Supplementary Material

A Notation

Let $\dot{\ell}_i(\boldsymbol{\beta}) \triangleq \dot{\ell}(y_i|\boldsymbol{x}_{\boldsymbol{i}}^\top \boldsymbol{\beta})$ and $\dot{\ell}(\boldsymbol{\beta}) \triangleq [\dot{\ell}(y_1|\boldsymbol{x}_{\boldsymbol{1}}^\top \boldsymbol{\beta}), \cdots, \dot{\ell}(y_n|\boldsymbol{x}_{\boldsymbol{n}}^\top \boldsymbol{\beta})]^\top$.

B Background material on Gaussian random variables, vectors and matrices

In this section, we review a few important results regarding the functions of Gaussian matrices and Gaussian vectors that are used in our examples. The first result is about the moments of the inverse of the minimum eigenvalue of a Wishart matrix.

Lemma 1. *(Lemma 19 of [Xu et al., 2019]) Let $X_{ij} \overset{i.i.d.}{\sim} N(0, \frac{1}{n})$, and suppose that $n, p \to \infty$ while $n/p = \delta_0$ for $\delta_0 > 1$. Then, for a fixed $r \geq 0$, we have*

$$\mathbb{E}\left[\frac{1}{\sigma^r_{\min}(\mathbf{X}^\top \mathbf{X})}\right] = O(1). \tag{10}$$

Our next two lemmas are concerned with the moments of a Gaussian and χ^2 random variables:

Lemma 2. *Let $Z \sim N(0, \sigma^2)$. Then, we have*

$$\mathbb{E}(|Z|^p) \leq \sigma^p (p-1)!!, \tag{11}$$

where the notation $p!!$ denotes the double factorial. Furthermore, when p is even the above inequality is in fact an equality.

The proof of this claim is straightforward and can be found in many standard statistics text books.

Lemma 3. *Let $Z \sim \chi^2_k$, i.e., it has a χ^2 distribution with k degrees of freedom. Then, for any integer $m \geq 1$ we have*

$$\mathbb{E}(Z^m) = k(k+2)(k+4)\dots(k+2m-2).$$

C Proof of Theorem 1

Define

$$V_1 \triangleq \frac{1}{n}\sum_{i=1}^n \phi(y_i, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\phi(y_i, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}],$$
$$V_2 \triangleq \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\phi(y_i, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}) | D_{/i}] - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}) \mid D].$$

Then,

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \phi(y_i, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}) - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}) \mid D]\right)^2$$
$$\leq \mathbb{E}\,(V_1 + V_2)^2 \leq \mathbb{E}\,V_1^2 + \mathbb{E}\,V_2^2 + 2\sqrt{\mathbb{E}\,V_1^2\,\mathbb{E}\,V_2^2}.$$

The proof concludes upon noting that Lemma 4 and 5 yield

$$\mathbb{E}\,V_1^2 \leq \frac{1}{n}\left(\mathbb{E}\,\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] + \left(\frac{c_0 c_1 \rho \delta^{1/2}}{\nu}\right)^2\right),$$
$$\mathbb{E}\,V_2^2 \leq \frac{1}{n}\left(\frac{c_0 c_1 \rho \delta^{1/2}}{\nu}\right)^2.$$

Lemma 4. *Under the assumptions of Theorem 1 we have that:*

$$\mathbb{E}\bigg(\frac{1}{n}\sum_{i=1}^{n}\phi(y_i, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}) - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\phi(y_i, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}]\bigg)^2 \leq \frac{1}{n}\bigg(\mathbb{E}\operatorname{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1})|D_{/1}] + \Big(\frac{c_0 c_1 \rho \delta^{1/2}}{\nu}\Big)^2\bigg).$$

Proof of Lemma 4.

$$\begin{aligned}\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\phi(y_i, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}) - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}]\right)^2 &= \frac{1}{n}\mathbb{E}\left(\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) - \mathbb{E}[\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]\right)^2 \\ &+ \frac{n-1}{n}\mathbb{E}\left[\left(\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) - \mathbb{E}[\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]\right)\left(\phi(y_2, \boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2}) - \mathbb{E}[\phi(y_2, \boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2}) \mid D_{/2}]\right)\right]\end{aligned}$$

Note that

$$\mathbb{E}\Big(\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) - \mathbb{E}[\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]\Big)^2 = \mathbb{E}\operatorname{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}].$$

Next we study

$$\mathbb{E}\Big[(\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) - \mathbb{E}[\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}])(\phi(y_2, \boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2}) - \mathbb{E}[\phi(y_2, \boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2}) \mid D_{/2}])\Big].$$

Recall $\hat{\boldsymbol{\beta}}_{/1,2} \triangleq \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\Big\{\sum_{k\geq 3}\ell(y_k \mid \boldsymbol{x}_k^\top \boldsymbol{\beta}) + \lambda r(\boldsymbol{\beta})\Big\}$. For some $t \in [0,1]$, the mean-value theorem yields

$$\begin{aligned}\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) &= \phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2}) + \dot{\phi}(y_1, t\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_1^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \\ \mathbb{E}[\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] &= \mathbb{E}[\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1}] + \mathbb{E}[\dot{\phi}(y_1, t\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_1^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1}] \\ &= \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1,2}] + \mathbb{E}[\dot{\phi}(y_o, t\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_o^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1}],\end{aligned}$$

where $(y_o, \boldsymbol{x}_o)$ is independent of D, leading to

$$\begin{aligned}\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) - \mathbb{E}[\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] &= \phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2}) - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1,2}] \\ &+ \dot{\phi}(y_1, t\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_1^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \\ &- \mathbb{E}[\dot{\phi}(y_o, t\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_o^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1}].\end{aligned}$$

Define the quantities A_0, B_0, C_0 as:

$$\begin{aligned}A_0 &\triangleq \mathbb{E}\left[(\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2}) - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1,2}])(\phi(y_2, \boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/1,2}) - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1,2}])\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\left(\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1,2}]\right)\left(\phi(\tilde{y}_o, \tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) - \mathbb{E}[\phi(\tilde{y}_o, \tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1,2}]\right) \,\Big|\, D_{/1,2}\right]\right] \\ &= 0.\end{aligned}$$

Likewise,

$$\begin{aligned}
B_0 \triangleq \mathbb{E}\Big\{ & \Big(\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2}) - \mathbb{E}\big[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/12}\big]\Big) \\
& \times \Big(\dot{\phi}(y_2, t\boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2} + (1-t)\boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_2^\top (\hat{\boldsymbol{\beta}}_{/2} - \hat{\boldsymbol{\beta}}_{/1,2}) \\
& - \mathbb{E}\Big[\dot{\phi}(y_2, t\boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2} + (1-t)\boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/12})\boldsymbol{x}_2^\top (\hat{\boldsymbol{\beta}}_{/2} - \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/2}\Big]\Big)\Big\} \\
= \mathbb{E}\Big[& \Big(\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2}) - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/12}]\Big) \\
& \times \Big(\dot{\phi}(\tilde{y}_o, t\tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/2} + (1-t)\tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/1,2})\tilde{\boldsymbol{x}}_o^\top (\hat{\boldsymbol{\beta}}_{/2} - \hat{\boldsymbol{\beta}}_{/1,2}) \\
& - \mathbb{E}\Big[\dot{\phi}(\tilde{y}_o, t\tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/j} + (1-t)\tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/1,2})\tilde{\boldsymbol{x}}_o^\top (\hat{\boldsymbol{\beta}}_{/2} - \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/2}\Big]\Big)\Big] \\
= \mathbb{E}\Big[\mathbb{E}\Big[& \Big(\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2}) - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/12}]\Big) \\
& \times \Big(\dot{\phi}(\tilde{y}_o, t\tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/2} + (1-t)\tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/1,2})\tilde{\boldsymbol{x}}_o^\top (\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \\
& - \mathbb{E}\big[\dot{\phi}(\tilde{y}_o, t\tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/2} + (1-t)\tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/1,2})\tilde{\boldsymbol{x}}_o^\top (\hat{\boldsymbol{\beta}}_{/2} - \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/2}\big]\Big) \,\Big|\, D_{/2}\Big]\Big] \\
= \mathbb{E}\Bigg\{ & \mathbb{E}\Big[\Big(\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2}) - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/12}]\Big) \,\Big|\, D_{/2}\Big] \\
& \times \mathbb{E}\Big[\Big(\dot{\phi}(\tilde{y}_o, t\tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/2} + (1-t)\tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/1,2})\tilde{\boldsymbol{x}}_o^\top (\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \\
& - \mathbb{E}\Big[\dot{\phi}(\tilde{y}_o, t\tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/2} + (1-t)\tilde{\boldsymbol{x}}_o^\top \hat{\boldsymbol{\beta}}_{/1,2})\tilde{\boldsymbol{x}}_o^\top (\hat{\boldsymbol{\beta}}_{/2} - \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/2}\Big]\Big) \,\Big|\, D_{/2}\Big]\Bigg\} \\
= 0. &
\end{aligned}$$

Likewise,

$$\begin{aligned}
C_0 \triangleq \mathbb{E}\Big[& \Big(\phi(y_2, \boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/1,2}) - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/12}]\Big) \\
& \times \Big(\dot{\phi}(y_1, t\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_1^\top (\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \\
& - \mathbb{E}\Big[\dot{\phi}(y_1, t\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/12})\boldsymbol{x}_1^\top (\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1}\Big]\Big)\Big] \\
= 0. &
\end{aligned}$$

