Bernoulli **32**(4), 2026, 2865–2893
https://doi.org/10.3150/26-BEJ1971

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*

We consider statistical inference based on the LASSO (cf. Tibshirani (*J. Roy. Statist. Soc. Ser. B, Methodol.* **58** (1996) 267–288)) in high dimensional regression problems. It is well known that the LASSO produces biased estimator of the regression parameter. The bias problem is further exacerbated when the LASSO has the variable selection consistency property (cf. Lahiri (*Ann. Statist.* **49** (2021) 820–844)). The centered and scaled point estimators are dominated by the bias term and fail to converge to *any* non-degenerate limit distribution. In contrast, we show that the LASSO interval estimators based on the Bootstrap are surprisingly accurate. Here, we consider two variants of the Bootstrap, namely, the Residual Bootstrap and the Perturbation Bootstrap, and show that under some regularity conditions, the Bootstrap approximations generated by both variants automatically adjust for the effects of the bias and are second order correct. This is an important finding as centered and scaled LASSO estimators under variable selection consistency fail to have a non-degenerate limit distribution and the Bootstrap approximations provide a viable way of constructing valid confidence intervals. Building on these second order results, we next consider construction of two-sided symmetric Bootstrap confidence intervals and show that with suitable choices of the studentized pivotal quantities, Bootstrap based two-sided symmetric confidence intervals attain a level of accuracy $O(n^{-2})$ where n is the sample size. Thus, even with penalization and with diverging model parameter dimension, the performance of the Bootstrap here remains comparable to its finite dimensional counterpart in the traditional Smooth Function model (cf. Hall (*The Bootstrap and Edgeworth Expansion* (1992) Springer-Verlag)). We also establish similar results for the heteroscedastic case for the Perturbation Bootstrap and report results from a moderate simulation study in support of the theoretical findings.

Keywords: Bootstrap; confidence interval; Edgeworth expansion; LASSO; second order correct; variable selection consistency

1. Introduction

Consider the multiple linear regression model

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n, \tag{1.1}$$

where $y_1, \dots, y_n$ are responses, $\epsilon_1, \dots, \epsilon_n$ are independent and identically distributed (iid) random variables with $E\epsilon_1 = 0$, and $E\epsilon_1^2 = \sigma^2 \in (0, \infty)$, $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are known non-random design vectors, and $\boldsymbol{\beta} \equiv \boldsymbol{\beta}_n = (\beta_{1,n}, \dots, \beta_{p,n})'$ is the p-dimensional vector of regression parameters. When the dimension p is large, particularly in the cases where $p > n$, the linear regression model (1.1) is over-parametrized and it is common to assume that only a few of the covariates are "active". In mathematical terms, this

1350-7265

means that the vector $\boldsymbol{\beta}_n$ is sparse, so that the set $\mathcal{A}_n = \{j : \beta_{j,n} \neq 0\}$ has cardinality $p_0 = |\mathcal{A}_n|$ which is much smaller than p. For notational simplicity, we shall set $\mathcal{A}_n = \{1, \dots, p_0\}$. Standard least squares approaches fail to recover the true sparsity pattern due to continuity of the least squares solutions. Traditionally, subset selection methods, such as AIC or BIC, have also been employed to identify the true sparse model but these are not computationally scalable to the large p regime. A widely used approach to handle the underlying sparsity, particularly for the $p > n$ case, is to add a penalty term to the least square criterion function, leading to penalized regression methods that aim to simultaneously select the true sparse model and also produce estimates of the nonzero regression coefficients corresponding to the active covariates.

A general definition of a penalized estimator $\hat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}_n$ is given by

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{t}} \left[\sum_{i=1}^{n} (y_i - \boldsymbol{x}_i' \boldsymbol{t})^2 + n \sum_{j=1}^{p} P_{\lambda_n, j}(|t_j|) \right] \tag{1.2}$$

where $\lambda_n > 0$ is the penalty parameter and $P_{\lambda_n,j}(\cdot) : [0, \infty) \to [0, \infty)$ is the penalty function corresponding to the jth coefficient of $\boldsymbol{\beta}_n$. For suitable choices of λ_n, the penalty terms $P_{\lambda_n,j}(\cdot)$ induce some sparsity to the estimated model. A large number of penalized estimators have been introduced in the literature, the most well known of them being the LASSO. The LASSO estimator, introduced by Tibshirani (1996), is simply the minimizer of l_1-penalized least-square criterion function where $P_{\lambda_n,j}(|t_j|) = n^{-1}\lambda_n |t_j|$ for all $j \in \{1, \dots, p\}$. Due to the convexity of the LASSO criterion function, it is computationally feasible in very high dimensional regression problems [cf. Efron et al. (2004), Friedman et al. (2007), Fu (1998), Osborne, Presnell and Turlach (2000)], and this makes the LASSO a popular choice among practitioners. The LASSO is well suited to the sparse setting because of its property that it sets some regression coefficients exactly equal to 0 and hence, it automatically leads to variable selection. With careful choice of the penalty parameter λ_n, the LASSO is variable selection consistent (VSC), i.e., with $\hat{\mathcal{A}}_n = \{1 \leq j \leq p : \hat{\beta}_{j,n} \neq 0\}$ denoting the set of selected covariates, as $n \to \infty$,

$$\mathbb{P}(\hat{\mathcal{A}}_n = \mathcal{A}_n) \to 1. \tag{1.3}$$

See Zhao and Yu (2006), Meinshausen and Bühlmann (2006), Wainwright (2009), and Lahiri (2021), and the references therein.

Building on the ideas of the LASSO, Fan and Li (2001) developed guidelines for constructing a penalty function that simultaneously guarantees variable selection consistency (also, abbreviated as VSC) as well as asymptotic normality of the penalized estimators of the nonzero $\beta_{j,n}$, leading to the concept of 'oracle property'. Specifically when p_0 is finite, a penalized regression method is said to satisfy the oracle property if (1.3) holds and if the asymptotic distribution of the estimators of the nonzero components is the same as that of the ordinary least squares (OLS) estimator under the true/"oracle" model that knows $\beta_{n,j} = 0, j = p_0 + 1, \dots, p$ *a priori*. The latter condition is equivalent to requiring

$$\sqrt{n}\boldsymbol{C}_{11,n}^{1/2}(\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n^{(1)}) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_{p_0}), \tag{1.4}$$

where $\boldsymbol{C}_{11,n}$ is the upper left $p_0 \times p_0$ submatrix of $\boldsymbol{C}_n = n^{-1}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'$, where $\boldsymbol{x}^{(1)}$ denotes the $p_0 \times 1$ subvector of the first p_0 elements of a p-vector $\boldsymbol{x}$, and where "$\xrightarrow{d}$" denotes convergence in distribution. Fan and Li (2001) developed the non-convex SCAD penalty following their own guidelines and showed that it satisfies the oracle property. There are several other penalization methods, such as the MCP (cf. Zhang (2010)), the Adaptive LASSO (or ALASSO, in short) (cf. Zou (2006)), the One-step weighted LASSO (cf. Zou and Li (2008)), and the Post-model selection OLS (cf. Belloni and Chernozhukov

(2013)) that are known to enjoy the oracle property under suitable conditions on the regression model (1.1). For penalization methods satisfying the oracle property, statistical inference on the nonzero regression coefficients can proceed using the limiting oracle normal distribution.

We next highlight the challenges to carrying out valid statistical inference based on the LASSO in high dimensions. In contrast to the penalization methods satisfying the oracle property, the LASSO typically does not satisfy the oracle property, ruling out approximations by the limiting oracle normal distribution. In a significant paper, Knight and Fu (2000) derived the limit distribution of the LASSO estimator assuming that p is fixed (as a function of the sample size n) and that $\lambda_n/\sqrt{n} \to c \in [0, \infty)$. That the model dimension p remains bounded as a function of the sample size is a very strong restriction that rules out most applications of practical interest where p grows to infinity and may even exceed n. Further, for λ_n satisfying the growth condition $\lambda_n \sim c\sqrt{n}$, some recent characterization results of Lahiri (2021) imply that the LASSO cannot select the true sparsity pattern with high probability and hence, it is not VSC. In such a scenario, on top of the fixed dimension restriction, the selected model would include spurious covariates, leading to a distorted description of the relation between the response and the covariates. Results in Lahiri (2021) also provide conditions on λ_n (outside the "$\lambda_n = O(\sqrt{n})$" regime) that allow the LASSO to be VSC, so that the LASSO selects the true model with probability tending to one. However, for such choices of λ_n, $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)$ is *not* tight (cf. Lahiri (2021)) and hence, it can no longer have a limit distribution, even for a finite p as in Knight and Fu (2000). Thus, for valid statistical inference on the model parameters based on the LASSO in high dimensions, neither the oracle limit distribution nor the non-normal limit distribution of Knight and Fu (2000) are applicable, prompting one to look for alternative distributional approximation methods such as the Bootstrap.

In this paper, we consider two popular Bootstrap methods, namely the Residual Bootstrap (RB) and the Perturbation Bootstrap (PB) methods for approximating the distributions of the LASSO estimators in high dimensions. Thus, we will allow p to grow to infinity with n and hence, in view of the discussions above, exclusively focus on the case where the LASSO has the VSC property. The main results of the paper elucidate on the error of normal approximation for the LASSO under VSC, supplementing the results of Lahiri (2021), and on higher order properties of the Bootstrap approximations generated by the RB and the PB. To describe the results more precisely, henceforth, we drop the assumption that p_0 is fixed (or bounded) and allow both p and p_0 to become unbounded with n. Define $\boldsymbol{T}_n = \sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ where $\boldsymbol{D}_n$ is a $q \times p$ matrix with $tr(\boldsymbol{D}_n\boldsymbol{D}_n') = O(1)$ (where q does not depend on n). Thus, $\boldsymbol{T}_n$ is a collection of q-many centered and scaled linear functions of $\hat{\boldsymbol{\beta}}_n$. It is easy to show that the oracle limit distribution of $\boldsymbol{T}_n$ is normal with mean zero and covariance matrix $\boldsymbol{\Sigma}_n = \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{D}_n^{(1)\prime}$ where $\boldsymbol{D}_n^{(1)}$ consists of the first p_0 columns of $\boldsymbol{D}_n$. As a measure of the error of oracle normal approximation, we consider the quantity

$$\Delta_n = \sup_{\boldsymbol{B} \in \mathcal{C}_q} |\mathbb{P}(\boldsymbol{T}_n \in \boldsymbol{B}) - \Phi(\boldsymbol{B}; \sigma^2\boldsymbol{\Sigma}_n)|$$

where $\mathcal{C}_q$ is the collection of Borel measurable convex subsets of $\mathcal{R}^q$ and $\Phi(\cdot; \boldsymbol{A})$ denotes the normal distribution on $\mathcal{R}^q$ with zero mean and variance matrix $\boldsymbol{A}$. We show that under some standard regularity conditions,

$$\Delta_n \to 1 \quad \text{as } n \to \infty.$$

Thus, the oracle normal approximation to the distribution of $\boldsymbol{T}_n$ fails. Indeed, it can be shown that in this case, $\boldsymbol{T}_n$ is no longer stochastically bounded and it can not be approximated by *any given* probability distribution (including the oracle normal distribution). On the other hand, denoting the difference between the distributions of $\boldsymbol{T}_n$ and its RB and PB approximations over the same class of sets $\mathcal{C}_q$ by Δ_n^* and Δ_n^{**}, respectively, we are able to show that

$$\Delta_n^* + \Delta_n^{**} = O_p(n^{-1/2}).$$

This shows that while the oracle normal approximation or approximations generated by any fixed distribution are not asymptotically valid, both of the Bootstrap methods closely track the variations in the distributions of $\boldsymbol{T}_n$ for different n's and are consistent. In particular, the RB and the PB can be used for constructing asymptotically valid tests and CIs for linear functions of the unknown regression parameter $\boldsymbol{\beta}_n$.

Next we consider studentized pivots based on $\boldsymbol{T}_n$. We show that with an appropriate bias-correction of the LASSO estimator, the errors of Bootstrap approximations can be improved to $o_p(n^{-1/2})$ uniformly over the class $\mathcal{C}_q$, making both RB and PB approximations second order correct (SOC). From these results, it follows that the Bootstrap methods enable, for example, the construction of higher order accurate one-sided CIs (with coverage error $O(n^{-1})$) for the nonzero regression coefficients even when dimension p is unbounded. Building on the results for the one-sided case, we push the boundary further and consider two-sided CIs. It is known (cf. Hall (1992)) that in the traditional fixed dimensional set up, one-sided Bootstrap CIs based on studentized statistics outperform their limit based counterparts due to the SOC phenomenon but in general a similar improvement is not possible for two-sided CIs. In an important work, Hall (1988) constructed two-sided symmetric Bootstrap CIs for parameters under the Smooth Function Model (cf. Hall (1992)) that achieved an even higher order improvement, leading to an impressive $O(n^{-2})$ level of accuracy in fixed dimensions. However, a direct application of Hall's (cf. Hall (1988)) construction does not yield the same level of accuracy for the LASSO studentized pivot considered above; Both the PB and the RB methods applied to this studentized pivot struggle to achieve the optimal rate due to sub-optimal performance of the residual based error variance estimator. Here, we develop a modified studentizing factor and show that the PB and the RB applied to the modified pivot attain the $O(n^{-2})$ rate. This is a striking result, particularly considering the fact that, for the LASSO with the VSC property, even a limit distribution does not exist but, with a careful construction of the pivots, the Bootstrap based statistical inference can achieve higher order accuracy levels comparable to their counterparts in the fixed dimensional Smooth Function Model even in the case where the dimension p of the regression model can grow at an exponential rate.

We conclude this section with a brief and selective literature review of Bootstrap methods for penalized regression. First we consider penalization methods that are VSC. Chatterjee and Lahiri (2013) established second order correctness of the RB for ALASSO, allowing p to grow as a polynomial function of n. First order results on the PB applied to the ALASSO and the SCAD were obtained by Minnier, Tian and Cai (2011). Recently Das, Gregory and Lahiri (2019) showed that the PB proposed by Minnier, Tian and Cai (2011) fails to be SOC and developed a modification which enables the PB to achieve second order correctness for the ALASSO. A good amount of work has been done on Bootstrapping the LASSO when it does not have the VSC property. Knight and Fu (2000) established $\sqrt{n}$-consistency and limit distribution of the LASSO estimators in fixed dimension assuming $n^{-1/2}\lambda_n \to c \in [0, \infty)$. In this case, the limit distribution is complicated and it is typically non-normal [cf. Knight and Fu (2000), Wagener and Dette (2013))]. Chatterjee and Lahiri (2010, 2011) showed that the usual RB fails for approximating the distribution of the LASSO in this setting and developed a modified RB based on thresholding the residuals when the errors are iid and design vectors are non-random. In the random design case, Camponovo (2015) developed a similar modified paired Bootstrap for the LASSO which also works in heteroscedastic regression models. Recently Das and Lahiri (2019a) developed a PB method for the LASSO which works for both random and non-random designs with heteroscedastic errors. As mentioned earlier, all of these results on the LASSO without the VSC property are in fixed p setting which is a serious limitation in most applications. The results in this paper are aimed at complementing the existing results and establishing Lasso as a viable variable selection method in high dimension with nice asymptotic properties.

Some variants of Lasso are also present in the literature. Lee et al. (2016) and Taylor and Tibshirani (2018) focused on post-selection inference by Lasso where they developed conditional inference methods in linear models after selecting a model by Lasso. More precisely, Lee et al. (2016) and Taylor and Tibshirani (2018) respectively considered Gaussian linear regression and generalized linear models and performed post-selection inference by Lasso when the dimension p is fixed and the sample size n grows to ∞. Zhang and Zhang (2014), van de Geer et al. (2014), Dezeure et al. (2015) and Javanmard and Montanari (2014) focused on building techniques to construct confidence intervals for low-dimensional parameters in high-dimensional setup without bothering about variable selection, similar to what least square estimator does in low dimensional regime. They developed a bias corrected Lasso estimator, called Debiased Lasso, for this purpose and established asymptotic normality for Debiased Lasso estimators of the individual parameters. Consistency of Wild Bootstrap methods for simultaneous CIs and testing based on Debiased LASSO estimators have also been established by Zhang and Zhang (2014) and Dezeure et al. (2015), in high dimensions. However, unlike our work, higher order accuracy of the Bootstrap remains unknown in all these cases.

The rest of the paper is organized as follows. For completeness, we describe the simpler RB method first and then the PB method in Section 2. We state the regularity conditions and assumptions in Section 3. Results on the rate of normal approximation are given in Section 4 while the second order correctness of the PB and the RB are established in Section 5, allowing p to grow at a polynomial rate and also at an exponential rate, under different moment assumptions on the error variables. Construction of the symmetric Bootstrap CIs based on the RB and the PB, and their properties are reported in Section 6. In Section 7, we give an extension of the higher order results to the heteroscedastic case, showing that the PB methods work but not the RB. Results from a simulation study are given in Section 8. Here we study finite sample performance of the two Bootstrap methods and also compare them to CIs based on the naive oracle normal approximation (which does not give a valid large sample approximation). Proofs of some of the results are presented in Section 9. Proofs of rest of the results are relegated to the supplementary material file Das, Chatterjee and Lahiri (2026). Additional simulation results are also presented in Das, Chatterjee and Lahiri (2026).

