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 inference methods for the "general linear model" with a diverging number of parameters.

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 \((X_{ni}, Y_i), \ldots, (X_{nq}, Y_q)\) be i.i.d. observations from some underlying distribution \((X_n, Y)\) with \(X_n = (X'_1, \ldots, X'_{p,n})\) 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

\[
m(x_n) = \mathbb{E}(Y \mid X_n = x_n) = F^{-1}(\mathbf{x}_n^T \hat{\beta}_{n,0}),
\]

(1)

and

\[
\text{var}(Y \mid X_n = x_n) = V(m(x_n)),
\]

(2)

where \( F \) is a known link function, \( \hat{\beta}_{n,0} \in \mathbb{R}^{p_n+1} \) is the vector of unknown true regression parameters, \( \mathbf{x}_n = (1, X_n^T)^T \) and \( V(\cdot) \) is a known function. Note that the conventional generalized linear model (GLM) satisfying Equations (1) and (2) assumes that \( Y \mid X_n = x_n \) 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 \( Z_n = (X_n^T, Y)^T \) and \( \tilde{Z}_n = (\mathbf{x}_n^T, Y)^T \).

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 \( Q_q(Y, \mu) = (Y - \mu)^2 \). 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 \( Q_q(Y, \mu) = -2\{Y \log(\mu) + (1 - Y) \log(1 - \mu)\} \); \( q(\mu) = 2.0\min\{\mu, 1 - \mu\} \) 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

\[
q(\mu) = \int_a^\mu \frac{s - \mu}{V(s)} ds,
\]

(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

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

(4)

where \( \psi(\cdot) \) is a bounded odd function, such as the Huber \( \psi \)-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

\[
G'_1(m(x_n)) = \mathbb{E}\{\psi(r(Y, m(x_n))) \mid X_n = x_n\}.
\]

Based on robust-BD, the estimator of \( \hat{\beta}_{n,0} \) proposed in [1] is defined as

\[
\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^n \rho_q(Y_i, F^{-1}(X_n^T \beta))w(X_{ni}) \right\},
\]

(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 \( \beta_{n,0} \) is formulated according to the “robust quasi-likelihood function” defined as

\[
Q_{\text{RQL}}(x_n, y, \mu) = \left\{ \int_{s_0}^{s_n} \psi(r(y, s)) / \sqrt{V(s)} ds \right\} w(x_n) - \frac{1}{n} \sum_{j=1}^{n} \int_{s_0}^{s_n} \left[ E\{\psi(Y_j, s)|X_{nj}\} / \sqrt{V(s)} \right] ds \right\} w(X_{nj}),
\]

where \( \mu = F^{-1}(\tilde{x}_n^T \beta) \) and \( \mu_j = F^{-1}(\tilde{x}_{nj}^T \beta) \), \( j = 1, \ldots, n \). 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”.

\[
\begin{array}{ccc}
\rho_q(y, \mu)w(x_n) & \psi(r) = r & Q_q(y, \mu) \\
\text{robust-BD} & w(x_n) \equiv 1 & \text{classical-BD} \\
\hline
\Psi \text{ in Equation (3)} & q \in \text{Equation (3)} & q \in \text{Equation (3)} \\
- \Psi_{\text{RQL}}(x_n, y, \mu) & \psi(r) = r & - \Psi_{\text{QL}}(y, \mu) \\
\text{robust quasi-likelihood} & w(x_n) \equiv 1 & \text{classical (negative) quasi-likelihood} \\
\end{array}
\]

For the robust-BD, assume that

\[
p_j(y; \theta) = \frac{\partial}{\partial \theta} \rho_q(y, F^{-1}(\theta)), \quad j = 0, 1, \ldots,
\]

exist finitely up to any order required. For example, for \( j = 1 \),

\[
p_1(y; \theta) = \{ \psi(r(y, \mu)) - G_1'(\mu) \} \{ q''(\mu) \sqrt{V(\mu)} / F'(\mu) \},
\]

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

\[
\frac{1}{n} \sum_{i=1}^{n} \Psi_{\text{RBD}}(z_i; \tilde{\beta}) = 0,
\]

where the score vector is

\[
\Psi_{\text{RBD}}(z_i; \tilde{\beta}) = p_1(y; \theta)w(x_n)\tilde{x}_n,
\]

with \( \theta = \tilde{x}_n^T \tilde{\beta} \). 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,

\[
H_0 : A_n \tilde{\beta}_{n,0} = 0 \quad \text{versus} \quad H_1 : A_n \tilde{\beta}_{n,0} \neq 0,
\]

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 \( g_0 \) is a known \( k \times 1 \) vector.

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

\[
W_n = n(A_n \tilde{\beta} - g_0)^T (A_n \tilde{H}_n^{-1} \tilde{\Omega}_n \tilde{H}_n^{-1} A_n^T)^{-1} (A_n \tilde{\beta} - g_0),
\]
We then define a functional for the estimator of $\Delta$ where $\rho$ is an arbitrary distribution and $z$ is the probability measure which puts mass 1 at the point $(\cdot, 0)$. From [23] (Equation (2.1.6) on pp. 84), the influence function of $T(K)$ is defined for each fixed $n$. Particularly, $T(K_{n,0}) = \hat{\beta}_{n,0}$. Similarly, $\rho(\cdot, \cdot)$ and $\psi_{\text{RBD}}(\cdot, \cdot)$ are defined in Equations (4) and (7), respectively. It’s worth noting that the solution to $E_K\{\psi_{\text{RBD}}(Z_n; \hat{\beta}) = 0\}$ may not be unique, i.e., $S_K$ may contain more than one element. We then define a functional for the estimator of $\hat{\beta}_{n,0}$ as follows,

$$
T(K) = \min_{\hat{\beta} \in S_K} ||\hat{\beta} - \hat{\beta}_{n,0}||. 
$$

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

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

$$
\text{IF}(z_n; T, K_{n,0}) = \frac{d}{dt} T((1-t)K_{n,0} + t\Delta z_n)\big|_{t=0} = \lim_{t \downarrow 0} \frac{T((1-t)K_{n,0} + t\Delta z_n) - \hat{\beta}_{n,0}}{t},
$$

where $\Delta z_n$ 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:

$$
\text{IF}(z_n; T, K_{n,0}) = -H_n^{-1}\psi_{\text{RBD}}(z_n; \hat{\beta}_{n,0}),
$$

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 $Z_n$ following the “general linear model” characterized by Equations (1) and (2). To facilitate the discussion of robustness properties, we consider the $\epsilon$-contamination,

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

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

For a generic distribution $K$ of $Z_n$, define

$$
\ell_K(\hat{\beta}) = E_K\{\rho_q(Y, F^{-1}(\hat{X}^T_n; \hat{\beta}))w(X_n)\},
$$

$$
S_K = \{\hat{\beta} : E_K\{\psi_{\text{RBD}}(Z_n; \hat{\beta}) = 0\}\},
$$

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

$$
T(K) = \min_{\hat{\beta} \in S_K} ||\hat{\beta} - \hat{\beta}_{n,0}||. 
$$

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

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

$$
\text{IF}(z_n; T, K_{n,0}) = \frac{d}{dt} T((1-t)K_{n,0} + t\Delta z_n)\big|_{t=0} = \lim_{t \downarrow 0} \frac{T((1-t)K_{n,0} + t\Delta z_n) - \hat{\beta}_{n,0}}{t},
$$

where $\Delta z_n$ 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:

$$
\text{IF}(z_n; T, K_{n,0}) = -H_n^{-1}\psi_{\text{RBD}}(z_n; \hat{\beta}_{n,0}),
$$

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

$$
\hat{\Omega}_n = \frac{1}{n} \sum_{i=1}^{n} p_i^2(Y_i; \hat{X}^T_{ni}\hat{\beta})w^2(X_{ni})\hat{X}_{ni}\hat{X}^T_{ni},
$$

$$
\hat{H}_n = \frac{1}{n} \sum_{i=1}^{n} p_i^2(Y_i; \hat{X}^T_{ni}\hat{\beta})w(X_{ni})\hat{X}_{ni}\hat{X}^T_{ni}.
$$
where \( H_n = E_{K_n,0} \{ p_2(Y; \tilde{X}_n^T \tilde{\beta}_{n,0}) w(X_n) \tilde{X}_n \tilde{X}_n^T \} \). 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:

\[
\begin{align*}
\Omega_n &= E_{K_n,0} \{ p_2^2(Y; \tilde{X}_n^T \tilde{\beta}_{n,0}) w^2(X_n) \tilde{X}_n \tilde{X}_n^T \}, \\
U_n &= A_n H_n^{-1} \Omega_n H_n^{-1} A_n^T.
\end{align*}
\]

### 3.1. Asymptotic Level of \( W_n \) under Contamination

We now investigate the asymptotic level of the Wald-type test \( W_n \) under the \( \epsilon \)-contamination.

**Theorem 1.** Assume Conditions A0–A9 and B4 in Appendix A. Suppose \( p_n^6/n \to 0 \) as \( n \to \infty \), \( \sup_{\epsilon > 0} \sup_{x_n} E_f(\|w(X_n)\tilde{X}_n\|) \leq C \). Denote by \( \alpha(K_{n,e}) \) the level of \( W_n = n \{ A_n T(K_n) - g_0 \}^T (A_n H_n^{-1} \Omega_n H_n^{-1} A_n^T)^{-1} \{ A_n T(K_n) - g_0 \} \) when the underlying distribution is \( K_{n,e} \) in Equation (10) and by \( \alpha_0 \) the nominal level. Under \( H_0 \) in Equation (8), it follows that

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

where

\[
D = \limsup_{n \to \infty} \| U_n^{-1/2} A_n E_f \{ IF(Z_n; T, K_{n,0}) \} \|^2 < \infty,
\]

\[
\mu_k = -\frac{\partial}{\partial \delta} H_k(\eta_1 - \epsilon_0; \delta)|_{\delta = 0}, \quad H_k(\cdot; \delta) \text{ is the cumulative distribution function of a } \chi^2_k(\delta) \text{ distribution, and } \eta_1 - \epsilon_0 \text{ is the } 1 - \alpha_0 \text{ quantile of the central } \chi^2_k \text{ distribution.}
\]

Theorem 1 indicates that if the influence function for \( T(\cdot) \) is bounded, then the asymptotic level of \( W_n \) under the \( \epsilon \)-contamination is also bounded and close to the nominal level when \( \epsilon \) 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 \( \epsilon \)-contamination, assumption \( p_n^5/n = o(1) \) 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 \( \epsilon \)-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, \( \sup_{x_n \in \mathbb{R}^n} \| w(x_n) x_n \| \leq C \) and \( \sup_{\mu < \epsilon} | q''(\mu) \sqrt{V(\mu)} / F'(\mu) | \leq C \).
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 \( K_{n,e} = (1 - \epsilon/\sqrt{n})K_0 + \epsilon/\sqrt{n}A_n \) with \( z \in \mathbb{R}^p \) a fixed point, under \( H_0 \) in Equation (8), it follows that

