SUPPLEMENT TO "MODEL-FREE FORWARD SCREENING VIA CUMULATIVE DIVERGENCE"

Tingyou Zhou[a], Liping Zhu[b], Chen Xu[c] and Runze Li[d]

[a] School of Data Sciences, Zhejiang University of Finance and Economics, Hangzhou, P. R. China. [b] Institute of Statistics and Big Data and Center for Applied Statistics, Renmin University of China, Beijing, P. R. China. [c] Department of Mathematics and Statistics University of Ottawa, Ottawa, Canada. [d] Department of Statistics and The Methodology Center, The Pennsylvania State University at University Park, U.S.A.

Some Lemmas

Stein's Lemma: Let $X \sim \mathcal{N}(0,1)$, and $g : \mathbb{R} \to \mathbb{R}$ be an indefinite integral of the Lebesgue measurable function g', the derivative of g. Suppose $E(\mid g'(X) \mid) < \infty$, then $E\{g'(X)\} = E\{Xg(X)\}$. Interested readers can refer to Stein (1981) for details.

Lemma 1. *Assume $X \sim \mathcal{N}(0,1)$, A and B are constants and all the moments involved exist. Denote $F(y \mid X) = pr(Y < y \mid X)$. It follows that*

$$\begin{aligned} E\left[\exp\{-A(X-B)^2\}\right] &= (2A+1)^{-1/2}\exp\left\{-AB^2/(2A+1)\right\} \text{ and} \\ E\left\{\partial F(y \mid X)/\partial X\right\} &= E\left\{F(y \mid X)X\right\} = E\left\{\mathbf{1}(Y<y)X\right\}. \end{aligned}$$

Proof of Lemma 1: The first statement is straightforward and the second is a direct application of Stein (1981)'s lemma. □

Proofs of Statement (2.2), Lemma 2, Proposition 1 and Theorem 1

Proof of Statement (2.2): We show the first equivalence. The $\Rightarrow$ part is obvious by noting that $E(Y \mid X < x_0) = E\left\{Y\mathbf{1}(X<x_0)\right\}/E\left\{\mathbf{1}(X<x_0)\right\}$. Next we show the $\Leftarrow$ part. Without loss of generality we assume $E(Y) = 0$ because otherwise we let $\widetilde{Y} = Y - E(Y)$. We need to prove that $E\left\{Y\mathbf{1}(X<x_0)\right\} = 0$ implies that $E(Y \mid X) = 0$.

By definition,

$$E\left\{Y\mathbf{1}(X<x_0)\right\} = E\left\{E(Y \mid X)\mathbf{1}(X<x_0)\right\} = \int_{-\infty}^{x_0} E(Y \mid X=x)f(x)dx.$$

*Liping Zhu is the corresponding author. Email: zhu.liping@ruc.edu.cn. Institute of Statistics and Big Data and Center for Applied Statistics, Renmin University of China, 59 Zhongguancun Avenue, Haidian District, Beijing 100872, P. R. China.

Thinking that $E\{Y\mathbf{1}(X<x_0)\}=0$ implies that the first derivative of $\int_{-\infty}^{x_0} E(Y \mid X = x)f(x)dx$ with respect to x_0, which is $E(Y \mid X = x_0)f(x_0)$, is also 0. By definition, $x_0 \in \mathrm{supp}(X)$ and hence $f(x_0) > 0$, $E(Y \mid X = x_0)$ must be 0 for all $x_0 \in \mathrm{supp}(X)$. This completes the proof of the first equivalence.

The second equivalence is obvious by using the fact that $E(Y \mid X < x_0) = E\{Y\mathbf{1}(X<x_0)\}/E\{\mathbf{1}(X<x_0)\}$. The third equivalence is also obvious. This completes the proof of Statement (2.2). □

Proof of Lemma 2: Define $\mathbf{\Sigma}_{k|\mathcal{F}} = E\left[\left\{g'_{k|\mathcal{F}}(\mathbf{x}_{\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\right\}\left\{g'_{k|\mathcal{F}}(\mathbf{x}_{\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\right\}^{\mathrm{T}}\right]$, we first prove that

$$\begin{aligned}(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}}) &= n^{-1}\mathbf{\Sigma}_{k|\mathcal{F}}^{-1}\sum_{i=1}^{n}\{X_{ik}-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\}g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\\ &+ o_p(n^{-1/2}). \end{aligned} \tag{S.1}$$