2. Description of the Bootstrap methods

Here we briefly describe the RB and the PB methods for the LASSO. While the original LASSO estimator is centered around the true value of the parameter, the Bootstrapped estimators need to be centered around suitable choices of a centering estimator $\breve{\boldsymbol{\beta}}_n$ (say) which is typically different from the original estimator. For centering the RB and the PB versions of the LASSO estimator, we are going to take $\breve{\boldsymbol{\beta}}_n$ to be the Post-LASSO OLS estimator (cf. Belloni and Chernozhukov (2013)) in the rest of the paper. This is an important step to ensure higher order accuracy of the resulting Bootstrap approximations; Intuitively, the main reason behind this consideration is to avoid the contribution of large bias of the original LASSO estimator under VSC to the centering term. See also the discussion on the effect of bias correction after the definition of pivotal quantities in Section 5.

2.1. Residual Bootstrap

First, let us briefly describe the RB method in LASSO. Define the residuals $\{\hat{\epsilon}_1, \dots, \hat{\epsilon}_n\}$ by $\hat{\epsilon}_i = y_i - x_i'\breve{\boldsymbol{\beta}}_n$. Suppose $\bar{\epsilon}_n$ is the mean of the residuals. Then, select a random sample $\{\epsilon_1^*, \dots, \epsilon_n^*\}$ with replacement from $\{(\hat{\epsilon}_1 - \bar{\epsilon}_n), \dots, (\hat{\epsilon}_n - \bar{\epsilon}_n)\}$ and define

$$y_i^* = x_i'\breve{\boldsymbol{\beta}}_n + \epsilon_i^*, \quad i = 1, \dots, n.$$

Then the RB version $\hat{\boldsymbol{\beta}}_n^*$ of the penalized estimator $\hat{\boldsymbol{\beta}}_n$ is defined as

$$\hat{\boldsymbol{\beta}}_n^* = \arg\min_{\boldsymbol{t}^*} \left\{ \sum_{i=1}^{n} (y_i^* - \boldsymbol{x}_i' \boldsymbol{t}^*)^2 + \lambda_n \sum_{j=1}^{p} |t_j^*| \right\}. \tag{2.1}$$

Note that here we have considered the actual residuals, not the thresholded residuals as in the construction of Chatterjee and Lahiri (2011). As pointed out earlier, we will use the Post-LASSO OLS estimator $\check{\boldsymbol{\beta}}_n$ to center $\hat{\boldsymbol{\beta}}_n^*$.

2.2. Perturbation Bootstrap

Let $G_1^*, \ldots, G_n^*$ be n independent copies of a non-degenerate random variable $G^* \in [0, \infty)$ having expectation μ_{G^*}. These quantities will serve as perturbation quantities in the construction of the PB version of the LASSO estimator. We define the PB version as the minimizer of a carefully constructed penalized objective function. Define the pseudo responses $\{z_1, \ldots, z_n\}$ by

$$z_i = \boldsymbol{x}_i' \check{\boldsymbol{\beta}}_n + \hat{\epsilon}_i \mu_{G^*}^{-1} (G_i^* - \mu_{G^*}),$$

where $\hat{\epsilon}_i = (y_i - \boldsymbol{x}_i' \check{\boldsymbol{\beta}}_n)$, $i = 1, \ldots, n$. Then the PB version $\hat{\boldsymbol{\beta}}_n^{**}$ of the penalized estimator $\hat{\boldsymbol{\beta}}_n$ is defined as

$$\hat{\boldsymbol{\beta}}_n^{**} = \arg\min_{\boldsymbol{t}^*} \left[\sum_{i=1}^{n} \left(z_i - \boldsymbol{x}_i' \boldsymbol{t}^*\right)^2 + \lambda_n \sum_{j=1}^{p} |t_j^*| \right]. \tag{2.2}$$

Some choices of the distribution G^* are given in Section 3 below. For the PB-versions of the pivots, we will use the Post-LASSO OLS estimator $\check{\boldsymbol{\beta}}_n$ to center $\hat{\boldsymbol{\beta}}_n^{**}$.

3. Assumptions

In this section we state the technical assumptions and briefly explain the justification behind these assumptions. We begin with some notation that will be used in the rest of the paper. We denote the true parameter vector as $\boldsymbol{\beta}_n = (\beta_{1,n}, \ldots, \beta_{p,n})'$, where the subscript n emphasizes that the dimension $p := p_n$ may grow with the sample size n. Let $\mathcal{A}_n = \{j : \beta_{j,n} \neq 0\} = \{1, \ldots, p_0\}$. For simplicity, we suppress the subscript n in p_n and p_{0n}. Consider the partition of $\boldsymbol{C}_n = n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'$ as

$$\boldsymbol{C}_n = \begin{bmatrix} \boldsymbol{C}_{11,n} & \boldsymbol{C}_{12,n} \\ \boldsymbol{C}_{21,n} & \boldsymbol{C}_{22,n} \end{bmatrix},$$

where $\boldsymbol{C}_{11,n}$ is of dimension $p_0 \times p_0$. and define $sgn(x) = -1, 0, 1$ according as $x < 0$, $x = 0$, $x > 0$, respectively. Let

$$\boldsymbol{S}_n = \begin{bmatrix} \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{D}_n^{(1)'} \cdot \sigma^2 & \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \bar{\boldsymbol{x}}_n^{(1)} \cdot \mu_3 \\ \bar{\boldsymbol{x}}_n^{(1)'} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{D}_n^{(1)'} \cdot \mu_3 & (\mu_4 - \sigma^4) \end{bmatrix},$$

where $\bar{\boldsymbol{x}}_n = n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i = (\bar{\boldsymbol{x}}_n^{(1)'}, \bar{\boldsymbol{x}}_n^{(2)'})'$ with $\bar{\boldsymbol{x}}_n^{(1)'}$ consisting of first p_0 components of $\bar{\boldsymbol{x}}_n'$, $\sigma^2 = \mathbb{V}(\epsilon_1) = \mathbb{E}(\epsilon_1^2)$, and where μ_3 and μ_4 are, respectively, the third and fourth central moments of ϵ_1.

$\boldsymbol{D}_n^{(1)}$ is the $q \times p_0$ submatrix of $\boldsymbol{D}_n$ consisting of first p_0 columns of $\boldsymbol{D}_n$. Define in addition the $q \times p_0$ matrix $\check{\boldsymbol{D}}_n^{(1)} = \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1/2}$ and the $p_0 \times 1$ vector $\check{\boldsymbol{x}}_i^{(1)} = \boldsymbol{C}_{11,n}^{-1/2} \boldsymbol{x}_i^{(1)}$. For any matrix $\boldsymbol{M}$, let $\boldsymbol{M}_{j\cdot}$ denote the jth row of $\boldsymbol{M}$. For any vector $\boldsymbol{l} \in \mathcal{R}^m$ and a real number $k \geq 1$, $\|\boldsymbol{l}\|_k^k = m^{-1} \sum_{j=1}^m |l_j|^k$ and $\|\boldsymbol{l}\|_\infty = \max\{|l_m| : 1 \leq j \leq m\}$ are respectively the L_k and the Sup norm. When $k = 2$ then we denote $\|\boldsymbol{l}\|_2$ by simply $\|\boldsymbol{l}\|$. By $\mathbb{P}^*$ and $\mathbb{E}_*$ we denote, respectively, the probability and the expectation conditional upon the observed data.

For a convex function $g : [0, \infty) \to [0, \infty)$ with $g(0) = 0$ and a zero mean random variable Y, define the g-Orlicz norm of Y as

$$\|Y\|_g = \inf\Big\{a > 0 : E\Big[g\Big(|Y|/a\Big)\Big] \leq 1\Big\}.$$

Let r be some positive number ≥ 1. Let $\delta_1 \in (0,1)$ and K_r be a constant depending only on r. We are going to use $r, \delta_1, K_r, \tilde{K}$ in stating the assumptions below. Let $g_\gamma(x) = e^{x^\gamma} - 1$. Assume that for any $\gamma > 0$, $C(\gamma) = \max\{\sqrt{2}, 2^{1/\gamma}\} \times \big[\sqrt{8} e^3 (2\pi)^{1/4} e^{1/24} \big(e^{2/e}/\gamma\big)^{1/\gamma}\big]$ when $\gamma < 1$ & $C(\gamma) = \max\{\sqrt{2}, 2^{1/\gamma}\} \times \big[4e + 2(\log 2)^{1/\gamma}\big]$ when $\gamma \geq 1$. Then define

$$\kappa(\gamma, r) = \begin{cases} 2e(C(\gamma) + e)\Big(\tilde{K}\|\epsilon\|_{g_\gamma} + \dfrac{4^{1+1/\gamma}}{\sqrt{2}}\Big) & \text{if } \gamma \in (0,1] \cup [2, \infty) \\ 2e(C(\gamma) + e)\Big(\tilde{K}\|\epsilon\|_{g_\gamma} + \dfrac{4^{1+1/\gamma}}{\sqrt{2}} K_r^{1/2r\gamma}\Big) & \text{if } \gamma \in (1,2). \end{cases}$$

We now list the assumptions. Except the assumptions (A.5) and (A.6), let all the below assumptions hold for all $n > \delta_1^{-1}$. The value of r will be specified in each theorem.

(A.1) $\max\{|(\boldsymbol{C}_{21,n})_{j\cdot} \boldsymbol{C}_{11,n}^{-1} sgn(\boldsymbol{\beta}_n^{(1)})| : 1 \leq j \leq (p - p_0)\} \leq 1 - \eta$ for some $\eta \in (0,1)$.

(A.2) (i) $\max\{n^{-1} \sum_{i=1}^n |x_{i,j}|^{2r} : 1 \leq j \leq p\} + \max\{n^{-1} \sum_{i=1}^n \big|(\boldsymbol{C}_{11,n}^{-1})_{j\cdot}\ \boldsymbol{x}_i^{(1)}\big|^{2r} : 1 \leq j \leq p_0\} + \max\{n^{-1} \sum_{i=1}^n |(\boldsymbol{C}_{21,n})_{j\cdot} \boldsymbol{C}_{11,n}^{-1} x_i^{(1)}|^{2r} : 1 \leq j \leq (p - p_0)\} + 1 \leq K_r$.

(i)′ $\max\{|x_{i,j}| : 1 \leq i \leq n, 1 \leq j \leq p\} + \max\{n^{-1} \sum_{i=1}^n \big|(\boldsymbol{C}_{11,n}^{-1})_{j\cdot} \boldsymbol{x}_i^{(1)}\big|^{2r} : 1 \leq j \leq p_0\} + \max\{|(\boldsymbol{C}_{21,n})_{j\cdot} \boldsymbol{C}_{11,n}^{-1}\ x_i^{(1)}| : 1 \leq i \leq n, 1 \leq j \leq (p - p_0)\} + 1 \leq \tilde{K}$ for some constant $\tilde{K}$.

(ii) $\boldsymbol{C}_{11,n}$ is non-singular and $\max\{c_{11,n}^{j,j} : 1 \leq j \leq p_0\} \leq K_r$, where $c_{11,n}^{j,j}$ is the (j,j)th element of $\boldsymbol{C}_{11,n}^{-1}$.

(A.3) (i) $\sup\{\boldsymbol{x}' \check{\boldsymbol{D}}_n^{(1)} \check{\boldsymbol{D}}_n^{(1)\prime} \boldsymbol{x} : \boldsymbol{x} \in \mathcal{R}^q, \|\boldsymbol{x}\| = 1\} < \delta_1^{-1}$.

(ii) $n^{-1} \sum_{i=1}^n \|\check{\boldsymbol{D}}_n^{(1)} \check{\boldsymbol{x}}_i^{(1)} \check{\boldsymbol{x}}_i^{(1)\prime} \check{\boldsymbol{D}}_n^{(1)\prime}\|^r \leq K_r$.

(iii) $\inf\{\boldsymbol{x}' \boldsymbol{S}_n \boldsymbol{x} : \boldsymbol{x} \in \mathcal{R}^{q+1}, \|\boldsymbol{x}\| = 1\} > \delta_1$.

(A.4) $\min\{|\beta_{j,n}| : j \in \mathcal{A}_n\} > \Big(\dfrac{K_r \delta_1^{-1}}{2}\Big)\Big(\dfrac{\lambda_n}{n}\Big) p_0$

(A.5) (i) $\mathbb{E}|\epsilon_1|^{2r} < \infty$. $\mathbb{E}\epsilon_1 = 0$.

(ii) $(\epsilon_1, \epsilon_1^2)$ satisfies Cramer's condition:

$\limsup_{\|(t_1,t_2)\| \to \infty} \mathbb{E}(exp(i(t_1 \epsilon_1 + t_2 \epsilon_1^2))) < 1.$

(A.6) (i) $\mathbb{E}_*(G_1^*)^{2r} < \infty$. $\mathbb{V}(G_1^*) = \sigma_{G^*}^2 = \mu_{G^*}^2$, $\mathbb{E}_*(G_1^* - \mu_{G^*})^3 = \mu_{G^*}^3$. (Only for PB)

(ii) G_i^* and ϵ_i are independent for all $1 \leq i \leq n$. (Only for PB)

(iii) $((G_1^* - \mu_{G^*}), (G_1^* - \mu_{G^*})^2)$ satisfies Cramer's condition:

$\limsup_{\|(t_1,t_2)\| \to \infty} \mathbb{E}_*(exp(i(t_1(G_1^* - \mu_{G^*}) + t_2(G_1^* - \mu_{G^*})^2))) < 1.$ (Only for PB)

(A.7) (i) $\dfrac{\lambda_n}{\sqrt{n \log n}} \geq \dfrac{re^{2r} K_r^{1/2}}{1-\eta}$ where η is defined in assumption (A.1).

(i)′ $\dfrac{\lambda_n}{n^{1/2+\delta}} \geq \dfrac{2\kappa(\gamma, r)}{1-\eta}$ for some $\delta \in (0, 1/2)$.

(ii) $\left(\dfrac{\lambda_n}{n}\right) p_0 \leq \delta_1 (\log n)^{-1}$.

If the condition (i)′ is considered instead of (i) in (A.2) or in (A.7) then we will write them as (A.2)′ or (A.7)′, respectively. Now we comment on the assumptions. Assumption (A.1) is the strong irrepresentable condition [cf. Zhao and Yu (2006)] that is required to achieve strong sign consistency as well as variable selection consistency (VSC) of the LASSO. Clearly this condition holds if $\max\{\|(\boldsymbol{C}_{21,n})_{j\cdot}\boldsymbol{C}_{11,n}^{-1}\| : 1 \leq j \leq p_0\} \leq 1-\eta$. Assumption (A.2) describes a set of regularity conditions on the design vectors. (A.2)(i) (or (i)′) and (ii) are needed to obtain moment bounds on weighted sums like $\left[\sum_{i=1}^n \boldsymbol{x}_i \epsilon_i\right]$, $\left[\boldsymbol{C}_{11,n}^{-1} \sum_{i=1}^n \boldsymbol{x}_i^{(1)} \epsilon_i\right]$, $\left[(\boldsymbol{C}_{21,n})_{j\cdot}\boldsymbol{C}_{11,n}^{-1} \sum_{i=1}^n \boldsymbol{x}_i^{(1)} \epsilon_i\right]$, etc. Also, for a general value of r, (A.2)(i), (ii) are much weaker than conditioning on L_r-norms of the design vectors. Here the value of r is determined by the order of the relevant Edgeworth expansion and will be specified in the statements of individual theorems. Assumptions (A.3)(i) and (iii) are conditions on the largest and the smallest eigen values of the respective matrices. These are necessary to obtain bounds needed in the studentized setup as well as to obtain certain Edgeworth expansions. (A.3)(ii) is specially used for moment bounds in the proofs. This is comparable to the condition (A.2)(i), but involves $\boldsymbol{D}_n$ as well. Assumption (A.4) is the beta-min condition. It is required to separate the relevant covariates from the non-relevant ones in the high dimensional regression model. More precisely, this condition along with (A.1) is ensuring the true support recovery (i.e., the VSC) property of the LASSO.