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

where

\[
D_1 = \sup_{z \in \mathbb{R}^p} \| U^{-1/2} A_I F(z; T, K_0) \|^2 < \infty.
\]

If \( p_n \) diverges with \( p_n^2/n \to 0 \), for \( K_{n,e} = (1 - \epsilon/\sqrt{n})K_{n,0} + \epsilon/\sqrt{n}A_n \) with \( z_n \in \mathbb{R}^{pn} \) a sequence of deterministic points, then, under \( H_0 \) in Equation (8),

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

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

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

In Corollary 1, conditions \( \sup_{x_n \in \mathbb{R}^{p_n}} \| w(x_n)x_n \| \leq C \) 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 \( w(x_n) \) downweights the high leverage points and can be chosen as, e.g., \( w(x_n) = 1/(1 + \| x_n \|) \). 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 \( \sigma^2 \) 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, \( | q''(\mu) \sqrt{V(\mu)/F'(\mu)} | = 2\{ (1 - \mu) \}^{1/2} \leq C \) since \( \mu \in [0, 1] \). Likewise, if \( q(\mu) = 2\{ (1 - \mu) \}^{1/2} \) (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 \( \epsilon \)-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 \( \| x_n \|_\infty \leq C_0 \), 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

\[
H_{1_n} : A_n \tilde{\beta}_{n,0} - g_0 = n^{-1/2} \epsilon_n
\]

where \( \epsilon = (\epsilon_1, \ldots, \epsilon_k)^T \neq 0 \) is fixed.

Theorem 2. Assume Conditions A0–A9 and B4 in Appendix A. Suppose \( p_n^2/n \to 0 \) as \( n \to \infty \), \( \sup_{n} E_j(\| w(X_n)X_n \|) \leq C \). Denote by \( \beta(K_{n,e}) \) the power of \( W_n = n\{ A_n T(K_n) - \beta(K_{n,e}) \} \).
When the underlying distribution is $K_{n,e}$ in Equation (10) and by $\beta_0$ the nominal power. Under $H_{1n}$, in Equation (14), it follows that

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

where

$$B = \liminf_{n \to \infty} 2\epsilon^T U_n^{-1} A_n \{IF(Z_n; T, K_{n,0})\},$$

with $|B| < \infty$, $v_k = -\frac{\partial}{\partial \delta} H_k(\eta_{1-e_0}; \delta)|_{\delta = \epsilon^T U_n^{-1} \epsilon}$ and $H_k(\cdot; \delta)$ and $\eta_{1-e_0}$ 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$_{x_n \in \mathbb{R}^p} \|w(x_n) x_n\| \leq C$ and sup$_{\mu \in \mathbb{R}} |q''(\mu) \sqrt{V(\mu)} / F'(\mu)| \leq C$.

(i) If $p_n = p$, $A_n \equiv A$, $\beta_{n,0} \equiv \beta_0$, $K_{n,0} \equiv K_0$ and $U_n \equiv U$ are fixed, then, for $K_{n,e} = (1 - \epsilon / \sqrt{n}) K_0 + \epsilon / \sqrt{n} \Delta z$ with $z \in \mathbb{R}^p$ a fixed point, under $H_{1n}$ in Equation (14), it follows that

$$\inf_{z \in \mathbb{R}^p} \liminf_{n \to \infty} \beta(K_{n,e}) = \beta_0 + \epsilon v_k B_1 + o(\epsilon) \quad \text{as } \epsilon \to 0,$$

where

$$B_1 = \inf_{z \in \mathbb{R}^p} 2\epsilon^T U_n^{-1} A_{IF}(z; T, K_0),$$

with $|B_1| < \infty$.

(ii) If $p_n$ diverges with $p_n^2 / n \to 0$, for $K_{n,e} = (1 - \epsilon / \sqrt{n}) K_{n,0} + \epsilon / \sqrt{n} \Delta z_n$ with $z_n \in \mathbb{R}^p$ a sequence of deterministic points, then, under $H_{1n}$ in Equation (14),

$$\liminf_{C_0 > 0} \liminf_{n \to \infty} \beta(K_{n,e}) = \beta_0 + \epsilon v_k B_2 + o(\epsilon) \quad \text{as } \epsilon \to 0,$$

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

$$B_2 = \inf_{C_0 > 0} \inf_{z_n \in S_{C_0}} \liminf_{n \to \infty} 2\epsilon^T U_n^{-1} A_n IF(Z_n; T, K_{n,0}),$$

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 $\epsilon$ varies.

The robust-BD estimation utilizes the Huber $\psi$-function $\psi(\cdot)$ with $c = 1.345$ and the weight function $w(x_n) = 1 / (1 + \|x_n\|)$. Comparisons are made with the classical non-robust counterparts corresponding to using $\psi(r) = r$ and $w(x_n) \equiv 1$. 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 \( p_n = [4(n^{1/5.5} - 1)] \) and \( \tilde{\beta}_{n,0} = (0,2,0,\ldots,0)^T \), where \([ \cdot ]\) denotes the floor function. Generate \( X_{ni} = (X_{i1},\ldots,X_{ip})^T \) by \( X_{ij} \sim \text{Unif}[-0.5, 0.5] \). 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 \( Y^*_i = Y_i I(Y_i > 20) + 20 I(Y_i \leq 20) \) for \( i = 1,\ldots,k \), 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 \( \{U_i\} \sim \text{Unif}(0,1) \). Then, the proportion of contaminated data, \( k/n \), is equal to \( \epsilon/\sqrt{n} \) as in Equation (10), which implies \( \epsilon = k/\sqrt{n} \).

Consider the null hypothesis \( H_0 : A_n \tilde{\beta}_{n,0} = 0 \) with \( A_n = (0,0,0,1,0,\ldots,0) \). Figure 1 plots the empirical level of \( W_n \) versus \( \epsilon \). 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](image1.png) Figure 1. Observed level of \( W_n \) versus \( \epsilon \) 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 \( c = (1,\ldots,1)^T \) 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](image2.png) Figure 2. Observed power of \( W_n \) versus \( \epsilon \) 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, $Y_i | X_n = x_n \sim \text{Bernoulli}(m(x_n))$, where $\logit\{m(x_n)\} = \tilde{x}_n^T \tilde{\beta}_{0,0}$. Let $p_n = 2$, $\tilde{\beta}_{n,0} = (0, 1, 1)^T$ and $X_n \overset{i.i.d.}{\sim} N(0, I_{p_n})$. 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 $i = 1, \ldots, k$ 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 $\epsilon$ in Figure 3. We find that the level of the non-robust method diverges fast as $\epsilon$ increases. It’s also clear that the empirical level of the robust method is close to the nominal level when $\epsilon$ is small and increases slightly with $\epsilon$, which coincides with our results in Theorem 1.

![Figure 3](image_url)

**Figure 3.** Observed level of $W_n$ versus $\epsilon$ 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 $c = (1, \ldots, 1)^T$ a vector of ones. Figure 4 plots the power of the Wald-type test versus $\epsilon$, 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 $\epsilon$ 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

\[
\Omega_n, K, T(K') = E_{K} \{ p_1^2 (Y; \tilde{X}_n^T (K')) w^2 (X_n) \tilde{X}_n \tilde{X}_n^T \},
\]

\[
H_n, K, T(K') = E_{K} \{ p_2 (Y; \tilde{X}_n^T (K')) w (X_n) \tilde{X}_n \tilde{X}_n^T \}.
\]

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

Define the following matrices,

\[
U(K_{n, \epsilon}) = A_n H_{n, \epsilon}^{-1} \Omega_{n, \epsilon} H_{n, \epsilon}^{-1} A_n^T,
\]

\[
U(\tilde{K}_n) = A_n \tilde{H}_n^{-1} \tilde{\Omega}_n \tilde{H}_n^{-1} \tilde{A}_n^T.
\]
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 \( \psi'(r), \psi'(r), \psi''(r), \psi''(r) \) and \( \psi''(r)r \) 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. \( \| X_n \|_\infty \leq C \) almost surely if the underlying distribution is \( K_{n,0}. \)
A5. \( E_{K_{n,0}}(\tilde{X}_n \tilde{X}_n^T) \) 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 \( F^{-1}(\tilde{X}_n^T \tilde{\beta}) \) is bounded for all \( \tilde{\beta} \) in the subset and all \( \tilde{x}_n \) such that \( \| \tilde{x}_n \|_\infty \leq C, \) where \( C > 0 \) is a large enough constant.
A7. \( H_n \) is positive definite, with eigenvalues uniformly bounded away from 0.
A8. \( \Omega_n \) is positive definite, with eigenvalues uniformly bounded away from 0.
A9. \( \| H_n^{-1} \Omega_n \| \) is bounded away from \( \infty. \)

**Condition B.**

B4. \( \| X_n \|_\infty \leq C \) 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** (\( \| T(K_{n,e}) - \tilde{\beta}_{n,0} \| \)). Assume Conditions A0–A7 and B4. For \( K_{n,e} \) in Equation (10), \( \ell_{K}(\cdot) \) in Equation (11) and \( T(\cdot) \) in Equation (12), if \( p_n/n \to 0 \) as \( n \to \infty, \) then \( T(K_{n,e}) \) is a local minimizer of \( \ell_{K_{n,e}}(\tilde{\beta}) \) such that \( \| T(K_{n,e}) - \tilde{\beta}_{n,0} \| = O(\sqrt{p_n/n}). \) Furthermore, \( T(K_{n,e}) \) is unique.