Let $\mathbf{\Omega}_{n,k|\mathcal{F}}(\boldsymbol{\beta}_{k|\mathcal{F}}) = \sum_{i=1}^{n}\{X_{ik}-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{k|\mathcal{F}})\}g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{k|\mathcal{F}})$, then we have $\mathbf{\Omega}_{n,k|\mathcal{F}}(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}) = \mathbf{0}$ for $\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}$ defined in (3.3). Applying Taylor's expansion, we get

$$\begin{aligned}\mathbf{0} &= \mathbf{\Omega}_{n,k|\mathcal{F}}(\boldsymbol{\beta}_{0,k|\mathcal{F}}) + \mathbf{\Omega}'_{n,k|\mathcal{F}}(\boldsymbol{\beta}_{0,k|\mathcal{F}})(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}})\\ &\qquad +(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}})^{\mathrm{T}}\mathbf{\Omega}''_{n,k|\mathcal{F}}(\boldsymbol{\beta}^*_{k|\mathcal{F}})(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}})/2,\end{aligned}$$

where $\boldsymbol{\beta}^*_{k|\mathcal{F}}$ lies between $\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}$ and $\boldsymbol{\beta}_{0,k|\mathcal{F}}$. Consequently we have

$$\begin{aligned}\mathbf{\Sigma}_{k|\mathcal{F}}(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}}) &= n^{-1}\mathbf{\Omega}_{n,k|\mathcal{F}}(\boldsymbol{\beta}_{0,k|\mathcal{F}}) + \left\{\mathbf{\Sigma}_{k|\mathcal{F}} + n^{-1}\mathbf{\Omega}'_{n,k|\mathcal{F}}(\boldsymbol{\beta}_{0,k|\mathcal{F}})\right\}(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}})\\ &+ (2n)^{-1}(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}})^{\mathrm{T}}\mathbf{\Omega}''_{n,k|\mathcal{F}}(\boldsymbol{\beta}^*_{k|\mathcal{F}})(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}}).\end{aligned}$$

Invoking assumptions on $g_{k|\mathcal{F}}(\cdot)$, we have

$$n^{-1}\sum_{i=1}^{n}\left[g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\left\{g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\right\}^{\mathrm{T}} - \mathbf{\Sigma}_{k|\mathcal{F}}\right] = O_p(n^{-1/2}s), \text{ and}$$

$$n^{-1}\sum_{i=1}^{n}\{X_{ik}-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\}g''_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}}) = O_p(n^{-1/2}s),$$

where $s = |\mathcal{F}|$. This leads to $\mathbf{\Sigma}_{k|\mathcal{F}} + n^{-1}\mathbf{\Omega}'_{n,k|\mathcal{F}}(\boldsymbol{\beta}_{0,k|\mathcal{F}}) = O_p(n^{-1/2}s)$. Using similar arguments, we have $\mathbf{\Omega}''_{n,k|\mathcal{F}}(\boldsymbol{\beta}^*_{k|\mathcal{F}}) = O_p(ns^{3/2})$ and $||\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}}|| = O_p(n^{-1/2}s^{1/2})$, where $||\cdot||$ denotes the Euclidean norm. Then we obtain

$$\boldsymbol{\Sigma}_{k|\mathcal{F}}(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}}) \;=\; n^{-1}\boldsymbol{\Omega}_{n,k|\mathcal{F}}(\boldsymbol{\beta}_{0,k|\mathcal{F}})+O_p(n^{-1}s^{3/2})+O_p(n^{-1}s^{5/2}).$$

This proves (S.1) when $s=o(n^{1/5})$.

We next introduce some notations which will be frequently used. For vector $\boldsymbol{\beta}$, let $||\boldsymbol{\beta}||_\infty=\max_j|\beta_j|$. For $m\times n$ matrix $\mathbf{M}$, define $||\mathbf{M}||_\infty=\max_{i=1,\ldots,m}\sum_{j=1}^{n}\left|M_{ij}\right|$. Note that

$$\mathrm{pr}\left\{(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}})^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}})>\varepsilon_n\right\}<\mathrm{pr}\left\{\left(||\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}}||_\infty\right)^2>\varepsilon_n/s\right\},$$

and $\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}}=n^{-1}\boldsymbol{\Sigma}_{k|\mathcal{F}}^{-1}\sum_{i=1}^{n}g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\delta_{i,k|\mathcal{F}}+o_p(n^{-1/2})$, where $\delta_{i,k|\mathcal{F}}=X_{ik}-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})$ are independent. Thus,

$$\mathrm{pr}\left\{||\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}}||_\infty>(\varepsilon_n/s)^{1/2}\right\}=\mathrm{pr}\left\{\left|\left|n^{-1}\boldsymbol{\Sigma}_{k|\mathcal{F}}^{-1}\sum_{i=1}^{n}g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\delta_{i,k|\mathcal{F}}\right|\right|_\infty>(\varepsilon_n/s)^{1/2}\right\}$$

