Supplementary Materials for "Integrating External Summary Information via James-Stein Shrinkage

Peisong Han, Haoyue Li, Jeremy M.G. Taylor

Biostatistics Innovation Group, Gilead Sciences

Department of Statistics, Pennsylvania State University

Department of Biostatistics, University of Michigan

1. Useful Facts about the Two Constrained MLEs

Using the Lagrange multipliers method, it is easy to show that the constrained MLE $\hat{\boldsymbol{\beta}}_{cmle-sp}$ defined in (2.2) is the corresponding component of $(\hat{\boldsymbol{\beta}}_{cmle-sp}, \hat{\boldsymbol{\gamma}}_{cmle-sp}, \hat{\boldsymbol{\rho}})$ that satisfies

$$\sum_{i=1}^{n} \boldsymbol{S}_i(\hat{\boldsymbol{\beta}}_{cmle-sp}, \hat{\boldsymbol{\gamma}}_{cmle-sp}) + \sum_{i=1}^{n} \frac{\partial \boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-sp}, \hat{\boldsymbol{\gamma}}_{cmle-sp}, \boldsymbol{\theta}_*)}{\partial(\boldsymbol{\beta}, \boldsymbol{\gamma})^{\mathrm{T}}} \hat{\boldsymbol{\rho}} = \mathbf{0}, \quad (1.1)$$

$$\sum_{i=1}^{n} \boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-sp}, \hat{\boldsymbol{\gamma}}_{cmle-sp}, \boldsymbol{\theta}_*) = \mathbf{0}, \quad (1.2)$$

where $\hat{\boldsymbol{\rho}}$ is the Lagrange multiplier. Using the Z-estimator theory (e.g., van der Vaart 1998), we have $(\hat{\boldsymbol{\beta}}_{cmle-sp}, \hat{\boldsymbol{\gamma}}_{cmle-sp}, \hat{\boldsymbol{\rho}}) \xrightarrow{p} (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathbf{0})$ as $n \to \infty$.

Using the Lagrange multipliers method again, below we show that the constrained MLE $\hat{\boldsymbol{\beta}}_{cmle-el}$ defined in (2.3) is the corresponding component of $(\hat{\boldsymbol{\beta}}_{cmle-el}, \hat{\boldsymbol{\gamma}}_{cmle-el}, \hat{\boldsymbol{\rho}})$ that satisfies

$$\sum_{i=1}^{n} \boldsymbol{S}_i(\hat{\boldsymbol{\beta}}_{cmle-el}, \hat{\boldsymbol{\gamma}}_{cmle-el}) + \sum_{i=1}^{n} \frac{\partial \boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-el}, \hat{\boldsymbol{\gamma}}_{cmle-el}, \boldsymbol{\theta}_*)/\partial(\boldsymbol{\beta}, \boldsymbol{\gamma})^{\mathrm{T}}}{1 - \hat{\boldsymbol{\rho}}^{\mathrm{T}} \boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-el}, \hat{\boldsymbol{\gamma}}_{cmle-el}, \boldsymbol{\theta}_*)} \hat{\boldsymbol{\rho}} = \mathbf{0}, \quad (1.3)$$

$$\sum_{i=1}^{n} \frac{\boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-el}, \hat{\boldsymbol{\gamma}}_{cmle-el}, \boldsymbol{\theta}_*)}{1 - \hat{\boldsymbol{\rho}}^{\mathrm{T}} \boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-el}, \hat{\boldsymbol{\gamma}}_{cmle-el}, \boldsymbol{\theta}_*)} = \mathbf{0}, \quad (1.4)$$

where we still use $\hat{\boldsymbol{\rho}}$ to denote the Lagrange multiplier. Using the Z-estimator theory again we have $(\hat{\boldsymbol{\beta}}_{cmle-el}, \hat{\boldsymbol{\gamma}}_{cmle-el}, \hat{\boldsymbol{\rho}}) \xrightarrow{p} (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathbf{0})$ as $n \to \infty$.