**Proof.** We follow the idea of the proof in [25]. Let \( r_n = \sqrt{p_n/n} \) and \( \tilde{u}_n = (u_{0,1}, \ldots, u_{p_n})^T \in \mathbb{R}^{p_n+1}. \) First, we show that there exists a sufficiently large constant \( C \) such that, for large \( n, \) we have

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

To show Equation (A1), consider

\[
\ell_{K_{n,e}}(\tilde{\beta}_{n,0} + r_n \tilde{u}_n) - \ell_{K_{n,e}}(\tilde{\beta}_{n,0}) = E_{K_{n,e}} \left\{ \rho_{\tilde{y}}(Y, F^{-1}(\tilde{X}_n^T \tilde{\beta}_{n,0} + r_n \tilde{X}_n^T \tilde{u}_n)) w(X_n) \right\} \\
- \rho_{\tilde{y}}(Y, F^{-1}(\tilde{X}_n^T \tilde{\beta}_{n,0})) w(X_n) \\
= I_1,
\]

where \( \| \tilde{u}_n \| = C. \)

By Taylor expansion,

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

where

\[
I_{1,1} = r_n E_{K_{n,e}} \left\{ p_1(Y, \tilde{X}_n^T \tilde{\beta}_{n,0}) w(X_n) \tilde{X}_n^T \tilde{u}_n \right\},
\]

\[
I_{1,2} = r_n^2 / 2 E_{K_{n,e}} \left\{ p_2(Y, \tilde{X}_n^T \tilde{\beta}_{n,0}) w(X_n) (\tilde{X}_n^T \tilde{u}_n)^2 \right\},
\]

\[
I_{1,3} = r_n^3 / 6 E_{K_{n,e}} \left\{ p_3(Y, \tilde{X}_n^T \tilde{\beta}_{n,0}) w(X_n) (\tilde{X}_n^T \tilde{u}_n)^3 \right\},
\]

for \( \tilde{\beta}_{n} \) located between \( \tilde{\beta}_{n,0} \) and \( \tilde{\beta}_{n,0} + r_n \tilde{u}_n. \) Hence

\[
| I_{1,1} | \leq r_n E_{K_{n,e}} \left\{ p_1(Y, \tilde{X}_n^T \tilde{\beta}_{n,0}) w(X_n) \tilde{X}_n^T \| \tilde{u}_n \|ight\}
\]
\[ \frac{\epsilon}{\sqrt{n}} \left\| E_f \{ p_1(Y; \tilde{X}_n \tilde{\beta}_{n,0}) w(X_n) \tilde{X}_n \} \right\| \tilde{u}_n \right\| \leq C r_n \sqrt{p_n/n} \left\| \tilde{u}_n \right\|, \]

since \( \left\| E_f \{ p_1(Y; \tilde{X}_n \tilde{\beta}_{n,0}) w(X_n) \tilde{X}_n \} \right\| = O(\sqrt{p_n}) \) and \( E_{k,n} \{ p_1(Y; \tilde{X}_n \tilde{\beta}_{n,0}) w(X_n) \tilde{X}_n \} = 0 \). For \( l_{1,2} \) in Equation (A2),

\[ l_{1,2} = \frac{r_n^2}{2} E_{k,n} \{ p_2(Y; \tilde{X}_n \tilde{\beta}_{n,0}) w(X_n) (\tilde{X}_n \tilde{u}_n)^2 \} + \frac{r_n^2}{2} E_{k,n} \{ p_1(Y; \tilde{X}_n \tilde{\beta}_{n,0}) w(X_n) (\tilde{X}_n \tilde{u}_n)^2 \} - E_{k,n} \{ p_2(Y; \tilde{X}_n \tilde{\beta}_{n,0}) w(X_n) (\tilde{X}_n \tilde{u}_n)^2 \} = l_{1,2,1} + l_{1,2,2}, \]

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

\[ |l_{1,2,2}| \leq \frac{r_n^2}{\sqrt{n}} \left\| E_{k,n} \{ p_2(Y; \tilde{X}_n \tilde{\beta}_{n,0}) w(X_n) (\tilde{X}_n \tilde{u}_n)^2 \} \right\| \| \tilde{u}_n \|^2 \leq Cr_n^2 p_n \left\| \tilde{u}_n \right\|^2 / \sqrt{n}, \]

where \( \left\| E_f \{ p_2(Y; \tilde{X}_n \tilde{\beta}_{n,0}) w(X_n) (\tilde{X}_n \tilde{u}_n)^2 \} \right\| = O(p_n) \) and \( \left\| E_{k,n} \{ p_2(Y; \tilde{X}_n \tilde{\beta}_{n,0}) w(X_n) (\tilde{X}_n \tilde{u}_n)^2 \} \right\| = O(p_n) \). Thus,

\[ l_{1,2} = 2^{-1} r_n^2 \tilde{u}_n^T \mathbf{H}_n \tilde{u}_n + O(r_n^2 p_n / \sqrt{n}) \left\| \tilde{u}_n \right\|^2. \]  

(A3)

For \( l_{1,3} \) in Equation (A2), we observe that

\[ |l_{1,3}| \leq C r_n^3 E_{k,n} \left\{ \left| p_3(Y; \tilde{X}_n \tilde{\beta}_{n,0}) w(X_n) (\tilde{X}_n \tilde{u}_n)^3 \right| \right\} = O(r_n^3 p_n^{3/2}) \left\| \tilde{u}_n \right\|^3. \]

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

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

\[ E_{k,n} \frac{\partial}{\partial \tilde{\beta}} \rho_q(Y, F^{-1}(\tilde{X}_n \tilde{\beta})) w(X_n) \right\| = E_{k,n} \| p_1(Y; \tilde{X}_n \tilde{\beta}) w(X_n) \tilde{X}_n \| \leq C \sqrt{p_n} \]

and hence,

\[ \frac{\partial}{\partial \tilde{\beta}} E_{k,n} \{ \rho_q(Y, F^{-1}(\tilde{X}_n \tilde{\beta})) w(X_n) \} = E_{k,n} \{ \frac{\partial}{\partial \tilde{\beta}} \rho_q(Y, F^{-1}(\tilde{X}_n \tilde{\beta})) w(X_n) \} \]

\[ \frac{\partial^2}{\partial \tilde{\beta}^2} E_{k,n} \{ \rho_q(Y, F^{-1}(\tilde{X}_n \tilde{\beta})) w(X_n) \} = E_{k,n} \{ \frac{\partial^2}{\partial \tilde{\beta}^2} \rho_q(Y, F^{-1}(\tilde{X}_n \tilde{\beta})) w(X_n) \}. \]

Therefore,

\[ \frac{\partial^2}{\partial \tilde{\beta}^2} E_{k,n} \{ \rho_q(Y, F^{-1}(\tilde{X}_n \tilde{\beta})) w(X_n) \} = E_{k,n} \{ p_2(Y; \tilde{X}_n \tilde{\beta}) w(X_n) \tilde{X}_n \tilde{X}_n^T \}. \]
We know that the minimum eigenvalues of $I_1^\ast$ are uniformly bounded away from 0,

$$||I_1^\ast|| = \|E_{K_{n,\epsilon}}\{p_3(Y;\tilde{X}_n^T\tilde{\beta}_{n,0})w(X_n)\tilde{X}_n\tilde{X}_n^T\tilde{\beta}_{n,0}\})\| \leq C_p/n^{1/4} = o(1)$$

$$||I_2^\ast|| \leq \epsilon/\sqrt{n}||E_{K_{n,\epsilon}}\{p_2(Y;\tilde{X}_n^T\tilde{\beta})w(X_n)\tilde{X}_n\tilde{X}_n^T\}|| + ||E_1\{p_2(Y;\tilde{X}_n^T\tilde{\beta})w(X_n)\tilde{X}_n\tilde{X}_n^T\}|| \leq C_p/\sqrt{n} = o(1).$$

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

$$0 = \frac{\partial}{\partial \tilde{\beta}}E_{K_{n,\epsilon}}\{\rho_\ell(Y, F^{-1}(\tilde{X}_n^T\tilde{\beta}))w(X_n)\} = E_{K_{n,\epsilon}}\{\frac{\partial}{\partial \tilde{\beta}}\rho_\ell(Y, F^{-1}(\tilde{X}_n^T\tilde{\beta}))\}_{\tilde{\beta} = \tilde{\beta}_{n,\epsilon}}w(X_n)$$

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. \hfill $\square$

**Lemma A2** ($||T(\tilde{K}_{n,\epsilon}) - T(K_{n,\epsilon})||$). 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 $p^{1/4}/n \rightarrow 0$ as $n \rightarrow \infty$ and the distribution of $(X_n, Y)$ is $K_{n,\epsilon}$, then there exists a unique local minimizer $\hat{\beta}_n$ of $\ell_{K_{n,\epsilon}}(\tilde{\beta})$ such that $||\hat{\beta}_n - T(K_{n,\epsilon})|| = O_p(\sqrt{p_n}/n)$. Furthermore, $||\hat{\beta}_n - \beta_{n,0}|| = O_p(\sqrt{p_n}/n)$ and $T(K_{n,\epsilon}) = \hat{\beta}$.

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

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

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

$$\ell_{K_{n,\epsilon}}(T(K_{n,\epsilon}) + r_n\tilde{u}_n) - \ell_{K_{n,\epsilon}}(T(K_{n,\epsilon})) = \frac{1}{n}\sum_{i=1}^n \{\rho_\ell(Y_i, F^{-1}(\tilde{X}_{ni}^T(T(K_{n,\epsilon}) + r_n\tilde{u}_n)))w(X_{ni})$$

$$- \rho_\ell(Y_i, F^{-1}(\tilde{X}_{ni}^T(T(K_{n,\epsilon}))))w(X_{ni})\}$$

$$= I_1,$$

where $||\tilde{u}_n|| = C_k$.

By Taylor expansion,

$$I_1 = I_{1,1} + I_{1,2} + I_{1,3}, \quad (A5)$$

where

$$I_{1,1} = r_n/n \sum_{i=1}^n p_1(Y_i; \tilde{X}_{ni}^T T(K_{n,\epsilon})\tilde{u}_n)w(X_{ni})\tilde{X}_{ni}^T\tilde{u}_n,$$
where \( \beta_n^* \) located between \( T(K_n, \varepsilon) \) and \( T(K_n, \varepsilon) + r_n \tilde{u}_n \).

