Abstract

DAS, DEBRAJ. Perturbation Bootstrap in Regression. (Under the supervision of Professor Soumendra Nath Lahiri)

Consider the multiple linear regression model

$$y_i = \boldsymbol{x_i'}\boldsymbol{\beta} + \epsilon_i, \quad i = 1, 2, \ldots, n$$

where $y_1, \ldots, y_n$ are responses, $\epsilon_1, \ldots, \epsilon_n$ are independent errors, $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}$ are design vectors and $\boldsymbol{\beta}$ is the p-dimensional vector of parameters. After introduction of perturbation bootstrap by Jin et al. (2001), the asymptotic properties of this method remains largely unexplored in the context of multiple linear regression. In this dissertation, we analyze the asymptotic properties of perturbation bootstrap method as a distribution approximation method for different estimators of $\boldsymbol{\beta}$.

In the first chapter, we establish second order results of the perturbation bootstrap approximation of the distribution of the M-estimator of $\boldsymbol{\beta}$ when the design vectors are non-random. Second order correctness is important for reducing the approximation error uniformly to $o(n^{-1/2})$ to get better inferences. We show that the classical studentized version of the bootstrapped estimator fails to be second order correct even when the errors are independent and identically distributed (iid). We introduce an innovative modification in the studentized version of the bootstrapped statistic and show that the modified bootstrapped pivot is second order correct for approximating the distribution of the studentized M-estimator. Additionally, we show that the Perturbation Bootstrap continues to be second order correct when the errors ϵ_i's are independent, but may not be identically distributed.

In chapter two, we consider sparsity in the underlying multiple linear regression model and subsequently investigate the asymptotic properties of perturbation bootstrap in case of Lasso. Least Absolute Shrinkage and Selection Operator or Lasso, introduced by Tibshirani (1996), is a popular estimation procedure in multiple linear regression when underlying design has a sparse structure, because of its property that it sets some regression coefficients exactly equal to 0. We develop a perturbation bootstrap method and establish its validity in approximating the distribution of the Lasso in heteroscedastic linear regression, or more generally when the errors are independent, but may not be identically distributed. We consider the underlying covariates, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, to be either random or non-random. We show that the proposed bootstrap method works irrespective of the nature of the covariates, unlike the resample-based bootstrap (residual and pairs bootstrap) of Freedman (1981).

In chapter three, we explore the asymptotic properties of perturbation bootstrap for the Adaptive Lasso (Alasso). Alasso was proposed by Zou (2006) as a modification of the Lasso for the purpose of simultaneous variable selection and estimation of $\boldsymbol{\beta}$. Zou (2006) established that the Alasso estimator is variable-selection consistent as well as asymptotically Normal in the indices corresponding to the nonzero regression coefficients in certain fixed-dimensional settings. In an influential paper, Minnier, Tian and Cai (2011) proposed a perturbation bootstrap method and established its distributional consistency for the Alasso estimator in the fixed-dimensional setting. In this chapter, however, we show that this (naive) perturbation bootstrap fails to achieve second order correctness in approximating the distribution of the Alasso estimator. We propose a modification to the perturbation bootstrap objective function and show that a suitably studentized version of our modified perturbation bootstrap Alasso

estimator achieves second-order correctness even when the dimension p is allowed to grow to infinity with the sample size n. We give simulation studies demonstrating good finite sample properties of our modified perturbation bootstrap method as well as an illustration of our method on a real data set.

Perturbation Bootstrap in Regression

by
Debraj Das

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2017

APPROVED BY:

Subhashis Ghoshal

Kazufumi Ito

Soumendra Nath Lahiri
Chair of Advisory Committee

Howard Bondell

Dedication

In memory of my grandmother Smt. Madhabi Das.

Biography

Debraj was born in Uttarpara, a small town in the eastern part of India. He is the only child of Mr. Arup Das and Mrs. Lina Das. Besides his parents, his grandmother, the late Smt. Madhabi Das, and his aunt Mrs. Soma Talapatra and uncle Mr. Amit Talapatra were significant part of his childhood. Debraj is very close to his brother Diptarka. They grew up together. He met with his love Poulami in the first year of M.Stat. in Indian Statistical Institute, Delhi. They got married in December, 2016.

Debraj went to Uttarpara Govt. High School and completed 10+2 level there. After that he went to Ramakrishna Mission Residential College, Narendrapur, to pursue Bachelors degree in Statistics. He completed his Master's degree from Indian Statistical Institute. Subsequently, he joined Department of Statistics, North Carolina State University, to pursue doctoral degree in Statistics.

Acknowledgments

First of all, I want to thank all of my family members for their love and encouragement throughout my life. Specially my parents, Arup Das and Lina Das, are the driving force behind each and every accomplishment of mine. Not to forget my charming wife Poulami, without her love and enthusiasm this dissertation would not be possible. Having said that, there is another person from past to whom I owe a lot for shaping my character and to turn me the person that I am today. She is my grandmother, the late Smt. Madhabi Das.

I want to thank my advisor Dr. Soumendra Nath Lahiri for his care, help and encouragement during my stay at NC State as a graduate student. His inputs and thoughts are invaluable throughout my dissertation research. I am particularly grateful to him for his enthusiastic clarifications of my doubts (sometimes silly) at the beginning of my research. His approach towards a potential research problem is a great learning to me. I have always tried to follow his way of thinking and continue to do so in future. I want to thank professors Subhashis Ghoshal, Howard Bondell and Kazufumi Ito for agreeing to serve as members of my committee and for their guidance. I want to thank specially professor Seth Sullivant for accepting the role of Graduate School Representative (GSR) in my final oral examination. Thanks are also due to Professor Karl Gregory of University of South Carolina for designing the simulation study presented in Chapter 3 and for implementing our proposed bootstrap method on a real dataset.

