Robust-BD Estimation and Inference for Varying-Dimensional General Linear Models

Chunming Zhang Xiao Guo Chen Cheng Zhengjun Zhang

University of Wisconsin-Madison

Supplementary Material

S1 Notation and Assumptions

For a matrix M, its eigenvalues, minimum eigenvalue, maximum eigenvalue and trace are labeled by $\lambda_j(M)$, $\lambda_{\min}(M)$, $\lambda_{\max}(M)$ and $\mathrm{tr}(M)$ respectively. Let $\|M\| = \sup_{\|\boldsymbol{x}_n\|=1} \|M\boldsymbol{x}_n\| = \{\lambda_{\max}(M^T M)\}^{1/2}$ be the matrix L_2 norm; let $\|M\|_F = \{\mathrm{tr}(M^T M)\}^{1/2}$ be the Frobenius norm. See Golub and Van Loan (1996) for details. Throughout the proof, C is used as a generic finite constant.

We first impose some regularity conditions, which are not the weakest possible but facilitate the technical derivations.

Condition A:

A0. $\sup_{n\ge 1} \|\widetilde{\boldsymbol{\beta}}_0\|_1 < \infty$.

A1. $\|\boldsymbol{X}_n\|_\infty = \max_{1\le j\le p_n} |X_j|$ is bounded almost surely.

A2. $E(\widetilde{\boldsymbol{X}}_n \widetilde{\boldsymbol{X}}_n^T)$ exists and is nonsingular.

A4. There is a large enough open subset of $\mathbb{R}^{p_n+1}$ which contains the true parameter point $\widetilde{\boldsymbol{\beta}}_0$, such that $F^{-1}(\widetilde{\boldsymbol{X}}_n^T \widetilde{\boldsymbol{\beta}})$ is bounded almost surely for all $\widetilde{\boldsymbol{\beta}}$ in the subset.

A5. $w(\cdot) \ge 0$ 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; $V(\cdot) > 0$, $V^{(2)}(\cdot)$ is continuous. The matrix $\mathbf{H}_n$ is positive definite, with eigenvalues uniformly bounded away from 0.

A6. $q^{(4)}(\cdot)$ is continuous, and $q^{(2)}(\cdot) < 0$. $G_1^{(3)}(\cdot)$ is continuous.

A7. $F(\cdot)$ is monotone and a bijection, $F^{(3)}(\cdot)$ is continuous, and $F^{(1)}(\cdot) \ne 0$.

Condition B:

B5. The matrices Ω_n and $\mathbf{H}_n$ are positive definite, with eigenvalues uniformly bounded away from 0. Also, $\|\mathbf{H}_n^{-1}\Omega_n\|$ is bounded away from ∞.

Condition C:

C4. There is a large enough open subset of $\mathbb{R}^{p_n+1}$ which contains the true parameter point $\widetilde{\boldsymbol{\beta}}_0$, such that $A_n\widetilde{\boldsymbol{\beta}}_0 = \boldsymbol{g}_0$, and $F^{-1}(\widetilde{\boldsymbol{X}}_n^T\widetilde{\boldsymbol{\beta}})$ is bounded almost surely for all $\widetilde{\boldsymbol{\beta}}$ in the subset.

Condition D:

D5. The eigenvalues of $\mathbf{H}_n$ are uniformly bounded away from 0. Also, $\|\mathbf{H}_n^{-1/2}\Omega_n^{1/2}\|$ is bounded away from ∞.

S2 Proofs of Main Results

Proof of Theorem 1

We follow the idea of the proof in Fan and Peng (2004). Let $r_n = \sqrt{p_n/n}$ and $\widetilde{\mathbf{u}}_n = (u_0, u_1, \ldots, u_{p_n})^T \in \mathbb{R}^{p_n+1}$. It suffices to show that for any given $\epsilon > 0$, there exists a sufficiently large constant C_ϵ such that, for large n we have

$$P\Big\{\inf_{\|\widetilde{\mathbf{u}}_n\|=C_\epsilon} \ell_n(\widetilde{\boldsymbol{\beta}}_0 + r_n\widetilde{\mathbf{u}}_n) > \ell_n(\widetilde{\boldsymbol{\beta}}_0)\Big\} \geq 1-\epsilon. \tag{S2.1}$$

This implies that with probability at least $1-\epsilon$, there exists a local minimizer $\widehat{\widetilde{\boldsymbol{\beta}}}$ of $\ell_n(\widetilde{\boldsymbol{\beta}})$ in the ball $\{\widetilde{\boldsymbol{\beta}}_0 + r_n\widetilde{\mathbf{u}}_n : \|\widetilde{\mathbf{u}}_n\| \leq C_\epsilon\}$ such that $\|\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_0\| = O_P(r_n)$. To show (S2.1), consider

$$\begin{aligned}\ell_n(\widetilde{\boldsymbol{\beta}}_0 + r_n\widetilde{\mathbf{u}}_n) - \ell_n(\widetilde{\boldsymbol{\beta}}_0) &= \frac{1}{n}\sum_{i=1}^n\{\rho_q(Y_i, F^{-1}(\widetilde{\boldsymbol{X}}_{ni}^T(\widetilde{\boldsymbol{\beta}}_0 + r_n\widetilde{\mathbf{u}}_n)))w(\boldsymbol{X}_{ni}) \\ &\quad -\rho_q(Y_i, F^{-1}(\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_0))w(\boldsymbol{X}_{ni})\} \\ &\equiv I_1,\end{aligned} \tag{S2.2}$$

where $\|\widetilde{\mathbf{u}}_n\| = C_\epsilon$.

By Taylor's expansion,

$$I_1 = I_{1,1} + I_{1,2} + I_{1,3}, \tag{S2.3}$$

where

$$\begin{aligned} I_{1,1} &= r_n/n\sum_{i=1}^n \mathrm{p}_1(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_0)w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\mathbf{u}}_n, \\ I_{1,2} &= r_n^2/(2n)\sum_{i=1}^n \mathrm{p}_2(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})w(\boldsymbol{X}_{ni})(\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\mathbf{u}}_n)^2,\end{aligned}$$

$$I_{1,3} = r_n^3/(6n)\sum_{i=1}^{n} \mathrm{p}_3(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_n^*) w(\boldsymbol{X}_{ni})(\widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\mathbf{u}}_n)^3$$

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

$$|I_{1,1}| \le r_n \left\| \frac{1}{n}\sum_{i=1}^{n} \mathrm{p}_1(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_{n;0}) w(\boldsymbol{X}_{ni}) \widetilde{\boldsymbol{X}}_{ni} \right\| \|\widetilde{\mathbf{u}}_n\| = O_P(r_n \sqrt{p_n/n}) \|\widetilde{\mathbf{u}}_n\|. \qquad \text{(S2.4)}$$

For $I_{1,2}$ in (S2.3),

$$\begin{aligned} I_{1,2} &= \frac{r_n^2}{2n}\sum_{i=1}^{n} E\{\mathrm{p}_2(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_{n;0}) w(\boldsymbol{X}_{ni})(\widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\mathbf{u}}_n)^2\} \\ &\quad + \frac{r_n^2}{2n}\sum_{i=1}^{n} [\mathrm{p}_2(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_{n;0}) w(\boldsymbol{X}_{ni})(\widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\mathbf{u}}_n)^2 - E\{\mathrm{p}_2(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_{n;0}) w(\boldsymbol{X}_{ni})(\widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\mathbf{u}}_n)^2\}] \\ &\equiv I_{1,2,1} + I_{1,2,2}, \end{aligned}$$

