Supplement to "Bootstrapping LASSO estimators under variable selection consistency in high dimensions and some higher order refinements"

DEBRAJ DAS[1,a], ARINDAM CHATTERJEE[2,b] and SOUMENDRA N. LAHIRI[3,c]

[1]*Department of Mathematics, Indian Institute of Technology, Bombay, Mumbai 400076, India,* [a]*debrajdas@math.iitb.ac.in*

[2]*Statistical Sciences Unit, Indian Statistical Institute Delhi, 7, S. J. S. Sansanwal Marg, New Delhi 110016, India,* [b]*cha@isid.ac.in*

[3]*Department of Mathematics and Statistics, Washington University in St. Louis, MO 63130, United States,* [c]*s.lahiri@wustl.edu*

Abstract. Additional simulation results are provided. The performance of the PB in the heteroscedastic errors setting is studied. Two tables comparing the coverages obtained by RB and PB based two-sided and two-sided symmetric CIs (in the homoscedastic case) are provided which clearly show the superior coverage of the two-sided symmetric CIs. Additional figures displaying the coverages of different one-sided and two-sided CIs at different parameter settings are displayed. Detailed proofs of Theorems 3, 7, 8 and 9 are presented. Statements and proofs of a set of preliminary lemmas (numbered Lemma 1 - 8) that are utilized for proving all the results are also stated and proved.

1. Additional Simulation results

Here we report some additional simulation results in the homoscedastic errors case, focusing on variations in the quality of Bootstrap CIs as we look at other nonzero components. As suggested by a reviewer, we also study empirical coverages for PB based CIs in the heteroscedastic errors case and the results are presented in Section 1.1.

As in the main paper, in order to study the coverages of the above CIs, we used various tuning parameters, which were of the form $\lambda_n = k\lambda_{\text{CV}}$, where the constant $k(\geq 1)$ is changed, and λ_{CV} is the five-fold cross validation (CV) based choice of the tuning parameter. The plots shown in Figure 1 (for $n = 150$) and Figure 2 (for $n = 300$) find the empirical coverage for the above mentioned collection of CIs for target parameter β_{10} (for both two-sided and one-sided cases) and average lengths for the two sided case, at various choices of k, when $F = F_2$. Similar results are obtained when the target parameter is β_1. In case of $F = F_1$, the performance of the bootstrap CIs were qualitatively similar to $F = F_2$ for both β_1 and β_{10}. Hence, they are not included. All simulated data, relevant code and other details related to the simulation study presented in this article are available at github.com/chaisid24/lasso-hd-rb-pb-bernoulli-.

In addition, to better illustrate the improvements in coverage obtained by using two sided RB and PB based symmetric CIs, we include Tables 1 and 2. In these tables, we provide the exact numerical values of empirical coverages attained by RB and PB based two-sided and symmetric two-sided CIs at a specific choice of (n, p), with a given target parameter and at a range of choices of k. Both these tables consider the homoscedastic errors setting. In Table 1, we provide the values for $(n, p) = (150, 500)$,

$F = F_1$ with target parameter β_1, while Table 2 considers the same parameters with $F = F_2$. The figures shown in these two tables clearly show that there is considerable gain in using two-sided symmetric CIs in comparison to two-sided CIs, for both RB and PB based settings. For other choices of (n, p) and other target parameters used in our simulation study, similar gains are observed while using the two-sided symmetric CIs, over a wide range of choices of k. The numbers obtained in Tables 1 and 2 can be obtained for other choices of (n, p) and other target parameters using the code and output data provided in the github page github.com/chaisid24/lasso-hd-rb-pb-bernoulli-.

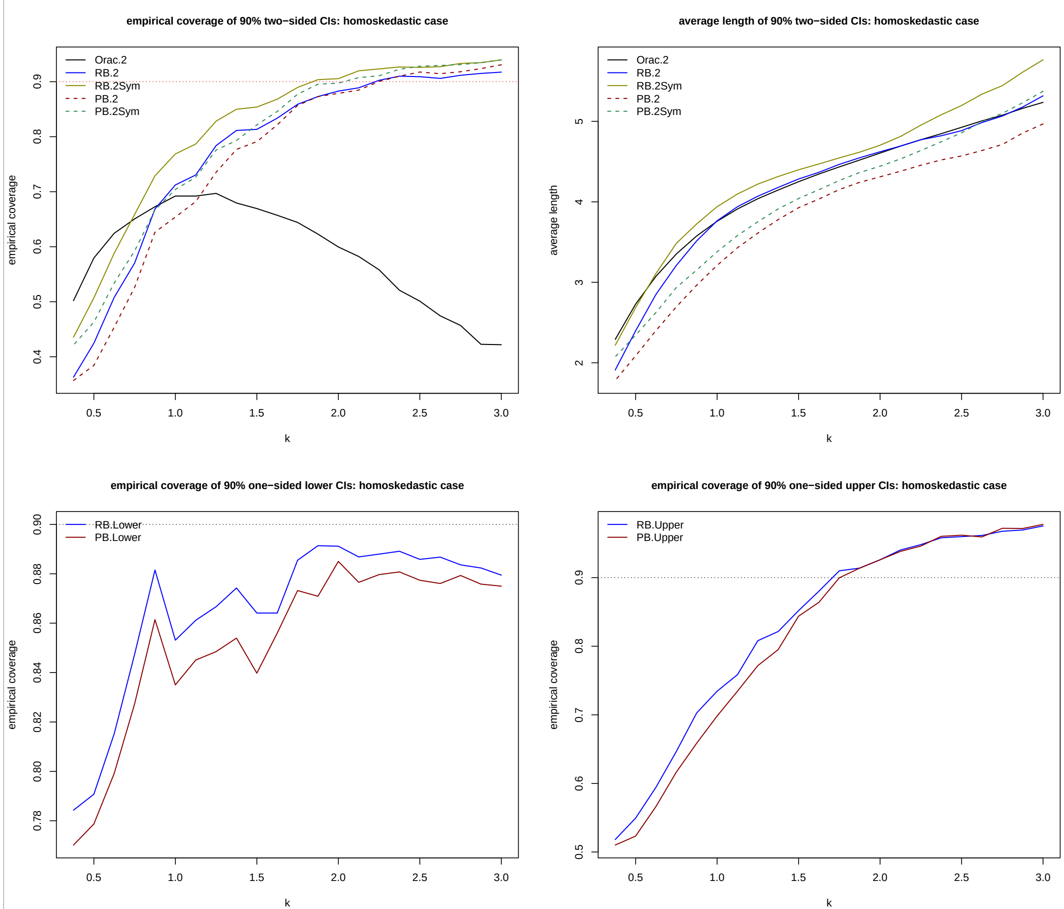

Figure 1: **Top left panel** shows empirical coverages of two sided 90% confidence intervals (CIs) for target parameter β_{10}, in the **homoscedastic** case as the constant k increases. These include the oracle based, RB based and PB based cases. The **top right panel** shows the average lengths of these two sided CIs. **Bottom left panel** shows the empirical coverages of RB and PB based one-sided lower CIs, while the **bottom right panel** shows the same for one-sided upper CIs. These plots are for $n = 150$, $p = 500$, $F = F_2$.

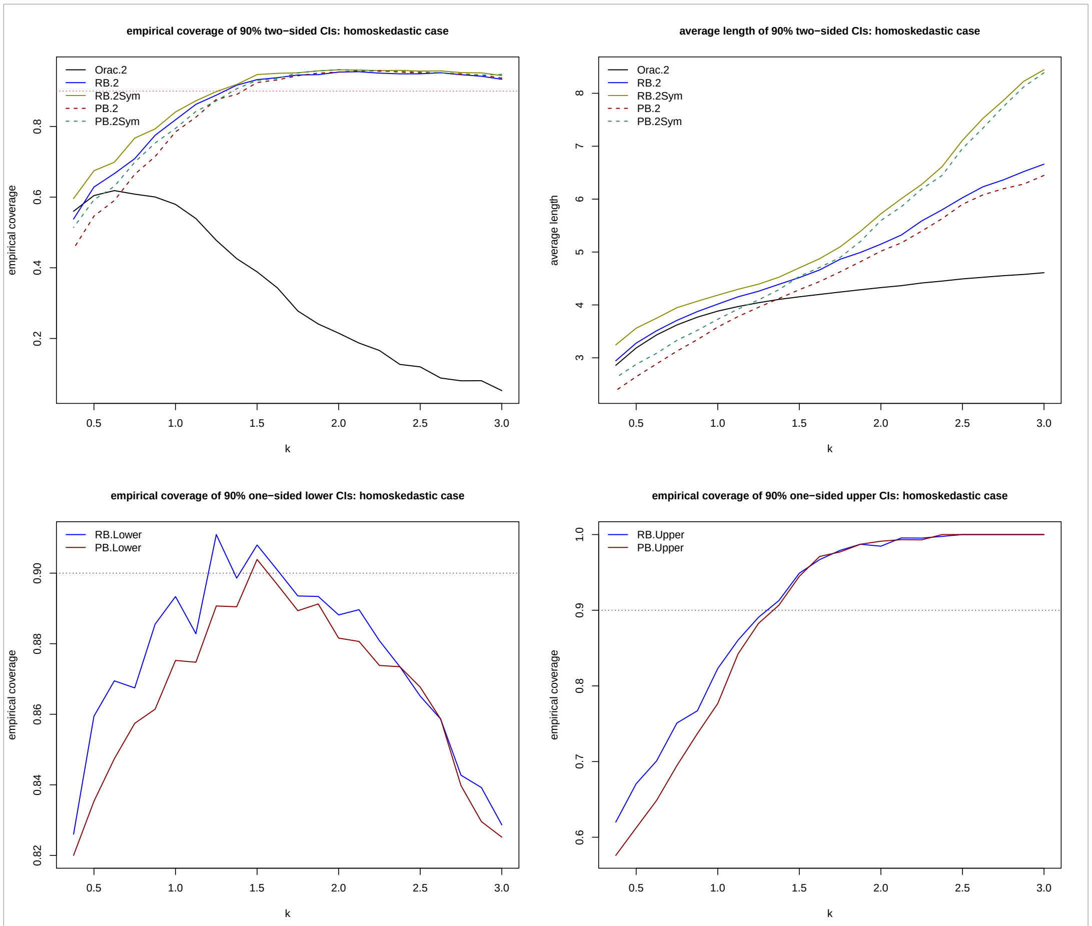

Figure 2: **Top left panel** shows empirical coverages of two sided 90% confidence intervals (CIs) for target parameter β_{10}, in the **homoscedastic** case as the constant k increases. These include the oracle based, RB based and PB based cases. The **top right panel** shows the average lengths of these two sided CIs. **Bottom left panel** shows the empirical coverages of RB and PB based one-sided lower CIs, while the **bottom right panel** shows the same for one-sided upper CIs. These plots are for $n = 300$, $p = 500$, $F = F_2$.

1.1. Simulation results for the heteroscedastic case

In this situation, the errors $\{\epsilon_i : 1 \le i \le n\}$ are considered to be heteroscedastic. Assume $(x_{i,1}, \dots, x_{i,p})$ denotes a p-dimensional covariate vector for the i-th observation. We generate $\epsilon_1, \dots, \epsilon_n$, independently with mean zero and variance σ_i^2, where

$$\sigma_i^2 = 3\sqrt{|x_{i,1}|} + |x_{i,2}| + 0.2, \quad \text{for all } i = 1, \dots, n.$$

In practice, we generated $\{\widetilde{\epsilon}_i : 1 \le i \le n\}$ as an i.i.d. sample from error distribution $F = F_2$, and then $\widetilde{\epsilon}_i$ was scaled by σ_i to obtain $\epsilon_i = \sigma_i \widetilde{\epsilon}_i$, for each $i = 1, \dots, n$. In this case we only considered

Table 1. Empirical coverages of various 90% two sided RB and PB based CIs. Target parameter is β_1, with $F = F_1$ and at $(n, p) = (150, 500)$.

		RB Based CIs		PB Based CIs	
k	Count[a]	RB two-sided	RB two-sided symmetric	PB two-sided	PB two-sided symmetric
3.000	313	0.8882	0.9105	0.8722	0.9042
2.750	350	0.8914	0.9114	0.8829	0.9029
2.500	388	0.8943	0.9046	0.8866	0.8918
2.250	425	0.8894	0.8871	0.8706	0.8729
2.000	448	0.8772	0.8683	0.8594	0.8616
1.750	468	0.8654	0.8739	0.8462	0.8568
1.500	485	0.8412	0.8495	0.8206	0.8268

[a]The number of times (out of 500 Monte-Carlo replications) in which LASSO estimate of β_1 was found to be non-zero, while using $\lambda_n = k\lambda_{\text{CV}}$.

Table 2. Empirical coverages of various 90% two sided RB and PB based CIs. Target parameter is β_1, with $F = F_2$ and at $(n, p) = (150, 500)$.

		RB Based CIs		PB Based CIs	
k	Count[a]	RB two-sided	RB two-sided symmetric	PB two-sided	PB two-sided symmetric
3.000	448	0.9174	0.9397	0.9308	0.9397
2.750	464	0.9116	0.9332	0.9181	0.9310
2.500	473	0.9091	0.9260	0.9175	0.9281
2.250	482	0.9025	0.9232	0.9004	0.9108
2.000	487	0.8830	0.9055	0.8789	0.8973
1.750	489	0.8589	0.8896	0.8569	0.8773
1.500	493	0.8134	0.8540	0.7911	0.8215

[a]The number of times (out of 500 Monte-Carlo replications) in which LASSO estimate of β_1 was found to be non-zero, while using $\lambda_n = k\lambda_{\text{CV}}$.

$(n, p) = (300, 500)$ and we found that with the above choice of σ_i's, we obtain $\sigma_i^2 \in (0.3, 2)$. The true regression coefficient and the design matrix remain unchanged, as in the homoscedastic case. Even in heteroscedastic case, to simplify notation we write, $\{\epsilon_i : 1 \leq i \leq n\}$ are independently generated from F_2. Figures 3 and 4 show the empirical coverages of two-sided oracle based, PB based, PB symmetric CIs, their average lengths and empirical coverages of one-sided lower and upper PB based CIs, for β_1 and β_{10} respectively. The plots show that very stable coverages are obtained in the two sided case, for both PB based CIs, while the oracle based CI performs very badly. The results are less stable in the one-sided case.

2. Details of Proofs of Theorems 3, 7, 8 and 9

First we define few additional notations. Then we move to state some preliminary lemmas before proving Theorems 3, 7, 8 and 9.