I am grateful to all the teachers that I have come across till now. I want to thank professors Arindam Chatterjee and Tapas Samanta from Indian Statistical institute and all the professors of Ramakrishna Mission Residential College, Narendrapur, specially Parthosarathi Chakrabarti and Tulsidas Mukhopadhyay, for giving me inspirations to pursue doctoral degree in Statistics. Another teacher, without whom I could not have come this far, is Mr. Abhijit Sengupta. He was inspirational during my school days. He was the one who backed me to pursue Bachelor's degree in Statistics.

I was fortunate to get in touch with my senior Rudrodip Mazumdar from Dept. of Nuclear Engineering in NC State. The helps and advices that I received from him, regarding my research and my general life, were priceless. Besides that, he is the one with whom I wrote my first research paper. I want to thank him for showing faith in me in writing a paper. There are two other persons to whom I owe a lot during my stay in Raleigh. They are Shuva Gupta and Suman Chakraborty. When I was new to NC State, Shuva da was the one who made me feel comfortable. I have seen very few people in my life who is always eager to help. Shuva da is certainly one of them. Then there is my friend Suman who always make me astound with his originality in mathematical thinking. I cannot describe how much I owe to him for his support during my stay in Raleigh.

I also want to thank Rajarshi da, Saswata da, Sayantan da, Prithwish da, Tirtho da, Abhra da, Rimli di, Sujatro da, APM da, Byana da, Priyam (Ala), Mithyl, Arkaprava (Gol), Indranil, Arnab (Hazra), Moumita, Sohini, Arkopal (Chapcho), Sayan, Arunava, Arnab (Chak), Abhishek for their love and support during my stay at NC State. Not to forget couple of my friends from Bachelor's and master's degree, Abhishek (Nasu), Noirrit, Kiranmoy, Rahul, Debmalya, Adhi, Angshuman (Mao), Jayabrata (Joga), for their friendship and love for me. Last but not the least, I am grateful to my childhood friend Kishalay for the friendship that is now standing in its eighteenth year.

Lastly, I want to thank Department of Statistics, North Carolina State University, for their constant support during my stay here as a graduate student.

Table of Contents

List of Tables

List of Figures

Chapter 1

Perturbation Bootstrap in Regression M-estimation

1.1 Introduction

Consider the multiple linear regression model :

$$y_i = \boldsymbol{x_i'\beta} + \epsilon_i, \quad i = 1, 2, \ldots, n \tag{1.1.1}$$

where $y_1, \ldots, y_n$ are responses, $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed (iid) random variables with common distribution F (say), $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}$ are known non random design vectors and $\boldsymbol{\beta}$ is the p-dimensional vector of parameters.

Suppose $\bar{\boldsymbol{\beta}}_n$ is the M-estimator of $\boldsymbol{\beta}$ corresponding to the objective function $\Lambda(\cdot)$ i.e. $\bar{\boldsymbol{\beta}}_{\boldsymbol{n}} = \arg\min_{\boldsymbol{t}} \sum_{i=1}^{n} \Lambda(y_i - \boldsymbol{x_i'}\boldsymbol{t})$. Now if $\psi(\cdot)$ is the derivative of $\Lambda(\cdot)$, then $\bar{\boldsymbol{\beta}}_{\boldsymbol{n}}$ is the M-estimator corresponding to the score function $\psi(\cdot)$ and is defined as the solution of the vector equation

$$\sum_{i=1}^{n} \boldsymbol{x_i}\psi(y_i - \boldsymbol{x_i'\beta}) = \boldsymbol{0}.$$

It is known [cf. Huber(1981)] that under some conditions on the objective function, design vectors and error distribution F; $(\bar{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ with proper scaling has an asymptotically normal distribution with mean $\mathbf{0}$ and dispersion matrix $\sigma^2 \boldsymbol{I_p}$ where $\sigma^2 = \boldsymbol{E}\psi^2(\epsilon_1)/\boldsymbol{E}^2\psi'(\epsilon_1)$.

After introduction of bootstrap by Efron in 1979 as a resampling technique, it has been widely used as a distributional approximation method. Resampling from the naive empirical distribution of the centered residuals in a regression setup, called residual bootstrap, was introduced by Freedman (1981). Freedman (1981) and Bickel and Freedman (1981b) had shown that given data, the conditional distribution of $\sqrt{n}(\boldsymbol{\beta}_n^* - \bar{\boldsymbol{\beta}}_n)$ converges to the same normal distribution as the distribution of $\sqrt{n}(\bar{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ when $\bar{\boldsymbol{\beta}}_n$ is the usual least square estimator of $\boldsymbol{\beta}$, that is, when $\Lambda(x) = x^2$. It implies that the residual bootstrap approximation to the exact distribution of the least square estimator is first order correct as in the case of normal approximation. The advantage of the residual bootstrap approximation over normal approximation for the distribution of linear contrasts of least square estimator for general p was first shown by Navidi (1989) by investigating the underlying Edgeworth Expansion (EE); although heuristics behind the same was given by Liu (1988) in restricted case $p = 1$. Consequently, EE for the general M-estimator of $\boldsymbol{\beta}$ was obtained by Lahiri (1989b) when $p = 1$; whereas the same for the multivariate least square estimator was found by Qumsiyeh (1990a). EE of standardized and studentized versions of the general M-estimator in multiple linear regression setup was first obtained by Lahiri (1992). Lahiri (1992) also established the second order results for residual bootstrap in regression M-estimation.

