


Article

Robustness Property of Robust-BD Wald-Type Test for Varying-Dimensional General Linear Models

Xiao Guo ^{1,*}  and Chunming Zhang ²

¹ Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China

² Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA; cmzhang@stat.wisc.edu

* Correspondence: xiaoguo@ustc.edu.cn; Tel.: +86-55163603167

Received: 12 January 2018; Accepted: 1 March 2018; Published: 5 March 2018

Abstract: An important issue for robust inference is to examine the stability of the asymptotic level and power of the test statistic in the presence of contaminated data. Most existing results are derived in finite-dimensional settings with some particular choices of loss functions. This paper re-examines this issue by allowing for a diverging number of parameters combined with a broader array of robust error measures, called “robust-BD”, for the class of “general linear models”. Under regularity conditions, we derive the influence function of the *robust*-BD parameter estimator and demonstrate that the *robust*-BD Wald-type test enjoys the robustness of validity and efficiency asymptotically. Specifically, the asymptotic level of the test is stable under a small amount of contamination of the null hypothesis, whereas the asymptotic power is large enough under a contaminated distribution in a neighborhood of the contiguous alternatives, thus lending supports to the utility of the proposed *robust*-BD Wald-type test.

Keywords: Bregman divergence; general linear model; hypothesis testing; influence function; robust; Wald-type test

1. Introduction

The class of varying-dimensional “general linear models” [1], including the conventional generalized linear model (GLM in [2]), is flexible and powerful for modeling a large variety of data and plays an important role in many statistical applications. In the literature, it has been extensively studied that the conventional maximum likelihood estimator for the GLM is nonrobust; for example, see [3,4]. To enhance the resistance to outliers in applications, many efforts have been made to obtain robust estimators. For example, Noh et al. [5] and Künsch et al. [6] developed robust estimator for the GLM, and Stefanski et al. [7], Bianco et al. [8] and Croux et al. [9] studied robust estimation for the logistic regression model with the deviance loss as the error measure.

Besides robust estimation for the GLM, robust inference is another important issue, which, however, receives relatively less attention. Basically, the study of robust testing includes two aspects: (i) establishing the stability of the asymptotic level under small departures from the null hypothesis (i.e., robustness of “validity”); and (ii) demonstrating that the asymptotic power is sufficiently large under small departures from specified alternatives (i.e., robustness of “efficiency”). In the literature, robust inference has been conducted for different models. For example, Heritier et al. [10] studied the robustness properties of the Wald, score and likelihood ratio tests based on M estimators for general parametric models. Cantoni et al. [11] developed a test statistic based on the robust deviance, and conducted robust inference for the GLM using quasi-likelihood as the loss function. A robust Wald-type test for the logistic regression model is studied in [12]. Ronchetti et al. [13]

concerned the robustness property for the generalized method of moments estimators. Basu et al. [14] proposed robust tests based on the density power divergence (DPD) measure for the equality of two normal means. Robust tests for parameter change have been studied using the density-based divergence method in [15,16]. However, the aforementioned methods based on the GLM mostly focus on situations where the number of parameters is fixed and the loss function is specific.

Zhang et al. [1] developed robust estimation and testing for the “general linear model” based on a broader array of error measures, namely Bregman divergence, allowing for a diverging number of parameters. The Bregman divergence includes a wide class of error measures as special cases, e.g., the (negative) quasi-likelihood in regression, the deviance loss and exponential loss in machine learning practice, among many other commonly used loss functions. Zhang et al. [1] studied the consistency and asymptotic normality of their proposed *robust*-BD parameter estimator and demonstrated the asymptotic distribution of the Wald-type test constructed from *robust*-BD estimators. Naturally, it remains an important issue to examine the robustness property of the *robust*-BD Wald-type test [1] in the varying-dimensional case, i.e., whether the test still has stable asymptotic level and power, in the presence of contaminated data.

This paper aims to demonstrate the robustness property of the *robust*-BD Wald-type test in [1]. Nevertheless, it is a nontrivial task to address this issue. Although the local stability for the Wald-type tests have been established for the M estimators [10], generalized method of moment estimators [13], minimum density power divergence estimator [17] and general M estimators under random censoring [18], their results for finite-dimensional settings are not directly applicable to our situations with a diverging number of parameters. Under certain regularity conditions, we provide rigorous theoretical derivation for robust testing based on the Wald-type test statistic. The essential results are approximations of the asymptotic level and power under contaminated distributions of the data in a small neighborhood of the null and alternative hypotheses, respectively.

- Specifically, we show in Theorem 1 that, if the influence function of the estimator is bounded, then the asymptotic level of the test is also bounded under a small amount of contamination.
- We also demonstrate in Theorem 2 that, if the contamination belongs to a neighborhood of the contiguous alternatives, then the asymptotic power is also stable.

Hence, we contribute to establish the robustness of validity and efficiency for the *robust*-BD Wald-type test for the “general linear model” with a diverging number of parameters.

The rest of the paper is organized as follows. Section 2 reviews the Bregman divergence (BD), *robust*-BD estimation and the Wald-type test statistic proposed in [1]. Section 3 derives the influence function of the *robust*-BD estimator and studies the robustness properties of the asymptotic level and power of the Wald-type test under a small amount of contamination. Section 4 conducts the simulation studies. The technical conditions and proofs are given in Appendix A. A list of notations and symbols is provided in Appendix B.

We will introduce some necessary notations. In the following, C and c are generic finite constants which may vary from place to place, but do not depend on the sample size n . Denote by $E_K(\cdot)$ the expectation with respect to the underlying distribution K . For a positive integer q , let $\mathbf{0}_q = (0, \dots, 0)^T \in \mathbb{R}^q$ be a $q \times 1$ zero vector and \mathbf{I}_q be the $q \times q$ identity matrix. For a vector $\mathbf{v} = (v_1, \dots, v_q)^T \in \mathbb{R}^q$, the L_1 norm is $\|\mathbf{v}\|_1 = \sum_{i=1}^q |v_i|$, L_2 norm is $\|\mathbf{v}\|_2 = (\sum_{i=1}^q v_i^2)^{1/2}$ and the L_∞ norm is $\|\mathbf{v}\|_\infty = \max_{i=1, \dots, q} |v_i|$. For a $q \times q$ matrix A , the L_2 and Frobenius norms of A are $\|A\|_2 = \{\lambda_{\max}(A^T A)\}^{1/2}$ and $\|A\|_F = \sqrt{\text{tr}(AA^T)}$, respectively, where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a matrix and $\text{tr}(\cdot)$ denotes the trace of a matrix.

2. Review of Robust-BD Estimation and Inference for “General Linear Models”

This section briefly reviews the *robust*-BD estimation and inference methods for the “general linear model” developed in [1]. Let $\{(\mathbf{X}_{n1}, Y_1), \dots, (\mathbf{X}_{nm}, Y_m)\}$ be i.i.d. observations from some underlying distribution (\mathbf{X}_n, Y) with $\mathbf{X}_n = (X_1, \dots, X_{p_n})^T \in \mathbb{R}^{p_n}$ the explanatory variables and Y the response

variable. The dimension p_n is allowed to diverge with the sample size n . The “general linear model” is given by

$$m(\mathbf{x}_n) \equiv E(Y | \mathbf{X}_n = \mathbf{x}_n) = F^{-1}(\tilde{\mathbf{x}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}), \tag{1}$$

and

$$\text{var}(Y | \mathbf{X}_n = \mathbf{x}_n) = V(m(\mathbf{x}_n)), \tag{2}$$

where F is a known link function, $\tilde{\boldsymbol{\beta}}_{n,0} \in \mathbb{R}^{p_n+1}$ is the vector of unknown true regression parameters, $\tilde{\mathbf{x}}_n = (1, \mathbf{x}_n^T)^T$ and $V(\cdot)$ is a known function. Note that the conventional generalized linear model (GLM) satisfying Equations (1) and (2) assumes that $Y | \mathbf{X}_n = \mathbf{x}_n$ follows a particular distribution in the exponential family. However, our “general linear model” does not require explicit form of distributions of the response. Hence, the “general linear model” includes the GLM as a special case. For notational simplicity, denote $\mathbf{Z}_n = (\mathbf{X}_n^T, Y)^T$ and $\tilde{\mathbf{Z}}_n = (\tilde{\mathbf{X}}_n^T, Y)^T$.

Bregman divergence (BD) is a class of error measures, which is introduced in [19] and covers a wide range of loss functions. Specifically, Bregman divergence is defined as a bivariate function,

$$Q_q(v, \mu) = -q(v) + q(\mu) + (v - \mu)q'(\mu),$$

where $q(\cdot)$ is the concave generating q -function. For example, $q(\mu) = a\mu - \mu^2$ for a constant a corresponds to the quadratic loss $Q_a(Y, \mu) = (Y - \mu)^2$. For a binary response variable Y , $q(\mu) = \min\{\mu, 1 - \mu\}$ gives the misclassification loss $Q_q(Y, \mu) = I\{Y \neq I(\mu > 0.5)\}$; $q(\mu) = -2\{\mu \log(\mu) + (1 - \mu) \log(1 - \mu)\}$ gives Bernoulli deviance loss $Q_q(Y, \mu) = -2\{Y \log(\mu) + (1 - Y) \log(1 - \mu)\}$; $q(\mu) = 2 \min\{\mu, 1 - \mu\}$ gives the hinge loss $Q_q(Y, \mu) = \max\{1 - (2Y - 1)\text{sign}(\mu - 0.5), 0\}$ for the support vector machine; $q(\mu) = 2\{\mu(1 - \mu)\}^{1/2}$ yields the exponential loss $Q_q(Y, \mu) = \exp[-(Y - 0.5) \log\{\mu/(1 - \mu)\}]$ used in AdaBoost [20]. Furthermore, Zhang et al. [21] showed that if

$$q(\mu) = \int_a^\mu \frac{s - \mu}{V(s)} ds, \tag{3}$$

where a is a finite constant such that the integral is well-defined, then $Q_q(y, \mu)$ is the “classical (negative) quasi-likelihood” function $-Q_{QL}(y, \mu)$ with $\partial Q_{QL}(y, \mu) / \partial \mu = (y - \mu) / V(\mu)$.

To obtain a robust estimator based on BD, Zhang et al. [1] developed the *robust*-BD loss function

$$\rho_q(y, \mu) = \int_y^\mu \psi(r(y, s)) \{q''(s) \sqrt{V(s)}\} ds - G(\mu), \tag{4}$$

where $\psi(\cdot)$ is a bounded odd function, such as the Huber ψ -function [22], $r(y, s) = (y - s) / \sqrt{V(s)}$ denotes the Pearson residual and $G(\mu)$ is the bias-correction term satisfying

$$G'(\mu) = G'_1(\mu) \{q''(\mu) \sqrt{V(\mu)}\},$$

with

$$G'_1(m(\mathbf{x}_n)) = E\{\psi(r(Y, m(\mathbf{x}_n))) | \mathbf{X}_n = \mathbf{x}_n\}.$$

Based on *robust*-BD, the estimator of $\tilde{\boldsymbol{\beta}}_{n,0}$ proposed in [1] is defined as

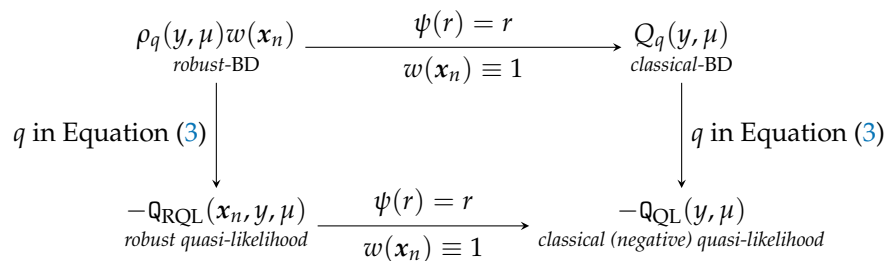
$$\hat{\boldsymbol{\beta}} = \arg \min_{\tilde{\boldsymbol{\beta}}} \left\{ \frac{1}{n} \sum_{i=1}^n \rho_q(Y_i, F^{-1}(\tilde{\mathbf{X}}_{ni}^T \tilde{\boldsymbol{\beta}})) w(\mathbf{X}_{ni}) \right\}, \tag{5}$$

where $w(\cdot) \geq 0$ is a known bounded weight function which downweights the high leverage points.

In [11], the “robust quasi-likelihood estimator” of $\tilde{\beta}_{n,0}$ is formulated according to the “robust quasi-likelihood function” defined as

$$Q_{\text{RQL}}(\mathbf{x}_n, \mathbf{y}, \mu) = \left\{ \int_{\mu_0}^{\mu} \psi(r(\mathbf{y}, s)) / \sqrt{V(s)} ds \right\} w(\mathbf{x}_n) - \frac{1}{n} \sum_{j=1}^n \int_{\mu_0}^{\mu_j} \left[E\{\psi(r(Y_j, s)) | \mathbf{X}_{nj}\} / \sqrt{V(s)} ds \right] w(\mathbf{X}_{nj}),$$

where $\mu = F^{-1}(\tilde{\mathbf{x}}_n^T \tilde{\beta})$ and $\mu_j = \mu_j(\tilde{\beta}) = F^{-1}(\tilde{\mathbf{X}}_{nj}^T \tilde{\beta})$, $j = 1, \dots, n$. To describe the intuition of the “robust-BD”, we use the following diagram from [1], which illustrates the relation among the “robust-BD”, “classical-BD”, “robust quasi-likelihood” and “classical (negative) quasi-likelihood”.



For the robust-BD, assume that

$$p_j(\mathbf{y}; \theta) = \frac{\partial^j}{\partial \theta^j} \rho_q(\mathbf{y}, F^{-1}(\theta)), \quad j = 0, 1, \dots,$$

exist finitely up to any order required. For example, for $j = 1$,

$$p_1(\mathbf{y}; \theta) = \{\psi(r(\mathbf{y}, \mu)) - G'_1(\mu)\} \{q''(\mu) \sqrt{V(\mu)}\} / F'(\mu), \tag{6}$$

where $\mu = F^{-1}(\theta)$. Explicit expressions for $p_j(\mathbf{y}; \theta)$ ($j = 2, 3$) can be found in Equation (3.7) of [1]. Then, the estimation equation for $\tilde{\beta}$ is

$$\frac{1}{n} \sum_{i=1}^n \psi_{\text{RBD}}(\mathbf{Z}_{ni}; \tilde{\beta}) = \mathbf{0},$$

where the score vector is

$$\psi_{\text{RBD}}(\mathbf{z}_n; \tilde{\beta}) = p_1(\mathbf{y}; \theta) w(\mathbf{x}_n) \tilde{\mathbf{x}}_n, \tag{7}$$

with $\theta = \tilde{\mathbf{x}}_n^T \tilde{\beta}$. The consistency and asymptotic normality of $\tilde{\beta}$ have been studied in [1]; see Theorems 1 and 2 therein.

Furthermore, to conduct statistical inference for the “general linear model”, the following hypotheses are considered,

$$H_0 : A_n \tilde{\beta}_{n,0} = \mathbf{g}_0 \quad \text{versus} \quad H_1 : A_n \tilde{\beta}_{n,0} \neq \mathbf{g}_0, \tag{8}$$

where A_n is a given $k \times (p_n + 1)$ matrix such that $A_n A_n^T \rightarrow \mathbb{G}$ with \mathbb{G} being a $k \times k$ positive-definite matrix, and \mathbf{g}_0 is a known $k \times 1$ vector.

To perform the test of Equation (8), Zhang et al. [1] proposed the Wald-type test statistic,

$$W_n = n(A_n \tilde{\beta} - \mathbf{g}_0)^T (A_n \hat{\mathbf{H}}_n^{-1} \hat{\Omega}_n \hat{\mathbf{H}}_n^{-1} A_n^T)^{-1} (A_n \tilde{\beta} - \mathbf{g}_0), \tag{9}$$

constructed from the *robust*-BD estimator $\widehat{\beta}$ in Equation (5), where

$$\begin{aligned} \widehat{\Omega}_n &= \frac{1}{n} \sum_{i=1}^n p_1^2(Y_i; \widetilde{X}_{ni}^T \widehat{\beta}) w^2(\mathbf{X}_{ni}) \widetilde{X}_{ni} \widetilde{X}_{ni}^T, \\ \widehat{H}_n &= \frac{1}{n} \sum_{i=1}^n p_2(Y_i; \widetilde{X}_{ni}^T \widehat{\beta}) w(\mathbf{X}_{ni}) \widetilde{X}_{ni} \widetilde{X}_{ni}^T. \end{aligned}$$

The asymptotic distributions of W_n under the null and alternative hypotheses have been developed in [1]; see Theorems 4–6 therein.

On the other hand, the issue on the robustness of W_n , used for possibly contaminated data, remains unknown. Section 3 of this paper will address this issue with detailed derivations.

3. Robustness Properties of W_n in Equation (9)

This section derives the influence function of the *robust*-BD Wald-type test and studies the influence of a small amount of contamination on the asymptotic level and power of the test. The proofs of the theoretical results are given in Appendix A.

Denote by $K_{n,0}$ the true distribution of \mathbf{Z}_n following the “general linear model” characterized by Equations (1) and (2). To facilitate the discussion of robustness properties, we consider the ϵ -contamination,

$$K_{n,\epsilon} = \left(1 - \frac{\epsilon}{\sqrt{n}}\right) K_{n,0} + \frac{\epsilon}{\sqrt{n}} J, \tag{10}$$

where J is an arbitrary distribution and $\epsilon > 0$ is a constant. Then, $K_{n,\epsilon}$ is a contaminated distribution of \mathbf{Z}_n with the amount of contamination converging to 0 at rate $1/\sqrt{n}$. Denote by \mathbb{K}_n the empirical distribution of $\{\mathbf{Z}_{ni}\}_{i=1}^n$.