Recall that we assume the infinity norm of the precision matrix is bounded. That is, there must exists a constant c_0 such that $||\boldsymbol{\Sigma}_{\mathcal{F}}^{-1}||_\infty<c_0$. Thus, there exists a positive constant c_1 such that

$$\begin{aligned}
&\mathrm{pr}\left\{||\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}}||_\infty>(\varepsilon_n/s)^{1/2}\right\}\\
\le\;&\mathrm{pr}\left\{||\boldsymbol{\Sigma}_{k|\mathcal{F}}^{-1}||_\infty\left|\left|n^{-1}\sum_{i=1}^{n}g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\delta_{i,k|\mathcal{F}}\right|\right|_\infty>(\varepsilon_n/s)^{1/2}\right\}\\
\le\;&s\max_{l\in\mathcal{F}}\mathrm{pr}\left\{\left|n^{-1}\sum_{i=1}^{n}g'_{l,k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\delta_{i,k|\mathcal{F}}\right|>(\varepsilon_n/s)^{1/2}/c_0\right\}\le 2s\exp(-c_1ns^{-1}\varepsilon_n),
\end{aligned}$$

where $g'_{l,k|\mathcal{F}}(\cdot)$ is the l-th element of $g'_{k|\mathcal{F}}(\cdot)$, and the last inequality holds by Lemma 1. This completes the proof of Lemma 2. □

Proof of Proposition 1: By definition, $\widehat{\omega}_{k|\mathcal{F}}=\widehat{\mathrm{CCov}}\left\{(X_k-\mu_{k|\mathcal{F}})\mid Y\right\}/\widehat{\mathrm{var}}(X_k-\mu_{k|\mathcal{F}})$ and $\omega_{k|\mathcal{F}}=\mathrm{CCov}\left\{(X_k-\mu_{k|\mathcal{F}})\mid Y\right\}/\mathrm{var}(X_k-\mu_{k|\mathcal{F}})$, where $\mu_{k|\mathcal{F}}\overset{\mathrm{def}}{=}E(X_k\mid\mathbf{x}_{\mathcal{F}})$. We decompose $\widehat{\omega}_{k|\mathcal{F}}-\omega_{k|\mathcal{F}}$ into four parts. In particular, $\widehat{\omega}_{k|\mathcal{F}}-\omega_{k|\mathcal{F}}=I_1+I_2+I_3+I_4$, where

$$
\begin{aligned}
I_1 &\stackrel{\text{def}}{=} \left[\widehat{\mathrm{CCov}}\{(X_k-\mu_{k|\mathcal{F}}) \mid Y\} - \mathrm{CCov}\{(X_k-\mu_{k|\mathcal{F}}) \mid Y\}\right]\{\mathrm{var}(X_k-\mu_{k|\mathcal{F}})\}^{-1},\\
I_2 &\stackrel{\text{def}}{=} \omega_{k|\mathcal{F}}\{\mathrm{var}(X_k-\mu_{k|\mathcal{F}})\}^{-1}\left\{\mathrm{var}(X_k-\mu_{k|\mathcal{F}}) - \widehat{\mathrm{var}}(X_k-\mu_{k|\mathcal{F}})\right\},\\
I_3 &\stackrel{\text{def}}{=} \left[\widehat{\mathrm{CCov}}\{(X_k-\mu_{k|\mathcal{F}}) \mid Y\} - \mathrm{CCov}\{(X_k-\mu_{k|\mathcal{F}}) \mid Y\}\right]\\
&\quad \left[\{\widehat{\mathrm{var}}(X_k-\mu_{k|\mathcal{F}})\}^{-1} - \{\mathrm{var}(X_k-\mu_{k|\mathcal{F}})\}^{-1}\right],\\
I_4 &\stackrel{\text{def}}{=} \omega_{k|\mathcal{F}}\left\{\mathrm{var}(X_k-\mu_{k|\mathcal{F}}) - \widehat{\mathrm{var}}(X_k-\mu_{k|\mathcal{F}})\right\}\left[\{\widehat{\mathrm{var}}(X_k-\mu_{k|\mathcal{F}})\}^{-1} - \{\mathrm{var}(X_k-\mu_{k|\mathcal{F}})\}^{-1}\right].
\end{aligned}
$$

We study $\widehat{\mathrm{CCov}}\{(X_k-\mu_{k|\mathcal{F}}) \mid Y\} - \mathrm{CCov}\{(X_k-\mu_{k|\mathcal{F}}) \mid Y\}$ and $\widehat{\mathrm{var}}(X_k-\mu_{k|\mathcal{F}}) - \mathrm{var}(X_k-\mu_{k|\mathcal{F}})$ respectively.

