Appendix for
"Information-Based Optimal Subdata Selection for Big Data Linear Regression"

HaiYing Wang, Min Yang, and John Stufken

A Proofs and Technical Details

A.1 Proof of Theorem 1

We will use the following convexity result (cf. Nordström, 2011) in the proof of Theorem 1.

Lemma 1. *For any positive definite matrices* $\mathbf{B}_1$ *and* $\mathbf{B}_2$ *of the same dimension,*

$$\{\alpha\mathbf{B}_1 + (1-\alpha)\mathbf{B}_2\}^{-1} \le \alpha\mathbf{B}_1^{-1} + (1-\alpha)\mathbf{B}_2^{-1} \tag{1}$$

in the Loewner ordering, where $0 \le \alpha \le 1$.

Proof of Theorem 1. The unbiasedness can be verified by direct calculation,

$$\mathrm{E}\{\tilde{\boldsymbol{\beta}}_L|\mathbf{Z}, I_\Delta(\boldsymbol{\eta}_L)=1\} = \mathrm{E}_{\boldsymbol{\eta}_L}[\mathrm{E}_{\mathbf{y}}\{\tilde{\boldsymbol{\beta}}_L|\mathbf{Z}, I_\Delta(\boldsymbol{\eta}_L)=1\}] = \mathrm{E}_{\boldsymbol{\eta}_L}(\boldsymbol{\beta}) = \boldsymbol{\beta}.$$

Let $\mathbf{W} = \mathrm{diag}(w_1\eta_{L1}, ..., w_n\eta_{Ln})$. The variance-covariance matrix of the sampling-based estimators can be written as

$$\begin{aligned}
\mathrm{V}\{\tilde{\boldsymbol{\beta}}_L|\mathbf{Z}, I_\Delta(\boldsymbol{\eta}_L)=1\} =& \mathrm{E}_{\boldsymbol{\eta}_L}[\mathrm{V}_{\mathbf{y}}\{\tilde{\boldsymbol{\beta}}_L|\mathbf{Z}, I_\Delta(\boldsymbol{\eta}_L)=1\}] + \mathrm{V}_{\boldsymbol{\eta}_L}[\mathrm{E}_{\mathbf{y}}\{\tilde{\boldsymbol{\beta}}_L|\mathbf{Z}, I_\Delta(\boldsymbol{\eta}_L)=1\}]\\
=&\sigma^2\mathrm{E}_{\boldsymbol{\eta}_L}\left\{\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}^2\mathbf{X}\right)\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\right\} + \mathrm{V}_{\boldsymbol{\eta}_L}(\boldsymbol{\beta})\\
=&\sigma^2\mathrm{E}_{\boldsymbol{\eta}_L}\left[\left\{\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}^2\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)\right\}^{-1}\right]\\
\ge&\sigma^2\left[\mathrm{E}_{\boldsymbol{\eta}_L}\left\{\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}^2\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)\right\}\right]^{-1}.
\end{aligned} \tag{2}$$

The last inequality is due to Lemma 1. Notice that $\mathbf{WX}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}^2\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W} = \mathrm{pr}(\mathbf{WX})$, the orthogonal projection matrix onto the column space of $\mathbf{WX}$. Define

$$\mathbf{B}_{WX} = \begin{bmatrix} w_1\eta_{L1}\mathbf{x}_1^{\mathrm{T}} & & \\ & \ddots & \\ & & w_n\eta_{Ln}\mathbf{x}_n^{\mathrm{T}} \end{bmatrix}.$$

Notice that the column-space of $\mathbf{WX} = (w_1\eta_{L1}\mathbf{x}_1, ..., w_n\eta_{Ln}\mathbf{x}_n)^{\mathrm{T}}$ is contained in the column-space of $\mathbf{B}_{WX}$. Hence we have $\mathrm{pr}(\mathbf{WX}) \leq \mathrm{pr}(\mathbf{B}_{WX})$ in the Loewner ordering, i.e.,

$$\mathbf{WX}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}^2\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W} \leq \begin{bmatrix} \mathbf{x}_1^{\mathrm{T}}\left(\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}\right)^{-}\mathbf{x}_1 I(\eta_{L1} > 0) & & \\ & \ddots & \\ & & \mathbf{x}_n^{\mathrm{T}}\left(\mathbf{x}_n\mathbf{x}_n^{\mathrm{T}}\right)^{-}\mathbf{x}_n I(\eta_{Ln} > 0) \end{bmatrix}.$$

where $I()$ is the indicator function. From this result, it can be shown that

$$\mathbf{X}^{\mathrm{T}}\mathbf{WX}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}^2\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{WX} \leq \sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} I(\eta_{Li} > 0). \tag{3}$$