For a generic distribution K of \mathbf{Z}_n , define

$$\begin{aligned} \ell_K(\widetilde{\beta}) &= E_K\{\rho_q(Y, F^{-1}(\widetilde{X}_n^T \widetilde{\beta})) w(\mathbf{X}_n)\}, \\ \mathcal{S}_K &= \{\widetilde{\beta} : E_K\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \widetilde{\beta})\} = \mathbf{0}\}, \end{aligned} \tag{11}$$

where $\rho_q(\cdot, \cdot)$ and $\boldsymbol{\psi}_{\text{RBD}}(\cdot; \cdot)$ are defined in Equations (4) and (7), respectively. It’s worth noting that the solution to $E_K\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \widetilde{\beta})\} = \mathbf{0}$ may not be unique, i.e., \mathcal{S}_K may contain more than one element. We then define a functional for the estimator of $\widetilde{\beta}_{n,0}$ as follows,

$$T(K) = \arg \min_{\widetilde{\beta} \in \mathcal{S}_K} \|\widetilde{\beta} - \widetilde{\beta}_{n,0}\|. \tag{12}$$

From the result of Lemma A1 in Appendix A, $T(K_{n,\epsilon})$ is the unique local minimizer of $\ell_{K_{n,\epsilon}}(\widetilde{\beta})$ in the $\sqrt{p_n/n}$ -neighborhood of $\widetilde{\beta}_{n,0}$. Particularly, $T(K_{n,0}) = \widetilde{\beta}_{n,0}$. Similarly, from Lemma A2 in Appendix A, $T(\mathbb{K}_n)$ is the unique local minimizer of $\ell_{\mathbb{K}_n}(\widetilde{\beta})$ which satisfies $\|T(\mathbb{K}_n) - \widetilde{\beta}_{n,0}\| = O_P(\sqrt{p_n/n})$.

From [23] (Equation (2.1.6) on pp. 84), the influence function of $T(\cdot)$ at $K_{n,0}$ is defined as

$$\text{IF}(z_n; T, K_{n,0}) = \left. \frac{\partial}{\partial t} T((1-t)K_{n,0} + t\Delta_{z_n}) \right|_{t=0} = \lim_{t \downarrow 0} \frac{T((1-t)K_{n,0} + t\Delta_{z_n}) - \widetilde{\beta}_{n,0}}{t},$$

where Δ_{z_n} is the probability measure which puts mass 1 at the point z_n . Since the dimension of $T(\cdot)$ diverges with n , its influence function is defined for each fixed n . From Lemma A8 in Appendix A, under certain regularity conditions, the influence function exists and has the following expression:

$$\text{IF}(z_n; T, K_{n,0}) = -\mathbf{H}_n^{-1} \boldsymbol{\psi}_{\text{RBD}}(z_n; \widetilde{\beta}_{n,0}), \tag{13}$$

where $\mathbf{H}_n = E_{K_{n,0}}\{p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0})w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^T\}$. The form of the influence function for diverging p_n in Equation(13) coincides with that in [23,24] for fixed p_n .

In our theoretical derivations, approximations of the asymptotic level and power of W_n will involve the following matrices:

$$\begin{aligned}\Omega_n &= E_{K_{n,0}}\{p_1^2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0})w^2(\mathbf{X}_n)\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^T\}, \\ U_n &= A_n \mathbf{H}_n^{-1} \Omega_n \mathbf{H}_n^{-1} A_n^T.\end{aligned}$$

3.1. Asymptotic Level of W_n under Contamination

We now investigate the asymptotic level of the Wald-type test W_n under the ϵ -contamination.

Theorem 1. Assume Conditions A0–A9 and B4 in Appendix A. Suppose $p_n^6/n \rightarrow 0$ as $n \rightarrow \infty$, $\sup_n E_J(\|w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\|) \leq C$. Denote by $\alpha(K_{n,\epsilon})$ the level of $W_n = n\{A_n \mathbf{T}(\mathbb{K}_n) - \mathbf{g}_0\}^T (A_n \hat{\mathbf{H}}_n^{-1} \hat{\Omega}_n \hat{\mathbf{H}}_n^{-1} A_n^T)^{-1} \{A_n \mathbf{T}(\mathbb{K}_n) - \mathbf{g}_0\}$ when the underlying distribution is $K_{n,\epsilon}$ in Equation (10) and by α_0 the nominal level. Under H_0 in Equation (8), it follows that

$$\limsup_{n \rightarrow \infty} \alpha(K_{n,\epsilon}) = \alpha_0 + \epsilon^2 \mu_k D + o(\epsilon^2) \quad \text{as } \epsilon \rightarrow 0,$$

where

$$D = \limsup_{n \rightarrow \infty} \|U_n^{-1/2} A_n E_J\{\text{IF}(\mathbf{Z}_n; \mathbf{T}, K_{n,0})\}\|^2 < \infty,$$

$\mu_k = -\frac{\partial}{\partial \delta} H_k(\eta_{1-\alpha_0}; \delta)|_{\delta=0}$, $H_k(\cdot; \delta)$ is the cumulative distribution function of a $\chi_k^2(\delta)$ distribution, and $\eta_{1-\alpha_0}$ is the $1 - \alpha_0$ quantile of the central χ_k^2 distribution.

Theorem 1 indicates that if the influence function for $T(\cdot)$ is bounded, then the asymptotic level of W_n under the ϵ -contamination is also bounded and close to the nominal level when ϵ is sufficiently small. As a comparison, the robustness property in [10] of the Wald-type test is studied based on M-estimator for general parametric models with a fixed dimension p_n . They assumed certain conditions that guarantee Fréchet differentiability which further implies the existence of the influence function and the asymptotic normality of the corresponding estimator. However, in the set-ups of our paper, it's difficult to check those conditions, due to the use of Bregman divergence and the diverging dimension p_n . Hence, the assumptions we make in Theorem 1 are different from those in [10], and are comparatively mild and easy to check. Moreover, the result of Theorem 1 cannot be easily derived from that of [10].

In Theorem 1, p_n is allowed to diverge with $p_n^6/n = o(1)$, which is slower than that in [1] with $p_n^5/n = o(1)$. Theoretically, the assumption $p_n^5/n = o(1)$ is required to obtain the asymptotic distribution of W_n in [1]. Furthermore, to derive the limit distribution of W_n under the ϵ -contamination, assumption $p_n^6/n = o(1)$ is needed (see Lemma A7 in Appendix A). Hence, the reason that our assumption is stronger than that in [1] is the consideration of the ϵ -contamination of the data. Practically, due to the advancement of technology and different forms of data gathering, large dimension becomes a common characteristic and hence the varying-dimensional model has a wide range of applications, e.g., brain imaging data, financial data, web term-document data and gene expression data. Even some of the classical settings, e.g. the Framingham heart study with $n = 25,000$ and $p_n = 100$, can be viewed as varying-dimensional cases.

As an illustration, we apply the general result of Theorem 1 to the special case of a point mass contamination.

Corollary 1. With the notations in Theorem 1, assume Conditions A0–A9 in Appendix A, $\sup_{\mathbf{x}_n \in \mathbb{R}^{p_n}} \|w(\mathbf{x}_n)\mathbf{x}_n\| \leq C$ and $\sup_{\mu \in \mathbb{R}} |q''(\mu)\sqrt{V(\mu)}/F'(\mu)| \leq C$.

- (i) If $p_n \equiv p$, $A_n \equiv A$, $\tilde{\beta}_{n,0} \equiv \tilde{\beta}_0$, $K_{n,0} \equiv K_0$ and $U_n \equiv U$ are fixed, then, for $K_{n,\epsilon} = (1 - \epsilon/\sqrt{n})K_0 + \epsilon/\sqrt{n}\Delta_z$ with $z \in \mathbb{R}^p$ a fixed point, under H_0 in Equation (8), it follows that

$$\sup_{z \in \mathbb{R}^p} \lim_{n \rightarrow \infty} \alpha(K_{n,\epsilon}) = \alpha_0 + \epsilon^2 \mu_k D_1 + o(\epsilon^2) \quad \text{as } \epsilon \rightarrow 0,$$

where

$$D_1 = \sup_{z \in \mathbb{R}^p} \|U^{-1/2} A \text{IF}(z; \mathbf{T}, K_0)\|^2 < \infty.$$

- (ii) If p_n diverges with $p_n^6/n \rightarrow 0$, for $K_{n,\epsilon} = (1 - \epsilon/\sqrt{n})K_{n,0} + \epsilon/\sqrt{n}\Delta_{z_n}$ with $z_n \in \mathbb{R}^{p_n}$ a sequence of deterministic points, then, under H_0 in Equation (8),

$$\sup_{C_0 > 0} \sup_{z_n \in S_{C_0}} \limsup_{n \rightarrow \infty} \alpha(K_{n,\epsilon}) = \alpha_0 + \epsilon^2 \mu_k D_2 + o(\epsilon^2) \quad \text{as } \epsilon \rightarrow 0,$$

where $S_{C_0} = \{z_n = (\mathbf{x}_n^T, y)^T : \|\mathbf{x}_n\|_\infty \leq C_0\}$, $C_0 > 0$ is a constant and

$$D_2 = \sup_{C_0 > 0} \sup_{z_n \in S_{C_0}} \limsup_{n \rightarrow \infty} \|U_n^{-1/2} A_n \text{IF}(z_n; \mathbf{T}, K_{n,0})\|^2 < \infty.$$

In Corollary 1, conditions $\sup_{\mathbf{x}_n \in \mathbb{R}^{p_n}} \|w(\mathbf{x}_n)\mathbf{x}_n\| \leq C$ and $\sup_{\mu \in \mathbb{R}} |q''(\mu)\sqrt{V(\mu)}/F'(\mu)| \leq C$ are needed to guarantee the boundedness of the score function in Equation (7). Particularly, the function $w(\mathbf{x}_n)$ downweights the high leverage points and can be chosen as, e.g., $w(\mathbf{x}_n) = 1/(1 + \|\mathbf{x}_n\|)$. The condition $\sup_{\mu \in \mathbb{R}} |q''(\mu)\sqrt{V(\mu)}/F'(\mu)| \leq C$ is needed to bound Equation (6), and is satisfied in many situations.

- For example, for the linear model with $q(\mu) = a\mu - \mu^2$, $V(\mu) = \sigma^2$ and $F(\mu) = \mu$, where a and σ^2 are constants, we observe $|q''(\mu)\sqrt{V(\mu)}/F'(\mu)| = 2\sigma \leq C$.
- Another example is the logistic regression model with binary response and $q(\mu) = -2\{\mu \log(\mu) + (1 - \mu) \log(1 - \mu)\}$ (corresponding to Bernoulli deviance loss), $V(\mu) = \mu(1 - \mu)$, $F(\mu) = \log\{\mu/(1 - \mu)\}$. In this case, $|q''(\mu)\sqrt{V(\mu)}/F'(\mu)| = 2\{\mu(1 - \mu)\}^{1/2} \leq C$ since $\mu \in [0, 1]$. Likewise, if $q(\mu) = 2\{\mu(1 - \mu)\}^{1/2}$ (for the exponential loss), then $|q''(\mu)\sqrt{V(\mu)}/F'(\mu)| = 1/2$.

Furthermore, the bound on $\psi(\cdot)$ is useful to control deviations in the Y -space, which ensures the stability of the robust-BD test if Y is arbitrarily contaminated.

Concerning the dimensionality p_n , Corollary 1 reveals the following implications. If p_n is fixed, then the asymptotic level of W_n under the ϵ -contamination is uniformly bounded for all $z \in \mathbb{R}^p$, which implies the robustness of validity of the test. This result coincides with that in Proposition 5 of [10]. When p_n diverges, the asymptotic level is still stable if the point contamination satisfies $\|\mathbf{x}_n\|_\infty \leq C_0$, where $C_0 > 0$ is an arbitrary constant. Although this condition may not be the weakest, it still covers a wide range of point mass contaminations.

3.2. Asymptotic Power of W_n under Contamination

Now, we will study the asymptotic power of W_n under a sequence of contiguous alternatives of the form

$$H_{1n} : A_n \tilde{\beta}_{n,0} - \mathbf{g}_0 = n^{-1/2} \mathbf{c}, \tag{14}$$

where $\mathbf{c} = (c_1, \dots, c_k)^T \neq \mathbf{0}$ is fixed.

Theorem 2. Assume Conditions A0–A9 and B4 in Appendix A. Suppose $p_n^6/n \rightarrow 0$ as $n \rightarrow \infty$, $\sup_n E_J(\|w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\|) \leq C$. Denote by $\beta(K_{n,\epsilon})$ the power of $W_n = n\{A_n \mathbf{T}(\mathbb{K}_n) -$

$\mathbf{g}_0\}^T (A_n \widehat{\mathbf{H}}_n^{-1} \widehat{\Omega}_n \widehat{\mathbf{H}}_n^{-1} A_n^T)^{-1} \{A_n \mathbf{T}(\mathbb{K}_n) - \mathbf{g}_0\}$ when the underlying distribution is $K_{n,\epsilon}$ in Equation (10) and by β_0 the nominal power. Under H_{1n} in Equation (14), it follows that

$$\liminf_{n \rightarrow \infty} \beta(K_{n,\epsilon}) = \beta_0 + \epsilon \nu_k B + o(\epsilon) \quad \text{as } \epsilon \rightarrow 0,$$

where

$$B = \liminf_{n \rightarrow \infty} 2\mathbf{c}^T U_n^{-1} A_n E_J \{ \text{IF}(\mathbf{Z}_n; \mathbf{T}, K_{n,0}) \},$$

with $|B| < \infty$, $\nu_k = -\frac{\partial}{\partial \delta} H_k(\eta_{1-\alpha_0}; \delta) |_{\delta=c^T U_n^{-1} \mathbf{c}}$ and $H_k(\cdot; \delta)$ and $\eta_{1-\alpha_0}$ being defined in Theorem 1.

The result for the asymptotic power is similar in spirit to that for the level. From Theorem 2, if the influence function is bounded, the asymptotic power is also bounded from below and close to the nominal power under a small amount of contamination. This means that the *robust*-BD Wald-type test enjoys the robustness of efficiency. In addition, the property of the asymptotic power can be obtained for a point mass contamination.

Corollary 2. *With the notations in Theorem 2, assume Conditions A0–A9 in Appendix A, $\sup_{\mathbf{x}_n \in \mathbb{R}^{p_n}} \|w(\mathbf{x}_n) \mathbf{x}_n\| \leq C$ and $\sup_{\mu \in \mathbb{R}} |q''(\mu) \sqrt{V(\mu)} / F'(\mu)| \leq C$.*

- (i) *If $p_n \equiv p$, $A_n \equiv A$, $\tilde{\beta}_{n,0} \equiv \tilde{\beta}_0$, $K_{n,0} \equiv K_0$ and $U_n \equiv U$ are fixed, then, for $K_{n,\epsilon} = (1 - \epsilon / \sqrt{n})K_0 + \epsilon / \sqrt{n} \Delta_{\mathbf{z}}$ with $\mathbf{z} \in \mathbb{R}^p$ a fixed point, under H_{1n} in Equation (14), it follows that*

$$\inf_{\mathbf{z} \in \mathbb{R}^p} \lim_{n \rightarrow \infty} \beta(K_{n,\epsilon}) = \beta_0 + \epsilon \nu_k B_1 + o(\epsilon) \quad \text{as } \epsilon \rightarrow 0,$$

where

$$B_1 = \inf_{\mathbf{z} \in \mathbb{R}^p} 2\mathbf{c}^T U^{-1} A \text{IF}(\mathbf{z}; \mathbf{T}, K_0),$$

with $|B_1| < \infty$.

- (ii) *If p_n diverges with $p_n^6/n \rightarrow 0$, for $K_{n,\epsilon} = (1 - \epsilon / \sqrt{n})K_{n,0} + \epsilon / \sqrt{n} \Delta_{\mathbf{z}_n}$ with $\mathbf{z}_n \in \mathbb{R}^{p_n}$ a sequence of deterministic points, then, under H_{1n} in Equation (14),*

$$\inf_{C_0 > 0} \inf_{\mathbf{z}_n \in S_{C_0}} \liminf_{n \rightarrow \infty} \beta(K_{n,\epsilon}) = \beta_0 + \epsilon \nu_k B_2 + o(\epsilon) \quad \text{as } \epsilon \rightarrow 0,$$

where $S_{C_0} = \{\mathbf{z}_n = (\mathbf{x}_n^T, \mathbf{y})^T : \|\mathbf{x}_n\|_\infty \leq C_0\}$, $C_0 > 0$ is a constant and

$$B_2 = \inf_{C_0 > 0} \inf_{\mathbf{z}_n \in S_{C_0}} \liminf_{n \rightarrow \infty} 2\mathbf{c}^T U_n^{-1} A_n \text{IF}(\mathbf{Z}_n; \mathbf{T}, K_{n,0}),$$

with $|B_2| < \infty$.

4. Simulation

Regarding the practical utility of W_n , numerical studies concerning the empirical level and power of W_n under a fixed amount of contamination have been conducted in Section 6 of [1]. To support the theoretical results in our paper, we conduct new simulations to check the robustness of validity and efficiency of W_n . Specifically, we will examine the empirical level and power of the test statistic as ϵ varies.

The *robust*-BD estimation utilizes the Huber ψ -function $\psi_c(\cdot)$ with $c = 1.345$ and the weight function $w(\mathbf{X}_n) = 1 / (1 + \|\mathbf{X}_n\|)$. Comparisons are made with the classical non-robust counterparts corresponding to using $\psi(r) = r$ and $w(\mathbf{x}_n) \equiv 1$. For each situation below, we set $n = 1000$ and conduct 400 replications.

4.1. Overdispersed Poisson Responses

Overdispersed Poisson counts Y , satisfying $\text{var}(Y|X_n = x_n) = 2m(x_n)$, are generated via a negative Binomial($m(x_n), 1/2$) distribution. Let $p_n = \lfloor 4(n^{1/5.5} - 1) \rfloor$ and $\tilde{\beta}_{n,0} = (0, 2, 0, \dots, 0)^T$, where $\lfloor \cdot \rfloor$ denotes the floor function. Generate $X_{ni} = (X_{i,1}, \dots, X_{i,p_n})^T$ by $X_{i,j} \stackrel{\text{i.i.d.}}{\sim} \text{Unif}[-0.5, 0.5]$. The log link function is considered and the (negative) quasi-likelihood is utilized as the BD, generated by the q -function in Equation (3) with $V(\mu) = \mu$. The estimator and test statistic are calculated by assuming Y follows Poisson distribution.