We first deal with $\widehat{\mathrm{CCov}}\{(X_k-\mu_{k|\mathcal{F}}) \mid Y\} - \mathrm{CCov}\{(X_k-\mu_{k|\mathcal{F}}) \mid Y\}$. It can be written as $L_1+L_2+L_3$, where

$$
\begin{aligned}
L_1 &= n^{-1}\sum_{j=1}^n\left[n^{-1}\sum_{i=1}^n \mathbf{1}(Y_i<Y_j)\{X_{ik}-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\}\right]^2 - \mathrm{CCov}\left\{(X_k-\mu_{k|\mathcal{F}}) \mid Y\right\},\\
L_2 &= 2n^{-1}\sum_{j=1}^n\Bigg(\left[n^{-1}\sum_{i=1}^n \mathbf{1}(Y_i<Y_j)\{X_{ik}-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\}\right]\\
&\qquad \left[n^{-1}\sum_{i=1}^n \mathbf{1}(Y_i<Y_j)\{g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}})\}\right]\Bigg),\\
L_3 &= n^{-1}\sum_{j=1}^n\left[n^{-1}\sum_{i=1}^n \mathbf{1}(Y_i<Y_j)\{g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}})\}\right]^2.
\end{aligned}
$$

Thus, for any $\varepsilon_n>0$, $\mathrm{pr}\left[\left|\widehat{\mathrm{CCov}}\{(X_k-\mu_{k|\mathcal{F}}) \mid Y\} - \mathrm{CCov}\{(X_k-\mu_{k|\mathcal{F}}) \mid Y\}\right| > 3\varepsilon_n\right] \le \mathrm{pr}\left(|L_1|>\varepsilon_n\right) + \mathrm{pr}\left(|L_2|>\varepsilon_n\right) + \mathrm{pr}\left(|L_3|>\varepsilon_n\right).$

We investigate these three probabilities separately. We first evaluate L_1. We write $n^3\{n(n-1)(n-2)\}^{-1}L_1$ as $U_{1,n} - \mathrm{CCov}\{(X_k-\mu_{k|\mathcal{F}}) \mid Y\} + \{(n-1)/n^2\}U_{2,n}$, where

$$
\begin{aligned}
U_{1,n} &= \binom{n}{3}^{-1}\sum_{i<l<j} h_1(X_{ik},\mathbf{x}_{i\mathcal{F}},Y_i;X_{lk},\mathbf{x}_{l\mathcal{F}},Y_l;X_{jk},\mathbf{x}_{j\mathcal{F}},Y_j),\\
U_{2,n} &= \binom{n}{2}^{-1}\sum_{i<j} h_2(X_{ik},\mathbf{x}_{i\mathcal{F}},Y_i;X_{jk},\mathbf{x}_{j\mathcal{F}},Y_j),
\end{aligned}
$$

$$
h_1(X_{ik},\mathbf{x}_{i\mathcal{F}},Y_i;X_{lk},\mathbf{x}_{l\mathcal{F}},Y_l;X_{jk},\mathbf{x}_{j\mathcal{F}},Y_j) = \omega_1(i,j,l)/3+\omega_1(i,l,j)/3+\omega_1(j,i,l)/3,
$$

$$
\begin{aligned}
\omega_1(i,j,l) \stackrel{\text{def}}{=} &\big[\mathbf{1}(Y_i<Y_j)\{X_{ik}-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\}\\
&+\mathbf{1}(Y_l<Y_j)\{X_{lk}-g_{k|\mathcal{F}}(\mathbf{x}_{l\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\}\big]/2 \qquad \text{(S.2)}
\end{aligned}
$$

and $h_2(X_{ik}, \mathbf{x}_{i\mathcal{F}}, Y_i; X_{jk}, \mathbf{x}_{j\mathcal{F}}, Y_j) = \Big(\big[\mathbf{1}(Y_i < Y_j)\{X_{ik} - g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}}, \boldsymbol{\beta}_{0,k|\mathcal{F}})\}\big]^2 + \big[\mathbf{1}(Y_j < Y_i)\{X_{jk} - g_{k|\mathcal{F}}(\mathbf{x}_{j\mathcal{F}}, \boldsymbol{\beta}_{0,k|\mathcal{F}})\}\big]^2\Big)\Big/2$.

Under H_0, $U_{1,n}$ is a degenerate U-statistic. The U-statistic theory for fixed p is systematically introduced in Serfling (1980). Zhong and Chen (2011, Section 3) studied the U-statistic theory for diverging p. They found that Hoeffding decomposition (Hoeffding, 1948) for U-statistic is still valid when p diverges, which plays a crucial role to generalize the U-statistic theory from the fixed p case to the diverging p case. We work with the Hoeffding decomposition (Hoeffding, 1948) which decomposes a U-statistic into a summation of i.i.d. random variables plus an asymptotically negligible term, which, together with the Lindeberg-Lévy central limit theorem, helps to derive the asymptotic normality even when $p \to \infty$. Another relevant work is Portnoy (1986). He showed that the central limit theorem is valid under some mild conditions when the covariate dimension is of order $o(n^{1/2})$, which is satisfied by Condition (B1). Following similar arguments used in Section 3 in Zhong and Chen (2011), we can obtain that $U_{1,n} - \mathrm{CCov}\{(X_k - \mu_{k|\mathcal{F}}) \mid Y\} = O_p(n^{-1})$ under H_0. Theorem 5.5.1 in Serfling (1980) yields $U_{1,n} - \mathrm{CCov}\{(X_k - \mu_{k|\mathcal{F}}) \mid Y\} = O_p(n^{-1/2})$ under H_1.

