Supplementary Material

for "Steinized Empirical Bayes Estimation for Heteroscedastic Data"

1 Additional discussion and results

1.1 Additional discussion

We discuss various issues associated with the use of SURE for selecting the tuning parameters (γ, β) in $\delta^{\mathrm{B}}_{\gamma,\beta}$ and in $\delta_{\lambda,\gamma,\beta}$, and associated with maximum likelihood estimation of (γ, β), as mentioned in Sections 3.1–3.2.

SURE tuning for $\delta^{B}_{\gamma,\beta}$. To investigate SURE tuning, we simulated 1000 data vectors Y of dimension 10 from (1), with θ the zero vector and $(d_1, \ldots, d_{10}) =$

Figure S1: A numerical example where nested optimization works properly for minimizing $\mathrm{SURE}(\delta^{\mathrm{B}}_{\gamma,\beta})$. On the top left is a plot of the 10 simulated observations; on the top right is $\mathrm{SURE}\{\delta^{\mathrm{B}}_{\gamma,\bar\beta(\gamma)}\}$ as a function of γ; on the bottom left is $\mathrm{SURE}(\delta^{\mathrm{B}}_{\gamma=0,\beta})$ and on the bottom right is $\mathrm{SURE}(\delta^{\mathrm{B}}_{\gamma=15,\beta})$, each as a function of β. The estimates $(\hat\gamma_{\mathrm{JX}}, \hat\beta_{\mathrm{JX}})$ are found correctly as (15.33, 3.32) by nested optimization, but incorrectly as (0, 0.43) by Nelder–Mead.

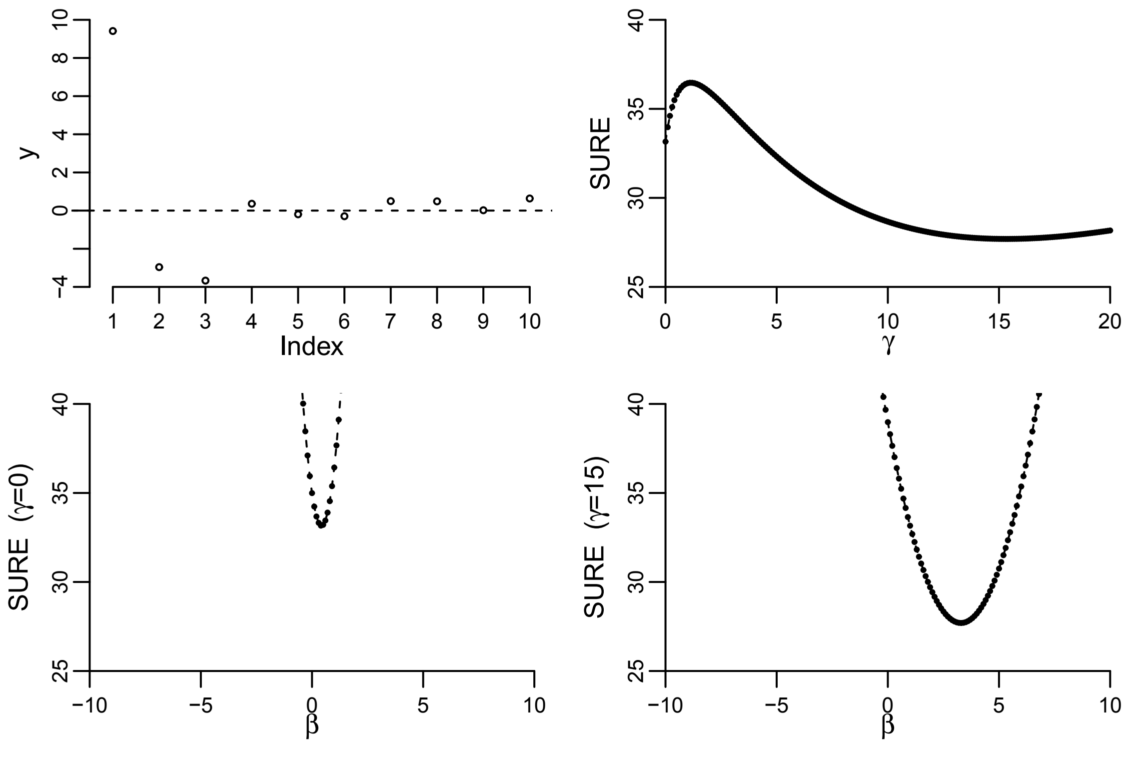
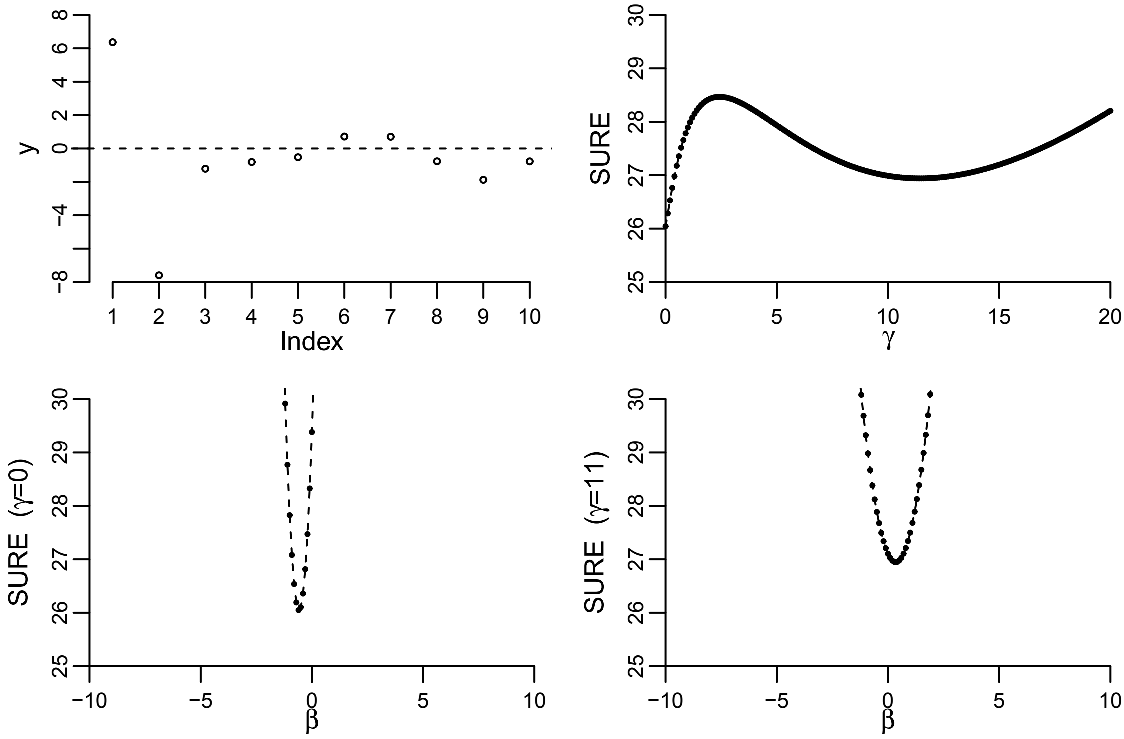


Figure S2: A numerical example where nested optimization fails for minimizing $\text{SURE}(\delta^{\text{B}}_{\gamma,\beta})$. On the top left is a plot of the 10 simulated observations; on the top right is $\text{SURE}\{\delta^{\text{B}}_{\gamma,\bar{\beta}(\gamma)}\}$ as a function of γ; on the bottom left is $\text{SURE}(\delta^{\text{B}}_{\gamma=0,\beta})$ and on the bottom right is $\text{SURE}(\delta^{\text{B}}_{\gamma=11,\beta})$, each as a function of β. The estimates $(\hat{\gamma}_{\text{JX}}, \hat{\beta}_{\text{JX}})$ are found incorrectly as (11.44, 0.39) by nested optimization, but correctly as (0, −0.58) by Nelder–Mead.



$(40, 20, 10, 1, \ldots, 1)$, where the last 7 variances are 1. For simplicity, suppose that 1 is used as the only covariate and hence both β and γ are scalars in the second-level model (2). Because the true values of θ_j are all zero, model (2) can be regarded as correctly specified, with the true values of (γ, β) being $(0, 0)$.

We computed the estimates $(\hat{\gamma}_{\text{JX}}, \hat{\beta}_{\text{JX}})$ in two ways: either minimizing $\text{SURE}\{\delta^{\text{B}}_{\gamma,\bar{\beta}(\gamma)}\}$ over γ by the one-dimensional optimization algorithm `optimize()` in R, or directly minimizing $\text{SURE}(\delta^{\text{B}}_{\gamma,\beta})$ over (γ, β) by the Nelder–Mead algorithm provided by `optim()` in R. The two methods gave different values of $\hat{\gamma}_{\text{JX}}$, by 0.1 or more, for 19 out of 1000 data vectors, in which the minimum SURE values found by the the nested optimization method are smaller for 14 data vectors, but are larger for 5 data vectors, than those from the Nelder-Mead method.

Figure S1 and S2 show two numerical examples where nested optimization correctly finds or, respectively, fails to find $(\hat{\gamma}_{\text{JX}}, \hat{\beta}_{\text{JX}})$ as a global minimizer of $\text{SURE}(\delta^{\text{B}}_{\gamma,\beta})$. For both examples, the profile function $\text{SURE}\{\delta^{\text{B}}_{\gamma,\bar{\beta}(\gamma)}\}$ is non-convex and admits both a local minimum and a local maximum over $(0, \infty)$. Then the local minimizer and the local maximizer are two solutions to equation (15).

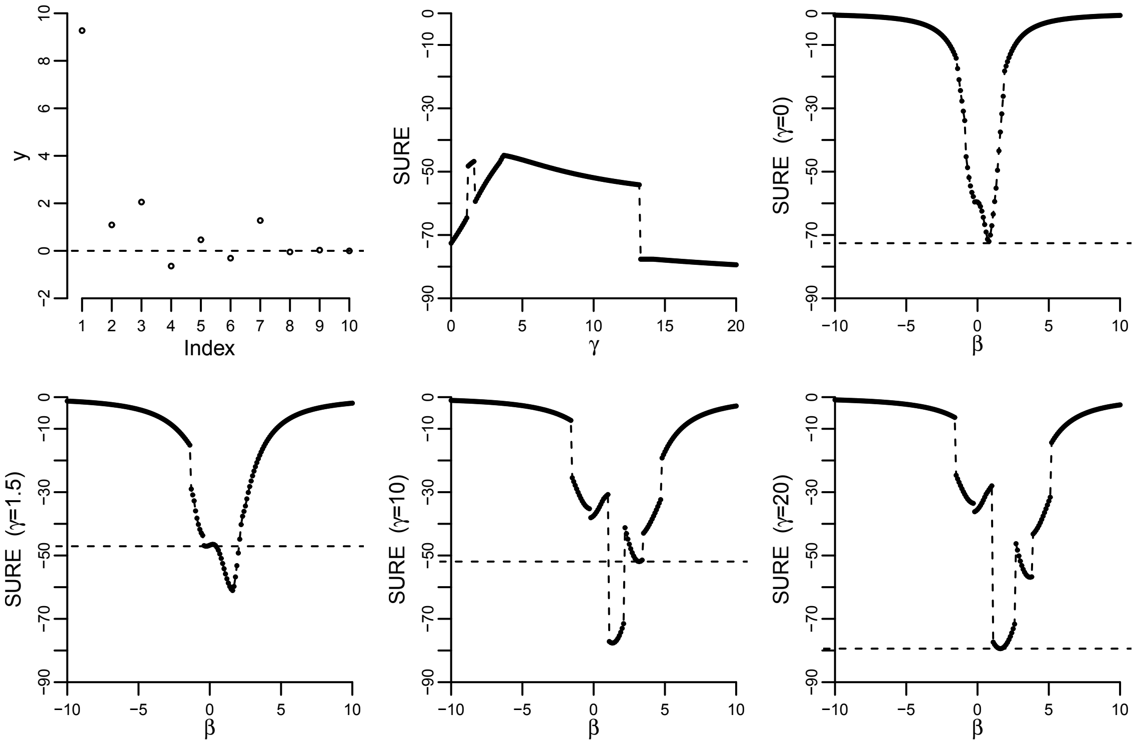
Figure S1 also serves to illustrate that the SURE-based estimates $(\hat{\gamma}_{\text{JX}}, \hat{\beta}_{\text{JX}})$ might appear unnatural in showing how the second-level model (2) could be fitted to the true values θ_j. For this example, the location estimate $\hat{\beta}_{\text{JX}} = 3.32$ is larger than 9 out of all 10 observations and overly pulled toward the single observation close to 10 with the largest variance 40, due to the fact that the observations are weighted in proportion to the variances in (14) as discussed in Section 3.2. To compensate for overestimation in $\hat{\beta}_{\text{JX}}$, the scale estimate $\hat{\gamma}_{\text{JX}} = 15.33$ is then inflated to a large extent. By comparison, the Fay–Herriot estimates $(\hat{\gamma}_{\text{FH}}, \hat{\beta}_{\text{FH}})$, found to be (0, 0.17) in this example, seems more reasonable than $(\hat{\gamma}_{\text{JX}}, \hat{\beta}_{\text{JX}})$ in reflecting the fact that the true values of θ_j are all zero. This phenomenon is reminiscent of that in the baseball example in Section 5 when 1 is used as the only covariate.

SURE tuning for $\boldsymbol{\delta_{\lambda,\gamma,\beta}}$. As mentioned in Section 3.1, it is computationally challenging to globally minimize $\text{SURE}(\delta_{\lambda,\gamma,\beta})$ as a function of (λ, γ, β). To illustrate this issue, we computed $\min_{0 \le \lambda \le 2} \text{SURE}(\delta_{\lambda,\gamma,\beta}) = \text{SURE}\{\delta_{\hat{\lambda}(\gamma,\beta),\gamma,\beta}\}$ as a function of β for fixed $\gamma \ge 0$ by the piecewise search method described in Section 3.1. Then we tried to minimize this function over $\beta \in \mathbb{R}$ for fixed $\gamma \ge 0$, by the one-dimensional optimization R algorithm `optimize()`.

Figure S3 demonstrates the complexity of $\text{SURE}(\delta_{\lambda,\gamma,\beta})$ for a particular data vector. The function $\text{SURE}\{\delta_{\hat{\lambda}(\gamma,\beta),\gamma,\beta}\}$ is non-smooth and multi-modal in β for a range of fixed $\gamma \ge 0$. As expected, minimizing this function by `optimize()` often fails to find a global minimum. Then the profile function plotted, $\min_{\beta \in \mathbb{R}} \text{SURE}\{\delta_{\hat{\lambda}(\gamma,\beta),\gamma,\beta}\}$, with the minimum over β computed by `optimize()` is incorrect. The approach of nested optimization over λ, β, and then γ does not work here.

Figure S3 also serves to illustrate a subtle issue in choosing (λ, γ, β) as a global minimizer of $\text{SURE}(\delta_{\lambda,\gamma,\beta})$, if correctly identified. For this example, the Fay–Herriot estimates are $(\hat{\gamma}_{\text{FH}}, \hat{\beta}_{\text{FH}}) = (0,\ 0.18)$, in agreement with the fact that the true values of θ_j are all zero. However, $\text{SURE}(\delta_{\lambda,\gamma,\beta})$ seems to achieve a global minimum at some β between 0 and 2 and γ greater than 20, even possibly $\gamma = \infty$. In contrast with

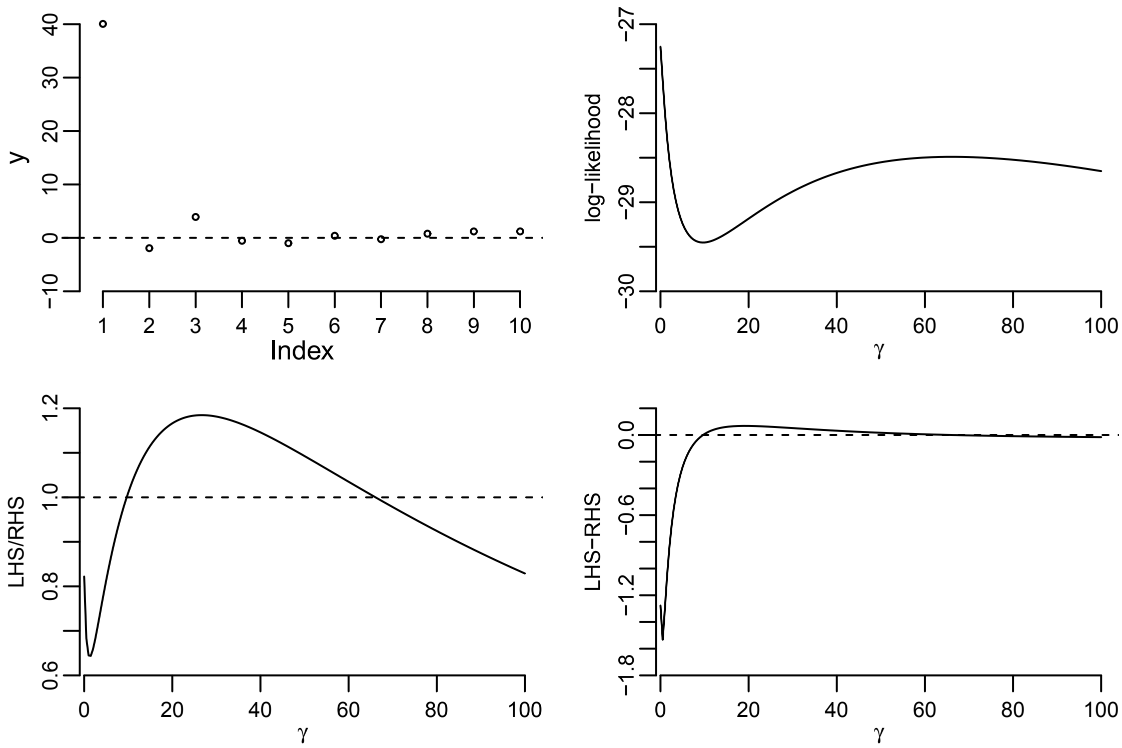
Figure S3: A numerical example illustrating the complexity of SURE($\delta_{\lambda,\gamma,\beta}$). On the top left is a plot of the 10 simulated observations; on the top middle is $\min_{\beta\in\mathbb{R}} \text{SURE}\{\delta_{\hat\lambda(\gamma,\beta),\gamma,\beta}\}$ as a function of γ, where the minimum over β is computed by the R function `optimize()` and may be a local or global minimum; the remaining four plots are $\text{SURE}\{\delta_{\hat\lambda(\gamma,\beta),\gamma,\beta}\}$ as a function of β for $\gamma = 0$, 1.5, 10, and 20 respectively, with a horizontal line placed at the local or global minimum over β found by `optimize()`.



the Bayes rule $\delta^{\text{B}}_{\gamma,\beta}$, an unusual feature of $\delta_{\lambda,\gamma,\beta}$ is that the variance parameter γ does not monotonically determine the magnitude of shrinkage, and $\delta_{\lambda,\gamma,\beta}$ remains a proper shrinkage estimator, different from $\delta_0 = Y$, at the limit as $\gamma \to \infty$. Therefore, the values of (γ, β) minimizing SURE($\delta_{\lambda,\gamma,\beta}$), similarly to $(\hat\gamma_{\text{JX}}, \hat\beta_{\text{JX}})$, might not reflect how the second-level model (2) could be properly fitted to the true values θ_j. Moreover, such choices of (γ, β) for $\delta_{\lambda,\gamma,\beta}$ can be more difficult to interpret than $(\hat\gamma_{\text{JX}}, \hat\beta_{\text{JX}})$, due to the nonstandard role of γ in the estimator $\delta_{\lambda,\gamma,\beta}$.