2.1. Notations

We denote by $\|\cdot\|$ and $\|\cdot\|_\infty$, respectively, the L^2 and L^∞ norm. For a non-negative integer-valued vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_l)'$ and a function $f = (f_1, f_2, \ldots, f_l) : \mathscr{R}^l \rightarrow \mathscr{R}^l$, $l \geq 1$, write $|\boldsymbol{\alpha}| = \alpha_1 + \ldots + \alpha_l$, $\boldsymbol{\alpha}! = \alpha_1! \ldots \alpha_l!$, $f^{\boldsymbol{\alpha}} = (f_1^{\alpha_1}) \ldots (f_l^{\alpha_l})$, and $D^{\boldsymbol{\alpha}} f_1 = D_1^{\alpha_1} \ldots D_l^{\alpha_l} f_1$, where $D_j f_1$ denotes the partial derivative of f_1 with respect to the jth component of the argument, $1 \leq j \leq l$. For $\boldsymbol{t} = (t_1, \ldots t_l)' \in \mathscr{R}^l$ and $\boldsymbol{\alpha}$ as above, define $t^{\boldsymbol{\alpha}} = t_1^{\alpha_1} \ldots t_l^{\alpha_l}$. Let $\boldsymbol{\Phi}_V$ denote the multivariate Normal distribution with mean $\mathbf{0}$ and dispersion matrix $\boldsymbol{V}$ having jth row $\boldsymbol{V}_{j\cdot}$ and let ϕ_V denote the density of $\boldsymbol{\Phi}_V$. We write $\boldsymbol{\Phi}_V = \boldsymbol{\Phi}$ and $\phi_V = \phi$ when $\boldsymbol{V}$ is the identity matrix. Also define the polynomial $\chi_{\boldsymbol{\alpha}}(\boldsymbol{y} : \boldsymbol{V})$ by the identity $(-D)^{\boldsymbol{\alpha}} \phi(\boldsymbol{y} : \boldsymbol{V}) = \chi_{\boldsymbol{\alpha}}(\boldsymbol{y} : \boldsymbol{V}) \phi(\boldsymbol{y} : \boldsymbol{V})$. We write $\chi_{\boldsymbol{\alpha}}(\boldsymbol{y} : \boldsymbol{V})$ as $\chi_{\boldsymbol{\alpha}}(\boldsymbol{y})$ when $\boldsymbol{V}$ is the identity matrix. For any set $B \subseteq \mathcal{R}^p$ and any $\boldsymbol{b} \in \mathcal{R}^p$, $B + \boldsymbol{b} = \{\boldsymbol{a} + \boldsymbol{b} : \boldsymbol{a} \in B\}$.

Define, $\boldsymbol{W}_n = n^{-1/2} \sum_{i=1}^n \boldsymbol{x}_i \epsilon_i$, $\dot{\boldsymbol{W}}_n = n^{-1/2} \sum_{i=1}^n \boldsymbol{C}_{11}^{-1} \boldsymbol{x}_i^{(1)} \epsilon_i$ and $\breve{\boldsymbol{W}}_n = n^{-1/2} \sum_{i=1}^n \boldsymbol{C}_{21} \boldsymbol{C}_{11}^{-1} \boldsymbol{x}_i^{(1)} \epsilon_i$. Also define $\boldsymbol{W}_n^* = n^{-1/2} \sum_{i=1}^n \boldsymbol{x}_i^* \epsilon_i^*$, $\dot{\boldsymbol{W}}_n^* = n^{-1/2} \sum_{i=1}^n \boldsymbol{C}_{11}^{-1} \boldsymbol{x}_i^{(1)} \epsilon_i^*$ and $\breve{\boldsymbol{W}}_n^* = n^{-1/2} \sum_{i=1}^n \boldsymbol{C}_{21} \boldsymbol{C}_{11}^{-1} \boldsymbol{x}_i^{(1)} \epsilon_i^*$ when underlying Bootstrap is RB. And when the PB is considered, define $\boldsymbol{W}_n^* = n^{-1/2} \sum_{i=1}^n \hat{\epsilon}_i \boldsymbol{x}_i \ (G_i^* - \mu_{G^*})$, $\dot{\boldsymbol{W}}_n^* = n^{-1/2} \sum_{i=1}^n \boldsymbol{C}_{11}^{-1} \boldsymbol{x}_i^{(1)} \epsilon_i (G_i^* - \mu_{G^*})$ and $\breve{\boldsymbol{W}}_n^* = n^{-1/2} \sum_{i=1}^n \boldsymbol{C}_{21} \boldsymbol{C}_{11}^{-1} \boldsymbol{x}_i^{(1)} \hat{\epsilon}_i (G_i^* - \mu_{G^*})$. Now define the sets $\boldsymbol{A}_n = \left\{\left\{\|\boldsymbol{W}_n\|_\infty \leq a_n\right\} \cap \left\{\|\dot{\boldsymbol{W}}_n\|_\infty \leq a_n\right\} \cap \left\{\|\breve{\boldsymbol{W}}_n\|_\infty \leq a_n\right\}\right\}$ where $a_n = K\sqrt{\log n}$ or Kn^δ according as p increases polynomially or exponentially. The constant δ is as defined in the assumptions in Section 3 in Das, Chatterjee and Lahiri (2022) and the constant K (independent of n, p) is specified by the applications of the concentration bounds stated in Lemma 1 and Lemma 2. Define the Bootstrap version of the set $\boldsymbol{A}_n$ as $\boldsymbol{A}_n^* = \left\{\left\{\|\boldsymbol{W}_n^*\|_\infty \leq a_n\right\} \cap \left\{\|\dot{\boldsymbol{W}}_n^*\|_\infty \leq a_n\right\} \cap \left\{\|\breve{\boldsymbol{W}}_n^*\|_\infty \leq a_n\right\}\right\}$ where $\breve{\boldsymbol{\beta}}^*$ is $\dot{\boldsymbol{\beta}}^*$ or $\dot{\boldsymbol{\beta}}^{**}$ according as the underlying Bootstrap is RB or PB. Arguments similar to the proofs of Lemma 8.1 of Chatterjee and Lahiri (2013) and Lemma 8.1 of Das, Gregory and Lahiri (2019) imply that $\mathbb{P}(\boldsymbol{A}_n) \geq 1 - o(n^{-1/2})$ and $\mathbb{P}^*(\boldsymbol{A}_n^*) \geq 1 - o_p(n^{-1/2})$, when p grows polynomially. When p grows exponentially, the same conclusions can be drawn based on Lemma 3 below.

2.2. Preliminary Lemmas

Lemmas necessary for the proofs of the results, are stated in this section, along with their proofs.

Lemma 1. *(Fuk-Nagaev Inequality) Suppose $Y_1, \ldots, Y_n$ are zero mean independent r.v.s with $\mathbb{E}(|Y_i|^t) < \infty$ for $i = 1, \ldots, n$ and $S_n = \sum_{i=1}^n Y_i$. Let $\sum_{i=1}^n \mathbb{E}(|Y_i|^t) = \sigma_t$, $c_t^{(1)} = \left(1 + \frac{2}{t}\right)^t$ and $c_t^{(2)} = 2(2+t)^{-1} e^{-t}$. Then, for any $t \geq 2$ and $x > 0$,*

$$P[|S_n| > x] \leq c_t^{(1)} \sigma_t x^{-t} + exp(-c_t^{(2)} x^2 / \sigma_2)$$

Proof of Lemma 1. This inequality was proved in Fuk and Nagaev (1971).

Lemma 2. *(Weibull Concentration Inequality) Suppose $Y_1, \ldots, Y_n$ are zero mean independent r.v.s with $\|Y_i\|_{g_\gamma} < \infty$ for some $\gamma > 0$, where $g_\gamma(x) = e^{x^\gamma} - 1$. Define $\boldsymbol{d}_n^Y = \left(a_1 \|Y_1\|_{g_\gamma}, \ldots, a_n \|Y_n\|_{g_\gamma}\right)'$ and $\chi(\gamma) = \infty$ when $\gamma \leq 1$ and $\gamma(\gamma-1)^{-1}$ for $\gamma > 1$. Then for any vector $(a_1, \ldots, a_n)' \in \mathcal{R}^n$ and for any $t > 0$ we have*

$$\mathbb{P}\left(\left|\sum_{i=1}^n a_i Y_i\right| \geq 2eC_Y(\gamma) \|\boldsymbol{d}_n^Y\| \sqrt{t} + 2eL_n^Y(\gamma) t^{1/\gamma} \|\boldsymbol{d}_n^Y\|_{\chi(\gamma)}\right) \leq 2e^{-t}.$$

Here $C_Y(\gamma) = \max\{\sqrt{2}, 2^{1/\gamma}\} \times \left[\sqrt{8}e^3(2\pi)^{1/4}e^{1/24}(e^{2/e}/\gamma)^{1/\gamma}\right]$ *when* $\gamma < 1$ *and* $C_Y(\gamma) = \max\{\sqrt{2}, 2^{1/\gamma}\} \times \left[4e + 2(\log 2)^{1/\gamma}\right]$ *when* $\gamma \geq 1$. *And* $L_n^Y(\gamma) = \dfrac{4^{1/\gamma}C_Y(\gamma)}{\sqrt{2}\|\boldsymbol{d}_n^Y\|}$ *if* $\gamma < 1$ & $L_n^Y(\gamma) = \dfrac{4^{1/\gamma}e}{\sqrt{2}\|\boldsymbol{d}_n^Y\|}$ *if* $\gamma \geq 1$.

Proof of Lemma 2. This inequality is giving us exponential tail bounds for a weighted sum of sub-Weibull random variables. This concentration bound is stated as Theorem 3.1 in Kuchibhotla and Chakrabortty (2022).

Lemma 3. *(Probability of the set* $\boldsymbol{A}_n$*) Suppose that the conditions of Theorem 4 are true. Then* $\mathbb{P}(\boldsymbol{A}_n) = 1 - o\big(n^{-1/2}\big)$ *and* $\mathbb{P}^*(\boldsymbol{A}_n^*) = 1 - o_p\big(n^{-1/2}\big)$*, provided* $p = o\big(n^{-1/2}\exp(\gamma_n)\big)$ *where* γ_n *is as defined before the statement of theorem 4 in Section 5.*

Proof of Lemma 3. Let us first show that $\mathbb{P}(\boldsymbol{A}_n) = 1 - o\big(n^{-1/2}\big)$. For this enough to show that for any fixed $j \in \{1, \ldots, p\}$, $\mathbb{P}\big(\big|n^{-1/2}\sum_{i=1}^n x_{ij}\epsilon_i\big| > Kn^{\delta}\big) \leq C\exp(\gamma_n)$ for some constant $C > 0$ (independent of n, p), when $\gamma \leq 2$. Similar bounds for the other parts of the set $\boldsymbol{A}_n$ can be obtained analogously. Simply apply Lemma 2 with $t = n^{\gamma\delta}$ and $a_i = n^{-1/2}x_{ij}$ to get

$$\mathbb{P}\Big(\Big|n^{-1/2}\sum_{i=1}^n x_{ij}\epsilon_i\Big| > 2eC_\epsilon(\gamma)\|\boldsymbol{d}_n^\epsilon\|n^{\gamma\delta/2} + 2eL_n^\epsilon(\gamma)n^{\delta}\|\boldsymbol{d}_n^\epsilon\|_{\mathcal{X}(\gamma)}\Big) \leq 2\exp(n^{-\gamma\delta}).$$

Since due to the assumed conditions $\left[2eC_\epsilon(\gamma)\|\boldsymbol{d}_n^\epsilon\|n^{\gamma\delta/2} + 2eL_n^\epsilon(\gamma)t^{\delta}\|\boldsymbol{d}_n^\epsilon\|_{\mathcal{X}(\gamma)}\right] \leq Kn^{\delta}$ for suitable choice of $K > 0$, hence we are done.

We prove $\mathbb{P}^*(\boldsymbol{A}_n^*) = 1 - o_p\big(n^{-1/2}\big)$ only for RB. For PB, the proof is similar to the original case upon applications of Lemma 1 and Lemma 4. Since $\{\epsilon_1^*, \ldots, \epsilon_n^*\}$ are conditionally bounded mean zero iid random variables, hence by Hoeffding's inequality we have

$$\mathbb{P}^*\Big(\Big|n^{-1/2}\sum_{i=1}^n x_{ij}\epsilon_i^*\Big| > Kn^{\delta}\Big) \leq 2\exp\Big(-\frac{2K^2n^{-(2\delta+1)}}{\sum_{i=1}^n x_{ij}^2\max_{j,k}|\hat{\epsilon}_j - \hat{\epsilon}_k|^2}\Big).$$

Now we can replace $\max_{j,k}|\hat{\epsilon}_j - \hat{\epsilon}_k|^2$ in the bound by $\max_{j,k}|\epsilon_j - \epsilon_k|^2$ due to Lemma 4 and assumption (A.2)(i). Again by the tail concentration bound in Lemma 2 we have

$$P\Big(|\epsilon_1| \geq \|\epsilon\|_{g_\gamma}(2\log n)^{1/\gamma}\Big) \leq 2\exp\Big(-2\log n\Big),$$

implying that

$$\mathbb{P}^*\Big(\Big|n^{-1/2}\sum_{i=1}^n x_{ij}\epsilon_i^*\Big| > Kn^{\delta}\Big) \leq 2\exp\Big(-\frac{K^2n^{-(2\delta+1)}}{\sum_{i=1}^n x_{ij}^2\|\epsilon\|_{g_\gamma}^2(2\log n)^{2/\gamma}}\Big).$$

Hence clearly a suitable choice of K will entail the desired bound provided $p = o\Big(n^{-1/2}\exp(\gamma_n)\Big)$.