Assumption (A.5)(i) is a moment condition on the error variables needed for valid Edgeworth expansions. Assumption (A.5)(ii) is Cramer's condition on the errors, which is very common in the literature of Edgeworth expansions; it is satisfied when the distribution of $(\epsilon_1, \epsilon_1^2)$ has a non-degenerate component which is absolutely continuous with respect to the Lebesgue measure [cf. Bhattacharya and Ghosh (1978)]. Assumptions (A.6)(i) and (iii) are the analogous conditions on the perturbing random quantities to get a valid Edgeworth expansion in case of the PB only. Assumption (A.6)(ii) is natural, since the ϵ_i are present already in the data generating process, whereas G_i^* are introduced by the user. One immediate choice of the distribution of G^* is the Beta$(1/2, 3/2)$, which is used in our simulations, presented in Section 8. Other choices of the distribution of G^* can be found in Das, Gregory and Lahiri (2019). Finally, Assumption (A.7) specifies conditions on the penalty parameter and on the number of relevant covariates p_0. Assumption (A.7)(i) will be used to obtain polynomial growth rate of the dimension p whereas (A.7)(i)′ will be used for proving the results when p grows at an exponential Weibull rate with n. The constants appearing in these bounds on λ_n are due to the applications of Lemma 1 and Lemma 2 (See Das, Chatterjee and Lahiri (2026)) respectively. Assumption (A.7)(ii) is ensuring that the minimum value of the true value of the parameter corresponding to relevant covariates in Assumption (A.4) can be very close to 0. Note that (A.7) implies that p_0 can grow at most at the rate $o\left(n^{1/2}(\log n)^{-3/2}\right)$. In a linear model with $p_0 > n$, the unknown parameter is non-identifiable even in the absence of error terms. The identifiability ensures that there is a 'true' $\boldsymbol{\beta}$ behind the data generating process. In our paper, the aim is to construct inferential procedures for the 'true' $\boldsymbol{\beta}$ based on Lasso. Moreover, a Lasso solution generally contains at most $\{n, p\}$ non-zero components, as is established by Tibshirani (2013) under some mild conditions. Therefore, $n > p_0$ is natural to assume while building an inferential procedure in Lasso. However, we need the stronger condition (A.7) to accommodate studentization.

4. Rates of normal and Bootstrap approximations before studentization

In this section, we explore the quality of normal and Bootstrap approximations to the distribution of $\hat{\boldsymbol{\beta}}_n$ in the "percentile CI case", i.e., without studentization. More precisely, here we are interested in approximating the distribution of $\boldsymbol{T}_n = \sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ where $\boldsymbol{D}_n$ is a $q \times p$ matrix with q being fixed. Define the RB version of $\boldsymbol{T}_n$ as $\boldsymbol{T}_n^* = \sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n^* - \check{\boldsymbol{\beta}}_n)$ and the PB version of $\boldsymbol{T}_n$ as $\boldsymbol{T}_n^{**} = \sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n^{**} - \check{\boldsymbol{\beta}}_n)$, where, recall that, $\check{\boldsymbol{\beta}}_n$ is the post-LASSO OLS estimator. To explore the quality of normal approximation we analyze the quantity $\Delta_n = \sup_{\boldsymbol{B} \in \mathcal{C}_q} |\mathbb{P}(\boldsymbol{T}_n \in \boldsymbol{B}) - \Phi(\boldsymbol{B}; \sigma^2 \boldsymbol{\Sigma}_n)|$ where $\mathcal{C}_q$ is the collection of Borel convex subsets of $\mathcal{R}^q$. On the other hand, for investigating the quality of Bootstrap approximations, define $\Delta_n^* = \sup_{\boldsymbol{B} \in \mathcal{C}_q} \left|\mathbb{P}^*(\boldsymbol{T}_n^* \in \boldsymbol{B}) - \mathbb{P}(\boldsymbol{T}_n \in \boldsymbol{B})\right|$ and $\Delta_n^{**} = \sup_{\boldsymbol{B} \in \mathcal{C}_q} \left|\mathbb{P}^*(\boldsymbol{T}_n^{**} \in \boldsymbol{B}) - \mathbb{P}(\boldsymbol{T}_n \in \boldsymbol{B})\right|$, respectively for the RB and the PB. Now we state the first result of this section.

Theorem 1. *Suppose that the conditions (A.1), (A.2)(i), (A.3)(ii), (A.3)(iii), (A.4), (A.5)(i) and (A.7) hold with $r = 3/2$. Also consider the matrix $\boldsymbol{D}_n$ such that $\max_j \left|\left(\boldsymbol{\Sigma}_n^{-1/2}\right)_{j\cdot}\left[\boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{c}_n^{(1)}\right]\right| > \kappa$ for some $\kappa > 0$, where $\boldsymbol{\Sigma}_n = \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{D}_n^{(1)\prime}$ and $\left(\boldsymbol{\Sigma}_n^{-1/2}\right)_{j\cdot}$ is the jth row of $\boldsymbol{\Sigma}_n^{-1/2}$. Then we have*

$$\Delta_n \to 1, \text{ as } n \to \infty.$$

Theorem 1 clearly shows that unlike ALASSO (cf. Zou (2006), Chatterjee and Lahiri (2013)), one cannot use the oracle based normal approximation for the purpose of inference for the LASSO. The reason is the bias of the LASSO estimator is much larger than that present in the ALASSO. To appreciate why, recall that 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 $\hat{\boldsymbol{u}}_n = \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)$, we have

$$\begin{aligned} \hat{\boldsymbol{u}}_n &= \arg\min_{\boldsymbol{v}} \left[\boldsymbol{v}' \boldsymbol{C}_n \boldsymbol{v} - 2\boldsymbol{v}' \boldsymbol{W}_n + \lambda_n \sum_{j=1}^{p} \left(\left|\beta_{j,n} + \frac{v_j}{\sqrt{n}}\right| - \left|\beta_{j,n}\right| \right) \right] \\ &= \arg\min_{\boldsymbol{v}} \boldsymbol{Z}_n(\boldsymbol{v}), \quad \text{(say)}. \end{aligned} \tag{4.1}$$

where $\boldsymbol{C}_n = n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'$ and $\boldsymbol{W}_n = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{x}_i \epsilon_i$. Note that $\boldsymbol{Z}_n(\boldsymbol{v})$ is convex in $\boldsymbol{v}$. Hence, the necessary and sufficient KKT condition corresponding to (4.1) is given by

$$2\boldsymbol{C}_n \boldsymbol{v} - 2\boldsymbol{W}_n + \frac{\lambda_n}{\sqrt{n}} \boldsymbol{l}_n = \boldsymbol{0} \tag{4.2}$$

for some $l_{j,n} \in [-1, 1]$ for all $j \in \{1, \ldots, p\}$, where $\boldsymbol{l}_n = (l_{1,n}, \ldots, l_{p,n})'$. Again note that $\mathcal{A}_n = \{j : \beta_{j,n} \neq 0\} = \{1, \ldots, p_0\}$ and $\boldsymbol{C}_{11,n}$ is assumed to be non-singular (cf. assumption (A.2)(ii)) where $\boldsymbol{C}_{11,n}$ is the upper left $p_0 \times p_0$ submatrix of $\boldsymbol{C}_n$. Therefore, using Lemma 2 of Tibshirani (2013), it can be shown from (4.2) that on a set with probability close to 1, $\boldsymbol{T}_n$ has the form (see the proof of Theorem 1 in Section 9):

$$\boldsymbol{T}_n = \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \Big[n^{-1/2} \sum_{i=1}^{n} \boldsymbol{x}_i^{(1)} \epsilon_i - \frac{\lambda_n}{2\sqrt{n}} \boldsymbol{c}_n^{(1)} \Big] \tag{4.3}$$

where $\boldsymbol{c}_n^{(1)}$ is a $p_0 \times 1$ vector with jth component $c_{j,n} = sgn(\beta_{j,n})$. We point out that the bias term in $\boldsymbol{T}_n$ becomes very large when the LASSO has the VSC property, resulting in a large error of approximation by the oracle normal distribution.

Next we consider quality of the Bootstrap approximations over the same class of sets, as measured by Δ_n^* and Δ_n^{**}. We summarize the Bootstrap approximation results for $\boldsymbol{T}_n$ in the next theorem.

Theorem 2. *Suppose that Assumptions (A.1), (A.2)(i), (A.3)(ii), (A.3) (iii), (A.4), (A.5)(i) and (A.7) hold with $r = 3$. Then we have*

$$\Delta_n^* = O_p(n^{-1/2}).$$

If, in addition, the Assumption (A.6)(i) holds with $r = 3$, then

$$\Delta_n^{**} = O_p(n^{-1/2}).$$

Theorem 2 is actually remarkable in contrast with Theorem 1. Theorem 2 implies that both the RB and PB approximations can actually achieve the classical Berry-Esseen rate although the normal approximation fails. In particular, one can use the RB and the PB to construct Bootstrap percentile CIs (cf. Hall (1992)) for linear functions of $\boldsymbol{\beta}$. In the next section we show that suitable studentization improves the rate of both RB and PB to $o_p(n^{-1/2})$, making both the Bootstrap methods second order correct (SOC).

5. Studentization and second order correctness of the Bootstrap

We have seen in the previous section that the oracle normal approximation is of no use in approximating the distribution of the LASSO estimator. However one can use either the RB or the PB approximation to the distribution of $\boldsymbol{T}_n = \sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)$ to make inferences with accuracy $O_p(n^{-1/2})$. In this section, we explore the possibility of obtaining second order correctness (also abbreviated as SOC) of the Bootstrap by introducing studentization. Recall from (4.3) that

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

where, as before, $\boldsymbol{c}_n^{(1)}$ is the $p_0 \times 1$ vector with its jth component being $sgn(\beta_{j,n})$. Here, the bias present $\boldsymbol{T}_n$ is $\boldsymbol{b}_n = -2^{-1} n^{-1/2} \lambda_n \sqrt{n} \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{c}_n^{(1)}$. Since $\lambda_n / \sqrt{n}$ diverges to ∞ faster than $\sqrt{\log n}$ (cf. Assumption (A.7)), the bias term $\boldsymbol{b}_n$ is too large to improve the error rate from $O_p(n^{-1/2})$ to $o_p(n^{-1/2})$ even with studentization. A way out is to first correct $\boldsymbol{T}_n$ for bias and then perform proper studentization.

Recall that $\mathcal{A}_n = \{j : \beta_{j,n} \neq 0\} = \{1, \ldots, p_0\}$ and $\check{\boldsymbol{\beta}}_n$ is the Post-LASSO OLS estimator. With $\hat{\mathcal{A}}_n = \{j : \hat{\beta}_{j,n} \neq 0\}$, define $\hat{p}_{0,n} = |\hat{\mathcal{A}}_n|$. Again, for notational simplicity and without loss of generality, we suppose that $\hat{\mathcal{A}}_n = \{1, \ldots, \hat{p}_{0,n}\}$. We then partition the matrix $\boldsymbol{C}_n = n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'$ as

$$\boldsymbol{C}_n = \begin{bmatrix} \hat{\boldsymbol{C}}_{11,n} & \hat{\boldsymbol{C}}_{12,n} \\ \hat{\boldsymbol{C}}_{21,n} & \hat{\boldsymbol{C}}_{22,n} \end{bmatrix},$$

where $\hat{\boldsymbol{C}}_{11,n}$ is of dimension $\hat{p}_{0,n} \times \hat{p}_{0,n}$. Similarly, we define $\hat{\boldsymbol{D}}_n^{(1)}$ as the matrix containing the first $\hat{p}_{0,n}$ columns of $\boldsymbol{D}_n$ and we define $\hat{\boldsymbol{x}}_i^{(1)}$ as the vector containing the first $\hat{p}_{0,n}$ entries of $\boldsymbol{x}_i$. Similarly, define $\mathcal{A}_n^* = \{j : \hat{\beta}_{j,n}^* \neq 0\}$ for RB and $\mathcal{A}_n^{**} = \{j : \hat{\beta}_{j,n}^{**} \neq 0\}$ for PB. Then, an estimator of the bias term $\boldsymbol{b}_n$ is given by $\dot{\boldsymbol{b}}_n = -2^{-1} n^{-1/2} \lambda_n \sqrt{n} \hat{\boldsymbol{D}}_n^{(1)} \hat{\boldsymbol{C}}_{11,n}^{-1} \hat{\boldsymbol{c}}_n^{(1)}$, where $\hat{\boldsymbol{c}}_n^{(1)}$ is a $\hat{p}_0 \times 1$ vector with jth component $sgn(\breve{\beta}_{j,n})$. Hence the bias corrected version of $\boldsymbol{T}_n$ is given by $\big(\boldsymbol{T}_n - \dot{\boldsymbol{b}}_n\big)$. For $n > \delta_1^{-1}$ it can be shown that $\big(\boldsymbol{T}_n - \dot{\boldsymbol{b}}_n\big) = n^{-1/2} \sum_{i=1}^n \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{x}_i^{(1)} \epsilon_i = n^{-1/2} \sum_{i=1}^n \boldsymbol{\xi}_i^{(1)} \epsilon_i$ (say) and hence, the asymptotic variance of $\big[\sigma^{-1}\big(\boldsymbol{T}_n - \dot{\boldsymbol{b}}_n\big)\big]$ is given by $\boldsymbol{\Sigma}_n = n^{-1} \sum_{i=1}^n \boldsymbol{\xi}_i^{(1)} \boldsymbol{\xi}_i^{(1)\prime}$. An estimator of $\boldsymbol{\Sigma}_n$ is $\hat{\boldsymbol{\Sigma}}_n = n^{-1} \sum_{i=1}^n \hat{\boldsymbol{\xi}}_i^{(1)} \hat{\boldsymbol{\xi}}_i^{(1)\prime}$ with $\hat{\boldsymbol{\xi}}_i^{(1)} = \hat{\boldsymbol{D}}_n^{(1)} \hat{\boldsymbol{C}}_{11,n}^{-1} \hat{\boldsymbol{x}}_i^{(1)}$. On the other hand, estimators of the error variance σ^2 are given by $\breve{\sigma}_n^2 = \sum_{i=1}^n \hat{\epsilon}_i^2$ and $\hat{\sigma}_n^2 = \sum_{i=1}^n (\hat{\epsilon}_i - \bar{\epsilon}_n)^2$ where $\hat{\epsilon}_i = y_i - \boldsymbol{x}_i' \breve{\boldsymbol{\beta}}_n$, $i \in \{1, \ldots, n\}$. Here $\breve{\boldsymbol{\beta}}_n$ is the post-LASSO estimator of $\boldsymbol{\beta}_n$ and $\bar{\epsilon}_n = n^{-1} \sum_{i=1}^n \hat{\epsilon}_i$. Therefore, the bias corrected studentized versions of $\boldsymbol{T}_n$ can be defined as

$$\breve{\boldsymbol{R}}_n = \hat{\sigma}_n^{-1} (\boldsymbol{T}_n - \dot{\boldsymbol{b}}_n) \quad \text{and} \quad \tilde{\boldsymbol{R}}_n = \breve{\sigma}_n^{-1} \hat{\Sigma}_n^{-1/2} (\boldsymbol{T}_n - \dot{\boldsymbol{b}}_n).$$

We point out that the pivots we are going to use for the RB and PB approximations are different, and are given by $\breve{\boldsymbol{R}}_n$ and $\tilde{\boldsymbol{R}}_n$, respectively, for the two methods. The main reason behind different studentizations for the RB and the PB is to ensure a close match of the variances in the original and in the Bootstrap scenarios which requires special care in the case of the PB, with $\tilde{\boldsymbol{R}}_n$ being more suitable for PB weight choices.

Next we are going to define the respective Bootstrapped versions of $\breve{\boldsymbol{R}}_n$ and $\tilde{\boldsymbol{R}}_n$. Using the Bootstrap KKT conditions, it can be shown that, with Bootstrap probability close to 1,

$$\{j : \beta_{j,n}^* \neq 0\} = \{j : \beta_{j,n}^{**} \neq 0\} = \mathcal{A}_n$$

and the RB and the PB versions of the LASSO estimator are respectively $\hat{\boldsymbol{\beta}}_n^* = \Big(\hat{\boldsymbol{\beta}}_n^{*(1)\prime}, \boldsymbol{0}'\Big)'$ and $\hat{\boldsymbol{\beta}}_n^{**} = \Big(\hat{\boldsymbol{\beta}}_n^{**(1)\prime}, \boldsymbol{0}'\Big)'$ for all $n > \delta_1^{-1}$, where

$$\hat{\boldsymbol{\beta}}_n^{*(1)} = \boldsymbol{C}_{11,n}^{-1} \Big[n^{-1} \sum_{i=1}^n \boldsymbol{x}_i^{(1)} y_i^* - \frac{\lambda_n}{2n} \hat{\boldsymbol{c}}_n^{(1)} \Big] \text{ and } \hat{\boldsymbol{\beta}}_n^{**(1)} = \boldsymbol{C}_{11,n}^{-1} \Big[n^{-1} \sum_{i=1}^n \boldsymbol{x}_i^{(1)} z_i - \frac{\lambda_n}{2n} \hat{\boldsymbol{c}}_n^{(1)} \Big]. \tag{5.1}$$

with $\hat{\boldsymbol{c}}_n^{(1)}$ is as defined before and $\{y_1^*, \ldots, y_n^*\}$ and $\{z_1, \ldots, z_n\}$ are as defined in Section 2. Hence the bias corrected Bootstrapped estimators are defined as $\dot{\boldsymbol{\beta}}_n^* = \hat{\boldsymbol{\beta}}_n^* + \big(\big(2^{-1} n^{-1} \lambda_n \hat{\boldsymbol{c}}^{(1)\prime} \hat{\boldsymbol{C}}_{11,n}^{-1}\big), \boldsymbol{0}'\big)'$ for the RB and $\dot{\boldsymbol{\beta}}_n^{**} = \hat{\boldsymbol{\beta}}_n^{**} + \big(\big(2^{-1} n^{-1} \lambda_n \hat{\boldsymbol{c}}^{(1)\prime} \hat{\boldsymbol{C}}_{11,n}^{-1}\big), \boldsymbol{0}'\big)'$ for the PB. Similarly, the bias corrected Bootstrap versions of $\boldsymbol{T}_n$ are $\big(\boldsymbol{T}_n^* - \dot{\boldsymbol{b}}_n\big)$ and $\big(\boldsymbol{T}_n^{**} - \dot{\boldsymbol{b}}_n\big)$ where $\dot{\boldsymbol{b}}_n$ is as defined earlier in this section. Using the bias corrected Bootstrapped estimators, define the RB estimator of the error variance as $\sigma_n^{*2} = n^{-1} \sum_{i=1}^n (y_i^* - \boldsymbol{x}_i' \dot{\boldsymbol{\beta}}_n^*)^2$ and the PB estimator of the error variance as $\sigma_n^{**2} = \mu_{G^*}^{-2} n^{-1} \sum_{i=1}^n (y_i - \boldsymbol{x}_i' \dot{\boldsymbol{\beta}}_n^{**})^2 (G_i^* - \mu_{G^*})^2$. Then define the RB version of $\breve{\boldsymbol{R}}_n$ and the PB version of $\tilde{\boldsymbol{R}}_n$ respectively as

$$\breve{\boldsymbol{R}}_n^* = \sigma_n^{*-1} (\boldsymbol{T}_n^* - \dot{\boldsymbol{b}}_n) \quad \text{and} \quad \tilde{\boldsymbol{R}}_n^* = \sigma_n^{**-1} \breve{\sigma}_n \tilde{\Sigma}_n^{-1/2} (\boldsymbol{T}_n^{**} - \dot{\boldsymbol{b}}_n),$$

where $\tilde{\boldsymbol{\Sigma}}_n = n^{-1} \sum_{i=1}^n \hat{\boldsymbol{\xi}}_i^{(1)} \hat{\boldsymbol{\xi}}_i^{(1)\prime} \hat{\epsilon}_i^2$. Now we are going to present the SOC results separately for two cases based on the growth rate of p with the sample size n. First we consider the case when p can grow polynomially in n and then we move to the situation when p can grow exponentially.