A natural generalization of sampling from the naive empirical distribution is to sample from a weighted empirical distribution to obtain the bootstrap sample residuals. Broadly, the resulting bootstrap procedure is called the weighted or generalized bootstrap. It was introduced by Mason and Newton (1992) for bootstrapping mean of a collection of iid random variables. Mason and Newton (1992) considered exchangeable weights and established its consistency. Lahiri (1992) established second order correctness of generalized bootstrap in approximating the distribution of the M-estimator for the model (1.1.1) when the weights are chosen in a particular fashion depending on the design vectors. Wellner and Zhan (1996) proved the consistency of infinite dimensional generalized bootstrapped M-estimators. Consequently, Chatterjee and Bose (2005) established distributional consistency of generalized bootstrap in estimating equations and showed that generalized bootstrap can be used in order to estimate the asymptotic variance of the original estimator. Chatterjee and Bose (2005) also mentioned the bias correction essential for achieving second order correctness. An important special case of generalized bootstrap is the bayesian bootstrap of Rubin (1981). Rao and Zhao (1992) showed that the distribution function of M-estimator for the model (1.1.1) can be approximated consistently by bayesian bootstrap. See the monograph of Barbe and Bertail (2012) for an extensive study of generalized bootstrap.

A close relative to the generalized bootstrap procedure is the wild bootstrap. It was introduced by Wu (1986) in multiple linear regression model (1.1.1) with errors ϵ_i's being heteroscedastic. Beran (1986) justified wild bootstrap method by pointing out that the distribution of the least square estimator can be approximated consistently by the wild bootstrap approximation. Second order results of wild bootstrap in heteroscedastic regression model was first established by Liu (1988) when $p = 1$. Liu (1988) also showed that usual residual bootstrap is not capable of approximating the distribution of the least square estimator upto second order in heteroscedastic setup and described a modification in resampling procedure which can establish second order correctness. For general p, the heuristics behind achieving second order correctness by wild bootstrap in homoscedastic least square regression were discussed in Mammen (1993). Recently, Kline and Santos (2011) developed a score based bootstrap method depending on wild bootstrap in M-estimation for the homoscedastic model (1.1.1) and established consistency of the procedure for Wald and Lagrange Multiplier type tests for a class of M-estimators under misspecification and clustering of data.

A novel bootstrap technique, called the perturbation bootstrap was introduced by Jin, Ying, and Wei (2001) as a resampling procedure where the objective function having a U-process structure was perturbed by non-negative random quantities. Jin, Ying, and Wei (2001) showed that in standardized setup, the conditional distribution of the perturbation resampling estimator given the data and the distribution of the original estimator have the same limiting distribution which means this resampling method is first order correct without studentization. In a recent work, Minnier, Tian, and Cai (2011) also applied this perturbation resampling method in penalized regression setup such as Adaptive Lasso, SCAD, l_q penalty and showed that the standardized perturbed penalized estimator is first order correct. But, second order properties of this new bootstrap method have remained largely unexplored in the context of multiple linear regression. In this current chapter, the perturbation bootstrap approximation is shown to be S.O.C. for the distribution of studentized M-estimator for the regression model (1.1.1). An extension to the case of independent and non-iid errors is also established, showing the robustness of perturbation bootstrap towards the presence of heteroscedasticity. Therefore, besides the existing bootstrap methods, the perturbation bootstrap method can also be used in regression M-estimation for making inferences regarding the regression parameters and higher order accuracy can be achieved than the normal approximation.

A classical way of studentization in bootstrap setup, in case of regression M-estimator and for iid errors, is to consider the studentization factor to be $\sigma_n^* = s_n^* \tau_n^{*-1}$, $\tau_n^* = n^{-1}\sum_{i=1}^{n}\psi'(\epsilon_i^*)$, $s_n^{*2} = n^{-1}\sum_{i=1}^{n}\psi^2(\epsilon_i^*)$ where $\epsilon_i^* = y_i - \boldsymbol{x_i'\beta_n^*}$, $i \in \{1,\ldots,n\}$, with $\boldsymbol{\beta_n^*}$ being the perturbation bootstrapped estimator of $\boldsymbol{\beta}$, defined in Section 1.2. Although the residual bootstrapped estimator is S.O.C. after straight-forward studentization, the same pivot fails to be S.O.C. in the case of perturbation bootstrap. Two important special cases are considered as examples in this respect. The reason behind this failure is that although the bootstrap residuals are sufficient in capturing the variability of the bootstrapped estimator in residual bootstrap, it is not enough in the case of perturbation resampling. Modifications have been proposed as remedies and are shown to be S.O.C. The modifications are based on the novel idea that the variability of the random perturbing quantities G_i^* $(1 \leq i \leq n)$ along with the bootstrap residuals are required to capture the variability of the perturbation bootstrapped estimator; whereas individually they are not sufficient. For technical details, see Section 1.4.3 and Section 1.5.1.