Maximum likelihood estimation of $(\boldsymbol{\gamma}, \boldsymbol{\beta})$**.** To investigate possible irregular behavior for score equation (13), we simulated 1000 data vectors Y of dimension 10 from (1) as before, except for $\theta = (20, 0, \ldots, 0)$ with the first element nonzero. Moreover, suppose that 1 is used as the only covariate as before. The second-level

Figure S4: A numerical example illustrating non-monotonicity associated with equation (13) for maximum likelihood estimation of γ. On the top left is a plot of the 10 simulated observations; on the top right is shown the profile log-likelihood of γ, $-\sum_{j=1}^{n}[\{Y_j - \hat{\beta}(\gamma)\}^2/(d_j+\gamma) + \log(d_j+\gamma)]/2$; on the bottom left and right are shown the difference and, respectively, the ratio between the two sides of (13).



model (2) can be seen as misspecified for the true values of θ_j.

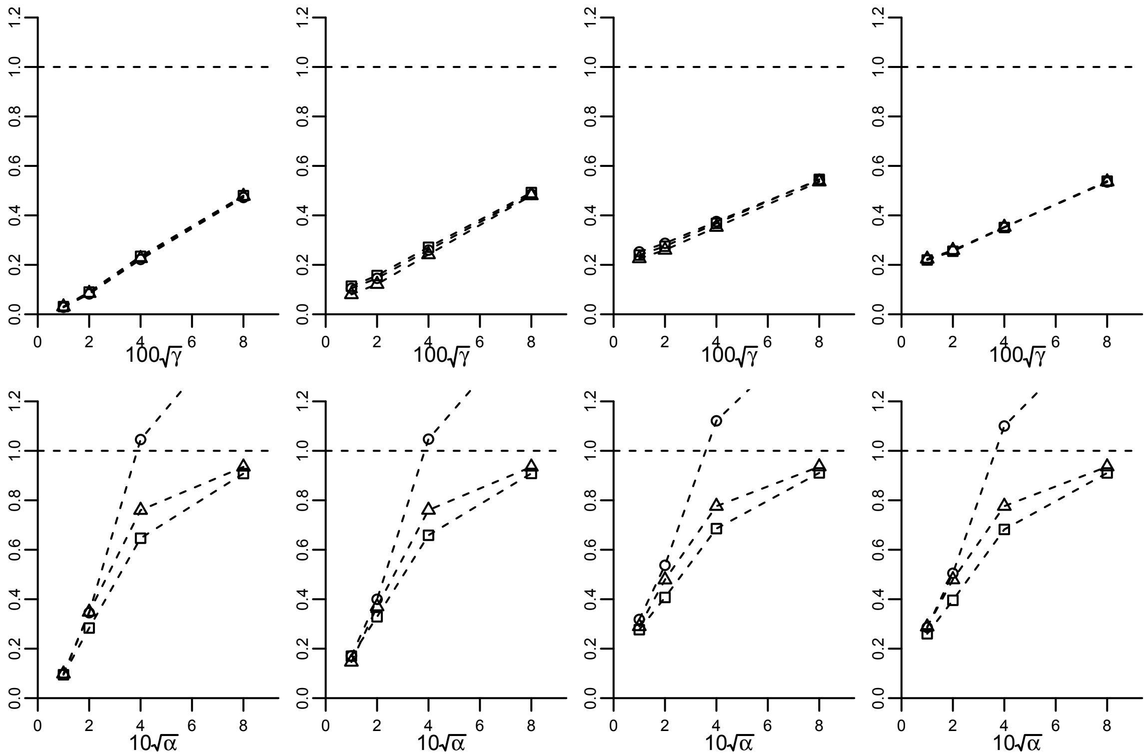
We tried to solve (13) by the following iteration:

$$\gamma_{t+1} = \sum_{j=1}^{n} \frac{\{Y_j - \hat{\beta}(\gamma_t)\}^2 - d_j}{(d_j+\gamma_t)^2} \Big/ \sum_{j=1}^{n} \frac{1}{(d_j+\gamma_t)^2}, \quad t = 0, 1, \ldots,$$

with the starting value $\gamma_0 = 0$ or 50. Comparison of the results obtained by this procedure with the two starting values indicates that there are two or more solutions to equation (13) for at least 9 out of 1000 data vectors.

Figure S4 shows the irregular behavior associated with equation (13) for a particular data vector. As mentioned in Section 3.2, neither the difference nor the ratio between the two sides of (13) is monotonic in $\gamma \geq 0$. The left-hand side of (13) is strictly less than the right-hand side at $\gamma = 0$. But there exist two solutions to (13), corresponding to a local minimizer and a local maximizer of the profile log-likelihood of γ, which has a global maximum at $\gamma = 0$.

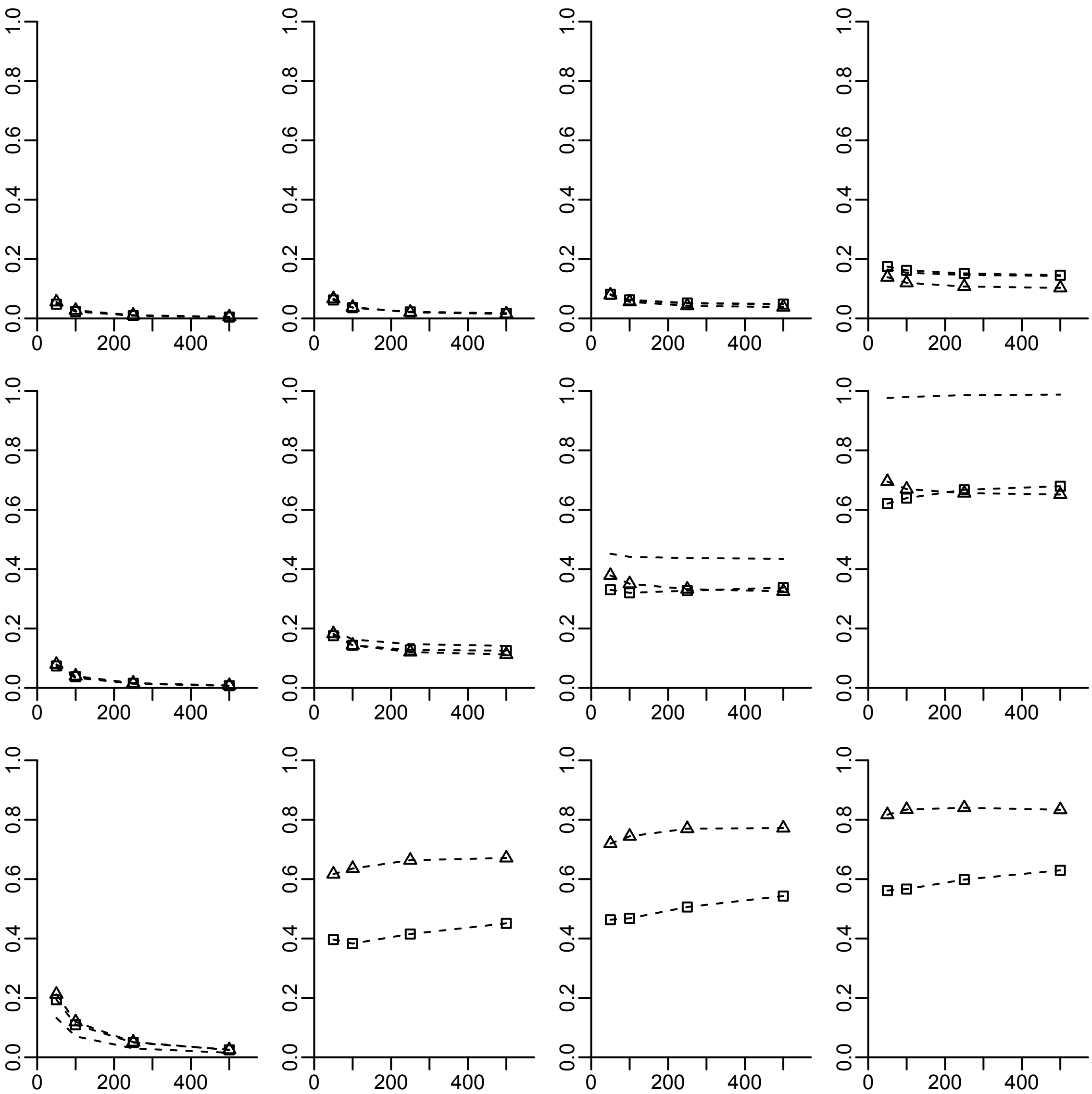
Figure S5: Relative Bayes risks of three estimators δ_{FH} ($\circ$), δ_{JX} ($\triangle$), and δ_{Res} ($\square$) similarly as in Figure 4, but based on simulated observations with a negative AB effect.



1.2 Additional simulation results

Figure S5 presents the additional simulation results mentioned at the end of Section 6. The observations are simulated from the homoscedastic prior (2) or from the heteroscedastic prior (24) similarly as in Figure 4, but the true value of β is set such that $x_j^{\text{T}}\beta = 0.5 - 0.0002(\text{AB}_j) - 0.1(\text{pitcher}_j)$.

Figure S6: Relative Bayes risks of three estimators δ_{FH} ($\circ$), δ_{JX} ($\triangle$), and δ_{Res} ($\square$) using the covariates $1+d$ (first column), $1+d^{-1}+d^{-2}$ (second column), $1+d^{-1}$ (third column), and 1 (fourth column), based on n = 50, 100, 250, and 500 observations from model (1) with $\theta_j = d_j$ and $d_j \sim \text{unif}(0.1, 1)$ (first row), $1/d_j \sim \text{unif}(0.1, 1)$ (second row), and $1/d_j \sim \text{unif}(0.1, 10)$ (third row). For the third row, the relative Bayes risks of δ_{FH} ($\circ$) are about 1.1, 1.4, and 1.7 respectively in the second to fourth plots.



We conducted additional simulations in the following settings where $Y_j \sim \text{N}(\theta_j, d_j)$, but the individual means and variances are highly dependent: $\theta_j = d_j$, and the individual variances are randomly generated: (i) $d_j \sim \text{unif}(0.1, 1)$, (ii) $1/d_j \sim \text{unif}(0.1, 1)$, and $1/d_j \sim \text{unif}(0.1, 10)$. Setting (i) is taken directly from Xie et al. (2012), and the other two settings (ii)–(iii) are introduced to allow greater variability in the variances $(d_1, \ldots, d_n)$. Note that $1/d_j$ can be interpreted as being proportional to the sample size, such as AB_j in the baseball example, underlying the observation Y_j.

For application of the estimators δ_{FH}, δ_{JX}, and δ_{Res}, the second-level model (2) is used, with x_j possibly depending on d_j: (i) $x_j = (1, d_j)^{\mathrm{T}}$, (ii) $x_j = (1, d_j^{-1}, d_j^{-2})^{\mathrm{T}}$, (iii) $x_j = (1, d_j^{-1})^{\mathrm{T}}$, and (iv) $x_j \equiv 1$. The first choice of x_j leads to a correctly specified model (2). The other three choices lead to a mean misspecification in (2), although the variance can be seen as correctly specified with $\gamma = 0$.

Figure S6 shows the relative Bayes risks of δ_{FH}, δ_{JX}, and δ_{Res} versus the naive estimator, obtained from 10^4 repeated simulations with sample sizes from $n = 50$ to 500. In the first setting where $d_j \sim \mathrm{unif}(0.1, 1)$, all the three estimators perform similarly to each other, except that δ_{JX} yields smaller Bayes risks than δ_{FH} and δ_{Res} when 1 is used as the only covariate. However, as the variances $(d_1, \ldots, d_n)$ become more variable in the second and third settings, the estimator δ_{Res} performs increasingly better than both δ_{FH} and δ_{JX} when the prior mean in (2) is misspecified. This comparison demonstrates a potential advantage of δ_{Res} over δ_{FH} and δ_{JX}, in the case where the individual variances $(d_1, \ldots, d_n)$ are highly variable.

2 Technical details

Throughout, a summation over an empty index set is 0; the index j or k runs from 1 to n in all summations unless otherwise stated; and $\bar{\lambda} = 2$.

2.1 Proof of Lemma 1

Sort the indices such that $d_1 \geq d_2 \geq \cdots \geq d_n$. If $1 \leq j \leq k \leq \nu$, then $a_j(\gamma)/a_k(\gamma) = d_k/d_j$. If $\nu + 1 \leq j \leq k \leq n$, then $1 \leq a_j(\gamma)/a_k(\gamma) = (d_j/d_k)\{(d_k + \gamma)/(d_j + \gamma)\} \leq d_j/d_k$. If $1 \leq j \leq \nu$ and $\nu + 1 \leq k \leq n$, then $a_j(\gamma)/a_k(\gamma) \leq d_j/(d_j + \gamma)/a_k(\gamma) \leq d_j/d_k$ by Corollary 3 in Tan (2014) and, moreover, $a_k(\gamma)/a_j(\gamma) \leq a_{\nu+1}(\gamma)/a_j(\gamma) \leq d_j/d_{\nu+1} \leq d_j/d_k$, because $d_{\nu+1}a_{\nu+1}(\gamma) < d_\nu a_\nu(\gamma)$ by the definition of ν and hence $a_{\nu+1}(\gamma)/a_j(\gamma) \leq (d_\nu a_\nu/d_{\nu+1})/a_j(\gamma) = d_j/d_{\nu+1}$. $\square$

2.2 Proof of Theorem 1

By the independence of $H_D Y$ and $(I - H_D)Y$, the unbiasedness of $H_D Y$ for $H_D\theta$, and the relationship that $L_2 L_2^{\mathrm{T}}(I - H_D) = L_2 V_2 L_2^{\mathrm{T}} D^{-1} = I - H_D$, we have

$$
\begin{aligned}
R(\delta^{\mathrm{S}}_{\lambda,\gamma}, \theta) &= \mathrm{tr}\left[E_\theta\left\{(\delta^{\mathrm{S}}_{\lambda,\gamma} - \theta)^{\otimes 2}\right\}\right] \\
=&\mathrm{tr}\left(E_\theta\left[\{H_d(Y-\theta)\}^{\otimes 2}\right]\right) + \mathrm{tr}\left\{E_\theta\left(\left[L_2\left\{\delta_{\lambda,\gamma,\beta=0}(\eta_2) - \psi_2\right\}\right]^{\otimes 2}\right)\right\} \\
=&\mathrm{tr}(H_D D H_D^{\mathrm{T}}) + \mathrm{tr}\left(E_\theta\left[\left\{\delta_{\lambda,\gamma,\beta=0}(\eta_2) - \psi_2\right\}^{\otimes 2}\right]\right),
\end{aligned}
$$

leading directly to the desired equation. Throughout, $y^{\otimes 2}$ denotes yy^{T} for a column vector y. Moreover, result (10) can extended as follows on the Bayes risk of $\delta^{\mathrm{S}}_{\lambda=1,\gamma}$. Let $\Gamma_\alpha = \alpha(V_2 + \gamma I_2) - V_2$ and $\alpha_0 = \max_{j=1,\ldots,n-q}\{v_j/(v_j+\gamma)\}$, with I_2 the $(n-q)\times(n-q)$ identity matrix and $(v_1, \ldots, v_{n-q})$ the diagonal elements of V_2. For $\alpha \geq \alpha_0$, consider the estimator of θ,

$$
\delta^{\mathrm{SB}}_{\alpha,\gamma} = X\tilde{\beta} + L_2\, \delta^{\mathrm{B}}_{\alpha,\gamma,\beta=0}\{L_2^{\mathrm{T}}(Y - X\tilde{\beta})\},
$$

where $\delta^{\mathrm{B}}_{\alpha,\gamma,\beta=0}(\eta_2)$ is the Bayes rule with the observation vector $\eta_2 = L_2^{\mathrm{T}}(Y - X\tilde{\beta})$ for estimating $\psi_2 = L_2^{\mathrm{T}}(I - H_D)\theta$, under the prior $\psi_2 \sim \mathrm{N}(0, \Gamma_\alpha)$. Similarly as above, the pointwise risk of $\delta^{\mathrm{SB}}_{\alpha,\gamma}$ is related to that of $\delta^{\mathrm{B}}_{\alpha,\gamma,\beta=0}(\eta_2)$ by

$$
R(\delta^{\mathrm{SB}}_{\alpha,\gamma}, \pi) = \mathrm{tr}\{X(X^{\mathrm{T}}D^{-1}X)^{-1}X^{\mathrm{T}}\} + R\{\delta^{\mathrm{B}}_{\alpha,\gamma,\beta=0}(\eta_2), \psi_2\}.
$$

Then for any prior π on θ such that $\psi_2 = L_2^{\mathrm{T}}(I - H_D)\theta \sim \mathrm{N}(0, \Gamma_\alpha)$ with $\alpha \geq \alpha_0$, the Bayes risk of $\delta^{\mathrm{S}}_{\lambda=1,\gamma}$ satisfies

$$R(\delta^{\mathrm{S}}_{\lambda=1,\gamma}, \pi) \leq R(\delta^{\mathrm{SB}}_{\alpha,\gamma}, \pi) + \alpha^{-1}(v_1^* + v_2^* + v_3^* + v_4^*),$$

where $v_j^* = v_j^2/(v_j + \gamma)$ and the indices are sorted such that $v_1^* \geq v_2^* \geq \cdots \geq v_{n-q}^*$. $\square$

2.3 Proof of Theorem 2

Recall that $\delta^{\mathrm{B}}_{\gamma,\beta} = Y_j - a_j^*(Y_j - x_j^{\mathrm{T}}\beta)$. Then

$$\begin{aligned} R(\delta^{\mathrm{B}}_{\gamma,\beta}, \theta) &= \sum_j d_j + \sum_j a_j^{*2}\{d_j + (\theta_j - x_j^{\mathrm{T}}\beta)^2\} - 2\sum_j a_j^* d_j \\ &\geq \sum_j d_j - \frac{(\sum_j a_j^* d_j)^2}{\sum_j a_j^{*2}\{d_j + (\theta_j - x_j^{\mathrm{T}}\beta^*)^2\}}. \end{aligned}$$

by minimization of the quadratic function $\sum_j d_j + \lambda^2 \sum_j a_j^{*2}\{d_j + (\theta_j - x_j^{\mathrm{T}}\beta)^2\} - 2\lambda \sum_j a_j^* d_j$ in λ. By inequality (4) and Jensen's inequality, we have

$$\begin{aligned} R(\delta_{\lambda=1,\gamma^*,\beta^*}, \theta) &\leq \sum_j d_j - \frac{\{\sum_j d_j a_j(\gamma^*) - 2\max_j d_j a_j(\gamma^*)\}^2}{\sum_j a_j^2(\gamma^*)\{d_j + (\theta_j - x_j^{\mathrm{T}}\beta^*)^2\}} \\ &\leq \sum_j d_j - \frac{\{\sum_j d_j a_j(\gamma^*) - 2\max_j d_j a_j(\gamma^*)\}^2}{\sum_j a_j^{*2}\{d_j + (\theta_j - x_j^{\mathrm{T}}\beta^*)^2\}} \end{aligned}$$

because $a_j(\gamma^*) \leq a_j^*$ for $j = 1, \ldots, n$ by Tan (2014, Corollary 3). Note that $(\max_j d_j a_j^*)/(\sum_j d_j a_j^*) \leq n^{-1}(\max_j d_j^2)/(\min_j d_j^2) = o(1)$. For n sufficiently large, we have $\sum_j d_j a_j^* - 2\max_j(d_j a_j^*) > 0$ and hence by the construction of $a_j(\gamma^*)$,

$$\frac{\{\sum_j d_j a_j(\gamma^*) - 2\max_j d_j a_j(\gamma^*)\}^2}{\sum_j a_j^2(\gamma^*)(d_j + \gamma^*)} \geq \frac{(\sum_j d_j a_j^* - 2\max_j d_j a_j^*)^2}{\sum_j a_j^{*2}(d_j + \gamma^*)}.$$