To conclude, note that:

$$\begin{aligned}
& \mathbb{E}\Big[\Big(\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) - \mathbb{E}[\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]\Big)\Big(\phi(y_2, \boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2}) - \mathbb{E}[\phi(y_2, \boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2}) \mid D_{/2}]\Big)\Big] = A_0 + B_0 + C_0 \\
& + \mathbb{E}\Bigg\{\Big(\dot{\phi}(y_1, t\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_1^\top (\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \\
& - \mathbb{E}\big[\dot{\phi}(y_1, t\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_1^\top (\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1}\big]\Big) \\
& \times \Bigg(\dot{\phi}(y_2, t\boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2} + (1-t)\boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_2^\top (\hat{\boldsymbol{\beta}}_{/2} - \hat{\boldsymbol{\beta}}_{/1,2}) \\
& - \mathbb{E}\big[\dot{\phi}(y_2, t\boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2} + (1-t)\boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_2^\top (\hat{\boldsymbol{\beta}}_{/2} - \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/2}\big]\Bigg)\Bigg\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{E}\,\mathrm{Var}\left[\dot\phi(y_1, t\boldsymbol{x}_1^\top\hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_1^\top\hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_1^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1}\right] \\
&\leq \mathbb{E}\Big(\mathbb{E}\big[\dot\phi(y_1, t\boldsymbol{x}_1^\top\hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_1^\top\hat{\boldsymbol{\beta}}_{/1,2})\boldsymbol{x}_1^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}) \mid D_{/1}\big]^2\Big) \\
&\leq \mathbb{E}\Big(\mathbb{E}\big[\dot\phi(y_1, t\boldsymbol{x}_1^\top\hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_1^\top\hat{\boldsymbol{\beta}}_{/1,2})^2 \mid D_{/1}\big]\,\mathbb{E}\big[\big(\boldsymbol{x}_1^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2})\big)^2 \mid D_{/1}\big]\Big) \\
&\leq c_1^2\,\mathbb{E}\Big(\mathbb{E}\big[\big(\boldsymbol{x}_1^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2})\big)^2 \mid D_1\big]\Big) \\
&= c_1^2\,\mathbb{E}\Big(\big(\boldsymbol{x}_1^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2})\big)^2\Big) \\
&= c_1^2\,\mathbb{E}\left((\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2})^\top\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2})\right) \\
&\leq c_1^2\frac{\rho}{p}\,\mathbb{E}\|\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}\|_2^2 \\
&\leq c_1^2 c_0^2\frac{\rho}{p\nu^2}\,\mathbb{E}\|\boldsymbol{x}_1\|_2^2,
\end{aligned}
$$

where the last inequality is due to Lemma 6. Using the fact that $\delta = n/p$, $\boldsymbol{x}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\sigma_{\max}(\boldsymbol{\Sigma}) = \rho/p$, we get

$$
\mathbb{E}\left[(\phi(y_i, \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i}) - \mathbb{E}[\phi(y_i, \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i})|D_{/i}])(\phi(y_j, \boldsymbol{x}_j^\top\hat{\boldsymbol{\beta}}_{/j}) - \mathbb{E}[\phi(y_j, \boldsymbol{x}_j^\top\hat{\boldsymbol{\beta}}_{/j})|D_{/j}])\right] \leq \frac{1}{n}\Big(\frac{c_0 c_1\rho\delta^{1/2}}{\nu}\Big)^2.
$$

□

Lemma 5. *Under the assumptions of Theorem 1, we have:*

$$
\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\phi(y_i, \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}] - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}}) \mid D]\right)^2 \leq \frac{1}{n}\left(\frac{c_0 c_1\rho\delta^{1/2}}{\nu}\right)^2.
$$

Proof of Lemma 5. Note that we have for all i:

$$
\mathbb{E}[\phi(y_i, \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}] = \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}] = \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}}_{/i}) \mid D].
$$

Therefore, using the mean-value Theorem, for some $t \in [0, 1]$, we get

$$
\begin{aligned}
&\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\phi(y_i, \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}] - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}}) \mid D]\right)^2 \\
&= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}}_{/i}) \mid D] - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}}) \mid D]\right)^2 \\
&= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}}_{/i}) - \phi(y_o, \boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}}) \mid D]\right)^2 \\
&= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\dot\phi(y_o, t\boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}} + (1-t)\boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}}_{/i})\boldsymbol{x}_o^\top(\hat{\boldsymbol{\beta}}_{/i} - \hat{\boldsymbol{\beta}}) \mid D]\right)^2 \\
&\leq \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \sqrt{\mathbb{E}[\dot\phi(y_o, t\boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}} + (1-t)\boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}}_{/i})^2 \mid D]}\sqrt{\mathbb{E}[(\boldsymbol{x}_o^\top(\hat{\boldsymbol{\beta}}_{/i} - \hat{\boldsymbol{\beta}}))^2) \mid D]}\right)^2
\end{aligned}
$$

$$
\begin{aligned}
&\leq c_1^2\, \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\mathbb{E}[(\boldsymbol{x}_o^\top(\hat{\boldsymbol{\beta}}_{/i}-\hat{\boldsymbol{\beta}}))^2) \mid D]}\right)^2 \\
&\leq c_1^2\, \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{(\hat{\boldsymbol{\beta}}_{/i}-\hat{\boldsymbol{\beta}})^\top\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_{/i}-\hat{\boldsymbol{\beta}})}\right)^2 \\
&\leq c_1^2\, \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{\rho}{p}}\|\hat{\boldsymbol{\beta}}_{/i}-\hat{\boldsymbol{\beta}}\|_2\right)^2 \\
&\leq c_1^2\frac{\rho}{p}\left(\frac{1}{n}\,\mathbb{E}\,\|\hat{\boldsymbol{\beta}}_{/1}-\hat{\boldsymbol{\beta}}\|_2^2+\frac{n-1}{n}\,\mathbb{E}\,\|\hat{\boldsymbol{\beta}}_{/1}-\hat{\boldsymbol{\beta}}\|_2\|\hat{\boldsymbol{\beta}}_{/2}-\hat{\boldsymbol{\beta}}\|_2\right) \\
&\leq c_1^2\frac{\rho}{p}\,\mathbb{E}\|\hat{\boldsymbol{\beta}}_{/1}-\hat{\boldsymbol{\beta}}\|_2^2 \\
&\leq c_1^2c_0^2\frac{\rho}{p\nu^2}\,\mathbb{E}\,\|\boldsymbol{x}_1\|_2^2,
\end{aligned}
$$

where the last inequality is due to Lemma 6. Using the fact that $\delta = n/p$, $\boldsymbol{x}_1 \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\sigma_{\max}(\boldsymbol{\Sigma}) = \rho/p$, we get

$$
\mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big[\phi(y_i, \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}\big] - \mathbb{E}\big[\phi(y_o, \boldsymbol{x}_o^\top\hat{\boldsymbol{\beta}}) \mid D\big]\Big)^2 \leq \frac{1}{n}\left(\frac{c_0c_1\rho\delta^{1/2}}{\nu}\right)^2.
$$

□

Lemma 6. *If both the loss function and the regularizer are twice differentiable, then for all $i = 1, \ldots, n$:*

$$
\begin{aligned}
\|\hat{\boldsymbol{\beta}}_{/i}-\hat{\boldsymbol{\beta}}\|_2^2 &\leq \left(\frac{\dot{\ell}_i(\hat{\boldsymbol{\beta}})}{\inf_{t\in[0,1]}\sigma_{\min}(\boldsymbol{J}_{/i}(t\hat{\boldsymbol{\beta}}+(1-t)\hat{\boldsymbol{\beta}}_{/i}))}\right)^2\|\boldsymbol{x_i}\|_2^2, \\
\|\hat{\boldsymbol{\beta}}_{/i}-\hat{\boldsymbol{\beta}}_{/ij}\|_2^2 &\leq \left(\frac{\dot{\ell}_j(\hat{\boldsymbol{\beta}}_{/i})}{\inf_{t\in[0,1]}\sigma_{\min}(\boldsymbol{J}_{/ij}(t\hat{\boldsymbol{\beta}}_{/i}+(1-t)\hat{\boldsymbol{\beta}}_{/ij}))}\right)^2\|\boldsymbol{x_j}\|_2^2.
\end{aligned}
$$

Proof of Lemma 6. The leave-one-out estimate, $\hat{\boldsymbol{\beta}}_{/i} = \hat{\boldsymbol{\beta}} + \boldsymbol{\Delta}_{/i}$, satisfies $\boldsymbol{f}_{/i}(\boldsymbol{\Delta}_{/i}) = 0$. The multivariate mean-value Theorem yields

$$
0 \quad = \quad \boldsymbol{f}_{/i}(\hat{\boldsymbol{\beta}}+\boldsymbol{\Delta}_{/i}) = \boldsymbol{f}_{/i}(\hat{\boldsymbol{\beta}}) + \left(\int_0^1 \boldsymbol{J}_{/i}(\hat{\boldsymbol{\beta}}+t\boldsymbol{\Delta}_{/i})dt\right)\boldsymbol{\Delta}_{/i} \tag{12}
$$

where the Jacobean is

$$
\boldsymbol{J}_{/i}(\boldsymbol{\theta}) \quad = \quad \lambda\boldsymbol{\nabla^2 r}(\boldsymbol{\theta}) + \boldsymbol{X}_{/i}^\top\,\mathrm{diag}[\ddot{\boldsymbol{\ell}}_{/i}(\boldsymbol{\theta})]\boldsymbol{X}_{/i}. \tag{13}
$$

Moreover, $\hat{\boldsymbol{\beta}}$ satisfies

$$
0 \quad = \quad \lambda\boldsymbol{\nabla r}(\hat{\boldsymbol{\beta}}) + \boldsymbol{X}^\top\dot{\boldsymbol{\ell}}(\hat{\boldsymbol{\beta}}) = \boldsymbol{f}_{/i}(\hat{\boldsymbol{\beta}}) + \dot{\ell}_i(\hat{\boldsymbol{\beta}})\boldsymbol{x_i}.
$$

We get

$$
\dot{\ell}_i(\hat{\boldsymbol{\beta}})\boldsymbol{x_i} \quad = \quad \left(\int_0^1 \boldsymbol{J}_{/i}(\hat{\boldsymbol{\beta}}+t\boldsymbol{\Delta}_{/i})dt\right)\boldsymbol{\Delta}_{/i},
$$

leading to

$$
\boldsymbol{\Delta}_{/i} \quad = \quad \dot{\ell}_i(\hat{\boldsymbol{\beta}})\left(\int_0^1 \boldsymbol{J}_{/i}(\hat{\boldsymbol{\beta}}+t\boldsymbol{\Delta}_{/i})dt\right)^{-1}\boldsymbol{x_i},
$$

and

$$
\|\boldsymbol{\Delta}_{/i}\|_2^2 \quad \leq \quad \left(\frac{\dot{\ell}_i(\hat{\boldsymbol{\beta}})}{\inf_{t\in[0,1]}\sigma_{\min}(\boldsymbol{J}_{/i}(t\hat{\boldsymbol{\beta}}+(1-t)\hat{\boldsymbol{\beta}}_{/i}))}\right)^2\|\boldsymbol{x_i}\|_2^2.
$$