Since \( \|T(K_n, \varepsilon) - \beta_n, 0\| = O(\sqrt{p_n/n}) = o(1) \), the large open set considered in Condition A6 contains \( T(K_n, \varepsilon) \) 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 \( T(K_n, \varepsilon) \) such that for all \( \beta \) in this set, \( \|p_1(Y; X_n^T \tilde{\beta})w(X_n) \tilde{X}_n \| \leq C \| \tilde{X}_n \| \) which is integrable with respect to \( K_n, \varepsilon \), where \( C \) is a positive constant. Thus, for \( n \geq N \),

\[
\theta = \frac{\partial}{\partial \beta} E_{K_n, \varepsilon} \{ p(Y, F^{-1}(X_n^T \tilde{\beta})) w(X_n) \} \bigg|_{\beta = T(K_n, \varepsilon)} = E_{K_n, \varepsilon} \{ p_1(Y; X_n^T T(K_n, \varepsilon)) w(X_n) \tilde{X}_n \}. \tag{A6}
\]

Hence,

\[
|I_{1,1}| \leq r_n \left\| \frac{1}{n} \sum_{i=1}^{n} p_1(Y_i; \tilde{X}_n^T T(K_n, \varepsilon)) w(X_n) \tilde{X}_n \right\| \| \tilde{u}_n \| = O_p(r_n \sqrt{p_n/n}) \| \tilde{u}_n \|.
\]

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

\[
I_{1,2} = \frac{r_n^2}{2n} \sum_{i=1}^{n} E_{K_n, \varepsilon} \{ p_2(Y_i; \tilde{X}_n^T T(K_n, \varepsilon)) w(X_n) (\tilde{X}_n^T \tilde{u}_n)^2 \}
+ \frac{r_n^2}{2n} \sum_{i=1}^{n} \left\| p_2(Y_i; \tilde{X}_n^T T(K_n, \varepsilon)) w(X_n) (\tilde{X}_n^T \tilde{u}_n)^2 \right\|.
\]

where \( I_{1,2,1} = 2^{-1/2} r_n^2 \tilde{u}_n^T H_{n, \varepsilon} \tilde{u}_n \). Meanwhile, we have

\[
|I_{1,2,2}| \leq \frac{r_n^2}{2} \left\| \frac{1}{n} \sum_{i=1}^{n} \left[p_2(Y_i; \tilde{X}_n^T T(K_n, \varepsilon)) w(X_n) \tilde{X}_n \tilde{X}_n^T \right. \right. \left. \left. - E_{K_n, \varepsilon} \{ p_2(Y_i; \tilde{X}_n^T T(K_n, \varepsilon)) w(X_n) \tilde{X}_n \tilde{X}_n^T \} \right] \right\| \| \tilde{u}_n \|^2
= r_n^2 O_p(p_n/\sqrt{n}) \| \tilde{u}_n \|^2.
\]

Thus,

\[
I_{1,2} = 2^{-1/2} r_n^2 \tilde{u}_n^T H_{n, \varepsilon} \tilde{u}_n + O_p(r_n^2 p_n/\sqrt{n}) \| \tilde{u}_n \|^2. \tag{A7}
\]

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

\[
|I_{1,3}| \leq C r_n^3 \frac{1}{n} \sum_{i=1}^{n} |p_3(Y_i; \tilde{X}_n^T \tilde{\beta}_n^*) w(X_n) (\tilde{X}_n^T \tilde{u}_n)|^2 = O_p(r_n^3 p_n^2/\sqrt{n}) \| \tilde{u}_n \|^3.
\]

We will show that the minimum eigenvalue of \( H_{n, \varepsilon} \) is uniformly bounded away from 0. \( H_{n, \varepsilon} = (1-\varepsilon/\sqrt{n}) H_{n, \varepsilon, T(K_n, \varepsilon)} + \varepsilon/\sqrt{n} H_{n, \varepsilon, T(K_n, \varepsilon)} \). Note

\[
\|H_{n, \varepsilon, T(K_n, \varepsilon)} - H_n\|
= \|E_{K_n, \varepsilon} \{ p_2(Y; \tilde{X}_n^T T(K_n, \varepsilon)) - p_2(Y; \tilde{X}_n^T \tilde{\beta}_n, 0) \} w(X_n) \tilde{X}_n \tilde{X}_n^T \|
= \|E_{K_n, \varepsilon} \{ p_2(Y; \tilde{X}_n^T \tilde{\beta}_n^*) w(X_n) \tilde{X}_n \tilde{X}_n^T \{ T(K_n, \varepsilon) - \tilde{\beta}_n, 0 \} \} \| = O(p_n^2/\sqrt{n}).
\]
Since the eigenvalues of $H_n$ are uniformly bounded away from 0, so are those of $H_{n, K_n, 0, T(K_{n,e})}$ and $H_{n,e}$.

We can choose some large $C_x$ 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 $\|\tilde{\beta} - T(K_{n,e})\| = O(n^{-1/4} p_n^{-1/2})$,

\[
\frac{1}{n} \sum_{i=1}^{n} p_2(y; \tilde{X}_n^T \tilde{\beta})w(x_{ni})\tilde{X}_n \tilde{X}_n^T
= E_{K_{n,0}} \{ p_2(y; \tilde{X}_n^T \tilde{\beta})w(x_n)\tilde{X}_n \tilde{X}_n^T \}
+ E_{K_{n,0}} \{ [p_2(y; \tilde{X}_n^T \tilde{\beta}) - p_2(y; \tilde{X}_n^T \tilde{\beta}_{n,0})]w(x_n)\tilde{X}_n \tilde{X}_n^T \}
+ E_{K_{n,0}} \{ p_2(y; \tilde{X}_n^T \tilde{\beta})w(x_n)\tilde{X}_n \tilde{X}_n^T - E_{K_{n,0}} \{ p_2(y; \tilde{X}_n^T \tilde{\beta})w(x_n)\tilde{X}_n \tilde{X}_n^T \} \}
+ 1/n \sum_{i=1}^{n} p_2(y; \tilde{X}_n^T \tilde{\beta})w(x_{ni})\tilde{X}_n \tilde{X}_n^T
= I_1 + I_2 + I_3 + I_4.
\]

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 $\|I_4^*\| = O(p_n / \sqrt{n})$.

Hence, for $n$ large enough, $\frac{\partial^2}{\partial \tilde{\beta}^2} \ell_{K_n}(\tilde{\beta})$ 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 $\ell_{K_n}(\tilde{\beta})$ in the $n^{-1/4} p_n^{-1/2}$ neighborhood of $T(K_{n,e})$ which covers $\tilde{\beta}$. □

**Lemma A3** ($\|A_n \{ T(K_{n,e}) - \tilde{\beta}_{n,0} \} \|$). Assume Conditions A0-A7 and B4. For $K_{n,e}$ in Equation (10) and $T(\cdot)$ in Equation (12), if $p_n^2/n \to 0$ as $n \to \infty$, the distribution of $(X_n, Y)$ is $K_{n,e}$ and $E_1(\|w(x_n)X_n\|) \leq C$, then

\[
\sqrt{n} A_n \{ T(K_{n,e}) - \tilde{\beta}_{n,0} \} = O(1),
\]

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

**Proof.** Taylor’s expansion yields

\[
0 = E_{K_{n,e}} \{ p_1(y; \tilde{X}_n^T T(K_{n,e}))w(x_n)\tilde{X}_n \}
= E_{K_{n,e}} \{ p_1(y; \tilde{X}_n^T \tilde{\beta}_{n,0})w(x_n)\tilde{X}_n \}
+ E_{K_{n,e}} \{ p_2(y; \tilde{X}_n^T \tilde{\beta}_{n,0})w(x_n)\tilde{X}_n \tilde{X}_n^T \} \{ T(K_{n,e}) - \tilde{\beta}_{n,0} \}
+ 1/2 E_{K_{n,e}} \{ p_3(y; \tilde{X}_n^T \tilde{\beta}_{n,0})w(x_n)\tilde{X}_n [\tilde{X}_n^T \{ T(K_{n,e}) - \tilde{\beta}_{n,0} \}]^2 \}
= I_1 + I_2 + I_3 + I_4,
\]

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

\[
\|I_1\| = O(1/\sqrt{n}), \quad \|I_2 - H_n\| = O(p_n / \sqrt{n}), \quad \|I_3\| = O(p_n^{5/2} / n).
\]

First, $\|I_1\| = c / \sqrt{n} \| E_1 \{ p_1(y; \tilde{X}_n^T T(K_{n,e}))w(x_n)\tilde{X}_n \} \| \leq C e / \sqrt{n} E_1(\|w(x_n)X_n\|) = O(1/\sqrt{n})$. Following the proof of $I_3^*$ in Lemma A1, $\|I_2 - H_n\| = O(p_n / \sqrt{n})$. Since $\|T(K_{n,e}) - \tilde{\beta}_{n,0}\| = O(\sqrt{p_n}/n)$, we have $\|I_3\| = O(p_n^{5/2} / n)$.

Therefore, $\sqrt{n} A_n \{ T(K_{n,e}) - \tilde{\beta}_{n,0} \} = -\sqrt{n} A_n H_n^{-1} I_4 + o(1)$, which completes the proof. □
Lemma A4 (asymptotic normality of $T(K_n) - T(K_{n,e})$). Assume Conditions A0–A8 and B4. If $p_n^5/n \to 0$ as $n \to \infty$ and the distribution of $(X_n, Y)$ is $K_{n,e}$, then

$$\sqrt{n}\{U(K_{n,e})\}^{-1/2} A_n \{T(K_n) - T(K_{n,e})\} \xrightarrow{D} N(0, I_k),$$

where $U(K_{n,e}) = A_n H_{n,e}^{-1} \Omega_{n,e} H_{n,e}^{-1} A_n^T$, $A_n$ is any given $k \times (p_n + 1)$ matrix such that $A_n A_n^T \to G$, with $G$ being a $k \times k$ positive-definite matrix, $k$ is a fixed integer.

**Proof.** We will first show that

$$T(K_n) - T(K_{n,e}) = -\frac{1}{n} H_{n,e}^{-1} \sum_{i=1}^{n} p_1(Y_i; \bar{X}_n^T T(K_{n,e})) w(X_n) \bar{X}_n + o_p(n^{-1/2}). \quad (A8)$$

From $\frac{\partial^2 \log(\tilde{p})}{\partial \beta^2} |_{\tilde{p}=T(K_n)} = 0$, Taylor’s expansion yields

$$0 = \left\{ \frac{1}{n} \sum_{i=1}^{n} p_1(Y_i; \bar{X}_n^T T(K_{n,e})) w(X_n) \bar{X}_n \right\} + \left\{ \frac{1}{n} \sum_{i=1}^{n} p_1(Y_i; \bar{X}_n^T T(K_{n,e})) w(X_n) \bar{X}_n \bar{X}_n^T \right\} \{ T(K_n) - T(K_{n,e}) \} + \frac{1}{2n} \sum_{i=1}^{n} p_3(Y_i; \bar{X}_n^T) \bar{X}_n \{ T(K_n) - T(K_{n,e}) \} \right\}^2 \bar{X}_n$$

$$\equiv \left\{ \frac{1}{n} \sum_{i=1}^{n} p_1(Y_i; \bar{X}_n^T T(K_{n,e})) w(X_n) \bar{X}_n \right\} + I_2 \{ T(K_n) - T(K_{n,e}) \} + I_3. \quad (A9)$$

where $\tilde{p}_n$ lies between $T(K_{n,e})$ and $T(K_n)$. Below, we will show

$$\|I_2 - H_{n,e}\| = O_p(p_n / \sqrt{n}), \quad \|I_3\| = O_p(p_n^{5/2} / n).$$

Similar arguments for the proof of $I_{1,2}$ of Lemma A2, we have $\|I_2 - H_{n,e}\| = O_p(p_n / \sqrt{n})$.

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 $\|T(K_n) - T(K_{n,e})\| = O_p(\sqrt{p_n/n})$, we see that

$$H_{n,e} \{ T(K_n) - T(K_{n,e}) \} = -\frac{1}{n} \sum_{i=1}^{n} p_1(Y_i; \bar{X}_n^T T(K_{n,e})) w(X_n) \bar{X}_n + u_n,$$

where $\|u_n\| = O_p(p_n^{5/2} / n) = o_p(n^{-1/2}).$ From the proof of Lemma A2, the eigenvalues of $H_{n,e}$ are uniformly bounded away from 0 and we complete the proof of Equation (A8).