By the proof of Tan (2014, Theorem 3), we have

$$\frac{\sum_j a_j^2(\gamma^*)(d_j + \gamma^*)}{\sum_j a_j^{*2}(d_j + \gamma^*)} \geq 1 - 4\frac{\max_j d_j a_j^*}{\sum_j d_j a_j^*}.$$

By simple manipulation, we have

$$(\sum_j d_j a_j^* - 2\max_j d_j a_j^*)^2 \geq \left(1 - 4\frac{\max_j d_j a_j^*}{\sum_j d_j a_j^*}\right)(\sum_j d_j a_j^*)^2.$$

Combining the preceding four inequalities shows that

$$R(\delta_{\lambda=1,\gamma^*,\beta^*},\theta) \le \sum_j d_j - \left(1 - 4\frac{\max_j d_j a_j^*}{\sum_j d_j a_j^*}\right)^2 \frac{(\sum_j a_j^* d_j)^2}{\sum_j a_j^{*2}\{d_j + (\theta_j - x_j^{\mathrm{T}}\beta^*)^2\}}$$
$$\le \sum_j d_j - \left(1 - 8\frac{\max_j d_j a_j^*}{\sum_j d_j a_j^*}\right) \frac{(\sum_j a_j^* d_j)^2}{\sum_j a_j^{*2}\{d_j + (\theta_j - x_j^{\mathrm{T}}\beta^*)^2\}}.$$

By the Cauchy–Schwartz inequality $(\sum_j d_j a_j^{*2})(\sum_j d_j) \ge (\sum_j d_j a_j^*)^2$, we have

$$R(\delta_{\lambda=1,\gamma^*,\beta^*},\theta) \le \left(1 + \frac{8}{n}\frac{\max_j d_j^2}{\min_j d_j^2}\right)\sum_j d_j - \frac{(\sum_j a_j^* d_j)^2}{\sum_j a_j^{*2}\{d_j + (\theta_j - x_j^{\mathrm{T}}\beta^*)^2\}},$$

which immediately leads to the desired inequalities. □

2.4 Proof of Theorem 3

We provide a proof of Theorem 3, based on Lemmas 1–3. For $\delta_{A,\lambda}$ with a data-independent choice of A, direct calculation yields

$$\mathrm{SURE}(\delta_{A,\lambda}) = \sum_j d_j + \sum_{j \notin J}(Y_j^2 - 2d_j)$$
$$+ \sum_{j \in J}\left\{\frac{\lambda^2 c^2\, a_j^2 Y_j^2}{(\sum_k a_k^2 Y_k^2)^2} - 2\frac{\lambda c\, d_j a_j}{\sum_k a_k^2 Y_k^2} + 4\frac{\lambda c\, d_j a_j^3 Y_j^2}{(\sum_k a_k^2 Y_k^2)^2}\right\},$$

where $J = \{1 \le j \le n : \sum_k a_k^2 x_k^2 > \lambda\, c\, a_j\}$ and $c = (\sum_j d_j a_j) - 2\max_j(d_j a_j)$.

Lemma 2. Write $Q_n = \sum_j a_j^2(\theta_j^2 + d_j)$. Under Assumption (A1), the following results hold for any constant $v_n > 0$:

$$\sup_{\theta\in\mathbb{R}^n} P\left\{n^{-1}\sup_{0\le\lambda\le\bar\lambda}\left|\sum_{j\in J}\lambda c\, a_j(d_j - \varepsilon_j^2)\right| \ge v_n Q_n/2\right\} \le \frac{4K_1\bar\lambda^2(\max_j d_j)(\sum_j d_j)}{n^2 v_n^2},$$

$$\sup_{\theta\in\mathbb{R}^n} P\left(n^{-1}\sup_{0\le\lambda\le\bar\lambda}\left|\sum_{j\in J}\lambda c\, a_j\theta_j\varepsilon_j\right| \ge v_n Q_n/2\right) \le \frac{4\bar\lambda^2(\max_j d_j)(\sum_j d_j)}{n^2 v_n^2},$$

$$\sup_{\theta\in\mathbb{R}^n} P\left(\sum_k a_k^2 Y_k^2 \le Q_n/2\right) \le \frac{(32 + 8K_1)\max_k(a_k^2 d_k)}{\sum_k a_k^2 d_k}.$$

Proof of Lemma 2. Sort the indices such that $a_1 \ge a_2 \ge \cdots \ge a_n$. To show the first inequality, we have, for all $\theta \in \mathbb{R}^n$,

$$P\left\{n^{-1}\sup_{0\le\lambda\le\bar\lambda}\left|\sum_{j\in J}\lambda c\, a_j(d_j - \varepsilon_j^2)\right| \ge v_n Q_n/2\right\}$$

$$\leq P\left\{n^{-1}\max_k\left|\sum_{j=k}^n \bar\lambda c\, a_j(d_j-\varepsilon_j^2)\right|\geq v_nQ_n/2\right\}$$
$$\leq \frac{\mathrm{var}(n^{-1}\sum_j \bar\lambda c\, a_j\varepsilon_j^2)}{v_n^2Q_n^2/4}\leq \frac{K_1\bar\lambda^2c^2\sum_j a_j^2d_j^2}{n^2\,v_n^2Q_n^2/4}\leq \frac{4K_1\bar\lambda^2(\max_j d_j)(\sum_j d_j)}{n^2\,v_n^2},$$

by Kolmogorov's maximal inequality, Assumption (A1), and $c^2 \leq (\sum_j a_jd_j)^2 \leq (\sum_j a_j^2d_j)(\sum_j d_j)$. To show the second inequality, we have, for all $\theta\in\mathbb{R}^n$,

$$P\left(n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}\left|\sum_{j\in J}\lambda c\, a_j\theta_j\varepsilon_j\right|\geq v_nQ_n/2\right)$$
$$\leq P\left(n^{-1}\max_k\left|\sum_{j=k}^n \bar\lambda c\, a_j\theta_j\varepsilon_j\right|\geq v_nQ_n/2\right)$$
$$\leq \frac{\mathrm{var}(n^{-1}\sum_j \bar\lambda c\, a_j\theta_j\varepsilon_j)}{v_n^2Q_n^2/4}=\frac{4\bar\lambda^2c^2\sum_j a_j^2\theta_j^2d_j}{n^2\,v_n^2Q_n^2}\leq\frac{4\bar\lambda^2(\max_j d_j)(\sum_j d_j)}{n^2\,v_n^2}.$$

To show the third inequality, we have, for all $\theta\in\mathbb{R}^n$,

$$P\left(\sum_k a_k^2Y_k^2\leq Q_n/2\right)\leq P\left(\left|\sum_k a_k^2Y_k^2-Q_n\right|\geq Q_n/2\right)$$
$$\leq\frac{\mathrm{var}(\sum_k a_k^2Y_k^2)}{Q_n^2/4}\leq\frac{8\sum_k a_k^4\theta_k^2d_k+2K_1\sum_k a_k^4d_k^2}{Q_n^2/4}\leq\frac{(32+8K_1)\max_k(a_k^2d_k)}{Q_n},$$

by Chebyshev's inequality, Assumption (A1), and $\mathrm{var}(\sum_k a_k^2Y_k^2)=\mathrm{var}(2\sum_k a_k^2\theta_k\varepsilon_k+\sum_k a_k^2\varepsilon_k^2)\leq 2\mathrm{var}(2\sum_k a_k^2\theta_k\varepsilon_k)+2\mathrm{var}(\sum_k a_k^2\varepsilon_k^2)$. □

Lemma 3. Write $R_n=(\max_k a_k)/(\min_k a_k)$. Under Assumption (A1), the following results hold for any constant $v_n>0$:

$$\sup_{\theta\in\mathbb{R}^n}P\left\{n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}\left|\sum_{j\notin J}(d_j-\varepsilon_j^2)\right|\geq v_n\right\}\leq\frac{K_1\sum_j d_j^2}{n^2\,v_n^2},$$
$$\sup_{\theta\in\mathbb{R}^n}P\left(n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}\left|\sum_{j\notin J}\theta_j\varepsilon_j\right|\geq v_n\right)\leq\max\left\{\frac{(32+8K_1)\max_k(a_k^2d_k)}{\sum_k a_k^2d_k},\frac{2\bar\lambda R_n^3(\max_k d_k)(\sum_k d_k)}{n^2v_n^2}\right\}.$$

Proof of Lemma 3. Sort the indices such that $a_1\geq a_2\geq\cdots\geq a_n$. The first inequality can be shown similarly to the first inequality in Lemma 2. To show the second inequality, we have

$$P\left(n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}\left|\sum_{j\notin J}\theta_j\varepsilon_j\right|\geq v_n\right)$$

$$\leq P\left\{\sum_k a_k^2 Y_k^2 \leq \bar\lambda c(\max_k a_k)\right\} \leq P\left\{\sum_k a_k^2 Y_k^2 \leq \bar\lambda R_n \sum_k a_k^2 d_k\right\},$$

because if there exists some $j \notin J$ then $\sum_k a_k^2 Y_k^2 \leq \lambda c(\max_k a_k)$. If $\sum_k a_k^2\theta_k^2 > 2\bar\lambda R_n \sum_k a_k^2 d_k$, then, from the preceding inequality, $P(\sup_{0\leq\lambda\leq\bar\lambda} n^{-1}|\sum_{j\notin J}\theta_j\varepsilon_j| \geq v_n) \leq P(\sum_j a_j^2 Y_j^2 \leq Q_n/2) \leq (32+8K_1)\max_k(a_k^2 d_k)/\sum_k a_k^2 d_k$ by the third inequality in Lemma 2. On the other hand, we have

$$P\left(n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}\left|\sum_{j\notin J}\theta_j\varepsilon_j\right| \geq v_n\right) \leq P\left(n^{-1}\max_k\left|\sum_{j=1}^k \theta_j\varepsilon_j\right| \geq v_n\right)$$
$$\leq \frac{\mathrm{var}(\sum_j \theta_j\varepsilon_j)}{n^2 v_n^2} = \frac{\sum_j \theta_j^2 d_j}{n^2 v_n^2} \leq \frac{(\max_j d_j)(\sum_j a_j^2\theta_j^2)}{n^2 v_n^2(\min_j a_j^2)},$$

by Kolmogorov's inequality. If $\sum_k a_k^2\theta_k^2 \leq 2\bar\lambda R_n \sum_k a_k^2 d_k$, then, from the preceding inequality, $P(\sup_{0\leq\lambda\leq\bar\lambda} n^{-1}|\sum_{j\notin J}\theta_j\varepsilon_j| \geq v_n) \leq 2\bar\lambda R_n^3(\max_k d_k)(\sum_k d_k)/(n^2 v_n^2)$. Combining the two cases gives the desired inequality. □

Lemma 4. Write $Z_{n,1} = n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}|\sum_{j\in J}\lambda c\, a_j(d_j - \varepsilon_j Y_j)/\sum_k a_k^2 Y_k^2|$, $Z_{n,2} = n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}|\sum_{j\notin J}(d_j - \varepsilon_j Y_j)|$, and $Z_{n,3} = n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}\{\lambda c\max_{j\in J}(d_j a_j)/\sum_k a_k^2 Y_k^2\}$. Then the following results hold:

$$Z_{n,1} \leq n^{-1}\sum_{j\in J} d_j + \bar\lambda^{1/2} R_n^{1/2}(n^{-1}\sum_j d_j)^{1/2}(n^{-1}\sum_j \varepsilon_j^2)^{1/2},$$
$$Z_{n,2} \leq n^{-1}\sum_{j\notin J} d_j + \bar\lambda^{1/2} R_n(n^{-1}\sum_j d_j)^{1/2}(n^{-1}\sum_j \varepsilon_j^2)^{1/2},$$
$$Z_{n,3} \leq n^{-1}(\max_j d_j).$$

Proof of Lemma 4. The third inequality follows because $\sum_k a_k^2 Y_k^2 > \lambda c a_j$ for $j \in J$. The first inequality follows because

$$\left|\sum_{j\in J}\frac{\lambda c\, a_j\varepsilon_j Y_j}{\sum_k a_k^2 Y_k^2}\right| \leq \frac{\bar\lambda c(\sum_{j\in J} a_j^2 Y_j^2)^{1/2}(\sum_{j\in J}\varepsilon_j^2)^{1/2}}{\sum_k a_k^2 Y_k^2}$$
$$\leq \frac{(\bar\lambda c)^{1/2}(\sum_j \varepsilon_j^2)^{1/2}}{\min_j(a_j^{1/2})} \leq \bar\lambda^{1/2} R_n^{1/2}\left(\sum_j d_j\right)^{1/2}\left(\sum_j \varepsilon_j^2\right)^{1/2},$$

by the Cauchy–Schwartz inequality and the fact that if $J \neq \emptyset$ then $\sum_k a_k^2 Y_k^2 > \lambda c(\min_k a_k)$. The second inequality follows because

$$\left|\sum_{j\notin J}\varepsilon_j Y_j\right| \leq \left\{\sum_{j\notin J}\left(\frac{\varepsilon_j}{a_j}\right)^2\right\}^{1/2}\left(\sum_{j\notin J} a_j^2 Y_j^2\right)^{1/2}$$

$$\le \frac{(\sum_j \varepsilon_j^2)^{1/2}}{\min_j a_j} \left\{\lambda c (\max_j a_j)\right\}^{1/2} \le \bar{\lambda}^{1/2} R_n \left(\sum_j d_j\right)^{1/2} \left(\sum_j \varepsilon_j^2\right)^{1/2},$$

by the Cauchy–Schwartz inequality and the fact that if there exists some $j \notin J$ then $\sum_k a_k^2 Y_k^2 \le \lambda c(\max_k a_k)$. □

Proof of Theorem 3. By direct calculation, we have

$$\sum_j \{(\delta_{A,\lambda})_j - \theta_j\}^2 = \sum_j \varepsilon_j^2 + \sum_{j \notin J} \left(Y_j^2 - 2\varepsilon_j Y_j\right) + \sum_{j \in J} \left\{ \frac{\lambda^2 c^2 a_j^2 Y_j^2}{(\sum_k a_k^2 Y_k^2)^2} - 2 \frac{\lambda c\, a_j \varepsilon_j Y_j}{\sum_k a_k^2 Y_k^2} \right\},$$

and hence

$$|\zeta_n(\lambda)| \le n^{-1} \left| \sum_j (d_j - \varepsilon_j^2) \right| + 2n^{-1} \left| \sum_{j \notin J} (d_j - \varepsilon_j Y_j) \right| + 2n^{-1} \left| \sum_{j \in J} \frac{\lambda c\, a_j (d_j - \varepsilon_j Y_j)}{\sum_k a_k^2 Y_k^2} \right| + 4n^{-1} \frac{\lambda c \, \max_{j \in J}(d_j a_j)}{\sum_k a_k^2 Y_k^2}.$$

(i) Take $v_n = (\tau_2/13) n^{-(1-4\eta)/2}$ and $n v_n^2 = (\tau_2/13)^2 n^{4\eta}$. By the triangle inequality,

$$\begin{aligned} P\left\{ \sup_{0 \le \lambda \le \bar{\lambda}} |\zeta_n(\lambda)| \ge \tau_2 n^{-(1-4\eta)/2} \right\} &\le P\left\{ n^{-1} \left| \sum_j (d_j - \varepsilon_j^2) \right| \ge v_n \right\} \\ &+ P\left\{ (nQ_n/2)^{-1} \sup_{0 \le \lambda \le \bar{\lambda}} \left| \sum_{j \in J} \lambda c\, a_j (d_j - \varepsilon_j^2) \right| \ge v_n \right\} \\ &+ P\left\{ (nQ_n/2)^{-1} \sup_{0 \le \lambda \le \bar{\lambda}} \left| \sum_{j \in J} \lambda c\, a_j \theta_j \varepsilon_j \right| \ge v_n \right\} + P\left(\sum_k a_k^2 Y_k^2 \le Q_n/2 \right) \\ &+ P\left\{ n^{-1} \sup_{0 \le \lambda \le \bar{\lambda}} \left| \sum_{j \notin J} (d_j - \varepsilon_j^2) \right| \ge v_n \right\} + P\left(n^{-1} \sup_{0 \le \lambda \le \bar{\lambda}} \left| \sum_{j \notin J} \theta_j \varepsilon_j \right| \ge v_n \right) \\ &+ P\left\{ n^{-1} \frac{\lambda c \, \max_{j \in J}(d_j a_j)}{\sum_k a_k^2 Y_k^2} \ge v_n \right\}. \end{aligned}$$

By Lemmas 1–2 and the third inequality in Lemma 4, $\sup_{\theta \in \mathbb{R}^n} P\{\sup_{0 \le \lambda \le \bar{\lambda}} |\zeta_n(\lambda)| \ge \tau_2 n^{-(1-4\eta)/2}\} = O\{n^{\eta} (n v_n^2)^{-1} + n^{3\eta} n^{-1} + n^{4\eta} (n v_n^2)^{-1}\} = O(\tau_2^{-2})$.

(ii) Recall the definitions of $Z_{n,1}$, $Z_{n,2}$, and $Z_{n,3}$ in Lemma 4. Then $\sup_{0 \le \lambda \le \bar{\lambda}} |\zeta_n(\lambda)| \le n^{-1} |\sum_j (d_j - \varepsilon_j^2)| + Z_{n,1} + Z_{n,2} + Z_{n,3}$. Note that $E^{1/2}[\{n^{-1} \sum_j (d_j - \varepsilon_j^2)\}^2] \le n^{-1} K_1^{1/2} (\sum_j d_j^2)^{1/2} = O\{n^{-(1-\eta)/2}\}$ and, by Lemma 4, $Z_{n,3} \le n^{-1}(\max_j d_j) = O\{ n^{-(1-\eta)}\}$. To complete the proof, we show below that $E(Z_{n,1}) = O\{n^{-(1-4\eta)/2}\}$ and $E(Z_{n,2}) = O\{n^{-(1-5\eta)/2}\}$ uniformly in $\theta \in \mathbb{R}^n$.