We turn to $U_{2,n}$. Since $U_{2,n} \xrightarrow{p} E\{h_2(X_{ik}, \mathbf{x}_{i\mathcal{F}}, Y_i; X_{jk}, \mathbf{x}_{j\mathcal{F}}, Y_j)\} < \infty$, we obtain that $U_{2,n}/(n-2) = O_p(n^{-1})$. Thus, for any $\varepsilon_n > 0$, $\mathrm{pr}\,(|L_1| > \varepsilon_n)$ is not greater than

$$\mathrm{pr}\left[\mid U_{1,n} - \mathrm{CCov}\{(X_k - \mu_{k|\mathcal{F}}) \mid Y\} \mid > \varepsilon_n/2\right] + \mathrm{pr}\left\{U_{2,n}/(n-2) > \varepsilon_n/2\right\}. \quad \text{(S.3)}$$

We evaluate the first term of (S.3). Following Theorem 2 in Zhu et al. (2011), we can similarly prove that for any $\varepsilon_n > 0$, there exists a sufficiently small constant $s_{1,\varepsilon_n} \in (0, 2/\varepsilon_n)$ satisfying $\mathrm{pr}[\mid U_{1,n} - \mathrm{CCov}\{(X_k - \mu_{k|\mathcal{F}}) \mid Y\} \mid > \varepsilon_n] \le 2\exp\{n\log(1 - \varepsilon_n s_{1,\varepsilon_n}/2)/3\}$. For the second term of (S.3), we have $\mathrm{pr}\,\{U_{2,n}/(n-2) > \varepsilon_n\} = \mathrm{pr}[U_{2,n} - \theta_{k,\mathcal{F}} > \{(n-2)\varepsilon_n - \theta_{k,\mathcal{F}}\}]$, where $0 < \theta_{k,\mathcal{F}} = E\{h_2(X_{ik}, \mathbf{x}_{i\mathcal{F}}, Y_i; X_{jk}, \mathbf{x}_{j\mathcal{F}}, Y_j)\} < \infty$. Similarly, we can obtain that for any $\varepsilon_n > 0$, there exists a sufficiently small constant $s_{2,\varepsilon_n} \in (0, 2/\varepsilon_n)$ satisfying $\mathrm{pr}(U_{2,n} - \theta_{k,\mathcal{F}} > \varepsilon_n) \le \exp\{n\log(1 - \varepsilon_n s_{2,\varepsilon_n}/2)/2\}$. Since for any $\varepsilon_n > 0$, it holds true that $\{(n-2)\varepsilon_n - \theta_{k,\mathcal{F}}\} > \varepsilon_n$ when n is sufficiently large. Thus we conclude that $\mathrm{pr}\{U_{2,n}/(n-2) > \varepsilon_n\} \le \exp\{n\log(1 - \varepsilon_n s_{2,\varepsilon_n}/2)/2\}$. Set $s_{\varepsilon_n} = \min\{s_{1,\varepsilon_n}, s_{2,\varepsilon_n}\}$. It follows that $\mathrm{pr}\,(|L_1| > \varepsilon_n) \le 3\exp\{n\log(1 - \varepsilon_n s_{\varepsilon_n}/2)/3\}$.

Next we deal with L_2. With Taylor's expansion and regularity condition (B3), we have $\mathrm{pr}\big(|L_2| > \varepsilon_n\big) < 2\ \mathrm{pr}\big(|L_{21}L_{22}L_{23}| > \varepsilon_n/4\big)$ where

$$\begin{aligned}
L_{21} &= n^{-1}\sum_{j=1}^{n}\left[n^{-1}\sum_{i=1}^{n}\mathbf{1}(Y_i < Y_j)\{X_{ik} - g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}}, \boldsymbol{\beta}_{0,k|\mathcal{F}})\}\right], \\
L_{22} &= \left[n^{-1}\sum_{i=1}^{n}\{g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}}, \boldsymbol{\beta}_{0,k|\mathcal{F}})\}^{\mathrm{T}} g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}}, \boldsymbol{\beta}_{0,k|\mathcal{F}})\right]^{1/2}, \\
L_{23} &= \left[(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}} - \boldsymbol{\beta}_{0,k|\mathcal{F}})^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}} - \boldsymbol{\beta}_{0,k|\mathcal{F}})\right]^{1/2}.
\end{aligned}$$