The data are contaminated by $X_{i, \text{mod}(i, p_n - 1) + 1}^* = 3\text{sign}(U_i - 0.5)$ and $Y_i^* = Y_i I(Y_i > 20) + 20I(Y_i \leq 20)$ for $i = 1, \dots, k$, with $k \in \{2, 4, 6, 8, 10, 12, 14, 16\}$ the number of contaminated data points, where $\text{mod}(a, b)$ is the modulo operation “ a modulo b ” and $\{U_i\} \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, 1)$. Then, the proportion of contaminated data, k/n , is equal to ϵ / \sqrt{n} as in Equation (10), which implies $\epsilon = k / \sqrt{n}$.

Consider the null hypothesis $H_0 : A_n \tilde{\beta}_{n,0} = 0$ with $A_n = (0, 0, 0, 1, 0, \dots, 0)$. Figure 1 plots the empirical level of W_n versus ϵ . We observe that the asymptotic nominal level 0.05 is approximately retained by the robust Wald-type test. On the other hand, under contaminations, the non-robust Wald-type test breaks in level, showing high sensitivity to the presence of outliers.

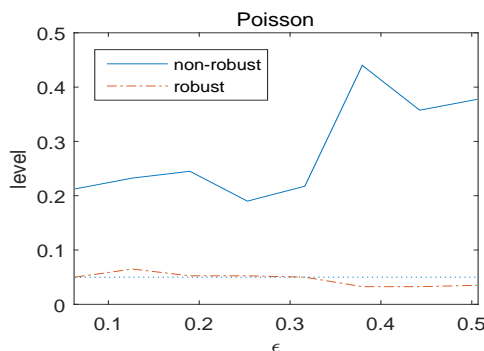


Figure 1. Observed level of W_n versus ϵ for overdispersed Poisson responses. The dotted line indicates the 5% significance level.

To assess the stability of the power of the test, we generate the original data from the true model, but with the true parameter $\tilde{\beta}_{n,0}$ replaced by $\tilde{\beta}_n = \tilde{\beta}_{n,0} + \delta c$ with $\delta \in \{-0.4, 0.4, -0.6, 0.6\}$ and $c = (1, \dots, 1)^T$ a vector of ones. Figure 2 plots the empirical rejection rates of the null model, which implies that the robust Wald-type test has sufficiently large power to detect the alternative hypothesis. In addition, the power of the robust method is generally larger than that of the non-robust method.

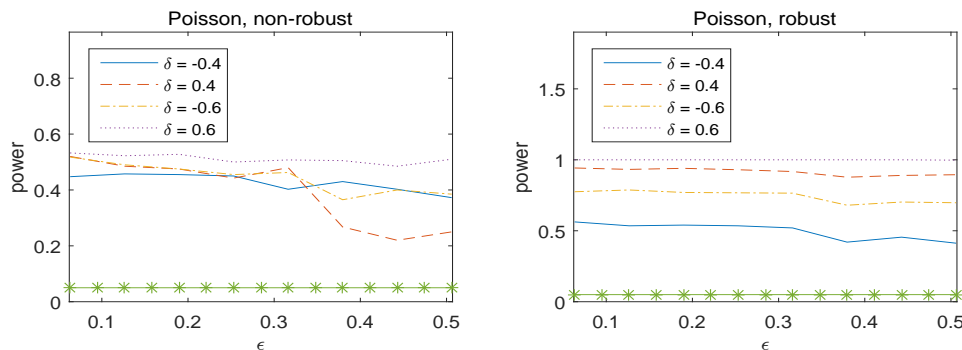


Figure 2. Observed power of W_n versus ϵ for overdispersed Poisson responses. The statistics in the left panel correspond to non-robust method and those in the right panel are for robust method. The asterisk line indicates the 5% significance level.

4.2. Bernoulli Responses

We generate data with two classes from the model, $Y|X_n = x_n \sim \text{Bernoulli}\{m(x_n)\}$, where $\text{logit}\{m(x_n)\} = \tilde{x}_n^T \tilde{\beta}_{n,0}$. Let $p_n = 2$, $\tilde{\beta}_{n,0} = (0, 1, 1)^T$ and $X_{ni} \stackrel{\text{i.i.d.}}{\sim} N(0, \mathbf{I}_{p_n})$. The null hypothesis is $H_0 : \tilde{\beta}_{n,0} = (0, 1, 1)^T$. Both the deviance loss and the exponential loss are employed as the BD. We contaminate the data by setting $X_{i,1}^* = 2 + i/8$ and $Y_i^* = 0$ for $i = 1, \dots, k$ with $k \in \{2, 4, 6, 8, 10, 12, 14, 16\}$. To investigate the robustness of validity of W_n , we plot the observed level versus ϵ in Figure 3. We find that the level of the non-robust method diverges fast as ϵ increases. It's also clear that the empirical level of the robust method is close to the nominal level when ϵ is small and increases slightly with ϵ , which coincides with our results in Theorem 1.

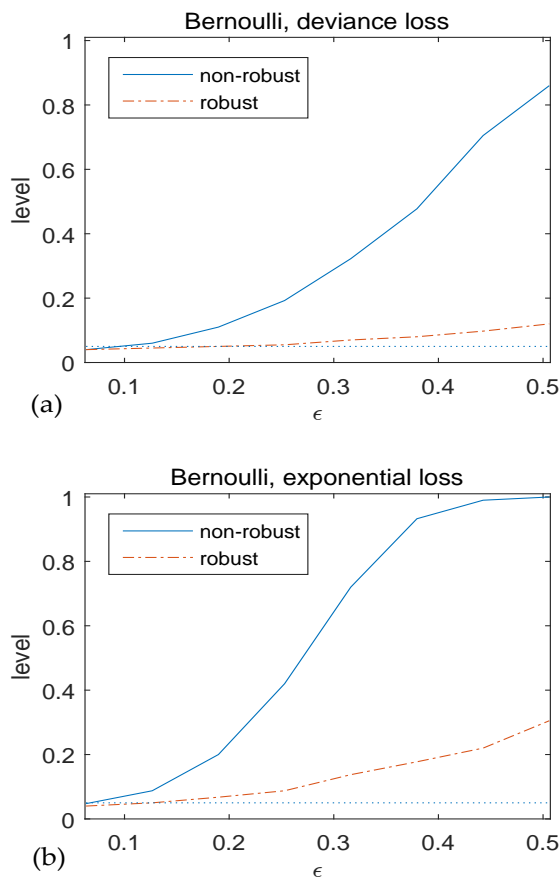


Figure 3. Observed level of W_n versus ϵ for Bernoulli responses. The statistics in (a) use deviance loss and those in (b) use exponential loss. The dotted line indicates the 5% significance level.

To assess the stability of the power of W_n , we generate the original data from the true model, but with the true parameter $\tilde{\beta}_{n,0}$ replaced by $\tilde{\beta}_n = \tilde{\beta}_{n,0} + \delta c$ with $\delta \in \{-0.1, 0.2, -0.3, 0.4\}$ and $c = (1, \dots, 1)^T$ a vector of ones. Figure 4 plots the power of the Wald-type test versus ϵ , which implies that the robust method has sufficiently large power, and hence supports the theoretical results in Theorem 2.

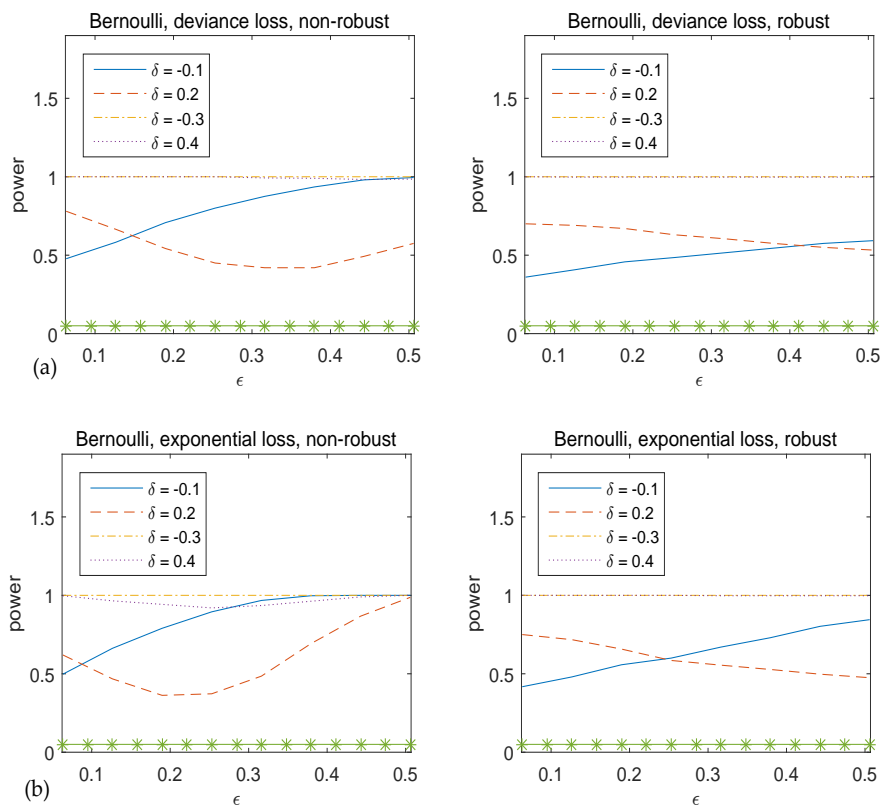


Figure 4. Observed power of W_n versus ϵ for Bernoulli responses. The top panels correspond to deviance loss while the bottom panels are for exponential loss. The statistics in the left panels are calculated using non-robust method and those in the right panels are from robust method. The asterisk line indicates the 5% significance level.

Acknowledgments: We thank the two referees for insightful comments and suggestions. Chunming Zhang’s research is supported by the U.S. NSF Grants DMS-1712418, DMS-1521761, the Wisconsin Alumni Research Foundation and the National Natural Science Foundation of China, grants 11690014. Xiao Guo’s research is supported by the Fundamental Research Funds for the Central Universities and the National Natural Science Foundation of China, grants 11601500, 11671374 and 11771418.

Author Contributions: Chunming Zhang conceived and designed the experiments; Xiao Guo performed the experiments; Xiao Guo analyzed the data; Chunming Zhang contributed to analysis tools; Chunming Zhang and Xiao Guo wrote the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Conditions and Proofs of Main Results

We first introduce some necessary notations used in the proof.

Notations. For arbitrary distributions K and K' of Z_n , define

$$\begin{aligned} \Omega_{n,K,T(K')} &= E_K\{p_1^2(Y; \tilde{X}_n^T T(K')) w^2(X_n) \tilde{X}_n \tilde{X}_n^T\}, \\ \mathbf{H}_{n,K,T(K')} &= E_K\{p_2(Y; \tilde{X}_n^T T(K')) w(X_n) \tilde{X}_n \tilde{X}_n^T\}. \end{aligned}$$

Therefore, $\Omega_n = \Omega_{n,K_{n,0},\tilde{\beta}_{n,0}}$, $\mathbf{H}_n = \mathbf{H}_{n,K_{n,0},\tilde{\beta}_{n,0}}$, $\hat{\Omega}_n = \Omega_{n,\mathbb{K}_n,T(\mathbb{K}_n)}$ and $\hat{\mathbf{H}}_n = \mathbf{H}_{n,\mathbb{K}_n,T(\mathbb{K}_n)}$. For notational simplicity, let $\Omega_{n,\epsilon} = \Omega_{n,K_{n,\epsilon},T(K_{n,\epsilon})}$ and $\mathbf{H}_{n,\epsilon} = \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}$.

Define the following matrices,

$$\begin{aligned} U(K_{n,\epsilon}) &= A_n \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1} A_n^T, \\ U(\mathbb{K}_n) &= A_n \hat{\mathbf{H}}_n^{-1} \hat{\Omega}_n \hat{\mathbf{H}}_n^{-1} A_n^T. \end{aligned}$$

The following conditions are needed in the proof, which are adopted from [1].

Condition A.

- A0. $\sup_{n \geq 1} \|\tilde{\beta}_{n,0}\|_1 < \infty$.
- A1. $w(\cdot)$ is a bounded function. Assume that $\psi(r)$ is a bounded, odd function, and twice differentiable, such that $\psi'(r)$, $\psi'(r)r$, $\psi''(r)$, $\psi''(r)r$ and $\psi''(r)r^2$ are bounded; $V(\cdot) > 0$, $V^{(2)}$ is continuous.
- A2. $q^{(4)}(\cdot)$ is continuous, and $q^{(2)}(\cdot) < 0$. $G_1^{(3)}$ is continuous.
- A3. $F(\cdot)$ is monotone and a bijection, $F^{(3)}(\cdot)$ is continuous, and $F^{(1)}(\cdot) \neq 0$.
- A4. $\|\mathbf{X}_n\|_\infty \leq C$ almost surely if the underlying distribution is $K_{n,0}$.
- A5. $E_{K_{n,0}}(\tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T)$ exists and is nonsingular.
- A6. There is a large enough open subset of \mathbb{R}^{p_n+1} which contains $\tilde{\beta}_{n,0}$, such that $F^{-1}(\tilde{\mathbf{x}}_n^T \tilde{\beta})$ is bounded for all $\tilde{\beta}$ in the subset and all $\tilde{\mathbf{x}}_n$ such that $\|\tilde{\mathbf{x}}_n\|_\infty \leq C$, where $C > 0$ is a large enough constant.
- A7. \mathbf{H}_n is positive definite, with eigenvalues uniformly bounded away from 0.
- A8. Ω_n is positive definite, with eigenvalues uniformly bounded away from 0.
- A9. $\|\mathbf{H}_n^{-1} \Omega_n\|$ is bounded away from ∞ .

Condition B.

- B4. $\|\mathbf{X}_n\|_\infty \leq C$ almost surely if the underlying distribution is J .

The following Lemmas A1–A9 are needed to prove the main theoretical results in this paper.

Lemma A1 ($\|T(K_{n,\epsilon}) - \tilde{\beta}_{n,0}\|$). Assume Conditions A0–A7 and B4. For $K_{n,\epsilon}$ in Equation (10), $\ell_K(\cdot)$ in Equation (11) and $T(\cdot)$ in Equation (12), if $p_n^4/n \rightarrow 0$ as $n \rightarrow \infty$, then $T(K_{n,\epsilon})$ is a local minimizer of $\ell_{K_{n,\epsilon}}(\tilde{\beta})$ such that $\|T(K_{n,\epsilon}) - \tilde{\beta}_{n,0}\| = O(\sqrt{p_n/n})$. Furthermore, $T(K_{n,\epsilon})$ is unique.

Proof. We follow the idea of the proof in [25]. Let $r_n = \sqrt{p_n/n}$ and $\tilde{\mathbf{u}}_n = (u_0, u_1, \dots, u_{p_n})^T \in \mathbb{R}^{p_n+1}$. First, we show that there exists a sufficiently large constant C such that, for large n , we have

$$\inf_{\|\tilde{\mathbf{u}}_n\|=C} \ell_{K_{n,\epsilon}}(\tilde{\beta}_{n,0} + r_n \tilde{\mathbf{u}}_n) > \ell_{K_{n,\epsilon}}(\tilde{\beta}_{n,0}). \tag{A1}$$

To show Equation (A1), consider

$$\begin{aligned} \ell_{K_{n,\epsilon}}(\tilde{\beta}_{n,0} + r_n \tilde{\mathbf{u}}_n) - \ell_{K_{n,\epsilon}}(\tilde{\beta}_{n,0}) &= E_{K_{n,\epsilon}} \{ \rho_q(Y, F^{-1}(\tilde{\mathbf{X}}_n^T \tilde{\beta}_{n,0} + r_n \tilde{\mathbf{X}}_n^T \tilde{\mathbf{u}}_n)) w(\mathbf{X}_n) \\ &\quad - \rho_q(Y, F^{-1}(\tilde{\mathbf{X}}_n^T \tilde{\beta}_{n,0})) w(\mathbf{X}_n) \} \\ &\equiv I_1, \end{aligned}$$

where $\|\tilde{\mathbf{u}}_n\| = C$.