With a view to establish second order correctness, we start with the standardized setup and then proceed to studentization. First, we find a two-term EE of the conditional density of a suitable stochastic approximation of the concerned bootstrapped pivot and then we show that it is the required two-term EE corresponding to the bootstrapped pivot. The result then follows by comparing the EE of the bootstrapped pivot with that of underlying original pivot. The techniques that are to be used in finding EE have been demonstrated and discussed in Bhattacharya and Ghosh (1978), Bhattacharya and Rao (1986), Navidi (1989) and Lahiri (1992).

A significant volume of chapter is available in bootstrapping M-estimators. We will conclude this section by briefly reviewing the literature. bootstrapping M-estimators in linear model has been studied by Navidi(1989), Lahiri (1992, 1996), Rao and Zhao (1992), Qumsiyeh (1994), Karabulut and Lahiri (1997), Jin, Ying and Wei (2001), Hu (2001), El Bantli (2004) among others. And in the applications other than linear model, bootstrapping in M-estimation and its subclasses has been investigated by Arcones and Giné (1992), Lahiri (1994), Wellner and Zhan (1996), Allen and Datta (1999), Hu and Kalbfleisch (2000), Hlavka (2003), Wang and Zhou (2004), Chatterjee and Bose (2005), Ma and Kosorok (2005), Lahiri and Zhu (2006), Cheng and Huang (2010), Feng et. al. (2011), Lee (2012), Cheng (2015), among others.

The rest of the chapter is organized as follows. Perturbation bootstrap is described briefly in Section 1.2. Section 1.3 states the assumptions and motivations behind considering those assumptions. Main results for iid case, along with the modification in bootstrap studentization, are stated in Section 1.4. An extension to the case of independent and non-iid errors is proposed in Section 1.5. Proofs are given in Section 1.6. Section 1.7 states concluding remarks.

1.2 Description of Perturbation Bootstrap

In the perturbation bootstrap, The objective function $\Lambda(\cdot)$ has been perturbed several times by a non-negative random quantity to get a bootstrapped estimate of $\boldsymbol{\beta}$. It has nothing to do with residuals in resampling stage, unlike the residual and weighted bootstrap. More precisely, the perturbation bootstrap estimator $\boldsymbol{\beta}_n^*$ is defined as

$$\boldsymbol{\beta}_n^* = \arg\min_{\boldsymbol{t}} \sum_{i=1}^{n} \Lambda(y_i - \boldsymbol{x}_i'\boldsymbol{t})G_i^*$$

or in terms of the score function $\psi(\cdot)$, as the solution of the vector equation

$$\sum_{i=1}^{n} \boldsymbol{x}_i \psi(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})G_i^* = \mathbf{0} \tag{1.2.1}$$

where $G_i^*, i \in \{1, \ldots, n\}$ are non-negative and non-degenerate completely known random variables, considered as perturbation quantities. Note that, if μ_{G^*} is the mean of G_1^*, then $\bar{\boldsymbol{\beta}}_n$ is the solution of $\mathbf{E}\Big(\sum_{i=1}^{n} \boldsymbol{x}_i \psi(\bar{\epsilon}_i)G_i^* | \epsilon_1, \ldots, \epsilon_n \Big) = \sum_{i=1}^{n} \boldsymbol{x}_i \psi(\bar{\epsilon}_i)\mu_{G^*} = 0$ where $\bar{\epsilon}_i = y_i - \boldsymbol{x}_i'\bar{\boldsymbol{\beta}}_n, i \in \{1, \ldots, n\}$, are the residuals corresponding to the M-estimator $\bar{\boldsymbol{\beta}}_n$. This observation will be helpful in finding a suitable stochastic approximation in bootstrap regime. For details, see Section 1.6.

The central idea of the perturbation bootstrap is to draw a relatively large collection of iid random samples $\{(G_1^{*b}, \ldots, G_n^{*b}) : b = 1, \ldots, B\}$ from the distribution of G_1^* and then to find the conditional empirical distribution of $\sqrt{n}(\boldsymbol{\beta}_n^* - \bar{\boldsymbol{\beta}}_n)$ given data $y_i : i = 1, \ldots, n$, by solving

$$\sum_{i=1}^{n} \boldsymbol{x}_i \psi(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})G_i^{*b} = \mathbf{0}$$

for each $b \in \{1, \ldots, B\}$; to approximate the distribution of $\sqrt{n}(\bar{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ asymptotically. As a result the bootstrapped distribution may be used as an approximation to the original distribution, just like the normal approximation, in constructing confidence intervals and testing of hypotheses regarding $\boldsymbol{\beta}$.

Now, in the perturbation bootstrap M-estimation, G_i^*'s can be thought of as weights corresponding to the ith data point $(\boldsymbol{x}_i, y_i)$. To make it easier to understand, consider the least square setup i.e. $\Lambda(x) = x^2$. In this case $\boldsymbol{\beta_n^*}$ takes the form

$$\boldsymbol{\beta_n^*} = \Big(\sum_{i=1}^{n} \boldsymbol{x_i x_i'} G_i^*\Big)^{-1}\Big(\sum_{i=1}^{n} \boldsymbol{x_i} y_i G_i^*\Big) \tag{1.2.2}$$

indicating that the perturbing quantities G_i^*'s can be thought of as weights.