As $\mathrm{pr}\left(|L_{21}L_{22}L_{23}| > \varepsilon_n\right)$ is equal to

$$
\begin{aligned}
& \mathrm{pr}\left\{|L_{21}L_{22}L_{23}| > \varepsilon_n, |L_{21}| > (2M_0+\varepsilon_n)\right\} + \mathrm{pr}\left\{|L_{21}L_{22}L_{23}| > \varepsilon_n, |L_{21}| \le (2M_0+\varepsilon_n)\right\} \\
\le\ & \mathrm{pr}\left\{|L_{21}| > (2M_0+\varepsilon_n)\right\} + \mathrm{pr}\left\{|L_{22}L_{23}| \ge \varepsilon_n/(2M_0+\varepsilon_n)\right\} \\
\le\ & \mathrm{pr}\left\{|L_{21}| > (2M_0+\varepsilon_n)\right\} + \mathrm{pr}\left\{|L_{22}| > (2M_n+\varepsilon_n)^{1/2}\right\} \\
+\ & \mathrm{pr}\left[|L_{23}| \ge \varepsilon_n/\{(2M_0+\varepsilon_n)(2M_n+\varepsilon_n)^{1/2}\}\right],
\end{aligned}
$$

where $M_0 \stackrel{\text{def}}{=} E\left[n^{-2}\sum_{j=1}^n\sum_{i=1}^n \mathbf{1}(Y_i<Y_j)\{X_{ik}-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\}\right]$, and $M_n \stackrel{\text{def}}{=} E\left[\{g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\}^{\mathrm{T}}g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\right]$. Then we have $M_0=0$ under H_0, $0<|M_0|<\infty$ under H_1 and $M_n=O(s)$, where $s=|\mathcal{F}|$. It follows from Lemma 1 that $\mathrm{pr}\{|L_{21}|>(2M_0+\varepsilon_n) \le \mathrm{pr}\{|L_{21}-M_0|>\varepsilon_n\} \le 2n\exp(-c_1n\varepsilon_n^2)$, where c_1 is a positive constant. Similarly $\mathrm{pr}\left\{|L_{22}|>(2M_n+\varepsilon_n)^{1/2}\right\} \le \mathrm{pr}\left\{|L_{22}^2-M_n|>\varepsilon_n\right\} \le 2s\exp(-c_2ns^{-2}\varepsilon_n^2)$, where c_2 is a positive constant. Lemma 2 yields that $\mathrm{pr}\Big(|L_{23}| > \varepsilon_n/\{(2M_0+\varepsilon_n)(2M_n+\varepsilon_n)^{1/2}\}\Big) \le 2s\exp(-c_3ns^{-2}\varepsilon_n^2)$. Thus we have $\mathrm{pr}\left(|L_2|>\varepsilon_n\right) < 4n\exp(-c_1n\varepsilon_n^2)+4s\exp(-c_2ns^{-2}\varepsilon_n^2)+4s\exp(-c_3ns^{-2}\varepsilon_n^2)$.

Next we deal with L_3. With Taylor's expansion, it is not difficult to show that $\mathrm{pr}\Big(|L_3|>\varepsilon_n\Big) < 2\ \mathrm{pr}\Big(|L_{31}L_{32}|>\varepsilon_n/2\Big)$ where

$$
\begin{aligned}
L_{31} &= n^{-1}\sum_{i=1}^n\left[\{g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\}^{\mathrm{T}}g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\right], \\
L_{32} &= \left[(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}})^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}}-\boldsymbol{\beta}_{0,k|\mathcal{F}})\right].
\end{aligned}
$$

As $\mathrm{pr}\left\{|L_{31}L_{32}|>\varepsilon_n\right\} \le \mathrm{pr}\left\{|L_{31}|>(2M_n+\varepsilon_n)\right\}+\mathrm{pr}\left\{|L_{32}|>\varepsilon_n/(2M_n+\varepsilon_n)\right\}$, by Lemma 1 and Lemma 2, we obtain that $\mathrm{pr}\left\{|L_3|>\varepsilon_n\right\} \le 4s\exp(-c_4ns^{-2}\varepsilon_n^2)+4s\exp(-c_5ns^{-2}\varepsilon_n)$, where c_4 and c_5 are some positive constants.