By Taylor expansion,

$$I_1 = I_{1,1} + I_{1,2} + I_{1,3}, \tag{A2}$$

where

$$\begin{aligned} I_{1,1} &= r_n E_{K_{n,\epsilon}} \{ p_1(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n^T \tilde{\mathbf{u}}_n \}, \\ I_{1,2} &= r_n^2 / 2 E_{K_{n,\epsilon}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}_{n,0}) w(\mathbf{X}_n) (\tilde{\mathbf{X}}_n^T \tilde{\mathbf{u}}_n)^2 \}, \\ I_{1,3} &= r_n^3 / 6 E_{K_{n,\epsilon}} \{ p_3(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}_{n,0}^*) w(\mathbf{X}_n) (\tilde{\mathbf{X}}_n^T \tilde{\mathbf{u}}_n)^3 \}, \end{aligned}$$

for $\tilde{\beta}_{n,0}^*$ located between $\tilde{\beta}_{n,0}$ and $\tilde{\beta}_{n,0} + r_n \tilde{\mathbf{u}}_n$. Hence

$$|I_{1,1}| \leq r_n \| E_{K_{n,\epsilon}} \{ p_1(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \} \| \|\tilde{\mathbf{u}}_n\|$$

$$\begin{aligned}
 &= r_n \frac{\epsilon}{\sqrt{n}} \| E_J \{ p_1(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \} \| \| \tilde{\mathbf{u}}_n \| \\
 &\leq Cr_n \sqrt{p_n/n} \| \tilde{\mathbf{u}}_n \|,
 \end{aligned}$$

since $\| E_J \{ p_1(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \} \| = O(\sqrt{p_n})$ and $E_{K_{n,0}} \{ p_1(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \} = \mathbf{0}$. For $I_{1,2}$ in Equation (A2),

$$\begin{aligned}
 I_{1,2} &= \frac{r_n^2}{2} E_{K_{n,0}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) (\tilde{\mathbf{X}}_n^T \tilde{\mathbf{u}}_n)^2 \} \\
 &\quad + \frac{r_n^2}{2} [E_{K_{n,\epsilon}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) (\tilde{\mathbf{X}}_n^T \tilde{\mathbf{u}}_n)^2 \} - E_{K_{n,0}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) (\tilde{\mathbf{X}}_n^T \tilde{\mathbf{u}}_n)^2 \}] \\
 &\equiv I_{1,2,1} + I_{1,2,2},
 \end{aligned}$$

where $I_{1,2,1} = 2^{-1} r_n^2 \tilde{\mathbf{u}}_n^T \mathbf{H}_n \tilde{\mathbf{u}}_n$. Meanwhile, we have

$$\begin{aligned}
 &|I_{1,2,2}| \\
 &\leq r_n^2 \| E_{K_{n,\epsilon}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \} - E_{K_{n,0}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \} \|_F \| \tilde{\mathbf{u}}_n \|^2 \\
 &= r_n^2 \frac{\epsilon}{\sqrt{n}} \| E_J \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \} - E_{K_{n,0}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \} \|_F \| \tilde{\mathbf{u}}_n \|^2 \\
 &\leq Cr_n^2 p_n \| \tilde{\mathbf{u}}_n \|^2 / \sqrt{n},
 \end{aligned}$$

where $\| E_J \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \} \|_F = O(p_n)$ and $\| E_{K_{n,0}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \} \|_F = O(p_n)$. Thus,

$$I_{1,2} = 2^{-1} r_n^2 \tilde{\mathbf{u}}_n^T \mathbf{H}_n \tilde{\mathbf{u}}_n + O(r_n^2 p_n / \sqrt{n}) \| \tilde{\mathbf{u}}_n \|^2. \tag{A3}$$

For $I_{1,3}$ in Equation (A2), we observe that

$$|I_{1,3}| \leq Cr_n^3 E_{K_{n,\epsilon}} \{ |p_3(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}^*) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n^T \tilde{\mathbf{u}}_n|^3 \} = O(r_n^3 p_n^{3/2}) \| \tilde{\mathbf{u}}_n \|^3.$$

We can choose some large C such that $I_{1,1}$, $I_{1,2,2}$ and $I_{1,3}$ are all dominated by the first term of $I_{1,2}$ in Equation (A3), which is positive by the eigenvalue assumption. This implies Equation (A1). Therefore, there exists a local minimizer of $\ell_{K_{n,\epsilon}}(\tilde{\boldsymbol{\beta}})$ in the $\sqrt{p_n/n}$ neighborhood of $\tilde{\boldsymbol{\beta}}_{n,0}$, and denote this minimizer by $\tilde{\boldsymbol{\beta}}_{n,\epsilon}$.

Next, we show that the local minimizer $\tilde{\boldsymbol{\beta}}_{n,\epsilon}$ of $\ell_{K_{n,\epsilon}}(\tilde{\boldsymbol{\beta}})$ is unique in the $\sqrt{p_n/n}$ neighborhood of $\tilde{\boldsymbol{\beta}}_{n,0}$. For all $\tilde{\boldsymbol{\beta}}$ such that $\| \tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}_{n,0} \| = O(n^{-1/4} p_n^{-1/2})$,

$$\begin{aligned}
 E_{K_{n,\epsilon}} \left\| \frac{\partial}{\partial \tilde{\boldsymbol{\beta}}} \rho_q(Y, F^{-1}(\tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}})) w(\mathbf{X}_n) \right\| &= E_{K_{n,\epsilon}} \| p_1(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \| \leq C \sqrt{p_n} \\
 E_{K_{n,\epsilon}} \left\| \frac{\partial^2}{\partial \tilde{\boldsymbol{\beta}}^2} \rho_q(Y, F^{-1}(\tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}})) w(\mathbf{X}_n) \right\| &= E_{K_{n,\epsilon}} \| p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \| \leq Cp_n
 \end{aligned}$$

and hence,

$$\begin{aligned}
 \frac{\partial}{\partial \tilde{\boldsymbol{\beta}}} E_{K_{n,\epsilon}} \{ \rho_q(Y, F^{-1}(\tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}})) w(\mathbf{X}_n) \} &= E_{K_{n,\epsilon}} \left\{ \frac{\partial}{\partial \tilde{\boldsymbol{\beta}}} \rho_q(Y, F^{-1}(\tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}})) w(\mathbf{X}_n) \right\} \\
 \frac{\partial^2}{\partial \tilde{\boldsymbol{\beta}}^2} E_{K_{n,\epsilon}} \{ \rho_q(Y, F^{-1}(\tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}})) w(\mathbf{X}_n) \} &= E_{K_{n,\epsilon}} \left\{ \frac{\partial^2}{\partial \tilde{\boldsymbol{\beta}}^2} \rho_q(Y, F^{-1}(\tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}})) w(\mathbf{X}_n) \right\}.
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 &\frac{\partial^2}{\partial \tilde{\boldsymbol{\beta}}^2} E_{K_{n,\epsilon}} \{ \rho_q(Y, F^{-1}(\tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}})) w(\mathbf{X}_n) \} \\
 &= E_{K_{n,\epsilon}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \}
 \end{aligned}$$

$$\begin{aligned}
 &= E_{K_{n,0}} \{p_2(Y; \tilde{X}_n^T \tilde{\beta}_{n,0}^T) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T\} \\
 &\quad + E_{K_{n,0}} [\{p_2(Y; \tilde{X}_n^T \tilde{\beta}) - p_2(Y; \tilde{X}_n^T \tilde{\beta}_{n,0}^T)\} w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T] \\
 &\quad + [E_{K_{n,\epsilon}} \{p_2(Y; \tilde{X}_n^T \tilde{\beta}) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T\} - E_{K_{n,0}} \{p_2(Y; \tilde{X}_n^T \tilde{\beta}) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T\}] \\
 &= I_1^* + I_2^* + I_3^*.
 \end{aligned}$$

We know that the minimum eigenvalues of I_1^* are uniformly bounded away from 0,

$$\begin{aligned}
 \|I_2^*\| &= \|E_{K_{n,0}} \{p_3(Y; \tilde{X}_n^T \tilde{\beta}^{**}) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T \tilde{X}_n^T (\tilde{\beta} - \tilde{\beta}_{n,0})\}\| \leq Cp_n/n^{1/4} = o(1) \\
 \|I_3^*\| &\leq \epsilon/\sqrt{n} [\|E_{K_{n,0}} \{p_2(Y; \tilde{X}_n^T \tilde{\beta}) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T\}\| + \|E_J \{p_2(Y; \tilde{X}_n^T \tilde{\beta}) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T\}\|] \\
 &\leq Cp_n/\sqrt{n} = o(1).
 \end{aligned}$$

Hence, for n large enough, $\frac{\partial^2}{\partial \tilde{\beta}^2} E_{K_{n,\epsilon}} \{\rho_q(Y, F^{-1}(\tilde{X}_n^T \tilde{\beta})) w(\mathbf{X}_n)\}$ is positive definite for all $\tilde{\beta}$ such that $\|\tilde{\beta} - \tilde{\beta}_{n,0}\| = O(n^{-1/4} p_n^{-1/2})$. Therefore, there exists a unique minimizer of $\ell_{K_{n,\epsilon}}(\tilde{\beta})$ in the $n^{-1/4} p_n^{-1/2}$ neighborhood of $\tilde{\beta}_{n,0}$ which covers $\tilde{\beta}_{n,\epsilon}$. From

$$\begin{aligned}
 \mathbf{0} &= \frac{\partial}{\partial \tilde{\beta}} E_{K_{n,\epsilon}} \{\rho_q(Y, F^{-1}(\tilde{X}_n^T \tilde{\beta})) w(\mathbf{X}_n)\} \Big|_{\tilde{\beta}=\tilde{\beta}_{n,\epsilon}} = E_{K_{n,\epsilon}} \left\{ \frac{\partial}{\partial \tilde{\beta}} \rho_q(Y, F^{-1}(\tilde{X}_n^T \tilde{\beta})) \Big|_{\tilde{\beta}=\tilde{\beta}_{n,\epsilon}} w(\mathbf{X}_n) \right\} \\
 &= E_{K_{n,\epsilon}} \{p_1(Y; \tilde{X}_n^T \tilde{\beta}_{n,\epsilon}) w(\mathbf{X}_n) \tilde{X}_n\},
 \end{aligned}$$

we know $T(K_{n,\epsilon}) = \tilde{\beta}_{n,\epsilon}$. From the definition of $T(\cdot)$, it's easy to see that $T(K_{n,\epsilon})$ is unique. \square

Lemma A2 ($\|T(\mathbb{K}_n) - T(K_{n,\epsilon})\|$). Assume Conditions A0–A7 and B4. For $K_{n,\epsilon}$ in Equation (10), $\ell_{\mathbb{K}}(\cdot)$ in Equation (11) and $T(\cdot)$ in Equation (12), if $p_n^4/n \rightarrow 0$ as $n \rightarrow \infty$ and the distribution of (\mathbf{X}_n, Y) is $K_{n,\epsilon}$, then there exists a unique local minimizer $\hat{\beta}_n$ of $\ell_{\mathbb{K}_n}(\tilde{\beta})$ such that $\|\hat{\beta}_n - T(K_{n,\epsilon})\| = O_P(\sqrt{p_n/n})$. Furthermore, $\|\hat{\beta}_n - \tilde{\beta}_{n,0}\| = O_P(\sqrt{p_n/n})$ and $T(\mathbb{K}_n) = \hat{\beta}_n$.

Proof. Let $r_n = \sqrt{p_n/n}$ and $\tilde{\mathbf{u}}_n = (u_0, u_1, \dots, u_{p_n})^T \in \mathbb{R}^{p_n+1}$. To show the existence of the estimator, it suffices to show that for any given $\kappa > 0$, there exists a sufficiently large constant C_κ such that, for large n we have

$$P \left\{ \inf_{\|\tilde{\mathbf{u}}_n\|=C_\kappa} \ell_{\mathbb{K}_n}(T(K_{n,\epsilon}) + r_n \tilde{\mathbf{u}}_n) > \ell_{\mathbb{K}_n}(T(K_{n,\epsilon})) \right\} \geq 1 - \kappa. \tag{A4}$$

This implies that with probability at least $1 - \kappa$, there exists a local minimizer $\hat{\beta}_n$ of $\ell_{\mathbb{K}_n}(\tilde{\beta})$ in the ball $\{T(K_{n,\epsilon}) + r_n \tilde{\mathbf{u}}_n : \|\tilde{\mathbf{u}}_n\| \leq C_\kappa\}$. To show Equation (A4), consider

$$\begin{aligned}
 \ell_{\mathbb{K}_n}(T(K_{n,\epsilon}) + r_n \tilde{\mathbf{u}}_n) - \ell_{\mathbb{K}_n}(T(K_{n,\epsilon})) &= \frac{1}{n} \sum_{i=1}^n \{ \rho_q(Y_i, F^{-1}(\tilde{X}_{ni}^T (T(K_{n,\epsilon}) + r_n \tilde{\mathbf{u}}_n))) w(\mathbf{X}_{ni}) \\
 &\quad - \rho_q(Y_i, F^{-1}(\tilde{X}_{ni}^T T(K_{n,\epsilon}))) w(\mathbf{X}_{ni}) \} \\
 &\equiv I_1,
 \end{aligned}$$

where $\|\tilde{\mathbf{u}}_n\| = C_\kappa$.

By Taylor expansion,

$$I_1 = I_{1,1} + I_{1,2} + I_{1,3}, \tag{A5}$$

where

$$I_{1,1} = r_n/n \sum_{i=1}^n p_1(Y_i; \tilde{X}_{ni}^T T(K_{n,\epsilon})) w(\mathbf{X}_{ni}) \tilde{X}_{ni}^T \tilde{\mathbf{u}}_n,$$

$$\begin{aligned}
 I_{1,2} &= r_n^2 / (2n) \sum_{i=1}^n p_2(Y_i; \tilde{\mathbf{X}}_{ni}^T \mathbf{T}(K_{n,\epsilon})) w(\mathbf{X}_{ni}) (\tilde{\mathbf{X}}_{ni}^T \tilde{\mathbf{u}}_n)^2, \\
 I_{1,3} &= r_n^3 / (6n) \sum_{i=1}^n p_3(Y_i; \tilde{\mathbf{X}}_{ni}^T \tilde{\boldsymbol{\beta}}_n^*) w(\mathbf{X}_{ni}) (\tilde{\mathbf{X}}_{ni}^T \tilde{\mathbf{u}}_n)^3
 \end{aligned}$$

for $\tilde{\boldsymbol{\beta}}_n^*$ located between $\mathbf{T}(K_{n,\epsilon})$ and $\mathbf{T}(K_{n,\epsilon}) + r_n \tilde{\mathbf{u}}_n$.

Since $\|\mathbf{T}(K_{n,\epsilon}) - \tilde{\boldsymbol{\beta}}_{n,0}\| = O(\sqrt{p_n/n}) = o(1)$, the large open set considered in Condition A6 contains $\mathbf{T}(K_{n,\epsilon})$ when n is large enough, say $n \geq N$ where N is a positive constant. Therefore, for any fixed $n \geq N$, there exists a bounded open subset of \mathbb{R}^{p_n+1} containing $\mathbf{T}(K_{n,\epsilon})$ such that for all $\tilde{\boldsymbol{\beta}}$ in this set, $\|p_1(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n\| \leq C \|\tilde{\mathbf{X}}_n\|$ which is integrable with respect to $K_{n,\epsilon}$, where C is a positive constant. Thus, for $n \geq N$,

$$\mathbf{0} = \frac{\partial}{\partial \tilde{\boldsymbol{\beta}}} E_{K_{n,\epsilon}} \{ \rho_q(Y, F^{-1}(\tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}})) w(\mathbf{X}_n) \} \Big|_{\tilde{\boldsymbol{\beta}}=\mathbf{T}(K_{n,\epsilon})} = E_{K_{n,\epsilon}} \{ p_1(Y; \tilde{\mathbf{X}}_n^T \mathbf{T}(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \}. \tag{A6}$$

Hence,

$$|I_{1,1}| \leq r_n \left\| \frac{1}{n} \sum_{i=1}^n p_1(Y_i; \tilde{\mathbf{X}}_{ni}^T \mathbf{T}(K_{n,\epsilon})) w(\mathbf{X}_{ni}) \tilde{\mathbf{X}}_{ni} \right\| \|\tilde{\mathbf{u}}_n\| = O_P(r_n \sqrt{p_n/n}) \|\tilde{\mathbf{u}}_n\|.$$

For $I_{1,2}$ in Equation (A5),

$$\begin{aligned}
 I_{1,2} &= \frac{r_n^2}{2n} \sum_{i=1}^n E_{K_{n,\epsilon}} \{ p_2(Y_i; \tilde{\mathbf{X}}_{ni}^T \mathbf{T}(K_{n,\epsilon})) w(\mathbf{X}_{ni}) (\tilde{\mathbf{X}}_{ni}^T \tilde{\mathbf{u}}_n)^2 \} \\
 &\quad + \frac{r_n^2}{2n} \sum_{i=1}^n [p_2(Y_i; \tilde{\mathbf{X}}_{ni}^T \mathbf{T}(K_{n,\epsilon})) w(\mathbf{X}_{ni}) (\tilde{\mathbf{X}}_{ni}^T \tilde{\mathbf{u}}_n)^2 \\
 &\quad - E_{K_{n,\epsilon}} \{ p_2(Y_i; \tilde{\mathbf{X}}_{ni}^T \mathbf{T}(K_{n,\epsilon})) w(\mathbf{X}_{ni}) (\tilde{\mathbf{X}}_{ni}^T \tilde{\mathbf{u}}_n)^2 \}] \\
 &\equiv I_{1,2,1} + I_{1,2,2},
 \end{aligned}$$

where $I_{1,2,1} = 2^{-1} r_n^2 \tilde{\mathbf{u}}_n^T \mathbf{H}_{n,\epsilon} \tilde{\mathbf{u}}_n$. Meanwhile, we have

$$\begin{aligned}
 |I_{1,2,2}| &\leq \frac{r_n^2}{2} \left\| \frac{1}{n} \sum_{i=1}^n [p_2(Y_i; \tilde{\mathbf{X}}_{ni}^T \mathbf{T}(K_{n,\epsilon})) w(\mathbf{X}_{ni}) \tilde{\mathbf{X}}_{ni} \tilde{\mathbf{X}}_{ni}^T \right. \\
 &\quad \left. - E_{K_{n,\epsilon}} \{ p_2(Y_i; \tilde{\mathbf{X}}_{ni}^T \mathbf{T}(K_{n,\epsilon})) w(\mathbf{X}_{ni}) \tilde{\mathbf{X}}_{ni} \tilde{\mathbf{X}}_{ni}^T \} \right\|_F \|\tilde{\mathbf{u}}_n\|^2 \\
 &= r_n^2 O_P(p_n / \sqrt{n}) \|\tilde{\mathbf{u}}_n\|^2.
 \end{aligned}$$

Thus,

$$I_{1,2} = 2^{-1} r_n^2 \tilde{\mathbf{u}}_n^T \mathbf{H}_{n,\epsilon} \tilde{\mathbf{u}}_n + O_P(r_n^2 p_n / \sqrt{n}) \|\tilde{\mathbf{u}}_n\|^2. \tag{A7}$$

For $I_{1,3}$ in Equation (A5), we observe that

$$|I_{1,3}| \leq C r_n^3 \frac{1}{n} \sum_{i=1}^n |p_3(Y_i; \tilde{\mathbf{X}}_{ni}^T \tilde{\boldsymbol{\beta}}_n^*)| w(\mathbf{X}_{ni}) |\tilde{\mathbf{X}}_{ni}^T \tilde{\mathbf{u}}_n|^3 = O_P(r_n^3 p_n^{3/2}) \|\tilde{\mathbf{u}}_n\|^3.$$

We will show that the minimum eigenvalue of $\mathbf{H}_{n,\epsilon}$ is uniformly bounded away from 0. $\mathbf{H}_{n,\epsilon} = (1 - \epsilon / \sqrt{n}) \mathbf{H}_{n,K_{n,0},\mathbf{T}(K_{n,\epsilon})} + \epsilon / \sqrt{n} \mathbf{H}_{n,J,\mathbf{T}(K_{n,\epsilon})}$. Note

$$\begin{aligned}
 &\| \mathbf{H}_{n,K_{n,0},\mathbf{T}(K_{n,\epsilon})} - \mathbf{H}_n \| \\
 &= \| E_{K_{n,0}} [\{ p_2(Y; \tilde{\mathbf{X}}_n^T \mathbf{T}(K_{n,\epsilon})) - p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) \} w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T] \| \\
 &= \| E_{K_{n,0}} [p_3(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_n^{**}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \tilde{\mathbf{X}}_n^T \{ \mathbf{T}(K_{n,\epsilon}) - \tilde{\boldsymbol{\beta}}_{n,0} \}] \| = O(p_n^2 / \sqrt{n}).
 \end{aligned}$$

Since the eigenvalues of \mathbf{H}_n are uniformly bounded away from 0, so are those of $\mathbf{H}_{n,K_{n,0},T(K_{n,\epsilon})}$ and $\mathbf{H}_{n,\epsilon}$.