Sort the indices such that $a_1 \geq a_2 \geq \cdots \geq a_n$. Write $B_{n,1} = \{\sum_k a_k^2 Y_k^2 \leq Q_n/2\}$. By Doob's L^2-maximal inequality, the first inequality in Lemma 3, and similar calculation to the proof of Lemma 2, we have, for all $\theta \in \mathbb{R}^n$,

$$
\begin{aligned}
&E(Z_{n,1}) = E(Z_{n,1} 1_{B_{n,1}^c}) + E(Z_{n,1} 1_{B_{n,1}}) \\
\leq & E^{1/2}\left[\left\{ n^{-1} \max_k \left| \sum_{j=k}^n \bar{\lambda} c\, a_j (d_j - \varepsilon_j Y_j) \right| / (Q_n/2) \right\}^2 \right] + E^{1/2}(Z_{n,1}^2) P^{1/2}(B_{n,1}) \\
\leq & \left\{ \frac{2(1+K_1)\bar{\lambda}^2 (\max_j d_j)(\sum_j d_j)}{n^2/4} \right\}^{1/2} + \{2(1+\bar{\lambda} R_n)\}^{1/2} (n^{-1} \sum_j d_j) P^{1/2}(B_{n,1}) \\
\leq & C_1 \left\{ n^{-(1-\eta)/2} + n^{\eta/2} n^{-(1-3\eta)/2} \right\},
\end{aligned} \tag{S1}
$$

where C_1 is a constant (free of θ). Therefore, $\sup_{\theta \in \mathbb{R}^n} E(Z_{n,1}) = O\{n^{-(1-4\eta)/2}\}$. By Doob's inequality and the proof of Lemma 3, we have

$$
\begin{aligned}
E(Z_{n,2}) &\leq E^{1/2}\left[\left\{ n^{-1} \max_k \left| \sum_{j=1}^k (d_j - \varepsilon_j Y_j) \right| \right\}^2 \right] \\
&\leq \left\{ \frac{2(K_1 \sum_j d_j^2 + \sum_j \theta_j^2 d_j)}{n^2} \right\}^{1/2}.
\end{aligned} \tag{S2}
$$

If $\sum_k a_k^2 \theta_k^2 \leq 2\bar{\lambda} R_n \sum_k a_k^2 d_k$, then, similarly as in the proof of Lemma 3, $\sum_j \theta_j^2 d_j \leq 2\bar{\lambda} R_n^3 (\max_k d_k)(\sum_k d_k)$ and hence $E(Z_{n,2}) \leq C_2\{n^{-(1-\eta)/2} + n^{-(1-4\eta)/2}\}$, where C_2 is a constant (free of θ). Moreover, write $B_{n,2} = \{\sum_k a_k^2 Y_k^2 \leq \bar{\lambda} c(\max_k a_k)\}$. By the second inequality in Lemma 4, we have

$$
\begin{aligned}
E(Z_{n,2}) &= E(Z_{n,2} 1_{B_{n,2}}) \leq E^{1/2}(Z_{n,2}^2) P^{1/2}(B_{n,2}) \\
&\leq \{2(1+\bar{\lambda} R_n^2)\}^{1/2} (n^{-1} \sum_j d_j) P^{1/2}(B_{n,2}).
\end{aligned} \tag{S3}
$$

If $\sum_k a_k^2 \theta_k^2 > 2\bar{\lambda} R_n \sum_k a_k^2 d_k$, then, by the proof of Lemma 3, $P(B_{n,2}) \leq P(\sum_j a_j^2 Y_j^2 \leq Q_n/2)$ and hence $E(Z_{n,2}) \leq C_3\, n^{\eta} n^{-(1-3\eta)/2}$, where C_3 is a constant (free of θ). Combining the two cases shows that $\sup_{\theta \in \mathbb{R}^n} E(Z_{n,2}) = O\{n^{-(1-5\eta)/2}\}$. □

2.5 Proofs of Propositions 1–2

Proof of Proposition 1. For simplicity, we write $\hat{\gamma}$ for $\hat{\gamma}_0$ and γ^* for γ_0^*, which should be distinguished from $\hat{\gamma}$ and γ^* in Proposition 2. Let $\tilde{\gamma} \in (-\min_j d_j, \infty)$ be a solution such that $n = \sum_j Y_j^2/(d_j + \tilde{\gamma})$. Then $\hat{\gamma} = \max(0, \tilde{\gamma})$. By simple manipulation of $0 = \sum_j Y_j^2/(d_j + \tilde{\gamma}) - \sum_j (d_j + \theta_j^2)/(d_j + \gamma^*)$, we have

$$\tilde{\gamma} - \gamma^* = \frac{\sum_j (Y_j^2 - d_j - \theta_j^2)/(d_j + \gamma^*)}{\sum_j Y_j^2/\{(d_j + \tilde{\gamma})(d_j + \gamma^*)\}},$$

and hence

$$|\hat{\gamma} - \gamma^*| \leq |\tilde{\gamma} - \gamma^*| \leq (\max_j d_j + \gamma^*) \left| n^{-1} \sum_j \frac{Y_j^2 - d_j - \theta_j^2}{d_j + \gamma^*} \right|.$$

(i) For any constant $v_n > 0$, we have, uniformly in $\theta \in \Theta_n$,

$$P(|\tilde{\gamma} - \gamma^*| \geq v_n) \leq P\left\{ \left| n^{-1} \sum_j \frac{Y_j^2 - (d_j + \theta_j^2)}{d_j + \gamma^*} \right| \geq \frac{v_n}{\min_j d_j + \gamma^*} \right\}$$
$$\leq \frac{(\max_j d_j + \gamma^*)^2}{v_n^2} \mathrm{var} \left\{ n^{-1} \sum_j \frac{Y_j^2}{d_j + \gamma^*} \right\}$$
$$\leq \frac{(\max_j d_j + \gamma^*)^2}{n^2 v_n^2} \sum_j \frac{8\theta_j^2 d_j + 2K_1 d_j^2}{(d_j + \gamma^*)^2}$$
$$\leq \frac{(8M + 2K_1)(\max_j d_j + \gamma^*)^2}{n v_n^2}$$

by Chebyshev's inequality. Note that $\gamma^* = n^{-1} \sum_j (d_j + \theta_j^2)\{\gamma^*/(d_j + \gamma^*)\} \leq n^{-1} \sum_j (d_j + \theta_j^2) \leq (1 + M)(\max_j d_j)$ for $\theta \in \Theta_n$. The preceding inequality shows that

$$\sup_{\theta \in \Theta_n} P(|\tilde{\gamma} - \gamma^*| \geq v_n) \leq \frac{C(\max_j d_j)^2}{n v_n^2}, \tag{S4}$$

for some constant C. Taking $v_n = \tau_2 n^{-(1-2\eta)/2}$ gives the desired result.

(ii) By the proof of (i), we have, uniformly in $\theta \in \Theta_n$,

$$E|\tilde{\gamma} - \gamma^*|^2 \leq \frac{(8M + 2K_1)(\max_j d_j + \gamma^*)^2}{n}.$$

This leads directly to the desired result. □

Proof of Proposition 2. Throughout, we write β_γ for $\hat{\beta}(\gamma)$ in (11), and hence $\hat{\beta} = \beta_{\hat{\gamma}} = \hat{\beta}(\hat{\gamma})$. Let $\tilde{\gamma} \in (-\min_j d_j, \infty)$ be a solution such that $n - q = \sum_j (Y_j - x_j^{\mathrm{T}} \beta_{\tilde{\gamma}})^2/(d_j + \tilde{\gamma})$. Then $\hat{\gamma} = \max(0, \tilde{\gamma})$. We make use of the following identities repeatedly. For $\gamma > -\min_j d_j$, direct calculation shows that

$$\sum_j (Y_j - x_j^{\mathrm{T}} \beta_\gamma)^2/(d_j + \gamma) - (n - q)$$

$$=S_{n,1}(\gamma,\gamma^*)(\gamma^*-\gamma)+S_{n,2}(\gamma,\gamma^*)(\gamma-\gamma^*)^2+T_{n,1}-T_{n,2} \tag{S5}$$

$$=S_{n,1}(\gamma^*,\gamma)(\gamma^*-\gamma)-S_{n,2}(\gamma^*,\gamma)(\gamma-\gamma^*)^2+T_{n,1}-T_{n,2} \tag{S6}$$

where $S_{n,1}(\gamma^*,\gamma)$ and $S_{n,2}(\gamma^*,\gamma)$ are defined as, respectively, $S_{n,1}(\gamma,\gamma^*)$ and $S_{n,2}(\gamma,\gamma^*)$ with γ and γ^* exchanged, and

$$S_{n,1}(\gamma,\gamma^*)=\sum_j \frac{(Y_j-x_j^{\mathrm{T}}\beta_\gamma)^2}{(d_j+\gamma)(d_j+\gamma^*)},$$

$$S_{n,2}(\gamma,\gamma^*)=\sum_j \frac{x_j^{\mathrm{T}}(Y_j-x_j^{\mathrm{T}}\beta_\gamma)}{(d_j+\gamma)(d_j+\gamma^*)}\left(\sum_j \frac{x_j x_j^{\mathrm{T}}}{d_j+\gamma^*}\right)^{-1}\sum_j \frac{x_j(Y_j-x_j^{\mathrm{T}}\beta_\gamma)}{(d_j+\gamma)(d_j+\gamma^*)},$$

$$T_{n,1}=\sum_j \frac{(Y_j-x_j^{\mathrm{T}}\beta^*)^2-d_j-(\theta_j-x_j^{\mathrm{T}}\beta^*)^2}{d_j+\gamma^*},$$

$$T_{n,2}=\sum_j \frac{x_j^{\mathrm{T}}(Y_j-\theta_j)}{d_j+\gamma^*}\left(\sum_j \frac{x_j x_j^{\mathrm{T}}}{d_j+\gamma^*}\right)^{-1}\sum_j \frac{x_j(Y_j-\theta_j)}{d_j+\gamma^*}.$$

(i) Take $v_n=\tau_2 n^{-(1-2\eta)/2}$. It suffices to show that

$$\sup_{\theta\in\Theta_n} P(|\tilde\gamma-\gamma^*|\geq v_n)\leq \frac{C_1(\max_j d_j)^2}{n v_n^2}, \tag{S7}$$

for some constant C_1. Taking $\gamma=\gamma^*+v_n$ in equation (S6) shows that $T_{n,1}-T_{n,2}\geq \sum_j(Y_j-x_j^{\mathrm{T}}\beta_\gamma)^2/(d_j+\gamma)-(n-q)+v_n S_{n,1}(\gamma^*,\gamma)$. By the monotonicity of $\sum_j(Y_j-x_j^{\mathrm{T}}\beta_\gamma)^2/(d_j+\gamma)$ in γ, we then have

$$P(\tilde\gamma\geq\gamma^*+v_n)=P\left\{\sum_j \frac{(Y_j-x_j^{\mathrm{T}}\beta_\gamma)^2}{d_j+\gamma}-(n-q)\geq 0 \text{ and } \tilde\gamma\geq\gamma^*+v_n\right\}$$

$$\leq P\left\{T_{n,1}-T_{n,2}\geq v_n\sum_j \frac{(Y_j-x_j^{\mathrm{T}}\beta_{\gamma^*})^2}{(d_j+\gamma^*)(d_j+\gamma)} \text{ and } \tilde\gamma\geq\gamma^*+v_n\right\}$$

$$\leq P\left\{T_{n,1}-T_{n,2}\geq \frac{(n-q)v_n}{\max_j d_j+\gamma}\right\}.$$

Taking $\gamma=\gamma^*-v_n$ in (S5) shows that $T_{n,1}-T_{n,2}\leq\sum_j(Y_j-x_j^{\mathrm{T}}\beta_\gamma)^2/(d_j+\gamma)-(n-q)-v_nS_{n,1}(\gamma,\gamma^*)$. By the definition of $S_{n,1}$, we have $\sum_j(Y_j-x_j^{\mathrm{T}}\beta_\gamma)^2/(d_j+\gamma)-(n-q)-v_nS_{n,1}(\gamma,\gamma^*)\leq\sum_j(Y_j-x_j^{\mathrm{T}}\beta_\gamma)^2/(d_j+\gamma)-(n-q)-v_n\sum_j\{(Y_j-x_j^{\mathrm{T}}\beta_\gamma)^2/(d_j+\gamma)\}/(\max_k d_k+\gamma^*)=\{1-v_n/(\max_k d_k+\gamma^*)\}\{\sum_j(Y_j-x_j^{\mathrm{T}}\beta_\gamma)^2/(d_j+\gamma)-(n-q)\}-(n-q)v_n/(\max_k d_k+\gamma^*)$. Then, for all large n such that $v_n\leq\max_j d_j/2$,

$$P(\tilde\gamma\leq\gamma^*-v_n)=P\left\{\sum_j \frac{(Y_j-x_j^{\mathrm{T}}\beta_\gamma)^2}{d_j+\gamma}-(n-q)\leq 0 \text{ and } \tilde\gamma\leq\gamma^*-v_n\right\}$$

$$\leq P\left[T_{n,1}-T_{n,2}\leq\frac{1}{2}\left\{\sum_j\frac{(Y_j-x_j^{\mathrm{T}}\beta_\gamma)^2}{d_j+\gamma}-(n-q)\right\}-\frac{(n-q)v_n}{\max_j d_j+\gamma^*}\text{ and }\tilde{\gamma}\leq\gamma^*-v_n\right]$$
$$\leq P\left\{T_{n,1}-T_{n,2}\leq-\frac{(n-q)v_n}{\max_j d_j+\gamma^*}\right\}.$$

Therefore, it suffices to show that

$$\sup_{\theta\in\Theta_n}P\left\{|T_{n,1}-T_{n,2}|\geq\frac{(n-q)v_n}{\max_j d_j+\gamma^*+v_n}\right\}\leq\frac{C_2(\max_j d_j)^2}{nv_n^2},$$

for some constant C_2. By Chebyshev's inequality, we have

$$P\left\{|T_{n,1}-T_{n,2}|\geq\frac{(n-q)v_n}{\max_j d_j+\gamma^*+v_n}\right\}$$
$$\leq\frac{2(\max_j d_j+\gamma^*+v_n)^2}{(n-q)^2v_n^2}\left\{E(T_{n,1}^2)+E(T_{n,2}^2)\right\}.$$

By the definition of γ^*, we have, for $\theta\in\Theta_n$,

$$\begin{aligned}\gamma^*&\leq(n-q)^{-1}\sum_j(d_j+\theta_j^2)\{\gamma^*/(d_j+\gamma^*)\}\\&\leq(n-q)^{-1}\sum_j(d_j+\theta_j^2)\leq(1-q/n)^{-1}(1+M)(\max_j d_j).\end{aligned}\tag{S8}$$

Moreover, direct calculation shows that

$$T_{n,1}=\sum_j\frac{\varepsilon_j^2-d_j+2\varepsilon_j(\theta_j-x_j^{\mathrm{T}}\beta^*)}{d_j+\gamma^*},$$

and hence

$$E(T_{n,1}^2)\leq\sum_j\frac{8(\theta_j-x_j^{\mathrm{T}}\beta^*)^2d_j+2K_1d_j^2}{(d_j+\gamma^*)^2}\leq(8+2K_1)(n-q),\tag{S9}$$

because $\sum_j(\theta_j-x_j^{\mathrm{T}}\beta^*)^2/(d_j+\gamma^*)=\sum_j\gamma^*/(d_j+\gamma^*)-q\leq n-q$. Let $\xi_j=V_{\gamma^*}^{-1/2}\{x_j(Y_j-\theta_j)/(d_j+\gamma^*)\}$ for $j=1,\ldots,n$. Then $T_{n,2}=\sum_j\xi_j^{\mathrm{T}}\sum_k\xi_k$ and

$$\begin{aligned}E(T_{n,2}^2)&=\sum_{j\neq k}E(\xi_j^{\mathrm{T}}\xi_j)E(\xi_k^{\mathrm{T}}\xi_k)+2\sum_{j\neq k}E\{(\xi_j^{\mathrm{T}}\xi_k)^2\}+\sum_jE\{(\xi_j^{\mathrm{T}}\xi_j)^2\}\\&\leq3\sum_{j\neq k}E(\xi_j^{\mathrm{T}}\xi_j)E(\xi_k^{\mathrm{T}}\xi_k)+\sum_jE\{(\xi_j^{\mathrm{T}}\xi_j)^2\}\\&\leq3\left[\sum_jE^{1/2}\{(\xi_j^{\mathrm{T}}\xi_j)^2\}\right]^2=3K_1\left\{\sum_j\frac{x_j^{\mathrm{T}}V_{\gamma^*}^{-1}x_j}{(d_j+\gamma^*)^2}d_j\right\}^2\leq3K_1q^2.\end{aligned}\tag{S10}$$

Combining the preceding results completes the proof.

(ii) Taking $\gamma = \tilde{\gamma}$ in equation (S5) shows that $S_{n,1}(\tilde{\gamma}, \gamma^*)(\tilde{\gamma} - \gamma^*) - S_{n,2}(\tilde{\gamma}, \gamma^*)(\tilde{\gamma} - \gamma^*)^2 = T_{n,1} - T_{n,2}$. If $\tilde{\gamma} \le \gamma^*$, then

$$0 \le \gamma^* - \tilde{\gamma} \le \frac{|T_{n,1} - T_{n,2}|}{S_{n,1}(\tilde{\gamma}, \gamma^*)} \le (\max_j d_j + \gamma^*) \frac{|T_{n,1} - T_{n,2}|}{n - q}.$$

Taking $\gamma = \tilde{\gamma}$ in equation (S6) shows that $S_{n,1}(\gamma^*, \tilde{\gamma})(\tilde{\gamma} - \gamma^*) + S_{n,2}(\gamma^*, \tilde{\gamma})(\tilde{\gamma} - \gamma^*)^2 = T_{n,1} - T_{n,2}$. If $\tilde{\gamma} > \gamma^*$, then

$$0 < \tilde{\gamma} - \gamma^* \le \frac{|T_{n,1} - T_{n,2}|}{S_{n,1}(\gamma^*, \tilde{\gamma})} \le (\max_j d_j + \tilde{\gamma}) \frac{|T_{n,1} - T_{n,2}|}{n - q},$$

and hence if further $|T_{n,1} - T_{n,2}| \le (n-q)/2$, then $\tilde{\gamma} - \gamma^* \le 2(\max_j d_j + \gamma^*)|T_{n,1} - T_{n,2}|/(n-q)$. Combining these results and using the bounds (S9)–(S10) on $E(T_{n,1}^2)$ and $E(T_{n,2}^2)$, we have, for all $\theta \in \Theta_n$,

$$\begin{aligned} E|\tilde{\gamma} - \gamma^*| &\le 2\frac{\max_j d_j + \gamma^*}{n - q} E|T_{n,1} - T_{n,2}| + E\left[|\tilde{\gamma} - \gamma^*| 1_{\{|T_{n,1} - T_{n,2}| > (n-q)/2\}}\right] \\ &\le 2\frac{\max_j d_j + \gamma^*}{n - q} E^{1/2}(|T_{n,1} - T_{n,2}|^2) + E^{1/2}(|\tilde{\gamma} - \gamma^*|^2) P^{1/2}\{|T_{n,1} - T_{n,2}| > (n-q)/2\} \\ &\le C_3(\min_j d_j) n^{-(1-2\eta)/2} + C_4 n^{-1/2} E^{1/2}(|\tilde{\gamma} - \gamma^*|^2), \end{aligned}$$

where C_3 and C_4 are constants. Because $\gamma^* = (\min_j d_j) O(n^\eta)$ uniformly in $\theta \in \Theta_n$, it suffices to show that $E(\tilde{\gamma}^2) = (\min_j d_j)^2 O(n^{2\eta})$ uniformly in $\theta \in \Theta_n$. In fact, $\tilde{\gamma} > -\min_j d_j$ by definition and if $\tilde{\gamma} \ge 0$, then $\tilde{\gamma} \le (n-q)^{-1} \sum_j Y_j^2 \{\tilde{\gamma}/(d_j + \tilde{\gamma})\} \le (n-q)^{-1} (\sum_j Y_j^2/d_j)(\max_j d_j)$. For all $\theta \in \Theta_n$, direct calculation shows that $n^{-1} \sum_j Y_j^2/d_j \le 1 + M + n^{-1}|\sum_j (\varepsilon_j^2 - d_j)/d_j| + 2n^{-1}|\sum_j \theta_j \varepsilon_j / d_j|$ and hence $E(\tilde{\gamma}^2) \le (\min_j d_j)^2 + 3(1 - q/n)^2 \{(1+M)^2 + K_1/n + 4M/n\} (\max_j d_j)^2 \le C_5 (\min_j d_j)^2 n^{2\eta}$ for a constant C_5.