Next, we evaluate $\left\{\widehat{\mathrm{var}}(X_k-\mu_{k|\mathcal{F}})-\mathrm{var}(X_k-\mu_{k|\mathcal{F}})\right\}$ by writing it into three parts as $M_1+M_2+M_3$, where

$$
\begin{aligned}
M_1 &\stackrel{\text{def}}{=} n^{-1}\sum_{i=1}^n\{X_{ik}-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\}^2-\mathrm{var}\{X_k-g_{k|\mathcal{F}}(\mathbf{x}_{\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\}, \\
M_2 &\stackrel{\text{def}}{=} 2n^{-1}\sum_{i=1}^n\{X_{ik}-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})\}\{g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}})\}, \\
M_3 &\stackrel{\text{def}}{=} n^{-1}\sum_{i=1}^n\{g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\boldsymbol{\beta}_{0,k|\mathcal{F}})-g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}},\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}})\}^2.
\end{aligned}
$$

By Lemma 1, we have $\mathrm{pr}(|M_1|>\varepsilon_n) \le 2\exp(-c_6n\varepsilon_n^2)$. Besides, we have $\mathrm{pr}(|M_2|>$

$\varepsilon_n) \leq 2\ \mathrm{pr}(|M_{21}M_{22}| > \varepsilon_n/4)$ where

$$
\begin{aligned}
M_{21} &= n^{-1}\sum_{i=1}^{n}\{X_{ik} - g_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}}, \boldsymbol{\beta}_{0,k|\mathcal{F}})\}\left[\{g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}}, \boldsymbol{\beta}_{0,k|\mathcal{F}})\}^{\mathrm{T}} g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}}, \boldsymbol{\beta}_{0,k|\mathcal{F}})\right]^{1/2}, \\
M_{22} &= \left[(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}} - \boldsymbol{\beta}_{0,k|\mathcal{F}})^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}} - \boldsymbol{\beta}_{0,k|\mathcal{F}})\right]^{1/2}.
\end{aligned}
$$

Similar to the arguments in evaluating L_2, we obtain $\mathrm{pr}(|M_{21}M_{22}| > \varepsilon_n/4) \leq 4s\exp(-c_7 ns^{-2}\varepsilon_n^2)$.

As $\mathrm{pr}(|M_3| > \varepsilon_n) \leq 2\mathrm{pr}(|M_{31}M_{32}| > \varepsilon_n/2)$ where

$$
\begin{aligned}
M_{31} &= n^{-1}\sum_{i=1}^{n}\left[\{g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}}, \boldsymbol{\beta}_{0,k|\mathcal{F}})\}^{\mathrm{T}} g'_{k|\mathcal{F}}(\mathbf{x}_{i\mathcal{F}}, \boldsymbol{\beta}_{0,k|\mathcal{F}})\right], \\
M_{32} &= (\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}} - \boldsymbol{\beta}_{0,k|\mathcal{F}})^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{k|\mathcal{F}} - \boldsymbol{\beta}_{0,k|\mathcal{F}}).
\end{aligned}
$$

Using Lemmas 1 and 2 repeatedly, we obtain $\mathrm{pr}(|M_3| > \varepsilon_n) \leq 4s\exp(-c_8 ns^{-2}\varepsilon_n^2) + 4s\exp(-c_9 ns^{-2}\varepsilon_n)$, where c_8, c_9 are positive constants.

In summary, for any $\varepsilon_n > 0$, there exist positive constants c_{10}, c_{11}, c_{12}, c_{13} and $s_{\varepsilon_n} \in (0, 2/\varepsilon_n)$ such that $\mathrm{pr}\{|\widehat{\omega}_{k|\mathcal{F}} - \omega_{k|\mathcal{F}}| > \varepsilon_n\} \leq O\Big[\exp\{n\log(1 - \varepsilon_n s_{\varepsilon_n}/2)/3\} + \exp(-c_{10}n\varepsilon_n^2) + n\exp(-c_{11}n\varepsilon_n^2) + s\exp(-c_{12}ns^{-2}\varepsilon_n^2) + s\exp(-c_{13}ns^{-2}\varepsilon_n)\Big]$.

Given a working index set $\mathcal{F}$, $\mathrm{pr}\big\{\max_{k\in\mathcal{F}^c}|\widehat{\omega}_{k|\mathcal{F}} - \omega_{k|\mathcal{F}}| > \varepsilon_n\big\} \leq (p-s)\max_{k\in\mathcal{F}^c}\mathrm{pr}\big\{|\widehat{\omega}_{k|\mathcal{F}} - \omega_{k|\mathcal{F}}| > \varepsilon_n\big\}$, which yields that $\mathrm{pr}\{\max_{k\in\mathcal{F}^c}|\widehat{\omega}_{k|\mathcal{F}} - \omega_{k|\mathcal{F}}| > \varepsilon_n\}$ is not greater than $O\Big[p\exp\{n\log(1-\varepsilon_n s_{\varepsilon_n}/2)/3\} + p\exp(-c_{10}n\varepsilon_n^2) + pn\exp(-c_{11}n\varepsilon_n^2) + ps\exp(-c_{12}ns^{-2}\varepsilon_n^2) + ps\exp(-c_{13}ns^{-2}\varepsilon_n)\Big]$. This completes the proof of Proposition 1.