For sampling with replacement,

$$P(\eta_{Li} > 0|\mathbf{Z}) = 1 - (1-\pi_i)^k = \pi_i\sum_{i=1}^{k}(1-\pi_i)^{i-1} \leq k\pi_i.$$

For sampling without replacement,

$$P(\eta_{Li} > 0|\mathbf{Z}) = P(\eta_{Li} = 1|\mathbf{Z}) = k\pi_i.$$

Thus, in either case, $P(\eta_{Li} > 0|\mathbf{Z}) \leq k\pi_i$. Therefore,

$$\begin{aligned} P\{\eta_{Li} > 0|\mathbf{Z}, I_\Delta(\boldsymbol{\eta}_L) = 1\} &= \frac{P\{\eta_{Li} > 0, I_\Delta(\boldsymbol{\eta}_L) = 1|\mathbf{Z}\}}{P\{I_\Delta(\boldsymbol{\eta}_L) = 1|\mathbf{Z}\}} \\ &\leq \frac{P(\eta_{Li} > 0|\mathbf{Z})}{P\{I_\Delta(\boldsymbol{\eta}_L) = 1|\mathbf{Z}\}} \leq \frac{k\pi_i}{P\{I_\Delta(\boldsymbol{\eta}_L) = 1|\mathbf{Z}\}}. \end{aligned} \tag{4}$$

Combining (2), (3) and (4), we have

$$\begin{aligned} \mathrm{V}\{\tilde{\boldsymbol{\beta}}_L|\mathbf{Z}, I_\Delta(\boldsymbol{\eta}_L) = 1\} &\geq \sigma^2\left[\mathrm{E}_{\boldsymbol{\eta}_L}\left\{\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} I(\eta_{Li} > 0)\right\}\right]^{-1} \\ &= \sigma^2\left[\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} P\{\eta_{Li} > 0|\mathbf{Z}, I_\Delta(\boldsymbol{\eta}_L) = 1\}\right]^{-1} \\ &\geq \frac{\sigma^2 P\{I_\Delta(\boldsymbol{\eta}_L) = 1|\mathbf{Z}\}}{k}\left\{\sum_{i=1}^{n}\pi_i\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right\}^{-1}. \end{aligned}$$

□

A.2 Proof of Theorem 2

Proof. Let $\breve{z}_{ij} = \{2z_{ij} - (z_{(n)j} + z_{(1)j})\}/(z_{(n)j} - z_{(1)j})$. Then we have,

$$\sum_{i=1}^{n} \delta_i \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} = k\mathbf{B}_3^{-1}\breve{\mathbf{M}}(\boldsymbol{\delta})(\mathbf{B}_3^{\mathrm{T}})^{-1}, \tag{5}$$

where

$$\breve{\mathbf{M}}(\boldsymbol{\delta}) = \begin{bmatrix} 1 & k^{-1}\sum_{i=1}^{n}\delta_i\breve{z}_{i1} & \dots & k^{-1}\sum_{i=1}^{n}\delta_i\breve{z}_{id} \\ k^{-1}\sum_{i=1}^{n}\delta_i\breve{z}_{i1} & k^{-1}\sum_{i=1}^{n}\delta_i\breve{z}_{i1}^2 & \dots & k^{-1}\sum_{i=1}^{n}\delta_i\breve{z}_{i1}\breve{z}_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ k^{-1}\sum_{i=1}^{n}\delta_i\breve{z}_{ip} & k^{-1}\sum_{i=1}^{n}\delta_i\breve{z}_{i1}\breve{z}_{ip} & \dots & k^{-1}\sum_{i=1}^{n}\delta_i\breve{z}_{ip}^2 \end{bmatrix},$$

and

$$\mathbf{B}_3 = \begin{bmatrix} 1 & & & \\ -\frac{z_{(n)1}+z_{(1)1}}{z_{(n)1}-z_{(1)1}} & \frac{2}{z_{(n)1}-z_{(1)1}} & & \\ \vdots & & \ddots & \\ -\frac{z_{(n)p}+z_{(1)p}}{z_{(n)p}-z_{(1)p}} & & & \frac{2}{z_{(n)p}-z_{(1)p}} \end{bmatrix} \tag{6}$$