Lemma 4. *(Concentration of Bias corrected estimators) Recall the post-LASSO OLS estimator* $\check{\boldsymbol{\beta}}_n$ *and its Bootstrap version* $\check{\boldsymbol{\beta}}_n^* = \dot{\boldsymbol{\beta}}_n^*$ *(for RB) or* $\check{\boldsymbol{\beta}}_n^* = \dot{\boldsymbol{\beta}}_n^{**}$ *(for PB). Also recall that* $a_n = K\sqrt{\log n}$ *or* n^{δ} *according as p increases polynomially or exponentially Then under the assumptions (A.1), (A.2)′, (A.5)(i) and (A.7) (or (A.7)′) we have on the set* $\boldsymbol{A}_n$

$$\sqrt{n}\|\dot{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\|_\infty \leq a_n \;\; and \;\; \mathbb{P}^*\Big(\sqrt{n}\|\check{\boldsymbol{\beta}}_n^* - \dot{\boldsymbol{\beta}}_n\|_\infty > a_n\Big) = o(n^{-1/2}).$$

Proof of Lemma 4. The LASSO estimator $\hat{\boldsymbol{\beta}}_n$ is defined as

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{t}} \left[\sum_{i=1}^{n} (y_i - \boldsymbol{x}_i'\boldsymbol{t})^2 + \lambda_n \sum_{j=1}^{p} (|t_j|) \right]$$

Now, writing $\hat{\boldsymbol{u}}_n = \sqrt{n}\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\big)$, we have

$$\begin{aligned} \hat{\boldsymbol{u}}_n &= \arg\min_{\boldsymbol{v}} \left[\boldsymbol{v}'\boldsymbol{C}_n\boldsymbol{v} - 2\boldsymbol{v}'\boldsymbol{W}_n + \lambda_n \sum_{j=1}^{p} \left(|\beta_{j,n} + \frac{v_j}{\sqrt{n}}| - |\beta_{j,n}| \right) \right] \\ &= \arg\min_{\boldsymbol{v}} \boldsymbol{Z}_n(\boldsymbol{v}) \quad \text{(say)}. \end{aligned} \tag{2.1}$$

Note that $\boldsymbol{Z}_n(\boldsymbol{v})$ is convex in $\boldsymbol{v}$. Hence, the KKT condition is necessary and sufficient. The KKT condition corresponding to (2.1) is given by

$$2\boldsymbol{C}_n\boldsymbol{v} - 2\boldsymbol{W}_n + \frac{\lambda_n}{\sqrt{n}}\boldsymbol{l}_n = \boldsymbol{0} \tag{2.2}$$

for some $l_{j,n} \in [-1,1]$ for all $j \in \{1,\dots,p\}$, where $\boldsymbol{l}_n = (l_{1,n},\dots,l_{p,n})'$. Since $\mathcal{A}_n = \{j : \beta_{n,j} \neq 0\} = \{1,\dots,p_0\}$ and $C_{11,n}$ is non-singular, Lemma 2 of Tibshirani (2013) (see also Theorem 1 of Wainwright (2009)) and noting the conditions (A.1), (A.4) imply that on the set $\boldsymbol{A}_n$, $\big((\hat{\boldsymbol{u}}_{3n}^{(1)})', \boldsymbol{0}'\big)'$, where $\hat{\boldsymbol{u}}_{3n}^{(1)} = \boldsymbol{C}_{11,n}^{-1}\big[\boldsymbol{W}_n^{(1)} - \frac{\lambda_n}{2\sqrt{n}}\boldsymbol{c}_n^{(1)}\big]$ is the unique solution of (2.2). Hence $\hat{\boldsymbol{u}}_{3n} = \big((\hat{\boldsymbol{u}}_{3n}^{(1)})', \boldsymbol{0}'\big)'$, is the unique solution of the minimization problem (2.1). Here $\boldsymbol{c}_n^{(1)} = (c_{1n},\dots,c_{p_0n})$ and $c_{j,n} = sgn(\beta_{j,n})$. Again on the set $\boldsymbol{A}_n$, $sgn(\breve{\beta}_{j,n}) = sgn(\beta_{j,n})$ for all j and hence on the set $\boldsymbol{A}_n$, the post-LASSO OLS estimator is given by

$$\sqrt{n}\big(\dot{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\big) = \boldsymbol{C}_{11,n}^{-1}\boldsymbol{W}_n^{(1)}.$$

Hence clearly $\sqrt{n}\|\dot{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\|_\infty \le a_n$ on the set $\boldsymbol{A}_n$.

Lemma 5. *($\sqrt{n}$–consistency of $\tilde{\boldsymbol{\Sigma}}_n$) The matrices $\boldsymbol{\Sigma}_n$, $\tilde{\boldsymbol{\Sigma}}_n$ are defined in Section 5. Under the conditions (A.2)(i), (A.3)(i), (A.3)(ii) & (A.5)(i) (or (A.2)(i), (A.3)(i)$''$, (A.3)(ii)$''$ & (A.5)(i)$''$ in non-IID case), we have*

$$\|\tilde{\boldsymbol{\Sigma}}_n - \boldsymbol{\Lambda}_n\| = O_p(n^{-1/2}).$$

Proof of Lemma 5. Note that

$$\tilde{\boldsymbol{\Sigma}}_n - \boldsymbol{\Lambda}_n = n^{-1}\sum_{i=1}^{n} \boldsymbol{\xi}_i^{(1)}\boldsymbol{\xi}_i^{(1)\prime}(\hat{\epsilon}_i^2 - \sigma^2) = \sum_{i=1}^{n} \boldsymbol{\xi}_i^{(1)}\boldsymbol{\xi}_i^{(1)\prime}(\epsilon_i^2 - \sigma^2) + \sum_{i=1}^{n} \boldsymbol{\xi}_i^{(1)}\boldsymbol{\xi}_i^{(1)\prime}(\hat{\epsilon}_i^2 - \epsilon_i^2),$$

where $\boldsymbol{\xi}_i^{(1)} = \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}$, $i \in \{1,\dots,n\}$. Now by Lemma 1, Lemma 4 and the assumed conditions, we have

$$\mathbb{P}\left(\left\{\left\|\sum_{i=1}^{n} \boldsymbol{\xi}_i^{(1)}\boldsymbol{\xi}_i^{(1)\prime}(\epsilon_i^2 - \sigma^2)\right\| > K.n^{1/2}\right\}\right) \to 0 \text{ as } K \to \infty$$

and

$$\mathbb{P}\Big(\Big\{\Big\|\sum_{i=1}^{n}\boldsymbol{\xi}_i^{(1)}\boldsymbol{\xi}_i^{(1)\prime}(\hat{\epsilon}_i^2-\epsilon_i^2)\Big\| > K.n^{1/2}\Big\}\Big) \to 0 \text{ as } K \to \infty$$

Therefore, Lemma 5 follows.

Lemma 6. *(Normal approximation over Convex sets) Suppose $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ are n random vectors in $\mathcal{R}^k$ satisfying $\mathbb{E}\boldsymbol{X}_j=0$, for all $j\in\{1,\ldots,n\}$, and $\boldsymbol{V}_n=n^{-1}\sum_{i=1}^{n}\mathbb{V}(\boldsymbol{X}_j)$ where γ_n is the smallest eigen value of $\boldsymbol{V}_n$. Define, $\rho_3=n^{-1}\sum_{i=1}^{n}\mathbb{E}\|\boldsymbol{X}_j\|^3$. If $\rho_3<\infty$, then we have*

$$\sup_{\boldsymbol{C}\in\mathcal{C}_q}\Big|\mathbb{P}\big(n^{-1/2}\sum_{i=1}^{n}\boldsymbol{X}_i\in\boldsymbol{C}\big)-\Phi\big(\boldsymbol{C};\sigma^2\boldsymbol{V}_n\big)\Big| \le k\gamma_n^{-3/2}\rho_3 n^{-1/2}$$

Proof of Lemma 6. This result is stated as Corollary 17.2 in Bhattacharya and Ranga Rao (1986).

Lemma 7. *(Cramer's condition for RB residuals) Suppose that all the assumptions of Theorem 3 except (A.3) hold. Then for any $\delta_2>0$ and $K\in(0,\infty)$, there exists $\delta_3\in(0,1)$ such that*

$$\sup\Big\{\hat{\omega}_n(t_1,t_2):\delta_2^2\le t_1^2+t_2^2\le n^K\Big\}=1-\delta_3+o_p(1),$$

where $\hat{\omega}_n(t_1,t_2)=\mathbb{E}_*\exp\Big(it_1\epsilon_1^*+it_2\epsilon_1^{*2}\Big)$.

Proof of Lemma 7. This follows through the same line arguments as in the proof of Lemma 2 in Babu and Singh (1984).

Lemma 8. *(Solution in Matrix Taylor's theorem) Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be positive definite matrices of same order. Then for some given matrix $\boldsymbol{C}$, the solution of the equation $\boldsymbol{AX}+\boldsymbol{XB}=\boldsymbol{C}$ can be expressed as*

$$\boldsymbol{X}=\int_0^\infty e^{-t\boldsymbol{A}}\boldsymbol{C}e^{-t\boldsymbol{B}}dt,$$

where $e^{-t\boldsymbol{A}}$ and $e^{-t\boldsymbol{B}}$ are defined in the usual way.

Proof of Lemma 8. This lemma is an easy consequence of Theorem VII.2.3 in Bhatia (1997).

2.3. Proofs of main results

In this section, we present detailed proofs of Theorems 3, 7, 8 and 9 of the main paper.

Proof of Theorem 3: Note that on the set $\boldsymbol{A}_n$, for sufficiently large n we have,

$$\boldsymbol{T}_n-\dot{\boldsymbol{b}}_n=n^{-1/2}\sum_{i=1}^{n}\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i-\frac{\lambda_n}{2\sqrt{n}}\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{c}_n^{(1)}+\hat{\boldsymbol{D}}_n^{(1)}\hat{\boldsymbol{C}}_{11,n}^{-1}\hat{\boldsymbol{c}}_n^{(1)}\frac{\lambda_n}{2\sqrt{n}}$$

$$= n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\xi}_i^{(1)}\epsilon_i + \frac{\lambda_n}{2\sqrt{n}}\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}(\hat{\boldsymbol{c}}_n^{(1)} - \boldsymbol{c}_n^{(1)})$$

$$= n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\xi}_i^{(1)}\epsilon_i + Q_{4n}, \quad (say) \tag{2.3}$$

where the jth element of $\hat{\boldsymbol{c}}_n^{(1)}$ is $sgn(\breve{\beta}_{j,n})$ with $\breve{\beta}_{j,n}$ being the jth component of the Post-LASSO OLS estimator $\breve{\boldsymbol{\beta}}_n$. Now since $\sqrt{n}\|\breve{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\|_\infty \le K(\log n)^{1/2}$ on the set $\boldsymbol{A}_n$ [cf. Lemma 4], one can conclude that on the set $\boldsymbol{A}_n$, $\mathbb{P}(\hat{\boldsymbol{c}}_n^{(1)} = \boldsymbol{c}_n^{(1)}) = 1$ for sufficiently large n. Hence we can conclude that $\mathbb{P}(\|Q_{4n}\| \neq 0) = o_p(n^{-1/2})$. Now expanding $\hat{\sigma}_n$ or $\breve{\sigma}_n$ around σ and by (2.3), one has

$$\breve{\boldsymbol{R}}_n = \left[\sigma^{-1} - 2^{-1}\sigma^{-3}\Big(n^{-1}\sum_{i=1}^{n}(\epsilon_i^2 - \sigma^2)\Big)\right]\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{W}_n^{(1)} + \boldsymbol{Q}_{5n}$$

$$= \breve{\boldsymbol{R}}_{1n} + Q_{5n}, \quad \text{(say)} \tag{2.4}$$

$$\tilde{\boldsymbol{R}}_n = \left[\sigma^{-1} - 2^{-1}\sigma^{-3}\Big(n^{-1}\sum_{i=1}^{n}(\epsilon_i^2 - \sigma^2)\Big)\right]\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{W}_n^{(1)} + Q_{6n}$$

$$= \tilde{\boldsymbol{R}}_{2n} + \boldsymbol{Q}_{6n}, \quad \text{(say)} \tag{2.5}$$

where on the set $\boldsymbol{A}_n$, $\|\boldsymbol{Q}_{5n}\| + \|\boldsymbol{Q}_{6n}\| = o(n^{-1/2})$.

Now let us concentrate on $\breve{\boldsymbol{R}}_n$. The two term EE of $\breve{\boldsymbol{R}}_n$ and $\breve{\boldsymbol{R}}_{1n}$ agree up to $o(n^{-1/2})$. Now we are going to use the transformation technique of Bhattacharya and Ghosh (1978) to find two term EE of $\breve{\boldsymbol{R}}_{1n}$. However before employing that technique we need to find two term EE of $\boldsymbol{R}_n^\dagger = \left[n^{-1/2}\sum_{i=1}^{n}\left(\boldsymbol{\xi}_i^{(1)\prime}\epsilon_i, (\epsilon_i^2 - \sigma^2)\right)'\right]$. First three cumulants of $\boldsymbol{t}'\boldsymbol{R}_n^\dagger$ are given by

$$\kappa_1\left(\boldsymbol{t}'\boldsymbol{R}_n^\dagger\right) = \mathbb{E}\left(\boldsymbol{t}'\boldsymbol{R}_n^\dagger\right) = \boldsymbol{0}$$

$$\kappa_2\left(\boldsymbol{t}'\boldsymbol{R}_n^\dagger\right) = \mathbb{V}\left(\boldsymbol{t}'\boldsymbol{R}_n^\dagger\right) = \boldsymbol{t}'\begin{bmatrix}\Big(n^{-1}\sum_{i=1}^{n}\boldsymbol{\xi}_i^{(1)}\boldsymbol{\xi}_i^{(1)\prime}\Big)\sigma^2 & \Big(n^{-1}\sum_{i=1}^{n}\boldsymbol{\xi}_i^{(1)}\Big)\mu_3 \\ \Big(n^{-1}\sum_{i=1}^{n}\boldsymbol{\xi}_i^{(1)\prime}\Big)\mu_3 & (\mu_4 - \sigma^4)\end{bmatrix}\boldsymbol{t} = \boldsymbol{t}'\boldsymbol{S}_n\boldsymbol{t}$$

$$\kappa_3\left(\boldsymbol{t}'\boldsymbol{R}_n^\dagger\right) = \mathbb{E}\left(\boldsymbol{t}'\boldsymbol{R}_n^\dagger\right)^3 - 3\mathbb{E}\left(\boldsymbol{t}'\boldsymbol{R}_n^\dagger\right)^2.\mathbb{E}\left(\boldsymbol{t}'\boldsymbol{R}_n^\dagger\right) + 2\Big(\mathbb{E}\left(\boldsymbol{t}'\boldsymbol{R}_n^\dagger\right)\Big)^3$$

$$= n^{-1/2}\mu_3\Bigg[\sum_{|\boldsymbol{\alpha}|\le 3}\boldsymbol{t}^{\boldsymbol{\alpha}}\bar{\boldsymbol{\xi}}_n^{(0)}(\boldsymbol{\alpha})\mathbb{E}\Big[\epsilon_1^{|\boldsymbol{\alpha}|}(\epsilon_1^2 - \sigma^2)^{3-|\boldsymbol{\alpha}|}\Big]\Bigg].$$