2. Derivation of (1.3) and (1.4)

The Lagrangian corresponding to (2.3) is

$$\mathcal{L} = \sum_{i=1}^{n} \log f_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) + \sum_{i=1}^{n} \log q_i + n\boldsymbol{\rho}^{\mathrm{T}} \sum_{i=1}^{n} q_i \boldsymbol{U}_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}_*) - \mu(\sum_{i=1}^{n} q_i - 1),$$

where $\boldsymbol{\rho}$ and μ are the Lagrange multipliers. At the solution $(\hat{\boldsymbol{\beta}}_{cmle-el}, \hat{\boldsymbol{\gamma}}_{cmle-el})$ and $\hat{q}_i$ we must have $\partial\mathcal{L}/\partial q_i = 0$ and $\partial\mathcal{L}/\partial(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbf{0}$ for some $\hat{\boldsymbol{\rho}}$ and $\hat{\mu}$. Multiplying both sides of $\partial\mathcal{L}/\partial q_i = 1/\hat{q}_i + n\hat{\boldsymbol{\rho}}^{\mathrm{T}} \boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-el}, \hat{\boldsymbol{\gamma}}_{cmle-el}, \boldsymbol{\theta}_*) - \hat{\mu} = 0$ by $\hat{q}_i$ and summing over i, the constraints in (2.3) lead to $\hat{\mu} = n$, which, combined with $\partial\mathcal{L}/\partial q_i = 0$ yields $\hat{q}_i = 1/[n\{1 - \hat{\boldsymbol{\rho}}^{\mathrm{T}} \boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-el}, \hat{\boldsymbol{\gamma}}_{cmle-el}, \boldsymbol{\theta}_*)\}]$. Then $\partial\mathcal{L}/\partial(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbf{0}$ gives (1.3) and the constraint $\sum_{i=1}^{n} \hat{q}_i \boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-el}, \hat{\boldsymbol{\gamma}}_{cmle-el}, \boldsymbol{\theta}_*) = \mathbf{0}$ gives (1.4).

3. Proof of Theorem 1

Let "$\xrightarrow{d_n}$" denote convergence in distribution as $n \to \infty$ along the sequence of distributions corresponding to the sequence of external study population distributions that are local to the internal study population distribution.

Lemma 1. For any estimator $\hat{\boldsymbol{\beta}}$ satisfying $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d_n} \boldsymbol{\psi}$ for some random variable $\boldsymbol{\psi}$ as $n \to \infty$, and for the weighted quadratic loss $l(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0) = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{\mathrm{T}} \boldsymbol{V}^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ with the weight $\boldsymbol{V}^{-1}$, the asymptotic risk for $\hat{\boldsymbol{\beta}}$ is $R(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0) = E(\boldsymbol{\psi}^{\mathrm{T}} \boldsymbol{V}^{-1} \boldsymbol{\psi})$.

Lemma 1 is Lemma 1 in Hansen (2016) and the proof is omitted.

Proof of (i). For the MLE $(\hat{\boldsymbol{\beta}}_{mle}, \hat{\boldsymbol{\gamma}}_{mle})$ we have

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\beta}}_{mle} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\gamma}}_{mle} - \boldsymbol{\gamma}_0 \end{pmatrix} = \boldsymbol{\Omega}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{S}_i(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0) + o_p(1) \xrightarrow{d_n} \boldsymbol{\Omega}^{-1} \boldsymbol{\Delta}_{\boldsymbol{S}}, \quad (3.5)$$

and thus $\sqrt{n}(\hat{\boldsymbol{\beta}}_{mle} - \boldsymbol{\beta}_0) \xrightarrow{d_n} \boldsymbol{P} \boldsymbol{\Omega}^{-1} \boldsymbol{\Delta}_{\boldsymbol{S}}$. The result then follows from Lemma 1 with $\boldsymbol{V} = \boldsymbol{V}_{\boldsymbol{\beta}, mle}$.