Proof of Theorem 1: The first two statements are obvious by noting that $\mathrm{cov}^2\big\{Y, \mathbf{1}(X < \widetilde{X}) \mid \widetilde{X}\big\} \leq \mathrm{var}(Y \mid \widetilde{X})\mathrm{var}\big\{\mathbf{1}(X < \widetilde{X}) \mid \widetilde{X}\big\}$, and $E\big\{\mathrm{var}(Y \mid \widetilde{X})\big\} \leq \mathrm{var}(Y)$, $\mathrm{var}\big\{\mathbf{1}(X < \widetilde{X}) \mid \widetilde{X}\big\} \leq 1/4$.

Next we prove the third assertion. Without loss of generality, we can assume both X and Y are standard normal. In other words,

$$
\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).
$$

Let $F(x \mid Y) = \mathrm{pr}(X < x \mid Y)$. We first note that $(X \mid Y) \sim \mathcal{N}(\rho Y, 1-\rho^2)$ and

$$
\begin{aligned}
\frac{\partial F(x \mid Y)}{\partial Y} &= \frac{\partial \mathrm{pr}\left\{\mathcal{N}(0,1) < (x - \rho Y)/\sqrt{1-\rho^2} \mid Y\right\}}{\partial Y} \\
&= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x-\rho Y)^2}{2(1-\rho^2)}\right\}\left(-\frac{\rho}{\sqrt{1-\rho^2}}\right).
\end{aligned}
$$

Lemma 1 yields that

$$E\left\{\frac{\partial F(x\mid Y)}{\partial Y}\right\}=-\frac{\rho}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\text{ and }E_{\widetilde{X}}\left[E_{Y\mid\widetilde{X}}\left\{\frac{\partial F(\widetilde{X}\mid Y)}{\partial Y}\right\}\right]^2=\frac{\rho^2}{2\sqrt{3}\pi}.$$

In general, if $\text{var}(X)=\sigma_X^2$, $\text{var}(Y)=\sigma_Y^2$, we can obtain $E_{\widetilde{X}}\left[E_{Y\mid\widetilde{X}}\left\{\partial F(\widetilde{X}\mid Y)/\partial Y\right\}\right]^2=\rho^2/(2\sqrt{3}\pi\sigma_Y^2)$. We further apply Stein (1981)'s lemma to get that $E\left\{\partial F(x\mid Y)/\partial Y\right\}=E\left\{F(x\mid Y)Y\right\}/\sigma_Y^2=\text{cov}\left\{\mathbf{1}(X<x),Y\right\}/\sigma_Y^2$, which indeed connects $E_{Y\mid\widetilde{X}}\big\{\partial F(\widetilde{X}\mid Y)/\partial Y\big\}$ with $\text{CD}(Y\mid X)$. To be precise, $\text{CD}(Y\mid X)=\rho^2/(2\sqrt{3}\pi)$.

It remains to prove the last assertion. Recall that we merely assume that $Y\sim N(0,\sigma^2)$, where $\sigma^2=\text{var}(Y)$. With integration by parts, we obtain that

$$\begin{aligned}E\left\{\partial F(x\mid Y)/\partial Y\right\}&=\int\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-y^2/(2\sigma^2)\right\}dF(x\mid y)\\&=E\{F(x\mid Y)Y\}/\sigma^2=\text{cov}\{\mathbf{1}(X<x),Y\}/\text{var}(Y).\end{aligned}$$

Accordingly, $E\left[E^2\left\{\partial F(\widetilde{X}\mid Y)/\partial Y|\widetilde{X}\right\}\right]=E\left[\text{cov}\left\{\mathbf{1}(X<\widetilde{X}),Y\mid\widetilde{X}\right\}\right]^2\Big/\text{var}^2(Y)$. This completes the proof of Theorem 1. □

REFERENCES

Hoeffding, W. (1948). "A class of statistics with asymptotically normal distribution." *The Annals of Mathematical Statistics*, **19**, 293–325.

Portnoy, S. (1986). "On the central limit theorem in $\mathbb{R}^p$ when $p\to\infty$." *Probability Theory and Related Fields*, **73**, 571–583.

Serfling, R. L. (1980) *Approximation Theorems in Mathematical Statistics.* New York: Wiley.

Stein, C. (1981) "Estimation of the mean of a multivariate normal distribution." *The Annals of Statistics*, **9**, 1135–1151.

Zhong, P. S. and Chen, S. X. (2011). "Tests for high-dimensional regression coefficients with factorial designs." *Journal of the American Statistical Association*, **106**, 260–274.

Zhu, L. P., Li, L., Li, R., and Zhu, L. X. (2011). "Model-free feature screening for ultrahigh dimensional data." *Journal of the American Statistical Association*, **106**, 1464–1475.