Remark 1.2.1 *Consider the least square estimator $\boldsymbol{\hat{\beta}_n}$. Then keeping the asymptotic properties fixed, the perturbation bootstrap version $\boldsymbol{\hat{\beta}_{1n}^*}$ of $\boldsymbol{\hat{\beta}_n}$ can be defined alternatively as the solution of*

$$\sum_{i=1}^{n} \boldsymbol{x_i}(y_i - \boldsymbol{x_i'\beta})\big(G_i^* - \mu_{G^*}\big) + \sum_{i=1}^{n} \boldsymbol{x_i x_i'}(\boldsymbol{\hat{\beta}_n} - \boldsymbol{\beta})\big(2\mu_{G^*} - G_i^*\big) = \mathbf{0}$$

which in turn implies that $\boldsymbol{\hat{\beta}_{1n}^*}$ is the solution of

$$\sum_{i=1}^{n} \boldsymbol{x_i}(z_i^* - \boldsymbol{x_i'\beta}) = \mathbf{0} \tag{1.2.3}$$

where $z_i^* = \boldsymbol{x_i'\hat{\beta}_n} + \hat{\epsilon}_i[\mu_{G^*}^{-1}(G_i^* - \mu_{G^*})]$, $\hat{\epsilon}_i = y_i - \boldsymbol{x_i'\hat{\beta}_n}$, $i \in \{1,\ldots,n\}$. On the other hand, the simple wild bootstrap version $\boldsymbol{\hat{\beta}_{2n}^*}$ of $\boldsymbol{\hat{\beta}_n}$ is defined as the solution of

$$\sum_{i=1}^{n} \boldsymbol{x_i}(y_i^* - \boldsymbol{x_i'\beta}) = \mathbf{0} \tag{1.2.4}$$

where $y_i^* = \boldsymbol{x_i'\hat{\beta}_n} + \hat{\epsilon}_i t_i$, $i \in \{1,\ldots,n\}$. $\{t_1,\ldots,t_n\}$ is a set of iid random variables independent of $\{\epsilon_1,\ldots,\epsilon_n\}$ with $\mathbf{E}t_1 = 0$, $\mathbf{Var}(t_1) = 1$. Additionally, one needs $\mathbf{E}(t_1^3) = 1$ for establishing second order correctness of wild bootstrap approximation [cf. Liu (1988), Mammen (1993)]. Now Looking at (1.2.3) and (1.2.4) and in view of assumption (A.5)(ii), it can be said that the perturbation bootstrap coincides with the wild bootstrap in least square setup. Therefore one can view perturbation bootstrap as a generalization of the wild bootstrap in regression M-estimation.

Remark 1.2.2 *There is a basic difference between perturbation bootstrap and weighted bootstrap with respect to the construction of the bootstrapped estimator. Whereas in the perturbation bootstrap, the bootstrapped estimator is defined through the non-negative and non-degenerate random perturbations of the objective function; in weighted bootstrap, the bootstrapped estimator is defined through bootstrap samples drawn from a weighted empirical distribution. See for example the construction of the weighted bootstrapped estimator corresponding to Theorem 2.3 of Lahiri (1992) and compare it with our construction as stated in Section 1.2. However, one can think the perturbation bootstrap defined in Section 1.2 as the weighted bootstrap version of some statistical functional if the design vectors are random. Suppose,* $\{(x_1, y_1) \ldots, (x_n, y_n)\}$ *are iid with underlying probability measure* $\mathbf{Q}$*. Then one can think*

$$\boldsymbol{\beta} = T(\mathbf{Q}) = \arg\min_{\boldsymbol{t}} \mathbf{E}\Big[\Lambda(y_i - \boldsymbol{x_i'}\boldsymbol{t})\Big]$$

for some statistical functional $T(\cdot)$*. Define the empirical measures* $\mathbf{Q}_n = n^{-1}\sum_{i=1}^{n} \mathbb{1}(x_i, y_i)$ *and* $\mathbf{Q}_{n,G^*} = n^{-1}\sum_{i=1}^{n} \mathbb{1}(x_i, y_i)G_i^*$ *where* $\mathbb{1}(\cdot)$ *is the indicator function. Then we have* $\bar{\boldsymbol{\beta}}_n = T(\mathbf{Q}_n)$ *and* $\boldsymbol{\beta}_n^* = T(\mathbf{Q}_{n,G^*})$*. This is the general setup of Barbe and Bertail (2012). Second order results are available only in standardized setup for differentiable statistical functionals, see Section 3.1 of Barbe and Bertail (2012). In particular, second order correctness of weighted bootstrap of standardized mean of iid random variables was established by Haeusler et. al. (1991). On the other hand, we have assumed the design vectors to be non-random, implying that our setup does not quite fit in the general statistical functional setup of Barbe and Bertail (2012); although Theorem 1.4.3 continue to hold when the design is random. Our main motivation is to explore second order results in studentized setup which is the common setup in practice. We have also extended second order correctness of perturbation bootstrap when error are independent, but may not be identically distributed. Hence the results in this* *chapter significantly extends the results available in the literature of second order correctness by bootstrap in regression M-estimation.*