Proof of (ii). We will first show the result for $\hat{\boldsymbol{\beta}}_{JS}$ based on $\hat{\boldsymbol{\beta}}_{cmle-el}$. Applying the mean-value theorem to (1.3) and (1.4) around $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathbf{0})$ leads to

$$
\begin{aligned}
& \mathbf{0} \\
= \quad & \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \boldsymbol{S}_i(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0) \\ \boldsymbol{U}_i(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta}_{n*}) \end{pmatrix} \\
& +\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \frac{\partial \boldsymbol{S}_i(\bar{\boldsymbol{\beta}},\bar{\boldsymbol{\gamma}})}{\partial(\boldsymbol{\beta},\boldsymbol{\gamma})}, & \frac{\partial \boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-el},\hat{\boldsymbol{\gamma}}_{cmle-el},\boldsymbol{\theta}_{n*})/\partial(\boldsymbol{\beta},\boldsymbol{\gamma})^{\mathrm{T}}}{1-\hat{\boldsymbol{\rho}}^{\mathrm{T}}\boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-el},\hat{\boldsymbol{\gamma}}_{cmle-el},\boldsymbol{\theta}_{n*})} \\ \frac{\partial \boldsymbol{U}_i(\bar{\boldsymbol{\beta}},\bar{\boldsymbol{\gamma}},\boldsymbol{\theta}_{n*})/\partial(\boldsymbol{\beta},\boldsymbol{\gamma})}{1-\bar{\boldsymbol{\rho}}^{\mathrm{T}}\boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-el},\hat{\boldsymbol{\gamma}}_{cmle-el},\boldsymbol{\theta}_{n*})}, & \frac{\boldsymbol{U}_i(\bar{\boldsymbol{\beta}},\bar{\boldsymbol{\gamma}},\boldsymbol{\theta}_{n*})\boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-el},\hat{\boldsymbol{\gamma}}_{cmle-el},\boldsymbol{\theta}_{n*})^{\mathrm{T}}}{\{1-\bar{\boldsymbol{\rho}}^{\mathrm{T}}\boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-el},\hat{\boldsymbol{\gamma}}_{cmle-el},\boldsymbol{\theta}_{n*})\}^2} \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{cmle-el}-\boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\gamma}}_{cmle-el}-\boldsymbol{\gamma}_0 \\ \hat{\boldsymbol{\rho}} \end{pmatrix},
\end{aligned}
$$

where $\bar{\boldsymbol{\beta}}$ is some value between $\hat{\boldsymbol{\beta}}_{cmle-el}$ and $\boldsymbol{\beta}_0$, $\bar{\boldsymbol{\gamma}}$ is some value between $\hat{\boldsymbol{\gamma}}_{cmle-el}$ and $\boldsymbol{\gamma}_0$, and $\bar{\boldsymbol{\rho}}$ is some value between $\hat{\boldsymbol{\rho}}$ and $\mathbf{0}$. Then we have

$$
\begin{aligned}
& \sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{cmle-el}-\boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\gamma}}_{cmle-el}-\boldsymbol{\gamma}_0 \\ \hat{\boldsymbol{\rho}} \end{pmatrix} \\
= \quad & -\begin{pmatrix} -\boldsymbol{\Omega}, & \boldsymbol{G}^{\mathrm{T}} \\ \boldsymbol{G}, & \boldsymbol{\Sigma} \end{pmatrix}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \boldsymbol{S}_i(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0) \\ \boldsymbol{U}_i(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta}_{n*}) \end{pmatrix}+o_p(1) \\
= \quad & -\begin{pmatrix} -\boldsymbol{\Omega}, & \boldsymbol{G}^{\mathrm{T}} \\ \boldsymbol{G}, & \boldsymbol{\Sigma} \end{pmatrix}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \boldsymbol{S}_i(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0) \\ \boldsymbol{U}_i(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta}_{n*})-E\{\boldsymbol{U}(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta}_{n*})\} \end{pmatrix}+\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\delta} \end{pmatrix}\right\}+o_p(1),
\end{aligned}
$$

which leads to

$$
\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{cmle-el}-\boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\gamma}}_{cmle-el}-\boldsymbol{\gamma}_0 \end{pmatrix}\xrightarrow{d_n}(\boldsymbol{\Omega}^{-1}-\boldsymbol{L}\boldsymbol{G}\boldsymbol{\Omega}^{-1},-\boldsymbol{L})\boldsymbol{\Delta}-\boldsymbol{L}\boldsymbol{\delta}=\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}_{\boldsymbol{S}}-\boldsymbol{L}(\boldsymbol{\Delta}_*+\boldsymbol{\delta}).
$$