Note that $\breve{z}_{ij} \in [-1, 1]$ for all $i = 1, ..., n$ and $j = 1, ..., p$, which implies $k^{-1}\sum_{i=1}^{n}\delta_i\breve{z}_{ij}^2 \le 1$ for all $1 \le j \le p$. Thus,

$$|\breve{\mathbf{M}}(\boldsymbol{\delta})| = \prod_{j=0}^{p}\lambda_j \le \left(\frac{\sum_{j=0}^{p}\lambda_j}{p+1}\right)^{p+1} = \left(\frac{1+\sum_{j=1}^{p}k^{-1}\sum_{i=1}^{n}\delta_i\breve{z}_{ij}^2}{p+1}\right)^{p+1} \le 1, \tag{7}$$

where λ_j, $j = 0, 1, ..., p$ are eigenvalues of $\breve{\mathbf{M}}(\boldsymbol{\delta})$. From (5), (6) and (7),

$$\left|\sum_{i=1}^{n}\delta_i\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right| = k^{p+1}|\mathbf{B}_3|^{-2}|\breve{\mathbf{M}}(\boldsymbol{\delta})| \le k^{p+1}\left|\prod_{j=1}^{p}\frac{2}{z_{(n)j}-z_{(1)j}}\right|^{-2} = \frac{k^{p+1}}{4^p}\prod_{j=1}^{p}(z_{(n)j}-z_{(1)j})^2.$$

If the subdata consists of the 2^p points $(a_1, \dots, a_p)^{\mathrm{T}}$ where $a_j = z_{(n)j}$ or $z_{(1)j}$, $j = 1, 2, ..., p$, each occurring equally often, then the $\boldsymbol{\delta}^{opt}$ corresponding to this subdata satisfies $\breve{\mathbf{M}}(\boldsymbol{\delta}) = \mathbf{I}$. This $\boldsymbol{\delta}^{opt}$ attains equality in (7) and corresponds therefore to D-optimal subdata. □

A.3 Proof of Theorem 3

Proof. As before, for $i = 1, ..., n$, $j = 1, ..., p$, let $z_{(i)j}$ be the ith order statistic for $z_{1j}, ..., z_{nj}$. For $l \neq j$, let $z_j^{(i)l}$ be the concomitant of $z_{(i)l}$ for z_j, i.e., if $z_{(i)l} = z_{sl}$ then $z_j^{(i)l} = z_{sj}$,

$i = 1, ..., n$. For the subdata obtained from Algorithm 1, let $\bar{z}_j^*$ and $\mathrm{var}(z_j^*)$ be the sample mean and sample variance for covariate z_j. From Algorithm 1, the values z_j, $j = 1, ..., p$, in the subdata consist of $z_{(m)j}$, and $z_j^{(m)l}$, $l = 1, ...j-1, j+1, ..., p$, $m = 1, ..., r$, $n-r+1, ..., n$. Note that the subdata may not contain exactly the r smallest and r largest values for each covariate since some data points may be removed in processing each covariate. However, since r is fixed when n goes to infinity, this will not affect the final result. Therefore, for easy of presentation, we abuse the notation and write the range of values of m as $1, ..., r$, $n - r + 1, ..., n$. The information matrix based on the subdata can be written as

$$(\mathbf{X}_{\mathrm{D}}^*)^{\mathrm{T}}\mathbf{X}_{\mathrm{D}}^* = \mathbf{B}_4^{-1}\begin{bmatrix} k & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & (k-1)\mathbf{R} \end{bmatrix}(\mathbf{B}_4^{\mathrm{T}})^{-1}, \tag{8}$$

where

$$\mathbf{B}_4 = \begin{bmatrix} 1 & & & \\ -\frac{\bar{z}_1^*}{\sqrt{\mathrm{var}(z_1^*)}} & \frac{1}{\sqrt{\mathrm{var}(z_1^*)}} & & \\ \vdots & & \ddots & \\ -\frac{\bar{z}_p^*}{\sqrt{\mathrm{var}(z_p^*)}} & & & \frac{1}{\sqrt{\mathrm{var}(z_p^*)}} \end{bmatrix}. \tag{9}$$

From (8) and (9),

$$|(\mathbf{X}_{\mathrm{D}}^*)^{\mathrm{T}}\mathbf{X}_{\mathrm{D}}^*| = k|(k-1)\mathbf{R}|\prod_{j=1}^{p}\mathrm{var}(z_j^*) \geq k(k-1)^p\lambda_{\min}^p(\mathbf{R})\prod_{j=1}^{p}\mathrm{var}(z_j^*). \tag{10}$$