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

$$A_n \{ T(K_n) - T(K_{n,e}) \} = -\frac{1}{n} A_n H_{n,e}^{-1} \sum_{i=1}^{n} p_1(Y_i; \bar{X}_n^T T(K_{n,e})) w(X_n) \bar{X}_n + o_p(n^{-1/2}).$$

It follows that

$$\sqrt{n}\{U(K_{n,e})\}^{-1/2} A_n \{ T(K_n) - T(K_{n,e}) \} = \sum_{i=1}^{n} R_{ni} + o_p(1),$$

where $R_{ni} = -n^{-1/2} \{U(K_{n,e})\}^{-1/2} A_n H_{n,e}^{-1} p_1(Y_i; \bar{X}_n^T T(K_{n,e})) w(X_n) \bar{X}_n.$ Following (A6) in Lemma A2, one can show that $E_{K_{n,e}}(R_{ni}) = 0$ for $n$ large enough.
To show \( \sum_{i=1}^{n} R_{ni} \xrightarrow{\mathcal{L}} N(0, I_k) \), we apply the Lindeberg-Feller central limit theorem in [26]. Specifically, we check (I) \( \sum_{i=1}^{n} \text{cov}_{n,i}(R_{ni}) \to I_k \); (II) \( \sum_{i=1}^{n} \mathbb{E}_{K_{ne}}(\|R_{ni}\|^{2+\delta}) = o(1) \) for some \( \delta > 0 \). Condition (I) is straightforward since \( \sum_{i=1}^{n} \text{cov}_{n,i}(R_{ni}) = \{U(K_{ne})\}^{-1/2}U(K_{ne})(U(K_{ne}))^{-1/2} = I_k \).

To check condition (II), we can show that \( \mathbb{E}_{K_{ne}}(\|R_{ni}\|^{2+\delta}) = O((p_{n}/n)^{(2+\delta)/2}) \). This yields \( \sum_{i=1}^{n} \mathbb{E}_{K_{ne}}(\|R_{ni}\|^{2+\delta}) \leq O(p_{n}^{2+\delta}/n^{\delta/2}) = o(1) \). Hence

\[
\sqrt{n}\{U(K_{ne})\}^{-1/2}A_{n}\{T(K_{n}) - T(K_{n,e})\} \xrightarrow{\mathcal{L}} N(0, I_k).
\]

Thus, we complete the proof.  

**Lemma A5** (asymptotic covariance matrices \( U(K_{ne}) \) and \( U_n \)). Assume Conditions A0–A9 and B4. If \( p_{n}^{2} / n \to 0 \) as \( n \to \infty \), then

\[
\|U_{n}^{1/2}\{U(K_{ne})\}^{1/2} - I_{k}\| = O(p_{n}^{1/4}),
\]

where \( U(K_{ne}) = A_{n}H^{-1}_{n,e}\Omega_{n,e}H_{n,e}^{-1}A_{n}^{T} \), \( A_{n} \) is any given \( k \times (p_{n} + 1) \) matrix such that \( A_{n}A_{n}^{T} \to G \), with \( G \) being a \( k \times k \) positive-definite matrix, and \( k \) is a fixed integer.

**Proof.** Note that

\[
\|\{U(K_{ne})\}^{1/2} - U_{n}^{1/2}\|^{2} \leq \|U(K_{ne}) - U_{n}\|^{2} \leq \|H_{n,e}^{-1}\Omega_{n,e}H_{n,e}^{-1} - H_{n}^{-1}\Omega_{n}H_{n}^{-1}\|\|A_{n}\|^{2}.
\]

Since \( A_{n} \frac{2}{n} \to \text{tr}(G) \), it suffices to prove that \( \|H_{n,e}^{-1}\Omega_{n,e}H_{n,e}^{-1} - H_{n}^{-1}\Omega_{n}H_{n}^{-1}\| = O(p_{n}^{2} / \sqrt{n}) \). First, we prove \( \|H_{n,e} - H_{n}\| = O(p_{n}^{2} / \sqrt{n}) \). Note that

\[
H_{n,e} - H_{n} = E_{K_{ne}}\left\{ \left[ \{p_{2}(Y; \tilde{X}_{n}^{T}T(K_{n,e})) - p_{2}(Y; \tilde{X}_{n}^{T}\tilde{\beta}_{n,0})\}w(X_{n})\tilde{X}_{n}\tilde{X}_{n}^{T} \right]
+ \left[ E_{K_{ne}}\left\{ \{p_{2}(Y; \tilde{X}_{n}^{T}\tilde{\beta}_{n,0})\}w(X_{n})\tilde{X}_{n}\tilde{X}_{n}^{T} \right\} \right] - H_{n}\right\}.
\]

We show that \( \|I_{1}\| = O(p_{n}^{2} / \sqrt{n}) \) and \( \|I_{2}\| = O(p_{n}^{2} / \sqrt{n}) \). Thus, \( \|I_{1}\| = O(p_{n}^{2} / \sqrt{n}) \). Second, we show \( \Omega_{n,e} - \Omega_{n} = O(p_{n}^{2} / \sqrt{n}) \). It is easy to see that

\[
\Omega_{n,e} - \Omega_{n} = E_{K_{ne}}\left\{ \left[ \{p_{3}(Y; \tilde{X}_{n}^{T}T(K_{n,e})) - p_{3}(Y; \tilde{X}_{n}^{T}\tilde{\beta}_{n,0})\}w^{2}(X_{n})\tilde{X}_{n}\tilde{X}_{n}^{T} \right]
+ \left[ E_{K_{ne}}\left\{ \{p_{3}(Y; \tilde{X}_{n}^{T}\tilde{\beta}_{n,0})\}w^{2}(X_{n})\tilde{X}_{n}\tilde{X}_{n}^{T} \right\} \right] - \Omega_{n}\right\} = \Delta_{1,1} + \Delta_{1,2},
\]

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

Third, we show \( \|H_{n,e}^{-1}\Omega_{n,e}H_{n,e}^{-1} - H_{n}^{-1}\Omega_{n}H_{n}^{-1}\| = O(p_{n}^{2} / \sqrt{n}) \). Note \( H_{n,e}^{-1}\Omega_{n,e}H_{n,e}^{-1} - H_{n}^{-1}\Omega_{n}H_{n}^{-1} = L_{1} + L_{2} + L_{3} \), where \( L_{1} = H_{n,e}^{-1}(\Omega_{n,e} - \Omega_{n})H_{n,e}^{-1}, L_{2} = H_{n,e}^{-1}(H_{n} - H_{n,e})H_{n,e}^{-1}\Omega_{n}H_{n,e}^{-1}, \) and \( L_{3} = H_{n,e}^{-1}\Omega_{n}H_{n,e}^{-1}(H_{n} - H_{n,e})H_{n,e}^{-1} \). Under Conditions A7 and A9, it is straightforward to see that \( \|H_{n,e}^{-1}\| = O(1), \|H_{n}^{-1}\| = O(1) \) and \( \|H_{n,e}^{-1}\Omega_{n}\| = O(1) \). Since \( \|L_{1}\| \leq \|H_{n,e}^{-1}\|\|\Omega_{n,e} - \Omega_{n}\||\|H_{n,e}^{-1}\| \), 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, \( \|H_{n,e}^{-1}\Omega_{n,e}H_{n,e}^{-1} - H_{n}^{-1}\Omega_{n}H_{n}^{-1}\| = O(p_{n}^{2} / \sqrt{n}) \).

Thus, we can conclude that \( \|U(K_{ne}) - U_{n}\| = O(p_{n}^{2} / \sqrt{n}) \) and that the eigenvalues of \( U(K_{ne}) \) and \( U_{n} \) are uniformly bounded away from 0 and \( \infty \). Consequently, \( \|\{U(K_{ne})\}^{1/2} - U_{n}^{1/2}\| = O(p_{n}^{1/4}) \) and proof is finished. \( \square \)
Lemma A6 (asymptotic covariance matrices $U(K_n)$ and $U(K_{n,e})$). Assume Conditions A0–A9 and B4. If $p_n^4/n \to 0$ as $n \to \infty$ and the distribution of $(X_n,Y)$ is $K_{n,e}$, then

$$||\{U(K_n)\}^{-1/2}\{U(K_{n,e})\}^{1/2} - I_k|| = O_p(p_n/n^{1/4}),$$

where $U(K_{n,e}) = A_nH_{n,e}^{-1}L_n^{-1}A_n^TA_n^T, U(K_n) = A_n\tilde{H}_n^{-1}\tilde{L}_n^{-1}\tilde{A}_n^T\tilde{A}_n^T, A_n$ is any given $k \times (p_n + 1)$ matrix such that $A_nA_n^T \to G$, with $G$ being a $k \times k$ positive-definite matrix, and $k$ is a fixed integer.

Proof. Note that $||\{U(K_n)\}^{1/2} - \{U(K_{n,e})\}^{1/2}||^2 \leq ||U(K_n) - U(K_{n,e})|| \leq ||\tilde{H}_n^{-1}\tilde{\Omega}_n\tilde{H}_n^{-1} - H_{n,e}^{-1}\Omega_{n,e}H_{n,e}^{-1}|| = O_p(p_n^2/\sqrt{n}).$

Following the proof of Proposition 1 in [1], we can show that $||\tilde{H}_n - H_{n,e}|| = O_p(p_n^2/\sqrt{n})$ and $||\tilde{\Omega}_n - \Omega_{n,e}|| = O_p(p_n^2/\sqrt{n}).$

To show $||H_{n,e}^{-1}\Omega_{n,e}H_{n,e}^{-1} - H_{n,e}^{-1}\Omega_{n,e}H_{n,e}^{-1}|| = O_p(p_n^2/\sqrt{n})$, note $\tilde{H}_n^{-1}\tilde{\Omega}_n\tilde{H}_n^{-1} - H_{n,e}^{-1}\Omega_{n,e}H_{n,e}^{-1} = L_1 + L_2 + L_3$, where $L_1 = \tilde{H}_n^{-1}(-\tilde{\Omega}_n - \Omega_{n,e})\tilde{H}_n^{-1}, L_2 = \tilde{H}_n^{-1}(\tilde{\Omega}_n - \tilde{H}_n)H_{n,e}^{-1}\tilde{\Omega}_n\tilde{H}_n^{-1}$ and $L_3 = \tilde{H}_n^{-1}\Omega_{n,e}H_{n,e}^{-1}(\tilde{H}_n - H_{n,e})\tilde{H}_n^{-1}$.