Before moving to second order correct results of RB and PB, let us analyse the effect of bias correction for improving the related concentration. Note that under the regularity conditions, $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)$ is of the form $\big((\boldsymbol{C}_{11,n}^{-1}\big[\boldsymbol{W}_n^{(1)} - \frac{\lambda_n}{2\sqrt{n}}\boldsymbol{c}_n^{(1)}\big])', \mathbf{0}'\big)'$ where as $\sqrt{n}(\check{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)$ is of the form $\big((\boldsymbol{C}_{11,n}^{-1}\boldsymbol{W}_n^{(1)})', \mathbf{0}'\big)'$, each with probability $1 - o(n^{-1/2})$. See for example the proof of Lemma 4 in Das, Chatterjee and Lahiri (2026). Hence with probability $1 - o(n^{-1/2})$, we have

$$\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\|_2 \ge c\frac{\lambda_n}{\sqrt{n}}\sqrt{p_0}\eta_{11,n}^{-1},$$

and $\|\check{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\|_2 \le C\sqrt{p_0 \log n}$ for some constants $c, C > 0$. $\eta_{11,n}$ is the largest eigen value of $C_{11,n}$. Clearly if the minimum eigen value of $C_{11,n}$ is bounded away from 0, then due to the regularity condition (A.7), we have

$$\mathbb{P}\big(\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\|_2 \gg \|\check{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\|_2\big) = 1 - o(n^{-1/2}),$$

where for two sequences $\{a_n\}_{n\ge1}$ and $\{b_n\}_{n\ge1}$, $a_n \gg b_n$ denotes that $b_n/a_n \to 0$ as $n \to \infty$. Similar conclusion can be drawn in Bootstrap setting as well. This justifies the bias correction that we considered in defining the aforementioned pivotal quantities.

5.1. *p* grows polynomially

Theorem 3. *Suppose that Assumptions (A.1)-(A.5) and (A.7) hold with some integer $r \ge 4$ and that $p = o\big(n^{(r-3)/2}\,(\log n)^{r/2}\big)$. Then we have*

$$\sup_{\boldsymbol{B}\in\mathcal{C}_q}\big|\mathbb{P}^*(\check{\boldsymbol{R}}_n^* \in \boldsymbol{B}) - \mathbb{P}(\check{\boldsymbol{R}}_n \in \boldsymbol{B})\big| = o_p(n^{-1/2}).$$

If, in addition, (A.6) holds (with the same value of r), then the following is also true:

$$\sup_{\boldsymbol{B}\in\mathcal{C}_q}\big|\mathbb{P}^*(\tilde{\boldsymbol{R}}_n^* \in \boldsymbol{B}) - \mathbb{P}(\tilde{\boldsymbol{R}}_n \in \boldsymbol{B})\big| = o_p(n^{-1/2}).$$

Theorem 3 states that both RB and PB approximations are SOC for the LASSO, even when p grows polynomially with n, depending on the existence of certain absolute moments of the error variable. To see why, let us consider derivation of second order Edgeworth expansions for the unbootstrapped statistics; Edgeworth expansions for both RB and PB versions have a similar behavior. It follows from the KKT condition corresponding to the LASSO optimization problem (1.2) and the condition (A.7)(i) that we must have

$$\mathbb{P}\Big(\max_{1\le j\le p}|W_{j,n}| > K_r\sqrt{\log n}\Big) + \mathbb{P}\Big(\max_{1\le j\le p}|\check{W}_{j,n}| > K_r\sqrt{\log n}\Big) = o(n^{-1/2}), \tag{5.2}$$

where $W_{j,n} = n^{-1/2}\sum_{i=1}^n \epsilon_i x_{i,j}$ and $\check{W}_{j,n} = n^{-1/2}\sum_{i=1}^n \big(\boldsymbol{C}_{21}\big)_{j\cdot}\boldsymbol{C}_{11}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i$. In view of Lemma 1 of Das, Chatterjee and Lahiri (2026), the following bound is needed to conclude (5.2):

$$p.\Big(\max_{1\le j\le p}\Big[\sum_{i=1}^n |x_{i,j}|^{2r}\Big] + \max_{1\le j\le p}\Big[\sum_{i=1}^n \big|\big(\boldsymbol{C}_{21}\big)_{j\cdot}\boldsymbol{C}_{11}^{-1}\boldsymbol{x}_i^{(1)}\big|^{2r}\Big]\Big)\Big(\mathbb{E}|\epsilon_1|^r\Big)^2 = o\Big(n^{(r-1)/2}(\log n)^{r/2}\Big).$$

Under Assumption (A.2) (i), this necessarily requires $p = o\big(n^{(r-3)/2}(\log n)^{r/2}\big)$ when $\mathbb{E}|\epsilon_1|^r < \infty$, for some integer $r \ge 4$.

5.2. p grows exponentially

To achieve exponential growth of p with n, existence of absolute moments of ϵ_1 of a given order is not enough. Here, we need to have some control over the moment generating function (MGF) of the errors. We are going to discuss the scenario when the errors are sub-Weibull. The error ϵ is sub-Weibull of order $\gamma > 0$ if

$$\|\epsilon\|_{g_\gamma} < \infty \text{ where } g_\gamma(x) = e^{x^\gamma} - 1, \tag{5.3}$$

where $\|\cdot\|_g$, the g-Orlicz norm, is as defined in the Section 3. Another equivalent definition of ϵ being Sub-Weibull of order $\gamma > 0$ is given by

$$\mathbb{P}\big(|\epsilon| \geq t\big) \leq 2\exp\Big(-(t/c)^\gamma\Big) \text{ for all } t > 0, \tag{5.4}$$

for some constant $c > 0$. It can be shown that $c = \|\epsilon\|_{g_\gamma}$. Sub-Weibull distributions generalizes sub-Gaussian and sub-Exponential distributions (which correspond to $\gamma = 2$ or 1 in (5.3), respectively). For different results on sub-Weibull distributions, see Kuchibhotla and Chakrabortty (2022) and Vladimirova et al. (2020). Before moving to state the results, we want to point out that the random variables $\{\epsilon_1^*, \ldots, \epsilon_n^*\}$ used to define the RB version of the LASSO estimator are always conditionally bounded iid random variables and hence iid sub-Gaussian, i.e. $\|\epsilon_1^*\|_{g_2} < \infty$, irrespective of the nature of the distribution of ϵ. On the other hand, for the PB, besides Assumption (A.5) one also needs some control over the MGF of $\big[G^* - \mu_{G^*}\big]$ to achieve exponential growth of p. Hence the problem of finding optimal exponential growth rate of p needs to be handled differently for RB and PB, prompting us to state the results for the RB and the PB separately.

For the sub-Weibull growth rate of p, first we explore SOC of the RB. To that end, define $r_\gamma = 6$ whenever $\gamma \in (0, 1] \cup [2, \infty)$ and define $r_\gamma = \max\{6, 1/(\gamma - 1)\}$ whenever $\gamma \in (1, 2)$. Additionally define

$$\gamma_n = \begin{cases} n^{\gamma\delta} & \text{if } 0 < \gamma < 2 \\ \dfrac{n^{2\delta}}{(\log n)^{2/\gamma}} & \text{if } \gamma \geq 2, \end{cases}$$

where δ is as in Assumption (A.7)$'$. Then, we have the following result.

Theorem 4. *Suppose that $\|\epsilon\|_{g_\gamma} < \infty$ for some $\gamma > 0$. Let (A.1), (A.2)$'$ and (A.3)-(A.5) and (A.7)$'$ hold with some integer $r \geq r_\gamma$. Then for $p = o\big(n^{-1/2}\exp(\gamma_n)\big)$ the conclusions of Theorem 3 hold.*

Clearly, SOC of the RB holds for the studentized LASSO estimator even when the dimension p grows exponentially. For example when $\gamma = 1$, i.e. when the errors are sub-Exponential, then p can grow like $\exp(n^\theta)$ for any $0 \leq \theta < \delta$. If the errors are sub-Gaussian then p can grow like $\exp(n^\theta)$ for any $0 \leq \theta < 2\delta$. The term $\exp\big(n^{\gamma\delta}\big)$ in the growth rate of p is due to the application of Lemma 2 of Das, Chatterjee and Lahiri (2026) in the unbootstrapped case for $\gamma < 2$. On the other hand, when $\gamma \geq 2$, an extra $(\log n)^{2/\gamma}$ term appears from handling the quantity $\max\{|\hat{\epsilon}_{i_1} - \hat{\epsilon}_{i_2}| : 1 \leq i_1, i_2 \leq n\}$, the conditional sample range of ϵ_i^*'s, using the bound in (5.4). This max term is appearing due to the exponential concentration of $\sum_{i=1}^n x_{i,j}\epsilon_i^*$ $(j \in \{1, \ldots, p\})$ and $\sum_{i=1}^n (\boldsymbol{C}_{21,n})_{j\cdot}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i^*$ $(j \in \{(p_0+1), \ldots, p\})$ upon an application of Hoeffding's inequality. As mentioned earlier, $\{\epsilon_1^*, \ldots, \epsilon_n^*\}$ are always conditionally bounded iid random variables and hence Hoeffding's inequality can be applied to handle the Bootstrap version.

Next we explore the exponential growth rate of p for SOC of the PB. We shall assume that $(G^* - \mu_{G^*})$ to be sub-Gaussian, i.e. sub-Weibull of order 2. One such choice of the distribution of G_1^* is $Beta(1/2, 3/2)$.

Theorem 5. *Suppose that* $\|\epsilon\|_{g_\gamma} < \infty$ *for some* $\gamma > 0$ *and* $\|G^* - \mu_{G^*}\|_{g_2} < \infty$. *Let (A.1), (A.2), (A.3)-(A.6) and (A.7)′ hold with an integer* $r \geq 6$. *Define* $\tilde{\gamma} = \min\{\gamma, 2\}$. *Then for* $p = o\big(n^{-1/2} \exp\big(n^{\tilde{\gamma}\delta}\big)\big)$ *the conclusions of Theorem 3 hold.*

On thing that we want to point out is that the assumption $\|G^* - \mu_{G^*}\|_{g_2} < \infty$ can be relaxed to $\|G^* - \mu_{G^*}\|_{g_\gamma} < \infty$ in Theorem 5 as well as in all the subsequent theorems on PB. The choice of the distribution of G^* is in our hand and we know that $Beta(1/2, 3/2)$ satisfies (A.6) and is also sub-Gaussian (i.e. $\|G^* - \mu_{G^*}\|_{g_2} < \infty$). Theorem 5 implies that exponential growth of p is achievable using the PB as well whenever the error distribution and G^* are sub-Weibull, in addition to Assumptions (A.5) and (A.6). Note that for $\gamma < 2$, the rate of growth of p are exactly same for the RB and the PB. The difference in the rates appear when $\gamma \geq 2$. Here, Lemma 2 of Das, Chatterjee and Lahiri (2026) is applied to handle both the original and Bootstrap cases, unlike the situation with the RB.

Remark 5.1. Theorem 3, 4 and 5 show that both the RB and PB approximations are SOC for approximating the distribution of the studentized LASSO estimator in high dimensions. Depending on the existence of absolute moments or the MGF of the distribution of errors, the rate of growth of p can be respectively made polynomial or exponential. Whenever ϵ is sub-Weibull of order γ, $\log p$ can grow like n^θ for any $0 \leq \theta < [\delta * \min\{\gamma, 2\}]$ where $\lambda_n/\sqrt{n}$ grows at least as fast as n^δ (cf. assumption (A.6)′). Note that the growth of $\log p$ cannot be improved from $n^{2\delta}$ even if $\gamma > 2$, i.e. even if the tail of ϵ is thinner than that of Normal distribution. We again point out that the SOC of the Bootstrap is very surprising and also very useful for statistical inference, in view of the failure of the normal approximation (cf. Theorem 1).

Remark 5.2. All the results in this section are valid in the almost sure sense provided the Assumptions hold with a higher value of r in each theorem. The only difference in the proof is the application of Lemma 1 of Das, Chatterjee and Lahiri (2026) with $t \geq 5$ instead of $t = 3$, as is done in all its applications, and then applying the Borel-Cantelli Lemma. On the other hand, the error rate of $o_p(n^{-1/2})$ can also be replaced with $O_p(n^{-1})$ in each of the results of this section with a slight change in the growth rate of p. To appreciate why, note that for the improved error-bound, one has to obtain three-term Edgeworth expansions of the pivots instead of the two-term ones. For example, under the assumptions of Theorem 3 with an integer $r \geq 8$, one can show that

$$\sup_{\boldsymbol{B} \in \mathcal{C}_q} \big|\mathbb{P}^*(\breve{\boldsymbol{R}}_n^* \in \boldsymbol{B}) - \mathbb{P}(\breve{\boldsymbol{R}}_n \in \boldsymbol{B})\big| + \sup_{\boldsymbol{B} \in \mathcal{C}_q} \big|\mathbb{P}^*(\tilde{\boldsymbol{R}}_n^* \in \boldsymbol{B}) - \mathbb{P}(\tilde{\boldsymbol{R}}_n \in \boldsymbol{B})\big| = O_p(n^{-1}),$$

provided $p = o(n^{(r-4)/2}(\log n)^{r/2})$.

Remark 5.3. The reason behind using the post-LASSO OLS estimator and the bias corrected Bootstrap estimators $\dot{\boldsymbol{\beta}}_n^*$ and $\dot{\boldsymbol{\beta}}_n^{**}$ for estimating the error variance is to avoid unwarranted contribution from the bias of the LASSO estimator. Indeed, if we used the LASSO estimator itself for variance estimation in the unbootstrapped statistic and the corresponding Bootstrap versions instead of their bias corrected versions, then the non-trivial contribution of bias in the variance estimation will destroy the SOC property of the Bootstrap approximation under the given regularity conditions. We will continue to use the pivots of this section in the next section as well.

6. Symmetric Bootstrap confidence intervals

The second order results of the previous section can dictate the accuracy of the one-sided Bootstrap CIs. However, the second order results are not accurate enough for analyzing accuracy of two-sided

Bootstrap CIs. For a detailed discussion on this issue, see Chapter 2 of Hall (1992). Among many variants of two-sided Bootstrap intervals, one of the most promising one is the symmetric Bootstrap CI, due to its coverage error of order $O(n^{-2})$ in the classical setup; See, for example, Hall (1988) for theoretical aspects of symmetric CIs based on Efron's Bootstrap under the smooth function model.