(iii) Similarly to the proof of (ii), we have, for all $\theta \in \Theta_n$,

$$\begin{aligned} E|\tilde{\gamma} - \gamma^*|^2 &\le 4\frac{(\max_j d_j + \gamma^*)^2}{(n - q)^2} E|T_{n,1} - T_{n,2}|^2 + E\left[|\tilde{\gamma} - \gamma^*|^2 1_{\{|T_{n,1} - T_{n,2}| > (n-q)/2\}}\right] \\ &\le 4\frac{(\max_j d_j + \gamma^*)^2}{(n - q)^2} E|T_{n,1} - T_{n,2}|^2 + E^{1/2}(|\tilde{\gamma} - \gamma^*|^4) P^{1/2}\{|T_{n,1} - T_{n,2}| > (n-q)/2\} \\ &\le C_6(\min_j d_j)^2 n^{-(1-2\eta)} + C_7 n^{-1/2} E^{1/2}(|\tilde{\gamma} - \gamma^*|^4), \end{aligned}$$

where C_6 and C_7 are constants. Then it suffices to show that $E(\tilde{\gamma}^4) = O(n^{4\eta})$ uniformly in $\theta \in \Theta_n$. This follows by similar calculation as above for $E(\tilde{\gamma}^2)$, using the fact that $E(\sum_j U_j)^4 = \sum_j E(U_j^4) + 3\sum_{j \ne k} E(U_j^2) E(U_k^2) \le 3\{\sum_j E^{1/2}(U_j^4)\}^2$, where $(U_1, \ldots, U_n)$ are independent variables, each with mean 0. □

2.6 Proofs of Theorem 4 and Corollary 1

We provide Lemma 5 on smoothness properties of $\{a_1(\gamma), \ldots, a_n(\gamma)\}$, determined from (8)–(9), as γ varies. Moreover, we give Proposition 3, which combined with Theorem 4 yields Corollary 1. Finally, we provide a proof of Theorem 4. Throughout, we write $\delta_{\lambda,\gamma}$ for $\delta_{\lambda,\gamma,\beta=0}$ and write $\hat{\gamma}$ for $\hat{\gamma}_0$ and γ^* for γ_0^*.

Lemma 5. Sort the indices such that $d_1 \geq d_2 \geq \cdots \geq d_n$. (i) For any $\gamma_1 \geq 0$ and $k = 1, \ldots, n$, the left and right derivatives

$$L_k(\gamma_1) = \lim_{\gamma_2 \uparrow \gamma_1} \frac{a_k(\gamma_2) - a_k(\gamma_1)}{\gamma_2 - \gamma_1}, \quad U_k(\gamma_1) = \lim_{\gamma_2 \downarrow \gamma_1} \frac{a_k(\gamma_2) - a_k(\gamma_1)}{\gamma_2 - \gamma_1}$$

exist and are finite.

(ii) For any $\gamma_1 \geq 0$,

$$0 \leq \frac{-L_1(\gamma_1)}{a_1(\gamma_1)} \leq \cdots \leq \frac{-L_n(\gamma_1)}{a_n(\gamma_1)} \leq (\min_j d_j)^{-1},$$
$$0 \leq \frac{-U_1(\gamma_1)}{a_1(\gamma_1)} \leq \cdots \leq \frac{-U_n(\gamma_1)}{a_n(\gamma_1)} \leq (\min_j d_j)^{-1}.$$

(iii) For any $0 \leq \gamma_1 \leq \gamma_2$,

$$1 \leq \frac{a_1(\gamma_1)}{a_1(\gamma_2)} \leq \cdots \leq \frac{a_n(\gamma_1)}{a_n(\gamma_2)} \leq 1 + \frac{\gamma_2 - \gamma_1}{\min_j d_j}.$$

(iv) For any $0 \leq \gamma_1 \leq \gamma_2$,

$$1 \leq \frac{c(\gamma_1)}{c(\gamma_2)} \leq 1 + \frac{\gamma_2 - \gamma_1}{\min_j d_j}.$$

Proof of Lemma 5. For $k = 3, \ldots, n-1$, let $r_k(\gamma) = \sum_{j=1}^k \{d_{k+1}^2/(d_{k+1} + \gamma)\}/\{d_j^2/(d_j + \gamma)\}$. Then $r_k(\gamma) \geq k-2$ for $3 \leq k \leq \nu(\gamma) - 1$ and $r_k(\gamma) < k-2$ for $\nu(\gamma) \leq k \leq n-1$ by Tan (2014, Corollary 2). Moreover, $r_k(\gamma)$ is non-increasing in γ for each k. To show (i)–(ii), consider the following three cases of γ_1.

Suppose that $r_k(\gamma_1) > k-2$ for $k = \nu(\gamma_1) - 1$ and hence for all $3 \leq k \leq \nu(\gamma_1) - 1$. By continuity of $r_k(\gamma)$ in γ, there exists $h > 0$ such that for any $\gamma_2 \in (\gamma_1 - h, \gamma_1 + h) \cap [0, \infty)$, $\nu(\gamma_2) = \nu(\gamma_1)$. Then $a_k(\gamma_2)/a_k(\gamma_1) = \{\sum_{j=1}^{\nu(\gamma_1)} (d_j + \gamma_1)/d_j^2\}/\{\sum_{j=1}^{\nu(\gamma_1)} (d_j + \gamma_2)/d_j^2\}$ for $1 \leq k \leq \nu(\gamma_1)$, and $a_k(\gamma_2)/a_k(\gamma_1) = (d_k + \gamma_1)/(d_k + \gamma_2)$ for $\nu(\gamma_1) + 1 \leq k \leq n$, which lead directly to the results (i)–(ii).

Suppose that $r_k(\gamma_1) = k-2$ for $k = \nu(\gamma_1) - 1$, but $r_k(\gamma_1) > k-2$ for $k = \nu(\gamma_1) - 2$ and hence for all $3 \leq k \leq \nu(\gamma_1) - 2$. By continuity and monotonicity of $r_k(\gamma)$ in γ, there exists $h_1 > 0$ such that for any $\gamma_2 \in (\gamma_1 - h_1, \gamma_1) \cap [0, \infty)$, $\nu(\gamma_2) = \nu(\gamma_1)$ and hence the desired results follow similarly to the first case. It remains to deal with $\gamma_2 \in (\gamma_1, \gamma_1 + h_2) \cap [0, \infty)$ for small $h_2 > 0$ such that $\nu(\gamma_2) = \nu(\gamma_1) - 1$. By direct calculation using $r_k(\gamma_1) = k - 2$ for $k = \nu(\gamma_1) - 1$, we have $a_k(\gamma_1) = [d^2_{\nu(\gamma_1)}/\{d_{\nu(\gamma_1)} + \gamma_1\}]/d_k$ for $1 \le k \le \nu(\gamma_1)$ and $a_k(\gamma_2) = w[d^2_{\nu(\gamma_1)}/\{d_{\nu(\gamma_1)} + \gamma_1\}]/d_k$ for $1 \le k \le \nu(\gamma_1) - 1$, where $w = \{\sum_{j=1}^{\nu(\gamma_1)-1}(d_j + \gamma_1)/d_j^2\}/\{\sum_{j=1}^{\nu(\gamma_1)-1}(d_j + \gamma_2)/d_j^2\}$. Then $a_k(\gamma_2)/a_k(\gamma_1) = w$ for $1 \le k \le \nu(\gamma_1) - 1$, and $a_k(\gamma_2)/a_k(\gamma_1) = (d_k + \gamma_1)/(d_k + \gamma_2)$ for $\nu(\gamma_1) \le k \le n$, which lead directly to the results (i)–(ii).

Suppose that for some $3 \le k_0 < \nu(\gamma_1) - 2$, $r_k(\gamma_1) = k - 2$ for $k = k_0 + 1, \ldots, \nu(\gamma_1) - 1$, but $r_k(\gamma_1) > k - 2$ for $k = k_0$ and hence for all $3 \le k \le k_0$. Then $d_{k0+2} = \cdots = d_{\nu(\gamma_1)}$ by Tan (2014, Corollary 2). The results (i)–(ii) follow similarly to the second case, which corresponds to $k_0 = \nu(\gamma_1) - 2$.

The result (iii) follows easily from (i)–(ii). Similarly, the result (iv) follows from corresponding results on the left and right derivatives of $c(\gamma_1)$. For example, for $\gamma_2 \in (\gamma_1, \gamma_1 + h_2) \cap [0, \infty)$ in the second case above, we have $c(\gamma_1) = \{\nu(\gamma_1) - 2\}d^2_{\nu(\gamma_1)}/\{d_{\nu(\gamma_1)} + \gamma_1\} + \sum_{j=\nu(\gamma_1)+1}^n d_j^2/(d_j + \gamma_1)$ and $c(\gamma_2) = w\{\nu(\gamma_1) - 3\}d^2_{\nu(\gamma_1)}/\{d_{\nu(\gamma_1)} + \gamma_1\} + \sum_{j=\nu(\gamma_1)}^n d_j^2/\ (d_j + \gamma_2)$, and hence $\{c(\gamma_1) - c(\gamma_2)\}/c(\gamma_1) \le (\gamma_2 - \gamma_1)/(\min_j d_j)$. □

Proposition 3. If Assumptions (A1)–(A3) hold with $0 \le \eta < 1/4$, then $\sup_{\theta \in \Theta_n} E\{\sup_{0 \le \lambda \le 2} n^{-1}|\text{SURE}(\delta_{\lambda,\hat{\gamma}}) - \text{SURE}(\delta_{\lambda,\gamma^*})|\} = O\{n^{-(1-4\eta)/2}\}$.

Proof of Proposition 3. The result follows from Proposition 1(ii) and the following inequality: for any $0 \le \gamma_1 \le \gamma_2$,

$$\begin{aligned}
&\sup_{0 \le \lambda \le \bar{\lambda}} |\text{SURE}(\delta_{\lambda,\gamma_1}) - \text{SURE}(\delta_{\lambda,\gamma_2})| \\
\le& 8(\max_j d_j) + 2(2\Delta + \Delta^2)(1 + \bar{\lambda} \sup_{0 \le \gamma \le \gamma_2} R_{n,\gamma}) \sum_j d_j,
\end{aligned}$$

where $\Delta = (\gamma_2 - \gamma_1)/(\min_j d_j)$ and $R_{n,\gamma} = \{\max_k a_k(\gamma)\}/\{\min_k a_k(\gamma)\}$. In fact, direct calculation shows that

$$\begin{aligned}
&\text{SURE}(\delta_{\lambda,\gamma_1}) - \text{SURE}(\delta_{\lambda,\gamma_2}) \\
=& \sum_{j \in J_{\lambda,\gamma_1}} \frac{4b_j(\gamma_1)d_j a_j^2(\gamma_1)Y_j^2}{\sum_k a_k^2(\gamma_1)Y_k^2} - \sum_{j \in J_{\lambda,\gamma_2}} \frac{4b_j(\gamma_2)d_j a_j^2(\gamma_2)Y_j^2}{\sum_k a_k^2(\gamma_2)Y_k^2}
\end{aligned}$$

$$+\sum_{j\in J_{\lambda,\gamma_1}\cup J_{\lambda,\gamma_2}}\left\{b'^2_j(\gamma_1)Y_j^2-2b'_j(\gamma_1)d_j-b'^2_j(\gamma_2)Y_j^2+2b'_j(\gamma_2)d_j\right\},\tag{S11}$$

where $b'_j(\gamma)=\min\{1,b_j(\gamma)\}$, $b_j(\gamma)=\lambda c(\gamma)a_j(\gamma)/\{\sum_k a_k^2(\gamma)Y_k^2\}$, and $J_{\lambda,\gamma}=\{j:b_j(\gamma)<1\}$. Notationally, the dependency of $b_j(\gamma)$ on λ is suppressed. Then $(1+\Delta)^{-2}\le b_j(\gamma_1)/b_j(\gamma_2)\le(1+\Delta)^2$ for $j=1,\ldots,n$ by Lemma 5(iii)–(iv). Moreover,

$$\begin{aligned}&|b'^2_j(\gamma_1)Y_j^2-2b'_j(\gamma_1)d_j-b'^2_j(\gamma_2)Y_j^2+2b'_j(\gamma_2)d_j|\\ \le&|b'_j(\gamma_1)-b'_j(\gamma_2)|\{b'_j(\gamma_1)+b'_j(\gamma_2)\}Y_j^2+2|b'_j(\gamma_1)-b'_j(\gamma_2)|d_j\\ \le&2|b'_j(\gamma_1)-b'_j(\gamma_2)|Y_j^2+2|b'_j(\gamma_1)-b'_j(\gamma_2)|d_j.\end{aligned}\tag{S12}$$

Combining the preceding results leads to

$$\begin{aligned}&|\mathrm{SURE}(\delta_{\lambda,\gamma_1})-\mathrm{SURE}(\delta_{\lambda,\gamma_2})|\\ \le&8(\max_j d_j)+2\{(1+\Delta)^2-1\}\sum_j\left\{b'_j(\gamma_1)Y_j^2+b'_j(\gamma_1)d_j\right\}.\end{aligned}$$

The desired result follows because $\sum_j b'_j(\gamma_1)d_j\le\sum_j d_j$ and $\sum_j b'_j(\gamma_1)Y_j^2\le\lambda c(\gamma_1)$ $\{\sum_j a_j(\gamma_1)Y_j^2\}/\ \{\sum_k a_k^2(\gamma_1)Y_k^2\}\le\lambda c(\gamma_1)/\{\min_j a_j(\gamma_1)\}\le\bar\lambda R_{n,\gamma_1}(\sum_j d_j)$. □

Proof of Theorem 4. Let $G_n=\{\gamma:|\gamma-\gamma^*|\le(\min_j d_j)/2\}$. Then $\sup_k|a_k(\gamma)/a_k(\gamma^*)-1|\le 1/2$ for $\gamma\in G_n$ by Lemma 5(iii). Moreover, $\sup_{\theta\in\Theta_n}P(\hat\gamma\notin G_n)\le C_1n^{-(1-2\eta)}$ for some constant C_1 by (S4) in the proof of Proposition 1.

(i) It suffices to show that $\sup_{0\le\lambda\le\bar\lambda,\gamma\in G_n}|\zeta_n(\lambda,\gamma)|=O_p\{n^{-(1-4\eta)/2}\}$, uniformly in $\theta\in\mathbb{R}^n$. This follows similarly to Theorem 3(i), based on suitable extensions of Lemmas 1–2.

Sort the indices such that $d_1\ge d_2\ge\cdots\ge d_n$. Then $a_1(\gamma)\le\cdots\le a_{\nu(\gamma)}(\gamma)$ and $a_{\nu(\gamma)+1}(\gamma)\ge\cdots\ge a_n(\gamma)$. By splitting the set $J_{\lambda,\gamma}=\{j:\sum_k a_k^2(\gamma)Y_k^2>\lambda c(\gamma)a_j(\gamma)\}$ into two subsets in $\{1,\ldots,\nu(\gamma)\}$ and $\{\nu(\gamma)+1,\ldots,n\}$ respectively, we have

$$\begin{aligned}&\sup_{0\le\lambda\le\bar\lambda,\gamma\in G_n}\left|\sum_{j\in J_{\lambda,\gamma}}\lambda c(\gamma)\,a_j(\gamma)(d_j-\varepsilon_j^2)\right|\\ \le&\sup_{\gamma\in G_n}\left\{\max_k\left|\sum_{j=1}^k\bar\lambda c(\gamma)\,a_j(\gamma)(d_j-\varepsilon_j^2)\right|+\max_k\left|\sum_{j=k}^n\bar\lambda c(\gamma)\,a_j(\gamma)(d_j-\varepsilon_j^2)\right|\right\}\\ \le&2\left(\frac{3}{2}\right)^2\left\{\max_k\left|\sum_{j=1}^k\bar\lambda c(\gamma^*)\,a_j(\gamma^*)(d_j-\varepsilon_j^2)\right|+\max_k\left|\sum_{j=k}^n\bar\lambda c(\gamma^*)\,a_j(\gamma^*)(d_j-\varepsilon_j^2)\right|\right\}.\end{aligned}$$

To see the last step, let $w_j = \{c(\gamma)/c(\gamma^*)\}\{a_j(\gamma)/a_j(\gamma^*)\}$ for $j = 1, \ldots, n$. If $\gamma < \gamma^*$ and $\gamma \in G_n$, then $1 \le w_1 \le \cdots \le w_n \le (3/2)^2$ by Lemma 5(iii)—(iv), and hence $|\sum_{j=1}^k \bar{\lambda} c(\gamma) a_j(\gamma)(d_j - \varepsilon_j^2)| = |\sum_{j=1}^k \bar{\lambda} c(\gamma^*) a_j(\gamma^*) w_j (d_j - \varepsilon_j^2)| \le (3/2)^2 \max_{1 \le l \le k} |\sum_{j=l}^k \bar{\lambda}(\gamma^*)\, a_j(\gamma^*)(d_j - \varepsilon_j^2)| \le 2(3/2)^2 \max_{1 \le l \le k} |\sum_{j=1}^l \bar{\lambda} c(\gamma^*)\, a_j(\gamma^*)(d_j - \varepsilon_j^2)|$ and $|\sum_{j=k}^n \bar{\lambda} c(\gamma)\, a_j(\gamma)(d_j - \varepsilon_j^2)| = |\sum_{j=k}^n \bar{\lambda} c(\gamma^*) a_j(\gamma^*) w_j (d_j - \varepsilon_j^2)| \le (3/2)^2 \max_{k \le l \le n} |\sum_{j=l}^n \bar{\lambda} c(\gamma^*)\, a_j(\gamma^*)\, (d_j - \varepsilon_j^2)|$, by the observation that

$$\sup_{0 \le w_1 \le \cdots \le w_n \le 1} \left| \sum_{j=1}^k w_j Y_j \right| = \max_{1 \le l \le k} \left| \sum_{j=l}^k Y_j \right|$$

for any real numbers $y_1, \ldots, y_n$ (Speckman 1985; Li 1985). Similarly, if $\gamma > \gamma^*$ and $\gamma \in G_n$, then $1 \ge w_1 \ge \cdots \ge w_n \ge (1/2)^2$, and hence $|\sum_{j=1}^k \bar{\lambda} c(\gamma)\, a_j(\gamma)(d_j - \varepsilon_j^2)| = |\sum_{j=1}^k \bar{\lambda} c(\gamma^*)\, a_j(\gamma^*) w_j (d_j - \varepsilon_j^2)| \le \max_{1 \le l \le k} |\sum_{j=1}^l \bar{\lambda} c(\gamma^*)\, a_j(\gamma^*)(d_j - \varepsilon_j^2)|$ and $|\sum_{j=k}^n \bar{\lambda} c(\gamma)\, a_j(\gamma)(d_j - \varepsilon_j^2)| = |\sum_{j=k}^n \bar{\lambda} c(\gamma^*)\, a_j(\gamma^*) w_j (d_j - \varepsilon_j^2)| \le \max_{k \le l \le n} |\sum_{j=k}^l \bar{\lambda} c(\gamma^*)\, a_j(\gamma^*)(d_j - \varepsilon_j^2)| \le 2 \max_{k \le l \le n} |\sum_{j=l}^n \bar{\lambda} c(\gamma^*)\, a_j(\gamma^*)(d_j - \varepsilon_j^2)|$.