Hence by Theorem 20.6 of Bhattacharya and Ranga Rao (1986) we have

$$\sup_{\boldsymbol{B}\in\mathcal{C}_{q+1}}\left|\mathbb{P}(\boldsymbol{R}_n^\dagger \in \boldsymbol{B}) - \int_{\boldsymbol{B}}\boldsymbol{\psi}_n^{(0)}(\boldsymbol{x})d\boldsymbol{x}\right| = o(n^{-1/2}) \tag{2.6}$$

where

$$\boldsymbol{\psi}_n^{(0)}(\boldsymbol{x}) = \phi(\boldsymbol{x}:\boldsymbol{\Sigma}_n)\left[1 + \frac{\mu_3}{\sqrt{n}}\sum_{|\boldsymbol{\alpha}|\le 3}\bar{\boldsymbol{\xi}}_n^{(0)}(\boldsymbol{\alpha})\mathbb{E}\Big[\epsilon_1^{|\boldsymbol{\alpha}|}(\epsilon_1^2 - \sigma^2)^{3-|\boldsymbol{\alpha}|}\Big]\chi_{\boldsymbol{\alpha}}(\boldsymbol{x}:\boldsymbol{S}_n)\right],$$

provided there exists $\delta_4 \in (0,1)$, independent of n, such that for all $\upsilon \le \delta_4$,

$$n^{-1}\sum_{i=1}^{n}\mathbb{E}\left\|(\boldsymbol{\xi}_i^{(1)\prime}\epsilon_i,(\epsilon_i^2-\sigma^2))'\right\|^3\mathbf{1}\left(\left\|(\boldsymbol{\xi}_i^{(1)\prime}\epsilon_i,(\epsilon_i^2-\sigma^2))'\right\|>\upsilon\sqrt{n}\right)=o(1) \tag{2.7}$$

and

$$\max_{|\alpha|\le q+2}\int_{\|\boldsymbol{t}\|\ge\upsilon\sqrt{n}}\left|D^{\alpha}\mathbb{E}\exp(i\boldsymbol{t}'\boldsymbol{R}_{1n}^{\dagger})\right|d\boldsymbol{t}=o\left(n^{-1/2}\right) \tag{2.8}$$

where $\boldsymbol{R}_{1n}^{\dagger}=n^{-1/2}\sum_{i=1}^{n}\left(\boldsymbol{Z}_i-\mathbb{E}\boldsymbol{Z}_i\right)$ with

$$\boldsymbol{Z}_i=(\boldsymbol{\xi}_i^{(1)\prime}\epsilon_i,(\epsilon_i^2-\sigma^2))'\mathbf{1}\left(\left\|(\boldsymbol{\xi}_i^{(1)\prime}\epsilon_i,(\epsilon_i^2-\sigma^2))'\right\|\le\upsilon\sqrt{n}\right).$$

First consider (2.7). Note that due to the condition (A.3)(ii) with $r=4$,

$$\max\left\{\|\boldsymbol{\xi}_i^{(1)}\|:i\in\{1,\ldots,n\}\right\}=O\left(n^{1/8}\right).$$

Therefore, due to (A.5)(i) with $r=4$, we have for any $\upsilon>0$,

$$\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\mathbb{E}\left\|(\boldsymbol{\xi}_i^{(1)\prime}\epsilon_i,(\epsilon_i^2-\sigma^2))'\right\|^3\mathbf{1}\left(\left\|(\boldsymbol{\xi}_i^{(1)\prime}\epsilon_i,(\epsilon_i^2-\sigma^2))'\right\|>\upsilon\sqrt{n}\right)\\
\le&n^{-1}\sum_{i=1}^{n}\mathbb{E}\left(\|\boldsymbol{\xi}_i^{(1)}\|^2\epsilon_1^2+(\epsilon_1^2-\sigma^2)^2\right)^{3/2}\mathbf{1}\left(\|\boldsymbol{\xi}_i^{(1)}\|^2\epsilon_1^2+(\epsilon_1^2-\sigma^2)^2>\upsilon^2 n\right)\\
\le&n^{-1}\sum_{i=1}^{n}\left(1+\|\boldsymbol{\xi}_i^{(1)}\|^2\right)^2\mathbb{E}\left[\left(\epsilon_1^2+(\epsilon_i^2-\sigma^2)^2\right)^{3/2}\mathbf{1}\left(\epsilon_1^2+(\epsilon_1^2-\sigma^2)^2>k\upsilon^2 n^{3/4}\right)\right]\\
=&o(1).
\end{aligned}$$

Now consider (2.8). Note that for any $|\boldsymbol{\alpha}|\le q+2$, $|D^{\alpha}\mathbb{E}\exp(i\boldsymbol{t}'\boldsymbol{R}_{1n}^{\dagger})|$ is bounded above by a sum of $n^{|\alpha|}$-terms, each of which is bounded above by

$$C(\alpha)\cdot n^{-|\alpha|/2}\max\{\mathbb{E}\|\boldsymbol{Z}_k-\mathbb{E}\boldsymbol{Z}_k\|^{|\alpha|}:k\in\boldsymbol{I}_n\}\cdot\prod_{k\in\boldsymbol{I}_n^{\mathbf{c}}}|\mathbb{E}\exp(i\boldsymbol{t}'\boldsymbol{Z}_k/\sqrt{n})| \tag{2.9}$$

where $\boldsymbol{I}_n\subset\{1,\ldots,n\}$ is of size $|\boldsymbol{\alpha}|$ and $\boldsymbol{I}_{\boldsymbol{n}}^{\mathbf{c}}=\{1,\ldots,n\}\backslash\boldsymbol{I}_n$ and $C(\boldsymbol{\alpha})$ is a constant which depends only on $\boldsymbol{\alpha}$. Now for any $\omega>0$ and $\boldsymbol{t}\in\mathcal{R}^q$, define the set

$$\boldsymbol{B}_n(\boldsymbol{t},\omega)=\left\{k:1\le k\le n\text{ and }\left(\boldsymbol{t}'\boldsymbol{\xi}_k^{(0)}\right)^2>\omega^2\right\}.$$

Hence for any $\boldsymbol{t}\in\mathcal{R}^{q+1}$ writing $\boldsymbol{t}=\left(\boldsymbol{t}_q',t_{q+1}\right)$, we have

$$\sup\left\{\prod_{k\in\boldsymbol{I}_n^{\mathbf{c}}}|\mathbb{E}\exp(i\boldsymbol{t}'\boldsymbol{Z}_k/\sqrt{n})|:\|\boldsymbol{t}\|\ge\upsilon\sqrt{n}\right\}$$

$$= \sup\Big\{ \prod_{k\in \boldsymbol{I}_n^c} |\mathbb{E}\exp(i\boldsymbol{t}'\boldsymbol{Z}_k)| : \|\boldsymbol{t}\|^2 \geq \upsilon^2 \Big\}$$

$$\leq \max\Bigg\{ \sup\Bigg\{ \prod_{k\in \boldsymbol{I}_n^c\cap \boldsymbol{B}_n\left(\frac{\boldsymbol{t}_q}{\|\boldsymbol{t}_q\|},\upsilon/\sqrt{2}\right)} \Big[|\mathbb{E}\exp\Big(i\boldsymbol{t}_q'\boldsymbol{\xi}_k^{(0)}\epsilon_k\Big)| + \mathbb{P}\Big(\epsilon_1^2 + (\epsilon_1^2-\sigma^2)^2 > k\upsilon^2 n^{3/4}\Big)\Big]$$

$$: \|\boldsymbol{t}_q\| \geq \upsilon/\sqrt{2}\Bigg\}, \sup\Bigg\{ \prod_{k\in \boldsymbol{I}_n^c} \Big[|\mathbb{E}\exp\Big(it_{q+1}(\epsilon_k^2-\sigma^2)\Big)| + \mathbb{P}\Big(\epsilon_1^2 + (\epsilon_1^2-\sigma^2)^2 > k\upsilon^2 n^{3/4}\Big)\Big]$$

$$: |t_{q+1}| \geq \upsilon/\sqrt{2}\Bigg\}\Bigg\}$$

Now since $\big|\boldsymbol{I}_n^c\big| \geq \big|\boldsymbol{I}_n^c \cap \boldsymbol{B}_n(\frac{\boldsymbol{t}_q}{\|\boldsymbol{t}_q\|},\upsilon/\sqrt{2})\big| \geq \big|\boldsymbol{B}_n(\frac{\boldsymbol{t}_q}{\|\boldsymbol{t}_q\|},\upsilon/\sqrt{2})\big| - |\alpha|$, due to Cramer's condition (A.5)(ii), for some $\theta \in (0,1)$, we have

$$\sup\Big\{ \prod_{k\in \boldsymbol{I}_n^c} |\mathbb{E}\exp(i\boldsymbol{t}'\boldsymbol{Z}_k/\sqrt{n})| : \|\boldsymbol{t}\| \geq \upsilon\sqrt{n}\Big\} \leq \theta^{\left|\boldsymbol{B}_n\left(\frac{\boldsymbol{t}_q}{\|\boldsymbol{t}_q\|},\upsilon/\sqrt{2}\right)\right| - |\alpha|} \tag{2.10}$$

Next note that for any $\boldsymbol{u} \in \mathcal{R}^q$ with $\|u\| = 1$, due to conditions (A.3)(ii) & (A.3)(iii), for sufficiently large n we have

$$\begin{aligned}
\frac{n\delta}{2} &\leq \sum_{i=1}^{n} \Big|\boldsymbol{u}'\boldsymbol{\xi}_i^{(1)}\Big|^2 \\
&\leq \max\Big\{\|\boldsymbol{\xi}_i^{(1)}\|^2 : 1 \leq i \leq n\Big\} \cdot |\boldsymbol{B}_n(\boldsymbol{u},\omega)| + \Big(n - |\boldsymbol{B}_n(\boldsymbol{u},\omega)|\Big)\cdot\omega^2 \\
&\leq n^{1/4}\cdot|\boldsymbol{B}_n(\boldsymbol{u},\omega)| + \Big(n - |\boldsymbol{B}_n(\boldsymbol{u},\omega)|\Big)\cdot\omega^2 \\
&\leq k\cdot n^{1/4}\cdot|\boldsymbol{B}_n(\boldsymbol{u},\omega)| + n\omega^2,
\end{aligned}$$

which implies $|\boldsymbol{B}_n(\boldsymbol{u},\omega)| \geq k_1\cdot n^{3/4}$ whenever $\omega < \sqrt{\delta/2}$. Therefore taking $\delta_4 = \sqrt{\delta/3}$, (2.8) follows from (2.9) and (2.10). Now we are ready to apply the transformation technique of Bhattacharya and Ghosh (1978) to find two term EE of $\check{\boldsymbol{R}}_{1n}$ and hence of $\check{\boldsymbol{R}}_n$. Now the first three cumulants of $\boldsymbol{t}'\check{\boldsymbol{R}}_{1n}$ are given by

$$\kappa_1\big(\boldsymbol{t}'\check{\boldsymbol{R}}_{1n}\big) = \mathbb{E}\big(\boldsymbol{t}'\check{\boldsymbol{R}}_{1n}\big) = -n^{-1/2}\frac{\mu_3}{2\sigma^3}\sum_{|\boldsymbol{\alpha}|=1} \boldsymbol{t}^{\boldsymbol{\alpha}}\bar{\boldsymbol{\xi}}_n^{(0)}(\boldsymbol{\alpha}) + o(n^{-1/2})$$

$$\kappa_2\big(\boldsymbol{t}'\check{\boldsymbol{R}}_{1n}\big) = \mathbb{V}\big(\boldsymbol{t}'\check{\boldsymbol{R}}_{1n}\big) = \boldsymbol{t}'\Big(n^{-1}\sum_{i=1}^{n}\boldsymbol{\xi}_i^{(1)}\boldsymbol{\xi}_i^{(1)\prime}\Big)\boldsymbol{t} = \boldsymbol{t}'\boldsymbol{\Sigma}_n\boldsymbol{t}$$

$$\begin{aligned}
\kappa_3\big(\boldsymbol{t}'\check{\boldsymbol{R}}_{1n}\big) &= \mathbb{E}\big(\boldsymbol{t}'\check{\boldsymbol{R}}_{1n}\big)^3 - 3\mathbb{E}\big(\boldsymbol{t}'\check{\boldsymbol{R}}_{1n}\big)^2.\mathbb{E}\big(\boldsymbol{t}'\check{\boldsymbol{R}}_{1n}\big) + 2\Big(\mathbb{E}\big(\boldsymbol{t}'\check{\boldsymbol{R}}_{1n}\big)\Big)^3 \\
&= n^{-1/2}\frac{\mu_3}{\sigma^3}\Big[\sum_{|\boldsymbol{\alpha}|=3}\boldsymbol{t}^{\boldsymbol{\alpha}}\bar{\boldsymbol{\xi}}_n^{(0)}(\boldsymbol{\alpha}) - 3\sum_{|\boldsymbol{\alpha}|=1}\sum_{|\boldsymbol{\gamma}|=2}\boldsymbol{t}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}}\bar{\boldsymbol{\xi}}_n^{(0)}(\boldsymbol{\alpha})\bar{\boldsymbol{\xi}}_n^{(0)}(\boldsymbol{\gamma})\Big]
\end{aligned}$$

where $\bar{\boldsymbol{\xi}}_n^{(0)}(\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^n (\boldsymbol{\xi}_i^{(1)})^{\boldsymbol{\alpha}}$. Therefore the Lebesgue density of two term EE of $\breve{\boldsymbol{R}}_{1n}$ is given by

$$\begin{aligned}\breve{\boldsymbol{\psi}}_{1n}(\boldsymbol{x}) =&\phi(\boldsymbol{x}:\boldsymbol{\Sigma}_n)\Bigg[1+\frac{1}{\sqrt{n}}\Bigg[-\frac{\mu_3}{2\sigma^3}\sum_{|\boldsymbol{\alpha}|=1}\bar{\boldsymbol{\xi}}_n(\boldsymbol{\alpha})\chi_{\boldsymbol{\alpha}}(\boldsymbol{x}:\boldsymbol{\Sigma}_n)\\ &+\frac{\mu_3}{6\sigma^3}\Big\{\sum_{|\boldsymbol{\alpha}|=3}\bar{\boldsymbol{\xi}}_n(\boldsymbol{\alpha})\chi_{\boldsymbol{\alpha}}(\boldsymbol{x}:\boldsymbol{\Sigma}_n)-3\sum_{|\boldsymbol{\alpha}|=3}\sum_{|\boldsymbol{\zeta}|=1}\bar{\boldsymbol{\xi}}_n(\boldsymbol{\alpha})\bar{\boldsymbol{\xi}}_n(\boldsymbol{\zeta})\chi_{\boldsymbol{\alpha}+\boldsymbol{\zeta}}(\boldsymbol{x}:\boldsymbol{\Sigma}_n)\Big\}\Bigg]\Bigg].\end{aligned}$$

Hence due to (2.4) and since $\mathbb{P}(\boldsymbol{A}) = 1 - o(n^{-1/2})$, we have

$$\sup_{\boldsymbol{B}\in\mathcal{C}_q}|\mathbb{P}(\breve{\boldsymbol{R}}_n\in\boldsymbol{B})-\int_{\boldsymbol{B}}\breve{\boldsymbol{\psi}}_{1n}(\boldsymbol{x})d\boldsymbol{x}| = o(n^{-1/2}) \tag{2.11}$$

Through the same line of arguments, it can be shown that

$$\sup_{\boldsymbol{B}\in\mathcal{C}_q}|\mathbb{P}(\tilde{\boldsymbol{R}}_n\in\boldsymbol{B})-\int_{\boldsymbol{B}}\tilde{\boldsymbol{\psi}}_{1n}(\boldsymbol{x})d\boldsymbol{x}| = o(n^{-1/2}) \tag{2.12}$$

where

$$\begin{aligned}\tilde{\boldsymbol{\psi}}_{1n}(\boldsymbol{x}) =&\phi(\boldsymbol{x})\Bigg[1+\frac{1}{\sqrt{n}}\Bigg[-\frac{\mu_3}{2\sigma^3}\sum_{|\boldsymbol{\alpha}|=1}\bar{\boldsymbol{\xi}}_n^{\dagger(0)}(\boldsymbol{\alpha})\chi_{\boldsymbol{\alpha}}(\boldsymbol{x})\\ &+\frac{\mu_3}{6\sigma^3}\Big\{\sum_{|\boldsymbol{\alpha}|=3}\bar{\boldsymbol{\xi}}_n^{\dagger(0)}(\boldsymbol{\alpha})\chi_{\boldsymbol{\alpha}}(\boldsymbol{x})-3\sum_{|\boldsymbol{\alpha}|=3}\sum_{|\boldsymbol{\zeta}|=1}\bar{\boldsymbol{\xi}}_n^{\dagger(0)}(\boldsymbol{\alpha})\bar{\boldsymbol{\xi}}_n^{\dagger(0)}(\boldsymbol{\zeta})\chi_{\boldsymbol{\alpha}+\boldsymbol{\zeta}}(\boldsymbol{x})\Big\}\Bigg]\Bigg],\end{aligned}$$

with $\bar{\boldsymbol{\xi}}_n^{\dagger(0)}(\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^n \big(\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\xi}_i^{(1)}\big)^{\boldsymbol{\alpha}}$.