For each sample variance,

$$\begin{aligned}
(k-1)\mathrm{var}(z_j^*) =& \sum_{i=1}^{k}\left(z_{ij}^* - \bar{z}_j^*\right)^2 \\
=& \left(\sum_{i=1}^{r} + \sum_{i=n-r+1}^{n}\right)\left(z_{(i)j} - \bar{z}_j^*\right)^2 + \sum_{l\neq j}\left(\sum_{i=1}^{r} + \sum_{i=n-r+1}^{n}\right)\left(z_j^{(i)l} - \bar{z}_j^*\right)^2 \\
\geq& \left(\sum_{i=1}^{r} + \sum_{i=n-r+1}^{n}\right)\left(z_{(i)j} - \bar{z}_j^{**}\right)^2 \\
=& \sum_{i=1}^{r}\left(z_{(i)j} - \bar{z}_j^{*l}\right)^2 + \sum_{i=n-r+1}^{n}\left(z_{(i)j} - \bar{z}_j^{*u}\right)^2 + \frac{r}{2}\left(\bar{z}_j^{*u} - \bar{z}_j^{*l}\right)^2 \\
\geq& \frac{r}{2}\left(\bar{z}_j^{*u} - \bar{z}_j^{*l}\right)^2
\end{aligned}$$

$$\geq \frac{r}{2}\left(z_{(n-r+1)j} - z_{(r)j}\right)^2 \tag{11}$$

where $\bar{z}_j^{**} = \left(\sum_{i=1}^{r} + \sum_{i=n-r+1}^{n}\right) z_{(i)j}/(2r)$, $\bar{z}_j^{*l} = \sum_{i=1}^{r} z_{(i)j}/r$, and $\bar{z}_j^{*u} = \sum_{i=n-r+1}^{n} z_{(i)j}/r$. From (11),

$$\mathrm{var}(z_j^*) \geq \frac{r\left(z_{(n)j} - z_{(1)j}\right)^2}{2(k-1)} \left(\frac{z_{(n-r+1)j} - z_{(r)j}}{z_{(n)j} - z_{(1)j}}\right)^2. \tag{12}$$

Thus,

$$\begin{aligned}
|(\mathbf{X}_{\mathrm{D}}^*)^{\mathrm{T}}\mathbf{X}_{\mathrm{D}}^*| \geq& k(k-1)^p \lambda_{\min}^p(\mathbf{R}) \prod_{j=1}^{p} \frac{r\left(z_{(n)j} - z_{(1)j}\right)^2}{2(k-1)} \left(\frac{z_{(n-r+1)j} - z_{(r)j}}{z_{(n)j} - z_{(1)j}}\right)^2 \\
=& \frac{r^p}{2^p} k \lambda_{\min}^p(\mathbf{R}) \prod_{j=1}^{p} (z_{(n)j} - z_{(1)j})^2 \times \prod_{j=1}^{p} \left(\frac{z_{(n-r+1)j} - z_{(r)j}}{z_{(n)j} - z_{(1)j}}\right)^2.
\end{aligned}$$

This shows that

$$\frac{|(\mathbf{X}_{\mathrm{D}}^*)^{\mathrm{T}}\mathbf{X}_{\mathrm{D}}^*|}{\frac{k^{p+1}}{4^p}\prod_{j=1}^{p}(z_{(n)j} - z_{(1)j})^2} \geq \frac{\lambda_{\min}^p(\mathbf{R})}{p^p} \times \prod_{j=1}^{p} \left(\frac{z_{(n-r+1)j} - z_{(r)j}}{z_{(n)j} - z_{(1)j}}\right)^2.$$

□

A.4 Proof of Theorem 4

Proof. From (8) and (9),

$$\mathrm{V}(\hat{\boldsymbol{\beta}}^{\mathrm{D}}|\mathbf{Z}) = \sigma^2\{(\mathbf{X}_{\mathrm{D}}^*)^{\mathrm{T}}\mathbf{X}_{\mathrm{D}}^*\}^{-1} = \sigma^2 \mathbf{B}_4^{\mathrm{T}} \begin{bmatrix} \frac{1}{k} & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & \frac{1}{k-1}\mathbf{R}^{-1} \end{bmatrix} \mathbf{B}_4.$$

Thus

$$\mathrm{V}(\hat{\beta}_0^{\mathrm{D}}|\mathbf{Z}) = \sigma^2 \left(\frac{1}{k} + \frac{1}{k-1}\mathbf{u}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{u}\right), \tag{13}$$

and

$$\mathrm{V}(\hat{\beta}_j^{\mathrm{D}}|\mathbf{Z}) = \frac{\sigma^2}{k-1}\frac{(\mathbf{R}^{-1})_{jj}}{\mathrm{var}(z_j^*)}, \tag{14}$$

where $\mathbf{u} = \left\{-\bar{z}_1^*/\sqrt{\mathrm{var}(z_1^*)}, ..., -\bar{z}_p^*/\sqrt{\mathrm{var}(z_p^*)}\right\}^{\mathrm{T}}$ and $(\mathbf{R}^{-1})_{jj}$ is the jth diagonal element of $\mathbf{R}^{-1}$.