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

$$\begin{aligned} |I_{1,2,2}| &\le r_n^2 \left\| \frac{1}{n}\sum_{i=1}^{n} \Big[\mathrm{p}_2(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_{n;0}) w(\boldsymbol{X}_{ni}) \widetilde{\boldsymbol{X}}_{ni} \widetilde{\boldsymbol{X}}_{ni}^T - E\{\mathrm{p}_2(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_{n;0}) w(\boldsymbol{X}_{ni}) \widetilde{\boldsymbol{X}}_{ni} \widetilde{\boldsymbol{X}}_{ni}^T\}\Big] \right\|_F \|\widetilde{\mathbf{u}}_n\|^2 \\ &= r_n^2 O_P(p_n/\sqrt{n}) \|\widetilde{\mathbf{u}}_n\|^2. \end{aligned}$$

Thus,

$$I_{1,2} = \frac{r_n^2}{2} \widetilde{\mathbf{u}}_n^T \mathbf{H}_n \widetilde{\mathbf{u}}_n + O_P(r_n^2 p_n/\sqrt{n}) \|\widetilde{\mathbf{u}}_n\|^2. \qquad \text{(S2.5)}$$

For $I_{1,3}$ in (S2.3), we observe that

$$|I_{1,3}| \le r_n^3 \frac{1}{n}\sum_{i=1}^{n} |\mathrm{p}_3(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_n^*)| w(\boldsymbol{X}_{ni}) |\widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\mathbf{u}}_n|^3 = O_P(r_n^3 p_n^{3/2}) \|\widetilde{\mathbf{u}}_n\|^3,$$

which follows from Conditions A0, A1, A4 and A5.

By (S2.4) and $p_n^4/n \to 0$, we can choose some large C_ϵ such that $I_{1,1}$ and $I_{1,3}$ are all dominated by the first term of $I_{1,2}$ in (S2.5), which is positive by the eigenvalue assumption on $\mathbf{H}_n$. This implies (S2.1). ■

Proof of Theorem 2

Notice the estimating equations $\frac{\partial \ell_n(\widetilde{\boldsymbol{\beta}})}{\partial \widetilde{\boldsymbol{\beta}}}|_{\widetilde{\boldsymbol{\beta}} = \widehat{\widetilde{\boldsymbol{\beta}}}} = \mathbf{0}$, since $\widehat{\widetilde{\boldsymbol{\beta}}}$ is a local minimizer of $\ell_n(\widetilde{\boldsymbol{\beta}})$. Taylor's expansion applied to the left side of the estimation equations yields

$$\mathbf{0} = \left\{ \frac{1}{n}\sum_{i=1}^{n} \mathrm{p}_1(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_{n;0}) w(\boldsymbol{X}_{ni}) \widetilde{\boldsymbol{X}}_{ni} \right\}$$

$$
\begin{aligned}
&+\left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{p}_2(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\widetilde{\boldsymbol{X}}_{ni}^T\right\}(\widehat{\widetilde{\boldsymbol{\beta}}}-\widetilde{\boldsymbol{\beta}}_{n;0})\\
&+\frac{1}{2n}\sum_{i=1}^{n}\mathrm{p}_3(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n}^*)w(\boldsymbol{X}_{ni})\{\widetilde{\boldsymbol{X}}_{ni}^T(\widehat{\widetilde{\boldsymbol{\beta}}}-\widetilde{\boldsymbol{\beta}}_{n;0})\}^2\widetilde{\boldsymbol{X}}_{ni}\\
\equiv&\left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{p}_1(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\right\}+K_2(\widehat{\widetilde{\boldsymbol{\beta}}}-\widetilde{\boldsymbol{\beta}}_{n;0})+K_3, \qquad \text{(S2.6)}
\end{aligned}
$$

where $\widetilde{\boldsymbol{\beta}}_n^*$ lies between $\widetilde{\boldsymbol{\beta}}_{n;0}$ and $\widehat{\widetilde{\boldsymbol{\beta}}}$. Below, we will show

$$\|K_2-\mathbf{H}_n\| = O_P(p_n/\sqrt{n}), \qquad \text{(S2.7)}$$
$$\|K_3\| = O_P(p_n^{5/2}/n). \qquad \text{(S2.8)}$$

First, to show (S2.7), note that $K_2-\mathbf{H}_n=K_2-E(K_2)\equiv L_1$. Similar arguments for the proof of $I_{1,2,2}$ in Theorem 1 give $\|L_1\|=O_P(p_n/\sqrt{n})$.

Second, a similar proof used for $I_{1,3}$ in (S2.3) completes (S2.8).

Third, by (S2.6)–(S2.8) and $\|\widehat{\widetilde{\boldsymbol{\beta}}}-\widetilde{\boldsymbol{\beta}}_{n;0}\|=O_P(\sqrt{p_n/n})$, we see that

$$\mathbf{H}_n(\widehat{\widetilde{\boldsymbol{\beta}}}-\widetilde{\boldsymbol{\beta}}_{n;0})=-\frac{1}{n}\sum_{i=1}^{n}\mathrm{p}_1(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}+\mathbf{u}_n, \qquad \text{(S2.9)}$$

where $\|\mathbf{u}_n\|=O_P(p_n^{5/2}/n)$. Note that by Condition B5,

$$
\begin{aligned}
&\|\sqrt{n}A_n\Omega_n^{-1/2}\mathbf{u}_n\|\le\sqrt{n}\|A_n\|_F\lambda_{\max}(\Omega_n^{-1/2})\|\mathbf{u}_n\|\\
&\quad=\sqrt{n}\{\mathrm{tr}(A_nA_n^T)\}^{1/2}/\lambda_{\min}^{1/2}(\Omega_n)\|\mathbf{u}_n\|=O_P(p_n^{5/2}/\sqrt{n})=o_P(1).
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\sqrt{n}A_n\Omega_n^{-1/2}\{\mathbf{H}_n(\widehat{\widetilde{\boldsymbol{\beta}}}-\widetilde{\boldsymbol{\beta}}_{n;0})\}\\
&\quad=-\frac{1}{\sqrt{n}}A_n\Omega_n^{-1/2}\sum_{i=1}^{n}\mathrm{p}_1(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}+o_P(1).
\end{aligned}
$$