Write $Q_n^* = \sum_j a_j^2(\gamma^*)(\theta_j^2 + d_j)$. By the preceding analysis, we have

$$\begin{aligned}
&\sup_{\theta \in \mathbb{R}^n} P\left\{ n^{-1} \sup_{0 \le \lambda \le \bar{\lambda}, \gamma \in G_n} \left| \sum_{j \in J_{\lambda,\gamma}} \lambda c(\gamma)\, a_j(\gamma)(d_j - \varepsilon_j^2) \right| \ge \frac{9}{2} Q_n^* v_n \right\} \\
\le & P\left\{ n^{-1} \max_k \left| \sum_{j=1}^k \bar{\lambda} c(\gamma^*) a_j(\gamma^*)(d_j - \varepsilon_j^2) \right| \ge Q_n^* v_n / 2 \right\} \\
& + P\left\{ n^{-1} \max_k \left| \sum_{j=k}^n \bar{\lambda} c(\gamma^*) a_j(\gamma^*)(d_j - \varepsilon_j^2) \right| \ge Q_n^* v_n / 2 \right\} \\
\le & \frac{8 K_1 \bar{\lambda}^2 (\max_j d_j)(\sum_j d_j)}{n^2\, v_n^2}, \qquad \text{(S13)}
\end{aligned}$$

in parallel to the first inequality in Lemma 2. Similarly, we have

$$\begin{aligned}
&\sup_{\theta \in \mathbb{R}^n} P\left\{ n^{-1} \sup_{0 \le \lambda \le \bar{\lambda}, \gamma \in G_n} \left| \sum_{j \in J_{\lambda,\gamma}} \lambda c(\gamma)\, a_j(\gamma) \theta_j \varepsilon_j \right| \ge \frac{9}{2} Q_n^* v_n \right\} \\
\le & \frac{8 \bar{\lambda}^2 (\max_j d_j)(\sum_j d_j)}{n^2\, v_n^2}, \qquad \text{(S14)} \\
&\sup_{\theta \in \mathbb{R}^n} P\left\{ \inf_{\gamma \in G_n} \sum_k a_k^2(\gamma) Y_k^2 \le \frac{1}{8} Q_n^* \right\} \\
\le & \sup_{\theta \in \mathbb{R}^n} P\left\{ \sum_k a_k^2(\gamma^*) Y_k^2 \le Q_n^* / 2 \right\} \le \frac{(32 + 8K_1) \max_k \{a_k^2(\gamma^*) d_k\}}{\sum_k a_k^2(\gamma^*) d_k}, \qquad \text{(S15)}
\end{aligned}$$

in parallel to the second and third inequalities in Lemma 2.

Write $R_{n,\gamma} = \{\max_k a_k(\gamma)\}/\{\min_k a_k(\gamma)\}$ and write $R_n = \sup_{\gamma\in G_n} R_{n,\gamma}$, which is bounded from above by $(\max_k d_k)/(\min_k d_k)$ by Lemma 1. By similar arguments to the preceding proof, we obtain the following extension of Lemma 3:

$$\sup_{\theta\in\mathbb{R}^n} P\left\{ n^{-1} \sup_{0\le\lambda\le\bar{\lambda},\gamma\in G_n} \left| \sum_{j\notin J_{\lambda,\gamma}} (d_j - \varepsilon_j^2) \right| \ge 2v_n \right\} \le \frac{2K_1 \sum_j d_j^2}{n^2\, v_n^2}, \tag{S16}$$

$$\begin{aligned} &\sup_{\theta\in\mathbb{R}^n} P\left(n^{-1} \sup_{0\le\lambda\le\bar{\lambda},\gamma\in G_n} \left| \sum_{j\notin J_{\lambda,\gamma}} \theta_j\varepsilon_j \right| \ge 2v_n \right) \\ \le &\max\left[\frac{(32+8K_1)\max_k\{a_k^2(\gamma^*)d_k\}}{\sum_k a_k^2(\gamma^*)d_k}, \frac{36\bar{\lambda}R_n^3(\max_k d_k)(\sum_k d_k)}{n^2 v_n^2} \right]. \end{aligned} \tag{S17}$$

To show the inequality (S17), we have

$$\begin{aligned} &P\left(n^{-1} \sup_{0\le\lambda\le\bar{\lambda},\gamma\in G_n} \left| \sum_{j\notin J_{\lambda,\gamma}} \theta_j\varepsilon_j \right| \ge 2v_n \right) \\ \le &P\left[\bigcup_{\gamma\in G_n} \left\{ \sum_k a_k^2(\gamma)Y_k^2 \le \lambda R_{n,\gamma} \sum_k a_k^2(\gamma)d_k \right\} \right] \\ \le &P\left\{ \sum_k a_k^2(\gamma^*)Y_k^2 \le 9\bar{\lambda}R_n \sum_k a_k^2(\gamma^*)d_k \right\}, \end{aligned}$$

and hence this is no greater than $P\{\sum_k a_k^2(\gamma^*)Y_k^2 \le Q_n^*/2\}$ if $\sum_k a_k^2(\gamma^*)\theta_k^2 > 18\bar{\lambda}R_n \sum_k a_k^2(\gamma^*)d_k$. On the other hand, we have

$$\begin{aligned} &P\left(n^{-1} \sup_{0\le\lambda\le\bar{\lambda},\gamma\in G_n} \left| \sum_{j\notin J_{\lambda,\gamma}} \theta_j\varepsilon_j \right| \ge 2v_n \right) \\ \le &P\left(n^{-1} \max_k \left| \sum_{j=1}^k \theta_j\varepsilon_j \right| \ge v_n \right) + P\left(n^{-1} \max_k \left| \sum_{j=k}^n \theta_j\varepsilon_j \right| \ge v_n \right) \\ \le &\frac{2(\max_j d_j)\{\sum_j a_j^2(\gamma^*)\theta_j^2\}}{n^2\, v_n^2\{\min_j a_j^2(\gamma^*)\}}, \end{aligned}$$

and hence this is no greater than $36\bar{\lambda}R_n^3(\max_k d_k)(\sum_k d_k)/(n^2v_n^2)$ if $\sum_k a_k^2(\gamma^*)\theta_k^2 \le 18\bar{\lambda}R_n \sum_k a_k^2(\gamma^*)d_k$. Combining the two cases gives the desired inequality.

Write $Z_{n,3} = n^{-1}\sup_{0\le\lambda\le\bar{\lambda},\gamma\in G_n}[\lambda c(\gamma)\max_{j\in J_{\lambda,\gamma}}\{d_j a_j(\gamma)\}/\sum_k a_k^2(\gamma)Y_k^2]$. Then $Z_{n,3} \le n^{-1}(\max_j d_j)$ by the proof of the third inequality in Lemma 4. Combining all the preceding results completes the proof of (i).

(ii) Write $Z_n = \sup_{0\le\lambda\le\bar\lambda} |\zeta_n(\lambda,\hat\gamma)|$. Applying Lemma 4 with $a_j = a_j(\hat\gamma)$ shows that $E(Z_n^2) \le C_2 n^{2\eta}$ for C_2 a constant (free of θ). Then by the Cauchy–Schwartz inequality, $E(Z_n 1_{G_n^c}) \le E^{1/2}(Z_{n,0}^2) P^{1/2}(G_n^c) \le (C_1C_2)^{1/2} n^{-(1-4\eta)/2}$ for all $\theta \in \Theta_n$. To complete the proof, it suffices to show that $E(Z_n 1_{G_n}) \le E\{\sup_{0\le\lambda\le\bar\lambda,\gamma\in G_n} |\zeta_n(\lambda,\gamma)|\}$ $= O\{n^{-(1-5\eta)/2}\}$ uniformly in $\theta \in \mathbb{R}^n$.

Write $Z_{n,1} = n^{-1}\sup_{0\le\lambda\le\bar\lambda,\gamma\in G_n} |\sum_{j\in J_{\lambda,\gamma}} \lambda c(\gamma) a_j(\gamma)(d_j - \varepsilon_j Y_j)/\{\sum_k a_k^2(\gamma) Y_k^2\}|$ and $Z_{n,2} = n^{-1}\sup_{0\le\lambda\le\bar\lambda,\gamma\in G_n} |\sum_{j\notin J_{\lambda,\gamma}} (d_j - \varepsilon_j Y_j)|$. Then $\sup_{0\le\lambda\le\bar\lambda,\gamma\in G_n} |\zeta_n(\lambda,\gamma)| \le n^{-1}$ $|\sum_j (d_j - \varepsilon_j^2)| + Z_{n,1} + Z_{n,2} + Z_{n,3}$. Note that $E^{1/2}[\{n^{-1}\sum_j (d_j - \varepsilon_j^2)\}^2] = O\{n^{-(1-\eta)/2}\}$ and, by Lemma 4, $Z_{n,3} \le n^{-1}(\max_j d_j) = O\{n^{-(1-\eta)}\}$. The desired result follows because $\sup_{\theta\in\mathbb{R}^n} E(Z_{n,1}) = O\{n^{-(1-4\eta)/2}\}$ and $\sup_{\theta\in\mathbb{R}^n} E(Z_{n,2}) = O\{n^{-(1-5\eta)/2}\}$ by similar arguments as in the proofs of (i) and Theorem 3(ii). □

2.7 Proofs of Theorem 5 and Corollary 2

In the following two lemmas, we provide an upper bound on $\text{SURE}(\delta_{A,\lambda})$ and then upper bounds on the differences between $x_j^{\text{T}}\beta_{\hat\gamma}$ and $x_j^{\text{T}}\beta^*$. Moreover, we give Proposition 4, which combined with Theorem 5 yields Corollary 2. Finally, we provide a proof of Theorem 5.

Lemma 6. If $c = \sum_j d_j a_j - 2\max_j(d_j a_j) \ge 0$, then

$$|\text{SURE}(\delta_{A,\lambda})| \le \sum_j d_j + \frac{4\sum_j d_j a_j}{\min_j a_j} \le (1+4R_n)\sum_j d_j.$$

Proof of Lemma 6. If $J \ne \emptyset$, then $\sum_k a_k^2 Y_k^2 \ge \lambda c(\max_{j\in J} a_k) \ge \lambda c(\min_j a_j)$. If $J^c \ne \emptyset$, then $\sum_{j\in J^c} Y_j^2 \le \lambda c/(\min_{j\in J^c} a_j) \le \lambda c/(\min_j a_j)$ because $(\min_{j\in J^c} a_j^2)\sum_{j\in J^c} Y_j^2$ $\le \sum_{j\in J^c} a_j^2 Y_j^2 \le \lambda c(\min_{j\in J^c} a_j)$. Using these bounds shows that

$$\text{SURE}(\delta_{A,\lambda}) \le \sum_j d_j + \sum_{j\in J^c} Y_j^2 + \sum_{j\in J}\left\{\frac{\lambda^2 c^2 a_j^2 Y_j^2}{(\sum_k a_k^2 Y_k^2)^2} + 4\frac{\lambda c\, d_j a_j^3 Y_j^2}{(\sum_k a_k^2 Y_k^2)^2}\right\}$$
$$\le \sum_j d_j + \frac{2\lambda c}{\min_j a_j} + \frac{4\max_j d_j a_j}{\min_j a_j} \le \sum_j d_j + 4\frac{\sum_j d_j a_j - \max_j(d_j a_j)}{\min_j a_j}.$$

On the other hand, because $\lambda c\sum_{j\in J} d_j a_j/\sum_k a_k^2 Y_k^2 \le \sum_{j\in J} d_j$, we have

$$\text{SURE}(\delta_{A,\lambda}) \ge \sum_j d_j - 2\sum_{j\in J^c} d_j - 2\sum_{j\in J}\frac{\lambda c\, d_j a_j}{\sum_k a_k^2 Y_k^2} \ge -\sum_j d_j.$$

The desired inequality then follows. $\square$

Lemma 7. Under Assumption (A5), the following results hold:

$$
\begin{aligned}
&|(Y_j - x_j^{\mathrm{T}}\beta_{\hat\gamma}) - (Y_j - x_j^{\mathrm{T}}\beta^*)| \le 2d_j^{1/2}(\Delta_2 + \Delta_3), \quad j = 1, \ldots, n,\\
&\frac{\sum_j a_j^2(\gamma^*)|(Y_j - x_j^{\mathrm{T}}\beta_{\hat\gamma})^2 - (Y_j - x_j^{\mathrm{T}}\beta^*)^2|}{\sum_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}}\beta^*)^2}\\
\le& 2(1 + 2\varrho^{1/2}\Delta_3 + \varrho\Delta_3^2)^{1/2}\varrho^{1/2}\Delta_2 + 2\varrho^{1/2}\Delta_3 + \varrho(\Delta_2^2 + \Delta_3^2),
\end{aligned}
$$

where $\varrho = \{\sum_j a_j^2(\gamma^*)d_j\}/\{\sum_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}}\beta^*)^2\}$, $\Delta_2 = K_5^{1/2}n^\eta(\min_k d_k)^{-1/2}(\min_k d_k + \hat\gamma)^{-1/2}|\hat\gamma - \gamma^*|$ and $\Delta_3 = K_5^{1/2}n^\eta \min_k^{-1/2}\{d_k/(d_k+\gamma^*)\}(n^{-1}T_{n,2})^{1/2}$ with $T_{n,2}$ defined in the proof of Proposition 2.

Proof of Lemma 7. Direct calculation leading to (S5) also shows that

$$
\begin{aligned}
&(Y_j - x_j^{\mathrm{T}}\beta_{\hat\gamma})^2 - (Y_j - x_j^{\mathrm{T}}\beta_{\gamma^*})^2\\
=&2(Y_j - x_j^{\mathrm{T}}\beta_{\gamma^*})x_j^{\mathrm{T}}V_{\gamma^*}^{-1}(X^{\mathrm{T}}D_{\gamma^*}^{-1}D_{\hat\gamma}^{-1}U_{\hat\gamma})(\hat\gamma - \gamma^*)\\
&+ (U_{\hat\gamma}^{\mathrm{T}}D_{\gamma^*}^{-1}D_{\hat\gamma}^{-1}X)V_{\gamma^*}^{-1}x_jx_j^{\mathrm{T}}V_{\gamma^*}^{-1}(X^{\mathrm{T}}D_{\gamma^*}^{-1}D_{\hat\gamma}^{-1}U_{\hat\gamma})(\hat\gamma - \gamma^*)^2
\end{aligned}
\tag{S18}
$$

and

$$
\begin{aligned}
&(Y_j - x_j^{\mathrm{T}}\beta_{\gamma^*})^2 - (Y_j - x_j^{\mathrm{T}}\beta^*)^2\\
=& - 2(Y_j - x_j^{\mathrm{T}}\beta^*)x_j^{\mathrm{T}}V_{\gamma^*}^{-1}(X^{\mathrm{T}}D_{\gamma^*}^{-1}\varepsilon)\\
&+ (\varepsilon^{\mathrm{T}}D_{\gamma^*}^{-1}X)V_{\gamma^*}^{-1}x_jx_j^{\mathrm{T}}V_{\gamma^*}^{-1}(X^{\mathrm{T}}D_{\gamma^*}^{-1}\varepsilon),
\end{aligned}
\tag{S19}
$$

where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}$, $D_\gamma = \mathrm{diag}(d_1 + \gamma, \ldots, d_n + \gamma)$, $V_\gamma = X^{\mathrm{T}}D_\gamma^{-1}X$, and $U_\gamma = \{x_1(Y_1 - x_1^{\mathrm{T}}\beta_\gamma), \ldots, x_n(Y_n - x_n^{\mathrm{T}}\beta_\gamma)\}^{\mathrm{T}}$. By the Cauchy–Schwartz inequality, $|x_j^{\mathrm{T}}V_{\gamma^*}^{-1}(X^{\mathrm{T}}D_{\gamma^*}^{-1}\mathcal{U})| \le (x_j^{\mathrm{T}}V_{\gamma^*}^{-1}x_j)^{1/2}(\mathcal{U}^{\mathrm{T}}D_{\gamma^*}^{-1}\mathcal{U})^{1/2}$ for $\mathcal{U} = D_{\hat\gamma}^{-1}U_{\hat\gamma}$ and $|x_j^{\mathrm{T}}V_{\gamma^*}^{-1}u| \le (x_j^{\mathrm{T}}V_{\gamma^*}^{-1}x_j)^{1/2}(u^{\mathrm{T}}V_{\gamma^*}^{-1}u)^{1/2}$ for $u = X^{\mathrm{T}}D_{\gamma^*}^{-1}\varepsilon$. By theory of linear regression, $D_{\gamma^*} - XV_{\gamma^*}^{-1}X^{\mathrm{T}} = D_{\gamma^*} - X(X^{\mathrm{T}}D_{\gamma^*}^{-1}X)^{-1}X^{\mathrm{T}}$ is nonnegative definite. Finally, $x_j^{\mathrm{T}}V_{\gamma^*}^{-1}x_j \le x_j^{\mathrm{T}}(X^{\mathrm{T}}D^{-1}X)^{-1}x_j \min_k^{-1}\{d_k/(d_k + \gamma^*)\} \le K_5n^{-(1-2\eta)}d_j \min_k^{-1}\{d_k/(d_k + \gamma^*)\}$ by Assumption (A5) and $U_{\hat\gamma}^{\mathrm{T}}D_{\gamma^*}^{-1}D_{\hat\gamma}^{-2}U_{\hat\gamma} \le (n - q)(\min_k d_k + \gamma^*)^{-1}(\min_k d_k + \hat\gamma)^{-1}$ by the definition of $\hat\gamma$. Then $(x_j^{\mathrm{T}}V_{\gamma^*}^{-1}x_j)(U_{\hat\gamma}^{\mathrm{T}}D_{\gamma^*}^{-1}D_{\hat\gamma}^{-2}U_{\hat\gamma}) \le K_5d_jn^{2\eta}(\min_k d_k)^{-1}(\min_k d_k + \hat\gamma)^{-1}$. Using these bounds and the expressions (S18)–(S19) yields

$$
\begin{aligned}
&|(Y_j - x_j^{\mathrm{T}}\beta_{\hat\gamma})^2 - (Y_j - x_j^{\mathrm{T}}\beta_{\gamma^*})^2|\\
\le&2|Y_j - x_j^{\mathrm{T}}\beta_{\gamma^*}|(x_j^{\mathrm{T}}V_{\gamma^*}^{-1}x_j)^{1/2}(U_{\hat\gamma}^{\mathrm{T}}D_{\gamma^*}^{-1}D_{\hat\gamma}^{-2}U_{\hat\gamma})^{1/2}|\hat\gamma - \gamma^*|
\end{aligned}
$$

$$
\begin{aligned}
&+ (x_j^{\mathrm{T}} V_{\gamma^*}^{-1} x_j)(U_{\hat\gamma}^{\mathrm{T}} D_{\gamma^*}^{-1} D_{\hat\gamma}^{-2} U_{\hat\gamma})(\hat\gamma - \gamma^*)^2 \\
\le & 2|Y_j - x_j^{\mathrm{T}} \beta_{\gamma^*}| d_j^{1/2} \Delta_2 + d_j \Delta_2^2, \qquad \text{(S20)}
\end{aligned}
$$

and

$$
\begin{aligned}
& |(Y_j - x_j^{\mathrm{T}} \beta_{\gamma^*})^2 - (Y_j - x_j^{\mathrm{T}} \beta^*)^2| \\
\le & 2|Y_j - x_j^{\mathrm{T}} \beta^*| (x_j^{\mathrm{T}} V_{\gamma^*}^{-1} x_j)^{1/2} \{(\varepsilon^{\mathrm{T}} D_{\gamma^*}^{-1} X) V_{\gamma^*}^{-1} (X^{\mathrm{T}} D_{\gamma^*}^{-1} \varepsilon)\}^{1/2} \\
& + (x_j^{\mathrm{T}} V_{\gamma^*}^{-1} x_j)\{(\varepsilon^{\mathrm{T}} D_{\gamma^*}^{-1} X) V_{\gamma^*}^{-1} (X^{\mathrm{T}} D_{\gamma^*}^{-1} \varepsilon)\} \\
\le & 2|Y_j - x_j^{\mathrm{T}} \beta^*| d_j^{1/2} \Delta_3 + d_j \Delta_3^2. \qquad \text{(S21)}
\end{aligned}
$$

The second desired result follows because

$$
\begin{aligned}
& \frac{\sum_j a_j^2(\gamma^*) |(Y_j - x_j^{\mathrm{T}} \beta_{\hat\gamma})^2 - (Y_j - x_j^{\mathrm{T}} \beta_{\gamma^*})^2|}{\sum_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}} \beta^*)^2} \\
\le & 2 \left\{ \frac{\sum_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}} \beta_{\gamma^*})^2}{\sum_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}} \beta^*)^2} \right\}^{1/2} \varrho^{1/2} \Delta_2 + \varrho \Delta_2^2,
\end{aligned}
$$

and

$$
\frac{\sum_j a_j^2(\gamma^*) |(Y_j - x_j^{\mathrm{T}} \beta_{\gamma^*})^2 - (Y_j - x_j^{\mathrm{T}} \beta^*)^2|}{\sum_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}} \beta^*)^2} \le 2\varrho^{1/2} \Delta_3 + \varrho \Delta_3^2,
$$

by the Cauchy–Schwartz inequality, $\sum_j a_j^2(\gamma^*)|Y_j - x_j^{\mathrm{T}} \beta_{\gamma^*}| d_j^{1/2} \le \{\sum_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}} \beta_{\gamma^*})^2\}^{1/2} \{\sum_j a_j^2(\gamma^*) d_j\}^{1/2}$ and $\sum_j a_j^2(\gamma^*)|Y_j - x_j^{\mathrm{T}} \beta^*| d_j^{1/2} \le \{\sum_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}} \beta^*)^2\}^{1/2}$ $\{\sum_j a_j^2(\gamma^*) d_j\}^{1/2}$.