Likewise,

$$\|\hat{\boldsymbol{\beta}}_{/i} - \hat{\boldsymbol{\beta}}_{/ij}\|_2^2 \leq \left(\frac{\dot{\ell}_j(\hat{\boldsymbol{\beta}}_{/i})}{\inf_{t\in[0,1]} \sigma_{\min}(\boldsymbol{J}_{/ij}(t\hat{\boldsymbol{\beta}}_{/i} + (1-t)\hat{\boldsymbol{\beta}}_{/ij}))} \right)^2 \|\boldsymbol{x_j}\|_2^2.$$

□

D Proof of Corollary 1

We would like to use Theorem 1 to prove this corollary. Toward this goal, we first have to prove that Assumptions 1, 2, and 3 hold, and that $\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}]$ is bounded. Given the fact that $\boldsymbol{x}_i$ is $N(0, \boldsymbol{\Sigma})$, Assumption 1 holds. As we discussed in Example 1, Assumption 2 holds as well with $\nu = \lambda$. Finally, given that the regularizer is ridge Assumption 3 holds too. Hence, the only remaining step is to check the boundedness of $\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}]$. In the rest of the proof we aim to prove that

$$\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}] \leq 6 + \frac{5\rho\delta}{\lambda}. \tag{14}$$

Note that

$$\begin{aligned}
&\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}] \\
&\quad = \mathrm{Var}[-y_o \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i} + \log(1 + e^{\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}}) \mid D_{/i}] \\
&\quad \leq \mathbb{E}[(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i})^2 \mid D_{/i}] + \mathbb{E}[\log^2(1 + e^{\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}}) \mid D_{/i}] \\
&\qquad + 2\sqrt{\mathbb{E}[(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i})^2 \mid D_{/i}]\, \mathbb{E}[\log^2(1 + e^{\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}}) \mid D_{/i}]} \\
&\quad \leq \mathbb{E}[(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i})^2 \mid D_{/i}] + \mathbb{E}[(1 + |\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}|)^2 \mid D_{/i}] \\
&\qquad + 2\sqrt{\mathbb{E}[(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i})^2 \mid D_{/i}]\, \mathbb{E}[(1 + |\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}|)^2 \mid D_{/i}]},
\end{aligned}$$

where to obtain the last inequality we have used $\log(1 + e^z) \leq 1 + |z|$. Furthermore, since $(1 + |z|)^2 \leq 2 + 2z^2$, we have

$$\begin{aligned}
&\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}] \\
&\quad \leq 2 + 3\,\mathbb{E}[(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i})^2 \mid D_{/i}] \\
&\qquad + 4(1 + \mathbb{E}[(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i})^2 \mid D_{/i}]) \\
&\quad = 6 + 7\,\mathbb{E}[(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i})^2 \mid D_{/i}] \\
&\quad = 6 + 7\hat{\boldsymbol{\beta}}_{/i}^\top \boldsymbol{\Sigma} \hat{\boldsymbol{\beta}}_{/i} \leq 6 + \frac{7\rho}{p} \hat{\boldsymbol{\beta}}_{/i}^\top \hat{\boldsymbol{\beta}}_{/i}.
\end{aligned} \tag{15}$$

Comparing $\hat{\boldsymbol{\beta}}_{/i}$ with the zero estimator yields, $n \log 2 \geq \boldsymbol{y}_{\boldsymbol{/i}}^\top \boldsymbol{X}_{\boldsymbol{/i}} \hat{\boldsymbol{\beta}}_{/i} + \boldsymbol{1}^\top \log(1 + e^{\boldsymbol{X}_{/i}\hat{\boldsymbol{\beta}}_{/i}}) + \lambda\|\hat{\boldsymbol{\beta}}_{/i}\|_2^2/2$. Since $\log(1 + e^z) - z \geq 0$ for any z, we get $\lambda\|\hat{\boldsymbol{\beta}}_{/i}\|_2^2 \leq n \log 2$. Therefore, we can say that for ridge regularized logistic regression

$$\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}) | D_{/i}] \leq 6 + \frac{5\rho\delta}{\lambda}.$$

To summarize, using the bound above and Theorem 1 (with $C_b = (\frac{c_0 c_1 \rho \delta^{1/2}}{\nu})^2)$, for ridge regularized logistic regression, we conclude that

$$\begin{aligned}
C_v &= \mathbb{E}\,\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] + 2C_b + 2C_b^{1/2} \sqrt{\mathbb{E}\,\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] + C_b} \\
&= 6 + \frac{5\rho\delta}{\lambda} + 2\left(\frac{c_0 c_1 \rho \delta^{1/2}}{\nu}\right)^2 + 2\left(\frac{c_0 c_1 \rho \delta^{1/2}}{\nu}\right) \sqrt{6 + \frac{5\rho\delta}{\lambda} + \left(\frac{c_0 c_1 \rho \delta^{1/2}}{\nu}\right)^2}, \\
&= 6 + \frac{5\rho\delta}{\lambda} + 2\left(\frac{4\rho\delta^{1/2}}{\lambda}\right)^2 + 2\left(\frac{4\rho\delta^{1/2}}{\lambda}\right) \sqrt{6 + \frac{5\rho\delta}{\lambda} + \left(\frac{4\rho\delta^{1/2}}{\lambda}\right)^2}
\end{aligned} \tag{16}$$

where the last equation is due to $c_0 = c_1 = 2$ (shown in Example 1) and $\nu = \lambda$.

E Proof of Corollary 2

We would like to use Theorem 1 to prove this corollary. Toward this goal we have to confirm Assumptions 1, 2, and 3 and prove the boundedness of $\mathbb{E}\left(\text{Var}[f_H(y_0, \boldsymbol{x}_0^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}]\right)$. Assumption 1 is already assumed in the corollary. Assumption 2 is also confirmed in Example 2 so that $c_0 = c_1 = \gamma$. Since the regularizer is assumed to be strongly convex, Assumption 3 is also automatically satisfied with $\nu = \nu_r$. Hence, the only remaining step is to obtain an upper bound for $\mathbb{E}\left(\text{Var}[f_H(y_0, \boldsymbol{x}_0^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}]\right)$. In the rest of the proof we prove that

$$\mathbb{E}\left(\text{Var}[f_H(y_0, \boldsymbol{x}_0^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}]\right) \leq 2\left(\gamma^4 + \frac{\rho\gamma^3\delta}{\nu_r}(\sigma_\epsilon + \sqrt{\rho b}) + \gamma^2(\rho b + \sigma_\epsilon^2)\right).$$

We have

$$\begin{aligned}
\text{Var}[f_H(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}] &= \gamma^4 \,\text{Var}\Big[\Big(\Big\{1 + (\frac{y_o - \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}}{\gamma})^2\Big\}^{1/2} - 1\Big)^2 \mid \hat{\boldsymbol{\beta}}_{/i}\Big] \\
\leq \;& \gamma^4\, \mathbb{E}\Big[\Big(\Big\{1 + (\frac{y_o - \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}}{\gamma})^2\Big\}^{1/2} - 1\Big)^2 \mid \hat{\boldsymbol{\beta}}_{/i}\Big] \\
\leq \;& \gamma^4\, \mathbb{E}\Big[2 + \frac{|y_o - \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}|^2}{\gamma^2} \mid \hat{\boldsymbol{\beta}}_{/i}\Big] \\
\leq \;& 2\gamma^4 + \gamma^2\, \mathbb{E}[|y_o - \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i}|^2 \mid \hat{\boldsymbol{\beta}}_{/i}] \\
\leq \;& 2\gamma^4 + 2\gamma^2\, \mathbb{E}[y_o^2 + (\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i})^2 \mid \hat{\boldsymbol{\beta}}_{/i}] \\
\leq \;& 2\gamma^4 + 2\gamma^2(\mathbb{E}[y_o^2 \mid \hat{\boldsymbol{\beta}}_{/i}] + \frac{\rho}{p}\|\hat{\boldsymbol{\beta}}_{/i}\|_2^2).
\end{aligned}$$

Furthermore, we have that $\mathbb{E}[y_o^2 \mid \hat{\boldsymbol{\beta}}_{/i}] \leq \frac{\rho \boldsymbol{\beta}^{*\top}\boldsymbol{\beta}^*}{p} + \text{Var}[\epsilon_0] \leq \rho b + \sigma_\epsilon^2$. Additionally, note that using the strong convexity of the regularizer, and by comparing the value of $\sum_{j\neq i} f_H(y_j - \boldsymbol{x}_j^\top \boldsymbol{\beta}) + \lambda r(\boldsymbol{\beta})$ at $\hat{\boldsymbol{\beta}}_{/i}$ and $\mathbf{0}$, we have that $\|\hat{\boldsymbol{\beta}}_{/i}\|_2^2 \leq \nu_r^{-1}\sum_{j\neq i}\gamma|y_j| \leq \nu_r^{-1}\gamma\|\boldsymbol{y}\|_1$.[4] Therefore,

$$\mathbb{E}\,\frac{\rho}{p}\|\hat{\boldsymbol{\beta}}_{/i}\|_2^2 \leq \frac{\rho\gamma}{\nu_r}\frac{1}{p}\,\mathbb{E}\,\|\boldsymbol{y}\|_1 = \frac{\rho\gamma\delta}{\nu_r}\,\mathbb{E}|y_1|.$$

We may bound this quantity explicitly in terms of the covariance of $\boldsymbol{x}$:

$$\begin{aligned}
\mathbb{E}|y_1| \leq \mathbb{E}|\epsilon_1| + |\boldsymbol{x}_1^\top \boldsymbol{\beta}^*| &\leq \sqrt{\mathbb{E}\,\epsilon_1^2} + \sqrt{\mathbb{E}(\boldsymbol{x}_1^\top \boldsymbol{\beta}^*)^2} \\
&\leq \sigma_\epsilon + \sqrt{\frac{\rho}{p}\|\boldsymbol{\beta}^*\|_2^2} \leq \sigma_\epsilon + \sqrt{\rho b}.
\end{aligned}$$

Hence,

$$\mathbb{E}\left(\text{Var}[f_H(y_0, \boldsymbol{x}_0^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}]\right) \leq 2\left(\gamma^4 + \frac{\rho\gamma^3\delta}{\nu_r}(\sigma_\epsilon + \sqrt{\rho b}) + \gamma^2(\rho b + \sigma_\epsilon^2)\right).$$

[4] We have used the fact that $f_H(a) \leq \gamma|a|$.