Now let us look into $\breve{\boldsymbol{R}}_n^*$, the RB versions of $\breve{\boldsymbol{R}}_n$ and $\tilde{\boldsymbol{R}}_n^*$, the PB versions of $\tilde{\boldsymbol{R}}_n$. Note that

$$\begin{aligned}\breve{\boldsymbol{R}}_n^* =&n^{-1/2}\sum_{i=1}^n\boldsymbol{\xi}_i^{(1)}\epsilon_i^*/\hat{\sigma}_n - 2^{-1}\hat{\sigma}_n^{-3}\Big[n^{-1}\sum_{i=1}^n(\epsilon_i^{*2}-\hat{\sigma}_n^2)\Big]n^{-1/2}\sum_{i=1}^n\boldsymbol{\xi}_i^{(1)}\epsilon_i^*+\boldsymbol{Q}_{2n}^*\\ =&\breve{\boldsymbol{R}}_{1n}^*+\boldsymbol{Q}_{2n}^* \end{aligned}\tag{2.13}$$

$$\begin{aligned}\tilde{\boldsymbol{R}}_n^* =&n^{-1/2}\sum_{i=1}^n\tilde{\boldsymbol{\Sigma}}_n^{-1/2}\boldsymbol{\xi}_i^{(1)}\hat{\epsilon}_i(G_i^*-\mu_{G^*})\mu_{G^*}^{-1}\\ &-2^{-1}\breve{\sigma}_n^{-2}\Big[\mu_{G^*}^{-2}n^{-1}\sum_{i=1}^n\hat{\epsilon}_i^2\big[(G_i^*-\mu_{G^*})^2-\sigma_{G^*}^2\big]\Big]n^{-1/2}\sum_{i=1}^n\tilde{\boldsymbol{\Sigma}}_n^{-1/2}\boldsymbol{\xi}_i^{(1)}\hat{\epsilon}_i(G_i^*-\mu_{G^*})\mu_{G^*}^{-1}+\boldsymbol{Q}_{3n}^*\\ =&\tilde{\boldsymbol{R}}_{1n}^*+\boldsymbol{Q}_{3n}^*\end{aligned}\tag{2.14}$$

where due to the underlying conditions and Lemma 4, we have $\mathbb{P}^*\big(\|\boldsymbol{Q}_{2n}^*\| = o(n^{-1/2})\big) = 1 - o_p(n^{-1/2})$ and $\mathbb{P}^*\big(\|\boldsymbol{Q}_{3n}^*\| = o(n^{-1/2})\big) = 1 - o_p(n^{-1/2})$. Therefore it is enough to find the two term EE of $\breve{\boldsymbol{R}}_{1n}^*$ and $\tilde{\boldsymbol{R}}_{1n}^*$. Due to the Cramer's condition (A.6)(ii) on $\big((G_i^*-\mu_{G^*}),(G_i^*-\mu_{G^*})^2\big)$, one can find the EE of $\breve{\boldsymbol{R}}_{1n}^*$ through the same line of arguments as in the original case, i.e. first finding two term EE of $\big(n^{-1/2}\sum_{i=1}^n\tilde{\boldsymbol{\Sigma}}_n^{-1/2}\boldsymbol{\xi}_i^{(1)\prime}\hat{\epsilon}_i(G_i^*-\mu_{G^*}), n^{-1/2}\sum_{i=1}^n\hat{\epsilon}_i^2\big[(G_i^*-\mu_{G^*})^2-\sigma_{G^*}^2\big]\big)'$ by applying Theorem 20.6

of Bhattacharya and Ranga Rao (1986) and then using the transformation technique of Bhattacharya and Ghosh (1978). However in case of RB, one can not use Theorem 20.6 of Bhattacharya and Ranga Rao (1986) directly to obtain two term EE of $\big(n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\xi}_i^{(1)\prime}\epsilon_i^*, n^{-1/2}\sum_{i=1}^{n}(\epsilon_i^{*2}-\hat{\sigma}_n^2)\big)$, since only the conditional Cramer's condition on $(\epsilon_1^*, \epsilon_1^{*2})$, viz Lemma 7, holds under (A.5)(ii). Here one needs to use a smoothing kernel which vanishes outside a compact set to take advantage of the conditional Cramer's condition, instead of using the smoothing kernel used in Theorem 20.6 of Bhattacharya and Ranga Rao (1986). The arguments of Theorem 20.8 of Bhattacharya and Ranga Rao (1986) or Theorem 2 of Babu and Singh (1984) can be followed to obtain the two term EE of $\big(n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\xi}_i^{(1)\prime}\epsilon_i^*, n^{-1/2}\sum_{i=1}^{n}(\epsilon_i^{*2} - \hat{\sigma}_n^2)\big)$and then, similar to the original case, one can use the transformation technique of Bhattacharya and Ghosh (1978) to come up with the two term EE of $\boldsymbol{R}_n^*$.

Writing $\hat{\mu}_3 = n^{-1}\sum_{i=1}^{n}(\hat{\epsilon}_i - \bar{\epsilon}_n)^3$, the first three conditional cumulants of $\boldsymbol{t}'\breve{\boldsymbol{R}}_{1n}^*$ given $\{\hat{\epsilon}_1,\dots,\hat{\epsilon}_n\}$ are given by

$$\kappa_1\big(\boldsymbol{t}'\breve{\boldsymbol{R}}_{1n}^*\big) = \mathbb{E}\big(\boldsymbol{t}'\breve{\boldsymbol{R}}_{1n}^*\big) = -n^{-1/2}\frac{\hat{\mu}_3}{2\hat{\sigma}_n^3}\sum_{|\boldsymbol{\alpha}|=1}\boldsymbol{t}^{\boldsymbol{\alpha}}\bar{\boldsymbol{\xi}}_n^{(0)}(\boldsymbol{\alpha}) + o(n^{-1/2})$$

$$\kappa_2\big(\boldsymbol{t}'\breve{\boldsymbol{R}}_{1n}^*\big) = \mathbb{V}\big(\boldsymbol{t}'\breve{\boldsymbol{R}}_{1n}^*\big) = \boldsymbol{t}'\Big(n^{-1}\sum_{i=1}^{n}\boldsymbol{\xi}_i^{(1)}\boldsymbol{\xi}_i^{(1)\prime}\Big)\boldsymbol{t} = \boldsymbol{t}'\boldsymbol{\Sigma}_n\boldsymbol{t}$$

$$\begin{aligned}\kappa_3\big(\boldsymbol{t}'\breve{\boldsymbol{R}}_{1n}^*\big) &= \mathbb{E}\big(\boldsymbol{t}'\breve{\boldsymbol{R}}_{1n}^*\big)^3 - 3\mathbb{E}\big(\boldsymbol{t}'\breve{\boldsymbol{R}}_{1n}^*\big)^2.\mathbb{E}\big(\boldsymbol{t}'\breve{\boldsymbol{R}}_{1n}^*\big) + 2\Big(\mathbb{E}\big(\boldsymbol{t}'\breve{\boldsymbol{R}}_{1n}^*\big)\Big)^3\\ &= n^{-1/2}\frac{\hat{\mu}_3}{\hat{\sigma}_n^3}\Big[\sum_{|\boldsymbol{\alpha}|=3}\boldsymbol{t}^{\boldsymbol{\alpha}}\bar{\boldsymbol{\xi}}_n^{(0)}(\boldsymbol{\alpha}) - 3\sum_{|\boldsymbol{\alpha}|=1}\sum_{|\boldsymbol{\gamma}|=2}\boldsymbol{t}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}}\bar{\boldsymbol{\xi}}_n^{(0)}(\boldsymbol{\alpha})\bar{\boldsymbol{\xi}}_n^{(0)}(\boldsymbol{\gamma})\Big]\end{aligned}$$

Therefore, the Lebesgue density of two term conditional EE of $\breve{\boldsymbol{R}}_{1n}^*$ is given by

$$\begin{aligned}\breve{\boldsymbol{\psi}}_{1n}^*(\boldsymbol{x}) =&\phi(\boldsymbol{x}:\boldsymbol{\Sigma}_n)\Bigg[1 + \frac{1}{\sqrt{n}}\Bigg[-\frac{\hat{\mu}_3}{2\hat{\sigma}_n^3}\sum_{|\boldsymbol{\alpha}|=1}\bar{\boldsymbol{\xi}}_n(\boldsymbol{\alpha})\chi_{\boldsymbol{\alpha}}(\boldsymbol{x}:\boldsymbol{\Sigma}_n)\\ &+\frac{\hat{\mu}_3}{6\hat{\sigma}_n^3}\Big\{\sum_{|\boldsymbol{\alpha}|=3}\bar{\boldsymbol{\xi}}_n(\boldsymbol{\alpha})\chi_{\boldsymbol{\alpha}}(\boldsymbol{x}:\boldsymbol{\Sigma}_n) - 3\sum_{|\boldsymbol{\alpha}|=3}\sum_{|\boldsymbol{\zeta}|=1}\bar{\boldsymbol{\xi}}_n(\boldsymbol{\alpha})\bar{\boldsymbol{\xi}}_n(\boldsymbol{\zeta})\chi_{\boldsymbol{\alpha}+\boldsymbol{\zeta}}(\boldsymbol{x}:\boldsymbol{\Sigma}_n)\Big\}\Bigg]\Bigg].\end{aligned}$$

Again the first three conditional cumulants of $\boldsymbol{t}'\tilde{\boldsymbol{R}}_{1n}^*$ given $\{\hat{\epsilon}_1,\dots,\hat{\epsilon}_n\}$ are given by

$\kappa_1\big(\boldsymbol{t}'\tilde{\boldsymbol{R}}_{1n}^*\big) = -\frac{1}{\sqrt{n}}.\frac{1}{2\breve{\sigma}_n^2}\sum_{|\boldsymbol{\alpha}|=1}\boldsymbol{t}^{\boldsymbol{\alpha}}\bar{\boldsymbol{\xi}}_n^{*(3)}(\boldsymbol{\alpha}) + o_p(n^{1/2})$

$\kappa_2\big(\boldsymbol{t}'\tilde{\boldsymbol{R}}_{1n}^*\big) = \mathbf{Var}_*\big(\boldsymbol{t}'\tilde{\boldsymbol{R}}_{1n}^*\big) = \boldsymbol{t}'\boldsymbol{t} + o_p(n^{-1/2})$

$\kappa_3\big(\boldsymbol{t}'\tilde{\boldsymbol{R}}_{1n}^*\big) = \mathbb{E}_*\big(\boldsymbol{t}'\tilde{\boldsymbol{R}}_{1n}^*\big)^3 - 3\mathbb{E}_*\big(\boldsymbol{t}'\tilde{\boldsymbol{R}}_{1n}^*\big)^2.\mathbb{E}_*\big(\boldsymbol{t}'\tilde{\boldsymbol{R}}_{1n}^*\big) + 2\Big(\mathbb{E}_*\big(\boldsymbol{t}'\tilde{\boldsymbol{R}}_{1n}^*\big)\Big)^3$

$= \frac{1}{\sqrt{n}}\Big[\sum_{|\boldsymbol{\alpha}|=3}\boldsymbol{t}^{\boldsymbol{\alpha}}\bar{\boldsymbol{\xi}}_n^{*(1)}(\boldsymbol{\alpha}) - \frac{3}{\hat{\sigma}_n^2}\sum_{|\boldsymbol{\alpha}|=1}\sum_{|\boldsymbol{\zeta}|=2}\boldsymbol{t}^{\boldsymbol{\alpha}+\boldsymbol{\zeta}}\bar{\boldsymbol{\xi}}_n^{*(3)}(\boldsymbol{\alpha})\bar{\boldsymbol{\xi}}_n^{*(1)}(\boldsymbol{\zeta})\Big] + o_p(n^{-1/2}),$

where $\bar{\boldsymbol{\xi}}_n^{*(j)}(\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^{n}\big(\tilde{\boldsymbol{\Sigma}}_n^{-1/2}\boldsymbol{\xi}_i^{(1)}\hat{\epsilon}_i^{j}\big)^{\boldsymbol{\alpha}}$, $j = 0,1,2,3$.