The calculation leading to (S18)–(S19) also shows that

$$
\begin{aligned}
& (Y_j - x_j^{\mathrm{T}} \beta_{\hat\gamma}) - (Y_j - x_j^{\mathrm{T}} \beta^*) \\
= & x_j^{\mathrm{T}} V_{\gamma^*}^{-1} (X^{\mathrm{T}} D_{\gamma^*}^{-1} D_{\hat\gamma}^{-1} U_{\hat\gamma})(\hat\gamma - \gamma^*) - x_j^{\mathrm{T}} V_{\gamma^*}^{-1} (X^{\mathrm{T}} D_{\gamma^*}^{-1} \varepsilon).
\end{aligned}
$$

The first desired inequality follows by the bounds used to obtain (S20)–(S21). □

Proposition 4. If Assumptions (A1)–(A3) and (A5) hold with $0 \le \eta < 1/6$. Then $\sup_{\theta \in \Theta_n} E\{\sup_{0 \le \lambda \le 2} n^{-1} |\mathrm{SURE}(\delta_{\lambda, \hat\gamma, \beta_{\hat\gamma}}) - \mathrm{SURE}(\delta_{\lambda, \gamma^*, \beta^*})|\} = O\{n^{-(1-6\eta)/2}\}$.

Proof of Proposition 4. Let $G_n = \{\gamma : |\gamma - \gamma^*| \le (\min_j d_j)/2\}$. Then $\sup_{\theta \in \Theta_n} P(\hat\gamma \notin G_n) \le C_1 n^{-(1-2\eta)}$ for a constant C_1 by (S4) in the proof of Proposition 1. Moreover, let $Q_n^* = \sum_j a_j^2(\gamma^*)\{d_j + (\theta_j - x_j^{\mathrm{T}} \beta^*)^2\}$ and $B_n = \{\sum_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}} \beta^*)^2 > Q_n^*/2\}$. Then $\sup_{\theta \in \mathbb{R}^n} P(B_n^c) \le C_2 n^{-(1-3\eta)}$ for a constant C_2 by the third inequality in Lemma 2. Finally, let $Z_n = \sup_{0 \le \lambda \le \bar\lambda} n^{-1} |\mathrm{SURE}(\delta_{\lambda, \hat\gamma, \beta_{\hat\gamma}}) - \mathrm{SURE}(\delta_{\lambda, \gamma^*, \beta^*})|$. Then $Z_n \le C_3 n^{\eta}$

for a constant C_3 by Lemma 6. Combining these results, we have $E(Z_n 1_{\{\hat\gamma \notin G_n\} \cup B_n^c}) \leq C_3 n^\eta \{P(\hat\gamma \notin G_n) + P(B_n^c)\} \leq C_3(C_1 + C_2) n^{-(1-4\eta)} = o\{n^{-(1-6\eta)/2}\}$ uniformly in $\theta \in \Theta_n$, for $0 \leq \eta < 1/6$. Therefore, it suffices to show that $E(Z_n 1_{\{\hat\gamma \in G_n\} \cap B_n}) = O\{n^{-(1-6\eta)/2}\}$ uniformly in $\theta \in \Theta_n$.

If the event B_n occurs, then $\sum_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}}\beta^*)^2 \geq \sum_j a_j^2(\gamma^*) d_j^2/2$ and applying Lemma 7 shows that for all $\theta \in \Theta_n$,

$$\frac{\sum_j a_j^2(\gamma^*)|(Y_j - x_j^{\mathrm{T}}\beta_{\hat\gamma})^2 - (Y_j - x_j^{\mathrm{T}}\beta^*)^2|}{\sum_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}}\beta^*)^2} \leq \Delta_4, \tag{S22}$$

where $\Delta_4 = 2^{3/2}(1+2^{3/2}\Delta_3+2\Delta_3^2)^{1/2}\Delta_2+2^{3/2}\Delta_3+2(\Delta_2^2+\Delta_3^2)$, $\Delta_2 = K_5^{1/2} n^\eta (\min_k d_k)^{-1} |\hat\gamma - \gamma^*|$, and $\Delta_3 = K_6\, n^{3\eta/2}(n^{-1}T_{n,2})^{1/2}$ for some constant K_6, depending on K_5 and M, by (S8) in the proof of Proposition 2.

Write $b_j(\gamma^*) = \lambda c(\gamma^*) a_j(\gamma^*)/\{\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta^*)^2\}$ and $b_j(\hat\gamma) = \lambda c(\hat\gamma) a_j(\hat\gamma)/ \{\sum_k a_k^2(\hat\gamma)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2\}$, where the dependency on λ is suppressed in the notation. On the event B_n, direct calculation using (S22) and Lemma 5 (iii)–(iv) shows that

$$\begin{aligned}
b_j(\gamma^*)/b_j(\hat\gamma) &= \frac{\lambda c(\gamma^*) a_j(\gamma^*)/\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta^*)^2}{\lambda c(\hat\gamma) a_j(\hat\gamma)/\sum_k a_k^2(\hat\gamma)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2} \\
&= \frac{\lambda c(\gamma^*) a_j(\gamma^*)/\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2}{\lambda c(\hat\gamma) a_j(\hat\gamma)/\sum_k a_k^2(\hat\gamma)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2} \cdot \frac{\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2}{\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta^*)^2} \\
&\leq \frac{\lambda c(\gamma^*) a_j(\gamma^*)/\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2}{\lambda c(\hat\gamma) a_j(\hat\gamma)/\sum_k a_k^2(\hat\gamma)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2} \\
&\quad \times \left\{1 + \frac{\sum_k a_k^2(\gamma^*)|(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2 - (Y_k - x_k^{\mathrm{T}}\beta^*)^2|}{\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta^*)^2}\right\} \\
&\leq (1+\Delta_1)^2(1+\Delta_4),
\end{aligned} \tag{S23}$$

for $j = 1, \ldots, n$, where $\Delta_1 = |\hat\gamma - \gamma^*|/(\min_j d_j)$.

By similar calculation to that leading to (S11), we have

$$\begin{aligned}
&\mathrm{SURE}(\delta_{\lambda,\hat\gamma,\beta_{\hat\gamma}}) - \mathrm{SURE}(\delta_{\lambda,\gamma^*,\beta^*}) \\
&= \sum_{j \in J_{\lambda,\hat\gamma}} \frac{4 b_j(\hat\gamma) d_j a_j^2(\hat\gamma)(Y_j - x_j^{\mathrm{T}}\beta_{\hat\gamma})^2}{\sum_k a_k^2(\hat\gamma)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2} - \sum_{j \in J_{\lambda,\gamma^*}} \frac{4 b_j(\gamma^*) d_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}}\beta^*)^2}{\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta^*)^2} \\
&\quad + \sum_{j \in J_{\lambda,\hat\gamma} \cup J_{\lambda,\gamma^*}} \left\{ b_j'^2(\hat\gamma)(Y_j - x_j^{\mathrm{T}}\beta_{\hat\gamma})^2 - 2b_j'(\hat\gamma) d_j - b_j'^2(\gamma^*)(Y_j - x_j^{\mathrm{T}}\beta^*)^2 + 2b_j'(\gamma^*) d_j \right\}.
\end{aligned} \tag{S24}$$

where $b'_j(\gamma) = \min\{1, b_j(\gamma)\}$, $J_{\lambda,\gamma^*} = \{j : b_j(\gamma^*) < 1\}$, and $J_{\lambda,\hat{\gamma}} = \{j : b_j(\hat{\gamma}) < 1\}$. By similar calculation to that leading to (S12), we have

$$
\begin{aligned}
&|b'^2_j(\hat{\gamma})(Y_j - x_j^{\mathrm{T}}\beta_{\hat{\gamma}})^2 - 2b'_j(\hat{\gamma})d_j - b'^2_j(\gamma^*)(Y_j - x_j^{\mathrm{T}}\beta^*)^2 + 2b'_j(\gamma^*)d_j| \\
\le& b'^2_j(\gamma^*)|(Y_j - x_j^{\mathrm{T}}\beta_{\hat{\gamma}})^2 - (Y_j - x_j^{\mathrm{T}}\beta^*)^2| \\
&+ |b'_j(\gamma^*) - b'_j(\hat{\gamma})|\{b'_j(\gamma^*) + b'_j(\hat{\gamma})\}(Y_j - x_j^{\mathrm{T}}\beta_{\hat{\gamma}})^2 + 2|b'_j(\gamma^*) - b'_j(\hat{\gamma})|d_j \\
\le& b'^2_j(\gamma^*)|(Y_j - x_j^{\mathrm{T}}\beta_{\hat{\gamma}})^2 - (Y_j - x_j^{\mathrm{T}}\beta^*)^2| \\
&+ 2|b'_j(\gamma^*) - b'_j(\hat{\gamma})|(Y_j - x_j^{\mathrm{T}}\beta_{\hat{\gamma}})^2 + 2|b'_j(\gamma^*) - b'_j(\hat{\gamma})|d_j. \qquad \text{(S25)}
\end{aligned}
$$

On the event B_n, combining the preceding results (S22)–(S25) yields

$$
\begin{aligned}
&\sum_j |b'^2_j(\hat{\gamma})(Y_j - x_j^{\mathrm{T}}\beta_{\hat{\gamma}})^2 - 2b'_j(\hat{\gamma})d_j - b'^2_j(\gamma^*)(Y_j - x_j^{\mathrm{T}}\beta_{\gamma^*})^2 + 2b'_j(\gamma^*)d_j| \\
\le& \frac{\bar{\lambda}c(\gamma^*)}{\min_j a(\gamma^*)}\Delta_4 + 2\left\{\frac{\bar{\lambda}c(\hat{\gamma})}{\min_j a(\tilde{\gamma})} + \sum_j d_j\right\}\left\{(1+\Delta_1)^2(1+\Delta_4) - 1\right\} \\
\le& \left(\bar{\lambda}R_{\gamma^* n}\sum_j d_j\right)\Delta_4 + \left\{2(1+\bar{\lambda}R_{\gamma^* n})\sum_j d_j\right\}\left\{(1+\Delta_1)^2(1+\Delta_4) - 1\right\},
\end{aligned}
$$

and hence

$$
\begin{aligned}
nZ_n \le& 8(\max_j d_j) + \left(\bar{\lambda}R_{\gamma^* n}\sum_j d_j\right)\Delta_4 \\
&+ \left\{2(1+\bar{\lambda}R_{\gamma^* n})\sum_j d_j\right\}\left\{(1+\Delta_1)^2(1+\Delta_4) - 1\right\}.
\end{aligned}
$$

If $\hat{\gamma} \in G_n$, then $\Delta_1 \le 1/2$ and hence $(1+\Delta_1)^2(1+\Delta_4) - 1 \le \{1 + (5/2)\Delta_1\}(1 + \Delta_4) - 1 \le (5/2)\Delta_1 + (9/4)\Delta_4$. To complete the proof, it suffices to show that $\sup_{\theta\in\Theta_n} E(\Delta_4 1_{\{\tilde{\gamma}\in G_n\}}) = O\{n^{-(1-4\eta)/2}\}$, with $\hat{\gamma}$ in Δ_2 replaced by $\tilde{\gamma}$ in the proof of Proposition 2.

First, $\sup_{\theta\in\Theta_n} E(\Delta_2) = O\{n^{-(1-4\eta)/2}\}$, because $E|\tilde{\gamma} - \gamma^*| \le C_4(\min_j d_j)n^{-(1-2\eta)/2}$ uniformly in $\theta \in \Theta_n$ for a constant C_4 by the proof of Proposition 2(ii). Similarly to the proof of Proposition 2(iii), we have, for all $\theta \in \Theta_n$,

$$
\begin{aligned}
&E(|\tilde{\gamma} - \gamma^*|^2 1_{\{\tilde{\gamma}\in G_n\}}) \\
\le& 4\frac{(\max_j d_j + \gamma^*)^2}{(n-q)^2}E|T_{n,1} - T_{n,2}|^2 + (\min_j d_j/2)^2 P\{|T_{n,1} - T_{n,2}| > (n-q)/2\} \\
\le& C_5(\min_j d_j)^2 n^{-(1-2\eta)} + C_6(\min_j d_j)^2 n^{-1},
\end{aligned}
$$

where C_5 and C_6 are constants. Then $\sup_{\theta\in\Theta_n} E(\Delta_2^2 1_{\{\tilde\gamma\in G_n\}}) = O\{n^{-(1-4\eta)}\} = o\{n^{-(1-4\eta)/2}\}$ for $0 \le \eta < 1/6$. Recall that $T_{n,2} = \sum_j \xi_j^{\mathrm{T}} \sum_k \xi_k$ in the proof of Proposition 2(i). By similar calculation leading to (S10), $E(T_{n,2}) = \sum_j E(\xi_j^{\mathrm{T}}\xi_j) \le q$ for all $\theta \in \mathbb{R}^n$. Then $\sup_{\theta\in\mathbb{R}^n} E(\Delta_3) = O\{n^{-(1-3\eta)/2}\} = O\{n^{-(1-4\eta)/2}\}$ for $0 \le \eta < 1/6$. By inequality (S10), $\sup_{\theta\in\mathbb{R}^n} E(\Delta_3^2) = O\{n^{-(1-3\eta)}\} = o\{n^{-(1-4\eta)/2}\}$ and hence $\sup_{\theta\in\mathbb{R}^n} E(\Delta_2\Delta_3 1_{\{\tilde\gamma\in G_n\}}) = o\{n^{-(1-4\eta)/2}\}$ for $0 \le \eta < 1/6$. Combining these results completes the proof. □

Proof of Theorem 5. Let $G_n = \{\gamma : |\gamma-\gamma^*| \le (\min_j d_j)/2 \ \& \ |\gamma-\gamma^*| \le (\min_j d_j)n^{-\eta}/(16K_5^{1/2})\}$. Then $\sup_{\theta\in\Theta_n} P(\hat\gamma \notin G_n) \le C_1 n^{-(1-4\eta)}$ for a constant C_1 by (S7) in the proof of Proposition 2. Moreover, let $Q_n^* = \sum_j a_j^2(\gamma^*)\{d_j + (\theta_j - x_j^{\mathrm{T}}\beta^*)^2\}$ and $D_n = \{(n^{-1}T_{n,2})^{1/2} \le n^{-3\eta/2}/(16K_6)\}$, where K_6 is a constant determined such that (S22) holds on the event $\{\sum_j a_j^2(\gamma^*)(Y_j - x_j^{\mathrm{T}}\beta^*)^2 > Q_n^*/2\}$. Then $\sup_{\theta\in\mathbb{R}^n} P(D_n^c) \le C_2 n^{-(1-3\eta)}$ for a constant C_2 by Chebyshev's inequality and the fact that $E(T_{n,2}) \le q$ from the proof of Proposition 4.

(i) It suffices to show that for any $\tau_1 > 0$, $\sup_{\theta\in\mathbb{R}^n} P[D_n \cap \{\hat\gamma \in G_n\} \cap \{\sup_{0\le\lambda\le\bar\lambda} |\zeta_n(\lambda,\hat\gamma,\beta_{\hat\gamma})| \ge \tau_2 n^{-(1-4\eta)/2}\}] \le \tau_1$ for all large enough τ_2 and n.

Take $v_n = \tau_2 n^{-(1-4\eta)/2}$. If $\hat\gamma \in G_n$, then $\sup_k |a_k(\hat\gamma)/a_k(\gamma^*) - 1| \le 1/2$ by Lemma 5(iii). By the proof of (S13), we have

$$\sup_{\theta\in\mathbb{R}^n} P\left\{\hat\gamma \in G_n \ \& \ n^{-1}\sup_{0\le\lambda\le\bar\lambda}\left|\sum_{j\in J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}} \lambda c(\hat\gamma)\, a_j(\hat\gamma)(d_j - \varepsilon_j^2)\right| \ge \frac{9}{2}Q_n^* v_n\right\} \le \frac{8K_1\bar\lambda^2(\max_j d_j)(\sum_j d_j)}{n^2\, v_n^2}, \tag{S26}$$

in parallel to the first inequality in Lemma 2.

If $D_n \cap \{\hat\gamma \in G_n\}$ occurs, then $\Delta_2 \le 1/16$ and $\Delta_3 \le 1/16$, where Δ_2 and Δ_3 are defined after (S22). If further $\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta^*)^2 > Q_n^*/2$, then (S22) holds with $\Delta_4 \le 1/2$ and hence $\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2 \ge \sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta^*)^2/2 > Q_n^*/4$. Therefore, if $D_n \cap \{\hat\gamma \in G_n\}$ occurs, then $\sum_k a_k^2(\gamma^*)\ (Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2 \le Q_n^*/4$ implies $\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta^*)^2 \le Q_n^*/2$. By this relationship and the proof of (S15), we have

$$\sup_{\theta\in\mathbb{R}^n} P\left[D_n \cap \{\hat\gamma \in G_n\} \cap \left\{\sum_k a_k^2(\hat\gamma)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2 \le \frac{1}{16}Q_n^*\right\}\right]$$

$$\leq \sup_{\theta\in\mathbb{R}^n} P\left[D_n \cap \{\hat\gamma \in G_n\} \cap \left\{\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2 \leq \frac{1}{4}Q_n^*\right\}\right]$$

$$\leq \sup_{\theta\in\mathbb{R}^n} P\left\{\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta^*)^2 \leq Q_n^*/2\right\} \leq \frac{(32+8K_1)\max_k\{a_k^2(\gamma^*)d_k\}}{\sum_k a_k^2(\gamma^*)d_k}, \quad \text{(S27)}$$

in parallel to the third inequality in Lemma 2.