To complete proving Theorem 2, we apply the Lindeberg-Feller central limit theorem (van der Vaart, 1998) to $\sum_{i=1}^n\boldsymbol{Z}_{ni}$, where $\boldsymbol{Z}_{ni}=-n^{-1/2}A_n\Omega_n^{-1/2}\mathrm{p}_1(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}$. It suffices to check (I) $\sum_{i=1}^n\mathrm{cov}(\boldsymbol{Z}_{ni})\to\mathbb{G}$; (II) $\sum_{i=1}^n E(\|\boldsymbol{Z}_{ni}\|^{2+\delta})=o(1)$ for some $\delta>0$. Condition (I) follows from the fact that $\mathrm{var}\{\mathrm{p}_1(Y;\widetilde{\boldsymbol{X}}_n^T\widetilde{\boldsymbol{\beta}}_{n;0})w(\boldsymbol{X}_n)\widetilde{\boldsymbol{X}}_n\}=\Omega_n$. To verify condition (II), notice that using Conditions B5 and A5,

$$
\begin{aligned}
E(\|\boldsymbol{Z}_{ni}\|^{2+\delta}) \le\ & n^{-(2+\delta)/2}E\Big\{\|A_n\|_F^{2+\delta}\Big[\|\Omega_n^{-1/2}\widetilde{\boldsymbol{X}}_n\|\\
&\Big|\{\psi(r(Y,m(\boldsymbol{X}_n)))-G_1'(m(\boldsymbol{X}_n))\}\frac{\{q''(m(\boldsymbol{X}_n))\sqrt{V(m(\boldsymbol{X}_n))}\}}{F'(m(\boldsymbol{X}_n))}w(\boldsymbol{X}_n)\Big|\Big]^{2+\delta}\Big\}\\
\le\ & Cn^{-(2+\delta)/2}E[\{\lambda_{\min}^{-1/2}(\Omega_n)\|\widetilde{\boldsymbol{X}}_n\|\}^{2+\delta}|\{\psi(r(Y,m(\boldsymbol{X}_n)))-G_1'(m(\boldsymbol{X}_n))\}\times
\end{aligned}
$$

$$
\begin{aligned}
& \{q''(m(\boldsymbol{X}_n))\sqrt{V(m(\boldsymbol{X}_n))}\}/F'(m(\boldsymbol{X}_n))|^{2+\delta}] \\
\leq\ & C p_n^{(2+\delta)/2} n^{-(2+\delta)/2} E[|\{\psi(r(Y, m(\boldsymbol{X}_n))) - G_1'(m(\boldsymbol{X}_n))\}\times \\
& \{q''(m(\boldsymbol{X}_n))\sqrt{V(m(\boldsymbol{X}_n))}\}/F'(m(\boldsymbol{X}_n))|^{2+\delta}] \\
\leq\ & O((p_n/n)^{(2+\delta)/2}).
\end{aligned}
$$

Thus, we get $\sum_{i=1}^n E(\|\boldsymbol{Z}_{ni}\|^{2+\delta}) \leq O(n(p_n/n)^{(2+\delta)/2}) = O(p_n^{(2+\delta)/2}/n^{\delta/2})$, which is $o(1)$. This verifies Condition (II). ■

Proposition 1 (covariance matrix estimation) *Assume* A0, A1, A2, A4, A5, B5, A6, *and* A7 *in the Appendix. Let* $V_n = \mathbf{H}_n^{-1}\Omega_n\mathbf{H}_n^{-1}$ *and* $\widehat{V}_n = \widehat{\mathbf{H}}_n^{-1}\widehat{\Omega}_n\widehat{\mathbf{H}}_n^{-1}$. *If* $p_n^4/n \to 0$ *as* $n \to \infty$, *then for any* $\sqrt{n/p_n}$*-consistent estimator* $\widehat{\widetilde{\boldsymbol{\beta}}}$ *of* $\widetilde{\boldsymbol{\beta}}_{n;0}$, *we have* $A_n(\widehat{V}_n - V_n)A_n^T \xrightarrow{P} \mathbf{0}$ *for any* $\mathsf{k} \times (p_n+1)$ *matrix* A_n *satisfying* $A_nA_n^T \to \mathbb{G}$, *where* $\mathbb{G}$ *is a* $\mathsf{k}\times\mathsf{k}$ *matrix and* k *is any fixed integer.*

Proof: Note $\|A_n(\widehat{V}_n - V_n)A_n^T\| \leq \|\widehat{V}_n - V_n\|\|A_n\|_F^2$. Since $\|A_n\|_F^2 \to \mathrm{tr}(\mathbb{G})$, it suffices to prove that $\|\widehat{V}_n - V_n\| = o_P(1)$.

First, we prove $\|\widehat{\mathbf{H}}_n - \mathbf{H}_n\| = o_P(1)$. Note that

$$
\begin{aligned}
\widehat{\mathbf{H}}_n - \mathbf{H}_n &= \frac{1}{n}\sum_{i=1}^n \{\mathrm{p}_2(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T\widehat{\widetilde{\boldsymbol{\beta}}}) - \mathrm{p}_2(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})\} w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\widetilde{\boldsymbol{X}}_{ni}^T \\
&\quad + \left\{\frac{1}{n}\sum_{i=1}^n \mathrm{p}_2(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0}) w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\widetilde{\boldsymbol{X}}_{ni}^T - \mathbf{H}_n\right\} \\
&\equiv I_1 + I_2.
\end{aligned}
$$

From the proof of (S2.7) in Theorem 2, we know that $\|I_2\| = O_P(p_n/\sqrt{n}) = o_P(1)$. We only need to consider the term I_1. Let $\widehat{m}_i = \widehat{m}(\boldsymbol{X}_{ni})$, $m_i = m(\boldsymbol{X}_{ni})$, $\widehat{r}_i = r(Y_i, \widehat{m}_i)$ and $r_i = r(Y_i, m_i)$. Then

$$
\begin{aligned}
I_1 &= \frac{1}{n}\sum_{i=1}^n [A_0(Y_i, \widehat{m}_i) + \{\psi(\widehat{r}_i) - G_1'(\widehat{m}_i)\}A_1(\widehat{m}_i) \\
&\quad - A_0(Y_i, m_i) - \{\psi(r_i) - G_1'(m_i)\}A_1(m_i)] w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\widetilde{\boldsymbol{X}}_{ni}^T \\
&= -\frac{1}{n}\sum_{i=1}^n \{G_1'(\widehat{m}_i)A_1(\widehat{m}_i) - G_1'(m_i)A_1(m_i)\} w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\widetilde{\boldsymbol{X}}_{ni}^T \\
&\quad + \frac{1}{n}\sum_{i=1}^n \{A_0(Y_i, \widehat{m}_i) + \psi(\widehat{r}_i)A_1(\widehat{m}_i) - A_0(Y_i, m_i) - \psi(r_i)A_1(m_i)\} w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\widetilde{\boldsymbol{X}}_{ni}^T \\
&\equiv I_{1,1} + I_{1,2}.
\end{aligned}
$$

Let $g(\cdot) = G_1'(\cdot)A_1(\cdot)$. By the assumptions, $g(\cdot)$ is differentiable. Thus

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^n |g(\widehat{m}_i) - g(m_i)| &= \frac{1}{n}\sum_{i=1}^n |(g \circ F^{-1})'(\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_n^*)\boldsymbol{X}_{ni}^T(\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0})| \\
&= O_P(1)O_P(\sqrt{p_n})O_P(\sqrt{p_n/n}) = O_P(p_n/\sqrt{n}),
\end{aligned}
$$

where $\widetilde{\boldsymbol{\beta}}_n^*$ is between $\widehat{\widetilde{\boldsymbol{\beta}}}$ and $\widetilde{\boldsymbol{\beta}}_{n;0}$. Thus

$$\left\|\frac{1}{n}\sum_{i=1}^{n}|g(\widehat{m}(\boldsymbol{X}_{ni}))-g(m(\boldsymbol{X}_{ni}))|w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\widetilde{\boldsymbol{X}}_{ni}^T\right\|_F = O_P(p_n/\sqrt{n})O_P(p_n) = O_P(p_n^2/\sqrt{n}).$$

Similar arguments give $\|I_{1,1}\| = O_P(p_n^2/\sqrt{n})$ and $\|I_{1,2}\| = O_P(p_n^2/\sqrt{n})$. Thus $\|I_1\| = O_P(p_n^2/\sqrt{n}) = o_P(1)$.