Therefore the Lebesgue density of two term conditional EE of $\tilde{\boldsymbol{R}}_{1n}^*$ is given by

$$\tilde{\boldsymbol{\psi}}_{1n}^*(\boldsymbol{x}) =\phi(\boldsymbol{x})\Bigg[1 + \frac{1}{\sqrt{n}}\Bigg[\frac{1}{6}\sum_{|\boldsymbol{\alpha}|=3}\bar{\boldsymbol{\xi}}_n^{*(1)}(\boldsymbol{\alpha})\chi_{\boldsymbol{\alpha}}(\boldsymbol{x})$$

$$-\frac{1}{2\hat{\sigma}_n^2}\Big\{\sum_{|\alpha|=1}\bar{\boldsymbol{\xi}}_n^{*(3)}(\boldsymbol{\alpha})\chi_\alpha(\boldsymbol{x})+\sum_{|\alpha|=1}\sum_{|\zeta|=2}\bar{\boldsymbol{\xi}}_n^{*(3)}(\boldsymbol{\alpha})\bar{\boldsymbol{\xi}}_n^{*(1)}(\boldsymbol{\zeta})\chi_{\alpha+\zeta}(\boldsymbol{x})\Big\}\Big]\Big].$$

Hence due to (2.13), (2.14) and since $\mathbb{P}^*(\boldsymbol{A}_n^*)=1-o_p(n^{-1/2})$, we have

$$\sup_{\boldsymbol{B}\in\mathcal{C}_q}|\mathbb{P}^*(\check{\boldsymbol{R}}_n^*\in\boldsymbol{B})-\int_{\boldsymbol{B}}\check{\boldsymbol{\psi}}_{1n}^*(\boldsymbol{x})d\boldsymbol{x}|=o_p(n^{-1/2})\quad and \tag{2.15}$$

$$\sup_{\boldsymbol{B}\in\mathcal{C}_q}|\mathbb{P}^*(\tilde{\boldsymbol{R}}_n^*\in\boldsymbol{B})-\int_{\boldsymbol{B}}\tilde{\boldsymbol{\psi}}_{1n}^*(\boldsymbol{x})d\boldsymbol{x}|=o_p(n^{-1/2}) \tag{2.16}$$

Now note that $|\hat{\mu}_3-\mu_3|=o_p(1)$, $|\hat{\sigma}_n^2-\sigma^2|=o_p(1)$ $\|\tilde{\Sigma}_n-\sigma^2\Sigma_n\|=o_p(1)$. Theorem X.3.8 of Bhatia (1997) and $\|\tilde{\Sigma}_n-\sigma^2\Sigma_n\|=o_p(1)$ imply that $\|\tilde{\Sigma}_n^{1/2}-\sigma\Sigma_n^{1/2}\|=o_p(1)$. This in turn implies that $\|\tilde{\Sigma}_n^{-1/2}-\sigma^{-1}\Sigma_n^{-1/2}\|=o_p(1)$ after applying

$$A^{-1/2}=B^{-1/2}+B^{-1/2}(B^{1/2}-A^{1/2})B^{-1/2}+B^{-1/2}(B^{1/2}-A^{1/2})B^{-1/2}(B^{1/2}-A^{1/2})A^{-1/2}.$$

Hence comparing (2.11) & (2.15) and comparing (2.12) & (2.16), Theorem 3 follows. $\square$

Proof of Theorem 7. We will only describe the proof when p grows polynomially, since the only difference between polynomial and exponential growth cases is the application of Lemma 1 or Lemma 2 (or Weibull concentration inequality), presented in Das, Chatterjee and Lahiri (2024), to conclude the desired rates $\mathbb{P}(\boldsymbol{A}_n)=1-o(n^{-2})$ and $\mathbb{P}^*(\boldsymbol{A}_n^*)=1-o_p(n^{-2})$. The pivotal pair is $(\tilde{\boldsymbol{R}}_n,\tilde{\boldsymbol{R}}_n^*)$ in terms of coverage accuracy of the symmetric Bootstrap confidence interval. Similar to the RB, we have

$$\begin{aligned}
&\Big|\mathbb{P}\big(\theta_n\in I^p_{n,(1-\alpha)}\big)-(1-\alpha)\Big|\\
=&\Big|\mathbb{P}\big(|\tilde{\boldsymbol{R}}_n|\le\hat{h}^p_{n,\alpha}\big)-\mathbb{P}\big(|\tilde{\boldsymbol{R}}_n|\le h^p_{n,\alpha}\big)\Big|\\
\le&\sup_x\Big|\Big[\mathbb{P}\Big(\tilde{\boldsymbol{R}}_n+n^{-3/2}V^p_{n,\alpha}\le x\Big)-\mathbb{P}\Big(\tilde{\boldsymbol{R}}_n-n^{-3/2}V^p_{n,\alpha}\le -x\Big)\Big]-\mathbb{P}\Big(|\tilde{\boldsymbol{R}}_n|\le x\Big)\Big|+O(n^{-2}),
\end{aligned}$$

where $V^p_{n,\alpha}=\sqrt{n}\{\hat{q}_2^p(z_\alpha)-q_2^p(z_\alpha)\}$. $q_2^p(x)\phi(x)$ is the coefficient of n^{-1} in the Edgeworth expansion of $\mathbb{P}(\tilde{\boldsymbol{R}}_n\le x)$. If $\kappa^p_{j,n}$ is the jth cumulant of $\tilde{\boldsymbol{R}}_n$, then from the theory of smooth function models we know

$$\kappa^p_{j,n}=O\big(n^{-(j-2)/2}\big),$$

see (1.21) in Bhattacharya and Ghosh (1978). Therefore $n^{-1}q_2^p(x)\phi(x)$ involves cumulants of H_n^p upto order 4. Similarly, $n^{-1}\hat{q}_2^p(x)\phi(x)$ involves cumulants of H_n^{p*} upto order 4 where $\hat{q}_2^p(x)\phi(x)$ is the coefficient of n^{-1} in the Edgeworth expansion of $\tilde{\boldsymbol{R}}_n^*$. Since $\big[\mu_{G^*}^{-4}\mathbb{E}(G_1^*-\mu_{G^*})^4\big]>1$, $\hat{q}_2^p(\cdot)$ is not in general asymptotically unbiased for $q_2^p(\cdot)$ and hence $V^p_{n,\alpha}$ is not properly centered. See also the discussion on this issue just before the statement of Theorem 7 in Section 6 of the main paper. Suppose $\tilde{q}_2^p(\cdot)$ is asymptotically unbiased for $q_2^p(\cdot)$. Then $\tilde{V}^p_{n,\alpha}=\sqrt{n}\{\tilde{q}_2^p(z_\alpha)-q_2^p(z_\alpha)\}$ is properly centered and aim should be to replace $V^p_{n,\alpha}$ by $\tilde{V}^p_{n,\alpha}$ in the calculations of coverage error. That is why $\tilde{I}^p_{n,(1-\alpha)}$ is the correct symmetric PB interval, not $I^p_{n,(1-\alpha)}$. Upon correction, we can show that the coverage error is $O(n^{-2})$ through the same arguments as in case of RB. Hence it is enough to find the correction term $C_n^p(z_\alpha) = n^{-1}\{\hat{q}_2^p(z_\alpha) - \tilde{q}_2^p(z_\alpha)\}$.

As mentioned earlier, only first four cumulants are important. And as the anomaly is only due to the fourth moment of $\mu_{G^*}^{-1}(G_1^* - \mu_{G^*})$, we are going to find terms of order n^{-1}, which involve fourth moment of ϵ_1, in the first four cumulants of $\tilde{\boldsymbol{R}}_n$. We will do the same for $\tilde{\boldsymbol{R}}_n^*$. On the set $\boldsymbol{A}_n$, we have

$$\boldsymbol{R}_n = \breve{\sigma}_n^{-1}\hat{\Sigma}_n^{-1/2}\boldsymbol{T}_n = \breve{\sigma}_n^{-1}\hat{\Sigma}_n^{-1/2}n^{-1/2}\sum_{i=1}^n \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i,$$

where $\boldsymbol{T}_n = \sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)$, $(\breve{\sigma}_n^2 - \sigma^2) = n^{-1}\sum_{i=1}^n(\hat{\epsilon}_i^2 - \epsilon_i)^2 + n^{-1}\sum_{i=1}^n(\epsilon_i^2 - \sigma^2)$

$$\text{and} \quad \breve{\sigma}_n^{-1} = \sigma^{-1} - (2\sigma^3)^{-1}(\breve{\sigma}_n^2 - \sigma^2) + 3(8\sigma^5)^{-1}(\breve{\sigma}_n^2 - \sigma^2)^2 + o(n^{-1}).$$

Therefore we have on the set $\boldsymbol{A}_n$,

$$\begin{aligned}
\tilde{\boldsymbol{R}}_n &= \sigma^{-1}\boldsymbol{\Sigma}_n^{-1/2}n^{-1/2}\sum_{i=1}^n \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i \\
&\quad - (2\sigma^3)^{-1}\Big[n^{-1}\sum_{i=1}^n(\epsilon_i^2 - \sigma^2)\Big]\Big[\boldsymbol{\Sigma}_n^{-1/2}n^{-1/2}\sum_{i=1}^n \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big] \\
&\quad - (2\sigma^3)^{-1}\Big[n^{-1}\sum_{i=1}^n \boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big]'\boldsymbol{C}_{11,n}\Big[n^{-1}\sum_{i=1}^n \boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big]\Big[\boldsymbol{\Sigma}_n^{-1/2}n^{-1/2}\sum_{i=1}^n \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big] \\
&\quad - (2\sigma^3)^{-1}\Big[n^{-1}\sum_{i=1}^n \boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big]'\Big[2n^{-1}\sum_{i=1}^n \boldsymbol{x}_i^{(1)}\epsilon_i\Big]\Big[\boldsymbol{\Sigma}_n^{-1/2}n^{-1/2}\sum_{i=1}^n \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big] \\
&\quad + 3(8\sigma^5)^{-1}\Big[n^{-1}\sum_{i=1}^n(\epsilon_i^2 - \sigma^2)\Big]^2\Big[\boldsymbol{\Sigma}_n^{-1/2}n^{-1/2}\sum_{i=1}^n \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big] \\
&\quad + o(n^{-1}) = \tilde{U}_n + o(n^{-1}) \quad \text{(say)}.
\end{aligned}$$

Hence enough to look into first four cumulants of $\tilde{U}_n$ instead of $\boldsymbol{R}_n$. Let us denote by $\tilde{\mu}_{j,n}$ (or $\tilde{\kappa}_{j,n}$) the collection of terms of order n^{-1} which involve fourth moment of ϵ_1, in jth raw moment (or cumulant) of $\tilde{U}_n$, $j = 1, 2, 3, 4$. Now writing $\tilde{U}_n = a + b + c + d + e$, it can be shown that

$$\begin{aligned}
&\tilde{\mu}_{1,n} = 0, \;\; \tilde{\mu}_{2,n} = \mathbb{E}(b^2 + 2ab + 2ae) = 0 \\
&\tilde{\mu}_{3,n} = 0, \;\; \tilde{\mu}_{4,n} = \mathbb{E}(a^4 + 4a^3b + 4a^3e + 6a^2b^2)
\end{aligned}$$

and hence

$$\begin{aligned}
\tilde{\kappa}_{1,n} &= \tilde{\mu}_{1,n} = 0 \\
\tilde{\kappa}_{2,n} &= \tilde{\mu}_{2,n} = 0 \\
\tilde{\kappa}_{3,n} &= \tilde{\mu}_{3,n} = 0 \\
\tilde{\kappa}_{4,n} = \tilde{\mu}_{4,n} - 3\tilde{\mu}_{2,n}^2 &= \mathbb{E}(a^4 + 4a^3b + 4a^3e + 6a^2b^2) - 6\mathbb{E}(b^2 + 2ab + 2ae)
\end{aligned}$$

$$= n^{-1}\left[\sigma^{-4}\Sigma_n^{-2}n^{-1}\sum_{i=1}^{n}(\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)})^4\mathbb{E}\epsilon_1^4\right]$$

Now similar to $\tilde{\boldsymbol{R}}_n$, it can be shown that $\tilde{\boldsymbol{R}}_n^* = \tilde{U}_n^* + o_p(n^{-1})$. Define $\tilde{\kappa}_{j,n}^*$, $j = 1, 2, 3, 4$, such that $\tilde{\kappa}_{j,n}^*$ to $\tilde{U}_n^*$ is what $\tilde{\kappa}_{j,n}$ is to $\tilde{U}_n$. Then in the same fashion it can be shown that

$$\tilde{\kappa}_{1,n}^* = 0, \quad \tilde{\kappa}_{3,n}^* = 0,$$

$$\tilde{\kappa}_{2,n}^* = n^{-1}\left[\left[\breve{\sigma}_n^{-4}n^{-1}\sum_{i=1}^{n}\hat{\epsilon}_i^4\right] - \left[\breve{\sigma}_n^{-2}\tilde{\boldsymbol{\Sigma}}_n^{-1}n^{-1}\sum_{i=1}^{n}\left(\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\right)^2\hat{\epsilon}_i^4\right]\right]\left[\frac{\mathbb{E}(G_1^* - \mu_{G^*})^4}{\mu_{G^*}^4} - 1\right],$$

$$\tilde{\kappa}_{4,n}^* = n^{-1}\left[\tilde{\boldsymbol{\Sigma}}_n^{-2}n^{-1}\sum_{i=1}^{n}\left(\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\right)^4\hat{\epsilon}_i^4\right]\frac{\mathbb{E}(G_1^* - \mu_{G^*})^4}{\mu_{G^*}^4}$$

Therefore the contribution of $\tilde{\kappa}_{j,n}$'s, $j = 1, 2, 3, 4$, in the characteristic function $\mathbb{E}\big(e^{it\tilde{\boldsymbol{R}}_n}\big)$ of $\tilde{\boldsymbol{R}}_n$ is

$$\left[\frac{1}{2!}\tilde{\kappa}_{2,n}(it)^2 + \frac{1}{4!}\tilde{\kappa}_{4,n}(it)^4\right]e^{-t^2/2}.$$

The Fourier inversion of this contribution is

$$l_2(x) = \left[\frac{1}{2!}\tilde{\kappa}_{2,n}(-1)^2 d^2/dx^2 + \frac{1}{4!}\tilde{\kappa}_{4,n}(-1)^4 d^4/dx^4\right]\phi(x).$$

For the time being assume that $n^{-1}q_2^p(x)\phi(x)$ appears in the Edgeworth expansion of $\tilde{\boldsymbol{R}}_n$ only due to $l_2(x)$. We can assume this abuse of the notation of q_2^p since we need to correct only the effect of the fourth moment. Then we have

$$q_2^p(x)\phi(x) = n\left[\frac{1}{2!}\tilde{\kappa}_{2,n}(-d/dx) + \frac{1}{4!}\tilde{\kappa}_{4,n}(-1)^3 d^3/dx^3\right]\phi(x),$$

that is $q_2^p(x) = -nx\left[\frac{1}{2!}\tilde{\kappa}_{2,n} + \frac{1}{4!}\tilde{\kappa}_{4,n}(x^2-3)\right] = -\frac{1}{4!}n\tilde{\kappa}_{4,n}x(x^2-3)$. Similarly abusing the notation $\hat{q}_2^p$, we have

$$\hat{q}_2^p(x) = -nx\left[\frac{1}{2!}\tilde{\kappa}_{2,n}^* + \frac{1}{4!}\tilde{\kappa}_{4,n}^*(x^2-3)\right].$$