Following the proof in Lemma A2, it is straightforward to verify that $||H_{n,e}^{-1}|| = O(1), ||H_{n,e}^{-1}|| = O(p_n^2/\sqrt{n})$. In addition, $||H_{n,e}^{-1}\Omega_{n,e}|| = ||(-H_{n,e}^{-1})\Omega_{n,e} + H_{n,e}^{-1}(\Omega_{n,e} - \Omega_{n})|| = ||H_{n,e}^{-1}||\Omega_{n,e}||H_{n,e}^{-1}|| + ||H_{n,e}^{-1}||\Omega_{n,e} - ||\Omega_{n,e}|| + ||H_{n,e}^{-1}|| = O(1)$. Since $L_1 \leq ||H_{n,e}^{-1}||\Omega_{n,e} - ||\Omega_{n,e}|| + ||H_{n,e}^{-1}||$, 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, $||H_{n,e}^{-1}\Omega_{n,e}H_{n,e}^{-1} - H_{n,e}^{-1}\Omega_{n,e}H_{n,e}^{-1}|| = O(p_n^2/\sqrt{n}).$

Thus, we can conclude that $||\{U(K_n) - U(K_{n,e})\}^{1/2}||^2 \leq ||U(K_n) - U(K_{n,e})||$.

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

$$\sqrt{n}\{||U(K_n)\}^{-1/2}A_n\{T(K_n) - \tilde{\beta}_{n,0}\} - U_n^{-1/2}A_n\{T(K_{n,e}) - \tilde{\beta}_{n,0}\} \overset{d}{\to} N(0, I_k),$$

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

Proof. Note that

$$\sqrt{n}\{||U(K_n)\}^{-1/2}A_n\{T(K_n) - \tilde{\beta}_{n,0}\} - U_n^{-1/2}A_n\{T(K_{n,e}) - \tilde{\beta}_{n,0}\} = \sqrt{n}\{||U(K_n)\}^{-1/2}A_n\{T(K_n) - T(K_{n,e})\} + \sqrt{n}\{||U(K_n)\}^{-1/2} - \{||U(K_n)\}^{-1/2}\}A_n\{T(K_{n,e}) - \tilde{\beta}_{n,0}\}$$

$$= I + II + III.$$

For term I, we obtain from Lemma A4 that $\sqrt{n}\{||U(K_n)\}^{-1/2}A_n\{T(K_n) - T(K_{n,e})\} \overset{d}{\to} N(0, I_k)$. From Lemma A6, we get $||\{U(K_n)\}^{-1/2} - \{U(K_{n,e})\}^{-1/2}|| = o_p(1)$. Thus, by Slutsky theorem,

$$I \overset{d}{\to} N(0, I_k). \quad (A10)$$

For term II, we see from Lemma A6 that

$$||\{U(K_n)\}^{-1/2} - \{U(K_{n,e})\}^{-1/2}|| = O(p_n/n^{1/4}).$$

Since

$$||A_n\{T(K_{n,e}) - \tilde{\beta}_{n,0}\}|| \leq ||A_n||||T(K_{n,e}) - \tilde{\beta}_{n,0}|| = O(\sqrt{p_n/n}).$$
Thus,
\[
\| \| \leq \sqrt{n}\| \{U(K_n)\}^{-1/2} - \{U(K_n,e)\}^{-1/2} \| A_n \| \| T(K_n,e) - \tilde{\beta}_{n,0} \| = O_P(p_n^{3/2}/n^{1/4}). \tag{A11}
\]

Similarly, \( \| III \| = o_P(1) \). 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 \( n \),
\[
\frac{\partial}{\partial \ell} T((1 - t)K_{n,0} + \ell f) \bigg|_{\ell = 0} = \frac{1}{t - t_0} T((1 - t_0)K_{n,0} + t_0 f) - T((1 - t_0)K_{n,0})
\]
\[
= -H_{n,K_{0}}^{-1}(T(K_{0})) \left[ \mathbb{E} \{ \psi_{RBD}(Z_n; T(K_{0})) \} - \mathbb{E}_{K_{0}} \{ \psi_{RBD}(Z_n; T(K_{0})) \} \right],
\]
where \( K_{0} = (1 - t_0)K_{n,0} + t_0 f \) 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, \( \| H_{n,K_{0}}^{-1}(T(K_{0})) \| \leq C \) 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
\[
\lim_{t \to t_0} \frac{T((1 - t)K_{n,0} + t f) - T((1 - t_0)K_{n,0})}{t - t_0} = \lim_{\Delta \to 0} \frac{T(K_{0} + \Delta(j - K_{n,0})) - T(K_{0})}{\Delta},
\]
where \( \Delta = t - t_0 \).

It suffices to prove that for any sequence \( \{\Delta_j\}_{j=1}^{\infty} \) such that \( \lim_{j \to \infty} \Delta_j = 0 \), we have
\[
\lim_{j \to \infty} \frac{T(K_{0} + \Delta_j(j - K_{n,0})) - T(K_{0})}{\Delta_j} = -H_{n,K_{0}}^{-1}(T(K_{0})) \left[ \mathbb{E} \{ \psi_{RBD}(Z_n; T(K_{0})) \} - \mathbb{E}_{K_{0}} \{ \psi_{RBD}(Z_n; T(K_{0})) \} \right].
\]
Following similar proofs in Lemma A1, we can show that for \( t_0 \) sufficiently small,
\[
\| \tilde{\beta}_{n,0} - T(K_{0}) \| \leq C t_0 \sqrt{p_n}. \tag{A12}
\]

Next we will show that the eigenvalues of \( H_{n,K_{0},T(K_{0})} \) are bounded away from 0.
\[
H_{n,K_{0},T(K_{0})} = (1 - t_0)H_{n,K_{0},T(K_{0})} + t_0 H_{n,l,T(K_{0})}
\]
\[
= (1 - t_0)H_n + t_0 H_{n,l,T(K_{0})} + (1 - t_0) \{ H_{n,K_{0},T(K_{0})} - H_n \}
\]
\[
+ t_0 \{ H_{n,l,T(K_{0})} - H_{n,l,T(K_{0})} \} = (1 - t_0)I_1 + t_0 I_2 + I_3 + I_4.
\]

First,
\[
\| I_3 \| \leq C \mathbb{E}_{K_{0}} [ \{ p_2(Y; X_n^T T(K_{0})) - p_2(Y; X_n^T \tilde{\beta}_{n,0}) \} w(X_n) X_n X_n^T ] \|
\leq C p_n^{3/2} \| T(K_{0}) - \tilde{\beta}_{n,0} \| \leq C p_n^{3/2}\| T(K_{0}) \|.
\]

Similarly, \( \| I_2 \| \leq C p_n t_0 \) and \( \| I_4 \| \leq C p_n^2 t_0^2 \). 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 \( H_{n,K_{0},T(K_{0})} \) are uniformly bounded away from 0.
Define $K_j = K_{t_0} + \Delta_j$. Following similar arguments for (A6) in Lemma A2, for $j$ large enough, $E_{K_j} \{ \psi_{\text{RBD}}(Z_n; T(K_j)) \} = 0$. We will only consider $j$ large enough below. The two term Taylor expansion yields

$$0 = E_{K_j} \{ \psi_{\text{RBD}}(Z_n; T(K_j)) \} = E_{K_j} \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} + H_{n, K_j, \beta_j^*} \{ T(K_j) - T(K_{t_0}) \}, \tag{A13}$$

where $\beta_j^*$ lies between $T(K_{t_0})$ and $T(K_j)$.

Thus, from (A13) and the fact $E_{K_j} \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} = \Delta_j [ E_f \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} - E_{K_{t_0}} \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} ]$, we have

$$0 = E_{K_j} \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} + H_{n, K_j, \beta_j^*} \{ T(K_j) - T(K_{t_0}) \} + \left[ E_{K_{t_0}} \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} - E_{K_{t_0}} \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} \right] \sum H_{n, K_j, \beta_j^*} - H_{n, K_{t_0}, T(K_{t_0})} \{ T(K_j) - T(K_{t_0}) \},$$

and we obtain that

$$T(K_j) - T(K_{t_0}) = -\Delta_j H_{n, K_j, \beta_j^*}^{-1} \{ E_f \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} - E_{K_{t_0}} \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} \} - H_{n, K_j, \beta_j^*} - H_{n, K_{t_0}, T(K_{t_0})} \{ T(K_j) - T(K_{t_0}) \}. \tag{A14}$$

Next, we will show that $\| H_{n, K_j, \beta_j^*} - H_{n, K_{t_0}, T(K_{t_0})} \| = o(1)$ as $j \to \infty$ for any fixed $n$. Since $\| \beta_j^* - T(K_{t_0}) \| \leq \| T(K_j) - T(K_{t_0}) \| = O(\Delta_j),$

$$\| H_{n, K_j, \beta_j^*} - H_{n, K_{t_0}, \beta_j^*} \| = \Delta_j \| E_f \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} - E_{K_{t_0}} \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} \| \leq C \Delta_j$$

as $j \to \infty$,

and also,

$$\| H_{n, K_j, \beta_j^*} - H_{n, K_{t_0}, T(K_{t_0})} \| = \| E_{K_j} \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} - E_{K_{t_0}} \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} \| = o(1) \quad \text{as } j \to \infty. \tag{A15}$$

From Equations (A15) and (A16),

$$\| H_{n, K_j, \beta_j^*} - H_{n, K_{t_0}, T(K_{t_0})} \| = o(1) \quad \text{as } j \to \infty$$

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

$$\| T(K_j) - T(K_{t_0}) + \Delta_j H_{n, K_j, \beta_j^*}^{-1} \{ E_f \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} - E_{K_{t_0}} \{ \psi_{\text{RBD}}(Z_n; T(K_{t_0})) \} \| = o(\Delta_j).$$

This completes the proof. □

**Lemma A9.** Assume Conditions A1–A8 and B4 and sup\(_n\) $E_f(\| w(X_n) \bar{X}_n \|) \leq C$. Let $H_k(\cdot; \delta)$ be the cumulative distribution function of $\chi^2_k(\delta)$ distribution with $\delta$ the noncentrality parameter. Denote $\delta(e) = 1/\{ A_n H_k(T(x_n)) - g_0 \}^2$. Let $b(e) = -H_k(x; \delta(e))$. Then, for any fixed $x > 0$, sup\(_e\in[0,C]\) lim sup\(_n\to\infty\) \(|b^{(3)}(e)| \leq C \) under $H_0$ and sup\(_e\in[0,C]\) lim sup\(_n\to\infty\) \(|b''(e)| \leq C \) under $H_1$. 