To summarize, using the bound above and Theorem 1 (with $C_b = (\frac{c_0 c_1 \rho \delta^{1/2}}{\nu})^2$), we conclude that

$$\begin{aligned} C_v &= \mathbb{E}\,\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] + 2C_b + 2C_b^{1/2}\sqrt{\mathbb{E}\,\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] + C_b} \\ &= 2\left(\gamma^4 + \frac{\rho\gamma^3\delta}{\nu_r}(\sigma_\epsilon + \sqrt{\rho b}) + \gamma^2(\rho b + \sigma_\epsilon^2)\right) + 2\left(\frac{c_0 c_1 \rho \delta^{1/2}}{\nu}\right)^2 \\ &+ 2\left(\frac{c_0 c_1 \rho \delta^{1/2}}{\nu}\right)\sqrt{2\left(\gamma^4 + \frac{\rho\gamma^3\delta}{\nu_r}(\sigma_\epsilon + \sqrt{\rho b}) + \gamma^2(\rho b + \sigma_\epsilon^2)\right) + \left(\frac{c_0 c_1 \rho \delta^{1/2}}{\nu}\right)^2} \\ &= 2\left(\gamma^4 + \frac{\rho\gamma^3\delta}{\nu_r}(\sigma_\epsilon + \sqrt{\rho b}) + \gamma^2(\rho b + \sigma_\epsilon^2)\right) + 2\left(\frac{\gamma^2 \rho \delta^{1/2}}{\nu_r}\right)^2 \\ &+ 2\left(\frac{\gamma^2 \rho \delta^{1/2}}{\nu_r}\right)\sqrt{2\left(\gamma^4 + \frac{\rho\gamma^3\delta}{\nu_r}(\sigma_\epsilon + \sqrt{\rho b}) + \gamma^2(\rho b + \sigma_\epsilon^2)\right) + \left(\frac{\gamma^2 \rho \delta^{1/2}}{\nu_r}\right)^2} \end{aligned}$$

where the last equation is due to $c_0 = c_1 = \gamma$ (shown in Example 2) and $\nu = \nu_r$.

F Proof of Example 1

It is straightforward to check that, for any i, we have:

$$\inf_{t \in [0,1]} \sigma_{\min}(\boldsymbol{A}_{t,/i}) \geq c\,\sigma_{\min}(\boldsymbol{X}_{/i}^\top \boldsymbol{X}_{/i}).$$

This implies that:

$$\mathbb{E}\Big(\inf_{t \in [0,1]} \sigma_{\min}(\boldsymbol{A}_{t,/i})\Big)^{-8} \leq \frac{1}{c^8}\mathbb{E}\sigma_{\min}^{-8}(\boldsymbol{X}_{/i}^\top \boldsymbol{X}_{/i}).$$

Define the vectors $\boldsymbol{z}_i = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{x}_i$. Hence, $\boldsymbol{z}_i \sim N(0, \boldsymbol{I})$. Furthermore, define the matrix $\boldsymbol{Z}$ as the matrix that has $\boldsymbol{z}_i$ as its rows. It is straightforward to check that:

$$\sigma_{\min}(\boldsymbol{X}_{/i}^\top \boldsymbol{X}_{/i}) = \sigma_{\min}(\sum_{j \neq i} \boldsymbol{x}_j \boldsymbol{x}_j^\top) \geq \frac{\rho}{p}\sigma_{\min}(\sum_{j \neq i} \boldsymbol{z}_j \boldsymbol{z}_j^\top) = \rho\delta\sigma_{\min}\left(\frac{\boldsymbol{Z}_{/i}^\top \boldsymbol{Z}_{/i}}{n}\right).$$

The fact that the quantity $\mathbb{E}\sigma_{\min}^{-8}\left(\frac{\boldsymbol{Z}_{/i}^\top \boldsymbol{Z}_{/i}}{n}\right)$ is lower bounded by a constant for large values of n, p when $n/p = \delta > 1$ is proved in [Xu et al., 2019]. See Lemma 1 in the supplementary material.

G Proof of Theorem 2

The proof of this result is very similar to the proof of Theorem 1. Hence, instead of rewriting the proof, we only emphasize on the differences between the proofs of Theorems 1 and 2. The strategy of the proof is exactly the same. We break the error between LO and $\mathrm{Err}_{\mathrm{out}}$ into V_1 and V_2 and try to bound the second moments of these quantities. The following lemma obtains an upper bound for the second moment of V_1.

Lemma 7. *Under the assumptions of Theorem 2 we have*

$$\mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^n \phi(y_i, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\phi(y_i, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}]\Big)^2 \leq \frac{1}{n}\Big(\mathbb{E}\,\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] + \tilde{c}_0 \tilde{c}_1^2 \rho\delta\tilde{\upsilon} c_4\Big)$$

Proof. The proof of this lemma is similar to the proof of Lemma 4. All the steps are exactly the same up to the point that is proved:

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) - \mathbb{E}[\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]\Big)\Big(\phi(y_2, \boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2}) - \mathbb{E}[\phi(y_2, \boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2}) \mid D_{/2}]\Big)\Big] \\
&\leq \mathbb{E}\Big(\mathbb{E}\big[\dot{\phi}(y_1, t\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1} + (1-t)\boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1,2})^2 \mid D_{/1}\big]\,\mathbb{E}\big[\big(\boldsymbol{x}_1^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2})\big)^2 \mid D_{/1}\big]\Big) \\
&\leq \tilde{c}_1^2\,\mathbb{E}\Big(\mathbb{E}\big[\big(\boldsymbol{x}_1^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2})\big)^2 \mid D_1\big]\Big) \\
&= \tilde{c}_1^2\,\mathbb{E}\Big(\big(\boldsymbol{x}_1^\top(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2})\big)^2\Big) \\
&= \tilde{c}_1^2\,\mathbb{E}\left((\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2})^\top \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2})\right) \\
&\leq \tilde{c}_1^2 \frac{\rho}{p}\,\mathbb{E}\|\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}\|_2^2.
\end{aligned}$$

However, the way we would like to bound $\mathbb{E}\|\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}\|_2^2$ here is slightly different from the approach used in the proof of Lemma 4. According to Lemma 6 we have:

$$\|\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}\|_2^2 \leq \left(\frac{\dot{\ell}_j(\hat{\boldsymbol{\beta}}_{/1})}{\inf_{t\in[0,1]} \sigma_{\min}(\boldsymbol{J}_{/1,2}(t\hat{\boldsymbol{\beta}}_{/1} + (1-t)\hat{\boldsymbol{\beta}}_{/1,2}))}\right)^2 \|\boldsymbol{x}_2\|_2^2.$$

Hence, by using the Cauchy-Schwarz inequality twice we obtain:

$$\mathbb{E}\,\|\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}_{/1,2}\|_2^2 \leq \mathbb{E}\,|(\dot{\ell}_j(\hat{\boldsymbol{\beta}}_{/1}))|^8\,\mathbb{E}\left[\frac{1}{(\inf_{t\in[0,1]} \sigma_{\min}(\boldsymbol{J}_{/1,2}(t\hat{\boldsymbol{\beta}}_{/1} + (1-t)\hat{\boldsymbol{\beta}}_{/1,2})))^8}\right]\mathbb{E}\,\|\boldsymbol{x}_2\|_2^4 \leq \tilde{c}_0\tilde{\nu}c_4.$$

Using the fact that $\delta = n/p$, $\boldsymbol{x}_1 \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\sigma_{\max}(\boldsymbol{\Sigma}) = \rho/p$, we get

$$\mathbb{E}\Big[\Big(\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) - \mathbb{E}[\phi(y_1, \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]\Big)\Big(\phi(y_2, \boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2}) - \mathbb{E}[\phi(y_2, \boldsymbol{x}_2^\top \hat{\boldsymbol{\beta}}_{/2}) \mid D_{/2}]\Big)\Big] \leq \frac{1}{n}(\tilde{c}_0\tilde{c}_1^2\rho\delta\tilde{v}c_4).$$

□

The second Lemma aims to obtain an upper bound for the second moment of V_2. This corresponds to Lemma 5 in the proof of Theorem 1.

Lemma 8. *Under the assumptions of Theorem 2, we have*

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\phi(y_i, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}] - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}) \mid D]\right)^2 \leq c_1^2 \frac{\rho}{p}\,\mathbb{E}\|\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}\|_2^2 \leq c_1^2\rho\delta_0\tilde{c}_0\tilde{v}c_4.$$

Proof. Again the proof follows very similar to the steps as the proof of Lemma 5. In fact, we follow exactly the same steps until it is proved that

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\phi(y_i, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}] - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}) \mid D]\right)^2 \leq c_1^2 \frac{\rho}{p}\,\mathbb{E}\|\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}\|_2^2.$$

Then, in order to bound $\mathbb{E}\|\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}\|_2^2$ we use a slightly different strategy. According to Lemma 6 we have

$$\|\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}\|_2^2 \leq \left(\frac{\dot{\ell}_1(\hat{\boldsymbol{\beta}})}{\inf_{t\in[0,1]} \sigma_{\min}(\boldsymbol{J}_{/1}(t\hat{\boldsymbol{\beta}} + (1-t)\hat{\boldsymbol{\beta}}_{/1}))}\right)^2 \|\boldsymbol{x}_1\|_2^2.$$

Hence, by using Cauchy-Schwarz inequality we have:

$$\frac{1}{p}\,\mathbb{E}\left(\|\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}\|_2^2\right) \leq \delta_o\tilde{c}_0\tilde{v}c_4,$$

from which we deduce:

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\phi(y_i, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}) \mid D_{/i}] - \mathbb{E}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}) \mid D]\right)^2 \leq c_1^2 \frac{\rho}{p}\,\mathbb{E}\|\hat{\boldsymbol{\beta}}_{/1} - \hat{\boldsymbol{\beta}}\|_2^2 \leq c_1^2\rho\delta_0\tilde{c}_0\tilde{v}c_4.$$

□

H Proof of Corollary 3

As is clear, we would like to use Theorem 2 to prove our claim. Toward this goal, we have to prove that Assumptions 1′, 2′, and 3′ hold. Furthermore, we have to obtain an upper bound for the constant $\tilde{C}_v$, which in turn requires us to bound $\mathbb{E}\operatorname{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]$. Given that $\boldsymbol{x}_i$ is Gaussian, Assumption 1′ is automatically satisfied. Furthermore, since the regularizer is elastic-net, it is straightforward to prove Assumption 3′. To see this, first note that, for all i, j, we have almost surely:

$$\boldsymbol{A}_{t,/i} \triangleq \boldsymbol{X}_{/i}^\top \operatorname{diag}[\ddot{\boldsymbol{\ell}}_{/i}(t\hat{\boldsymbol{\beta}}_{/i} + (1-t)\hat{\boldsymbol{\beta}})]\boldsymbol{X}_{/i} + \lambda \boldsymbol{\nabla^2 r}(t\hat{\boldsymbol{\beta}}_{/i} + (1-t)\hat{\boldsymbol{\beta}}),$$
$$\boldsymbol{A}_{t,/i,j} \triangleq \boldsymbol{X}_{/ij}^\top \operatorname{diag}[\ddot{\boldsymbol{\ell}}_{/ij}(t\hat{\boldsymbol{\beta}}_{/ij} + (1-t)\hat{\boldsymbol{\beta}}_{/i})]\boldsymbol{X}_{/ij} + \lambda \boldsymbol{\nabla^2 r}(t\hat{\boldsymbol{\beta}}_{/ij} + (1-t)\hat{\boldsymbol{\beta}}_{/i}),$$

where $r(\beta) = \gamma\beta^2 + (1-\gamma)r^\alpha(\beta)$. Hence, it is straightforward to see that

$$\sigma_{\min}(\boldsymbol{A}_{t,/i}) \geq \lambda\gamma,$$
$$\sigma_{\min}(\boldsymbol{A}_{t,/i,j}) \geq \lambda\gamma.$$

Hence, the only remaining steps are to prove Assumption 2′ and bound the term $\mathbb{E}\operatorname{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]$. Given that $\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) = \frac{1}{2}(y_o - \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1})^2$, we have

$$\operatorname{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] \leq \frac{1}{4}\mathbb{E}[(y_o - \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1})^4 \mid D_{/1}].$$

Hence,

$$\mathbb{E}\operatorname{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] \leq \frac{1}{4}\mathbb{E}[(y_o - \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1})^4] \leq \frac{1}{4}\left(\mathbb{E}[(y_o - \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1})^8]\right)^{0.5}.$$

Hence, if we prove Assumption 2′, we have also proved that

$$\mathbb{E}\operatorname{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] \leq \frac{1}{4}\left(\mathbb{E}[(y_o - \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1})^8]\right)^{0.5} \leq \frac{\tilde{c}_0^{0.5}}{4}.$$

In the rest of this section, we focus on the proof of Assumption 2′. Note that $\dot{\ell}(y, \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}) = y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}$. Under these assumptions, we prove that there exists a fixed number $\tilde{c}_0$ such that $\mathbb{E}(y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}})^8 \leq \tilde{c}_0$, and $\mathbb{E}(y_o - \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/i})^8 \leq \tilde{c}_0$.

Consider the following definitions:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} f(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} \sum_{j=1}^n \frac{(y_j - \boldsymbol{x}_j^\top \boldsymbol{\beta})^2}{2} + \lambda \sum_{i=1}^p r(\beta_i),$$
$$\hat{\boldsymbol{\beta}}_{/i} = \arg\min_{\boldsymbol{\beta}} f_{/i}(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} \sum_{j=1, j\neq i}^n \frac{(y_j - \boldsymbol{x}_j^\top \boldsymbol{\beta})^2}{2} + \lambda \sum_{i=1}^p r(\beta_i) \tag{17}$$

Furthermore, define $r_{0.5}(\beta) = \frac{\gamma}{2}\beta^2 + (1-\gamma)r^\alpha(\beta)$. Our optimization problem can be written as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} f(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} \sum_{j=1}^n \frac{(y_j - \boldsymbol{x}_j^\top \boldsymbol{\beta})^2}{2} + \lambda \sum_{i=1}^p r_{0.5}(\beta_i) + \frac{\lambda\gamma}{2}\sum_{i=1}^p \beta_i^2.$$

Since $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathrm{N}(0, \sigma_\epsilon^2 \boldsymbol{I})$, the optimality conditions yield

$$\boldsymbol{X}^\top(\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{y}) + \lambda\gamma\hat{\boldsymbol{\beta}} + \lambda\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}) = 0.$$

Hence,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top\boldsymbol{y} - \lambda(\boldsymbol{X}^\top\boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}).$$

It is then straightforward to prove that

$$\begin{aligned}
\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} &= (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top)\boldsymbol{y} + \lambda\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}) \\
&= (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top)\boldsymbol{X}\boldsymbol{\beta}^* + (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top)\boldsymbol{\epsilon} \\
&\quad + \lambda\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}).
\end{aligned} \tag{18}$$

Our goal is to show that all the "finite" moments of the elements of $\boldsymbol{y}_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}$, including the 8^{th} moment required in our example, are $O(1)$. From (18) we have

$$\begin{aligned}\mathbb{E}|\boldsymbol{y}_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}|^k \quad \leq \quad & 3^{k-1}\Big(\mathbb{E}(1 - \boldsymbol{x}_i^\top (\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma \boldsymbol{I})^{-1}\boldsymbol{X}^\top)\boldsymbol{X}\boldsymbol{\beta}^*)^k + \mathbb{E}|1 - \boldsymbol{x}_i^\top (\boldsymbol{X}^\top \boldsymbol{X}\lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top)\boldsymbol{\epsilon}|^k \\ & + \lambda^k \mathbb{E}|\boldsymbol{x}_i^\top (\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})|^k\Big) \end{aligned} \tag{19}$$

Hence, we bound each of the above three terms separately in the following lemmas:

Lemma 9. *Under the assumptions of Example 3 we have*

$$\mathbb{E}(1 - \boldsymbol{x}_i^\top (\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top)\boldsymbol{X}\boldsymbol{\beta}^*)^k \leq \left(\frac{\rho}{p\lambda^2\gamma^2}\|\boldsymbol{\beta}^*\|_2\right)^{2k} k!!.$$

Proof. First note that

$$(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top)\boldsymbol{X}\boldsymbol{\beta}^* = \lambda\gamma\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{\beta}^*. \tag{20}$$

Hence,

$$1 - \boldsymbol{x}_i^\top (\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top)\boldsymbol{X}\boldsymbol{\beta}^* = \lambda\gamma\boldsymbol{x}_i^\top (\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{\beta}^*.$$

Define $\boldsymbol{D}_i = (\boldsymbol{X}_{/i}^\top \boldsymbol{X}_{/i} + \lambda\gamma\boldsymbol{I})^{-1}$. According to the matrix inversion lemma we have

$$\boldsymbol{x}_i^\top (\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{\beta}^* = \boldsymbol{x}_i^\top \boldsymbol{D}_i\boldsymbol{\beta}^* - \frac{\boldsymbol{x}_i^\top \boldsymbol{D}_i\boldsymbol{x}_i\boldsymbol{x}_i^\top \boldsymbol{D}_i\boldsymbol{\beta}^*}{1 + \boldsymbol{x}_i^\top \boldsymbol{D}_i\boldsymbol{x}_i} = \frac{\boldsymbol{x}_i^\top \boldsymbol{D}_i\boldsymbol{\beta}^*}{1 + \boldsymbol{x}_i^\top \boldsymbol{D}_i\boldsymbol{x}_i}. \tag{21}$$

Note that conditioned on $\boldsymbol{X}_{/i}$ the distribution of $\boldsymbol{x}_i^\top \boldsymbol{D}_i\boldsymbol{\beta}^*$ is a zero mean Gaussian random variable with variance $v_i = \|\boldsymbol{\Sigma}^{1/2}\boldsymbol{D}_i\boldsymbol{\beta}^*\|_2^2 \leq \frac{\rho}{p\lambda^2\gamma^2}\|\boldsymbol{\beta}^*\|_2^2$. Hence, (21) and the moments of a Gaussian random variable (see Lemma 2) lead to

$$\mathbb{E}(|\boldsymbol{x}_i^\top (\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{\beta}^*|^k \mid \boldsymbol{X}_{/i}) \leq \nu_i^k (k-1)!!. \tag{22}$$

Hence, by the law of iterated expectation, we obtain

$$\mathbb{E}(|\boldsymbol{x}_i^\top (\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{\beta}^*|^k) \leq \nu_i^k (k-1)!! \leq \left(\frac{\rho}{p\lambda^2\gamma^2}\|\boldsymbol{\beta}^*\|_2\right)^{2k} k!!.$$

□

Lemma 10. *Under the assumptions of Example 3, if $\boldsymbol{\epsilon} \sim \mathrm{N}(0, \sigma_\epsilon^2\boldsymbol{I})$, then*

$$\mathbb{E}|1 - \boldsymbol{x}_i^\top (\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top)\boldsymbol{\epsilon}|^k \leq \sigma_\epsilon^k (k-1)!!.$$

Proof. Note that conditioned on $\boldsymbol{X}$, the distribution of $\boldsymbol{v} = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top)\boldsymbol{\epsilon}$ is multivariate Gaussian with mean zero and covariance matrix $\sigma_\epsilon^2(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top)^2$. We have

$$(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top)^2 = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{X}^\top - \lambda\gamma\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-2}\boldsymbol{X}^\top. \tag{23}$$