1.3 Assumptions

Suppose, $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \ldots, x_{ip})'$. Define, $\boldsymbol{D_n} \equiv \boldsymbol{D} = (\sum_{i=1}^{n} \boldsymbol{x_i x_i'})^{1/2}$, $\boldsymbol{A_n} = n^{-1}\boldsymbol{D^2}$, $\boldsymbol{d_i} = \boldsymbol{D^{-1} x_i}$, $1 \leq i \leq n$ and $q = \dfrac{p(p+1)}{2}$. Also define, $q \times 1$ vector $\boldsymbol{z_i} = (x_{i1}^2, x_{i1}x_{i2}, \ldots, x_{i1}x_{ip}$ $, x_{i2}^2, x_{i2}x_{i3}, \ldots, x_{i2}x_{ip}, \ldots, x_{ip}^2)'$. Note that for any constants $a_i, \ldots, a_n \in \mathscr{R}$, $\sum_{i=1}^{n} a_i \boldsymbol{z_i} = \boldsymbol{0}$ which implies and is implied by $\sum_{i=1}^{n} a_i \boldsymbol{x_i x_i'} = \boldsymbol{0}$. Hence, $\{\boldsymbol{z_1}, \ldots, \boldsymbol{z_n}\}$ are linearly independent if and only if $\{\boldsymbol{x_i x_i'} : 1 \leq i \leq n\}$ are linearly independent. Therefore, $r_n =$ the rank of $\sum_{i=1}^{n} \boldsymbol{z_i z_i'}$ is nondecreasing in n. So, if $r = \max\{r_n : n \geq 1\}$ then without loss of generality (w.l.g.), we can assume that $r_n = r$ for all $n \geq q$. Consider canonical decomposition of $\sum_{i=1}^{n} \boldsymbol{z_i z_i'}$ as

$$\boldsymbol{L}(\sum_{i=1}^{n} \boldsymbol{z_i z_i'})\boldsymbol{L'} = \begin{bmatrix} \boldsymbol{I_r} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

where $\boldsymbol{L}$ is a $q \times q$ non-singular matrix. Partition $\boldsymbol{L}$ as $\boldsymbol{L'} = [\boldsymbol{L_1'} \quad \boldsymbol{L_2'}]$, where $\boldsymbol{L_1}$ is of order $r \times q$. Define $r \times 1$ vector $\boldsymbol{\tilde{z}_i}$ by

$$\boldsymbol{\tilde{z}_i} = \boldsymbol{L_1 z_i}, \quad 1 \leq i \leq n$$

Note that $\sum_{i=1}^{n} \boldsymbol{\tilde{z}_i \tilde{z}_i'} = \boldsymbol{L_1}(\sum_{i=1}^{n} \boldsymbol{z_i z_i'})\boldsymbol{L_1'} = \boldsymbol{I_r}$. Suppose, $\boldsymbol{v_i} = (\boldsymbol{x_i'}\psi(\epsilon_1), \boldsymbol{z_i'}\psi'(\epsilon_1))'$. $\boldsymbol{\check{z}_i} = (\boldsymbol{z_i'}, n^{-1})'$.

Let, $\boldsymbol{\Phi_V}$ denotes the normal distribution with mean $\boldsymbol{0}$ and dispersion matrix $\boldsymbol{V}$ and $\boldsymbol{\phi_V}$ is the density of $\boldsymbol{\Phi_V}$. Write $\boldsymbol{\Phi_V} = \boldsymbol{\Phi}$ and $\boldsymbol{\phi_V} = \boldsymbol{\phi}$ when $\boldsymbol{V}$ is the identity matrix. h', h'' denote respectively first and second derivatives of real valued function

h that is twice differentiable. Also $||.||$ denotes euclidean norm.For any set $B \in \mathscr{R}^p$ and $\epsilon > 0$, δB denotes the boundary of B, $|B|$ denotes the cardinality of B and $B^\epsilon = \{\boldsymbol{x} : \boldsymbol{x} \in \mathscr{R}^p$ and $d(\boldsymbol{x}, B) < \epsilon\}$ where $d(\boldsymbol{x}, B) = \inf\{||\boldsymbol{x} - \boldsymbol{y}|| : \boldsymbol{y} \in B\}$. For a function $f : \mathscr{R}^l \rightarrow \mathscr{R}$ and a non-negative integral vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_l)'$, $D^{\boldsymbol{\alpha}} f = D_1^{\alpha_1} \ldots D_l^{\alpha_l} f$, where $D_j^{\alpha_j} f$ denotes α_j times partial derivative of f with respect to the jth component of its argument, $1 \leq j \leq l$. Also assume that $(\boldsymbol{e_1}, \ldots, \boldsymbol{e_p})'$ is the standard basis of $\mathscr{R}^p$. Let, $\mathbf{P}_*$ and $\mathbf{E}_*$ respectively denote conditional bootstrap probability and conditional expectation of G_1^* given data. The class of sets $\mathscr{B}$ denotes the collection of borel subsets of $\mathscr{R}^p$ satisfying

$$\sup_{B \in \mathscr{B}} \boldsymbol{\Phi}((\delta B)^\epsilon) = O(\epsilon) \quad as\ \epsilon \downarrow 0 \tag{1.3.1}$$

Next we state the assumptions:

(A.1) $\psi(\cdot)$ is twice differentiable and $\psi''(\cdot)$ satisfies a Lipschitz condition of order α for some $0 < 2\alpha \leq 1$.

(A.2) (i) $\boldsymbol{A_n} \rightarrow \boldsymbol{A_1}$ as $n \rightarrow \infty$ for some positive definite matrix $\boldsymbol{A_1}$.

(ii) $\mathbf{E}(n^{-1} \sum_{i=1}^n \boldsymbol{v_i v_i'}) \rightarrow \boldsymbol{A_2}$ as $n \rightarrow \infty$ for some non-singular matrix $\boldsymbol{A_2}$, where expectation is with respect to F.