To extend the second inequality in Lemma 2, we have

$$P\left\{\hat\gamma \in G_n \;\&\; n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}\left|\sum_{j\in J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}} \lambda c(\hat\gamma)\, a_j(\hat\gamma)(\theta_j - x_j^{\mathrm{T}}\beta_{\hat\gamma})\varepsilon_j\right| \geq \frac{27}{2}Q_n^* v_n\right\}$$

$$\leq P\left\{\hat\gamma \in G_n \;\&\; n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}\left|\sum_{j\in J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}} \lambda c(\hat\gamma)\, a_j(\hat\gamma)(\theta_j - x_j^{\mathrm{T}}\beta^*)\varepsilon_j\right| \geq \frac{9}{2}Q_n^* v_n\right\}$$

$$+ P\left\{n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}\left|\sum_{j\in J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}} \lambda c(\hat\gamma)\, a_j(\hat\gamma)\varepsilon_j x_j^{\mathrm{T}}(\beta^* - \beta_{\hat\gamma})\right| \geq 9Q_n^* v_n\right\}. \quad \text{(S28)}$$

The first term on the right-hand side is, uniformly in $\theta \in \mathbb{R}^n$, no greater than $8\bar\lambda^2(\max_j d_j)(\sum_j d_j)/(n^2 v_n^2)$, similarly to (S14). The second term is, uniformly in $\theta \in \mathbb{R}^n$, $O\{\tau_2^{-2} + n^{-(1-\eta)}\}$. In fact, by Proposition 4 and the Cauchy–Schwartz inequality, the second term is, for all $\theta \in \mathbb{R}^n$, no greater than

$$P\left\{n^{-1}\sum_j \bar\lambda c(\gamma^*)\, a_j(\gamma^*)\left|\varepsilon_j x_j^{\mathrm{T}}(\beta^* - \beta_{\hat\gamma})\right| \geq 4Q_n^* v_n\right\}$$

$$\leq P\left\{n^{-1}\sum_j \bar\lambda c(\gamma^*)\, a_j(\gamma^*) d_j^{1/2}|\varepsilon_j|(\Delta_2 + \Delta_3) \geq 2Q_n^* v_n\right\}$$

$$\leq P\left\{(\Delta_2+\Delta_3) \geq v_n/(2^{1/2}K_2)\right\} + P\left\{n^{-1}\sum_j \bar\lambda c(\gamma^*)\, a_j(\gamma^*) d_j^{1/2}|\varepsilon_j| \geq 2^{3/2}K_2 Q_n^*\right\}$$

$$\leq P\left\{(\Delta_2+\Delta_3) \geq v_n/(2^{1/2}K_2)\right\}$$

$$+ P\left[n^{-2}c^2(\gamma^*)\left\{\sum_j a_j^2(\gamma^*)d_j\right\}\left(\sum_j \varepsilon_j^2\right) \geq 2K_2^2 Q_{\gamma^* n}^2\right].$$

By (S7) and Chebyshev's inequality, the first term on the right-hand side of the last inequality is no greater than $P\{|\hat\gamma - \gamma^*| \geq (\min_k d_k) n^{-\eta} v_n/(2^{3/2}K_2^{1/2}K_2)\} + P\{(n^{-1}T_{n,2})^{1/2} \geq n^{-3\eta/2}v_n/(2^{3/2}K_5K_2)\} = O\{n^{4\eta}(nv_n^2) + n^{3\eta}(nv_n^2)\} = O(\tau_2^{-2})$. By Chebyshev's inequality and the fact that $n^{-2}c^2(\gamma^*)\{\sum_j a_j^2(\gamma^*)d_j\}(\sum_j d_j) \leq K_2^2 Q_n^{*2}$, the second term is no greater than $P[n^{-2}c^2(\gamma^*)\{\sum_j a_j^2(\gamma^*)d_j\}\{\sum_j(\varepsilon_j^2 - d_j^2)\} \geq K_2^2 Q_n^{*2}] \leq P\{\sum_j(\varepsilon_j^2 - d_j^2) \geq \sum_j d_j\} \leq K_1(\sum_j d_j^2)/(\sum_j d_j)^2 = O\{n^{-(1-\eta)}\}$.

Write $R_{n,\gamma} = \{\max_k a_k(\gamma)\}/\{\min_k a_k(\gamma)\}$ and $R_n = \sup_{\gamma\in G_n} R_{n,\gamma}$. By the proofs of (S16)–(S17), we have

$$\sup_{\theta\in\mathbb{R}^n} P\left\{ n^{-1} \sup_{0\le\lambda\le\bar\lambda} \left| \sum_{j\notin J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}} (d_j - \varepsilon_j^2) \right| \ge 2v_n \right\} \le \frac{2K_1 \sum_j d_j^2}{n^2\, v_n^2}, \tag{S29}$$

$$\sup_{\theta\in\mathbb{R}^n} P\left[D_n \cap \{\hat\gamma \in G_n\} \cap \left\{ n^{-1} \sup_{0\le\lambda\le\bar\lambda} \left| \sum_{j\notin J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}} (\theta_j - x_j^{\mathrm{T}}\beta^*)\varepsilon_j \right| \ge 2v_n \right\}\right]$$
$$\le \max\left[\frac{(32+8K_1)\max_k\{a_k^2(\gamma^*)d_k\}}{\sum_k a_k^2(\gamma^*)d_k},\ \frac{72\bar\lambda R_n^3(\max_k d_k)(\sum_k d_k)}{n^2 v_n^2} \right]. \tag{S30}$$

To show the inequality (S30), we have

$$P\left[D_n \cap \{\hat\gamma \in G_n\} \cap \left\{ n^{-1} \sup_{0\le\lambda\le\bar\lambda} \left| \sum_{j\notin J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}} (\theta_j - x_j^{\mathrm{T}}\beta^*)\varepsilon_j \right| \ge 2v_n \right\}\right]$$
$$\le P\left[D_n \cap \{\hat\gamma \in G_n\} \cap \left\{ \sum_k a_k^2(\hat\gamma)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2 \le \bar\lambda R_{n,\hat\gamma} \sum_k a_k^2(\hat\gamma) d_k \right\}\right]$$
$$\le P\left[D_n \cap \{\hat\gamma \in G_n\} \cap \left\{ \sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2 \le 9\bar\lambda R_n \sum_k a_k^2(\gamma^*) d_k \right\}\right].$$

If $\sum_k a_k^2(\gamma^*)(\theta_k - x_k^{\mathrm{T}}\beta^*)^2 > 36\bar\lambda R_n \sum_k a_k^2(\gamma^*)d_k$, then this is no greater than $P[D_n \cap \{\hat\gamma \in G_n\} \cap \{\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2 \le Q_n^*/4\}] \le P\{\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta^*)^2 \le Q_n^*/2\}$ by the proof of (S27). On the other hand, we have

$$P\left\{ n^{-1} \sup_{0\le\lambda\le\bar\lambda} \left| \sum_{j\notin J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}} (\theta_j - x_j^{\mathrm{T}}\beta^*)\varepsilon_j \right| \ge 2v_n \right\}$$
$$\le P\left(n^{-1} \max_k \left| \sum_{j=1}^k (\theta_j - x_j^{\mathrm{T}}\beta^*)\varepsilon_j \right| \ge v_n \right) + P\left(n^{-1} \max_k \left| \sum_{j=k}^n (\theta_j - x_j^{\mathrm{T}}\beta^*)\varepsilon_j \right| \ge v_n \right)$$
$$\le \frac{2(\max_j d_j)\{\sum_j a_j^2(\gamma^*)(\theta_j - x_j^{\mathrm{T}}\beta^*)^2\}}{n^2\, v_n^2 \{\min_j a_j^2(\gamma^*)\}},$$

If $\sum_k a_k^2(\gamma^*)(\theta_k - x_k^{\mathrm{T}}\beta^*)^2 \le 36\bar\lambda R_n \sum_k a_k^2(\gamma^*)d_k$, then this is no greater than $72\bar\lambda R_n^3$ $(\max_k d_k)(\sum_k d_k)/(n^2 v_n^2)$. Combining the two cases gives the desired inequality.

To extend the second inequality in Lemma 3, we have

$$P\left[D_n \cap \{\hat\gamma \in G_n\} \cap \left\{ n^{-1} \sup_{0\le\lambda\le\bar\lambda} \left| \sum_{j\notin J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}} (\theta_j - x_j^{\mathrm{T}}\beta_{\hat\gamma})\varepsilon_j \right| \ge 6v_n \right\}\right]$$

$$\leq P\left[D_n \cap \{\hat\gamma \in G_n\} \cap \left\{ n^{-1} \sup_{0\leq\lambda\leq\bar\lambda} \left| \sum_{j\notin J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}} (\theta_j - x_j^{\mathrm{T}}\beta^*)\varepsilon_j \right| \geq 2v_n \right\}\right]$$
$$+ P\left\{ n^{-1} \sup_{0\leq\lambda\leq\bar\lambda} \left| \sum_{j\notin J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}} (x_j^{\mathrm{T}}\beta^* - x_j^{\mathrm{T}}\beta_{\hat\gamma})\varepsilon_j \right| \geq 4v_n \right\}. \qquad \text{(S31)}$$

The first term on the right-hand side is bounded by (S30). The second term is, uniformly in $\theta \in \mathbb{R}^n$, $O\{\tau_2^{-2} + n^{-(1-\eta)}\}$. In fact, by Lemma 7 and the Cauchy-Schwartz inequality, the second term is, for all $\theta \in \mathbb{R}^n$, no greater than

$$P\left\{ n^{-1} \sum_j |\varepsilon_j x_j^{\mathrm{T}}(\beta^* - \beta_{\hat\gamma})| \geq 4v_n \right\}$$
$$\leq P\left\{ n^{-1} \sum_j d_j^{1/2} |\varepsilon_j| (\Delta_2 + \Delta_3) \geq 2v_n \right\}$$
$$\leq P\left\{ (\Delta_2 + \Delta_3) \geq 2^{1/2} v_n / K_2 \right\} + P\left(n^{-1} \sum_j d_j^{1/2} |\varepsilon_j| \geq 2^{1/2} K_2 \right)$$
$$\leq P\left\{ (\Delta_2 + \Delta_3) \geq 2^{1/2} v_n / K_2 \right\} + P\left\{ n^{-2} \left(\sum_j d_j\right)\left(\sum_j \varepsilon_j^2\right) \geq 2K_2^2 \right\}.$$

The first term on the right-hand side of the last inequality is $O(\tau_2^{-2})$, as shown when handling the last term in (S28). By Chebyshev's inequality and the fact that $n^{-1}\sum_j d_j \leq K_2$, the second term is no greater than $P\{\sum_j(\varepsilon_j^2 - d_j) \geq \sum_j d_j\} \leq K_1(\sum_j d_j^2)/(\sum_j d_j)^2 = O\{n^{-(1-\eta)}\}$.

Write $Z_{n,3} = n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}[\lambda c(\hat\gamma) \max_{j\in J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}}\{d_j a_j(\hat\gamma)\}/\{\sum_k a_k^2(\hat\gamma)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2\}]$. Then $Z_{n,3} \leq n^{-1}(\max_j d_j)$ by the proof of the third inequality in Lemma 4. Combining the preceding results (S26)–(S31) completes the proof of (i).

(ii) Write $Z_n = \sup_{0\leq\lambda\leq\bar\lambda}|\zeta_n(\lambda, \hat\gamma, \beta_{\hat\gamma})|$. Applying Lemma 4 with $a_j = a_j(\hat\gamma)$ and Y_j replaced by $Y_j - x_j^{\mathrm{T}}\beta_{\hat\gamma}$ shows that $E(Z_n^2) \leq C_3 n^{2\eta}$ for a constant C_3. By the Cauchy–Schwartz inequality, $E(Z_n 1_{D_n^c \cup \{\hat\gamma \notin G_n\}}) \leq E^{1/2}(Z_n^2) P^{1/2}(D_n^c \cup \{\hat\gamma \notin G_n\}) \leq (C_1 + C_2)^{1/2} C_3^{1/2} n^{-(1-6\eta)/2}$ for all $\theta \in \Theta_n$. To complete the proof, it suffices to show that $E(Z_n 1_{D_n \cap \{\hat\gamma \in G_n\}}) = O\{n^{-(1-5\eta)/2}\}$ uniformly in $\theta \in \mathbb{R}^n$.

Write $Z_{n,1} = n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}|\sum_{j\in J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}} \lambda c(\hat\gamma) a_j(\hat\gamma)\{d_j - \varepsilon_j(Y_j - x_j^{\mathrm{T}}\beta_{\hat\gamma})\}/\{\sum_k a_k^2(\hat\gamma)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2\}|$ and $Z_{n,2} = n^{-1}\sup_{0\leq\lambda\leq\bar\lambda}|\sum_{j\notin J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}}\{d_j - \varepsilon_j(Y_j - x_j^{\mathrm{T}}\beta_{\hat\gamma})\}|$. Then $\sup_{0\leq\lambda\leq\bar\lambda} |\zeta_n(\lambda, \hat\gamma, \beta_{\hat\gamma})| \leq n^{-1}|\sum_j(d_j - \varepsilon_j^2)| + Z_{n,1} + Z_{n,2} + Z_{n,3}$. Note that $E^{1/2}[\{n^{-1}\sum_j(d_j - \varepsilon_j^2)\}^2] = O\{n^{-(1-\eta)/2}\}$ and, by Lemma 4, $Z_{n,3}$ is bounded from above by $O\{n^{-(1-\eta)}\}$.

Following the proof of Theorem 3(ii), we show that $E(Z_{n,1}1_{D_n\cap\{\hat\gamma\in G_n\}}) = O\{n^{-(1-5\eta)/2}\}$ and $E(Z_{n,2}1_{D_n\cap\{\hat\gamma\in G_n\}}) = O\{n^{-(1-5\eta)/2}\}$ uniformly in $\theta \in \mathbb{R}^n$.

Write $B_{n,1} = \{\sum_k a_k^2(\hat\gamma)(Y_k - x_k^{\mathrm{T}}\beta_{\hat\gamma})^2 \le Q_n^*/16\}$. Similarly to inequalities (S1) and (S28), we have, for all $\theta \in \mathbb{R}^n$,

$$
\begin{aligned}
&E(Z_{n,1}1_{D_n\cap\{\hat\gamma\in G_n\}}) = E(Z_{n,1}1_{B_{n,1}^c\cap D_n\cap\{\hat\gamma\in G_n\}}) + E(Z_{n,1}1_{B_{n,1}\cap D_n\cap\{\hat\gamma\in G_n\}}) \\
\le& E^{1/2}\left(1_{\{\hat\gamma\in G_n\}}\sup_{0\le\lambda\le\bar\lambda}\left[n^{-1}\sum_{j\in J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}}\lambda c(\hat\gamma)a_j(\hat\gamma)\{d_j-\varepsilon_j(Y_j-x_j^{\mathrm{T}}\beta^*)\}/(Q_n^*/16)\right]^2\right) \\
&+ E\left\{1_{\{\hat\gamma\in G_n\}}\sup_{0\le\lambda\le\bar\lambda}\left|n^{-1}\sum_{j\in J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}}\lambda c(\hat\gamma)\,a_j(\hat\gamma)\varepsilon_j x_j^{\mathrm{T}}(\beta^*-\beta_{\hat\gamma})/(Q_n^*/16)\right|\right\} \\
&+ E^{1/2}(Z_{n,1}^2)P^{1/2}(B_{n,1}\cap D_n\cap\{\hat\gamma\in G_n\}).
\end{aligned}
$$

By the proofs of (S26)–(S27), the first and third terms are, respectively, $O\{n^{-(1-\eta)/2}\}$ and $O\{n^{-(1-5\eta)/2}\}$. By the Cauchy–Schwartz inequality, the second term is no greater than $(3/2)^2E[1_{\{\hat\gamma\in G_n\}}n^{-1}\sum_j \bar\lambda c(\gamma^*)\,a_j(\gamma^*)\left|\varepsilon_j x_j^{\mathrm{T}}(\beta^*-\beta_{\hat\gamma})\right|/(Q_n^*/16)] \le 2(3/2)^2E^{1/2}[1_{\{\hat\gamma\in G_n\}}(\Delta_2+\Delta_3)^2]E^{1/2}[n^{-2}\bar\lambda^2c^2(\gamma^*)\{\sum_j a_j^2(\gamma^*)d_j\}(\sum_j \varepsilon_j^2)/(Q_n^*/16)^2] = O\{n^{-(1-4\eta)/2}\}$, because $c^2(\gamma^*) \le \{\sum_j a_j^2(\gamma^*)d_j\}(\sum_j d_j)$, and $\sup_{\theta\in\Theta_n} E(\Delta_2^2 1_{\{\hat\gamma\in G_n\}}) = O\{n^{-(1-4\eta)}\}$ and $\sup_{\theta\in\mathbb{R}^n} E(\Delta_3^2) = O\{n^{-(1-4\eta)}\}$ by the proof of Proposition 4. Combining the three cases shows that $\sup_{\theta\in\mathbb{R}^n} E(Z_{n,1}1_{D_n\cap\{\hat\gamma\in G_n\}}) = O\{n^{-(1-5\eta)/2}\}$.

Similarly to inequalities (S2) and (S31), we have

$$
\begin{aligned}
&E(Z_{n,2}1_{D_n\cap\{\hat\gamma\in G_n\}}) \\
\le& E^{1/2}\left(\sup_{0\le\lambda\le\bar\lambda}\left[n^{-1}\sum_{j\notin J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}}\{d_j-\varepsilon_j(Y_j-x_j^{\mathrm{T}}\beta^*)\}\right]^2\right) \\
&+ E\left\{1_{\{\hat\gamma\in G_n\}}\sup_{0\le\lambda\le\bar\lambda}\left|n^{-1}\sum_{j\notin J_{\lambda,\hat\gamma,\beta_{\hat\gamma}}}(x_j^{\mathrm{T}}\beta^*-x_j^{\mathrm{T}}\beta_{\hat\gamma})\varepsilon_j\right|\right\} \\
\le& \left[\frac{2\{K_1\sum_j d_j^2+\sum_j(\theta_j-x_j^{\mathrm{T}}\beta^*)^2d_j\}}{n^2}\right]^{1/2} \\
&+ 2E\left\{1_{\{\hat\gamma\in G_n\}}n^{-1}\sum_j d_j^{1/2}|\varepsilon_j|(\Delta_2+\Delta_3)\right\}.
\end{aligned}
$$

By the Cauchy-Schwartz inequality, the second term on the right-hand side of the last inequality is no greater than $2E^{1/2}[1_{\{\hat\gamma\in G_n\}}\,(\Delta_2+\Delta_3)^2]E^{1/2}\{n^{-2}(\sum_j d_j)(\sum_j \varepsilon_j^2)\} =$

$O\{n^{-(1-4\eta)/2}\}$. If $\sum_k a_k^2(\gamma^*)(\theta_k - x_k^{\mathrm{T}}\beta^*)^2 \leq 36\bar{\lambda}R_n \sum_k a_k^2(\gamma^*)d_k$, then, similarly as in the proof of Lemma 3, the first term is no greater than $C_4\{n^{-(1-\eta)/2} + n^{-(1-4\eta)/2}\}$, where C_4 is a constant (free of θ). Moreover, write $B_{n,2} = \{\sum_k a_k^2(\hat{\gamma})(Y_k - x_k^{\mathrm{T}}\beta_{\hat{\gamma}})^2 \leq \bar{\lambda}R_{n,\hat{\gamma}} \sum_k a_k^2(\hat{\gamma})d_k\}$. Similarly to inequality (S3), we have

$$\begin{aligned} &E(Z_{n,2}1_{D_n\cap\{\hat{\gamma}\in G_n\}}) = E(Z_{n,2}1_{B_{n,2}\cap D_n\cap\{\hat{\gamma}\in G_n\}}) \\ \leq &E^{1/2}(Z_{n,2}^2)P^{1/2}(B_{n,2}\cap D_n \cap \{\hat{\gamma}\in G_n\}). \end{aligned}$$

If $\sum_k a_k^2(\gamma^*)(\theta_k - x_k^{\mathrm{T}}\beta^*)^2 > 36\bar{\lambda}R_n \sum_k a_k^2(\gamma^*)d_k$, then, by the proof of (S30), $P(B_{n,2}\cap D_n\cap\{\hat{\gamma}\in G_n\}) \leq P\{\sum_k a_k^2(\gamma^*)(Y_k - x_k^{\mathrm{T}}\beta^*)^2 \leq Q_n^*/2\}$ and hence $E(Z_{n,2}1_{D_n\cap\{\hat{\gamma}\in G_n\}}) \leq C_5 n^{-(1-5\eta)/2}$, where C_5 is a constant (free of θ). Combining the two cases shows that $\sup_{\theta\in\mathbb{R}^n} E(Z_{n,2}1_{D_n\cap\{\hat{\gamma}\in G_n\}}) = O\{n^{-(1-5\eta)/2}\}$. □

References

Speckman, P. (1985) "Spline smoothing and optimal rates of convergence in nonparametric regression models," *Annals of Statistics*, 13, 970–983.