Second, we show $\|\widehat{\Omega}_n - \Omega_n\| = o_P(1)$. It is easy to see that

$$\begin{aligned}\widehat{\Omega}_n - \Omega_n &= \frac{1}{n}\sum_{i=1}^{n}\{\mathrm{p}_1^2(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widehat{\widetilde{\boldsymbol{\beta}}}) - \mathrm{p}_1^2(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})\}w^2(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\widetilde{\boldsymbol{X}}_{ni}^T \\ &\quad + \left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{p}_1^2(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})w^2(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\widetilde{\boldsymbol{X}}_{ni}^T - \Omega_n\right\} \\ &= \Delta_{1,1} + \Delta_{1,2},\end{aligned}$$

where $\|\Delta_{1,1}\| = O_P(p_n^2/\sqrt{n})$ and $\|\Delta_{1,2}\| = O_P(p_n/\sqrt{n})$. We observe that $\|\widehat{\Omega}_n - \Omega_n\| = O_P(p_n^2/\sqrt{n}) = o_P(1)$.

Third, we show $\|\widehat{V}_n - V_n\| = o_P(1)$. Note $\widehat{V}_n - V_n = L_1 + L_2 + L_3$, where $L_1 = \widehat{\mathbf{H}}_n^{-1}(\widehat{\Omega}_n - \Omega_n)\widehat{\mathbf{H}}_n^{-1}$, $L_2 = \widehat{\mathbf{H}}_n^{-1}(\mathbf{H}_n - \widehat{\mathbf{H}}_n)\mathbf{H}_n^{-1}\Omega_n\widehat{\mathbf{H}}_n^{-1}$ and $L_3 = \mathbf{H}_n^{-1}\Omega_n\widehat{\mathbf{H}}_n^{-1}(\mathbf{H}_n - \widehat{\mathbf{H}}_n)\mathbf{H}_n^{-1}$. By Assumption B5, it is straightforward to verify that $\|\mathbf{H}_n^{-1}\| \le O(1)$, $\|\widehat{\mathbf{H}}_n^{-1}\| \le O_P(1)$ and $\|\mathbf{H}_n^{-1}\Omega_n\| \le O(1)$. Since $\|L_1\| \le \|\widehat{\mathbf{H}}_n^{-1}\|\|\widehat{\Omega}_n - \Omega_n\|\|\widehat{\mathbf{H}}_n^{-1}\|$, we conclude $\|L_1\| = o_P(1)$, and similarly $\|L_2\| = o_P(1)$ and $\|L_3\| = o_P(1)$. Hence $\widehat{V}_n - V_n = o_P(1)$. ■

Proof of Theorem 3

For the matrix A_n in (4.3), there exists a $(p_n + 1 - \mathsf{k}) \times (p_n + 1)$ matrix B_n satisfying $B_n B_n^T = \mathbf{I}_{p_n+1-\mathsf{k}}$ and $A_n B_n^T = \mathbf{0}$. Therefore, $A_n\widetilde{\boldsymbol{\beta}}_n = \boldsymbol{g}_0$ is equivalent to $\widetilde{\boldsymbol{\beta}}_n = B_n^T\boldsymbol{\gamma}_n + \boldsymbol{c}_0$, where $\boldsymbol{\gamma}_n$ is a $(p_n + 1 - \mathsf{k}) \times 1$ vector and $\boldsymbol{c}_0 = A_n^T\mathbb{G}^{-1}\boldsymbol{g}_0$. Thus under H_0 in (4.3), we have $\widetilde{\boldsymbol{\beta}}_{n;0} = B_n^T\boldsymbol{\gamma}_{n;0} + \boldsymbol{c}_0$. Then minimizing $\ell_n(\widetilde{\boldsymbol{\beta}}_n)$ subject to $A_n\widetilde{\boldsymbol{\beta}}_n = \boldsymbol{g}_0$ is equivalent to minimizing $\ell_n(B_n^T\boldsymbol{\gamma}_n + \boldsymbol{c}_0)$ with respect to $\boldsymbol{\gamma}_n$, and we denote by $\widehat{\boldsymbol{\gamma}}_n$ the minimizer. Note that under (4.4), $\widehat{\widetilde{\boldsymbol{\beta}}}$ is the unique minimizer of $\ell_n(\widetilde{\boldsymbol{\beta}}_n)$. Hence $\Lambda_n = 2n\{\ell_n(B_n^T\widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0) - \ell_n(\widehat{\widetilde{\boldsymbol{\beta}}})\}$. Before showing Theorem 3, we need Lemma 1.

Lemma 1 *Assume conditions of Theorem* 3*. Then under H_0 in (4.3), we have that* $B_n^T(\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_{n;0}) = -n^{-1}B_n^T(B_n\mathbf{H}_nB_n^T)^{-1}B_n\sum_{i=1}^n \mathrm{p}_1(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni} + o_P(n^{-1/2})$, *and* $2n\{\ell_n(B_n^T\widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0) - \ell_n(\widehat{\widetilde{\boldsymbol{\beta}}})\} = n(B_n^T\widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}})^T\mathbf{H}_n(B_n^T\widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}}) + o_P(1)$.

Proof: To obtain the first part, following the proof of (S2.9) in Theorem 2, we have a similar expression for $\widehat{\boldsymbol{\gamma}}_n$,

$$B_n \mathbf{H}_n B_n^T (\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_{n;0}) = -\frac{1}{n} B_n \sum_{i=1}^n \mathrm{p}_1(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_{n;0}) w(\boldsymbol{X}_{ni}) \widetilde{\boldsymbol{X}}_{ni} + \mathbf{w}_n,$$

with $\|\mathbf{w}_n\| = o_P(n^{-1/2})$. As a result,

$$B_n^T (\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_{n;0}) = -\frac{1}{n} B_n^T (B_n \mathbf{H}_n B_n^T)^{-1} B_n \sum_{i=1}^n \mathrm{p}_1(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_{n;0}) w(\boldsymbol{X}_{ni}) \widetilde{\boldsymbol{X}}_{ni} + B_n^T (B_n \mathbf{H}_n B_n^T)^{-1} \mathbf{w}_n.$$

We notice that

$$\|B_n^T (B_n \mathbf{H}_n B_n^T)^{-1} \mathbf{w}_n\| \le \|(B_n \mathbf{H}_n B_n^T)^{-1}\| \|\mathbf{w}_n\| \le \|\mathbf{w}_n\| / \lambda_{\min}(\mathbf{H}_n) = o_P(n^{-1/2}),$$

in which the fact $\lambda_{\min}(B_n \mathbf{H}_n B_n^T) \ge \lambda_{\min}(\mathbf{H}_n)$ is used.