This, together with (3.5), implies that $\sqrt{n}(\hat{\boldsymbol{\beta}}_{mle}-\hat{\boldsymbol{\beta}}_{cmle-el})\xrightarrow{d_n}\boldsymbol{P}\boldsymbol{L}(\boldsymbol{\Delta}_*+\boldsymbol{\delta})$, which then leads to

$$n(\hat{\boldsymbol{\beta}}_{mle} - \hat{\boldsymbol{\beta}}_{cmle-el})^{\mathrm{T}} \hat{\boldsymbol{V}}^{-1}_{\boldsymbol{\beta},mle} (\hat{\boldsymbol{\beta}}_{mle} - \hat{\boldsymbol{\beta}}_{cmle-el}) \xrightarrow{d_n} \xi = (\boldsymbol{\Delta}_* + \boldsymbol{\delta})^{\mathrm{T}} \boldsymbol{B} (\boldsymbol{\Delta}_* + \boldsymbol{\delta}).$$

Therefore, we have $\hat{w} \xrightarrow{d_n} w = (1 - \tau/\xi)_+$, and thus

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{JS} - \boldsymbol{\beta}_0) \xrightarrow{d_n} \boldsymbol{\psi}_{JS} = w \boldsymbol{P}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}_{\boldsymbol{S}} + (1-w)\boldsymbol{P}\{\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}_{\boldsymbol{S}} - \boldsymbol{L}(\boldsymbol{\Delta}_* + \boldsymbol{\delta})\}.$$

From Lemma 1, the asymptotic risk for $\hat{\boldsymbol{\beta}}_{JS}$ based on $\hat{\boldsymbol{\beta}}_{cmle-el}$ is $R(\hat{\boldsymbol{\beta}}_{JS}, \boldsymbol{\beta}_0) = E(\boldsymbol{\psi}_{JS}^{\mathrm{T}} \boldsymbol{V}^{-1}_{\boldsymbol{\beta},mle} \boldsymbol{\psi}_{JS})$.

Define the random variable $\boldsymbol{\psi}^*_{JS}$ that is analogous to $\boldsymbol{\psi}_{JS}$ without the positive part trimming in the weight w:

$$\begin{aligned} \boldsymbol{\psi}^*_{JS} &= (1 - \tau/\xi)\boldsymbol{P}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}_{\boldsymbol{S}} + \tau/\xi \boldsymbol{P}\{\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}_{\boldsymbol{S}} - \boldsymbol{L}(\boldsymbol{\Delta}_* + \boldsymbol{\delta})\} \\ &= \boldsymbol{P}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}_{\boldsymbol{S}} - \tau/\xi \boldsymbol{P}\boldsymbol{L}(\boldsymbol{\Delta}_* + \boldsymbol{\delta}). \end{aligned}$$

Then from Lemma 2 of Hansen (2015), we have

$$R(\hat{\boldsymbol{\beta}}_{JS}, \boldsymbol{\beta}_0) = E(\boldsymbol{\psi}_{JS}^{\mathrm{T}} \boldsymbol{V}^{-1}_{\boldsymbol{\beta},mle} \boldsymbol{\psi}_{JS}) < E(\boldsymbol{\psi}_{JS}^{*\mathrm{T}} \boldsymbol{V}^{-1}_{\boldsymbol{\beta},mle} \boldsymbol{\psi}^*_{JS}). \tag{3.6}$$