We can define $\tilde{q}_2^p(\cdot)$ as

$$\tilde{q}_2^p(x) = -nx\left[\frac{1}{2!}\hat{\kappa}_{2,n}^* + \frac{1}{4!}\hat{\kappa}_{4,n}^*(x^2-3)\right],$$

where

$$\hat{\kappa}_{2,n}^* = 0 \text{ and } \hat{\kappa}_{4,n}^* = n^{-1}\left[\tilde{\boldsymbol{\Sigma}}_n^{-2}n^{-1}\sum_{i=1}^{n}\left(\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\right)^4\hat{\epsilon}_i^4\right].$$

Note that $\tilde{\boldsymbol{\Sigma}}_n \to \sigma^2\boldsymbol{\Sigma}_n$ (cf. Lemma 5 of Das, Chatterjee and Lahiri (2024)). Again $|\hat{\epsilon}_i^2 - \epsilon_i^2| \leq \|\boldsymbol{x}_i^{(1)}\|\|(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)\|$ where by Lemma 4 of Das, Chatterjee and Lahiri (2024), $\sqrt{n}\|(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)\| =$

$O_p\left(\sqrt{logn}\right)$. Hence $\tilde{q}_2^p(\cdot)$ is asymptotically unbiased for $q_2^p(\cdot)$. Therefore the correction needed for constructing symmetric confidence interval is

$$\begin{aligned}
C_n^p(z_\alpha) &= n^{-1}\left\{\hat{q}_2^p(z_\alpha) - \tilde{q}_2^p(z_\alpha)\right\} \\
&= -x\left[\frac{1}{2!}\left(\tilde{\kappa}^*_{2,n} - \hat{\kappa}^*_{2,n}\right) + \frac{1}{4!}\left(\tilde{\kappa}^*_{4,n} - \hat{\kappa}^*_{4,n}\right)(x^2-3)\right] \\
&= -n^{-1}x\left[\frac{\omega_2}{2} + \frac{\omega_4}{24}(x^2-3)\right],
\end{aligned}$$

where

$$\omega_2 = \left[\breve{\sigma}_n^{-4}\left[n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^4\right] - \breve{\sigma}_n^{-2}\tilde{\boldsymbol{\Sigma}}_n^{-1}\left[n^{-1}\sum_{i=1}^n\left\{\hat{D}_n^{(1)}\hat{C}_{11,n}^{-1}\hat{\boldsymbol{x}}_i^{(1)}\right\}^2\hat{\epsilon}_i^4\right]\right]\left[\frac{\mathbb{E}(G_1^* - \mu_{G^*})^4}{\mu_{G^*}^4} - 1\right],$$

$$\omega_4 = \tilde{\boldsymbol{\Sigma}}_n^{-2}\left[n^{-1}\sum_{i=1}^n\left\{\hat{D}_n^{(1)}\hat{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\right\}^4\hat{\epsilon}_i^4\right]\left[\frac{\mathbb{E}(G_1^* - \mu_{G^*})^4}{\mu_{G^*}^4} - 1\right].$$

Hence we are done. □

Proof of Theorem 8. The polynomial growth rate of p is dictated by the application of Lemma 1 of Das, Chatterjee and Lahiri (2024) whereas the exponential growth rate of p is dictated by Lemma 2 of Das, Chatterjee and Lahiri (2024). These are similar to the homoscedastic cases discussed in Section 5 of the main paper. Here the proof mostly follows the same line as in the proof of Theorem 3 of the main paper. The only difference is in finding the leading term $\dot{\boldsymbol{R}}_{1n}$ of $\dot{\boldsymbol{R}}_n$ where $\|\dot{\boldsymbol{R}}_n - \dot{\boldsymbol{R}}_{1n}\| = o_p(n^{-1/2})$ and similarly in case of $\dot{\boldsymbol{R}}_n^*$. Note that by equation (5) at page 52 of Turnbull (1930) we have

$$\tilde{\boldsymbol{\Sigma}}_n^{-1/2} - \boldsymbol{\Lambda}_n^{-1/2} = -\boldsymbol{\Lambda}_n^{-1/2}\boldsymbol{Z}_{1n}\boldsymbol{\Lambda}_n^{-1/2} + \boldsymbol{Z}_{2n}, \tag{2.17}$$

where $\left(\tilde{\boldsymbol{\Sigma}}_n - \boldsymbol{\Lambda}_n\right) = \boldsymbol{\Lambda}_n^{1/2}\boldsymbol{Z}_{1n} + \boldsymbol{Z}_{1n}\boldsymbol{\Lambda}_n^{1/2}$ and $\|\boldsymbol{Z}_{2n}\| \le \|\tilde{\boldsymbol{\Sigma}}_n - \boldsymbol{\Lambda}_n\|^2$. Similar to Lemma 5 of Das, Chatterjee and Lahiri (2024), it is easy to show that $\|\tilde{\boldsymbol{\Sigma}}_n - \boldsymbol{\Lambda}_n\| = O\big(n^{-1/2}(\log n)\big)$ and hence $\|\boldsymbol{Z}_{2n}\| = O\big(n^{-1}(\log n)^2\big)$ on the set $\boldsymbol{A}_n$. Therefore, the matrix Taylor's formula in Lemma 8 (cf. Das, Chatterjee and Lahiri (2024)) will imply that

$$\dot{\boldsymbol{R}}_n = \left[\boldsymbol{\Lambda}_n^{-1/2} - \boldsymbol{\Lambda}_n^{-1/2}\Big[\int_0^\infty e^{-t\boldsymbol{\Lambda}_n^{1/2}}(\tilde{\boldsymbol{\Sigma}}_n - \boldsymbol{\Lambda}_n)e^{-t\boldsymbol{\Lambda}_n^{1/2}}dt\Big]\boldsymbol{\Lambda}_n^{-1/2}\right]\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{W}_n^{(1)} + \boldsymbol{Q}_{7n}, \tag{2.18}$$

where $\mathbb{P}\big(\|\boldsymbol{Q}_{7n}\| = o(n^{-1/2})\big) = 1 - o\big(n^{-1/2}\big)$. Now define $\boldsymbol{W}_i = \big(\epsilon_i\boldsymbol{\xi}_i^{(0)\prime}, \big[\epsilon_i^2 - \mathbb{E}\epsilon_i^2\big]\boldsymbol{\xi}_i^{(1)\prime}\big)'$ for $i \in \{1,\dots,n\}$. Here $\boldsymbol{\xi}_i^{(2)\prime} = ((\xi_{i,1}^{(1)})^2, \xi_{i,1}^{(1)}\xi_{i,2}^{(1)}, \dots, \xi_{i,1}^{(1)}\xi_{ip}^{(1)}, (\xi_{i,2}^{(1)})^2, \xi_{i,2}^{(1)}\xi_{i3}^{(1)}, \dots, \xi_{i,2}^{(1)}\xi_{ip}^{(1)}, \dots, (\xi_{ip}^{(1)})^2)'$.

Now writing $\boldsymbol{W}_i = (\boldsymbol{W}_{i,1}', \boldsymbol{W}_{i,2}')'$ and $\bar{\boldsymbol{W}}_n = n^{-1}\sum_{i=1}^n \boldsymbol{W}_i = (\bar{\boldsymbol{W}}_{n,1}', \bar{\boldsymbol{W}}_{n,2}')'$ with $\boldsymbol{W}_{i,1}$ has first q components of $\boldsymbol{W}_i$ for all $i \in \{1,\dots,n\}$, we have

$$\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{W}_n^{(1)} = \sqrt{n}\bar{\boldsymbol{W}}_{n,1} \quad \text{and} \quad j\text{th row of } \left(\tilde{\boldsymbol{\Sigma}}_n - \boldsymbol{\Lambda}_n\right) = \bar{\boldsymbol{W}}_{n,2}'\boldsymbol{E}_{j,n} \tag{2.19}$$

where $\boldsymbol{E}_{j,n}$ is a matrix of order $q(q+1)/2 \times q$ with $\|\boldsymbol{E}_{j,n}\| \le q$, $j \in \{1,\dots,q\}$. Therefore, we have

$$\boldsymbol{\Lambda}_n^{-1/2}\Big[\int_0^\infty e^{-t\boldsymbol{\Lambda}_n^{1/2}}(\tilde{\boldsymbol{\Sigma}}_n - \boldsymbol{\Lambda}_n)e^{-t\boldsymbol{\Lambda}_n^{1/2}}dt\Big]\boldsymbol{\Lambda}_n^{-1/2}\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{W}_n^{(1)}$$

$$= \sqrt{n}\Big(\bar{\boldsymbol{W}}_{n,2}'\bar{\boldsymbol{M}}_1\bar{\boldsymbol{W}}_{n,1},\dots,\bar{\boldsymbol{W}}_{n,2}'\bar{\boldsymbol{M}}_p\bar{\boldsymbol{W}}_{n,1}\Big)', \tag{2.20}$$

where $\bar{\boldsymbol{M}}_k = \int_0^\infty \big[\sum_{j=1}^q m_{kjn}(t)\check{\boldsymbol{M}}_j(t)\big]dt$ with $m_{kjn}(t)$ being the (k,j)th element of the matrix $\boldsymbol{\Lambda}_n^{-1/2}e^{-t\boldsymbol{\Lambda}_n^{1/2}}$ and $\check{\boldsymbol{M}}_j(t) = \boldsymbol{E}_{j,n}e^{-t\boldsymbol{\Lambda}_n^{1/2}}\boldsymbol{\Lambda}_n^{-1/2}$, $k,j \in \{1,\dots,q\}$. Now define the $[q+q(q+1)/2]\times[q+q(q+1)/2]$ matrices $\{\boldsymbol{M}_1^\dagger,\dots,\boldsymbol{M}_p^\dagger\}$ where $\boldsymbol{M}_k^\dagger = \begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \bar{\boldsymbol{M}}_k & \boldsymbol{0}\end{bmatrix}$. Therefore from (2.18)-(2.20) we have

$$\dot{\boldsymbol{R}}_n = \sqrt{n}\bigg[\Big(\boldsymbol{\Lambda}_n^{-1/2}\ \ \boldsymbol{0}\Big)\bar{\boldsymbol{W}}_n + \Big(\bar{\boldsymbol{W}}_n'\boldsymbol{M}_1^\dagger\tilde{W}_n,\dots,\tilde{W}_n'\boldsymbol{M}_p^\dagger\bar{\boldsymbol{W}}_n\Big)'\bigg] + \boldsymbol{Q}_{7n} = \sqrt{n}H\big(\bar{\boldsymbol{W}}_n\big) + \boldsymbol{Q}_{7n}, \quad \text{(say)}, \tag{2.21}$$

where the function $H(\cdot)$ has continuous partial derivatives of all orders.

Through the same line of arguments, writing $\bar{\boldsymbol{W}}_{n,1}^* = n^{-1}\sum_{i=1}^n \boldsymbol{W}_{i,1}^* = n^{-1}\sum_{i=1}^n \boldsymbol{\xi}_i^{(1)}\mu_{G^*}^{-1}(G_i^* - \mu_{G^*})$ and $\bar{\boldsymbol{W}}_{n,2}^* = n^{-1}\sum_{i=1}^n \boldsymbol{W}_{i,2}^* = n^{-1}\sum_{i=1}^n \boldsymbol{\xi}_i^{(2)}\big[\mu_{G^*}^{-2}(G_i^* - \mu_{G^*})^2 - 1\big]$, it can be shown that

$$\dot{\boldsymbol{R}}_n^* = \sqrt{n}\bigg[\Big(\tilde{\boldsymbol{\Sigma}}_n^{-1/2}\ \ \boldsymbol{0}\Big)\bar{\boldsymbol{W}}_n^* + \Big(\bar{\boldsymbol{W}}_n^{*\prime}\boldsymbol{M}_1^{*\dagger}\bar{\boldsymbol{W}}_n^*,\dots,\bar{\boldsymbol{W}}_n^{*\prime}\boldsymbol{M}_p^{*\dagger}\bar{\boldsymbol{W}}_n^*\Big)'\bigg] + \boldsymbol{Q}_{7n}^* = \sqrt{n}\hat{H}\big(\bar{\boldsymbol{W}}_n^*\big) + \boldsymbol{Q}_{7n}^*, \tag{2.22}$$

where the function $\hat{H}(\cdot)$ has continuous partial derivatives of all orders and $\mathbb{P}^*\big(\|\boldsymbol{Q}_{7n}^*\| = o(n^{-1/2})\big) = 1 - o_p\big(n^{-1/2}\big)$. Again, $\bar{\boldsymbol{W}}_n^* = (\bar{\boldsymbol{W}}_{n,1}^{*\prime}, \bar{\boldsymbol{W}}_{n,2}^{*\prime})'$, $\boldsymbol{M}_k^{*\dagger} = \begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \bar{\boldsymbol{M}}_k^* & \boldsymbol{0}\end{bmatrix}$ are $[q+q(q+1)/2]\times[q+q(q+1)/2]$ matrices with $\bar{\boldsymbol{M}}_k^* = \int_0^\infty \big[\sum_{j=1}^q m_{kjn}^*\check{\boldsymbol{M}}_j^*(t)\big]dt$. $m_{kjn}^*(t)$ is the (k,j)th element of the matrix $\tilde{\boldsymbol{\Sigma}}_n^{-1/2}e^{-t\tilde{\boldsymbol{\Sigma}}_n^{1/2}}$ and $\check{\boldsymbol{M}}_j^*(t) = \boldsymbol{E}_{j,n}e^{-t\tilde{\boldsymbol{\Sigma}}_n^{1/2}}\tilde{\boldsymbol{\Sigma}}_n^{-1/2}$, $j,k \in \{1,\dots,q\}$.