We can choose some large C_κ such that $I_{1,1}$ and $I_{1,3}$ are both dominated by the first term of $I_{1,2}$ in Equation (A7), which is positive by the eigenvalue assumption. This implies Equation (A4).

Next we show the uniqueness of $\hat{\beta}$. For all $\tilde{\beta}$ such that $\|\tilde{\beta} - T(K_{n,\epsilon})\| = O(n^{-1/4}p_n^{-1/2})$,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n p_2(Y_i; \tilde{\mathbf{X}}_{ni}^T \tilde{\beta}) w(\mathbf{X}_{ni}) \tilde{\mathbf{X}}_{ni} \tilde{\mathbf{X}}_{ni}^T \\ = & E_{K_{n,0}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \mathbf{X}_n^T \} \\ & + E_{K_{n,0}} \{ [p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}) - p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}_{n,0})] w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \mathbf{X}_n^T \} \\ & + [E_{K_{n,\epsilon}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \mathbf{X}_n^T \} - E_{K_{n,0}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \mathbf{X}_n^T \}] \\ & + \left[\frac{1}{n} \sum_{i=1}^n p_2(Y_i; \tilde{\mathbf{X}}_{ni}^T \tilde{\beta}) w(\mathbf{X}_{ni}) \tilde{\mathbf{X}}_{ni} \tilde{\mathbf{X}}_{ni}^T - E_{K_{n,\epsilon}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \mathbf{X}_n^T \} \right] \\ = & I_1^* + I_2^* + I_3^* + I_4^*. \end{aligned}$$

We know that the minimum eigenvalues of I_1^* are uniformly bounded away from 0. Following the proof of Lemma A1, we have $\|I_2^*\| = o(1)$ and $\|I_3^*\| = o(1)$. It's easy to see $\|I_4^*\| = O_P(p_n/\sqrt{n})$.

Hence, for n large enough, $\frac{\partial^2}{\partial \beta^2} \ell_{\mathbb{K}_n}(\tilde{\beta})$ is positive definite with high probability for all $\tilde{\beta}$ such that $\|\tilde{\beta} - \tilde{\beta}_{n,0}\| = O(n^{-1/4}p_n^{-1/2})$. Therefore, there exists a unique minimizer of $\ell_{\mathbb{K}_n}(\tilde{\beta})$ in the $n^{-1/4}p_n^{-1/2}$ neighborhood of $T(K_{n,\epsilon})$ which covers $\hat{\beta}$. \square

Lemma A3 ($\|A_n\{T(K_{n,\epsilon}) - \tilde{\beta}_{n,0}\}\|$). Assume Conditions A0–A7 and B4. For $K_{n,\epsilon}$ in Equation (10) and $T(\cdot)$ in Equation (12), if $p_n^5/n \rightarrow 0$ as $n \rightarrow \infty$, the distribution of (\mathbf{X}_n, Y) is $K_{n,\epsilon}$ and $E_J(\|w(\mathbf{X}_n)\mathbf{X}_n\|) \leq C$, then

$$\sqrt{n}A_n\{T(K_{n,\epsilon}) - \tilde{\beta}_{n,0}\} = O(1),$$

where A_n is any given $k \times (p_n + 1)$ matrix such that $A_n A_n^T \rightarrow \mathbb{G}$, with \mathbb{G} being a $k \times k$ positive-definite matrix and k is a fixed integer.

Proof. Taylor's expansion yields

$$\begin{aligned} \mathbf{0} &= E_{K_{n,\epsilon}} \{ p_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \} \\ &= E_{K_{n,\epsilon}} \{ p_1(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \} \\ &\quad + E_{K_{n,\epsilon}} \{ p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \} \{ T(K_{n,\epsilon}) - \tilde{\beta}_{n,0} \} \\ &\quad + 1/2 E_{K_{n,\epsilon}} \{ p_3(Y; \tilde{\mathbf{X}}_n^T \tilde{\beta}_{n,0}^*) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n [\tilde{\mathbf{X}}_n^T \{ T(K_{n,\epsilon}) - \tilde{\beta}_{n,0} \}]^2 \} \\ &= I_1 + I_2 \{ T(K_{n,\epsilon}) - \tilde{\beta}_{n,0} \} + I_3, \end{aligned}$$

where $\tilde{\beta}_{n,0}^*$ lies between $T(K_{n,\epsilon})$ and $\tilde{\beta}_{n,0}$. Below, we will show

$$\|I_1\| = O(1/\sqrt{n}), \quad \|I_2 - \mathbf{H}_n\| = O(p_n/\sqrt{n}), \quad \|I_3\| = O(p_n^{5/2}/n).$$

First, $\|I_1\| = \epsilon/\sqrt{n} \|E_J\{p_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n\}\| \leq C\epsilon/\sqrt{n} E_J(\|w(\mathbf{X}_n)\mathbf{X}_n\|) = O(1/\sqrt{n})$. Following the proof of I_3^* in Lemma A1, $\|I_2 - \mathbf{H}_n\| = O(p_n/\sqrt{n})$. Since $\|T(K_{n,\epsilon}) - \tilde{\beta}_{n,0}\| = O(\sqrt{p_n/n})$, we have $\|I_3\| = O(p_n^{5/2}/n)$.

Therefore, $\sqrt{n}A_n\{T(K_{n,\epsilon}) - \tilde{\beta}_{n,0}\} = -\sqrt{n}A_n\mathbf{H}_n^{-1}I_1 + o(1)$, which completes the proof. \square

Lemma A4 (asymptotic normality of $T(\mathbb{K}_n) - T(K_{n,\epsilon})$). Assume Conditions A0–A8 and B4. If $p_n^5/n \rightarrow 0$ as $n \rightarrow \infty$ and the distribution of (\mathbf{X}_n, Y) is $K_{n,\epsilon}$, then

$$\sqrt{n}\{U(K_{n,\epsilon})\}^{-1/2}A_n\{T(\mathbb{K}_n) - T(K_{n,\epsilon})\} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_k),$$

where $U(K_{n,\epsilon}) = A_n\mathbf{H}_{n,\epsilon}^{-1}\Omega_{n,\epsilon}\mathbf{H}_{n,\epsilon}^{-1}A_n^T$, A_n is any given $k \times (p_n + 1)$ matrix such that $A_nA_n^T \rightarrow \mathbb{G}$, with \mathbb{G} being a $k \times k$ positive-definite matrix, k is a fixed integer.

Proof. We will first show that

$$T(\mathbb{K}_n) - T(K_{n,\epsilon}) = -\frac{1}{n}\mathbf{H}_{n,\epsilon}^{-1}\sum_{i=1}^n p_1(Y_i; \tilde{\mathbf{X}}_{ni}^T T(K_{n,\epsilon}))w(\mathbf{X}_{ni})\tilde{\mathbf{X}}_{ni} + o_P(n^{-1/2}). \tag{A8}$$

From $\frac{\partial \ell_{\mathbb{K}_n}(\tilde{\beta})}{\partial \tilde{\beta}} \Big|_{\tilde{\beta}=T(\mathbb{K}_n)} = \mathbf{0}$, Taylor’s expansion yields

$$\begin{aligned} \mathbf{0} &= \left\{ \frac{1}{n} \sum_{i=1}^n p_1(Y_i; \tilde{\mathbf{X}}_{ni}^T T(K_{n,\epsilon}))w(\mathbf{X}_{ni})\tilde{\mathbf{X}}_{ni} \right\} \\ &+ \left\{ \frac{1}{n} \sum_{i=1}^n p_2(Y_i; \tilde{\mathbf{X}}_{ni}^T T(K_{n,\epsilon}))w(\mathbf{X}_{ni})\tilde{\mathbf{X}}_{ni}\tilde{\mathbf{X}}_{ni}^T \right\} \{T(\mathbb{K}_n) - T(K_{n,\epsilon})\} \\ &+ \frac{1}{2n} \sum_{i=1}^n p_3(Y_i; \tilde{\mathbf{X}}_{ni}^T \tilde{\beta}_n^*)w(\mathbf{X}_{ni})[\tilde{\mathbf{X}}_{ni}^T \{T(\mathbb{K}_n) - T(K_{n,\epsilon})\}]^2 \tilde{\mathbf{X}}_{ni} \\ &\equiv \left\{ \frac{1}{n} \sum_{i=1}^n p_1(Y_i; \tilde{\mathbf{X}}_{ni}^T T(K_{n,\epsilon}))w(\mathbf{X}_{ni})\tilde{\mathbf{X}}_{ni} \right\} + I_2 \{T(\mathbb{K}_n) - T(K_{n,\epsilon})\} + I_3, \end{aligned} \tag{A9}$$

where $\tilde{\beta}_n^*$ lies between $T(K_{n,\epsilon})$ and $T(\mathbb{K}_n)$. Below, we will show

$$\|I_2 - \mathbf{H}_{n,\epsilon}\| = O_P(p_n/\sqrt{n}), \quad \|I_3\| = O_P(p_n^{5/2}/n).$$

Similar arguments for the proof of $I_{1,2}$ of Lemma A2, we have $\|I_2 - \mathbf{H}_{n,\epsilon}\| = O_P(p_n/\sqrt{n})$.

Second, a similar proof used for $I_{1,3}^*$ in Equation (A5) gives $\|I_3\| = O_P(p_n^{5/2}/n)$.

Third, by Equation (A9) and $\|T(\mathbb{K}_n) - T(K_{n,\epsilon})\| = O_P(\sqrt{p_n/n})$, we see that

$$\mathbf{H}_{n,\epsilon}\{T(\mathbb{K}_n) - T(K_{n,\epsilon})\} = -\frac{1}{n}\sum_{i=1}^n p_1(Y_i; \tilde{\mathbf{X}}_{ni}^T T(K_{n,\epsilon}))w(\mathbf{X}_{ni})\tilde{\mathbf{X}}_{ni} + \mathbf{u}_n,$$

where $\|\mathbf{u}_n\| = O_P(p_n^{5/2}/n) = o_P(n^{-1/2})$. From the proof of Lemma A2, the eigenvalues of $\mathbf{H}_{n,\epsilon}$ are uniformly bounded away from 0 and we complete the proof of Equation (A8).

Following the proof for the bounded eigenvalues of $\mathbf{H}_{n,\epsilon}$ in Lemma A2, we can show that the eigenvalues of $\Omega_{n,\epsilon}$ are uniformly bounded away from 0. Hence, the eigenvalues of $\mathbf{H}_{n,\epsilon}^{-1}\Omega_{n,\epsilon}\mathbf{H}_{n,\epsilon}^{-1}$ are uniformly bounded away from 0, as are the eigenvalues of $U(K_{n,\epsilon})$. From Equation (A8), we see that

$$A_n\{T(\mathbb{K}_n) - T(K_{n,\epsilon})\} = -\frac{1}{n}A_n\mathbf{H}_{n,\epsilon}^{-1}\sum_{i=1}^n p_1(Y_i; \tilde{\mathbf{X}}_{ni}^T T(K_{n,\epsilon}))w(\mathbf{X}_{ni})\tilde{\mathbf{X}}_{ni} + o_P(n^{-1/2}).$$

It follows that

$$\sqrt{n}\{U(K_{n,\epsilon})\}^{-1/2}A_n\{T(\mathbb{K}_n) - T(K_{n,\epsilon})\} = \sum_{i=1}^n \mathbf{R}_{ni} + o_P(1),$$

where $\mathbf{R}_{ni} = -n^{-1/2}\{U(K_{n,\epsilon})\}^{-1/2}A_n\mathbf{H}_{n,\epsilon}^{-1}p_1(Y_i; \tilde{\mathbf{X}}_{ni}^T T(K_{n,\epsilon}))w(\mathbf{X}_{ni})\tilde{\mathbf{X}}_{ni}$. Following (A6) in Lemma A2, one can show that $E_{K_{n,\epsilon}}(\mathbf{R}_{ni}) = \mathbf{0}$ for n large enough.

To show $\sum_{i=1}^n \mathbf{R}_{ni} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_k)$, we apply the Lindeberg-Feller central limit theorem in [26]. Specifically, we check (I) $\sum_{i=1}^n \text{cov}_{K_{n,\epsilon}}(\mathbf{R}_{ni}) \rightarrow \mathbf{I}_k$; (II) $\sum_{i=1}^n E_{K_{n,\epsilon}}(\|\mathbf{R}_{ni}\|^{2+\delta}) = o(1)$ for some $\delta > 0$. Condition (I) is straightforward since $\sum_{i=1}^n \text{cov}_{K_{n,\epsilon}}(\mathbf{R}_{ni}) = \{U(K_{n,\epsilon})\}^{-1/2} U(K_{n,\epsilon}) \{U(K_{n,\epsilon})\}^{-1/2} = \mathbf{I}_k$. To check condition (II), we can show that $E_{K_{n,\epsilon}}(\|\mathbf{R}_{ni}\|^{2+\delta}) = O((p_n/n)^{(2+\delta)/2})$. This yields $\sum_{i=1}^n E_{K_{n,\epsilon}}(\|\mathbf{R}_{ni}\|^{2+\delta}) \leq O(p_n^{(2+\delta)/2}/n^{\delta/2}) = o(1)$. Hence

$$\sqrt{n}\{U(K_{n,\epsilon})\}^{-1/2} A_n \{T(\mathbb{K}_n) - T(K_{n,\epsilon})\} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_k).$$

Thus, we complete the proof. \square

Lemma A5 (asymptotic covariance matrices $U(K_{n,\epsilon})$ and U_n). Assume Conditions A0–A9 and B4. If $p_n^4/n \rightarrow 0$ as $n \rightarrow \infty$, then

$$\|U_n^{-1/2}\{U(K_{n,\epsilon})\}^{1/2} - \mathbf{I}_k\| = O(p_n/n^{1/4}),$$

where $U(K_{n,\epsilon}) = A_n \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1} A_n^T$, A_n is any given $k \times (p_n + 1)$ matrix such that $A_n A_n^T \rightarrow \mathbb{G}$, with \mathbb{G} being a $k \times k$ positive-definite matrix, and k is a fixed integer.

Proof. Note that

$$\begin{aligned} & \| \{U(K_{n,\epsilon})\}^{1/2} - U_n^{1/2} \|^2 \leq \|U(K_{n,\epsilon}) - U_n\| \\ & \leq \| \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1} - \mathbf{H}_n^{-1} \Omega_n \mathbf{H}_n^{-1} \| \|A_n\|_F^2. \end{aligned}$$

Since $\|A_n\|_F^2 \rightarrow \text{tr}(\mathbb{G})$, it suffices to prove that $\| \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1} - \mathbf{H}_n^{-1} \Omega_n \mathbf{H}_n^{-1} \| = O(p_n^2/\sqrt{n})$. First, we prove $\| \mathbf{H}_{n,\epsilon} - \mathbf{H}_n \| = O(p_n^2/\sqrt{n})$. Note that

$$\begin{aligned} \mathbf{H}_{n,\epsilon} - \mathbf{H}_n &= E_{K_{n,\epsilon}}[\{p_2(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) - p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0})\} w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T] \\ &+ [E_{K_{n,\epsilon}}\{p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0})\} w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T - \mathbf{H}_n] \\ &= E_{K_{n,\epsilon}}[p_3(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}^*) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \tilde{\mathbf{X}}_n^T \{T(K_{n,\epsilon}) - \tilde{\boldsymbol{\beta}}_{n,0}\}] \\ &+ [E_{K_{n,\epsilon}}\{p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0})\} w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T - \mathbf{H}_n] \\ &\equiv I_1 + I_2. \end{aligned}$$

We know that $\|I_1\| = O(p_n^2/\sqrt{n})$ and $\|I_2\| = O(p_n/\sqrt{n})$. Thus, $\|I_1\| = O(p_n^2/\sqrt{n})$. Second, we show $\| \Omega_{n,\epsilon} - \Omega_n \| = O(p_n^2/\sqrt{n})$. It is easy to see that

$$\begin{aligned} \Omega_{n,\epsilon} - \Omega_n &= E_{K_{n,\epsilon}}[\{p_1^2(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) - p_1^2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0})\} w^2(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T] \\ &+ [E_{K_{n,\epsilon}}\{p_1^2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0})\} w^2(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T - \Omega_n] \\ &= \Delta_{1,1} + \Delta_{1,2}, \end{aligned}$$

where $\|\Delta_{1,1}\| = O(p_n^2/\sqrt{n})$ and $\|\Delta_{1,2}\| = O(p_n/\sqrt{n})$. We observe that $\| \Omega_{n,\epsilon} - \Omega_n \| = O(p_n^2/\sqrt{n})$.