The proof of the second part proceeds in three steps. In Step 1, we use the following Taylor expansion for $\ell_n(B_n^T \widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0) - \ell_n(\widehat{\widetilde{\boldsymbol{\beta}}})$,

$$\begin{aligned} \ell_n(B_n^T \widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0) - \ell_n(\widehat{\widetilde{\boldsymbol{\beta}}}) &= \frac{1}{2n} \sum_{i=1}^n \mathrm{p}_2(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widehat{\widetilde{\boldsymbol{\beta}}}) w(\boldsymbol{X}_{ni}) \{\widetilde{\boldsymbol{X}}_{ni}^T (B_n^T \widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}})\}^2 \\ &\quad + \frac{1}{6n} \sum_{i=1}^n \mathrm{p}_3(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_n^*) w(\boldsymbol{X}_{ni}) \{\widetilde{\boldsymbol{X}}_{ni}^T (B_n^T \widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}})\}^3 \\ &\equiv I_1 + I_2, \end{aligned}$$

where $\widetilde{\boldsymbol{\beta}}_n^*$ lies between $\widehat{\widetilde{\boldsymbol{\beta}}}$ and $B_n^T \widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0$.

In Step 2, we analyze the stochastic order of $B_n^T \widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}}$. For a matrix X whose column vectors are linearly independent, set $P_X = X(X^T X)^{-1} X^T$. Define $H_n = \mathbf{I}_{p_n+1} - P_{\mathbf{H}_n^{1/2} B_n^T} = P_{\mathbf{H}_n^{-1/2} A_n^T}$. Then $\mathbf{H}_n^{-1} - B_n^T (B_n \mathbf{H}_n B_n^T)^{-1} B_n = \mathbf{H}_n^{-1/2} H_n \mathbf{H}_n^{-1/2}$. By (S2.9) and the first part of Lemma 1, we see immediately that

$$\begin{aligned} B_n^T \widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}} &= B_n^T (\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_{n;0}) - (\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0}) \\ &= \mathbf{H}_n^{-1/2} H_n \mathbf{H}_n^{-1/2} \left\{ \frac{1}{n} \sum_{i=1}^n \mathrm{p}_{1,i} w(\boldsymbol{X}_{ni}) \widetilde{\boldsymbol{X}}_{ni} \right\} + o_P(n^{-1/2}), \end{aligned} \tag{S2.10}$$

where $\mathrm{p}_{1,i} = \mathrm{p}_1(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_{n;0})$. Note that $\|\mathbf{H}_n^{-1/2} H_n \mathbf{H}_n^{-1/2} \{n^{-1} \sum_{i=1}^n \mathrm{p}_{1,i} w(\boldsymbol{X}_{ni}) \widetilde{\boldsymbol{X}}_{ni}\}\| = O_P(1/\sqrt{n})$. This gives

$$\|B_n^T \widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}}\| = O_P(1/\sqrt{n}). \tag{S2.11}$$

In Step 3, we conclude from (S2.11) that $I_2 = O_P\{(p_n/n)^{3/2}\} = o_P(1/n)$. Then $2n\{\ell_n(B_n^T \widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0) - \ell_n(\widehat{\widetilde{\boldsymbol{\beta}}})\} = 2nI_1 + o_P(1)$. Similar to the proof of Proposition 1, it is straightforward to see that

$$2nI_1 = n(B_n^T \widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}})^T \left\{ \frac{1}{n} \sum_{i=1}^n \mathrm{p}_2(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widehat{\widetilde{\boldsymbol{\beta}}}) w(\boldsymbol{X}_{ni}) \widetilde{\boldsymbol{X}}_{ni} \widetilde{\boldsymbol{X}}_{ni}^T \right\} (B_n^T \widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}})$$

$$
\begin{aligned}
&= n(B_n^T\widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}})^T\left\{\frac{1}{n}\sum_{i=1}^n \mathrm{p}_2(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\widetilde{\boldsymbol{X}}_{ni}^T\right\}(B_n^T\widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}}) + o_P(1)\\
&= n(B_n^T\widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}})^T E\{\mathrm{p}_2(Y_n;\widetilde{\boldsymbol{X}}_n^T\widetilde{\boldsymbol{\beta}}_{n;0})w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_n\widetilde{\boldsymbol{X}}_n^T\}(B_n^T\widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}}) + o_P(1)\\
&= n(B_n^T\widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}})^T\mathbf{H}_n(B_n^T\widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0 - \widehat{\widetilde{\boldsymbol{\beta}}}) + o_P(1).
\end{aligned}
$$

Then the second part of Lemma 1 is proved. ■

We now show Theorem 3. For part (i), a direct use of Lemma 1 and (S2.10) leads to

$$
\begin{aligned}
&2n\{\ell_n(B_n^T\widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0) - \ell_n(\widehat{\widetilde{\boldsymbol{\beta}}})\}\\
&\quad = \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{p}_{1,i}w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\right\}^T \mathbf{H}_n^{-1/2}H_n\mathbf{H}_n^{-1/2}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{p}_{1,i}w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\right\} + o_P(1).
\end{aligned}
$$

Since H_n is idempotent of rank k, it can be written as $H_n = C_n^TC_n$, where C_n is a $\mathsf{k}\times(p_n+1)$ matrix satisfying $C_nC_n^T = \mathbf{I}_{\mathsf{k}}$. Then

$$
\begin{aligned}
&2n\{\ell_n(B_n^T\widehat{\boldsymbol{\gamma}}_n + \boldsymbol{c}_0) - \ell_n(\widehat{\widetilde{\boldsymbol{\beta}}})\}\\
&= \left\{\frac{1}{\sqrt{n}}C_n\mathbf{H}_n^{-1/2}\sum_{i=1}^n \mathrm{p}_{1,i}w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\right\}^T\left\{\frac{1}{\sqrt{n}}C_n\mathbf{H}_n^{-1/2}\sum_{i=1}^n \mathrm{p}_{1,i}w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni}\right\} + o_P(1).
\end{aligned}
$$

Now consider part (ii). If $\psi(r) = r$ and the q-function satisfies (4.5), then $\mathrm{p}_1(y;\theta) = \mathrm{q}_1(y;\theta)$, $\mathrm{p}_2(y;\theta) = \mathrm{q}_2(y;\theta)$ and $\mathbf{H}_n = \Omega_n/C$, where $\mathrm{q}_j(y;\theta) = \frac{\partial^j}{\partial\theta^j}Q_q(y, F^{-1}(\theta))$. In this case, similar arguments for Theorem 2 yield

$$
\frac{1}{\sqrt{n}}C_n\mathbf{H}_n^{-1/2}\sum_{i=1}^n \mathrm{q}_1(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni} \xrightarrow{\mathcal{L}} N(\mathbf{0}, C\mathbf{I}_{\mathsf{k}}),
$$

which completes the proof. ■

Proof of Theorem 4

Before showing Theorem 4, Lemma 2 is needed.

Lemma 2 *Assume conditions of Theorem* 4. *Then*

$$
\begin{aligned}
\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0} &= -\frac{1}{n}\mathbf{H}_n^{-1}\sum_{i=1}^n \mathrm{p}_1(Y_i;\widetilde{\boldsymbol{X}}_{ni}^T\widetilde{\boldsymbol{\beta}}_{n;0})w(\boldsymbol{X}_{ni})\widetilde{\boldsymbol{X}}_{ni} + o_P(n^{-1/2}),\\
&\sqrt{n}(A_n\widehat{\mathbf{H}}_n^{-1}\widehat{\Omega}_n\widehat{\mathbf{H}}_n^{-1}A_n^T)^{-1/2}A_n(\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_k).
\end{aligned}
$$

Proof: Following (S2.9) in the proof of Theorem 2, we observe that $\|\mathbf{u}_n\| = O_P(p_n^{5/2}/n) = o_P(n^{-1/2})$. Condition B5 completes the proof for the first part.