From (13), $\mathrm{V}(\hat{\beta}_0^{\mathrm{D}}|\mathbf{Z}) \geq \sigma^2/k$ because $\mathbf{u}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{u} \geq 0$.

Denote the spectral decomposition of $\mathbf{R}$ as $\mathbf{R} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}}$. Since $\boldsymbol{\Lambda}^{-1} \leq \lambda_{\min}^{-1}(\mathbf{R})\mathbf{I}_p$, $\mathbf{R}^{-1} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^{\mathrm{T}} \leq \mathbf{V}\lambda_{\min}^{-1}(\mathbf{R})\mathbf{I}_p\mathbf{V}^{\mathrm{T}} = \lambda_{\min}^{-1}(\mathbf{R})\mathbf{I}_p^{\mathrm{T}}$. Thus $\mathbf{R}_{jj}^{-1} \leq \lambda_{\min}^{-1}(\mathbf{R})$ for all j. From this fact, and (14) and (12), we have

$$\mathrm{V}(\hat{\beta}_j^{\mathrm{D}}|\mathbf{Z}) = \frac{\sigma^2}{k-1}\frac{(\mathbf{R}^{-1})_{jj}}{\mathrm{var}(z_j^*)} \leq \frac{4p\sigma^2}{k\lambda_{\min}(\mathbf{R})(z_{(n-r+1)j} - z_{(r)j})^2}. \tag{15}$$

Similarly, we have

$$\mathrm{V}(\hat{\beta}_j^{\mathrm{D}}|\mathbf{Z}) = \frac{\sigma^2}{k-1}\frac{(\mathbf{R}^{-1})_{jj}}{\mathrm{var}(z_j^*)} \geq \frac{4\sigma^2}{k\lambda_{\max}(\mathbf{R})(z_{(n)j} - z_{(1)j})^2}. \tag{16}$$

Here we utilize the following inequality

$$\mathrm{var}(z_j^*) \leq \frac{1}{k-1}\sum_{i=1}^{k}\left(z_{ij}^* - \frac{z_{(n)j} + z_{(1)j}}{2}\right)^2 \leq \frac{k}{4(k-1)}\left(z_{(n)j} - z_{(1)j}\right)^2, \tag{17}$$

where the last inequality is due to the fact $|z_{ij}^* - \frac{z_{(n)j}+z_{(1)j}}{2}| \leq \frac{z_{(n)j}-z_{(1)j}}{2}$ for all $i = 1, \ldots, k$. □

A.5 Proof of Theorem 5

Proof. For (21), it is a direct result from (20).

For (22), we consider the five cases in the following. For the first case that r is fixed, from results in Theorems 2.8.1 and 2.8.2 in Galambos (1987), we have that

$$\frac{z_{(n-r+1)j} - z_{(r)j}}{z_{(n)j} - z_{(1)j}} = O_P(1) \quad \text{and} \quad \frac{z_{(n)j} - z_{(1)j}}{z_{(n-r+1)j} - z_{(r)j}} = O_P(1). \tag{18}$$

Combining (21) and (18), (22) follows.

For the second case when $r \to \infty$, $r/n \to 0$, and the support of F_j is bounded, (18) can be easily verified.

For the third case when the upper endpoint for the support of F_j is ∞ and the lower endpoint for the support of F_j is finite, and $r \to \infty$ slow enough such that (23) holds, if we can show that $z_{(n-r+1)j}/z_{(n)j} = 1 + o_P(1)$, then the result in (22) follows. Let $b_{n,j} = F_j^{-1}(1 - n^{-1})$. From Hall (1979), we only need to show that $z_{(n-r+1)j}/b_{n,j} = 1 + o_P(1)$ in order to show that $z_{(n-r+1)j}/z_{(n)j} = 1 + o_P(1)$. For this, from the proof of Theorem 1 of Hall (1979), it suffices to show that

$$\left[\frac{1 - F_j(b_{n,j})}{1 - F_j\{(1-\epsilon)b_{n,j}\}}\right]^{-1/2}\left[1 - \frac{r\{1 - F_j(b_{n,j})\}}{1 - F_j\{(1-\epsilon)b_{n,j}\}}\right] \to \infty,$$

which holds by directly applying the assumption in (23) and the fact that $r \to \infty$.

For the fourth case, it can be proved by using an approach similar to the one used for the third case. It can also be proved by noting that $z_{(r)j} = -(-z)_{(n-r+1)j}$, $z_{(1)j} = -(-z)_{(n)j}$, and the fact that the condition in (24) on $\mathbf{z}$ becomes the condition in (23) on $-\mathbf{z}$.