We define $\sigma_i^2(\boldsymbol{X}) = \left(1 - \boldsymbol{x}_i^\top (\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-1}\boldsymbol{x}_i - \lambda\gamma\boldsymbol{x}_i^\top (\boldsymbol{X}^\top \boldsymbol{X} + \lambda\gamma\boldsymbol{I})^{-2}\boldsymbol{x}_i\right)\sigma_\epsilon^2$. Clearly $\sigma_i^2(\boldsymbol{X}) \leq \sigma_\epsilon^2$, hence,

$$\mathbb{E}(|v_i|^k \mid \boldsymbol{X}) \leq \sigma_i^k(X)(k-1)!! \leq \sigma_\epsilon^k(k-1)!!, \tag{24}$$

where the first inequality is due to Lemma 2. Hence, again by the law of iterated expectation, we have

$$\mathbb{E}(|v_i|^k) \leq \sigma_\epsilon^k (k-1)!!.$$

□

Lemma 11. *Under the assumptions of Example 3 we have*

$$\mathbb{E}|\boldsymbol{x}_i^\top(\boldsymbol{X}^\top\boldsymbol{X}+\frac{\lambda\gamma}{2}\boldsymbol{I})^{-1}\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})|^k$$
$$\leq 2^{2k-\frac{3}{2}}1.5^{\frac{k}{2}}\left(\frac{1}{\lambda^2\gamma}\left(1+\frac{\alpha(1-\gamma)}{2\gamma}\right)\right)^k\left(\rho\frac{\boldsymbol{\beta}^\top\boldsymbol{\beta}}{p}+\sigma_\epsilon^2\right)^{\frac{k}{2}}\sqrt{(2k)!!(1+\left(\frac{1.5c}{\sqrt{2\lambda\gamma}}\right)^{2k})+\zeta^{\frac{k}{2}}}. \tag{25}$$

Proof. Since $f_{/i}(\hat{\boldsymbol{\beta}}_{/i})\leq f_{/i}(\boldsymbol{0})$, we have

$$2\lambda\gamma\|\hat{\boldsymbol{\beta}}_{/i}\|_2^2\leq\|\boldsymbol{y}_{/i}\|_2^2. \tag{26}$$

Furthermore, due to $\ddot{r}_{0.5}(\beta)\leq\gamma+\frac{\alpha(1-\gamma)}{2}$, $\dot{r}_{0.5}(0)=0$, and (26), we have

$$\|\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i})\|_2^2\leq\left(\gamma+\frac{\alpha(1-\gamma)}{2}\right)\|\hat{\boldsymbol{\beta}}_{/i}\|_2^2\leq\left(\frac{1}{2\lambda}+\frac{\alpha(1-\gamma)}{4\lambda\gamma}\right)\|\boldsymbol{y}_{/i}\|_2^2. \tag{27}$$

The first order optimality condition yields

$$\boldsymbol{X}^\top\boldsymbol{X}(\hat{\boldsymbol{\beta}}_{/i}-\hat{\boldsymbol{\beta}})+\lambda\dot{\boldsymbol{r}}(\hat{\boldsymbol{\beta}}_{/i})-\lambda\dot{\boldsymbol{r}}(\hat{\boldsymbol{\beta}})=-\boldsymbol{x}_i(y_i-\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i}).$$

Since the minimum eigenvalue of the Hessian of $\boldsymbol{r}(\boldsymbol{\beta})$ is 2γ, therefore the minimum eigenvalue of $\boldsymbol{X}^\top\boldsymbol{X}+\lambda\,\mathrm{diag}[\ddot{\boldsymbol{r}}(\boldsymbol{\beta})]$ (for all $\boldsymbol{\beta}$) is greater than $2\lambda\gamma$, leading to

$$\|\hat{\boldsymbol{\beta}}_{/i}-\hat{\boldsymbol{\beta}}\|_2\leq\frac{|y_i-\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i}|}{2\lambda\gamma}\|\boldsymbol{x}_i\|_2.$$

This together with $\ddot{r}_{0.5}(\beta)\leq\gamma+\frac{\alpha(1-\gamma)}{2}$ yields

$$\|\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i})-\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})\|_2\leq\left(\gamma+\frac{\alpha(1-\gamma)}{2}\right)\|\hat{\boldsymbol{\beta}}_{/i}-\hat{\boldsymbol{\beta}}\|_2\leq\left(\frac{1}{2\lambda}+\frac{\alpha(1-\gamma)}{4\lambda\gamma}\right)|y_i-\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i}|\|\boldsymbol{x}_i\|_2.$$

Define $\boldsymbol{D}_i=(\boldsymbol{X}_{/i}^\top\boldsymbol{X}_{/i}+\lambda\gamma\boldsymbol{I})^{-1}$. According to the matrix inversion lemma we have

$$\boldsymbol{x}_i^\top(\boldsymbol{X}^\top\boldsymbol{X}+\lambda\gamma\boldsymbol{I})^{-1}\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})=\boldsymbol{x}_i^\top\boldsymbol{D}_i\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})-\frac{\boldsymbol{x}_i^\top\boldsymbol{D}_i\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{D}_i\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})}{1+\boldsymbol{x}_i^\top\boldsymbol{D}_i\boldsymbol{x}_i}=\frac{\boldsymbol{x}_i^\top\boldsymbol{D}_i\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})}{1+\boldsymbol{x}_i^\top\boldsymbol{D}_i\boldsymbol{x}_i}. \tag{28}$$

Furthermore, we have

$$|\boldsymbol{x}_i^\top\boldsymbol{D}_i\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})|\leq|\boldsymbol{x}_i^\top\boldsymbol{D}_i\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i})|+|\boldsymbol{x}_i^\top\boldsymbol{D}_i(\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})-\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i}))|. \tag{29}$$

Note that for two random variables a and b we have

$$\mathbb{E}(a+b)^k\leq 2^{k-1}\mathbb{E}(a^k+b^k).$$

Hence,

$$\mathbb{E}(|\boldsymbol{x}_i^\top\boldsymbol{D}_i\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})|)^k\leq 2^{k-1}\left(\mathbb{E}|\boldsymbol{x}_i^\top\boldsymbol{D}_i\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i})|^k+\mathbb{E}|\boldsymbol{x}_i^\top\boldsymbol{D}_i(\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})-\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i}))|^k\right). \tag{30}$$

First note that, since the maximum eigenvalue of $\boldsymbol{D}_i$ is $\lambda\gamma$ we have

$$\begin{aligned}&|\boldsymbol{x}_i^\top\boldsymbol{D}_i(\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})-\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i}))|\\ &\leq\frac{1}{\lambda\gamma}\|\boldsymbol{x}_i\|_2\|\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})-\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i})\|_2\leq\frac{1}{2\lambda^2\gamma}\|\boldsymbol{x}_i\|_2^2\left(1+\frac{\alpha(1-\gamma)}{2\gamma}\right)|y_i-\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i}|\\ &\leq\frac{1}{2\lambda^2\gamma}\left(1+\frac{\alpha(1-\gamma)}{2\gamma}\right)\|\boldsymbol{x}_i\|_2^2(|y_i|+|\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i}|).\end{aligned} \tag{31}$$

Hence,

$$\begin{aligned}\mathbb{E}(|\boldsymbol{x}_i^\top\boldsymbol{D}_i(\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}})-\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i}))|)^k&\leq\left(\frac{1}{\lambda^2\gamma}\left(1+\frac{\alpha(1-\gamma)}{2\gamma}\right)\right)^k\sqrt{\mathbb{E}(\|\boldsymbol{x}_i\|_2)^{2k}\mathbb{E}(|y_i|+|\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i}|)^{2k}}\\ &\leq\left(\frac{1}{2\lambda^2\gamma}\left(1+\frac{\alpha(1-\gamma)}{2\gamma}\right)\right)^k 2^{(2k-1)/2}\sqrt{\mathbb{E}(\|\boldsymbol{x}_i\|_2)^{2k}(\mathbb{E}|y_i|^{2k}+\mathbb{E}|\boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}_{/i}|^{2k})}\end{aligned} \tag{32}$$

Furthermore, we have

1. According to Lemma 3, $\mathbb{E}\|\boldsymbol{x}_i\|_2^\ell = \frac{p(p+2)\dots(p+\ell-2)}{p^{\frac{\ell}{2}}} \leq \left(1+\frac{\ell-2}{p}\right)^{\frac{\ell}{2}} \leq 1.5^{\frac{\ell}{2}}$, where the last inequality is according to the assumption $p > 2(\ell-2)$.

2. Note that $y_i \sim N(0, \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta} + \sigma_\epsilon^2)$. Furthermore, $\boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta} + \sigma_\epsilon^2 \leq \rho\frac{\boldsymbol{\beta}^\top\boldsymbol{\beta}}{p} + \sigma_\epsilon^2$. Hence, using the the moments of Gaussian (see Lemma 2), we have
$$\mathbb{E}|y_i|^\ell \leq \left(\rho\frac{\boldsymbol{\beta}^\top\boldsymbol{\beta}}{p} + \sigma_\epsilon^2\right)^{\ell/2} \ell!!. \tag{33}$$

3. Given $\boldsymbol{X}_{/i}, \boldsymbol{y}_{/i}$, the distribution of $\boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}$ is $N(0, \hat{\boldsymbol{\beta}}_{/i}^\top \boldsymbol{\Sigma}\hat{\boldsymbol{\beta}}_{/i})$. Furthermore, $\hat{\boldsymbol{\beta}}_{/i}^\top \boldsymbol{\Sigma}\hat{\boldsymbol{\beta}}_{/i} \leq \frac{c\hat{\boldsymbol{\beta}}_{/i}^\top\hat{\boldsymbol{\beta}}_{/i}}{n} \leq \frac{c\|\boldsymbol{y}_{/i}\|_2^2}{2n\lambda\gamma}$, where the last inequality is due to (26). Hence, we have
$$\mathbb{E}(|\boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}|^\ell \mid \boldsymbol{X}_{/i}, \boldsymbol{y}_{/i}) \leq \left(\frac{c\|\boldsymbol{y}_{/i}\|_2^2}{2n\lambda\gamma}\right)^{\ell/2} \ell!!. \tag{34}$$

   Since $y_i \overset{i.i.d.}{\sim} N(0, \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta} + \sigma_\epsilon^2)$, and $\boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta} + \sigma_\epsilon^2 \leq \frac{\rho\boldsymbol{\beta}^\top\boldsymbol{\beta}}{p} + \sigma_\epsilon^2$, we have
$$\begin{aligned}\mathbb{E}(|\boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_{/i}|^\ell) &\leq \left(\frac{c^\ell \mathbb{E}(\|\boldsymbol{y}_{/i}\|_2^\ell)}{(2n\lambda\gamma)^{\ell/2}}\right)\ell!! \leq \frac{c^\ell\left(\frac{\rho\|\boldsymbol{\beta}^\top\boldsymbol{\beta}\|_2^2}{p} + \sigma_\epsilon^2\right)^\ell}{(2\gamma\lambda)^{\frac{\ell}{2}}}\ell!!\frac{n(n+2)\dots(n+\ell-2)}{n^{\ell/2}} \\ &\leq \frac{c^\ell\left(\frac{\rho\|\boldsymbol{\beta}^\top\boldsymbol{\beta}\|_2^2}{p} + \sigma_\epsilon^2\right)^\ell}{(2\gamma\lambda)^{\frac{\ell}{2}}} 1.5^\ell \ell!!, \end{aligned} \tag{35}$$

   where for the last inequality we assumed that $n > 2\ell$.