Third, we show $\| \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1} - \mathbf{H}_n^{-1} \Omega_n \mathbf{H}_n^{-1} \| = O(p_n^2/\sqrt{n})$. Note $\mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1} - \mathbf{H}_n^{-1} \Omega_n \mathbf{H}_n^{-1} = L_1 + L_2 + L_3$, where $L_1 = \mathbf{H}_{n,\epsilon}^{-1} (\Omega_{n,\epsilon} - \Omega_n) \mathbf{H}_{n,\epsilon}^{-1}$, $L_2 = \mathbf{H}_{n,\epsilon}^{-1} (\mathbf{H}_n - \mathbf{H}_{n,\epsilon}) \mathbf{H}_n^{-1} \Omega_n \mathbf{H}_n^{-1}$ and $L_3 = \mathbf{H}_n^{-1} \Omega_n \mathbf{H}_n^{-1} (\mathbf{H}_n - \mathbf{H}_{n,\epsilon}) \mathbf{H}_n^{-1}$. Under Conditions A7 and A9, it is straightforward to see that $\| \mathbf{H}_{n,\epsilon}^{-1} \| = O(1)$, $\| \mathbf{H}_n^{-1} \| = O(1)$ and $\| \mathbf{H}_n^{-1} \Omega_n \| = O(1)$. Since $\|L_1\| \leq \| \mathbf{H}_{n,\epsilon}^{-1} \| \| \Omega_{n,\epsilon} - \Omega_n \| \| \mathbf{H}_{n,\epsilon}^{-1} \|$, we conclude $\|L_1\| = O(p_n^2/\sqrt{n})$, and similarly $\|L_2\| = O(p_n^2/\sqrt{n})$ and $\|L_3\| = O(p_n^2/\sqrt{n})$. Hence, $\| \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1} - \mathbf{H}_n^{-1} \Omega_n \mathbf{H}_n^{-1} \| = O(p_n^2/\sqrt{n})$.

Thus, we can conclude that $\|U(K_{n,\epsilon}) - U_n\| = O(p_n^2/\sqrt{n})$ and that the eigenvalues of $U(K_{n,\epsilon})$ and U_n are uniformly bounded away from 0 and ∞ . Consequently, $\| \{U(K_{n,\epsilon})\}^{1/2} - U_n^{1/2} \| = O(p_n/n^{1/4})$ and proof is finished. \square

Lemma A6 (asymptotic covariance matrices $U(\mathbb{K}_n)$ and $U(K_{n,\epsilon})$). Assume Conditions A0–A9 and B4. If $p_n^4/n \rightarrow 0$ as $n \rightarrow \infty$ and the distribution of (\mathbf{X}_n, Y) is $K_{n,\epsilon}$, then

$$\|\{U(\mathbb{K}_n)\}^{-1/2}\{U(K_{n,\epsilon})\}^{1/2} - \mathbf{I}_k\| = O_P(p_n/n^{1/4}),$$

where $U(K_{n,\epsilon}) = A_n \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1} A_n^T$, $U(\mathbb{K}_n) = A_n \widehat{\mathbf{H}}_n^{-1} \widehat{\Omega}_n \widehat{\mathbf{H}}_n^{-1} A_n^T$, A_n is any given $k \times (p_n + 1)$ matrix such that $A_n A_n^T \rightarrow \mathbb{G}$, with \mathbb{G} being a $k \times k$ positive-definite matrix, and k is a fixed integer.

Proof. Note that $\|\{U(\mathbb{K}_n)\}^{1/2} - \{U(K_{n,\epsilon})\}^{1/2}\|^2 \leq \|U(\mathbb{K}_n) - U(K_{n,\epsilon})\| \leq \|\widehat{\mathbf{H}}_n^{-1} \widehat{\Omega}_n \widehat{\mathbf{H}}_n^{-1} - \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1}\| \|A_n\|_F^2$. Since $\|A_n\|_F^2 \rightarrow \text{tr}(\mathbb{G})$, it suffices to prove that $\|\widehat{\mathbf{H}}_n^{-1} \widehat{\Omega}_n \widehat{\mathbf{H}}_n^{-1} - \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1}\| = O_P(p_n^2/\sqrt{n})$.

Following the proof of Proposition 1 in [1], we can show that $\|\widehat{\mathbf{H}}_n - \mathbf{H}_{n,\epsilon}\| = O_P(p_n^2/\sqrt{n})$ and $\|\widehat{\Omega}_n - \Omega_{n,\epsilon}\| = O_P(p_n^2/\sqrt{n})$.

To show $\|\widehat{\mathbf{H}}_n^{-1} \widehat{\Omega}_n \widehat{\mathbf{H}}_n^{-1} - \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1}\| = O_P(p_n^2/\sqrt{n})$, note $\widehat{\mathbf{H}}_n^{-1} \widehat{\Omega}_n \widehat{\mathbf{H}}_n^{-1} - \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1} = L_1 + L_2 + L_3$, where $L_1 = \widehat{\mathbf{H}}_n^{-1} (\widehat{\Omega}_n - \Omega_{n,\epsilon}) \widehat{\mathbf{H}}_n^{-1}$, $L_2 = \widehat{\mathbf{H}}_n^{-1} (\mathbf{H}_{n,\epsilon} - \widehat{\mathbf{H}}_n) \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \widehat{\mathbf{H}}_n^{-1}$ and $L_3 = \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \widehat{\mathbf{H}}_n^{-1} (\mathbf{H}_{n,\epsilon} - \widehat{\mathbf{H}}_n) \mathbf{H}_{n,\epsilon}^{-1}$. Following the proof in Lemma A2, it is straightforward to verify that $\|\mathbf{H}_{n,\epsilon}^{-1}\| = O(1)$, $\|\widehat{\mathbf{H}}_n^{-1}\| = O_P(1)$. In addition, $\|\mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon}\| = \|(\mathbf{H}_{n,\epsilon}^{-1} - \mathbf{H}_n^{-1}) \Omega_{n,\epsilon} + \mathbf{H}_n^{-1} (\Omega_{n,\epsilon} - \Omega_n) + \mathbf{H}_n^{-1} \Omega_n\| \leq \|\mathbf{H}_{n,\epsilon}^{-1}\| \|\mathbf{H}_{n,\epsilon} - \mathbf{H}_n\| \|\Omega_{n,\epsilon}\| + \|\mathbf{H}_n^{-1}\| \|\Omega_{n,\epsilon} - \Omega_n\| + \|\mathbf{H}_n^{-1} \Omega_n\| = O(1)$.

Since $\|L_1\| \leq \|\widehat{\mathbf{H}}_n^{-1}\| \|\widehat{\Omega}_n - \Omega_{n,\epsilon}\| \|\widehat{\mathbf{H}}_n^{-1}\|$, we conclude $\|L_1\| = O_P(p_n^2/\sqrt{n})$, and similarly $\|L_2\| = O_P(p_n^2/\sqrt{n})$ and $\|L_3\| = O_P(p_n^2/\sqrt{n})$. Hence, $\|\widehat{\mathbf{H}}_n^{-1} \widehat{\Omega}_n \widehat{\mathbf{H}}_n^{-1} - \mathbf{H}_{n,\epsilon}^{-1} \Omega_{n,\epsilon} \mathbf{H}_{n,\epsilon}^{-1}\| = O_P(p_n^2/\sqrt{n})$.

Thus, we can conclude that $\|U(\mathbb{K}_n) - U(K_{n,\epsilon})\| = O_P(p_n^2/\sqrt{n})$ and the eigenvalues of $U(\mathbb{K}_n)$ are uniformly bounded away from 0 and ∞ with probability tending to 1. Noting that $\|\{U(\mathbb{K}_n)\}^{1/2} - \{U(K_{n,\epsilon})\}^{1/2}\|^2 \leq \|U(\mathbb{K}_n) - U(K_{n,\epsilon})\|$. \square

Lemma A7 (asymptotic distribution of test statistic). Assume Conditions A0–A9 and B4. If $p_n^6/n \rightarrow 0$ as $n \rightarrow \infty$ and the distribution of (\mathbf{X}_n, Y) is $K_{n,\epsilon}$, then

$$\sqrt{n}[\{U(\mathbb{K}_n)\}^{-1/2} A_n \{\mathbf{T}(\mathbb{K}_n) - \widetilde{\boldsymbol{\beta}}_{n,0}\} - U_n^{-1/2} A_n \{\mathbf{T}(K_{n,\epsilon}) - \widetilde{\boldsymbol{\beta}}_{n,0}\}] \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_k),$$

where A_n is any given $k \times (p_n + 1)$ matrix such that $A_n A_n^T \rightarrow \mathbb{G}$, with \mathbb{G} being a $k \times k$ positive-definite matrix, and k is a fixed integer.

Proof. Note that

$$\begin{aligned} & \sqrt{n}[\{U(\mathbb{K}_n)\}^{-1/2} A_n \{\mathbf{T}(\mathbb{K}_n) - \widetilde{\boldsymbol{\beta}}_{n,0}\} - U_n^{-1/2} A_n \{\mathbf{T}(K_{n,\epsilon}) - \widetilde{\boldsymbol{\beta}}_{n,0}\}] \\ &= \sqrt{n}\{U(\mathbb{K}_n)\}^{-1/2} A_n \{\mathbf{T}(\mathbb{K}_n) - \mathbf{T}(K_{n,\epsilon})\} \\ & \quad + \sqrt{n}[\{U(\mathbb{K}_n)\}^{-1/2} - \{U(K_{n,\epsilon})\}^{-1/2}] A_n \{\mathbf{T}(K_{n,\epsilon}) - \widetilde{\boldsymbol{\beta}}_{n,0}\} \\ & \quad + \sqrt{n}[\{U(K_{n,\epsilon})\}^{-1/2} - U_n^{-1/2}] A_n \{\mathbf{T}(K_{n,\epsilon}) - \widetilde{\boldsymbol{\beta}}_{n,0}\} \\ & \equiv \text{I} + \text{II} + \text{III}. \end{aligned}$$

For term I, we obtain from Lemma A4 that $\sqrt{n}\{U(K_{n,\epsilon})\}^{-1/2} A_n (\mathbf{T}(\mathbb{K}_n) - \mathbf{T}(K_{n,\epsilon})) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_k)$. From Lemma A6, we get $\|\{U(\mathbb{K}_n)\}^{-1/2}\{U(K_{n,\epsilon})\}^{1/2} - \mathbf{I}_k\| = o_P(1)$. Thus, by Slutsky theorem,

$$\text{I} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_k). \tag{A10}$$

For term II, we see from Lemma A6 that

$$\|\{U(\mathbb{K}_n)\}^{-1/2} - \{U(K_{n,\epsilon})\}^{-1/2}\| = O_P(p_n/n^{1/4}).$$

Since

$$\|A_n \{\mathbf{T}(K_{n,\epsilon}) - \widetilde{\boldsymbol{\beta}}_{n,0}\}\| \leq \|A_n\| \|\mathbf{T}(K_{n,\epsilon}) - \widetilde{\boldsymbol{\beta}}_{n,0}\| = O(\sqrt{p_n/n}).$$

Thus,

$$\|\text{II}\| \leq \sqrt{n} \|\{U(\mathbb{K}_n)\}^{-1/2} - \{U(K_{n,\epsilon})\}^{-1/2}\| \|A_n\| \|\mathbf{T}(K_{n,\epsilon}) - \tilde{\boldsymbol{\beta}}_{n,0}\| = O_P(p_n^{3/2}/n^{1/4}). \tag{A11}$$

Similarly, $\|\text{III}\| = o_P(1)$. Combining (A10) and (A11) with Slutsky theorem completes the proof. \square

Lemma A8 (Influence Function IF). *Assume Conditions A1–A8 and B4. For any fixed sample size n ,*

$$\begin{aligned} & \left. \frac{\partial}{\partial t} \mathbf{T}((1-t)K_{n,0} + tJ) \right|_{t=t_0} \equiv \lim_{t \rightarrow t_0} \frac{\mathbf{T}((1-t)K_{n,0} + tJ) - \mathbf{T}((1-t_0)K_{n,0} + t_0J)}{t - t_0} \\ & = -\mathbf{H}_{n,K_{t_0},T(K_{t_0})}^{-1} [E_J\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\} - E_{K_{n,0}}\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\}], \end{aligned}$$

where $K_{t_0} = (1 - t_0)K_{n,0} + t_0J$ and t_0 is a positive constant such that $t_0 \leq c/p_n^2$ with $c > 0$ a sufficiently small constant. In addition, $\|\mathbf{H}_{n,K_{t_0},T(K_{t_0})}^{-1}\| \leq C$ uniformly for all n and t_0 such that $t_0 \leq c/p_n^2$ with $c > 0$ a sufficiently small constant.

Proof. We follow the proof of Theorem 5.1 in [27]. Note

$$\begin{aligned} & \lim_{t \rightarrow t_0} \frac{\mathbf{T}((1-t)K_{n,0} + tJ) - \mathbf{T}((1-t_0)K_{n,0} + t_0J)}{t - t_0} \\ & = \lim_{\Delta \rightarrow 0} \frac{\mathbf{T}(K_{t_0} + \Delta(J - K_{n,0})) - \mathbf{T}(K_{t_0})}{\Delta}, \end{aligned}$$

where $\Delta = t - t_0$.

It suffices to prove that for any sequence $\{\Delta_j\}_{j=1}^\infty$ such that $\lim_{j \rightarrow \infty} \Delta_j = 0$, we have

$$\begin{aligned} & \lim_{j \rightarrow \infty} \frac{\mathbf{T}(K_{t_0} + \Delta_j(J - K_{n,0})) - \mathbf{T}(K_{t_0})}{\Delta_j} \\ & = -\mathbf{H}_{n,K_{t_0},T(K_{t_0})}^{-1} [E_J\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\} - E_{K_{n,0}}\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\}]. \end{aligned}$$

Following similar proofs in Lemma A1, we can show that for t_0 sufficiently small,

$$\|\tilde{\boldsymbol{\beta}}_{n,0} - \mathbf{T}(K_{t_0})\| \leq Ct_0\sqrt{p_n}. \tag{A12}$$

Next we will show that the eigenvalues of $\mathbf{H}_{n,K_{t_0},T(K_{t_0})}$ are bounded away from 0.

$$\begin{aligned} \mathbf{H}_{n,K_{t_0},T(K_{t_0})} & = (1 - t_0)\mathbf{H}_{n,K_{n,0},T(K_{t_0})} + t_0\mathbf{H}_{n,J,T(K_{t_0})} \\ & = (1 - t_0)\mathbf{H}_n + t_0\mathbf{H}_{n,J,\tilde{\boldsymbol{\beta}}_{n,0}} + (1 - t_0)\{\mathbf{H}_{n,K_{n,0},T(K_{t_0})} - \mathbf{H}_n\} \\ & \quad + t_0\{\mathbf{H}_{n,J,T(K_{t_0})} - \mathbf{H}_{n,J,\tilde{\boldsymbol{\beta}}_{n,0}}\} = (1 - t_0)I_1 + t_0I_2 + I_3 + I_4. \end{aligned}$$

First,

$$\begin{aligned} \|I_3\| & \leq C E_{K_{n,0}} \|\{p_2(Y; \tilde{\mathbf{X}}_n^T \mathbf{T}(K_{t_0})) - p_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0})\} w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T\| \\ & \leq Cp_n^{3/2} \|\mathbf{T}(K_{t_0}) - \tilde{\boldsymbol{\beta}}_{n,0}\| \leq Cp_n^2 t_0. \end{aligned}$$

Similarly, $\|I_2\| \leq Cp_n t_0$ and $\|I_4\| \leq Cp_n^2 t_0^2$. Since the eigenvalues of I_1 are bounded away from zero, $\|I_2\|$, $\|I_3\|$ and $\|I_4\|$ could be sufficiently small, we conclude that for $t_0 \leq c/p_n^2$ when c is sufficiently small, the eigenvalues of $\mathbf{H}_{n,K_{t_0},T(K_{t_0})}$ are uniformly bounded away from 0.

Define $K_j = K_{t_0} + \Delta_j(J - K_{n,0})$. Following similar arguments for (A6) in Lemma A2, for j large enough, $E_{K_j}\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_j))\} = \mathbf{0}$. We will only consider j large enough below. The two term Taylor expansion yields

$$\mathbf{0} = E_{K_j}\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_j))\} = E_{K_j}\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\} + \mathbf{H}_{n,K_j,\tilde{\boldsymbol{\beta}}_j^*}\{\mathbf{T}(K_j) - \mathbf{T}(K_{t_0})\}, \tag{A13}$$

where $\tilde{\boldsymbol{\beta}}_j^*$ lies between $\mathbf{T}(K_{t_0})$ and $\mathbf{T}(K_j)$.