Suppose that H_n is a pivotal quantity for a parameter of interest θ_n and H_n^* is the Bootstrap version of the original pivot H_n. Define the original quantile $h_{n,\alpha} = \inf\{x : \mathbb{P}(|H_n| \le x) \ge 1-\alpha\}$ and the Bootstrap quantile $\hat{h}_{n,\alpha} = \inf\{x : \mathbb{P}^*(|H_n^*| \le x) \ge 1-\alpha\}$, for $\alpha \in (0,1)$. Without loss of generality, we can simply define $h_{n,\alpha}$ and $\hat{h}_{n,\alpha}$ respectively as solutions of $\mathbb{P}(|H_n| \le x) \approx 1-\alpha$ and $\mathbb{P}^*(|H_n^*| \le x) \approx 1-\alpha$, due to Cramer's conditions [cf. assumption (A.5)(ii) for ϵ_i's, and (A.6)(iii) on G_1^*'s for the PB and the restricted Cramer's condition (cf. Lemma 7 in Das, Chatterjee and Lahiri (2026)) for ϵ_i^*'s in case of the RB]. Indeed, Cramer's condition implies that the heaviest atom of the distribution of the underlying pivot has mass $O(e^{-\delta_6 n})$ for some $\delta_6 > 0$ [cf. Theorem 2.3 in Hall (1992)], which is negligible compared to the desired coverage error $O(n^{-2})$. This is the same reason why there is no difference between openness and closedness of any interval considered in this paper. By symmetric Bootstrap CI of θ_n based on H_n and H_n^*, here we mean the interval $I_{n,(1-\alpha)}$ with the property that the event $\{\theta \in I_{n,(1-\alpha)}\}$ is same as the event $\{|H_n| \le \hat{h}_{n,\alpha}\}$. Clearly, the event $\{|H_n| \le \hat{h}_{n,\alpha}\}$ is an estimated version of the ideal event $\{|H_n| \le h_{n,\alpha}\}$ which corresponds to the exact symmetric CI. In most situations, $H_n = a_n(\hat{\theta}_n - \theta_n)$ for some estimator $\hat{\theta}_n$ and scaling a_n, resulting $I_{n,(1-\alpha)}$ to be symmetric around $\hat{\theta}_n$. We denote $(H_n, h_{n,\alpha}, I_{n,(1-\alpha)})$ by $(H_n^r, h_{n,\alpha}^r, I_{n,(1-\alpha)}^r)$ or $(H_n^p, h_{n,\alpha}^p, I_{n,(1-\alpha)}^p)$ when the underlying Bootstrap is the RB or the PB, respectively.

In this section, we are interested in the coverage accuracy of symmetric Bootstrap CIs for a linear combination of the components of $\boldsymbol{\beta}_n$. Hence, we have $\theta_n = \boldsymbol{D}_n\boldsymbol{\beta}_n$ with $\boldsymbol{D}_n$ being a p-dimensional row vector (i.e., a $q \times p$ matrix $q = 1$) throughout this section. While exploring symmetric CIs based on the RB, we are going to consider the pair (H_n, H_n^*) to be $(\breve{\boldsymbol{R}}_n, \breve{\boldsymbol{R}}_n^*)$ and for the PB, (H_n, H_n^*) is $(\tilde{\boldsymbol{R}}_n, \tilde{\boldsymbol{R}}_n^*)$. We show that the symmetric Bootstrap CIs based on the pair $(\breve{\boldsymbol{R}}_n, \breve{\boldsymbol{R}}_n^*)$ result in two-sided intervals with coverage error $O(n^{-2})$, whereas the coverage error remains $O(n^{-1})$ for the symmetric CIs based on $(\tilde{\boldsymbol{R}}_n, \tilde{\boldsymbol{R}}_n^*)$. We introduce some further correction terms in the form of the symmetric CIs based on $(\tilde{\boldsymbol{R}}_n, \tilde{\boldsymbol{R}}_n^*)$ and show that the modified symmetric CIs based on the PB achieve an error of the order $O(n^{-2})$ like their RB counterparts.

6.1. Results on symmetric CIs based on the RB

Here we present the results on the higher order accurate symmetric Bootstrap CIs based on the RB. As compared to two-sided normal CIs, symmetric Bootstrap CIs can produce a two-fold correction even when p grows exponentially with n. The following result gives conditions for the validity of this result.

Theorem 6. *Let $\alpha \in (0,1)$ and $\theta_n = \boldsymbol{D}_n\boldsymbol{\beta}_n$ where $\boldsymbol{D}_n$ is a p-dimensional row vector. The intervals $I_{n,(1-\alpha)}^r$ are as defined above. Then we have following results depending on the growth rate of p.*

(A) *Suppose the conditions (A.1)-(A.5) and (A.7) hold and $p = o(n^{(r-6)/2}(\log n)^{r/2})$ with any integer $r \ge 7$. Then we have*

$$\mathbb{P}(\theta_n \in I_{n,(1-\alpha)}^r) = 1 - \alpha + O(n^{-2}).$$

(B) *Suppose that $\|\epsilon\|_{g_\gamma} < \infty$ for some $\gamma > 0$. Let (A.1), (A.2)′ and (A.3)-(A.5) and (A.7)′ hold with some integer $r \ge \max\{r_\gamma, 9\}$. Also assume that $p = o(n^{-2}\exp(\gamma_n))$, where r_γ and γ_n are as* *defined just before Theorem 4. Then we have*

$$\mathbb{P}\big(\theta_n \in I^r_{n,(1-\alpha)}\big) = 1 - \alpha + O(n^{-2}).$$

Thus, it follows that the symmetric CIs based on the RB attains $O(n^{-2})$ accuracy even in high dimensions.

6.2. Results on symmetric CIs based on the PB

Here we consider the performance of the symmetric CIs based on the PB. As mentioned earlier, the naive symmetrized CIs $I^p_{n,(1-\alpha)}$ based on $\big(\tilde{\boldsymbol{R}}_n, \tilde{\boldsymbol{R}}^*_n\big)$ may not achieve the $O(n^{-2})$ accuracy level, without further modification. The modification is needed because the fourth moment of $(G^*_1 - \mu_{G^*})/\mu_{G^*}$ is always greater than 1, due to the condition (A.6)(i) on G^*, and this results in a mismatch in the $O(n^{-1})$ terms. Define $\tilde{I}^p_{n,(1-\alpha)}$ as the corrected symmetric PB CI such that $\{\theta \in I^p_{n,(1-\alpha)}\}$ is same as $\{|H_n| \le h_{n,\alpha} + C^p_n(z_\alpha)\}$ where $\Phi(z_\alpha) = 1 - \alpha/2$. The correction factor $C^p_n(\cdot)$ is given by

$$C^p_n(x) = -n^{-1} x\Big[\frac{\omega_2}{2} + \frac{\omega_4}{24}(x^2 - 3)\Big] \tag{6.1}$$

where (noting that $q = 1$ and hence, $\tilde{\boldsymbol{\Sigma}}_n$ is a scalar)

$$\omega_2 = \bigg[\breve{\sigma}_n^{-4}\Big[n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^4\Big] - \breve{\sigma}_n^{-2}\tilde{\boldsymbol{\Sigma}}_n^{-1}\Big[n^{-1}\sum_{i=1}^n \big\{\hat{\boldsymbol{D}}^{(1)}_n \hat{\boldsymbol{C}}^{-1}_{11,n}\boldsymbol{x}^{(1)}_i\big\}^2 \hat{\epsilon}_i^4\Big]\bigg]\bigg[\frac{\mathbb{E}(G^*_1 - \mu_{G^*})^4}{\mu^4_{G^*}} - 1\bigg],$$

$$\omega_4 = \tilde{\boldsymbol{\Sigma}}_n^{-2}\Big[n^{-1}\sum_{i=1}^n \big\{\hat{\boldsymbol{D}}^{(1)}_n \hat{\boldsymbol{C}}^{-1}_{11,n}\boldsymbol{x}^{(1)}_i\big\}^4 \hat{\epsilon}_i^4\Big]\bigg[\frac{\mathbb{E}(G^*_1 - \mu_{G^*})^4}{\mu^4_{G^*}} - 1\bigg].$$

As an example, when $G^*_1 \sim Beta(1/2, 3/2)$, then $\mathbb{E}(G^*_1 - \mu_{G^*})^4/\mu^4_{G^*} = 3$ and hence

$$\omega_2 = -2\breve{\sigma}_n^{-2}\tilde{\boldsymbol{\Sigma}}_n^{-1}\Big[n^{-1}\sum_{i=1}^n \big\{\hat{\boldsymbol{D}}^{(1)}_n \hat{\boldsymbol{C}}^{-1}_{11,n}\boldsymbol{x}^{(1)}_i\big\}^2 \hat{\epsilon}_i^4\Big] + \breve{\sigma}_n^{-4}\Big[n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^4\Big]$$

and

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

With the modification described above, we are now ready to formally state the result on $O(n^{-2})$ accuracy of the PB symmetric CIs.

Theorem 7. *Let $\alpha \in (0,1)$ and $\theta_n = \boldsymbol{D}_n\boldsymbol{\beta}_n$ where $\boldsymbol{D}_n$ is a p-dimensional row vector. The intervals $\tilde{I}^p_{n,(1-\alpha)}$ are as defined above. Then we have following results depending on the growth rate of p.*

(A) *Suppose the conditions (A.1)-(A.7) hold and $p = o\big(n^{(r-6)/2}(\log n)^{r/2}\big)$ with any integer $r \ge 7$. Then we have*

$$\mathbb{P}\big(\theta_n \in \tilde{I}^p_{n,(1-\alpha)}\big) = 1 - \alpha + O(n^{-2}).$$

(B) *Suppose that* $\|\epsilon\|_{g_\gamma} < \infty$ *for some* $\gamma > 0$ *and* $\|G^* - \mu_{G^*}\|_{g_2} < \infty$. *Let (A.1)-(A.6) and* (A.7)$'$ *hold with an integer* $r \geq 9$. *Let* $p = o\big(n^{-2}\exp\big(n^{\tilde{\gamma}\delta}\big)\big)$ *where* $\tilde{\gamma} = \min\{\gamma, 2\}$. *Then we have*

$$\mathbb{P}\big(\theta_n \in \tilde{I}^p_{n,(1-\alpha)}\big) = 1 - \alpha + O(n^{-2}).$$

7. Extensions to independent and non-identically distributed errors

In this section, we obtain higher accuracy results for the PB when the errors $\epsilon_1, \dots, \epsilon_n$ are independent and non-identically distributed [hereafter referred to as non-IID]. Clearly the case of non-IID errors includes the situation when the regression errors are heteroscedastic. In many practical situations, the measurements obtained have different variability due to a number of reasons and hence it is crucial for an inference procedure to be robust towards the presence of heterosckedasticity. We show that the PB can approximate the exact distribution of the LASSO estimator $\hat{\boldsymbol{\beta}}_n$ up to second order and guarantees coverage accuracy of order $O(n^{-2})$ for the symmetric CIs, even when the errors are non-IID, provided we define the pivotal quantities appropriately. On the other hand, the RB is not capable of achieving SOC in the non-IID case. This can be shown by considering a heteroscedastic simple linear regression model (with $p = 1 = p_0$) and using the arguments given by Wu (1986) and Liu (1988).

We now state some modified versions of the assumptions needed to handle the non-IID case. Let $\bar{\boldsymbol{W}}_n = \big[n^{-1}\sum_{i=1}^n \big(\boldsymbol{\xi}_i^{(1)\prime}\epsilon_i, \boldsymbol{\xi}_i^{(1)\prime}(\epsilon_i^2 - \sigma^2)\big)'\big]$ and $\boldsymbol{U}_n = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i^{(1)}\boldsymbol{x}_i^{(1)\prime}\mathbb{E}\epsilon_i^2$. Recall that $\boldsymbol{\xi}_i^{(1)} = \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}$ and $\check{\boldsymbol{D}}_n^{(1)} = \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1/2}$. Define $q(q+1)/2 \times 1$ vector as $\boldsymbol{\xi}_i^{(2)} = ((\xi_{i,1}^{(1)})^2, \xi_{i,1}^{(1)}\xi_{i,2}^{(1)}, \dots, \xi_{i,1}^{(1)}\xi_{ip}^{(1)}, (\xi_{i,2}^{(1)})^2, \xi_{i,2}^{(1)}\xi_{i3}^{(1)}, \dots, \xi_{i,2}^{(1)}\xi_{ip}^{(1)}, \dots, (\xi_{ip}^{(1)})^2)'$. Now consider the following assumptions.

(A.3) (i)$''$ $n^{-1}\sum_{i=1}^n \|\check{\boldsymbol{D}}_n^{(1)}\check{\boldsymbol{x}}_i^{(1)}\check{\boldsymbol{x}}_i^{(1)\prime}\check{\boldsymbol{D}}_n^{(1)\prime}\|^{2r} \leq K_r$.

(A.3) (ii)$''$ $\inf\{\boldsymbol{x}'Var\big(\sqrt{n}\bar{\boldsymbol{W}}_n\big)\boldsymbol{x} : \boldsymbol{x} \in \mathcal{R}^{q+1}, \|\boldsymbol{x}\| = 1\} > \delta_1$.

(A.5) (i)$''$ $\mathbb{E}\epsilon_i = 0$ for all $i \in \{1, \dots, n\}$ and $n^{-1}\sum_{i=1}^n \mathbb{E}|\epsilon_i|^{2r} = O(1)$.

(A.5) (ii)$''$ $\big(\epsilon_n, \epsilon_n^2\big)_{n=1}^\infty$ satisfies Cramer's condition in a uniform sense i.e. for any positive b,

$$\limsup_{n\to\infty} \sup_{\|(t_1,t_2)\|>b} \Big|\mathbb{E}\big(\exp\big(it_1\epsilon_n + it_2\epsilon_n^2\big)\big)\Big| < 1.$$

(A.8) $\boldsymbol{U}_n$ is non-singular and $\max\{U_n^{j,j} : 1 \leq j \leq p_0\} \leq K_r$, where $U_n^{j,j}$ is the (j, j)th element of $\boldsymbol{U}_n^{-1}$.

We will denote the assumptions (A.3), (A.5) respectively by (A.3)$''$ and (A.5)$''$ when they are defined with (i)$''$, (ii)$''$ in place of (i) and (ii) respectively.

7.1. Second order correctness of the PB approximation

Recall the quantities $\boldsymbol{T}_n = \sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)$ and its PB version $\boldsymbol{T}_n^{**} = \sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n^{**} - \check{\boldsymbol{\beta}}_n)$. Note that when the regression errors are non-IID, $\boldsymbol{T}_n$ has asymptotic variance $\boldsymbol{\Lambda}_n = \big[\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\,\boldsymbol{U}_n\boldsymbol{C}_{11,n}^{-1}\boldsymbol{D}_n^{(1)}\big]$. An estimator of $\boldsymbol{U}_n$ is $\hat{\boldsymbol{U}}_n = n^{-1}\sum_{i=1}^n \hat{\boldsymbol{x}}_i^{(1)}\hat{\boldsymbol{x}}_i^{(1)\prime}\hat{\epsilon}_i^2$ where $\hat{\epsilon}_i = y_i - \boldsymbol{x}_i'\check{\boldsymbol{\beta}}_n$, $i \in \{1, \dots, n\}$, $\check{\boldsymbol{\beta}}_n$ being the post-LASSO OLS estimator of $\boldsymbol{\beta}_n$ and $\hat{\boldsymbol{x}}_i^{(1)}$ is as defined in Section 5. Also recall the estimator $\hat{\boldsymbol{C}}_{11,n}$ of $\boldsymbol{C}_{11,n}$ from Section 5. Hence we can define the original studentized pivot as

$$\dot{\boldsymbol{R}}_n = \tilde{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{T}_n - \dot{\boldsymbol{b}}_n),$$

where $\tilde{\boldsymbol{\Sigma}}_n = \big[\hat{\boldsymbol{D}}_n^{(1)}\hat{\boldsymbol{C}}_{11,n}^{-1}\hat{\boldsymbol{U}}_n\hat{\boldsymbol{C}}_{11,n}^{-1}\hat{\boldsymbol{D}}_n^{(1)}\big] = n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\xi}}_i^{(1)}\hat{\boldsymbol{\xi}}_i^{(1)\prime}\hat{\epsilon}_i^2$ and where $\dot{\boldsymbol{b}}_n$ is an estimator of the bias term, defined in Section 5. Define the corresponding PB version as

$$\dot{\boldsymbol{R}}_n^* = \tilde{\boldsymbol{\Sigma}}_n^{**-1/2}\big(\boldsymbol{T}_n^{**} - \dot{\boldsymbol{b}}_n\big),$$

where $\tilde{\boldsymbol{\Sigma}}_n^{**} = \mu_{G^*}^{-2} n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\xi}}_i^{(1)}\hat{\boldsymbol{\xi}}_i^{(1)\prime}(y_i - \boldsymbol{x}_i'\dot{\boldsymbol{\beta}}_n^{**})^2(G_i^* - \mu_{G^*})^2$ with $\dot{\boldsymbol{\beta}}_n^{**}$ being the PB version of the bias corrected LASSO estimator, defined in Section 5. Note the difference between the pair of pivots $(\tilde{\boldsymbol{R}}_n, \tilde{\boldsymbol{R}}_n^*)$, defined in Section 5, and $(\dot{\boldsymbol{R}}_n, \dot{\boldsymbol{R}}_n^*)$, defined here. We point out that the pivot-pair $(\dot{\boldsymbol{R}}_n, \dot{\boldsymbol{R}}_n^*)$ is more general than $(\tilde{\boldsymbol{R}}_n, \tilde{\boldsymbol{R}}_n^*)$ in the sense that $(\dot{\boldsymbol{R}}_n, \dot{\boldsymbol{R}}_n^*)$ still works when errors are iid. However, unlike in case of $\tilde{\boldsymbol{R}}_n^*$, $\dot{\boldsymbol{R}}_n^*$ requires computation of a new matrix at each Bootstrap iteration. This cannot be avoided since the regression errors may not have same variance in the non-IID case. Now we are ready to state SOC of the PB in the non-IID case.