It is easy to see that

$$\begin{aligned} E(\boldsymbol{\psi}_{JS}^{*\mathrm{T}} \boldsymbol{V}^{-1}_{\boldsymbol{\beta},mle} \boldsymbol{\psi}^*_{JS}) &= R(\hat{\boldsymbol{\beta}}_{mle}, \boldsymbol{\beta}_0) + \tau^2 E\left\{ \frac{(\boldsymbol{\Delta}_* + \boldsymbol{\delta})^{\mathrm{T}} \boldsymbol{L}^{\mathrm{T}} \boldsymbol{P}^{\mathrm{T}} \boldsymbol{V}^{-1}_{\boldsymbol{\beta},mle} \boldsymbol{P}\boldsymbol{L}(\boldsymbol{\Delta}_* + \boldsymbol{\delta})}{\xi^2} \right\} \\ &\quad -2\tau E\left\{ \frac{(\boldsymbol{\Delta}_* + \boldsymbol{\delta})^{\mathrm{T}} \boldsymbol{L}^{\mathrm{T}} \boldsymbol{P}^{\mathrm{T}} \boldsymbol{V}^{-1}_{\boldsymbol{\beta},mle} \boldsymbol{P}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}_{\boldsymbol{S}}}{\xi} \right\} \\ &= R(\hat{\boldsymbol{\beta}}_{mle}, \boldsymbol{\beta}_0) + \tau^2 E(1/\xi) - 2\tau E\{\boldsymbol{g}(\boldsymbol{D}\boldsymbol{\Delta} + \boldsymbol{\delta})^{\mathrm{T}} \boldsymbol{K}\boldsymbol{\Delta}\}, \end{aligned} \tag{3.7}$$

where $\boldsymbol{D} = (\boldsymbol{G}\boldsymbol{\Omega}^{-1}, \boldsymbol{I})$ and thus $\boldsymbol{\Delta}_* = \boldsymbol{D}\boldsymbol{\Delta}$, $\boldsymbol{K} = \boldsymbol{L}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{V}_{\boldsymbol{\beta},mle}^{-1}\boldsymbol{P}(\boldsymbol{\Omega}^{-1}, \boldsymbol{0})$, and

$$\boldsymbol{g}(\boldsymbol{D}\boldsymbol{\Delta} + \boldsymbol{\delta}) = \frac{(\boldsymbol{D}\boldsymbol{\Delta} + \boldsymbol{\delta})}{(\boldsymbol{D}\boldsymbol{\Delta} + \boldsymbol{\delta})^{\mathrm{T}}\boldsymbol{B}(\boldsymbol{D}\boldsymbol{\Delta} + \boldsymbol{\delta})}.$$

Let $\phi_{\boldsymbol{\Delta}}(\boldsymbol{x})$ denote the density for $\boldsymbol{\Delta}$, which is multivariate normal with mean $\boldsymbol{0}$ and variance $\boldsymbol{V}_{\boldsymbol{\Delta}}$. Then