Now the rest follows similarly as in case of the proof of Theorem 3 of the main paper. First apply Theorem 20.6 of Bhattacharya and Ranga Rao (1986) to obtain two term Edgeworth expansion of the distribution of $\sqrt{n}\boldsymbol{W}_i$ (or of the conditional distribution of $\sqrt{n}\boldsymbol{W}_i^*$ in case of PB) and then apply the transformation technique of Bhattacharya and Ghosh (1978) to get two term Edgeworth expansion of the distribution of $\dot{\boldsymbol{R}}_n$ (or of the conditional distribution of $\dot{\boldsymbol{R}}_n^*$ in case of PB). Then compare these Edgeworth expansions and apply Lemma 5 to get the desired second order result. □

Proof of Theorem 9. This theorem can be established by retracting the proof of Theorem 7 of the main paper. It is easy to show that on the set $\boldsymbol{A}_n$,

$$\begin{aligned}
\dot{\boldsymbol{R}}_n &= \boldsymbol{\Lambda}_n^{-1/2}n^{-1/2}\sum_{i=1}^n \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i \\
&\quad - (2\boldsymbol{\Lambda}_{\boldsymbol{n}}^{3/2})^{-1}\Big[n^{-1}\sum_{i=1}^n(\epsilon_i^2 - \sigma^2)\Big]\Big[n^{-1/2}\sum_{i=1}^n \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big] \\
&\quad - (2\boldsymbol{\Lambda}_n^{3/2})^{-1}\Big[n^{-1}\sum_{i=1}^n \boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big]'\boldsymbol{C}_{11,n}\Big[n^{-1}\sum_{i=1}^n \boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big]\Big[n^{-1/2}\sum_{i=1}^n \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big] \\
&\quad - (2\boldsymbol{\Lambda}_n^{3/2})^{-1}\Big[n^{-1}\sum_{i=1}^n \boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big]'\Big[2n^{-1}\sum_{i=1}^n \boldsymbol{x}_i^{(1)}\epsilon_i\Big]\Big[n^{-1/2}\sum_{i=1}^n \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big] \\
&\quad + 3(8\boldsymbol{\Lambda}_n^{5/2})^{-1}\Big[n^{-1}\sum_{i=1}^n(\epsilon_i^2 - \sigma^2)\Big]^2\Big[n^{-1/2}\sum_{i=1}^n \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\Big]
\end{aligned}$$

$$+o(n^{-1})$$
$$= \dot{a} + \dot{b} + \dot{c} + \dot{d} + \dot{e} + o(n^{-1}) \quad \text{(say)}.$$

The first four cumulants of $\dot{U}_n = \dot{a} + \dot{b} + \dot{c} + \dot{d} + \dot{e}$ (and hence of $\dot{\boldsymbol{R}}_n$) involving fourth moment of ϵ_1 as the coefficient of n^{-1} are given by

$$\dot{\kappa}_{1,n} = \dot{\kappa}_{2,n} = \dot{\kappa}_{3,n} = 0,$$
$$\dot{\kappa}_{4,n} = \mathbb{E}(\dot{a}^4 + 4\dot{a}^3\dot{b} + 4\dot{a}^3\dot{e} + 6\dot{a}^2\dot{b}^2) - 6\mathbb{E}(\dot{b}^2 + 2\dot{a}\dot{b} + 2\dot{a}\dot{e})$$
$$= n^{-1}\Big[\boldsymbol{\Lambda}_n^{-2} n^{-1} \sum_{i=1}^{n} (\boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{x}_i^{(1)})^4 \mathbb{E}\epsilon_1^4\Big].$$

Similarly the first four cumulants of $\dot{\boldsymbol{R}}_n^*$ involving fourth moment of G^* as the coefficient of n^{-1} can be shown to be

$$\dot{\kappa}^*_{1,n} = \dot{\kappa}^*_{2,n} = \dot{\kappa}^*_{3,n} = 0,$$
$$\dot{\kappa}^*_{4,n} = n^{-1}\Big[\tilde{\boldsymbol{\Sigma}}_n^{-2} n^{-1} \sum_{i=1}^{n} (\boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{x}_i^{(1)})^4 \hat{\epsilon}_i^4\Big] \mathbb{E}(G_1^* - \mu_{G^*})^4 / \mu_{G^*}^4.$$

Therefore clearly the correction factor $\dot{C}_n^p(\cdot)$ is given by

$$\dot{C}_n^p(x) = -n^{-1} x(x^2 - 3)\frac{\omega_4}{24}, \tag{2.23}$$

where

$$\omega_4 = \tilde{\boldsymbol{\Sigma}}_n^{-2}\Big[n^{-1} \sum_{i=1}^{n} \{\hat{\boldsymbol{D}}_n^{(1)} \hat{\boldsymbol{C}}_{11,n}^{-1} \hat{\boldsymbol{x}}_i^{(1)}\}^4 \hat{\epsilon}_i^4\Big]\Big[\frac{\mathbb{E}(G_1^* - \mu_{G^*})^4}{\mu_{G^*}^4} - 1\Big].$$

This proves the theorem. □

References

BABU, G. J. and SINGH, K. (1984). On one term Edgeworth correction by Efron's bootstrap. *Sankhyā Ser. A* **46** 219–232. MR778872

BHATIA, R. (1997). *Matrix analysis*. *Graduate Texts in Mathematics* **169**. Springer-Verlag, New York. https://doi.org/10.1007/978-1-4612-0653-8 MR1477662

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451. MR0471142 (57 ##10880)

BHATTACHARYA, R. N. and RANGA RAO, R. (1986). *Normal Approximation and Asymptotic Expansions*. Robert E. Krieger Publishing Co. Inc., Melbourne, FL. Reprint of the 1976 original. MR855460 (87k:60062)

CHATTERJEE, A. and LAHIRI, S. N. (2013). Rates of convergence of the adaptive LASSO estimators to the oracle distribution and higher order refinements by the bootstrap. *Ann. Statist.* **41** 1232–1259. https://doi.org/10.1214/13-AOS1106 MR3113809

DAS, D., CHATTERJEE, A. and LAHIRI, S. N. (2022). Bootstrapping Lasso estimators under variable selection consistency in high dimensions and some higher order refinements. Manuscript.

DAS, D., CHATTERJEE, A. and LAHIRI, S. N. (2024). Supplement to 'Bootstrapping LASSO estimators under variable selection consistency in high dimensions and some higher order refinements'. (Available Online).

DAS, D., GREGORY, K. and LAHIRI, S. N. (2019). Perturbation bootstrap in Adaptive Lasso. *Ann. Statist.* **47** 2080–2116. https://doi.org/10.1214/18-AOS1741 MR3953445

FUK, D. K. and NAGAEV, S. V. (1971). Probability Inequalities for Sums of Independent Random Variables. *Theory of Probability & Its Applications* **16** 643-660. https://doi.org/10.1137/1116071

KUCHIBHOTLA, A. K. and CHAKRABORTTY, A. (2022). Moving beyond sub-Gaussianity in high-dimensional statistics: applications in covariance estimation and linear regression. *Information and Inference: A Journal of the IMA*. iaac012. https://doi.org/10.1093/imaiai/iaac012

TIBSHIRANI, R. J. (2013). The lasso problem and uniqueness. *Electron. J. Stat.* **7** 1456–1490. https://doi.org/10.1214/13-EJS815 MR3066375

TURNBULL, H. W. (1930). A Matrix Form of Taylor's Theorem. *Proc. Edinb. Math. Soc.* **2** 33–54. https://doi.org/10.1017/S0013091500007537

WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inform. Theory* **55** 2183–2202. https://doi.org/10.1109/TIT.2009.2016018 MR2729873

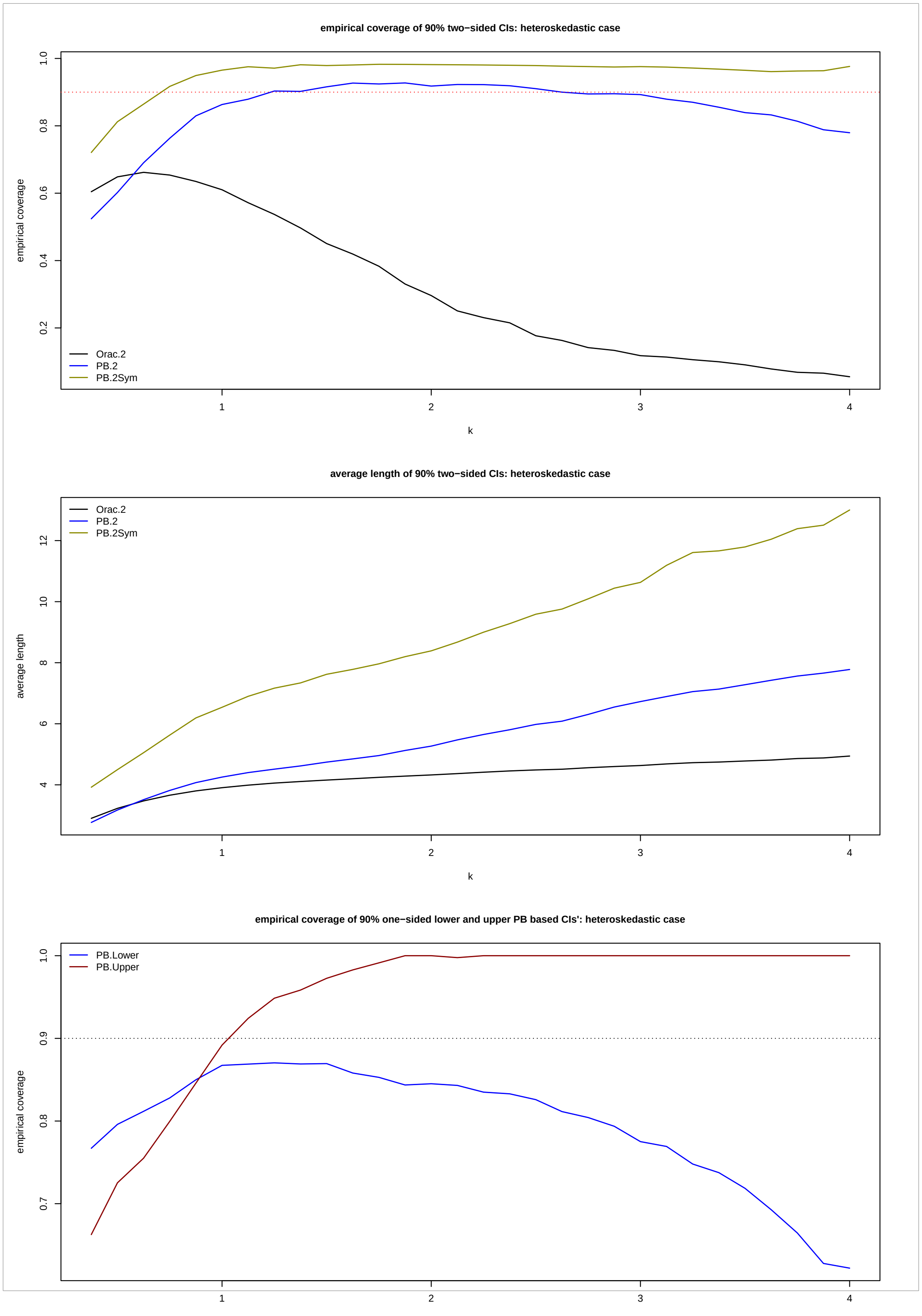

Figure 3: **Top panel** shows empirical coverages of two sided 90% confidence intervals (CIs) for target parameter β_1, in the **heteroscedastic** case as the constant k increases. These include the oracle based and PB based cases. The **middle panel** shows the average lengths of these two sided CIs. **Bottom panel** shows the empirical coverages of PB based one-sided lower and upper CIs These plots are for $n = 300$, $p = 500$, $F = F_2$ (see description in Section 1.1).

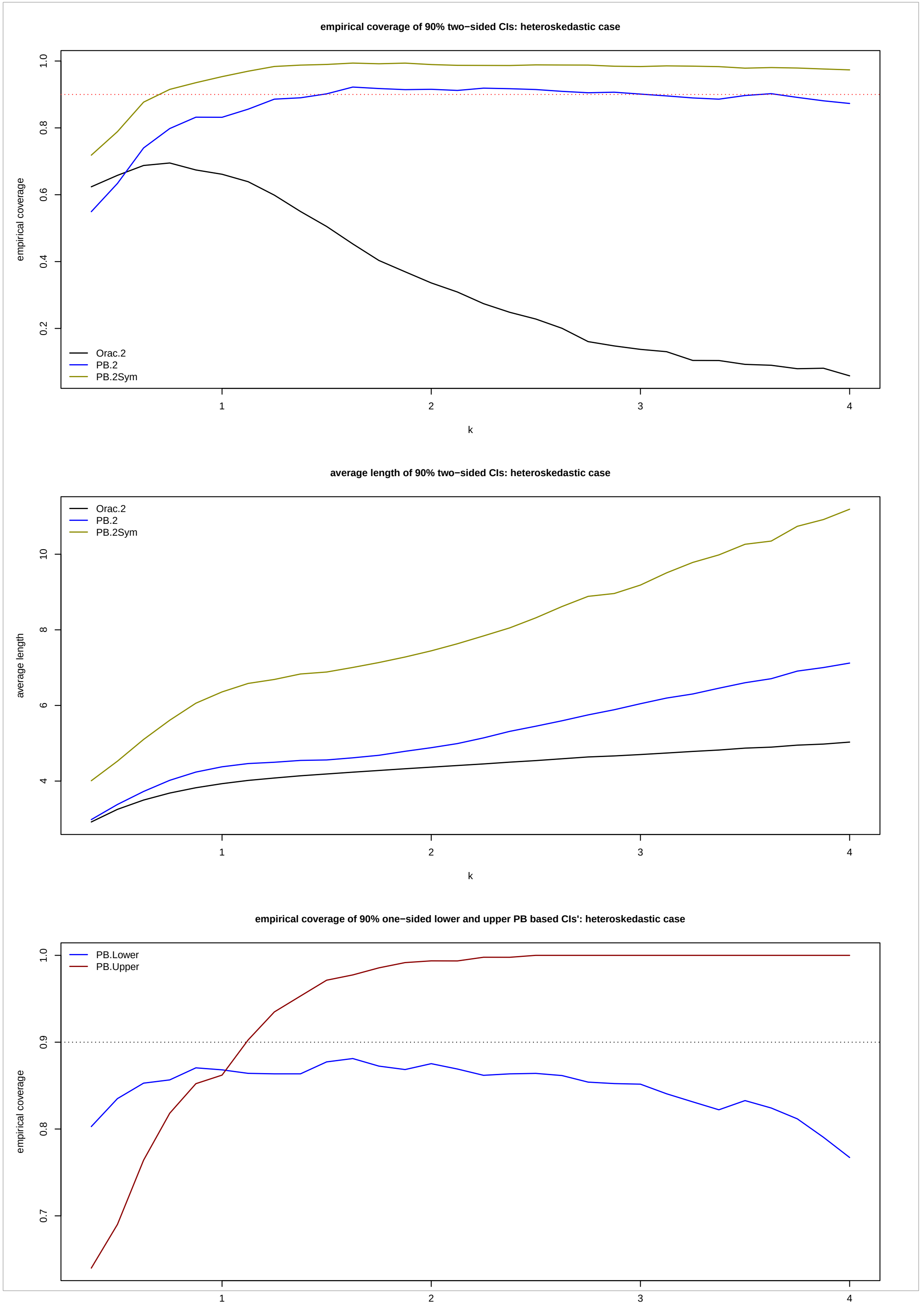

Figure 4: **Top panel** shows empirical coverages of two sided 90% confidence intervals (CIs) for target parameter β_{10}, in the **heteroscedastic** case as the constant k increases. These include the oracle based and PB based cases. The **middle panel** shows the average lengths of these two sided CIs. **Bottom panel** shows the empirical coverages of PB based one-sided lower and upper CIs These plots are for $n = 300$, $p = 500$, $F = F_2$ (see description in Section 1.1).