Theorem 8. *Suppose that the assumptions (A.1), (A.2), (A.3)$''$, (A.4), (A.5)$''$, (A.6), (A.7) (or (A.7)$'$ (for part (b) only)) and (A.8) hold.*

(a) *(p grows polynomially) Assume that $p = o\big(n^{(r-3)/2}(\log n)^{r/2}\big)$ and r is an integer ≥ 6. Then we have*

$$\sup_{\boldsymbol{B}\in\mathcal{C}_q}\big|\mathbb{P}^*(\dot{\boldsymbol{R}}_n^* \in \boldsymbol{B}) - \mathbb{P}(\dot{\boldsymbol{R}}_n \in \boldsymbol{B})\big| = o_p(n^{-1/2}).$$

(b) *(p grows exponentially) Assume that $\|\epsilon_i\|_{g_\gamma} < \infty$ for all $i \in \{1,\dots,n\}$ for some $\gamma > 0$ and $\|G^* - \mu_{G^*}\|_{g_2} < \infty$. r is an integer ≥ 8. Define $\tilde{\gamma} = \min\{\gamma, 2\}$. Then for $p = o\big(n^{-1/2}\exp\big(n^{\tilde{\gamma}\delta}\big)\big)$ the conclusion of part (a) holds.*

Clearly, the PB remains second order correct for high dimensional LASSO even when the errors are heteroscedastic. This is not in general true for the RB. Similar differences have also been observed in the literature, as in Liu (1988), Das and Lahiri (2019a,b). Therefore, the PB is a more widely applicable Bootstrap technique than the RB in high dimensional. linear regression.

7.2. Higher order accuracy of symmetric CIs based on the PB

In this section, we consider the performance of the symmetric CIs based on the PB when the errors are non-IID. We show that the accuracy of $O(n^{-2})$ hold in the non-IID case as well, as long as the pivot-pair $(\dot{\boldsymbol{R}}_n, \dot{\boldsymbol{R}}_n^*)$ is considered. However similar to the IID case, the naive CIs based only on $\big(\dot{\boldsymbol{R}}_n, \dot{\boldsymbol{R}}_n^*\big)$ may not achieve the $O(n^{-2})$ accuracy level, due to the fourth moment of $(G_1^* - \mu_{G^*})/\mu_{G^*}$ being strictly greater than 1, as noted before. Define $\dot{I}_{n,(1-\alpha)}^p$ as the symmetric PB CI such that $\{\theta \in \dot{I}_{n,(1-\alpha)}^p\}$ is same as $\{|H_n^p| \leq \dot{h}_{n,\alpha}^p + \dot{C}_n^p(z_\alpha)\}$ where $\dot{h}_{n,\alpha}^p$ is the $(1-\alpha)$th quantile of $|\dot{\boldsymbol{R}}_n^*|$, i.e. $\dot{h}_{n,\alpha}^p = \inf\{x : \mathbb{P}^*\big(|\dot{\boldsymbol{R}}_n^*| \leq x\big) \geq 1-\alpha\}$, and $\Phi(z_\alpha) = 1 - \alpha/2$. The correction factor $\dot{C}_n^p(\cdot)$ is now given by

$$\dot{C}_n^p(x) = -n^{-1}x(x^2-3)\frac{\omega_4}{24}, \tag{7.1}$$

where

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

Note that the form of the correction factor is simpler in the non-IID case than in the IID case. When $G_1^* \sim Beta(1/2, 3/2)$, then the correction factor becomes

$$\dot{C}_n^p(x) = -\frac{x(x^2-3)}{12n}\Big[n^{-1}\sum_{i=1}^{n}\{\hat{\boldsymbol{D}}_n^{(1)}\hat{\boldsymbol{C}}_{11,n}^{-1}\hat{\boldsymbol{x}}_i^{(1)}\}^4\hat{\epsilon}_i^4\Big]\tilde{\boldsymbol{\Sigma}}_n^{-2}.$$

We are now ready to formally state the results for the symmetric PB CIs in the non-IID case.

Theorem 9. *Let $\alpha \in (0,1)$ and $\theta_n = \boldsymbol{D}_n\boldsymbol{\beta}_n$ where $\boldsymbol{D}_n$ is a p-dimensional row vector. The intervals $\dot{I}^p_{n,(1-\alpha)}$ are as defined above. Then we have the following results depending on the growth rate of p.*

(A) *Suppose that $p = o\big(n^{(r-6)/2}(\log n)^{r/2}\big)$ and the conditions (A.1), (A.2), (A.3)$''$, (A.4), (A.5)$''$, (A.6), (A.7) and (A.8) hold with some integer $r \geq 7$. Then we have*

$$\mathbb{P}\big(\theta_n \in \dot{I}^p_{n,(1-\alpha)}\big) = 1 - \alpha + O(n^{-2}).$$

(B) *Suppose that $\|\epsilon\|_{g_\gamma} < \infty$ for some $\gamma > 0$ and $\|G^* - \mu_{G^*}\|_{g_2} < \infty$. Let (A.1), (A.2), (A.3)$''$, (A.4), (A.5)$''$, (A.6), (A.7)$'$ and (A.8) hold with some integer $r \geq 9$. Let $p = o\big(n^{-2}\exp\big(n^{\tilde{\gamma}\delta}\big)\big)$ where $\tilde{\gamma} = \min\{\gamma, 2\}$. Then we have*

$$\mathbb{P}\big(\theta_n \in \dot{I}^p_{n,(1-\alpha)}\big) = 1 - \alpha + O(n^{-2}).$$

Remark 7.1. As suggested by the referee, we can use duality between hypothesis testing and confidence interval to carry out tests on θ_n. For example to test $H_0 : \theta_n = \theta_{n0}$, for some specified parameter value $\theta_{n0} \in \mathcal{R}$, one can check if θ_{n0} lies outside the confidence intervals $I^r_{n,(1-\alpha)}$ or $\tilde{I}^p_{n,(1-\alpha)}$ or $\dot{I}^p_{n,(1-\alpha)}$, the level $(1-\alpha)$ symmetric confidence intervals for θ_n developed in this paper, to guarantee that H_0 can be rejected at asymptotic significance level α. Moreover, the actual significance level will be atmost $O(n^{-2})$ away from α, as a consequence of Theorems 6, 7 and 9. One can also perform tests for one-sided null hypothesis of the form $H_0 : \theta_n \leq \theta_{n0}$ or $H_0 : \theta_n \geq \theta_{n0}$, based on the second order results of Bootstrap presented in Section 2 in homoscedastic case and based on Theorem 8 in heteroscedastic case. However, the accuracy in the significance level for such tests will only be $O(n^{-1})$.

8. Simulation results

In this section, we present simulation results for the LASSO estimator. We compare the following two sided CIs, (i) bias corrected two-sided RB CIs based on the pivotal quantities $(\breve{\boldsymbol{R}}_n, \breve{\boldsymbol{R}}_n^*)$, (ii) symmetric two-sided RB CIs based on the pivotal quantities $(|\breve{\boldsymbol{R}}_n|, |\breve{\boldsymbol{R}}_n^*|)$, (iii) bias corrected two-sided PB CIs based on the pivotal quantities $(\tilde{\boldsymbol{R}}_n, \tilde{\boldsymbol{R}}_n^*)$, (iv) symmetric two-sided PB CIs based on the pivotal quantities $(|\tilde{\boldsymbol{R}}_n|, |\tilde{\boldsymbol{R}}_n^*|)$ and (v) two-sided CIs based on the oracle normal approximation. For one-sided CIs, we found lower and upper CIs based on the bias-corrected RB and PB. Empirical coverages of 90% confidence intervals (and their average lengths, in case of two-sided CIs) are computed. The constructions of the aforementioned pivotal quantities are presented in Section 5 and 6.

In order to obtain the design matrix, initially n samples (rows) were generated independently from a p dimensional Gaussian distribution with zero mean and covariance values $\sigma_{i,j} = (0.6)^{|i-j|}$ for all $1 \leq i, j \leq p$. Once generated, these values were transformed by centering and scaling each column. The resulting matrix, with centered and scaled columns, is used as the design matrix to generate responses according to (1.1). This design matrix remains unchanged through all Monte Carlo replications, since it is non-stochastic. We consider two choices of (n, p), *viz.*, $(n, p) = (150, 500)$ and $(n, p) = (300, 500)$. We fix $p_0 = 10$ for both these choices. We set $\beta_j = 5$, for $1 \le j \le p_0$, and $\beta_j = 0$, for $j > p_0$. In the homoscedastic errors setting, within each Monte Carlo replication, the errors $\epsilon_1, \dots, \epsilon_n$ are obtained as i.i.d. realizations from an underlying error distribution function F. Two choices of F were considered: $F_1 = N(0, 1)$ and $F_2 = (\chi_1^2 - 1)/\sqrt{2}$, corresponding to a centered and scaled χ_1^2 random variable. The simulation results are based on 500 Monte-Carlo replications and 750 bootstrap iterations within each Monte-Carlo replication. Due to similarity in the relative performances between F_1 and F_2, the results for $F = F_1$ are not displayed to save space. However, the github repository (see below) contains results (and plots for empirical coverages) for both types of error distribution. We have used the `glmnet` package in R (cf. Friedman, Hastie and Tibshirani (2010)) to compute LASSO solutions. In response to a question raised by a reviewer, we wish to point out that the design matrix entries were generated from a continuous distribution (multivariate normal), hence the LASSO solution will be unique (see Tibshirani (2013)) and the maximum number of non-zero estimated coefficients will be $\le \min\{n, p\}$.

In order to study the empirical coverages of the above CIs we use tuning parameters of the form $\lambda_n = k\lambda_{\mathrm{CV}}$, where the constant $k(> 0)$ is changed, and λ_{CV} is the five-fold cross validation (CV) based choice of the tuning parameter obtained initially from the observed data-set. The plots shown in Figures 1 (for $n = 150$) and 2 (for $n = 300$) find the empirical coverage (and average lengths) for the above mentioned collection of CIs (for both two-sided and one-sided cases) for the true non-zero coefficient β_1, at various choices of k.

Using a CV based tuning parameter λ_n is not an optimal approach in case of the LASSO estimator, especially when the primary goal is variable selection, in order to carry out the proposed post selection inference. However, the CV based approach is most frequently used in practice to choose a tuning parameter. The plots displayed in Figures 1 and 2 indicate that with a choice of the constant k over a range of values, we can achieve close to nominal coverage using the penalty parameter $\lambda_n = k\lambda_{\mathrm{CV}}$. Finding a data based optimal choice of k is beyond the scope of this article and we hope to investigate this in a future work.

The following additional simulation results are presented in the Supplementary Materials (Das, Chatterjee and Lahiri (2026)). Results are provided for the case where the target parameter is β_{10} (in case of $F = F_2$). As suggested by a reviewer, we have also included the empirical coverages for PB based CIs in the heteroscedastic case (when $(n, p) = (300, 500)$). The details of the heteroscedastic simulation setup are provided in Das, Chatterjee and Lahiri (2026). Also, in order to illustrate the improvement achieved by the two-sided symmetric RB and PB based CIs, we have compared the exact numerical values of the coverages attained by both RB and PB based two-sided and two-sided symmetric CIs in two different settings. Tables 1 and 2 in Das, Chatterjee and Lahiri (2026) provide this comparison and clearly show the improvements. Similar improvements were observed for all choices of (n, p), error distribution and target parameters. Details are available in Das, Chatterjee and Lahiri (2026). All simulated data, relevant code and other details related to this simulation study are available at github.com/chaisid24/lasso-hd-rb-pb-bernoulli-.

It is evident from the plots (see Figures 1 and 2 and additional simulations provided in Das, Chatterjee and Lahiri (2026)) that $k > 1$ is required for all the CIs to achieve close to nominal coverages which is expected since λ_{CV} does not satisfy the assumption (A.7)(i), as implied by the results of Choudhury and Das (2025). Clearly, a choice of $k \in [1.5, 2]$ ensures that the two-sided RB and PB based CIs attain close nominal coverage. The symmetric RB and PB based two-sided CIs obtain a slightly better close-to-nominal coverage compared to the usual two-sided RB and PB based CIs, at all choices of k. However, both RB and PB based two-sided symmetric CIs have larger lengths, compared to the usual RB and PB based two-sided CIs. This is a well known fact for two-sided symmetric CIs (see Hall (1992)). For the one-sided CIs, the coverage accuracy is not as tight, partly owing to the slower accuracy level, $O(n^{-1})$ in the one-sided case versus $O(n^{-2})$ in the two-sided case. Further, the coverage patterns in the one-sided

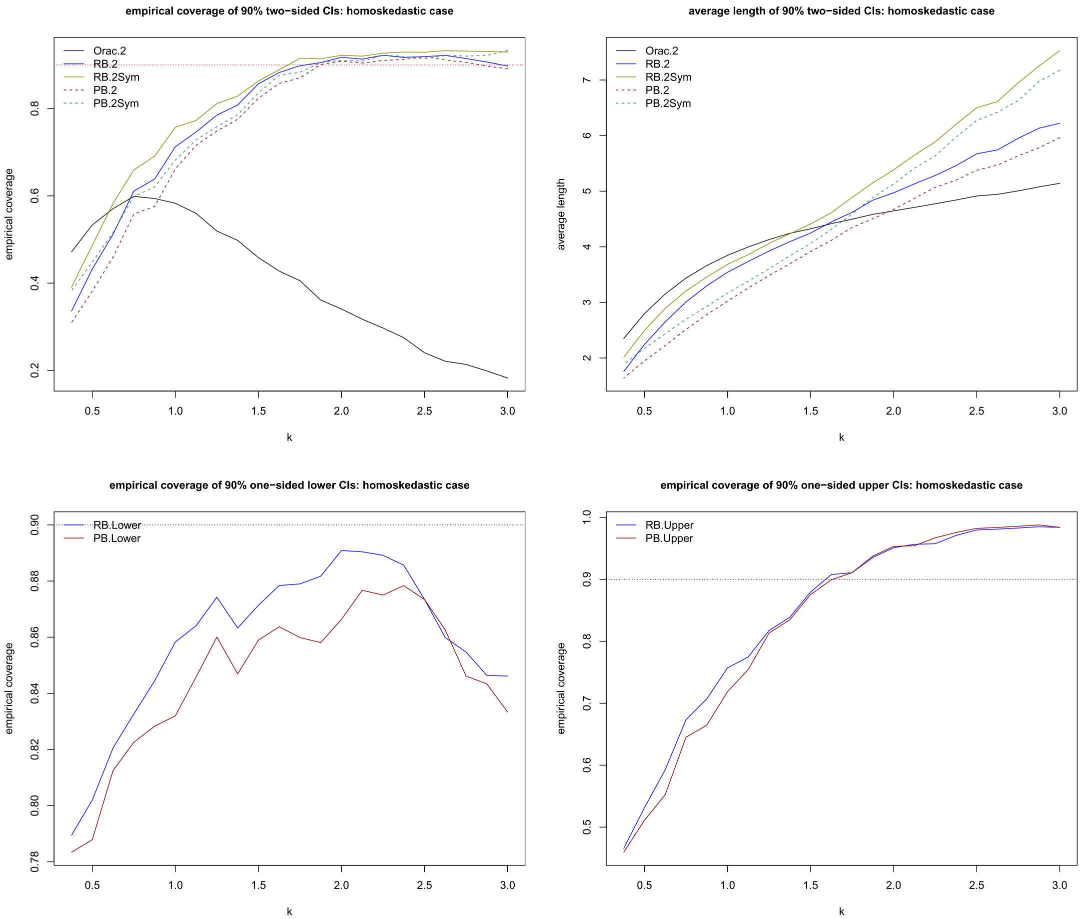

Figure 1. **Top left panel** shows empirical coverages of two sided 90% confidence intervals (CIs) for target parameter β_1, in the **homoscedastic** case as the constant k increases. These include the oracle based, RB based and PB based cases. The **top right panel** shows the average lengths of these 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 $n = 150$, $p = 500$, $F = F_2$.

case varied depending on whether one considers an upper or a lower one-sided CI. The difference in the coverage of the one-sided lower and upper CIs is also due to the ‘sign’ of the third-order cumulant. A similar phenomenon has been reported in the literature; see Chatterjee and Lahiri (2013). There is a price to pay for very large k, both in terms of support recovery, and also, the lengths of the symmetric Bootstrap CIs tend to become larger as k increases. Further, as expected in view of the results of Section 4, the Oracle based CIs perform very poorly at all levels of λ_n, with the performance deteriorating with a larger k. For the one sided CIs the best performance happens at different ranges of k, in comparison to the two sided case.