(ii)′ $\mathbf{E}(n^{-1} \sum_{i=1}^n \boldsymbol{\tilde{v}_i \tilde{v}_i'}) \rightarrow \boldsymbol{A_3}$ as $n \rightarrow \infty$ for some non-singular matrix $\boldsymbol{A_3}$ where $\boldsymbol{\tilde{v}_i}$ is defined as same way as $\boldsymbol{v_i}$ with $\boldsymbol{z_i}$ being replaced by $\boldsymbol{\tilde{z}_i}$.

(iii) $n^{\alpha/2} (\sum_{i=1}^n ||\boldsymbol{d_i}||^{6+2\alpha})^{1/2} + \sum_{i=1}^n ||\boldsymbol{\tilde{z}_i}||^4 = O(n^{-1})$

(A.3) (i) $\mathbf{E}\psi(\epsilon_1) = 0$ and $\sigma^2 = \mathbf{E}\psi^2(\epsilon_1) / \mathbf{E}(\psi'(\epsilon_1)) \in (0, \infty)$.

(ii) $\mathbf{E}|\psi(\epsilon_1)|^4 + \mathbf{E}|\psi'(\epsilon_1)|^4 + \mathbf{E}|\psi''(\epsilon_1)|^2 < \infty$.

(A.4) G_i^* and ϵ_i are independent for all $1 \leq i \leq n$.

(A.5) (i) $\mathbf{E}G_1^{*3} < \infty$

(ii) $\mathbf{Var}(G_1^*) = \mu_{G^*}^2$, $\mathbf{E}(G_1^* - \mu_{G^*})^3 = \mu_{G^*}^3$.

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

$$\limsup_{|t|\to\infty} \left|\mathbf{E}\big(exp\big(it\big(G_1^* - \mu_{G^*}\big)\big)\big)\right| < 1.$$

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

$$\limsup_{||(t_1,t_2)||\to\infty} \left|\mathbf{E}\Big(exp\Big(it_1\big(G_1^* - \mu_{G^*}\big) + it_2\big(G_1^* - \mu_{G^*}\big)^2\Big)\Big)\right| < 1$$

(A.6) (i) $\big(\psi(\epsilon_1), \psi'(\epsilon_1)\big)$ satisfies Cramer's condition:

$$\limsup_{||(t_1,t_2)||\to\infty} \left|\mathbf{E}\big(exp\big(it_1\psi(\epsilon_1) + it_2\psi'(\epsilon_1)\big)\big)\right| < 1$$

(i)$'$ $\big(\psi(\epsilon_1), \psi'(\epsilon_1), \psi^2(\epsilon_1)\big)$ satisfies Cramer's condition:

$$\limsup_{||(t_1,t_2,t_3)||\to\infty} \left|\mathbf{E}\big(exp\big(it_1\psi(\epsilon_1) + it_2\psi'(\epsilon_1) + it_3\psi^2(\epsilon_1)\big)\big)\right| < 1$$

Define $\bar{\boldsymbol{v}}_{\boldsymbol{i}} = (\bar{\boldsymbol{x}}'_i, \bar{\boldsymbol{z}}'_{\boldsymbol{i}})'$ where $\bar{\boldsymbol{x}}_{\boldsymbol{i}} = \boldsymbol{x}_{\boldsymbol{i}}\psi(\bar{\epsilon}_i)$, $\bar{\boldsymbol{z}}_{\boldsymbol{i}} = \boldsymbol{z}_{\boldsymbol{i}}\psi'(\bar{\epsilon}_i)$; $\{\bar{\epsilon}_1, \ldots, \bar{\epsilon}_n\}$ being the set of residuals. Also, define $\bar{\boldsymbol{A}}_{\mathbf{2}\boldsymbol{n}} = n^{-1}\sum_{i=1}^n \bar{\boldsymbol{x}}_{\boldsymbol{i}}\bar{\boldsymbol{x}}'_{\boldsymbol{i}}$ and $\bar{\boldsymbol{A}}_{\mathbf{1}\boldsymbol{n}} = n^{-1}\sum_{i=1}^n \boldsymbol{x}_{\boldsymbol{i}}\boldsymbol{x}'_{\boldsymbol{i}}\psi'(\bar{\epsilon}_i)$. Note that $n^{-1}\sum_{i=1}^n \bar{\boldsymbol{v}}_i\bar{\boldsymbol{v}}'_{\boldsymbol{i}}$ is an estimate of the matrix $\mathbf{E}(n^{-1}\sum_{i=1}^n \boldsymbol{v}_{\boldsymbol{i}}\boldsymbol{v}'_{\boldsymbol{i}})$ and due to assumption (A.2)(ii), $\sum_{i=1}^n \bar{\boldsymbol{v}}_i\bar{\boldsymbol{v}}'_{\boldsymbol{i}}$ is non-singular for sufficiently large n. Hence, without loss of generality the canonical decomposition of $\sum_{i=1}^n \bar{\boldsymbol{v}}_i\bar{\boldsymbol{v}}'_{\boldsymbol{i}}$ can be assumed as

$$\boldsymbol{B}\Big(\sum_{i=1}^n \bar{\boldsymbol{v}}_i\bar{\boldsymbol{v}}'_i\Big)\boldsymbol{B}' = \boldsymbol{I}_{\boldsymbol{k}}$$

where $k = p + q$ and $\boldsymbol{B}$ is a $k \times k$ non-singular matrix. Define $k \times 1$ vector $\breve{\boldsymbol{v}}_{\boldsymbol{i}}$ by

$$\breve{\boldsymbol{v}}_{\boldsymbol{i}} = \boldsymbol{B}\bar{\boldsymbol{v}}_{\boldsymbol{i}}, \quad 1 \leq i \leq n$$

To find valid EE in the perturbation bootstrap regime, the following condition [cf.