To show the second part, denote $U_n = A_n \mathbf{H}_n^{-1} \Omega_n \mathbf{H}_n^{-1} A_n^T$ and $\widehat{U}_n = A_n \widehat{\mathbf{H}}_n^{-1} \widehat{\Omega}_n \widehat{\mathbf{H}}_n^{-1} A_n^T$. Notice that the eigenvalues of $\mathbf{H}_n^{-1} \Omega_n \mathbf{H}_n^{-1}$ are uniformly bounded away from 0. So are the eigenvalues of U_n. From the first part, we see that

$$A_n(\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0}) = -\frac{1}{n} A_n \mathbf{H}_n^{-1} \sum_{i=1}^n \mathrm{p}_1(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_{n;0}) w(\boldsymbol{X}_{ni}) \widetilde{\boldsymbol{X}}_{ni} + o_P(n^{-1/2}).$$

It follows that

$$\sqrt{n} U_n^{-1/2} A_n(\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0}) = \sum_{i=1}^n \boldsymbol{Z}_{ni} + o_P(1),$$

where $\boldsymbol{Z}_{ni} = -n^{-1/2} U_n^{-1/2} A_n \mathbf{H}_n^{-1} \mathrm{p}_1(Y_i; \widetilde{\boldsymbol{X}}_{ni}^T \widetilde{\boldsymbol{\beta}}_{n;0}) w(\boldsymbol{X}_{ni}) \widetilde{\boldsymbol{X}}_{ni}$. To show $\sum_{i=1}^n \boldsymbol{Z}_{ni} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_{\mathsf{k}})$, similar to the proof for Theorem 2, we check (III) $\sum_{i=1}^n \mathrm{cov}(\boldsymbol{Z}_{ni}) \to \mathbf{I}_{\mathsf{k}}$; (IV) $\sum_{i=1}^n E(\|\boldsymbol{Z}_{ni}\|^{2+\delta}) = o(1)$ for some $\delta > 0$. Condition (III) is straightforward since $\sum_{i=1}^n \mathrm{cov}(\boldsymbol{Z}_{ni}) = U_n^{-1/2} U_n U_n^{-1/2} = \mathbf{I}_{\mathsf{k}}$. To check condition (IV), similar arguments used in the proof of Theorem 2 give that $E(\|\boldsymbol{Z}_{ni}\|^{2+\delta}) = O((p_n/n)^{(2+\delta)/2})$. This and the boundedness of ψ yield $\sum_{i=1}^n E(\|\boldsymbol{Z}_{ni}\|^{2+\delta}) \le O(p_n^{(2+\delta)/2}/n^{\delta/2}) = o(1)$. Hence

$$\sqrt{n} U_n^{-1/2} A_n(\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_{\mathsf{k}}). \tag{S2.12}$$

From the proof of Proposition 1, it can be concluded that $\|\widehat{U}_n - U_n\| = o_P(1)$ and that the eigenvalues of $\widehat{U}_n$ are uniformly bounded away from 0 and ∞ with probability tending to one. Consequently,

$$\|\widehat{U}_n^{-1/2} U_n^{1/2} - \mathbf{I}_{\mathsf{k}}\| = o_P(1). \tag{S2.13}$$

Combining (S2.12), (S2.13) and Slutsky's theorem completes the proof that $\sqrt{n}\widehat{U}_n^{-1/2} A_n(\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_{\mathsf{k}})$. ■

We now show Theorem 4, which follows directly from H_0 in (4.3) and the second part of Lemma 2. This completes the proof. ■

Proof of Theorem 5

Note that W_n can be decomposed into three additive terms,

$$\begin{aligned} I_1 &= n\{A_n(\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0})\}^T (A_n \widehat{V}_n A_n^T)^{-1} \{A_n(\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0})\}, \\ I_2 &= 2n(A_n \widetilde{\boldsymbol{\beta}}_{n;0} - \boldsymbol{g}_0)^T (A_n \widehat{V}_n A_n^T)^{-1} \{A_n(\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0})\}, \\ I_3 &= n(A_n \widetilde{\boldsymbol{\beta}}_{n;0} - \boldsymbol{g}_0)^T (A_n \widehat{V}_n A_n^T)^{-1} (A_n \widetilde{\boldsymbol{\beta}}_{n;0} - \boldsymbol{g}_0), \end{aligned}$$

where $\widehat{V}_n = \widehat{\mathbf{H}}_n^{-1} \widehat{\Omega}_n \widehat{\mathbf{H}}_n^{-1}$. We observe that $I_1 \xrightarrow{\mathcal{L}} \chi^2_{\mathsf{k}}$ following the second part of Lemma 2; $I_3 = n(A_n \widetilde{\boldsymbol{\beta}}_{n;0} - \boldsymbol{g}_0)^T \mathbf{M}^{-1} (A_n \widetilde{\boldsymbol{\beta}}_{n;0} - \boldsymbol{g}_0)\{1 + o_P(1)\}$ by Proposition 1; $I_2 = O_P(\sqrt{n})$ by Cauchy-Schwartz inequality. Thus

$$n^{-1} I_3 \ge \lambda_{\min}(\mathbf{M}^{-1}) \|A_n \widetilde{\boldsymbol{\beta}}_{n;0} - \boldsymbol{g}_0\|^2 \{1 + o_P(1)\} = \lambda_{\max}^{-1}(\mathbf{M}) \|A_n \widetilde{\boldsymbol{\beta}}_{n;0} - \boldsymbol{g}_0\|^2 + o_P(1).$$

These complete the proof for W_n. ■

Proof of Theorem 6

Following the second part of Lemma 2, we observe that $\sqrt{n}(A_n \widehat{V}_n A_n^T)^{-1/2}(A_n \widehat{\widetilde{\boldsymbol{\beta}}} - \boldsymbol{g}_0) \stackrel{\mathcal{L}}{\longrightarrow} N(\mathbf{M}^{-1/2}\boldsymbol{c}, \mathbf{I}_{\mathsf{k}})$, which completes the proof. ■

Proof of Theorem 7

We first need to show Lemma 3.

Lemma 3 *Suppose that* $(\boldsymbol{X}_n^o, Y^o)$ *follows the distribution of* $(\boldsymbol{X}_n, Y)$ *and is independent of the training set* $\mathcal{T}_n$*. If* $\mathtt{Q}$ *is a* BD*, then*

$$E\{\mathtt{Q}(Y^o, \widehat{m}(\boldsymbol{X}_n^o))\} = E\{\mathtt{Q}(Y^o, m(\boldsymbol{X}_n^o))\} + E\{\mathtt{Q}(m(\boldsymbol{X}_n^o), \widehat{m}(\boldsymbol{X}_n^o))\}.$$

Proof: Let $\mathtt{q}$ be the generating function of $\mathtt{Q}$. Then

$$\begin{aligned}\mathtt{Q}(Y^o, \widehat{m}(\boldsymbol{X}_n^o)) &= [\mathtt{q}(m(\boldsymbol{X}_n^o)) - E\{\mathtt{q}(Y^o) \mid \mathcal{T}_n, \boldsymbol{X}_n^o\}] + [E\{\mathtt{q}(Y^o) \mid \mathcal{T}_n, \boldsymbol{X}_n^o\} \\ &\quad - \mathtt{q}(Y^o)] - \mathtt{q}(m(\boldsymbol{X}_n^o)) + \mathtt{q}(\widehat{m}(\boldsymbol{X}_n^o)) + \{Y^o - \widehat{m}(\boldsymbol{X}_n^o)\}\mathtt{q}'(\widehat{m}(\boldsymbol{X}_n^o)). \quad \text{(S2.14)}\end{aligned}$$