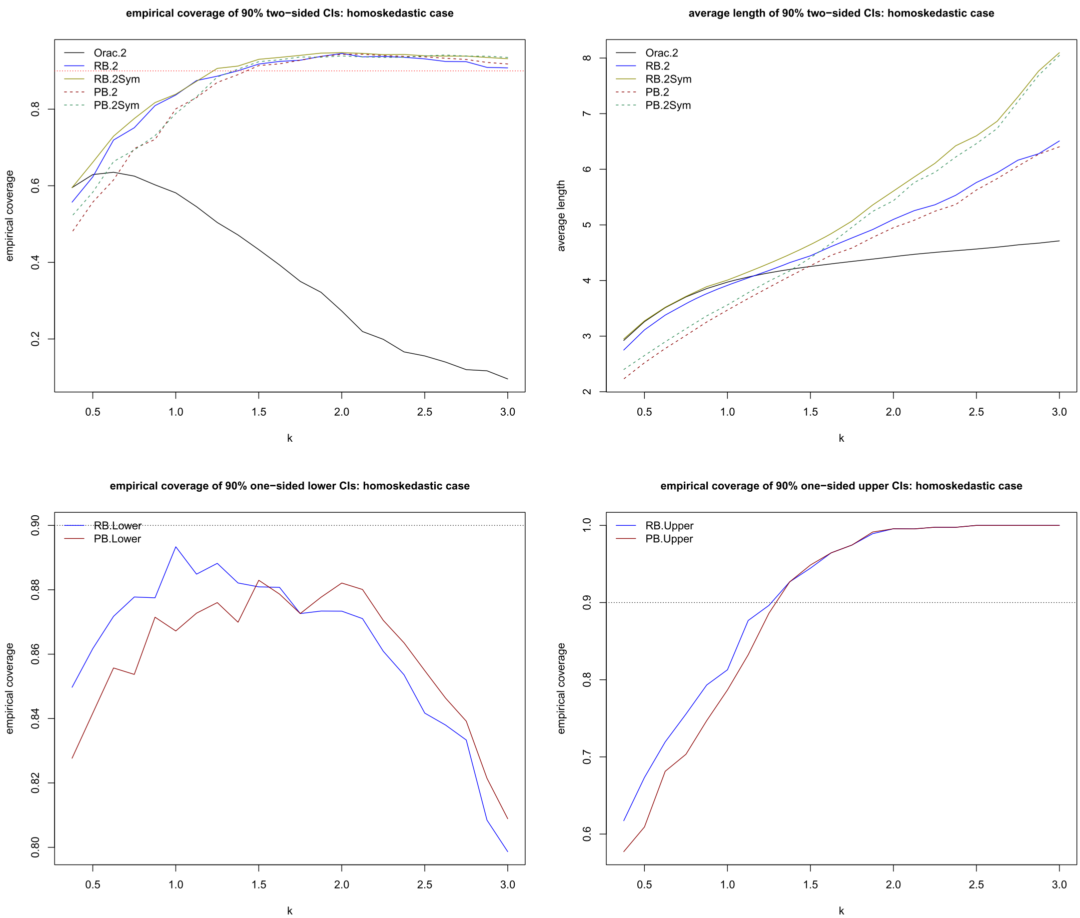

Figure 2. **Top left panel** shows empirical coverages of two sided 90% confidence intervals (CIs) for target parameter β_1, in the **homoscedastic** case as the constant k increases. These include the oracle based, RB based and PB based cases. The **top right panel** shows the average lengths of these 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 $n = 300$, $p = 500$, $F = F_2$.

9. Proofs

9.1. Notation

First we introduce some additional notation. 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 $\chi_\alpha(\boldsymbol{y}:\boldsymbol{V})$ by the identity $(-D)^\alpha\phi(\boldsymbol{y}:\boldsymbol{V}) = \chi_\alpha(\boldsymbol{y}:\boldsymbol{V})\phi(\boldsymbol{y}:\boldsymbol{V})$. We write $\chi_\alpha(\boldsymbol{y}:\boldsymbol{V})$ as $\chi_\alpha(\boldsymbol{y})$ 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 $\boldsymbol{A}_n = \big\{\{\|\boldsymbol{W}_n\|_\infty \le a_n\}\cap\{\|\dot{\boldsymbol{W}}_n\|_\infty \le a_n\}\cap\{\|\breve{\boldsymbol{W}}_n\|_\infty \le a_n\}\big\}$ 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 and the constant K (independent of n,p) is specified by the applications of the concentration bounds stated in Lemmas 1 and 2 of Das, Chatterjee and Lahiri (2026). Define the Bootstrap version of the set $\boldsymbol{A}_n$ as $\boldsymbol{A}_n^* = \big\{\{\|\boldsymbol{W}_n^*\|_\infty \le a_n\}\cap\{\|\dot{\boldsymbol{W}}_n^*\|_\infty \le a_n\}\cap\{\|\breve{\boldsymbol{W}}_n^*\|_\infty \le a_n\}\big\}$ 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 $\mathbb{P}(\boldsymbol{A}_n) \ge 1-o(n^{-1/2})$ and $\mathbb{P}^*(\boldsymbol{A}_n^*) \ge 1-o_p(n^{-1/2})$, when p grows polynomially. When p grows exponentially, same conclusions can be drawn based on Lemma 3 of Das, Chatterjee and Lahiri (2026).

9.2. Proofs of the main results

Here we present proofs of Theorems 1, 2 and 6. Proofs of Theorems 3, 7, 8 and 9 are presented in the supplementary file Das, Chatterjee and Lahiri (2026) to meet the space requirements. We omit the proofs of Theorems 4 and 5 entirely since the only difference from the proof of Theorem 3 is that Lemma 2 of Das, Chatterjee and Lahiri (2026) needs to be used to conclude the desired rates $\mathbb{P}(\boldsymbol{A}_n) = 1-o(n^{-2})$ and $\mathbb{P}^*(\boldsymbol{A}_n^*) = 1-o_p(n^{-2})$ instead of Lemma 1 of Das, Chatterjee and Lahiri (2026). Statements and proofs of a set of preliminary lemmas (numbered Lemma 1 - 8) that are used for proving all the results are given in Das, Chatterjee and Lahiri (2026).

Proof of Theorem 1. As is shown in the proof of Lemma 4 of Das, Chatterjee and Lahiri (2026), $\hat{\boldsymbol{u}}_n = \big((\hat{\boldsymbol{u}}_{3n}^{(1)})', \boldsymbol{0}'\big)'$ 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 the LASSO minimization problem. Here $\boldsymbol{c}_n^{(1)} = (c_{1n},\ldots,c_{p_0n})$ and $c_{j,n} = sgn(\beta_{j,n})$ on the set $\boldsymbol{A}_n$. Therefore on the set $\boldsymbol{A}_n$ we have

$$\boldsymbol{T}_n = \sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_i^{(1)}\epsilon_i - \frac{\lambda_n}{2\sqrt{n}}\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{c}_n^{(1)} = \boldsymbol{T}_{3n} + \boldsymbol{b}_n.$$

Under the conditions (A.3)(ii), (A.3)(iii), (A.5)(i) with $r = 3/2$, Lemma 6 of Das, Chatterjee and Lahiri (2026) implies that

$$\sup_{\boldsymbol{B}\in\mathcal{C}_q} |\mathbb{P}(\boldsymbol{T}_{3n}\in\boldsymbol{B}) - \Phi(\boldsymbol{B};\sigma^2\boldsymbol{\Sigma}_n)| = O(n^{-1/2}), \tag{9.1}$$

where $\boldsymbol{\Sigma}_n = \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{D}_n^{(1)}$. Since we have assumed $\max\big|(\boldsymbol{\Sigma}_n^{-1/2})_{j\cdot}\big[\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{c}_n^{(1)}\big]\big| > \kappa$ for some $\kappa > 0$, without loss of generality we can consider $\big|(\boldsymbol{\Sigma}_n^{-1/2})_{1\cdot}\big[\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{c}_n^{(1)}\big]\big| > \kappa$, where $(\boldsymbol{\Sigma}_n^{-1/2})_{j\cdot}$ is the jth row of $\mathbf{\Sigma}_n^{-1/2}$ and $\mathbf{\Sigma}_n = \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{D}_n^{(1)}$. Then consider a set $\boldsymbol{B}_n$ in $\mathcal{R}^q$ as

$$\boldsymbol{B}_n = \begin{cases} \left(-\infty, -\frac{\kappa\lambda_n}{4\sqrt{n}}\right) \times \mathcal{R} \times \cdots \times \mathcal{R}, & \text{if } \left(\mathbf{\Sigma}_n^{-1/2}\right)_{1\cdot} \left[\boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{c}_n^{(1)}\right] > \kappa \\ \left(\frac{\kappa\lambda_n}{4\sqrt{n}}, \infty\right) \times \mathcal{R} \times \cdots \times \mathcal{R}, & \text{if } \left(\mathbf{\Sigma}_n^{-1/2}\right)_{1\cdot} \left[\boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{c}_n^{(1)}\right] < -\kappa \end{cases},$$

Since $\boldsymbol{B}_n$ is convex, $\tilde{\boldsymbol{B}}_n = \mathbf{\Sigma}_n^{1/2} \boldsymbol{B}_n = \{\boldsymbol{y} \in \mathcal{R}^q : \boldsymbol{y} = \mathbf{\Sigma}_n^{1/2} \boldsymbol{x} \text{ for some } \boldsymbol{x} \in \boldsymbol{B}_n\}$ is also a convex set. Therefore due to (9.1) and the fact that $\mathbb{P}(\boldsymbol{A}_n) = 1 - o(n^{-1/2})$, we have

$$\begin{aligned} \Delta_n &= \sup_{\boldsymbol{B} \in \mathcal{C}_q} \left|\mathbb{P}(\boldsymbol{T}_n \in \boldsymbol{B}) - \Phi(\boldsymbol{B}; \sigma^2 \mathbf{\Sigma}_n)\right| \geq \left|\mathbb{P}(\boldsymbol{T}_n \in \tilde{\boldsymbol{B}}_n) - \Phi(\tilde{\boldsymbol{B}}_n; \sigma^2 \mathbf{\Sigma}_n)\right| \\ &\geq \left|\mathbb{P}\left(\{\boldsymbol{T}_n \in \tilde{\boldsymbol{B}}_n\} \cap \boldsymbol{A}_n\right) - \Phi(\tilde{\boldsymbol{B}}_n; \sigma^2 \mathbf{\Sigma}_n)\right| - \mathbb{P}(\boldsymbol{A}_n^c) \geq \left|\mathbb{P}\left(\boldsymbol{T}_{3n} + \boldsymbol{b}_n \in \tilde{\boldsymbol{B}}_n\right) - \Phi(\tilde{\boldsymbol{B}}_n; \sigma^2 \mathbf{\Sigma}_n)\right| - 2\mathbb{P}(\boldsymbol{A}_n^c) \\ &\geq \left|\Phi(\boldsymbol{B}_n - \mathbf{\Sigma}_n^{-1/2} \boldsymbol{b}_n; \sigma^2 \boldsymbol{I}_q) - \Phi(\boldsymbol{B}_n; \sigma^2 \boldsymbol{I}_q)\right| - o(n^{-1/2}). \end{aligned}$$

Again note that $\left|(\mathbf{\Sigma}_n^{-1/2})_{1\cdot} \boldsymbol{b}_n\right| \geq \frac{\kappa\lambda_n}{2\sqrt{n}} \geq \kappa\eta^{-1} K_{3/2} \sqrt{\log n}$, due to condition (A.7)(ii). Therefore we have

$$\left|\Phi(\boldsymbol{B}_n - \mathbf{\Sigma}_n^{-1/2} \boldsymbol{b}_n; \sigma^2 \boldsymbol{I}_q) - \Phi(\boldsymbol{B}_n; \sigma^2 \boldsymbol{I}_q)\right| \to 1 \quad \text{as } n \to \infty.$$

Therefore Theorem 1 follows. See the Das, Chatterjee and Lahiri (2026) for the details of the arguments. □

Proof of Theorem 2. It can be shown that on the set $\boldsymbol{A}_n^*$,

$$\begin{aligned} \boldsymbol{T}_n^* &= \sqrt{n} \boldsymbol{D}_n (\hat{\boldsymbol{\beta}}_n^* - \check{\boldsymbol{\beta}}_n) = n^{-1/2} \sum_{i=1}^n \hat{\boldsymbol{\xi}}_i^{(1)} \epsilon_i^* - \frac{\lambda_n}{2\sqrt{n}} \hat{\boldsymbol{D}}_n^{(1)} \hat{\boldsymbol{C}}_{11,n}^{-1} \hat{\boldsymbol{c}}_n^{(1)} \equiv \boldsymbol{T}_{3n}^* + \dot{\boldsymbol{b}}_n \quad \text{and} \\ \boldsymbol{T}_n^{**} &= \sqrt{n} \hat{\boldsymbol{D}}_n (\hat{\boldsymbol{\beta}}_n^{**} - \check{\boldsymbol{\beta}}_n) = n^{-1/2} \sum_{i=1}^n \hat{\boldsymbol{\xi}}_i^{(1)} \hat{\epsilon}_i (G_i^* - \mu_{G^*}) \mu_{G^*}^{-1} - \frac{\lambda_n}{2\sqrt{n}} \hat{\boldsymbol{D}}_n^{(1)} \hat{\boldsymbol{C}}_{11,n}^{-1} \hat{\boldsymbol{c}}_n^{(1)} \equiv \boldsymbol{T}_{3n}^{**} + \dot{\boldsymbol{b}}_n, \end{aligned}$$

where $\hat{\boldsymbol{\xi}}_i^{(1)} = \hat{\boldsymbol{D}}_n^{(1)} \hat{\boldsymbol{C}}_{11,n}^{-1} \hat{\boldsymbol{x}}_i^{(1)}$, $i \in \{1, \ldots, n\}$ and $\hat{\boldsymbol{c}}_n^{(1)} = (\hat{c}_{1n}, \ldots, \hat{c}_{p_0 n})$ with $\hat{c}_{j,n} = sgn(\hat{\beta}_{j,n})$. Note that on the set $\boldsymbol{A}_{3n}$ with $\mathbb{P}(\boldsymbol{A}_n) = 1 - o(n^{-1/2})$, $\hat{c}_{j,n} = c_{j,n}$ for all $j \in \{1, \ldots, p_0\}$ and hence $\dot{\boldsymbol{b}}_n = \boldsymbol{b}_n$. Therefore for any $k > 0$ and sufficiently large n,

$$\begin{aligned} &\mathbb{P}\left[\Delta_n^* = \sup_{\boldsymbol{B} \in \mathcal{C}_q} |\mathbb{P}^*(\boldsymbol{T}_n^* \in \boldsymbol{B}) - \mathbb{P}(\boldsymbol{T}_n \in \boldsymbol{B})| > k \cdot n^{-1/2}\right] \\ &\leq \mathbb{P}\left[\sup_{\boldsymbol{B} \in \mathcal{C}_q} \left|\mathbb{P}^*\left(\{\boldsymbol{T}_n^* \in \boldsymbol{B}\} \cap \boldsymbol{A}_n^*\right) - \mathbb{P}\left(\{\boldsymbol{T}_n \in \boldsymbol{B}\} \cap \boldsymbol{A}_n\right)\right| + \mathbb{P}_*(\boldsymbol{A}_n^{*c}) + \mathbb{P}(\boldsymbol{A}_n^c) > k \cdot n^{-1/2}\right] \\ &\leq \mathbb{P}\left[\sup_{\boldsymbol{B} \in \mathcal{C}_q} \left|\mathbb{P}^*\left(\{\boldsymbol{T}_n^* \in \boldsymbol{B}\}\right) - \mathbb{P}\left(\{\boldsymbol{T}_n \in \boldsymbol{B}\}\right)\right| + 2\mathbb{P}_*(\boldsymbol{A}_n^{*c}) + 2\mathbb{P}(\boldsymbol{A}_n^c) > k \cdot n^{-1/2}\right] \\ &\leq \mathbb{P}\left[\sup_{\boldsymbol{B} \in \mathcal{C}_q} \left|\mathbb{P}^*\left(\boldsymbol{T}_{3n}^* + \dot{\boldsymbol{b}}_n \in \boldsymbol{B}\right) - \mathbb{P}\left(\boldsymbol{T}_{3n} + \boldsymbol{b}_n \in \boldsymbol{B}\right)\right| + 2\mathbb{P}_*(\boldsymbol{A}_n^{*c}) > (k/2) \cdot n^{-1/2}\right] \end{aligned}$$

$$\leq \mathbb{P}\Big[\Big\{\sup_{\boldsymbol{B}\in\mathcal{C}_q}\Big|\mathbb{P}^*\Big(\boldsymbol{T}^*_{3n}+\boldsymbol{b}_n\in\boldsymbol{B}\Big)-\mathbb{P}\Big(\boldsymbol{T}_{3n}+\boldsymbol{b}_n\in\boldsymbol{B}\Big)\Big|>(k/4)\cdot n^{-1/2}\Big\}\Big]+o(n^{-1/2})$$

$$=\mathbb{P}\Big[\Big\{\sup_{\boldsymbol{B}\in\mathcal{C}_q}\Big|\mathbb{P}^*\Big(\boldsymbol{T}^*_{3n}\in\boldsymbol{B}\Big)-\mathbb{P}\Big(\boldsymbol{T}_{3n}\in\boldsymbol{B}\Big)\Big|>(k/4)\cdot n^{-1/2}\Big\}\Big]+o(n^{-1/2}).$$