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

\[
b'(e) = -\frac{\partial}{\partial e} H_k(x; \delta(e)) = \left\{ \frac{\partial}{\partial \delta} H_k(x; \delta) \right\}_{\delta = \delta(e)} \left\{ \frac{\partial \delta(e)}{\partial e} \right\} \left\{ \frac{\partial^2 \delta(e)}{\partial e^2} \right\} \]

\[
b''(e) = \left\{ \frac{\partial^2}{\partial \delta^2} H_k(x; \delta) \right\}_{\delta = \delta(e)} \left\{ \frac{\partial \delta(e)}{\partial e} \right\}^2 + \left\{ \frac{\partial}{\partial \delta} H_k(x; \delta) \right\}_{\delta = \delta(e)} \left\{ \frac{\partial^2 \delta(e)}{\partial e^2} \right\}
+ \left\{ \frac{\partial}{\partial \delta} H_k(x; \delta) \right\}_{\delta = \delta(e)} \left\{ \frac{\partial^3 \delta(e)}{\partial e^3} \right\}.
\]

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

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

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,

\[
\frac{\partial}{\partial \delta} H_k(x; \delta) = -\frac{e^{-\delta/2}}{2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{\gamma(j + k/2, x/2)}{\Gamma(j + k/2)} + \frac{e^{-\delta/2}}{2} \sum_{j=1}^{\infty} \frac{(\delta/2)^{j-1}}{j!} \frac{\gamma(j + k/2, x/2)}{\Gamma(j + k/2)}
\]

Since

\[
\frac{\gamma(j + 1 + k/2, x/2)}{\Gamma(j + 1 + k/2)} = \frac{\gamma(j + k/2, x/2)}{\Gamma(j + k/2)} - \frac{(x/2)^{j+k/2-1/2} e^{-x/2}}{\Gamma(j + k/2)}
\]

we have

\[
\frac{\partial}{\partial \delta} H_k(x; \delta) = -\frac{1}{2} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^{j+k/2-1/2} e^{-x/2}}{\Gamma(j + 1 + k/2)}
\]

\[
\frac{\partial^2}{\partial \delta^2} H_k(x; \delta) = \frac{1}{4} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^{j+k/2-1/2} e^{-x/2}}{\Gamma(j + 1 + k/2)} - \frac{1}{4} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^{j+k/2-1/2} e^{-x/2}}{\Gamma(j + 1 + k/2)}
\]

\[
\frac{\partial^3}{\partial \delta^3} H_k(x; \delta) = -\frac{1}{8} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^{j+k/2-1/2} e^{-x/2}}{\Gamma(j + 1 + k/2)} + \frac{1}{8} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^{j+k/2-1/2} e^{-x/2}}{\Gamma(j + 1 + k/2)}
\]

\[
\frac{1}{8} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} \frac{(x/2)^{j+k/2-1/2} e^{-x/2}}{\Gamma(j + 1 + k/2)}
\]
we conclude that $\sqrt{n} \| \delta'(\epsilon) \| (i = 1, 2, 3)$ are bounded as $n \to \infty$ for all $\epsilon \in [0, C]$. Now, we consider the derivatives of $\delta(\epsilon)$,

$$
\frac{\partial \delta(\epsilon)}{\partial \epsilon} = 2n \left\{ A_n \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \right\}^T U_n^{-1} \left\{ A_n T(K_{n,\epsilon}) - g_0 \right\}
$$

$$
\frac{\partial^2 \delta(\epsilon)}{\partial \epsilon^2} = 2n \left\{ A_n \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \right\}^T U_n^{-1} \left\{ A_n \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \right\} + 2n \left\{ A_n \frac{\partial^2 T(K_{n,\epsilon})}{\partial \epsilon^2} \right\}^T U_n^{-1} \left\{ A_n T(K_{n,\epsilon}) - g_0 \right\}
$$

$$
\frac{\partial^3 \delta(\epsilon)}{\partial \epsilon^3} = 6n \left\{ A_n \frac{\partial^2 T(K_{n,\epsilon})}{\partial \epsilon^2} \right\}^T U_n^{-1} \left\{ A_n \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \right\} + 2n \left\{ A_n \frac{\partial^3 T(K_{n,\epsilon})}{\partial \epsilon^3} \right\}^T U_n^{-1} \left\{ A_n T(K_{n,\epsilon}) - g_0 \right\}.
$$

To complete the proof, we only need to show that $\sqrt{n} \| \delta'/\partial \epsilon(\epsilon) \| (i = 1, 2, 3)$ are bounded as $n \to \infty$ for all $\epsilon \in [0, C]$, and $\sqrt{n} \| A_n T(K_{n,\epsilon}) - g_0 \| \text{ is bounded under } H_0 \text{ and } H_{1n} \text{ as } n \to \infty \text{ for all } \epsilon \in [0, C]$. The result for $\sqrt{n} \| A_n T(K_{n,\epsilon}) - g_0 \|$ is straightforward from Lemma A3.

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

$$
\sqrt{n} \frac{\partial}{\partial \epsilon} T(K_{n,\epsilon}) = -H^{-1}_{n,K_{n,\epsilon}} \{ E_f \{ p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(X_n) \tilde{X}_n \} - E_{K_{n,\epsilon}} \{ p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(X_n) \tilde{X}_n \} \}.
$$

Since $\| H^{-1}_{n,K_{n,\epsilon}} \| \leq C$, $\| E_f \{ p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(X_n) \tilde{X}_n \} \| \leq C \| w(X_n) \tilde{X}_n \| \leq C$ and

$$
\| E_{K_{n,\epsilon}} \{ p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(X_n) \tilde{X}_n \} \| \leq C \| w(X_n) \tilde{X}_n \| \leq C
$$

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

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

$$
\sqrt{n} \frac{\partial^2}{\partial \epsilon^2} T((1 - \epsilon / \sqrt{n}) K_{n,0} + \epsilon / \sqrt{n})
$$

$$
= -\frac{\partial}{\partial \epsilon} H^{-1}_{n,K_{n,\epsilon}} \{ E_f \{ p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(X_n) \tilde{X}_n \} - E_{K_{n,\epsilon}} \{ p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(X_n) \tilde{X}_n \} \} - H^{-1}_{n,K_{n,\epsilon}} \{ E_f \{ p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(X_n) \tilde{X}_n \} - E_{K_{n,\epsilon}} \{ p_1(Y; \tilde{X}_n^T T(K_{n,\epsilon})) w(X_n) \tilde{X}_n \} \} \}
$$

with

$$
\frac{\partial}{\partial \epsilon} H^{-1}_{n,K_{n,\epsilon}} = -H^{-1}_{n,K_{n,\epsilon}} \frac{\partial}{\partial \epsilon} H^{-1}_{n,K_{n,\epsilon}} \frac{H^{-1}_{n,K_{n,\epsilon}}}{H^{-1}_{n,K_{n,\epsilon}}}
$$

$$
\frac{\partial}{\partial \epsilon} H^{-1}_{n,K_{n,\epsilon}} = -\frac{1}{\sqrt{n}} E_{K_{n,\epsilon}} \{ p_2(Y; \tilde{X}_n T(K_{n,\epsilon})) w(X_n) \tilde{X}_n \tilde{X}_n^T \} + \frac{1}{\sqrt{n}} E_f \{ p_2(Y; \tilde{X}_n T(K_{n,\epsilon})) w(X_n) \tilde{X}_n \tilde{X}_n^T \}.
$$
\[(1 - \epsilon / \sqrt{n}) E_{K_{\epsilon}} \left\{ p_3(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{\epsilon,e}) \right\} \]

\[\sqrt{n} E_{\epsilon} \left\{ p_3(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{\epsilon,e}) \right\}.

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

\[\|\frac{\partial}{\partial \epsilon} \left[ E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} - E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} \right]\]

\[= \left\| E_{\epsilon} \left\{ p_2(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{\epsilon,e}) \right\} \right\| - E_{\epsilon} \left\{ p_2(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{\epsilon,e}) \right\} \right\| \]

\[\leq C \epsilon n / \sqrt{n}.\]

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

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

\[\sqrt{n} \frac{\partial^3}{\partial \epsilon^3} T((1 - \epsilon / \sqrt{n}) K_{\epsilon,0} + \epsilon / \sqrt{n}) \]

\[= \frac{\partial^2}{\partial \epsilon^2} H^{-1}_{n,K_{\epsilon,e},T(K_{\epsilon,e})} - 2 \frac{\partial}{\partial \epsilon} \left[ E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} - E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} \right] \]

\[\cdot \frac{\partial}{\partial \epsilon} \left[ E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} - E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} \right] \]

\[\cdot \frac{\partial}{\partial \epsilon} \left[ E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} - E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} \right]. \]

Note:

\[\frac{\partial^2}{\partial \epsilon^2} H^{-1}_{n,K_{\epsilon,e},T(K_{\epsilon,e})} = - \frac{\partial}{\partial \epsilon} \left[ E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} - E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} \right] \frac{\partial}{\partial \epsilon} \left[ E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} - E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} \right] \frac{\partial}{\partial \epsilon} \left[ E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} - E_{\epsilon} \left\{ p_1(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \right\} \right], \]

where

\[\frac{\partial^2}{\partial \epsilon^2} H^{-1}_{n,K_{\epsilon,e},T(K_{\epsilon,e})} \]

\[= - \frac{2}{\sqrt{n}} E_{K_{\epsilon}} \left\{ p_3(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{\epsilon,e}) \right\} \]

\[+ \frac{2}{\sqrt{n}} E_{\epsilon} \left\{ p_3(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{\epsilon,e}) \right\} \]

\[+ (1 - \epsilon / \sqrt{n}) E_{K_{\epsilon}} \left\{ p_4(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{\epsilon,e}) \right\} \]

\[+ \frac{\epsilon}{\sqrt{n}} E_{\epsilon} \left\{ p_4(Y; \tilde{X}_n^T T(K_{\epsilon,e})) w(X_n) \tilde{X}_n \tilde{X}_n^T \frac{\partial}{\partial \epsilon} T(K_{\epsilon,e}) \right\} \].
Hence, \( \| \frac{\partial^2}{\partial c^2} H_{n,k_n,e,T(K_n,e)} \| \leq C p_n^2 / n \) which implies that \( \| \frac{\partial^2}{\partial c^2} H_{n,k_n,e,T(K_n,e)} \| = o(1) \) for all \( c \in [0, C] \). In addition,

\[
\begin{align*}
\frac{\partial^2}{\partial c^2} \left[ E_f \{ p_1(Y; \hat{X}_n^T T(K_n,e))w(X_n) \hat{X}_n \} - E_{K_n,0} \{ p_1(Y; \hat{X}_n^T T(K_n,e))w(X_n) \hat{X}_n \} \right] \\
= \frac{\partial}{\partial c} \left[ E_f \{ p_2(Y; \hat{X}_n^T T(K_n,e))w(X_n) \hat{X}_n \hat{X}_n^T \frac{\partial}{\partial c} T(K_n,e) \} \\
- E_{K_n,0} \{ p_2(Y; \hat{X}_n^T T(K_n,e))w(X_n) \hat{X}_n \hat{X}_n^T \frac{\partial}{\partial c} T(K_n,e) \} \right] \\
= E_f \left[ p_3(Y; \hat{X}_n^T T(K_n,e))w(X_n) \hat{X}_n \left( \hat{X}_n^T \frac{\partial}{\partial c} T(K_n,e) \right)^2 \right] \\
+ E_f \left[ p_2(Y; \hat{X}_n^T T(K_n,e))w(X_n) \hat{X}_n \hat{X}_n^T \frac{\partial^2}{\partial c^2} T(K_n,e) \right] \\
- E_{K_n,0} \left[ p_3(Y; \hat{X}_n^T T(K_n,e))w(X_n) \hat{X}_n \left( \hat{X}_n^T \frac{\partial}{\partial c} T(K_n,e) \right)^2 \right] \\
- E_{K_n,0} \left[ p_2(Y; \hat{X}_n^T T(K_n,e))w(X_n) \hat{X}_n \hat{X}_n^T \frac{\partial^2}{\partial c^2} T(K_n,e) \right].
\end{align*}
\]

Hence, \( \| \frac{\partial^2}{\partial c^2} \left[ E_f \{ p_1(Y; \hat{X}_n^T T(K_n,e))w(X_n) \hat{X}_n \} - E_{K_n,0} \{ p_1(Y; \hat{X}_n^T T(K_n,e))w(X_n) \hat{X}_n \} \| \leq C p_n / \sqrt{n} \).