Since $(\boldsymbol{X}_n^o, Y^o)$ is independent of $\mathcal{T}_n$, we deduce from Chow and Teicher (1989, Corollary 3, p. 223) that

$$E\{\mathtt{q}(Y^o) \mid \mathcal{T}_n, \boldsymbol{X}_n^o\} = E\{\mathtt{q}(Y^o) \mid \boldsymbol{X}_n^o\}. \quad \text{(S2.15)}$$

Similarly,

$$E\{Y^o \mathtt{q}'(\widehat{m}(\boldsymbol{X}_n^o)) \mid \mathcal{T}_n, \boldsymbol{X}_n^o\} = E(Y^o \mid \boldsymbol{X}_n^o)\mathtt{q}'(\widehat{m}(\boldsymbol{X}_n^o)) = m(\boldsymbol{X}_n^o)\mathtt{q}'(\widehat{m}(\boldsymbol{X}_n^o)). \quad \text{(S2.16)}$$

Applying (S2.15) and (S2.16) to (S2.14) results in

$$E\{\mathtt{Q}(Y^o, \widehat{m}(\boldsymbol{X}_n^o)) \mid \mathcal{T}_n, \boldsymbol{X}_n^o\} = E\{\mathtt{Q}(Y^o, m(\boldsymbol{X}_n^o)) \mid \boldsymbol{X}_n^o\} + \mathtt{Q}(m(\boldsymbol{X}_n^o), \widehat{m}(\boldsymbol{X}_n^o))$$

and thus the conclusion. ■

Now show Theorem 7. Setting $\mathtt{Q}$ in Lemma 3 to be the misclassification loss gives

$$\begin{aligned}1/2[E\{R(\widehat{\phi}_n)\} - R(\phi_{n,\mathrm{B}})] &\le E[|m(\boldsymbol{X}_n^o) - .5|\mathrm{I}\{m(\boldsymbol{X}_n^o) \le .5,\ \widehat{m}(\boldsymbol{X}_n^o) > .5\}] \\ &\quad + E[|m(\boldsymbol{X}_n^o) - .5|\mathrm{I}\{m(\boldsymbol{X}_n^o) > .5,\ \widehat{m}(\boldsymbol{X}_n^o) \le .5\}] \\ &= I_1 + I_2.\end{aligned}$$

For any $\epsilon > 0$, it follows that

$$\begin{aligned}I_1 &= E[|m(\boldsymbol{X}_n^o) - .5|\mathrm{I}\{m(\boldsymbol{X}_n^o) < .5 - \epsilon,\ \widehat{m}(\boldsymbol{X}_n^o) > .5\}] \\ &\quad + E[|m(\boldsymbol{X}_n^o) - .5|\mathrm{I}\{.5 - \epsilon \le m(\boldsymbol{X}_n^o) \le .5,\ \widehat{m}(\boldsymbol{X}_n^o) > .5\}] \\ &\le P\{|\widehat{m}(\boldsymbol{X}_n^o) - m(\boldsymbol{X}_n^o)| > \epsilon\} + \epsilon\end{aligned}$$

and similarly, $I_2 \le \epsilon + P\{|\widehat{m}(\boldsymbol{X}_n^o) - m(\boldsymbol{X}_n^o)| \ge \epsilon\}$. Recall that

$$|\widehat{m}(\boldsymbol{X}_n^o) - m(\boldsymbol{X}_n^o)| = |F^{-1}(\widetilde{\boldsymbol{X}_n^o}^T \widehat{\widetilde{\boldsymbol{\beta}}}) - F^{-1}(\widetilde{\boldsymbol{X}_n^o}^T \widetilde{\boldsymbol{\beta}}_{n;0})| \le |(F^{-1})'(\widetilde{\boldsymbol{X}_n^o}^T \widetilde{\boldsymbol{\beta}}_n^*)| \|\boldsymbol{X}_n^o\| \|\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0}\|,$$

for some $\widetilde{\boldsymbol{\beta}}_n^*$ between $\widetilde{\boldsymbol{\beta}}_{n;0}$ and $\widehat{\widetilde{\boldsymbol{\beta}}}$, where $\widetilde{\boldsymbol{X}_n^o} = (1, \boldsymbol{X}_n^{o\,T})^T$. By Condition A4, we conclude that $(F^{-1})'(\widetilde{\boldsymbol{X}_n^o}^T \widetilde{\boldsymbol{\beta}}_n^*) = O_P(1)$. This along with $\|\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}_{n;0}\| = O_P(1)$ and $\|\widetilde{\boldsymbol{X}_n^o}\| = O_P(\sqrt{p_n})$ implies that $|\widehat{m}(\boldsymbol{X}_n^o) - m(\boldsymbol{X}_n^o)| = O_P(r_n\sqrt{p_n}) = o_P(1)$. Therefore $I_1 \to 0$ and $I_2 \to 0$, which completes the proof. ■

S3 Figures 7–10 in Section 6.2

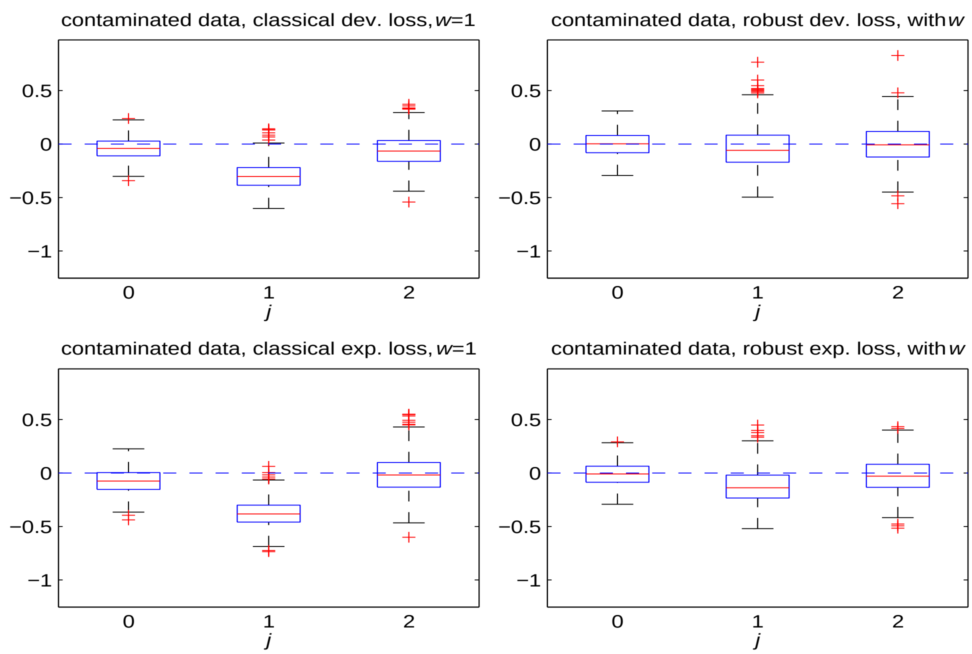

Figure 7: (Simulated Bernoulli response data with contamination) Boxplots of $\widehat{\beta}_j - \beta_{j;0}$, $j = 0, 1, \ldots, p_n$ (from left to right in each panel). Left panels: the non-robust estimates; right panels: the robust estimates.