Fourth inequality follows from the fact that $\mathbb{P}_*\Big(\boldsymbol{A}^{*c}_n\Big)=o(n^{-1/2})$ on the set $\boldsymbol{A}_{3n}$. Therefore, to prove the result corresponding to RB of Theorem 2, it is enough to show that

$$\sup_{\boldsymbol{B}\in\mathcal{C}_q}\Big|\mathbb{P}^*\Big(\boldsymbol{T}^*_{3n}\in\boldsymbol{B}\Big)-\mathbb{P}\Big(\boldsymbol{T}_{3n}\in\boldsymbol{B}\Big)\Big|=O_p(n^{-1/2}).$$

Similar arguments will also ensure that the following is enough to establish the result corresponding to PB:

$$\sup_{\boldsymbol{B}\in\mathcal{C}_q}\Big|\mathbb{P}^*\Big(\boldsymbol{T}^{**}_{3n}\in\boldsymbol{B}\Big)-\mathbb{P}\Big(\boldsymbol{T}_{3n}\in\boldsymbol{B}\Big)\Big|=O_p(n^{-1/2}).$$

Again due to Lemma 6 of Das, Chatterjee and Lahiri (2026) we have

$$\sup_{\boldsymbol{B}\in\mathcal{C}_q}\Big|\mathbb{P}\big(\boldsymbol{T}_{3n}\in\boldsymbol{B}\big)-\Phi\big(\boldsymbol{B};\sigma^2\boldsymbol{\Sigma}_n\big)\Big|=O\big(n^{-1/2}\big), \tag{9.2}$$

$$\sup_{\boldsymbol{B}\in\mathcal{C}_q}\Big|\mathbb{P}^*\big(\boldsymbol{T}^*_{3n}\in\boldsymbol{B}\big)-\Phi\big(\boldsymbol{B};\hat{\sigma}_n^2\boldsymbol{\Sigma}_n\big)\Big|=O_p\big(n^{-1/2}\big)\ \text{ and }\ \sup_{\boldsymbol{B}\in\mathcal{C}_q}\Big|\mathbb{P}^*\big(\boldsymbol{T}^{**}_{3n}\in\boldsymbol{B}\big)-\Phi\big(\boldsymbol{B};\tilde{\boldsymbol{\Sigma}}_n\big)\Big|=O_p\big(n^{-1/2}\big),$$

where $\boldsymbol{\Sigma}_n=n^{-1}\sum_{i=1}^n\boldsymbol{\xi}_i^{(1)}\boldsymbol{\xi}_i^{(1)\prime}=\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{D}_n^{(1)}$, $\hat{\sigma}_n^2=n^{-1}\sum_{i=1}^n(\hat{\epsilon}_i-\bar{\epsilon}_n)^2$ and $\tilde{\boldsymbol{\Sigma}}_n=n^{-1}\sum_{i=1}^n\boldsymbol{\xi}_i^{(1)}\boldsymbol{\xi}_i^{(1)\prime}\hat{\epsilon}_i^2$. Clearly we are done from (9.2) by noting that $|\hat{\sigma}_n^2-\sigma^2|=O_p(n^{-1/2})$ and $\|\tilde{\boldsymbol{\Sigma}}_n-\sigma^2\boldsymbol{\Sigma}_n\|=O_p(n^{-1/2})$ (cf. Lemma 5 of Das, Chatterjee and Lahiri (2026)). □

Proof of Theorem 6. 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 (2026), to conclude the desired rates $\mathbb{P}\big(\boldsymbol{A}_n\big)=1-o(n^{-2})$ and $\mathbb{P}^*\big(\boldsymbol{A}^*_n\big)=1-o_p\big(n^{-2}\big)$. As soon as we have these bounds, there is no role of the original dimension p. All the analysis will be in the reduced dimension p_0 which has same growth rate irrespective of whether p grows polynomially or exponentially. Note that for the RB, the pivotal pair is $\big(\breve{\boldsymbol{R}}_n,\breve{\boldsymbol{R}}^*_n\big)$. As in the proof of Theorem 3, for sufficiently large n we have,

$$\breve{\boldsymbol{R}}_n=\hat{\sigma}_n^{-1}n^{-1/2}\sum_{i=1}^n\boldsymbol{D}_n^{(1)}C_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i\ \text{ and }\ \breve{\boldsymbol{R}}^*_n=\sigma_n^{*-1}n^{-1/2}\sum_{i=1}^n\boldsymbol{D}_n^{(1)}C_{11,n}^{-1}\boldsymbol{x}_i^{(1)}\epsilon_i^*,$$

respectively on the sets $\boldsymbol{A}_n$ and $\boldsymbol{A}^*_n$. Since $\mathbb{P}\big(\boldsymbol{A}_n\big)=1-o(n^{-2})$ and $\mathbb{P}^*\big(\boldsymbol{A}^*_n\big)=1-o_p\big(n^{-2}\big)$ and the desired error rate is $O(n^{-2})$, without loss of generality we can assume that $\boldsymbol{A}_{0n}$ and $\boldsymbol{A}^*_{0n}$ both have probability equal to 1. Again note that Cramer's condition holds for $(\epsilon_1,\epsilon_1^2)'$ (cf. assumption (A.4)(ii)) and conditional Cramer's condition holds for (r_1^*,r_1^{*2}) (cf. Lemma 7 of Das, Chatterjee and Lahiri (2026)). Therefore the set up for RB fits with the regression set up considered in Hall (1988). Hence the arguments through (3.4) to (3.7) of Hall (1988) (with error $o(n^{-2})$ instead of $O(n^{-5/2})$) can be followed to prove Theorem 6. Without loss of generality assume that the asymptotic variances of the pivots $\breve{\boldsymbol{R}}_n$ and $\breve{\boldsymbol{R}}_n^*$ are all equal to 1. This will not allow us to lose generality since asymptotic variances of $\breve{\boldsymbol{R}}_n$ and $\breve{\boldsymbol{R}}_n^*$ are actually bounded below due to the condition (A.2)(iii). To that end, by Edgeworth expansion theory we have

$$\mathbb{P}\big(\breve{\boldsymbol{R}}_n \le x\big) = \Phi(x) + n^{-1/2} q_1^r(x)\phi(x) + n^{-1} q_2^r(x)\phi(x) + n^{-3/2} q_3^r(x)\phi(x) + O(n^{-2})$$

uniformly in x. Here q_i^r is even or odd polynomial if i is odd or even. Also the coefficients of q_i^r depends on the moments of $(\epsilon_1, \epsilon_1^2)$ of order $i+2$ or less, $i = 1, 2, 3$. Therefore,

$$\mathbb{P}\big(|\breve{\boldsymbol{R}}_n| \le x\big) = 2\Phi(x) - 1 + 2n^{-1} q_2^r(x)\phi(x) + O(n^{-2})$$

uniformly in x. If $\Phi(z_\alpha) = 1 - \alpha/2$, then inverting the above expression we have $h_{n,\alpha}^r = z_\alpha - n^{-1} q_2^r(z_\alpha) + O(n^{-2})$. We can find the Edgeworth expansion with error $O_p(n^{-2})$ for $\breve{\boldsymbol{R}}_n^*$ by using Lemma 4 and Lemma 7 (cf. Das, Chatterjee and Lahiri (2026)) which will similarly imply $\hat{h}_{n,\alpha}^r = z_\alpha - n^{-1} \hat{q}_2^r(z_\alpha) + O_p(n^{-2})$ for some odd polynomial $\hat{q}_2^r(\cdot)$. Hence $\hat{h}_{n,\alpha}^r = h_{n,\alpha}^r - n^{-3/2} V_{n,\alpha}^r + O_p(n^{-2})$ where $V_{n,\alpha}^r = \sqrt{n}\big\{\hat{q}_2^r(z_\alpha) - q_2^r(z_\alpha)\big\}$. Therefore upon application of delta method for Edgeworth expansions [cf. Section 2.7 of Hall (1992)], we have

$$\mathbb{P}\big(\theta_n \in I_{n,(1-\alpha)}^r\big) = \mathbb{P}\Big(\breve{\boldsymbol{R}}_n + n^{-3/2} V_{n,\alpha}^r \le h_{n,\alpha}^r\Big) - \mathbb{P}\Big(\breve{\boldsymbol{R}}_n - n^{-3/2} V_{n,\alpha}^r \le -h_{n,\alpha}^r\Big) + O(n^{-2}). \tag{9.3}$$

Now following the arguments similar to Hall (1988), it can be established that

$$\mathbb{P}\Big(\breve{\boldsymbol{R}}_n + n^{-3/2} V_{n,\alpha}^r \le x\Big) = \mathbb{P}\Big(\breve{\boldsymbol{R}}_n \le x\Big) + n^{-3/2} s_1^r(x) + O(n^{-2})$$
$$\mathbb{P}\Big(\breve{\boldsymbol{R}}_n - n^{-3/2} V_{n,\alpha}^r \le x\Big) = \mathbb{P}\Big(\breve{\boldsymbol{R}}_n \le x\Big) - n^{-3/2} s_1^r(x) + O(n^{-2}),$$

both uniformly in x, since the quantity $V_{n,\alpha}^r$ is properly centered and scaled. Here $s_1^r(\cdot)$ is an odd polynomial. Hence, for any $\alpha \in (0, 1)$ we have

$$\begin{aligned}
&\Big|\mathbb{P}\big(\theta_n \in I_{n,(1-\alpha)}^r\big) - (1-\alpha)\Big| = \Big|\mathbb{P}\big(|\breve{\boldsymbol{R}}_n| \le \hat{h}_{n,\alpha}^r\big) - \mathbb{P}\big(|\breve{\boldsymbol{R}}_n| \le h_{n,\alpha}^r\big)\Big| \\
&= \Big|\Big[\mathbb{P}\Big(\breve{\boldsymbol{R}}_n + n^{-3/2} V_{n,\alpha}^r \le h_{n,\alpha}^r\Big) - \mathbb{P}\Big(\breve{\boldsymbol{R}}_n - n^{-3/2} V_{n,\alpha}^r \le -h_{n,\alpha}^r\Big)\Big] - \mathbb{P}\Big(|\breve{\boldsymbol{R}}_n| \le h_{n,\alpha}^r\Big)\Big| + O(n^{-2}) \\
&= \Big|\Big[\mathbb{P}\Big(\breve{\boldsymbol{R}}_n \le h_{n,\alpha}^r\Big) + n^{-3/2} s_1^r\big(h_{n,\alpha}^r\big)\Big] - \Big[\mathbb{P}\Big(\breve{\boldsymbol{R}}_n \le -h_{n,\alpha}^r\Big) \\
&\quad - n^{-3/2} s_1^r\big(-h_{n,\alpha}^r\big)\Big] - \mathbb{P}\Big(|\breve{\boldsymbol{R}}_n| \le h_{n,\alpha}^r\Big)\Big| + O(n^{-2}) = O(n^{-2}).
\end{aligned}$$

Therefore we are done. □

Acknowledgments

The authors would like to thank an anonymous referee for some constructive comments that improved the simulation section and overall quality of the presentation.

Funding

The first author was supported in part by the DST fellowship DST/INSPIRE/04/2018/001290. The second author was supported by SERB-MATRICS Grant MTH/2017/000224, from Department of Science and Technology, Government of India. The third author was supported by NSF grant DMS-2006475.

Supplementary Material

Supplement to "Bootstrapping LASSO estimators under variable selection consistency in high dimensions and some higher order refinements" (DOI: 10.3150/26-BEJ1971SUPP; .pdf). The supplementary materials file Das, Chatterjee and Lahiri (2026) contains additional simulation results. The performance of the PB in the heteroscedastic errors setting is studied. Additional tables and figures comparing the coverages obtained by RB and PB based one-sided, 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. Auxiliary Lemmas (numbered Lemma 1 - 8) and detailed proofs of Theorems 3, 7, 8 and 9 are also presented there.

References

Belloni, A. and Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19** 521–547. MR3037163 https://doi.org/10.3150/11-BEJ410

Bhattacharya, R.N. and Ghosh, J.K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451. MR0471142

Camponovo, L. (2015). On the validity of the pairs bootstrap for lasso estimators. *Biometrika* **102** 981–987. MR3431568 https://doi.org/10.1093/biomet/asv039

Chatterjee, A. and Lahiri, S.N. (2010). Asymptotic properties of the residual bootstrap for Lasso estimators. *Proc. Amer. Math. Soc.* **138** 4497–4509. MR2680074 https://doi.org/10.1090/S0002-9939-2010-10474-4

Chatterjee, A. and Lahiri, S.N. (2011). Bootstrapping lasso estimators. *J. Amer. Statist. Assoc.* **106** 608–625. MR2847974 https://doi.org/10.1198/jasa.2011.tm10159

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. MR3113809 https://doi.org/10.1214/13-AOS1106

Choudhury, M. and Das, D. (2025). Does K-fold CV based penalty perform variable selection or does it lead to $n^{1/2}$-consistency in Lasso? arXiv preprint, arXiv:2507.12457.

Das, D., Chatterjee, A. and Lahiri, S.N. (2026). Supplement to "Bootstrapping LASSO estimators under variable selection consistency in high dimensions and some higher order refinements." https://doi.org/10.3150/26-BEJ1971SUPP

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

Das, D. and Lahiri, S.N. (2019a). Distributional consistency of the lasso by perturbation bootstrap. *Biometrika* **106** 957–964. MR4031208 https://doi.org/10.1093/biomet/asz029

Das, D. and Lahiri, S.N. (2019b). Second order correctness of perturbation bootstrap M-estimator of multiple linear regression parameter. *Bernoulli* **25** 654–682. MR3892332 https://doi.org/10.3150/17-bej1001

Dezeure, R., Bühlmann, P., Meier, L. and Meinshausen, N. (2015). High-dimensional inference: Confidence intervals, p-values and R-software hdi. *Statist. Sci.* **30** 533–558. MR3432840 https://doi.org/10.1214/15-STS527

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. With discussion, and a rejoinder by the authors. MR2060166

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581

Friedman, J.H., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22. https://doi.org/10.18637/jss.v033.i01

Friedman, J., Hastie, T., Höfling, H. and Tibshirani, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.* **1** 302–332. MR2415737 https://doi.org/10.1214/07-AOAS131

Fu, W.J. (1998). Penalized regressions: The bridge versus the lasso. *J. Comput. Graph. Statist.* **7** 397–416. MR1646710 https://doi.org/10.2307/1390712

Hall, P. (1988). On symmetric bootstrap confidence intervals. *J. Roy. Statist. Soc. Ser. B, Methodol.* **50** 35–45. MR0954730

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. New York: Springer-Verlag. MR1145237

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909.

Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *Ann. Statist.* **28** 1356–1378. MR1805787

Kuchibhotla, A.K. and Chakrabortty, A. (2022). Moving beyond sub-Gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *Inf. Inference* **iaac012**. https://doi.org/10.1093/imaiai/iaac012

Lahiri, S.N. (2021). Necessary and sufficient conditions for variable selection consistency of the LASSO in high dimensions. *Ann. Statist.* **49** 820–844. MR4255109 https://doi.org/10.1214/20-aos1979

Lee, J.D., Sun, D.L., Sun, Y. and Taylor, J.E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. MR3485948 https://doi.org/10.1214/15-AOS1371

Liu, R.Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16** 1696–1708. https://doi.org/10.1214/aos/1176351062 MR964947

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

Minnier, J., Tian, L. and Cai, T. (2011). A perturbation method for inference on regularized regression estimates. *J. Amer. Statist. Assoc.* **106** 1371–1382. MR2896842 https://doi.org/10.1198/jasa.2011.tm10382

Osborne, M.R., Presnell, B. and Turlach, B.A. (2000). On the LASSO and its dual. *J. Comput. Graph. Statist.* **9** 319–337. MR1822089

Taylor, J. and Tibshirani, R. (2018). Post-selection inference for ℓ_1-penalized likelihood models. *Canad. J. Statist.* **46** 41–61. MR3767165 https://doi.org/10.1002/cjs.11313

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B, Methodol.* **58** 267–288. MR1379242

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

van de Geer, S., Bühlmann, P., Ritov, Y. and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. https://doi.org/10.1214/14-AOS1221

Vladimirova, M., Girard, S., Nguyen, H. and Arbel, J. (2020). Sub-Weibull distributions: Generalizing sub-Gaussian and sub-exponential properties to heavier tailed distributions. *Stat* **9** e318. https://doi.org/10.1002/sta4.318

Wagener, J. and Dette, H. (2013). The adaptive Lasso in high-dimensional sparse heteroscedastic models. *Math. Methods Statist.* **22** 137–154. MR3071959 https://doi.org/10.3103/S106653071302004X

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

Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1295. https://doi.org/10.1214/aos/1176350142

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

Zhang, C.-H. and Zhang, S.S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

Zhao, P. and Yu, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. MR2274449

Zou, H. (2006). The Adaptive Lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469
Zou, H. and Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443 https://doi.org/10.1214/009053607000000802

Received May 2023 and revised January 2026