For the fifth case, it can be proved by combining the results in the third case and the fourth case. $\square$

A.6 Proof of Theorem 6

Let σ_j and $\rho_{j_1 j_2}$ be the jth diagonal element of $\mathbf{\Phi}$ and entry (j_1, j_2) of $\boldsymbol{\rho}$, respectively, for $j, j_1, j_2 = 1, ..., p$. As described in the proof of Theorem 3, from Algorithm 1, the values z_j, $j = 1, ..., p$, in the subdata consist of $z_{(i)j}$, and $z_j^{(i)l}$, $l = 1, ...j-1, j+1, ..., p$, $i = 1, ..., r$, $n-r+1, ..., n$, where $z_j^{(i)l}$ are the concomitants for z_j.

Let $\mathbf{v} = (\mathbf{Z}_{\mathrm{D}}^*)^{\mathrm{T}}\mathbf{1}$ and $\mathbf{\Omega} = (\mathbf{Z}_{\mathrm{D}}^*)^{\mathrm{T}}\mathbf{Z}_{\mathrm{D}}^*$. Then

$$(\mathbf{X}_{\mathrm{D}}^*)^{\mathrm{T}}\mathbf{X}_{\mathrm{D}}^* = \begin{bmatrix} k & \mathbf{v}^{\mathrm{T}} \\ \mathbf{v} & \mathbf{\Omega} \end{bmatrix}. \tag{19}$$

The jth diagonal element of $\mathbf{\Omega}$ is

$$\Omega_{jj} = \left(\sum_{i=1}^{r} + \sum_{i=n-r+1}^{n}\right) z_{(i)j}^2 + \sum_{l \neq j}\left(\sum_{i=1}^{r} + \sum_{i=n-r+1}^{n}\right)\left(z_j^{(i)l}\right)^2, \tag{20}$$

while entry (j_1, j_2), $j_1 \neq j_2$, is

$$\Omega_{j_1 j_2} = \left(\sum_{i=1}^{r} + \sum_{i=n-r+1}^{n}\right)\left(z_{(i)j_1} z_{j_2}^{(i)j_1} + z_{(i)j_2} z_{j_1}^{(i)j_2}\right) + \sum_{l \neq j_1 j_2}\left(\sum_{i=1}^{r} + \sum_{i=n-r+1}^{n}\right) z_{j_1}^{(i)l} z_{j_2}^{(i)l}. \tag{21}$$

The jth element of $\mathbf{v}$ is

$$v_j = \left(\sum_{i=1}^{r} + \sum_{i=n-r+1}^{n}\right) z_{(i)j} + \sum_{l \neq j}\left(\sum_{i=1}^{r} + \sum_{i=n-r+1}^{n}\right) z_j^{(i)l}. \tag{22}$$

Now we consider the two specific distributions in Theorem 6 and prove the corresponding results in (26) and (27).

A.6.1 Proof of equation (26) in Theorem 6

Proof. When $\mathbf{z}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, using the results in Example 2.8.1 of Galambos (1987), we obtain

$$\begin{aligned} z_{(i)j} &= \mu_j - \sigma_j \sqrt{2\log n} + o_P(1), \quad i = 1, ..., r, \\ z_{(i)j} &= \mu_j + \sigma_j \sqrt{2\log n} + o_P(1), \quad i = n-r+1, ..., n. \end{aligned} \tag{23}$$

Using an approach similar to Example 5.5.1 of Galambos (1987), we obtain

$$\begin{aligned} z_j^{(i)l} &= \mu_j - \rho_{lj}\sigma_j \sqrt{2\log n} + O_P(1), \quad i = 1, ..., r, \\ z_j^{(i)l} &= \mu_j + \rho_{lj}\sigma_j \sqrt{2\log n} + O_P(1), \quad i = n-r+1, ..., n. \end{aligned} \tag{24}$$

Using (23) and (24), from (20), (21) and (22), we obtain that

$$\Omega_{jj} = 4r\log n \sigma_j^2 \sum_{l=1}^{p} \rho_{lj}^2 + O_P(\sqrt{\log n}), \tag{25}$$

$$\Omega_{j_1 j_2} = 4r\log n \sigma_{j_1}\sigma_{j_2} \sum_{l=1}^{p} \rho_{lj_1}\rho_{lj_2} + O_P(\sqrt{\log n}) \tag{26}$$

$$v_j = O_P(1), \tag{27}$$

respectively. From (25), (26) and (27), we have

$$\boldsymbol{\Omega} = 4r\log n \boldsymbol{\Phi}\boldsymbol{\rho}^2\boldsymbol{\Phi} + O_P(\sqrt{\log n}) \quad \text{and} \quad \mathbf{v} = O_P(1). \tag{28}$$