$$\begin{aligned}
& E\{\boldsymbol{g}(\boldsymbol{D}\boldsymbol{\Delta} + \boldsymbol{\delta})^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\Delta}\} \\
= & \int_{-\infty}^{\infty} \boldsymbol{g}(\boldsymbol{D}\boldsymbol{x} + \boldsymbol{\delta})^{\mathrm{T}}\boldsymbol{K}\boldsymbol{x}\phi_{\boldsymbol{\Delta}}(\boldsymbol{x})d\boldsymbol{x} = -\int_{-\infty}^{\infty} \boldsymbol{g}(\boldsymbol{D}\boldsymbol{x} + \boldsymbol{\delta})^{\mathrm{T}}\boldsymbol{K}\boldsymbol{V}_{\boldsymbol{\Delta}}d\phi_{\boldsymbol{\Delta}}(\boldsymbol{x}) \\
= & \int_{-\infty}^{\infty} \mathrm{tr}\left\{\frac{d}{d\boldsymbol{x}}\boldsymbol{g}(\boldsymbol{D}\boldsymbol{x} + \boldsymbol{\delta})^{\mathrm{T}}\boldsymbol{K}\boldsymbol{V}_{\boldsymbol{\Delta}}\right\}\phi_{\boldsymbol{\Delta}}(\boldsymbol{x})d\boldsymbol{x} \\
= & \int_{-\infty}^{\infty} \mathrm{tr}\left\{\frac{\boldsymbol{D}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{V}_{\boldsymbol{\Delta}}}{(\boldsymbol{D}\boldsymbol{x} + \boldsymbol{\delta})^{\mathrm{T}}\boldsymbol{B}(\boldsymbol{D}\boldsymbol{x} + \boldsymbol{\delta})} - \frac{2\boldsymbol{D}^{\mathrm{T}}\boldsymbol{B}(\boldsymbol{D}\boldsymbol{x} + \boldsymbol{\delta})(\boldsymbol{D}\boldsymbol{x} + \boldsymbol{\delta})^{\mathrm{T}}\boldsymbol{K}\boldsymbol{V}_{\boldsymbol{\Delta}}}{\{(\boldsymbol{D}\boldsymbol{x} + \boldsymbol{\delta})^{\mathrm{T}}\boldsymbol{B}(\boldsymbol{D}\boldsymbol{x} + \boldsymbol{\delta})\}^2}\right\}\phi_{\boldsymbol{\Delta}}(\boldsymbol{x})d\boldsymbol{x} \\
= & E\mathrm{tr}\left(\frac{\boldsymbol{D}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{V}_{\boldsymbol{\Delta}}}{\xi}\right) - 2E\mathrm{tr}\left(\frac{\boldsymbol{D}^{\mathrm{T}}\boldsymbol{B}(\boldsymbol{D}\boldsymbol{\Delta} + \boldsymbol{\delta})(\boldsymbol{D}\boldsymbol{\Delta} + \boldsymbol{\delta})^{\mathrm{T}}\boldsymbol{K}\boldsymbol{V}_{\boldsymbol{\Delta}}}{\xi^2}\right) \\
= & E\left(\frac{\mathrm{tr}(\boldsymbol{J}_2)}{\xi}\right) - 2E\left(\frac{(\boldsymbol{D}\boldsymbol{\Delta} + \boldsymbol{\delta})^{\mathrm{T}}\boldsymbol{K}\boldsymbol{V}_{\boldsymbol{\Delta}}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{B}(\boldsymbol{D}\boldsymbol{\Delta} + \boldsymbol{\delta})}{\xi^2}\right) \\
= & E\left(\frac{\mathrm{tr}(\boldsymbol{J}_2)}{\xi}\right) - 2E\left(\frac{(\boldsymbol{D}\boldsymbol{\Delta} + \boldsymbol{\delta})^{\mathrm{T}}\boldsymbol{B}_1^{\mathrm{T}}\boldsymbol{J}_2\boldsymbol{B}_1(\boldsymbol{D}\boldsymbol{\Delta} + \boldsymbol{\delta})}{\xi^2}\right) \\
\geq & E\left(\frac{\mathrm{tr}(\boldsymbol{J}_2) - 2 \parallel \boldsymbol{J}_2 \parallel}{\xi}\right)
\end{aligned} \tag{3.8}$$

where $\boldsymbol{B}_1 = \boldsymbol{V}_{\boldsymbol{\beta},mle}^{-1/2}\boldsymbol{P}\boldsymbol{L}$. In the above display the third equality follows from integration by parts, the sixth and seventh equalities follow some matrix algebra, and the last inequality follows the fact that $\boldsymbol{B}_1^{\mathrm{T}}\boldsymbol{B}_1 = \boldsymbol{B}$.