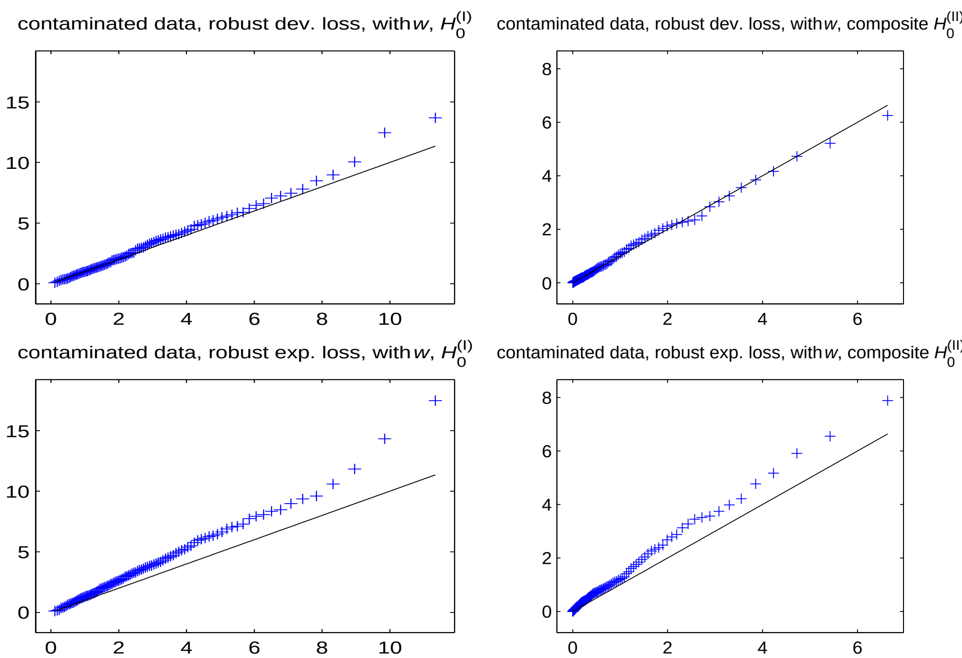

Figure 8: (Simulated Bernoulli response data with contamination) Empirical quantiles (on the y-axis) of test statistics W_n versus quantiles (on the x-axis) of the χ^2_k distribution. Solid line: the 45 degree reference line. Left panels: for testing $H_0^{(I)}$; right panels: for testing $H_0^{(II)}$.

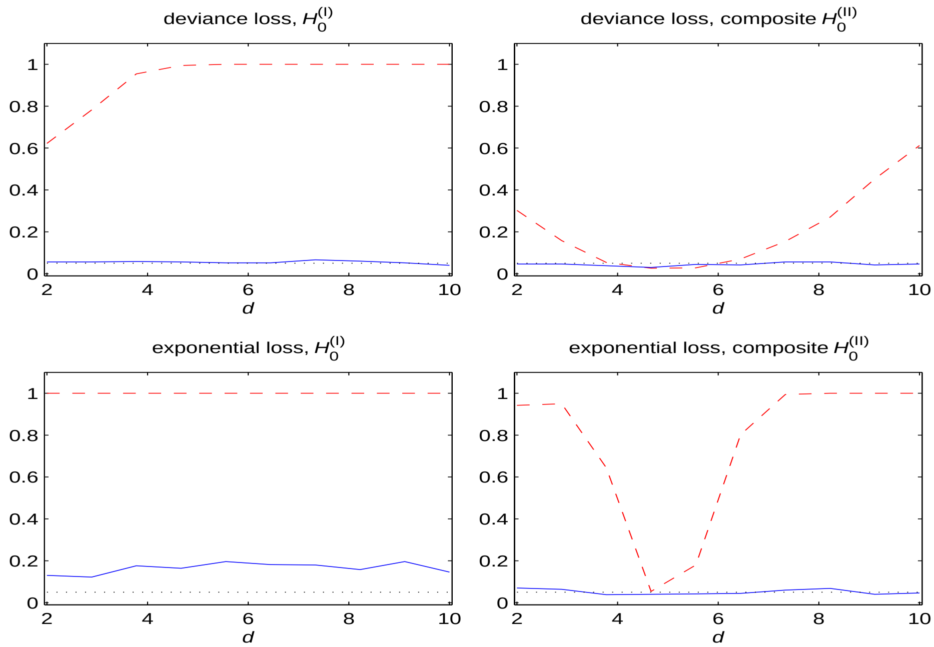

Figure 9: Level of tests for the Bernoulli response data. The dashed line corresponds to the non-robust Wald-type test; the solid line corresponds to the robust Wald-type test; the dotted line indicates the 5% nominal level. Left panels: for testing $H_0^{(\mathrm{I})}$; right panels: for testing $H_0^{(\mathrm{II})}$.

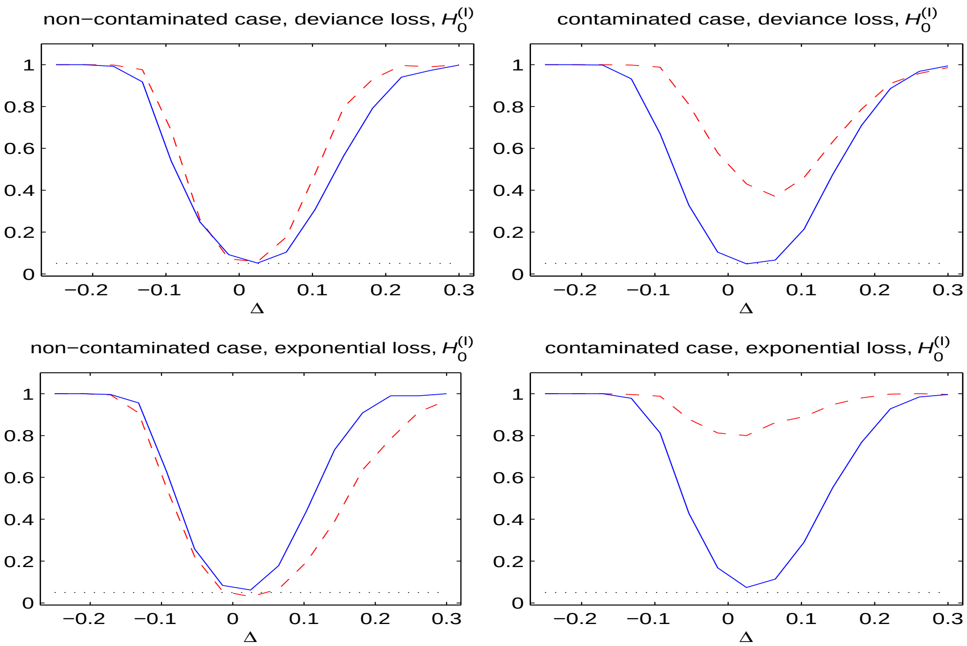

Figure 10: Observed power functions of tests for the Bernoulli response data. The dashed line corresponds to the non-robust Wald-type test; the solid line corresponds to the robust Wald-type test; the dotted line indicates the 5% nominal level. Left panels: non-contaminated case; right panels: contaminated case.