The variance,

$$\mathrm{V}(\hat{\boldsymbol{\beta}}^{\mathrm{D}}|\mathbf{X}) = \sigma^2 \begin{bmatrix} k & \mathbf{v}^{\mathrm{T}} \\ \mathbf{v} & \boldsymbol{\Omega} \end{bmatrix}^{-1} = \frac{\sigma^2}{c} \begin{bmatrix} 1 & -\mathbf{v}^{\mathrm{T}}\boldsymbol{\Omega}^{-1} \\ -\boldsymbol{\Omega}^{-1}\mathbf{v} & c\boldsymbol{\Omega}^{-1} + \boldsymbol{\Omega}^{-1}\mathbf{v}\mathbf{v}^{\mathrm{T}}\boldsymbol{\Omega}^{-1} \end{bmatrix}, \tag{29}$$

where $c = k - \mathbf{v}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{v} = k + O_P(1/\log n)$ and the second equality is from (28). Note that from (28) $\boldsymbol{\Omega}^{-1} = O_P(1/\log n)$, so

$$\begin{aligned} \boldsymbol{\Omega}^{-1} - (4r\log n \boldsymbol{\Phi}\boldsymbol{\rho}^2\boldsymbol{\Phi})^{-1} &= \boldsymbol{\Omega}^{-1}(4r\log n \boldsymbol{\Phi}\boldsymbol{\rho}^2\boldsymbol{\Phi} - \boldsymbol{\Omega})(4r\log n \boldsymbol{\Phi}\boldsymbol{\rho}^2\boldsymbol{\Phi})^{-1} \\ &= O_P\left(\frac{1}{\log n}\right) O_P(\sqrt{\log n}) O\left(\frac{1}{\log n}\right) = O_P\left\{\frac{1}{(\log n)^{3/2}}\right\}. \end{aligned}$$

Thus

$$\boldsymbol{\Omega}^{-1} = \frac{1}{4r\log n}(\boldsymbol{\Phi}\boldsymbol{\rho}^2\boldsymbol{\Phi})^{-1} + O_P\left\{\frac{1}{(\log n)^{3/2}}\right\}. \tag{30}$$

Combining (19), (29) and (30), and using that $k = 2rp$

$$\mathrm{V}(\hat{\boldsymbol{\beta}}^{\mathrm{D}}|\mathbf{X}) = \sigma^2 \begin{bmatrix} \frac{1}{k} + O_P\left(\frac{1}{\log n}\right) & O_P\left(\frac{1}{\log n}\right) \\ O_P\left(\frac{1}{\log n}\right) & \frac{1}{4r\log n}(\boldsymbol{\Phi}\boldsymbol{\rho}^2\boldsymbol{\Phi})^{-1} + O_P\left\{\frac{1}{(\log n)^{3/2}}\right\} \end{bmatrix}.$$

□

A.6.2 Proof of equation (27) in Theorem 6

Proof. When $\mathbf{z}_i \sim LN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $z_{ij} = \exp(U_{ij})$ with $\mathbf{U}_i = (U_{i1}, ..., U_{ip})^{\mathrm{T}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. From (23),

$$\begin{aligned} z_{(i)j} &= \exp(U_{(i)j}) = \exp(-\sigma_j\sqrt{2\log n})O_P(1) = o_P(1), \quad i = 1, ..., r, \\ z_{(i)j} &= \exp(U_{(i)j}) = \exp(\sigma_j\sqrt{2\log n})\{e^{\mu_j} + o_P(1)\}, \quad i = n-r+1, ..., n. \end{aligned} \tag{31}$$

Without loss of generality, assume that $\rho_{lj} \geq 0$, $l, j = 1, ..., p$. From (24),

$$\begin{aligned} z_j^{(i)l} &= \exp(U_j^{(i)l}) = \exp(-\rho_{lj}\sigma_j\sqrt{2\log n})O_P(1) = o_P(1), \quad i = 1, ..., r, \\ z_j^{(i)l} &= \exp(U_j^{(i)l}) = \exp\{\sigma_j\sqrt{2\log n} - (1-\rho_{lj})\sigma_j\sqrt{2\log n} + \mu_j + O_P(1)\} \\ &= \exp(\sigma_j\sqrt{2\log n})o_P(1), \quad i = n-r+1, ..., n. \end{aligned} \tag{32}$$