Finally, we compute an upper bound on $|\boldsymbol{x}_i^\top \boldsymbol{D}_i \dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i})|$. Since $\boldsymbol{x}_i$ is independent of $\boldsymbol{y}_{/i}$ and $\boldsymbol{X}_{/i}$, we conclude that given $\boldsymbol{X}_{/i}$ and $\boldsymbol{y}_{/i}$, $\boldsymbol{x}_i^\top \boldsymbol{D}_i \dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i})$ is a Gaussian random variable with mean zero and variance
$$\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{D}_i \dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i})\|_2^2 \leq \frac{4\rho_{\max}}{\lambda^2\gamma^2}\|\dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i})\|_2^2 \leq \frac{2\rho_{\max}}{\lambda^3\gamma^2}\left(1+\frac{\alpha(1-\gamma)}{2\gamma}\right)\|\boldsymbol{y}_{/i}\|_2^2 = \frac{\zeta\|\boldsymbol{y}_{/i}\|_2^2}{n},$$
where $\zeta = \frac{2c}{\lambda^3\gamma^2}\left(1+\frac{\alpha(1-\gamma)}{2\gamma}\right)$, and the second inequality is due to (27). Hence,
$$\mathbb{E}\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{D}_i \dot{\boldsymbol{r}}_{0.5}(\hat{\boldsymbol{\beta}}_{/i})\|_2^\ell \leq \zeta^{\ell/2}\frac{n(n+2)\dots(n+\frac{\ell}{2}-2)}{n^{\ell/2}} \leq (1.5\zeta)^{\ell/2}.$$

□

I Proof of Corollary 4

The goal of this section is to use Theorem 2 to prove corollary 4. Hence, we have to confirm that Assumptions 1′, 2′, and 3′ hold, and that $\mathbb{E}\,\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]$ is bounded. Similar to what we did at the beginning of Section H, it is straightforward to check the validity of Assumptions 1′ and 3′. Hence, we only focus on proving Assumption 2′ and finding an upper bound for $\mathbb{E}\,\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]$.

Regarding Assumption 2′, we first prove that under the assumptions of this corollary, there exists a fixed number $\tilde{c}_0$, such that $\mathbb{E}(\dot{\ell}(y_i \mid \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}))^8 \leq \tilde{c}_0$ and $\mathbb{E}(\dot{\ell}(y_0 \mid \boldsymbol{x}_0^\top \hat{\boldsymbol{\beta}}_{/i}))^8 \leq \tilde{c}_0$. Since $\ell(y \mid z) = f(z) - y\log f(z)$, we have
$$\dot{\ell}(y_i \mid \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}) = f'(\boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}) - y_i f'(\boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}})/f(\boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}),$$
where $f'(z) = 1/(1+e^{-z})$. We have that, for all $z \in \mathbb{R}$, $f'(z) \leq 1$ and $0 \leq f'(z)/f(z) \leq 1$, from which we deduce that:
$$|\dot{\ell}(y_i \mid \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}})| \leq 1 + y_i. \tag{36}$$

In particular, we have that:

$$
\begin{aligned}
\mathbb{E}|\dot{\ell}(y_i \mid \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}})|^8 &\leq \mathbb{E}(1+y_i)^8 \\
&\leq \mathbb{E}\, e^{8y_i} = \mathbb{E}\,\mathbb{E}[e^{8y_i} \mid \boldsymbol{x}_i^\top \boldsymbol{\beta}^*] \\
&\overset{(a)}{=} \mathbb{E}\exp\{(e^8-1)\boldsymbol{x}_i^\top \boldsymbol{\beta}^*\} \\
&\overset{(b)}{\leq} \exp\Big\{\frac{\rho}{2p}\|\boldsymbol{\beta}^*\|_2^2 (e^8-1)^2\Big\} \\
&= \exp\Big\{\frac{(e^8-1)^2}{2}\frac{\rho}{p}\|\boldsymbol{\beta}^*\|_2^2\Big\}, \\
&\leq \exp\Big\{\frac{(e^8-1)^2}{2}\rho b\Big\}.
\end{aligned}
$$

To obtain equality (a) we have used the moment generating function of the Poisson distribution with $y_i \sim Poisson(f(\boldsymbol{x}_i^\top \boldsymbol{\beta}^*))$. To obtain inequality (b) we have used the moment generating function of a Gaussian distribution and the fact that $\mathbb{E}(\boldsymbol{x}_i^\top \boldsymbol{\beta}^*)^2 \leq \frac{\rho}{p}\|\boldsymbol{\beta}^*\|_2^2$. Given that the upper bound we derived in (36) for the derivative of the loss function does not depend on the second input argument of the loss, that is $\boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}$, the proof that Poisson loss satisfies the other conditions of Assumption 2′ for $\phi(y,z) = \ell(y \mid z)$ will be exactly similar and hence is skipped. In particular, we have verified the conditions of Assumption 2′ for any convex regularizer.

Now we turn our attention to bounding $\mathbb{E}\,\mathrm{Var}[\ell(y_o \mid \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]$. First note that

$$
\mathrm{Var}[\ell(y_o|\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] \leq \mathbb{E}[\ell^2(y_o|\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]. \tag{37}
$$

Furthermore, from the mean value theorem we have:

$$
\ell(y_o|\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) = \ell(y_o \mid \boldsymbol{x}_o^\top \boldsymbol{\beta}^*) + \dot{\ell}(y_o|\tilde{z})(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1} - \boldsymbol{x}_o^\top \boldsymbol{\beta}^*),
$$

Hence, we have:

$$
\ell^2(y_o \mid \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \leq 2\ell^2(y_o \mid \boldsymbol{x}_o^\top \boldsymbol{\beta}^*) + 2(1+y_o^2)(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1} - \boldsymbol{x}_o^\top \boldsymbol{\beta}^*)^2. \tag{38}
$$

To complete the proof we have to show that both $\mathbb{E}\,\ell^2(y_o, \boldsymbol{x}_o^\top \boldsymbol{\beta}^*)$ and $\mathbb{E}(1+y_i^2)(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1} - \boldsymbol{x}_o^\top \boldsymbol{\beta}^*)^2$ are bounded. First note that, using $\ell(y \mid z) = f(z) - y\log f(z)$ and, for any $a, b \in \mathrm{R}$, $(a+b)^2 \leq 2a^2 + 2b^2$, yields

$$
\ell^2(y_o \mid \boldsymbol{x}_o^\top \boldsymbol{\beta}^*) \leq 2f^2(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*) + 2y_o^2 \log^2 f(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*). \tag{39}
$$

Hence,

$$
\mathbb{E}\,\ell^2(y_o \mid \boldsymbol{x}_o^\top \boldsymbol{\beta}^*) \leq 2\,\mathbb{E}\,f^2(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*) + 2\,\mathbb{E}(f(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*) + f^2(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*))\log^2 f(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*). \tag{40}
$$

The following facts will help us bound these terms:

$$
\begin{aligned}
f(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*) &\geq 0 \\
f(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*) &\leq 1 + |\boldsymbol{x}_o^\top \boldsymbol{\beta}^*|,
\end{aligned} \tag{41}
$$

On the other hand, it is straightforward to check that for any $\gamma > 0$ we have

$$
\begin{aligned}
\gamma \log^2 \gamma &\leq 1 + \gamma^2, \\
\gamma^2 \log^2 \gamma &\leq 1 + \gamma^3.
\end{aligned} \tag{42}
$$

By combining these equations we obtain:

$$
\begin{aligned}
\mathbb{E}\,\ell^2(y_o \mid \boldsymbol{x}_o^\top \boldsymbol{\beta}^*) &\leq 2\,\mathbb{E}\,f^2(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*) + 2\,\mathbb{E}(1 + f^2(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*)) + 2\,\mathbb{E}(1 + f^3(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*)) \\
&\leq 4 + 4\,\mathbb{E}\,f^2(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*) + 2\,\mathbb{E}\,f^3(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*) \\
&\leq 4 + 4\,\mathbb{E}(1 + |\boldsymbol{x}_o^\top \boldsymbol{\beta}^*|)^2 + 2\,\mathbb{E}(1 + |\boldsymbol{x}_o^\top \boldsymbol{\beta}^*|)^3.
\end{aligned} \tag{43}
$$

Note that $\boldsymbol{x}_o^\top \boldsymbol{\beta}^*$ is a Gaussian random variable with mean zero and variance $(\boldsymbol{\beta}^*)^\top \boldsymbol{\Sigma} \boldsymbol{\beta}^* \leq \rho b$. Hence, $\mathbb{E}\, \ell^2(y_o, \boldsymbol{x}_o^\top \boldsymbol{\beta}^*)$ is bounded by a constant.