Therefore, \( \| \sqrt{n} \frac{\partial}{\partial c} T((1 - e / \sqrt{n}) K_n,0 + e / \sqrt{n}) \| = o(1) \) for all \( e \in [0, C] \). Hence, we complete the proof. \( \Box \)

**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 \( \chi_k^2 \) with noncentrality parameter \( \delta(e) = n \| U_n^{-1/2} [ A_n T(K_n,e) - g_0 ] \| ^2 \). Therefore, \( a(K_n,e) = P(W_n > \eta_{1-\alpha}; H_0) = 1 - H_k(\eta_{1-\alpha}; \delta(e)) + h(n,e) \) where \( h(n,e) = a(K_n,e) - 1 + H_k(\eta_{1-\alpha}; \delta(e)) \rightarrow 0 \) as \( n \rightarrow \infty \) for any fixed \( e \).

Let \( b(e) = -H_k(\eta_{1-\alpha}; \delta(e)) \). Then, for \( e \) close to 0, we have

\[
\begin{align*}
\quad a(K_n,e) - a_0 = b(e) - b(0) + h(n,e) - h(n,0) \\
= e b'(0) + \frac{1}{2} e^2 b''(0) + \frac{1}{6} e^3 b''(0) + h(n,e) - h(n,0),
\end{align*}
\]

where \( 0 < e^* < e \). Note that under \( H_0 \)

\[
b'(0) = \mu_k \left\{ \frac{\partial \delta(e)}{\partial e} \bigg|_{e=0} \right\} = 2 \mu_k n \left\{ \frac{\partial T(K_n,e)}{\partial e} \right\}^T U_n^{-1} \{ A_n \tilde{\eta}_{0,0} - g_0 \} = 0.
\]

From Lemma A8, under \( H_0 \)

\[
\frac{\partial T(K_n,e)}{\partial e} \bigg|_{e=0} = 1 / \sqrt{n} E_f \{ IF(Z_n; T, K_n,0) \}.
\]

Thus,

\[
b''(0) = \mu_k \left\{ \frac{\partial^2 \delta(e)}{\partial e^2} \bigg|_{e=0} \right\} = 2 \mu_k || U_n^{-1/2} A_n E_f \{ IF(Z_n; T, K_n,0) \} ||^2.
\]

Since from Lemma A8, \( IF(z_n; T, K_n,0) = -H_n^{-1} E_f \{ \Psi_0(z_n; \tilde{\eta}_{0,0}) \} \) is uniformly bounded, we have

\[
D = \limsup_{n \to \infty} || U_n^{-1/2} A_n E_f \{ IF(Z_n; T, K_n,0) \} ||^2 < \infty.
\]
From Equation (A17)
\[
\lim_{n \to \infty} \alpha(n, \delta) = \alpha_0 + \epsilon^2 \mu_k D + o(\epsilon^2),
\]

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

**Proof of Corollary 1.** For Part (i), following the proof of Theorem 1, for any fixed \(z\),
\[
\lim_{n \to \infty} \alpha(n, \epsilon) = \alpha_0 + \epsilon^2 \mu_k ||U^{-1/2} A \psi(z; T, K_0)||^2 + d(z, \epsilon),
\]
where \(d(z, \epsilon) = o(\epsilon^2)\). From the assumption that \(\sup_{x \in [0, C]} ||w(x)|| \leq C\) and \(\sup_{\mu \in \mathbb{R}} |\eta''(\mu)| \sqrt{V(\mu)/F(\mu)}| \leq C\), 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 \(J = \Delta z_n\). □

**Proof of Theorem 2.** Lemma A7 implies that
\[
\sqrt{n} \left\{ \{U(\mathbb{K}_n)\}^{-1/2} \{A_n T(\mathbb{K}_n) - g_0\} - \{U(\mathbb{K}_n)\}^{-1/2} \{A_n \hat{\beta}_{n,0} - g_0\} \right\} \xrightarrow{F} (0, I_k).
\]

From Lemmas A5 and A6,
\[
\sqrt{n} \left\{ \{U(\mathbb{K}_n)\}^{-1/2} \{A_n T(\mathbb{K}_n) - g_0\} - U_n^{-1/2} \{A_n T(K_{n,\epsilon}) - g_0\} \right\} \xrightarrow{F} (0, I_k).
\]

Then, \(W_n\) is asymptotically \(\chi^2_k(\delta(\epsilon))\) with \(\delta(\epsilon) = n ||U_n^{-1/2} \{A_n T(K_{n,\epsilon}) - g_0\}||^2\) under \(H_{1n}\). Therefore, \(\beta(K_{n,\epsilon}) = P(W_n > \eta_{1-\epsilon}, \epsilon \mid H_{1n}) = 1 - H_k(\eta_{1-\epsilon}, \epsilon \mid h(n, \epsilon))\), where \(h(n, \epsilon) = \beta(K_{n,\epsilon}) - 1 + H_k(\eta_{1-\epsilon}, \epsilon \mid h(n, \epsilon)) \to 0\) as \(n \to \infty\) for any fixed \(\epsilon\). Let \(b(\epsilon) = -H_k(\eta_{1-\epsilon}, \epsilon \mid h(n, \epsilon))\). Then, for \(\epsilon\) close to 0, we have
\[
\beta(K_{n,\epsilon}) - \beta_0 = b(\epsilon) - b(0) + h(n, \epsilon) - h(n, 0) = \epsilon b'(0) + \frac{\epsilon^2 b''(\epsilon^*)}{2} + h(n, \epsilon) - h(n, 0),
\]
where \(0 < \epsilon^* < \epsilon\). Note that under \(H_{1n}\), \(\delta(0) = n ||U_n^{-1/2} \{A_n \hat{\beta}_{n,0} - g_0\}||^2 = c^T U_n^{-1} c\). Then,
\[
b'(0) = \frac{-\partial H_k(\eta_{1-\epsilon}, \epsilon \mid h(n, \epsilon))}{\partial \epsilon} \bigg|_{\epsilon=0} = 2\nu_k \{A_n \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \}_{\epsilon=0} U_n^{-1} \{A_n \hat{\beta}_{n,0} - g_0\}
\]
\[
= 2\nu_k \sqrt{n} \left\{ A_n \frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \right\}_{\epsilon=0} U_n^{-1} c.
\]

From Lemma A8,
\[
\frac{\partial T(K_{n,\epsilon})}{\partial \epsilon} \bigg|_{\epsilon=0} = 1/\sqrt{n} E_f(\{\psi(Z_n; T, K_{n,\epsilon})\}),
\]
and hence,
\[
b'(0) = 2\nu_k c^T U_n^{-1} A_n E_f(\{\psi(Z_n; T, K_{n,\epsilon})\}).
\]

Since \(\sup_{x \in [0, C]} \limsup_{n \to \infty} |b''(\epsilon)| \leq C\) under \(H_{1n}\) by Lemma A9, we have \(\liminf_{n \to \infty} 1/2c^T b''(\epsilon^*) = o(\epsilon)\) as \(\epsilon \to 0\).

Since from Lemma A8, \(IF(z_n; T, K_{n,0}) = -H_k^{-1} E_f(\psi_{RBD}(z_n; \hat{\beta}_{n,0}))\) is uniformly bounded,
\[
|B| = |\liminf_{n \to \infty} 2c^T U_n^{-1} A_n E_f(\{IF(Z_n; T, K_{n,0})\})| < \infty.
\]
From Equation (A18), we complete the proof. □

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

**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 \( A_n A_n^{-1} \), i.e. \( A_n A_n^{-1} \mathbb{G} \)
- \( H_n : H_n = E_{K_{n0}} \left\{ p_2(Y; \tilde{X}_n \tilde{\beta}_{n0}) w(X_n) \tilde{X}_n \tilde{X}_n^T \right\} \)
- \( IF(\cdot; \cdot) \): influence function
- \( J \): an arbitrary distribution in the contamination of Equation (10)
- \( K_{n0} \): true parametric distribution of \( Z_n \)
- \( K_{n \epsilon} : K_{n \epsilon} = (1 - \frac{\epsilon}{\sqrt{n}}) K_{n0} + \frac{\epsilon}{\sqrt{n}} I \), \( \epsilon \)-contamination in Equation (10)
- \( \mathbb{K}_n \): empirical distribution of \( \{ Z_n \}_{i=1}^n \)
- \( \ell_K(\cdot) \): 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 \( \beta \)
- \( p_i(\cdot; \cdot) \): \( i \)th order derivative of robust-BD
- \( q(\cdot) \): generating \( q \)-function of BD
- \( T(\cdot) \): vector, a functional of estimator in Equation (12)
- \( U_n : U_n = A_n H_n^{-1} \Omega_n H_n^{-1} A_n^T \)
- \( V(\cdot) \): 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
- \( Z_n = (X_n^T, Y)^T \)
- \( a(\cdot) \): level of the test
- \( \beta(\cdot) \): power of the test
- \( \beta_{n0} \): true regression parameter
- \( \Delta_{z_n} \): probability measure which puts mass 1 at the point \( z_n \)
- \( \epsilon \): amount of contamination in Equation (10), positive constant
- \( \psi_{\text{RBD}} (\cdot; \cdot) \): score vector in Equation (7)
- \( \Omega_n : \Omega_n = E_{K_{n0}} \left\{ p_2^2(Y; \tilde{X}_n \tilde{\beta}_{n0}) w^2(X_n) \tilde{X}_n \tilde{X}_n^T \right\} \)
- \( p_q (\cdot; \cdot) \): robust-BD in Equation (4)

**References**


© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).