Using (31) and (32), from (20), (21) and (22), we obtain that

$$\Omega_{jj} = r\exp(2\sigma_j\sqrt{2\log n})\{e^{2\mu_j} + o_P(1)\}, \tag{33}$$

$$\Omega_{j_1j_2} = 2r\exp\left\{(\sigma_{j_1} + \sigma_{j_2})\sqrt{2\log n}\right\}o_P(1), \tag{34}$$

$$v_j = r\exp(\sigma_j\sqrt{2\log n})\{e^{\mu_j} + o_P(1)\}. \tag{35}$$

From (19), (33)-(35), for $\mathbf{A}_n = \mathrm{diag}\left\{1, \exp\left(\sigma_1\sqrt{2\log n}\right), ..., \exp\left(\sigma_p\sqrt{2\log n}\right)\right\}$,

$$\mathbf{A}_n^{-1}(\mathbf{X}_{\mathrm{D}}^*)^{\mathrm{T}}\mathbf{X}_{\mathrm{D}}^*\mathbf{A}_n^{-1} = \mathbf{A}_n^{-1}\begin{bmatrix} k & \mathbf{v}^{\mathrm{T}} \\ \mathbf{v} & \boldsymbol{\Omega} \end{bmatrix}\mathbf{A}_n^{-1} = \begin{bmatrix} k & r\mathbf{v}_1^{\mathrm{T}} \\ r\mathbf{v}_1 & r\mathbf{B}_5, \end{bmatrix} + o_P(1) \tag{36}$$

where $\mathbf{v}_1 = (e^{\mu_1}, ..., e^{\mu_p})^{\mathrm{T}}$ and $\mathbf{B}_5 = \mathrm{diag}(e^{2\mu_1}, ..., e^{2\mu_p})$. From (36),

$$\begin{aligned} \mathrm{V}(\mathbf{A}_n\hat{\boldsymbol{\beta}}^{\mathrm{D}}|\mathbf{X}) = \sigma^2\mathbf{A}_n\{(\mathbf{X}_{\mathrm{D}}^*)^{\mathrm{T}}\mathbf{X}_{\mathrm{D}}^*\}^{-1}\mathbf{A}_n &= \sigma^2\begin{bmatrix} k & r\mathbf{v}_1^{\mathrm{T}} \\ r\mathbf{v}_1 & r\mathbf{B}_5, \end{bmatrix}^{-1} + o_P(1) \\ &= \frac{2\sigma^2}{k}\begin{bmatrix} 1 & -\mathbf{u}^{\mathrm{T}} \\ -\mathbf{u} & p\boldsymbol{\Lambda} + \mathbf{u}\mathbf{u}^{\mathrm{T}}, \end{bmatrix} + o_P(1). \end{aligned}$$

□

A.7 Proof of results in Table 1

When the covariate has a t distribution, from Theorem 4, for simple linear model, the variance of the estimator of β_1 using the D-OPT IBOSS approach is of the same order as $(z_{(n)1} - z_{(1)1})^{-2}$. From Theorems 2.1.2 and 2.9.2 of Galambos (1987), we obtain that $z_{(n)1} - z_{(1)1} \asymp_P n^{1/\nu}$. Thus, the variance is of the order $n^{-2/\nu}$.

For the full data approach, the variance of the estimator of β_1 is of the same order as $(\sum_{i=1}^n z_{i1}^2)^{-1}$. When z_1 has a t distribution with degrees of freedom $\nu > 2$, from Kolmogorov's strong law of large numbers (SLLN), $\sum_{i=1}^n z_{i1}^2 = O(n)$ almost surely. If $\nu \le 2$, $\mathrm{E}[\{z_{i1}^2\}^{1/(2/\nu+\alpha)}] < \infty$ for any $\alpha > 0$. Thus, from Marcinkiewicz-Zygmund SLLN (Theorem 2 of Section 5.2 of Chow and Teicher, 2003), $\sum_{i=1}^n z_{ij}^2 = o(n^{2/\nu+\alpha})$ almost surely for any $\alpha > 0$. This shows that the order of $(\sum_{i=1}^n z_{i1}^2)^{-1}$ is slower than $n^{-(2/\nu+\alpha)}$ for any $\alpha > 0$.

For the UNI approach, the lower bound for the variance of the estimator of β_1 is of the same order as $n(\sum_{i=1}^n z_{i1}^2)^{-1}$, which is of order $O(1)$ when $\nu > 2$ and is slower than $n^{2/\nu-1+\alpha}$ for any $\alpha > 0$ when $\nu \le 2$.

For the intercept β_0, the variance of the estimator is of the same order as the inverse of the sample size used in each method.

References

Chow, Y. S. C. and Teicher, H. (2003). *Probability Theory: Independence, Interchangeability, Martingales*. Springer, New York.

Galambos, J. (1987). *The asymptotic theory of extreme order statistics*. Florida: Robert E. Krieger.

Hall, P. (1979). On the relative stability of large order statistics. In *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 86, 467–475. Cambridge Univ Press.

Nordström, K. (2011). Convexity of the inverse and Moore–Penrose inverse. *Linear Algebra and its Applications* **434**, 6, 1489–1512.