From (3.6), (3.7) and (3.8) we have

$$\begin{aligned} R(\hat{\boldsymbol{\beta}}_{JS},\boldsymbol{\beta}_0) &< R(\hat{\boldsymbol{\beta}}_{mle},\boldsymbol{\beta}_0) - \tau E\left(\frac{2\{\mathrm{tr}(\boldsymbol{J}_2) - 2 \parallel \boldsymbol{J}_2 \parallel\} - \tau}{\xi}\right) \\ &\leq R(\hat{\boldsymbol{\beta}}_{mle},\boldsymbol{\beta}_0) - \tau \frac{2\{\mathrm{tr}(\boldsymbol{J}_2) - 2 \parallel \boldsymbol{J}_2 \parallel\} - \tau}{E\xi}, \end{aligned}$$

where the second inequality follows from $d > 2$ and Jensen's inequality. This is the desired result for $\hat{\boldsymbol{\beta}}_{JS}$ based on $\hat{\boldsymbol{\beta}}_{cmle-el}$.

Now we show the result for $\hat{\boldsymbol{\beta}}_{JS}$ based on $\hat{\boldsymbol{\beta}}_{cmle-sp}$. Applying the mean-value theorem to (1.1) and (1.2) around $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \mathbf{0})$ leads to

$$\begin{aligned} \mathbf{0} &= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \boldsymbol{S}_i(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0) \\ \boldsymbol{U}_i(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta}_{n*}) \end{pmatrix} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \frac{\partial \boldsymbol{S}_i(\bar{\boldsymbol{\beta}},\bar{\boldsymbol{\gamma}})}{\partial(\boldsymbol{\beta},\boldsymbol{\gamma})}, & \frac{\partial \boldsymbol{U}_i(\hat{\boldsymbol{\beta}}_{cmle-sp},\hat{\boldsymbol{\gamma}}_{cmle-sp},\boldsymbol{\theta}_{n*})}{\partial(\boldsymbol{\beta},\boldsymbol{\gamma})^{\mathrm{T}}} \\ \frac{\partial \boldsymbol{U}_i(\bar{\boldsymbol{\beta}},\bar{\boldsymbol{\gamma}},\boldsymbol{\theta}_{n*})}{\partial(\boldsymbol{\beta},\boldsymbol{\gamma})}, & \mathbf{0} \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{cmle-sp} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\gamma}}_{cmle-sp} - \boldsymbol{\gamma}_0 \\ \hat{\boldsymbol{\rho}} \end{pmatrix}, \end{aligned}$$

where $\bar{\boldsymbol{\beta}}$ is some value between $\hat{\boldsymbol{\beta}}_{cmle-sp}$ and $\boldsymbol{\beta}_0$ and $\bar{\boldsymbol{\gamma}}$ is some value between $\hat{\boldsymbol{\gamma}}_{cmle-sp}$ and $\boldsymbol{\gamma}_0$. Then we have

$$\begin{aligned} &\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{cmle-sp} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\gamma}}_{cmle-sp} - \boldsymbol{\gamma}_0 \\ \hat{\boldsymbol{\rho}} \end{pmatrix} \\ &= -\begin{pmatrix} -\boldsymbol{\Omega}, & \boldsymbol{G}^{\mathrm{T}} \\ \boldsymbol{G}, & \mathbf{0} \end{pmatrix}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \boldsymbol{S}_i(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0) \\ \boldsymbol{U}_i(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta}_{n*}) - E\{\boldsymbol{U}(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\boldsymbol{\theta}_{n*})\} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\delta} \end{pmatrix}\right\} + o_p(1). \end{aligned}$$

The desired result then follows the same arguments as those for $\hat{\boldsymbol{\beta}}_{JS}$ based on $\hat{\boldsymbol{\beta}}_{cmle-el}$.

References

Hansen, B. (2015). Shrinkage efficiency bounds. *Econometric Theory 31*, 860–879.

Hansen, B. E. (2016). Efficient shrinkage in parametric models. *Journal of Econometrics 190*, 115–132.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.

Biostatistics Innovation Group, Gilead Sciences

E-mail: peisong.han2@gilead.com

Department of Statistics, Pennsylvania State University

E-mail: hpl5434@psu.edu

Department of Biostatistics, University of Michigan

E-mail: jmgt@umich.edu