Thus, from (A13) and the fact $E_{K_j}\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\} = \Delta_j[E_J\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\} - E_{K_{n,0}}\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\}]$, we have

$$\begin{aligned} \mathbf{0} &= E_{K_j}\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\} + \mathbf{H}_{n,K_{t_0},\mathbf{T}(K_{t_0})}\{\mathbf{T}(K_j) - \mathbf{T}(K_{t_0})\} \\ &\quad + \{\mathbf{H}_{n,K_j,\tilde{\boldsymbol{\beta}}_j^*} - \mathbf{H}_{n,K_{t_0},\mathbf{T}(K_{t_0})}\}\{\mathbf{T}(K_j) - \mathbf{T}(K_{t_0})\} \\ &= \Delta_j[E_J\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\} - E_{K_{n,0}}\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\}] \\ &\quad + \mathbf{H}_{n,K_{t_0},\mathbf{T}(K_{t_0})}\{\mathbf{T}(K_j) - \mathbf{T}(K_{t_0})\} + (\mathbf{H}_{n,K_j,\tilde{\boldsymbol{\beta}}_j^*} - \mathbf{H}_{n,K_{t_0},\mathbf{T}(K_{t_0})})\{\mathbf{T}(K_j) - \mathbf{T}(K_{t_0})\}, \end{aligned}$$

and we obtain that

$$\begin{aligned} &\mathbf{T}(K_j) - \mathbf{T}(K_{t_0}) \\ &= -\Delta_j \mathbf{H}_{n,K_{t_0},\mathbf{T}(K_{t_0})}^{-1} [E_J\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\} - E_{K_{n,0}}\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\}] \\ &\quad - \mathbf{H}_{n,K_{t_0},\mathbf{T}(K_{t_0})}^{-1} \{\mathbf{H}_{n,K_j,\tilde{\boldsymbol{\beta}}_j^*} - \mathbf{H}_{n,K_{t_0},\mathbf{T}(K_{t_0})}\} \{\mathbf{T}(K_j) - \mathbf{T}(K_{t_0})\}. \end{aligned} \tag{A14}$$

Next, we will show that $\|\mathbf{H}_{n,K_j,\tilde{\boldsymbol{\beta}}_j^*} - \mathbf{H}_{n,K_{t_0},\mathbf{T}(K_{t_0})}\| = o(1)$ as $j \rightarrow \infty$ for any fixed n . Since $\|\tilde{\boldsymbol{\beta}}_j^* - \mathbf{T}(K_{t_0})\| \leq \|\mathbf{T}(K_j) - \mathbf{T}(K_{t_0})\| = O(\Delta_j)$,

$$\begin{aligned} &\|\mathbf{H}_{n,K_j,\tilde{\boldsymbol{\beta}}_j^*} - \mathbf{H}_{n,K_{t_0},\tilde{\boldsymbol{\beta}}_j^*}\| \\ &= \Delta_j \|E_J\{\mathbf{p}_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_j^*) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T\} - E_{K_{n,0}}\{\mathbf{p}_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_j^*) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T\}\| \\ &= O(\Delta_j) = o(1) \quad \text{as } j \rightarrow \infty, \end{aligned} \tag{A15}$$

and also,

$$\begin{aligned} &\|\mathbf{H}_{n,K_{t_0},\tilde{\boldsymbol{\beta}}_j^*} - \mathbf{H}_{n,K_{t_0},\mathbf{T}(K_{t_0})}\| \\ &= \|E_{K_{t_0}}\{\{\mathbf{p}_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_j^*) - \mathbf{p}_2(Y; \tilde{\mathbf{X}}_n^T \mathbf{T}(K_{t_0}))\} w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T\}\| \\ &= o(1) \quad \text{as } j \rightarrow \infty. \end{aligned} \tag{A16}$$

From Equations (A15) and (A16),

$$\|\mathbf{H}_{n,K_j,\tilde{\boldsymbol{\beta}}_j^*} - \mathbf{H}_{n,K_{t_0},\mathbf{T}(K_{t_0})}\| = o(1) \quad \text{as } j \rightarrow \infty$$

which, together with Equations (A12) and (A14), implies that

$$\|\mathbf{T}(K_j) - \mathbf{T}(K_{t_0}) + \Delta_j \mathbf{H}_{n,K_{t_0},\mathbf{T}(K_{t_0})}^{-1} [E_J\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\} - E_{K_{n,0}}\{\boldsymbol{\psi}_{\text{RBD}}(\mathbf{Z}_n; \mathbf{T}(K_{t_0}))\}]\| = o(\Delta_j).$$

This completes the proof. \square

Lemma A9. Assume Conditions A1–A8 and B4 and $\sup_n E_J(\|w(\mathbf{X}_n) \tilde{\mathbf{X}}_n\|) \leq C$. Let $H_k(\cdot; \delta)$ be the cumulative distribution function of $\chi_k^2(\delta)$ distribution with δ the noncentrality parameter. Denote $\delta(\epsilon) = n\|U_n^{-1/2}\{A_n \mathbf{T}(K_{n,\epsilon}) - \mathbf{g}_0\}\|^2$. Let $b(\epsilon) = -H_k(x; \delta(\epsilon))$. Then, for any fixed $x > 0$, $\sup_{\epsilon \in [0,C]} \limsup_{n \rightarrow \infty} |b^{(3)}(\epsilon)| \leq C$ under H_0 and $\sup_{\epsilon \in [0,C]} \limsup_{n \rightarrow \infty} |b''(\epsilon)| \leq C$ under H_{1n} .

Proof. Since $b(\epsilon) = -H_k(x; \delta(\epsilon))$, we have

$$\begin{aligned} b'(\epsilon) &= -\frac{\partial}{\partial \epsilon} H_k(x; \delta(\epsilon)) = \left\{ -\frac{\partial}{\partial \delta} H_k(x; \delta) \Big|_{\delta=\delta(\epsilon)} \right\} \left\{ \frac{\partial \delta(\epsilon)}{\partial \epsilon} \right\} \\ b''(\epsilon) &= \left\{ -\frac{\partial^2}{\partial \delta^2} H_k(x; \delta) \Big|_{\delta=\delta(\epsilon)} \right\} \left\{ \frac{\partial \delta(\epsilon)}{\partial \epsilon} \right\}^2 + \left\{ -\frac{\partial}{\partial \delta} H_k(x; \delta) \Big|_{\delta=\delta(\epsilon)} \right\} \left\{ \frac{\partial^2 \delta(\epsilon)}{\partial \epsilon^2} \right\} \\ b^{(3)}(\epsilon) &= \left\{ -\frac{\partial^3}{\partial \delta^3} H_k(x; \delta) \Big|_{\delta=\delta(\epsilon)} \right\} \left\{ \frac{\partial \delta(\epsilon)}{\partial \epsilon} \right\}^3 \\ &\quad + 3 \left\{ -\frac{\partial^2}{\partial \delta^2} H_k(x; \delta) \Big|_{\delta=\delta(\epsilon)} \right\} \left\{ \frac{\partial \delta(\epsilon)}{\partial \epsilon} \right\} \left\{ \frac{\partial^2 \delta(\epsilon)}{\partial \epsilon^2} \right\} \\ &\quad + \left\{ -\frac{\partial}{\partial \delta} H_k(x; \delta) \Big|_{\delta=\delta(\epsilon)} \right\} \left\{ \frac{\partial^3 \delta(\epsilon)}{\partial \epsilon^3} \right\}. \end{aligned}$$

To complete the proof, we only need to show that $\partial^i / \partial \delta^i H_k(x; \delta) \Big|_{\delta=\delta(\epsilon)}$ and $\partial^i \delta(\epsilon) / \partial \epsilon^i$ ($i = 1, 2, 3$) are bounded as $n \rightarrow \infty$ for all $\epsilon \in [0, C]$. Note that

$$H_k(x; \delta) = e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{\gamma(j+k/2, x/2)}{\Gamma(j+k/2)},$$

where $\Gamma(\cdot)$ is the Gamma function, and $\gamma(\cdot, \cdot)$ is the lower incomplete gamma function $\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt$, which satisfies $\gamma(s, x) = (s-1)\gamma(s-1, x) - x^{s-1} e^{-x}$. Therefore,

$$\begin{aligned} \frac{\partial}{\partial \delta} H_k(x; \delta) &= -\frac{e^{-\delta/2}}{2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{\gamma(j+k/2, x/2)}{\Gamma(j+k/2)} + \frac{e^{-\delta/2}}{2} \sum_{j=1}^{\infty} \frac{(\delta/2)^{j-1}}{(j-1)!} \frac{\gamma(j+k/2, x/2)}{\Gamma(j+k/2)} \\ &= \frac{1}{2} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \left\{ -\frac{\gamma(j+k/2, x/2)}{\Gamma(j+k/2)} + \frac{\gamma(j+1+k/2, x/2)}{\Gamma(j+1+k/2)} \right\}. \end{aligned}$$

Since

$$\begin{aligned} \frac{\gamma(j+1+k/2, x/2)}{\Gamma(j+1+k/2)} &= \frac{(j+k/2)\gamma(j+k/2, x/2) - (x/2)^{j+k/2} e^{-x/2}}{\Gamma(j+1+k/2)} \\ &= \frac{\gamma(j+k/2, x/2)}{\Gamma(j+k/2)} - \frac{(x/2)^{j+k/2} e^{-x/2}}{\Gamma(j+1+k/2)}, \end{aligned}$$

we have

$$\begin{aligned} \frac{\partial}{\partial \delta} H_k(x; \delta) &= -\frac{1}{2} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^{j+k/2} e^{-x/2}}{\Gamma(j+1+k/2)} \\ \frac{\partial^2}{\partial \delta^2} H_k(x; \delta) &= \frac{1}{4} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^{j+k/2} e^{-x/2}}{\Gamma(j+1+k/2)} - \frac{1}{4} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^{j+1+k/2} e^{-x/2}}{\Gamma(j+2+k/2)} \\ &= \frac{1}{4} (x/2)^{k/2} e^{-x/2} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \left\{ \frac{(x/2)^j}{\Gamma(j+1+k/2)} - \frac{(x/2)^{j+1}}{\Gamma(j+2+k/2)} \right\} \\ &= \frac{1}{4} (x/2)^{k/2} e^{-x/2} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^j}{\Gamma(j+1+k/2)} \left\{ 1 - \frac{(x/2)}{j+1+k/2} \right\} \\ \frac{\partial^3}{\partial \delta^3} H_k(x; \delta) &= -\frac{1}{8} (x/2)^{k/2} e^{-x/2} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^j}{\Gamma(j+1+k/2)} \left\{ 1 - \frac{(x/2)}{j+1+k/2} \right\} \\ &\quad + \frac{1}{8} (x/2)^{k/2} e^{-x/2} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^{j+1}}{\Gamma(j+2+k/2)} \left\{ 1 - \frac{(x/2)}{j+2+k/2} \right\} \\ &= \frac{1}{8} (x/2)^{k/2} e^{-x/2} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^j}{\Gamma(j+1+k/2)} \end{aligned}$$

$$\cdot \left[\frac{(x/2)}{j+1+k/2} \left\{ 1 - \frac{(x/2)}{j+2+k/2} \right\} - \left\{ 1 - \frac{(x/2)}{j+1+k/2} \right\} \right].$$

From the results of Lemma A3, that $|\delta(\epsilon)|$ is bounded as $n \rightarrow \infty$ for all $\epsilon \in [0, C]$ under both H_0 and H_{1n} , so are $\partial^i / \partial \delta^i H_k(x; \delta)|_{\delta=\delta(\epsilon)}$ ($i = 1, 2, 3$). Now, we consider the derivatives of $\delta(\epsilon)$,

$$\begin{aligned} \frac{\partial \delta(\epsilon)}{\partial \epsilon} &= 2n \left\{ A_n \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \right\}^T U_n^{-1} \{ A_n T(K_{n,\epsilon}) - \mathbf{g}_0 \} \\ \frac{\partial^2 \delta(\epsilon)}{\partial \epsilon^2} &= 2n \left\{ A_n \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \right\}^T U_n^{-1} \left\{ A_n \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \right\} \\ &\quad + 2n \left\{ A_n \frac{\partial^2 T(K_{n,\epsilon})}{\partial \epsilon^2} \right\}^T U_n^{-1} \{ A_n T(K_{n,\epsilon}) - \mathbf{g}_0 \} \\ \frac{\partial^3 \delta(\epsilon)}{\partial \epsilon^3} &= 6n \left\{ A_n \frac{\partial^2 T(K_{n,\epsilon})}{\partial \epsilon^2} \right\}^T U_n^{-1} \left\{ A_n \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \right\} \\ &\quad + 2n \left\{ A_n \frac{\partial^3 T(K_{n,\epsilon})}{\partial \epsilon^3} \right\}^T U_n^{-1} \{ A_n T(K_{n,\epsilon}) - \mathbf{g}_0 \}. \end{aligned}$$

To complete the proof, we only need to show that $\sqrt{n} \|\partial^i / \partial \epsilon^i T(K_{n,\epsilon})\|$ ($i = 1, 2, 3$) are bounded as $n \rightarrow \infty$ for all $\epsilon \in [0, C]$, and $\sqrt{n} \|A_n T(K_{n,\epsilon}) - \mathbf{g}_0\|$ is bounded under H_0 and H_{1n} as $n \rightarrow \infty$ for all $\epsilon \in [0, C]$. The result for $\sqrt{n} \|A_n T(K_{n,\epsilon}) - \mathbf{g}_0\|$ is straightforward from Lemma A3.

First, for the first order derivative of $T(K_{n,\epsilon})$,

$$\begin{aligned} &\sqrt{n} \frac{\partial}{\partial \epsilon} T(K_{n,\epsilon}) \\ &= -\mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1} [E_J \{ \mathbf{p}_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \} - E_{K_{n,0}} \{ \mathbf{p}_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \}]. \end{aligned}$$

Since $\|\mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1}\| \leq C$, $\|E_J \{ \mathbf{p}_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \}\| \leq C E_J \|w(\mathbf{X}_n) \tilde{\mathbf{X}}_n\| \leq C$ and

$$\begin{aligned} &\|E_{K_{n,0}} \{ \mathbf{p}_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \}\| \\ &= \|E_{K_{n,0}} \{ \mathbf{p}_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \} - E_{K_{n,0}} \{ \mathbf{p}_1(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}_{n,0}) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \}\| \\ &= \|E_{K_{n,0}} [\mathbf{p}_2(Y; \tilde{\mathbf{X}}_n^T \tilde{\boldsymbol{\beta}}^*) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \{ T(K_{n,\epsilon}) - \tilde{\boldsymbol{\beta}}_{n,0} \}]\| \\ &\leq Cp_n^{3/2} / \sqrt{n}, \end{aligned}$$

we conclude that $\sqrt{n} \|\partial / \partial \epsilon T(K_{n,\epsilon})\|$ is uniformly bounded for all $\epsilon \in [0, C]$ as $n \rightarrow \infty$.

Second, for the second order derivative of $T(K_{n,\epsilon})$,

$$\begin{aligned} &\sqrt{n} \frac{\partial^2}{\partial \epsilon^2} T((1 - \epsilon / \sqrt{n})K_{n,0} + \epsilon / \sqrt{n}J) \\ &= -\frac{\partial \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1}}{\partial \epsilon} \\ &\quad \cdot [E_J \{ \mathbf{p}_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \} - E_{K_{n,0}} \{ \mathbf{p}_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \}] \\ &\quad - \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1} \\ &\quad \cdot \frac{\partial}{\partial \epsilon} [E_J \{ \mathbf{p}_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \} - E_{K_{n,0}} \{ \mathbf{p}_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \}] \end{aligned}$$

with

$$\begin{aligned} \frac{\partial}{\partial \epsilon} \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1} &= -\mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1} \frac{\partial \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}}{\partial \epsilon} \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1} \\ \frac{\partial \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}}{\partial \epsilon} &= -\frac{1}{\sqrt{n}} E_{K_{n,0}} \{ \mathbf{p}_2(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \} \\ &\quad + \frac{1}{\sqrt{n}} E_J \{ \mathbf{p}_2(Y; \tilde{\mathbf{X}}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{\mathbf{X}}_n \tilde{\mathbf{X}}_n^T \} \end{aligned}$$

$$\begin{aligned}
 & + (1 - \epsilon/\sqrt{n}) E_{K_{n,0}} \left\{ p_3(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{n,\epsilon}) \right\} \\
 & + \epsilon/\sqrt{n} E_J \left\{ p_3(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{n,\epsilon}) \right\}.
 \end{aligned}$$

Therefore, $\|\partial/\partial\epsilon \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1}\| \leq C \|\partial/\partial\epsilon \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}\| \leq Cp_n^{3/2}/\sqrt{n}$. In addition,

$$\begin{aligned}
 & \left\| \frac{\partial}{\partial \epsilon} [E_J \{p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n\} - E_{K_{n,0}} \{p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n\}] \right\| \\
 = & \left\| E_J \left\{ p_2(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{n,\epsilon}) \right\} \right. \\
 & \left. - E_{K_{n,0}} \left\{ p_2(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{n,\epsilon}) \right\} \right\| \\
 \leq & Cp_n/\sqrt{n}.
 \end{aligned}$$

Therefore, $\|\sqrt{n} \frac{\partial^2}{\partial \epsilon^2} T((1 - \epsilon/\sqrt{n})K_{n,0} + \epsilon/\sqrt{n}J)\| = o(1)$ for all $\epsilon \in [0, C]$.

Finally, for the third order derivative of $T(K_{n,\epsilon})$,

$$\begin{aligned}
 & \sqrt{n} \frac{\partial^3}{\partial \epsilon^3} T((1 - \epsilon/\sqrt{n})K_{n,0} + \epsilon/\sqrt{n}J) \\
 = & - \frac{\partial^2 \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1}}{\partial \epsilon^2} \\
 & \cdot [E_J \{p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n\} - E_{K_{n,0}} \{p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n\}] \\
 & - 2 \frac{\partial \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1}}{\partial \epsilon} \\
 & \cdot \frac{\partial}{\partial \epsilon} [E_J \{p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n\} - E_{K_{n,0}} \{p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n\}] \\
 & - \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1} \\
 & \cdot \frac{\partial^2}{\partial \epsilon^2} [E_J \{p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n\} - E_{K_{n,0}} \{p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n\}].
 \end{aligned}$$

Note:

$$\begin{aligned}
 & \frac{\partial^2}{\partial \epsilon^2} \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1} \\
 = & - \frac{\partial \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1}}{\partial \epsilon} \frac{\partial \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}}{\partial \epsilon} \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1} \\
 & - \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1} \frac{\partial^2 \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}}{\partial \epsilon^2} \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1} \\
 & - \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1} \frac{\partial \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}}{\partial \epsilon} \frac{\partial \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})}^{-1}}{\partial \epsilon},
 \end{aligned}$$

where

$$\begin{aligned}
 & \frac{\partial^2}{\partial \epsilon^2} \mathbf{H}_{n,K_{n,\epsilon},T(K_{n,\epsilon})} \\
 = & - \frac{2}{\sqrt{n}} E_{K_{n,0}} \left\{ p_3(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{n,\epsilon}) \right\} \\
 & + \frac{2}{\sqrt{n}} E_J \left\{ p_3(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{n,\epsilon}) \right\} \\
 & + (1 - \epsilon/\sqrt{n}) E_{K_{n,0}} \left\{ p_4(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T (\tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{n,\epsilon}))^2 \right\} \\
 & + \epsilon/\sqrt{n} E_J \left\{ p_4(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T (\tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{n,\epsilon}))^2 \right\}.
 \end{aligned}$$

