0000  |
| $PCvM_n$, $\alpha = 0.01$  | 0.00    | 0.0080  | 0.0120  | 0.0100  | 0.0090  |
|                  | 0.25    | 0.3810  | 0.7230  | 0.9820  | 1.0000  |
| $ICM_n$, $\alpha = 0.10$  | 0.00    | 0.0160  | 0.0010  | 0.0000  | 0.0000  |
|                  | 0.25    | 0.1220  | 0.0060  | 0.0000  | 0.0000  |
| $ICM_n$, $\alpha = 0.05$  | 0.00    | 0.0040  | 0.0000  | 0.0000  | 0.0000  |
|                  | 0.25    | 0.0470  | 0.0010  | 0.0000  | 0.0000  |
| $ICM_n$, $\alpha = 0.01$  | 0.00    | 0.0000  | 0.0000  | 0.0000  | 0.0000  |
|                  | 0.25    | 0.0070  | 0.0000  | 0.0000  | 0.0000  |
| $T^ZH_n$, $\alpha = 0.10$  | 0.00    | 0.1100  | 0.1020  | 0.0960  | 0.1110  |
|                  | 0.25    | 0.1420  | 0.1370  | 0.1550  | 0.1545  |
| $T^ZH_n$, $\alpha = 0.05$  | 0.00    | 0.0400  | 0.0410  | 0.0365  | 0.0390  |
|                  | 0.25    | 0.0710  | 0.0700  | 0.0610  | 0.0550  |
| $T^ZH_n$, $\alpha = 0.01$  | 0.00    | 0.0045  | 0.0035  | 0.0040  | 0.0035  |
|                  | 0.25    | 0.0140  | 0.0075  | 0.0065  | 0.0035  |
| $T^GWZ_n$, $\alpha = 0.10$  | 0.00    | 0.1135  | 0.1045  | 0.1115  | 0.1240  |
|                  | 0.25    | 0.5275  | 0.8140  | 0.9860  | 0.9995  |
| $T^GWZ_n$, $\alpha = 0.05$  | 0.00    | 0.0790  | 0.0760  | 0.0775  | 0.0750  |
|                  | 0.25    | 0.4625  | 0.7300  | 0.9610  | 1.0000  |
| $T^GWZ_n$, $\alpha = 0.01$  | 0.00    | 0.0340  | 0.0345  | 0.0310  | 0.0305  |
|                  | 0.25    | 0.3175  | 0.6015  | 0.9295  | 0.9985  |
Table 6: Empirical sizes and powers of $ACM^2_n$, $T^n_{SZ}$, $PCvM_n$, $ICM_n$, $T^n_{ZH}$ and $T^n_{GWZ}$ for $H_0$ vs. $H_{22}$ in Study 2.

|                      | $n=100$ | $n=200$ | $n=400$ | $n=800$ |
|----------------------|---------|---------|---------|---------|
|                      | $p=7$   | $p=10$  | $p=12$  | $p=16$  |
| $ACM^2_n$, $\alpha = 0.10$ | 0.0     | 0.1180  | 0.1190  | 0.1095  | 0.1060  |
|                      | 0.5     | 0.2255  | 0.3090  | 0.4805  | 0.7390  |
| $ACM^2_n$, $\alpha = 0.05$ | 0.0     | 0.0575  | 0.0550  | 0.0585  | 0.0530  |
|                      | 0.5     | 0.1295  | 0.1895  | 0.3030  | 0.5790  |
| $ACM^2_n$, $\alpha = 0.01$ | 0.0     | 0.0110  | 0.0135  | 0.0115  | 0.0120  |
|                      | 0.5     | 0.0325  | 0.0605  | 0.1155  | 0.2830  |
| $T^n_{SZ}$, $\alpha = 0.10$ | 0.0     | 0.1110  | 0.1075  | 0.0980  | 0.1010  |
|                      | 0.5     | 0.1335  | 0.1480  | 0.1580  | 0.1920  |
| $T^n_{SZ}$, $\alpha = 0.05$ | 0.0     | 0.0650  | 0.0535  | 0.0550  | 0.0550  |
|                      | 0.5     | 0.0755  | 0.0970  | 0.0835  | 0.1195  |
| $T^n_{SZ}$, $\alpha = 0.01$ | 0.0     | 0.0085  | 0.0140  | 0.0095  | 0.0120  |
|                      | 0.5     | 0.0205  | 0.0285  | 0.0180  | 0.0330  |
| $PCvM_n$, $\alpha = 0.10$ | 0.0     | 0.1110  | 0.1160  | 0.1010  | 0.1180  |
|                      | 0.5     | 0.2370  | 0.3480  | 0.4730  | 0.6630  |
| $PCvM_n$, $\alpha = 0.05$ | 0.0     | 0.0470  | 0.0560  | 0.0690  | 0.0510  |
|                      | 0.5     | 0.1310  | 0.2000  | 0.2760  | 0.4450  |
| $PCvM_n$, $\alpha = 0.01$ | 0.0     | 0.0070  | 0.0100  | 0.0240  | 0.0100  |
|                      | 0.5     | 0.0430  | 0.0580  | 0.0930  | 0.1700  |
| $ICM_n$, $\alpha = 0.10$ | 0.0     | 0.0200  | 0.0000  | 0.0000  | 0.0000  |
|                      | 0.5     | 0.0980  | 0.0140  | 0.0030  | 0.0020  |
| $ICM_n$, $\alpha = 0.05$ | 0.0     | 0.0050  | 0.0000  | 0.0000  | 0.0000  |
|                      | 0.5     | 0.0210  | 0.0020  | 0.0000  | 0.0000  |
| $ICM_n$, $\alpha = 0.01$ | 0.0     | 0.0000  | 0.0000  | 0.0000  | 0.0000  |
|                      | 0.5     | 0.0000  | 0.0000  | 0.0000  | 0.0000  |
| $T^n_{ZH}$, $\alpha = 0.10$ | 0.0     | 0.0940  | 0.0915  | 0.0985  | 0.1135  |
|                      | 0.5     | 0.1325  | 0.1455  | 0.1625  | 0.1455  |
| $T^n_{ZH}$, $\alpha = 0.05$ | 0.0     | 0.0445  | 0.0365  | 0.0410  | 0.0380  |
|                      | 0.5     | 0.0690  | 0.0765  | 0.0770  | 0.0545  |
| $T^n_{ZH}$, $\alpha = 0.01$ | 0.0     | 0.0050  | 0.0035  | 0.0020  | 0.0020  |
|                      | 0.5     | 0.0125  | 0.0090  | 0.0070  | 0.0040  |
| $T^n_{GWZ}$, $\alpha = 0.10$ | 0.0     | 0.1015  | 0.1020  | 0.0995  | 0.1125  |
|                      | 0.5     | 0.2380  | 0.3745  | 0.5450  | 0.8265  |
| $T^n_{GWZ}$, $\alpha = 0.05$ | 0.0     | 0.0615  | 0.0675  | 0.0670  | 0.0580  |
|                      | 0.5     | 0.1700  | 0.2750  | 0.4560  | 0.7725  |
| $T^n_{GWZ}$, $\alpha = 0.01$ | 0.0     | 0.0240  | 0.0270  | 0.0290  | 0.0335  |
|                      | 0.5     | 0.1015  | 0.1655  | 0.3360  | 0.6260  |
Figure 1: Scatter plots of the response $Y$ 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 2: Scatter plot of the response $Y$ against the projected covariates $(\hat{\beta}_1^T X, \hat{\beta}_2^T X)$ where the directions $(\hat{\beta}_1, \hat{\beta}_2)$ are obtained by CSE.