For the second term in (38) we have

$$\begin{aligned}
\mathbb{E}(1+y_o^2)(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1} - \boldsymbol{x}_o^\top \boldsymbol{\beta}^*)^2 &= \mathbb{E}\left(1 + f(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*) + f^2(\boldsymbol{x}_o^\top \boldsymbol{\beta}^*)\right)(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1} - \boldsymbol{x}_o^\top \boldsymbol{\beta}^*)^2 \\
&\leq \mathbb{E}\left(1 + (1+|\boldsymbol{x}_o^\top \boldsymbol{\beta}^*|) + (1+|\boldsymbol{x}_o^\top \boldsymbol{\beta}^*|^2)\right)(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1} - \boldsymbol{x}_o^\top \boldsymbol{\beta}^*)^2. \qquad (44)
\end{aligned}$$

Note that in order to show that this term is bounded from above by a constant, we only need to show that terms of the form:

$$\mathbb{E}\, |\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}|^{k_1} |\boldsymbol{x}_o^\top \boldsymbol{\beta}^*|^{k_2} \leq (\mathbb{E}\, |\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}|^{2k_1}\, \mathbb{E}\, |\boldsymbol{x}_o^\top \boldsymbol{\beta}^*|^{2k_2})^{1/2}$$

are bounded for $k_1 \leq 2$ and $k_1 + k_2 \leq 4$. As previously, we note that $\boldsymbol{x}_0^\top \boldsymbol{\beta}^*$ is a Gaussian random variable with variance $\boldsymbol{\beta}^{*\top} \boldsymbol{\Sigma} \boldsymbol{\beta}^* \leq \frac{\rho}{p} \|\boldsymbol{\beta}^*\|_2^2 \leq \rho b$, and hence $(\mathbb{E}\, |\boldsymbol{x}_o^\top \boldsymbol{\beta}^*|^{2k_2})^{1/2}$ is bounded. Hence, the only remaining step is to prove the boundedness of $\mathbb{E}\, |\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}|^{2k_1}$, where k_1 is at most 2. Note that conditioned on $D_{/1}$ the random variable $\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}$ is Gaussian with the variance that is bounded by $\frac{\rho}{p} \hat{\boldsymbol{\beta}}_{/1}^\top \hat{\boldsymbol{\beta}}_{/1}$. Hence, using Lemma 2 we have

$$\mathbb{E}\, |\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}|^{2k_1} \leq (2k_1 - 1)!!\mathbb{E}\left(\frac{\rho}{p} \hat{\boldsymbol{\beta}}_{/1}^\top \hat{\boldsymbol{\beta}}_{/1}\right)^{k_1}.$$

The definition of $\hat{\boldsymbol{\beta}}_{/1}$ (and comparing it with $\boldsymbol{\beta}^*$) yields

$$\sum_{j \neq i} \ell(y_j \mid \boldsymbol{x}_j^\top \hat{\boldsymbol{\beta}}_{/1}) + \lambda r(\hat{\boldsymbol{\beta}}_{/1}) \leq \sum_{j \neq i} \ell(y_j \mid \boldsymbol{x}_j^\top \boldsymbol{\beta}^*) + \lambda r(\boldsymbol{\beta}^*),$$

The γ-strong convexity of the smoothed elastic-net regularizer r, and the fact that $\ell \geq 0$, leads to

$$\lambda \gamma \|\hat{\boldsymbol{\beta}}_{/1}\|_2^2 \leq \sum_{j \neq i} \ell(y_j \mid \boldsymbol{x}_j^\top \boldsymbol{\beta}^*) + \lambda r(\boldsymbol{\beta}^*).$$

Since $k_1 \leq 2$, we only prove that $\mathbb{E}\, |\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}|^4$ is bounded. Toward this goal we have:

$$\begin{aligned}
\mathbb{E}\Big(\frac{\lambda\gamma}{p} \|\hat{\boldsymbol{\beta}}_{/1}\|_2^2\Big)^2 &\leq \frac{1}{p^2}\, \mathbb{E}\Big(\sum_{j \neq i} \ell(y_j \mid \boldsymbol{x}_j^\top \boldsymbol{\beta}^*) + \lambda r(\boldsymbol{\beta}^*)\Big)^2 \\
&\leq \frac{2}{p^2}\, \mathbb{E}(\sum_{j \neq i} \ell(y_j \mid \boldsymbol{x}_j^\top \boldsymbol{\beta}^*))^2 + \mathbb{E}(\lambda r(\boldsymbol{\beta}^*))^2 \\
&\leq \frac{2n(n-1)}{p^2}\, \mathbb{E}\, \ell^2(y_1 \mid \boldsymbol{x}_1^\top \boldsymbol{\beta}^*) + \frac{\lambda^2 r^2(\boldsymbol{\beta}^*)}{p^2} \\
&\leq 2\delta^2\, \mathbb{E}\, \ell^2(y_1 \mid \boldsymbol{x}_1^\top \boldsymbol{\beta}^*) + \frac{\lambda^2 r^2(\boldsymbol{\beta}^*)}{p^2}. \qquad (45)
\end{aligned}$$

Hence, we have to prove that $\mathbb{E}\, \ell^2(y_1 \mid \boldsymbol{x}_1^\top \boldsymbol{\beta}^*)$ and $\frac{\lambda^2 r^2(\boldsymbol{\beta}^*)}{p^2}$ are bounded. First note we proved in (43) that:

$$\ell^2(y_o \mid \boldsymbol{x}_o^\top \boldsymbol{\beta}^*) \leq 4 + 4(1 + |\boldsymbol{x}_o^\top \boldsymbol{\beta}^*|)^2 + 2(1 + |\boldsymbol{x}_o^\top \boldsymbol{\beta}^*|)^3. \qquad (46)$$

Note that $\boldsymbol{x}_0^\top \boldsymbol{\beta}^*$ is a Gaussian random variable with variance $\boldsymbol{\beta}^{*\top} \boldsymbol{\Sigma} \boldsymbol{\beta}^* \leq \frac{\rho}{p} \|\beta^*\|_2^2 \leq \rho b$, and hence $\mathbb{E}\, \ell^2(y_o \mid \boldsymbol{x}_o^\top \boldsymbol{\beta}^*)$ is bounded. On the other hand,

$$r(\boldsymbol{\beta}^*) = \gamma(\boldsymbol{\beta}^*)^\top \boldsymbol{\beta}^* + (1-\gamma) \sum_{i=1}^p r^\alpha(\beta_i^*). \qquad (47)$$

It is straightforward to prove that $\dot{r}^\alpha(z) = \frac{e^{\alpha z} - e^{-\alpha z}}{e^{\alpha z} + e^{-\alpha z} + 1} < 1$. Hence,

$$r(\boldsymbol{\beta}^*) = \gamma(\boldsymbol{\beta}^*)^\top \boldsymbol{\beta}^* + (1-\gamma) \sum_{i=1}^p r^\alpha(\beta_i^*) < \gamma(\boldsymbol{\beta}^*)^\top \boldsymbol{\beta}^* + (1-\gamma) \sum_{i=1}^p (\frac{2 \log 2}{\alpha} + |\beta_i^*|), \qquad (48)$$

where to obtain the last inequality we used the mean value theorem

$$r^{\alpha}(|z|) = r\alpha(0) + \dot{r}^{\alpha}(\tilde{z})|z|,$$

where $\tilde{z} \in (0, |z|)$, and the facts that $\dot{r}^{\alpha}(\tilde{z}) \leq 1$ and $r^{\alpha}(0) = \frac{2\log 2}{\alpha}$. Using (48) we obtain:

$$\begin{aligned}\frac{1}{p} r(\boldsymbol{\beta}^*) &\leq \frac{\gamma(\boldsymbol{\beta}^*)^\top \boldsymbol{\beta}^*}{p} + \frac{(1-\gamma)2\log 2}{\alpha} + \frac{1-\gamma}{p}\sum_{i=1}^{p}|\beta_i^*| \\ &\leq \frac{\gamma(\boldsymbol{\beta}^*)^\top \boldsymbol{\beta}^*}{p} + \frac{(1-\gamma)2\log 2}{\alpha} + (1-\gamma)\sqrt{\frac{\sum_{i=1}^{p}|\beta_i^*|^2}{p}} \\ &\leq \gamma b + \frac{(1-\gamma)2\log 2}{\alpha} + (1-\gamma)\sqrt{b}. \end{aligned} \tag{49}$$

J Proof of Corollary 5

Similar to the proofs of Corollaries 4, 3, we would like to use Theorem 2 to prove our claim. Toward this goal, We have to prove that Assumptions 1′, 2′, and 3′ hold. Furthermore, we have to obtain an upper bound for the constant $\tilde{C}_v$, which in turn requires us to bound $\mathbb{E}\,\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]$. Again, the proofs of Assumptions 1′ and 3′ are exactly the same as we presented in the last two sections. Hence, we only focus on Assumption 2′ and $\mathbb{E}\,\mathrm{Var}[\phi(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]$. We would like to prove that the conditions of Assumption 2′ are satisfied with $\tilde{c}_0 = \tilde{c}_1 = 2^8(\kappa + \alpha^{-8})$.

It we compute the derivative of the log-likelihood, we will obtain

$$|\dot{\ell}(y \mid z)| = \left| -y + (y+\alpha^{-1})\frac{\alpha \mathrm{e}^z}{1+\alpha \mathrm{e}^z} \right| \leq y + \alpha^{-1}. \tag{50}$$

We thus deduce that:

$$\mathbb{E}|\dot{\ell}(y_1 \mid \boldsymbol{x}_1^\top \hat{\boldsymbol{\beta}})|^8 \leq \mathbb{E}(y+\alpha^{-1})^8 \leq 2^8(\kappa + \alpha^{-8}).$$

As the bound (50) is free of z, the same argument above applies to the other requirements in Assumption 2′.

Now we turn our attention to the calculation of $\mathbb{E}\,\mathrm{Var}[\ell(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}]$. Note that

$$\mathbb{E}\,\mathrm{Var}[\ell(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}) \mid D_{/1}] \leq \mathbb{E}\,\ell^2(y_o, \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}).$$

Note that by removing the constant from the log-likelihood we obtain

$$\begin{aligned}|\ell(y_o \mid \boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1})| &= |(y_o+\alpha^{-1})\log(1+\alpha \mathrm{e}^{\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}}) - y_o(\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1})| \leq |y_o + \alpha^{-1}|(1+|\log\alpha| + |\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}|) + y_o|\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}| \\ &\leq 2y_o|\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}| + \alpha^{-1}(1+|\log\alpha| + |\boldsymbol{x}_o^\top \hat{\boldsymbol{\beta}}_{/1}|). \end{aligned} \tag{51}$$

The rest of the proof is very similar to the proof that we presented for Corollary 4. Hence, we skip it.