Hence, $\|\frac{\partial^2}{\partial \epsilon^2} \mathbf{H}_{n, K_{n, \epsilon}, T(K_{n, \epsilon})}\| \leq Cp_n^2/n$ which implies that $\|\frac{\partial^2}{\partial \epsilon^2} \mathbf{H}_{n, K_{n, \epsilon}, T(K_{n, \epsilon})}^{-1}\| = o(1)$ for all $\epsilon \in [0, C]$. In addition,

$$\begin{aligned} & \frac{\partial^2}{\partial \epsilon^2} [E_J\{p_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n, \epsilon}))w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\} - E_{K_{n,0}}\{p_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n, \epsilon}))w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\}] \\ &= \frac{\partial}{\partial \epsilon} \left[E_J\left\{p_2(Y; \tilde{\mathbf{X}}_n^T T(K_{n, \epsilon}))w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^T \frac{\partial}{\partial \epsilon} T(K_{n, \epsilon})\right\} \right. \\ & \quad \left. - E_{K_{n,0}}\left\{p_2(Y; \tilde{\mathbf{X}}_n^T T(K_{n, \epsilon}))w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^T \frac{\partial}{\partial \epsilon} T(K_{n, \epsilon})\right\} \right] \\ &= E_J\left\{p_3(Y; \tilde{\mathbf{X}}_n^T T(K_{n, \epsilon}))w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\left(\tilde{\mathbf{X}}_n^T \frac{\partial}{\partial \epsilon} T(K_{n, \epsilon})\right)^2\right\} \\ & \quad + E_J\left\{p_2(Y; \tilde{\mathbf{X}}_n^T T(K_{n, \epsilon}))w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^T \frac{\partial^2}{\partial \epsilon^2} T(K_{n, \epsilon})\right\} \\ & \quad - E_{\tilde{\beta}_{n,0}}\left\{p_3(Y; \tilde{\mathbf{X}}_n^T T(K_{n, \epsilon}))w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\left(\tilde{\mathbf{X}}_n^T \frac{\partial}{\partial \epsilon} T(K_{n, \epsilon})\right)^2\right\} \\ & \quad - E_{\tilde{\beta}_{n,0}}\left\{p_2(Y; \tilde{\mathbf{X}}_n^T T(K_{n, \epsilon}))w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^T \frac{\partial^2}{\partial \epsilon^2} T(K_{n, \epsilon})\right\}. \end{aligned}$$

Hence, $\|\frac{\partial^2}{\partial \epsilon^2} [E_J\{p_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n, \epsilon}))w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\} - E_{K_{n,0}}\{p_1(Y; \tilde{\mathbf{X}}_n^T T(K_{n, \epsilon}))w(\mathbf{X}_n)\tilde{\mathbf{X}}_n\}]\| \leq Cp_n/\sqrt{n}$. Therefore, $\|\sqrt{n} \frac{\partial^3}{\partial \epsilon^3} T((1 - \epsilon/\sqrt{n})K_{n,0} + \epsilon/\sqrt{n}J)\| = o(1)$ for all $\epsilon \in [0, C]$. Hence, we complete the proof. \square

Proof of Theorem 1. We follow the idea of the proof in [10]. Lemma A7 implies that the Wald-type test statistic W_n is asymptotically noncentral χ_k^2 with noncentrality parameter $\delta(\epsilon) = n\|U_n^{-1/2}\{A_n T(K_{n, \epsilon}) - \mathbf{g}_0\}\|^2$. Therefore, $\alpha(K_{n, \epsilon}) = P(W_n > \eta_{1-\alpha_0} | H_0) = 1 - H_k(\eta_{1-\alpha_0}; \delta(\epsilon)) + h(n, \epsilon)$ where $h(n, \epsilon) = \alpha(K_{n, \epsilon}) - 1 + H_k(\eta_{1-\alpha_0}; \delta(\epsilon)) \rightarrow 0$ as $n \rightarrow \infty$ for any fixed ϵ . Let $b(\epsilon) = -H_k(\eta_{1-\alpha_0}; \delta(\epsilon))$. Then, for ϵ close to 0, we have

$$\begin{aligned} \alpha(K_{n, \epsilon}) - \alpha_0 &= b(\epsilon) - b(0) + h(n, \epsilon) - h(n, 0) \\ &= \epsilon b'(0) + \frac{1}{2}\epsilon^2 b''(0) + \frac{1}{6}\epsilon^3 b^{(3)}(\epsilon^*) + h(n, \epsilon) - h(n, 0), \end{aligned} \tag{A17}$$

where $0 < \epsilon^* < \epsilon$. Note that under H_0

$$b'(0) = \mu_k \left\{ \frac{\partial \delta(\epsilon)}{\partial \epsilon} \Big|_{\epsilon=0} \right\} = 2\mu_k n \left\{ A_n \frac{\partial T(K_{n, \epsilon})}{\partial \epsilon} \Big|_{\epsilon=0} \right\}^T U_n^{-1} \{A_n \tilde{\beta}_{n,0} - \mathbf{g}_0\} = 0.$$

From Lemma A8, under H_0

$$\frac{\partial T(K_{n, \epsilon})}{\partial \epsilon} \Big|_{\epsilon=0} = 1/\sqrt{n} E_J\{\text{IF}(\mathbf{Z}_n; \mathbf{T}, K_{n,0})\}.$$

Thus,

$$b''(0) = \mu_k \left\{ \frac{\partial^2 \delta(\epsilon)}{\partial \epsilon^2} \Big|_{\epsilon=0} \right\} = 2\mu_k \|U_n^{-1/2} A_n E_J\{\text{IF}(\mathbf{Z}_n; \mathbf{T}, K_{n,0})\}\|^2.$$

Since from Lemma A8, $\text{IF}(z_n; \mathbf{T}, K_{n,0}) = -\mathbf{H}_n^{-1} E_J\{\psi_{\text{RBD}}(z_n; \tilde{\beta}_{n,0})\}$ is uniformly bounded, we have

$$D = \limsup_{n \rightarrow \infty} \|U_n^{-1/2} A_n E_J\{\text{IF}(\mathbf{Z}_n; \mathbf{T}, K_{n,0})\}\|^2 < \infty.$$

From Equation (A17)

$$\limsup_{n \rightarrow \infty} \alpha(K_{n,\epsilon}) = \alpha_0 + \epsilon^2 \mu_k D + o(\epsilon^2),$$

since $\sup_{\epsilon \in [0,C]} \limsup_{n \rightarrow \infty} |b^{(3)}(\epsilon)| \leq C$ from Lemma A9. We complete the proof. \square

Proof of Corollary 1. For Part (i), following the proof of Theorem 1, for any fixed z ,

$$\lim_{n \rightarrow \infty} \alpha(K_{n,\epsilon}) = \alpha_0 + \epsilon^2 \mu_k \|U^{-1/2} A \text{IF}(z; T, K_0)\|^2 + d(z, \epsilon),$$

where $d(z, \epsilon) = o(\epsilon^2)$. From the assumption that $\sup_{x \in \mathbb{R}^p} \|w(x)x\| \leq C$ and $\sup_{\mu \in \mathbb{R}} |q''(\mu)\sqrt{V(\mu)}/F'(\mu)| \leq C$, we know $D_1 \leq \infty$. Following the proof of Lemma A9, $\sup_{z \in \mathbb{R}} |d(z, \epsilon)| = o(\epsilon^2)$. We finished the proof of part (i).

Part (ii) is straightforward by applying Theorem 1 with $J = \Delta_{z_n}$. \square

Proof of Theorem 2. Lemma A7 implies that

$$\begin{aligned} & \sqrt{n}[\{U(\mathbb{K}_n)\}^{-1/2}\{A_n T(\mathbb{K}_n) - \mathbf{g}_0\} - \{U(\mathbb{K}_n)\}^{-1/2}(A_n \tilde{\beta}_{n,0} - \mathbf{g}_0) \\ & - U_n^{-1/2} A_n \{T(K_{n,\epsilon}) - \tilde{\beta}_{n,0}\}] \xrightarrow{\mathcal{L}} (\mathbf{0}, \mathbf{I}_k). \end{aligned}$$

From Lemmas A5 and A6,

$$\sqrt{n}[\{U(\mathbb{K}_n)\}^{-1/2}\{A_n T(\mathbb{K}_n) - \mathbf{g}_0\} - U_n^{-1/2}\{A_n T(K_{n,\epsilon}) - \mathbf{g}_0\}] \xrightarrow{\mathcal{L}} (\mathbf{0}, \mathbf{I}_k).$$

Then, W_n is asymptotically $\chi_k^2(\delta(\epsilon))$ with $\delta(\epsilon) = n\|U_n^{-1/2}\{A_n T(K_{n,\epsilon}) - \mathbf{g}_0\}\|^2$ under H_{1n} . Therefore, $\beta(K_{n,\epsilon}) = P(W_n > \eta_{1-\alpha_0} | H_{1n}) = 1 - H_k(\eta_{1-\alpha_0}; \delta(\epsilon)) + h(n, \epsilon)$, where $h(n, \epsilon) = \beta(K_{n,\epsilon}) - 1 + H_k(\eta_{1-\alpha_0}; \delta(\epsilon)) \rightarrow 0$ as $n \rightarrow \infty$ for any fixed ϵ . Let $b(\epsilon) = -H_k(\eta_{1-\alpha_0}; \delta(\epsilon))$. Then, for ϵ close to 0, we have

$$\begin{aligned} \beta(K_{n,\epsilon}) - \beta_0 &= b(\epsilon) - b(0) + h(n, \epsilon) - h(n, 0) \\ &= \epsilon b'(0) + \frac{1}{2} \epsilon^2 b''(\epsilon^*) + h(n, \epsilon) - h(n, 0), \end{aligned} \tag{A18}$$

where $0 < \epsilon^* < \epsilon$. Note that under H_{1n} , $\delta(0) = n\|U_n^{-1/2}(A_n \tilde{\beta}_{n,0} - \mathbf{g}_0)\|^2 = \mathbf{c}^T U_n^{-1} \mathbf{c}$. Then,

$$\begin{aligned} b'(0) &= \left. \frac{-\partial H_k(\eta_{1-\alpha_0}; \delta)}{\partial \delta} \right|_{\delta=\delta(0)} \left. \frac{\partial \delta(\epsilon)}{\partial \epsilon} \right|_{\epsilon=0} = 2\nu_k n \left\{ A_n \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \right\}^T \Big|_{\epsilon=0} U_n^{-1} \{A_n \tilde{\beta}_{n,0} - \mathbf{g}_0\} \\ &= 2\nu_k \sqrt{n} \left\{ A_n \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \right\}^T \Big|_{\epsilon=0} U_n^{-1} \mathbf{c}. \end{aligned}$$

From Lemma A8,

$$\left. \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \right|_{\epsilon=0} = 1/\sqrt{n} E_J \{ \text{IF}(\mathbf{Z}_n; T, K_{n,0}) \},$$

and hence,

$$b'(0) = 2\nu_k \mathbf{c}^T U_n^{-1} A_n E_J \{ \text{IF}(\mathbf{Z}_n; T, K_{n,0}) \}.$$

Since $\sup_{\epsilon \in [0,C]} \limsup_{n \rightarrow \infty} |b''(\epsilon)| \leq C$ under H_{1n} by Lemma A9, we have $\liminf_{n \rightarrow \infty} 1/2\epsilon^2 b''(\epsilon^*) = o(\epsilon)$ as $\epsilon \rightarrow 0$.

Since from Lemma A8, $\text{IF}(z_n; T, K_{n,0}) = -\mathbf{H}_n^{-1} E_J \{ \psi_{\text{RBD}}(z_n; \tilde{\beta}_{n,0}) \}$ is uniformly bounded,

$$|B| = \left| \liminf_{n \rightarrow \infty} 2\mathbf{c}^T U_n^{-1} A_n E_J \{ \text{IF}(\mathbf{Z}_n; T, K_{n,0}) \} \right| < \infty.$$

From Equation (A18), we complete the proof. \square

Proof of Corollary 2. The proof is similar to that for Corollary 1, using the results in Theorem 2. \square

Appendix B. List of Notations and Symbols

- A_n : $k \times (p_n + 1)$ matrix in hypotheses Equations (8) and (14)
- c : k dimensional vector in H_{1n} in Equation (14)
- $F(\cdot)$: link function
- G : bias-correction term in “robust-BD”
- \mathbb{G} : limit of $A_n A_n^T$, i.e. $A_n A_n^T \xrightarrow{n \rightarrow \infty} \mathbb{G}$
- \mathbf{H}_n : $\mathbf{H}_n = E_{K_{n,0}} \{p_2(Y; \tilde{X}_n^T \tilde{\beta}_{n,0}) w(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T\}$
- $\text{IF}(\cdot; \cdot, \cdot)$: influence function
- J : an arbitrary distribution in the contamination of Equation (10)
- $K_{n,0}$: true parametric distribution of \mathbf{Z}_n
- $K_{n,\epsilon}$: $K_{n,\epsilon} = (1 - \frac{\epsilon}{\sqrt{n}})K_{n,0} + \frac{\epsilon}{\sqrt{n}}J$, ϵ -contamination in Equation (10)
- \mathbb{K}_n : empirical distribution of $\{\mathbf{Z}_{ni}\}_{i=1}^n$
- $\ell_K(\cdot)$: expectation of robust-BD in Equation (11)
- $m(\cdot)$: conditional mean of Y given X_n in Equation (1)
- n : sample size
- p_n : dimension of β
- $p_i(\cdot; \cdot)$: i th order derivative of robust-BD
- $q(\cdot)$: generating q -function of BD
- $\mathbf{T}(\cdot)$: vector, a functional of estimator in Equation (12)
- U_n : $U_n = A_n \mathbf{H}_n^{-1} \Omega_n \mathbf{H}_n^{-1} A_n^T$
- $V(\cdot)$: conditional variance of Y given X_n in Equation (2)
- W_n : Wald-type test statistic in Equation (9)
- $w(\cdot)$: weight function
- X_n : explanatory variables
- Y : response variable
- $\mathbf{Z}_n = (\mathbf{X}_n^T, Y)^T$
- $\alpha(\cdot)$: level of the test
- $\beta(\cdot)$: power of the test
- $\tilde{\beta}_{n,0}$: true regression parameter
- Δ_{z_n} : probability measure which puts mass 1 at the point z_n
- ϵ : amount of contamination in Equation (10), positive constant
- $\psi_{\text{RBD}}(\cdot; \cdot)$: score vector in Equation (7)
- Ω_n : $\Omega_n = E_{K_{n,0}} \{p_1^2(Y; \tilde{X}_n^T \tilde{\beta}_{n,0}) w^2(\mathbf{X}_n) \tilde{X}_n \tilde{X}_n^T\}$
- $\rho_q(\cdot, \cdot)$: robust-BD in Equation (4)

References

1. Zhang, C.M.; Guo, X.; Cheng, C.; Zhang, Z.J. Robust-BD estimation and inference for varying-dimensional general linear models. *Stat. Sin.* **2012**, *24*, 653–673.
2. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*, 2nd ed.; Chapman & Hall: London, UK, 1989.
3. Morgenthaler, S. Least-absolute-deviations fits for generalized linear models. *Biometrika* **1992**, *79*, 747–754.
4. Ruckstuhl, A.F.; Welsh, A.H. Robust fitting of the binomial model. *Ann. Stat.* **2001**, *29*, 1117–1136.
5. Noh, M.; Lee, Y. Robust modeling for inference from generalized linear model classes. *J. Am. Stat. Assoc.* **2007**, *102*, 1059–1072.
6. Künsch, H.R.; Stefanski, L.A.; Carroll, R.J. Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models. *J. Am. Stat. Assoc.* **1989**, *84*, 460–466.

7. Stefanski, L.A.; Carroll, R.J.; Ruppert, D. Optimally bounded score functions for generalized linear models with applications to logistic regression. *Biometrika* **1986**, *73*, 413–424.
8. Bianco, A.M.; Yohai, V.J. Robust estimation in the logistic regression model. In *Robust Statistics, Data Analysis, and Computer Intensive Methods*; Springer: New York, NY, USA, 1996; pp. 17–34.
9. Croux, C.; Haesbroeck, G. Implementing the Bianco and Yohai estimator for logistic regression. *Comput. Stat. Data Anal.* **2003**, *44*, 273–295.
10. Heritier, S.; Ronchetti, E. Robust bounded-influence tests in general parametric models. *J. Am. Stat. Assoc.* **1994**, *89*, 897–904.
11. Cantoni, E.; Ronchetti, E. Robust inference for generalized linear models. *J. Am. Stat. Assoc.* **2001**, *96*, 1022–1030.
12. Bianco, A.M.; Martínez, E. Robust testing in the logistic regression model. *Comput. Stat. Data Anal.* **2009**, *53*, 4095–4105.
13. Ronchetti, E.; Trojani, F. Robust inference with GMM estimators. *J. Econom.* **2001**, *101*, 37–69.
14. Basu, A.; Mandal, N.; Martin, N.; Pardo, L. Robust tests for the equality of two normal means based on the density power divergence. *Metrika* **2015**, *78*, 611–634.
15. Lee, S.; Na, O. Test for parameter change based on the estimator minimizing density-based divergence measures. *Ann. Inst. Stat. Math.* **2005**, *57*, 553–573.
16. Kang, J.; Song, J. Robust parameter change test for Poisson autoregressive models. *Stat. Probab. Lett.* **2015**, *104*, 14–21.
17. Basu, A.; Ghosh, A.; Martin, N.; Pardo, L. Robust Wald-type tests for non-homogeneous observations based on minimum density power divergence estimator. *ArXiv Pre-print* **2017**, arXiv:1707.02333.
18. Ghosh, A.; Basu, A.; Pardo, L. Robust Wald-type tests under random censoring. *ArXiv* **2017**, arXiv:1708.09695.
19. Brègman, L.M. A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming. *U.S.S.R. Comput. Math. Math. Phys.* **1967**, *7*, 620–631.
20. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer: Berlin, Germany, 2001.
21. Zhang, C.M.; Jiang, Y.; Shang, Z. New aspects of Bregman divergence in regression and classification with parametric and nonparametric estimation. *Can. J. Stat.* **2009**, *37*, 119–139.
22. Huber, P. Robust estimation of a location parameter. *Ann. Math. Statist.* **1964**, *35*, 73–101.
23. Hampel, F.R.; Ronchetti, E.M.; Rousseeuw, P.J.; Stahel, W.A. *Robust Statistics: The Application Based on Influence Function*; John Wiley: New York, NY, USA, 1986.
24. Hampel, F.R. The influence curve and its role in robust estimation. *J. Am. Stat. Assoc.* **1974**, *69*, 383–393.
25. Fan, J.; Peng, H. Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Stat.* **2004**, *32*, 928–961.
26. van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press: Cambridge, UK, 1998.
27. Clarke, B.R. Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *Ann. Stat.* **1983**, *11*, 1196–1205.