Navidi (1989)] is also required:

(A.7) There exists a $\delta > 0$ such that $-K_n(\delta)/log\gamma_n \to \infty$ where $\boldsymbol{B_n}(\delta) = \{1 \leq i \leq n : (\breve{\boldsymbol{v}}_i'\boldsymbol{t})^2 > \delta\gamma_n^2$ for all $\boldsymbol{t} \in \mathscr{R}^k$ with $||\boldsymbol{t}||^2 = 1\}$, $K_n(\delta) = |\boldsymbol{B_n}(\delta)|$, the cardinality of the set $\boldsymbol{B_n(\delta)}$, and $\gamma_n = (\sum_{i=1}^n ||\breve{\boldsymbol{v}}_i||^4)^{1/2}$.

But note that the condition (A.7) has already been satisfied in our set up due to Lemma 1.6.4 and the proposition in Lahiri (1992).

Now we briefly explain the assumptions. Assumption (A.1) is smoothness condition on the score function $\psi(\cdot)$. This condition is essential for obtaining a Taylor's expansion of $\psi(\cdot)$ around regression errors. Assumption (A.2) presents the regularity conditions on the design vectors necessary to find EE. For the validity of asymptotic normality of the regression M-estimator, only (A.2)(i) is enough [cf. Huber (1981)]; whereas additional condition (A.2)(ii) is required for the validity of the EE. (A.2)(iii) states atmost how fast the L^2 norm of the design vectors can increase to get a valid EE. This condition is somewhat stronger than the condition (C.6) assumed in Lahiri (1992); although there was a reduction in accuracy of bootstrap approximation due to this relaxation. This type of conditions are quite common in the literature of edgeworth expansions in regression setup; see for example Navidi (1989), Qumsiyeh (1990a). We now state an example where assumption (A.2) (iii) is fulfilled.

Example 1.3.1 *Suppose,* $\{\boldsymbol{X^{(1)}}, \ldots, \boldsymbol{X^{(p)}}\}$ *is a set of independent random vectors where* $\boldsymbol{X}^{(j)} = (X_{1j}, \ldots, X_{nj})'$ *is a vector of n iid copies of the non-degenerate random variable* X_{1j}, $j \in \{1, \ldots, p\}$. *Define,* $p \times p$ *matrix* $\boldsymbol{M} = ((m_{jk}))_{j,k=1,\ldots,p}$ *where* $m_{jk} = \mathbf{E}(X_{1j}^2 X_{1k}^2)$ *and* $n \times p$ *matrix* $\boldsymbol{X} = \big(\boldsymbol{X^{(1)}}, \ldots, \boldsymbol{X^{(p)}}\big)$. *Assume,* $\mathbf{E}(X_{1j}) = \mathbf{E}(X_{1j}^3) = 0$ *and* $\mathbf{E}|X_{1j}|^8 < \infty$ *for* *all $j \in \{1, \ldots, p\}$ and* $\det(M) \neq 0$*. Then for the design matrix $\boldsymbol{X}$, assumption* (A.2) (iii) *holds with probability 1 (w.p. 1).*

proof :

For the design matrix $\boldsymbol{X}$, $\boldsymbol{x_i} = (X_{i1}, X_{i2}, \ldots, X_{ip})'$ and $\boldsymbol{z_i} = (X_{i1}^2, X_{i1}X_{i2}, \ldots, X_{i1}X_{ip}, X_{i2}^2, X_{i2}X_{i3}, \ldots, X_{i2}X_{ip}, \ldots, X_{ip}^2)'$ for $i \in \{1, \ldots, n\}$.

First note that if all the entries of $\boldsymbol{X}$ are iid then the condition $\det(M) \neq 0$ is redundant. By Kolmogorov strong law of large numbers, $\boldsymbol{A_n} = n^{-1}\boldsymbol{D^2} \rightarrow diag\big(\mathbf{E}(X_{11}^2), \ldots, \mathbf{E}(X_{1p}^2)\big)$ and $n^{-1}\sum_{i=1}^{n} ||\boldsymbol{x_i}||^{6+2\alpha} \rightarrow \mathbf{E}||\boldsymbol{x_1}||^{6+2\alpha}$ both $w.p.1$ and hence

$$
\begin{aligned}
n^{\alpha/2}\Big(\sum_{i=1}^{n} ||\boldsymbol{d_i}||^{6+2\alpha}\Big)^{1/2} &\leq n^{\alpha/2}||\boldsymbol{D^{-1}}||^{3+\alpha}\Big(\sum_{i=1}^{n} ||\boldsymbol{x_i}||^{6+2\alpha}\Big)^{1/2} \\
&= O(n^{-1}) \quad w.p.1 \qquad (1.3.2)
\end{aligned}
$$

Again, since $\boldsymbol{M}$ is a non-singular matrix, $n^{-1}\sum_{i=1}^{n} \boldsymbol{z}_i\boldsymbol{z}_i' \rightarrow \boldsymbol{N}$ $w.p.1$, for some positive definite matrix $\boldsymbol{N}$. This implies that $||\boldsymbol{L}|| = O(n^{-1/2})$ $w.p.1$ and hence

$$
\begin{aligned}
\sum_{i=1}^{n} ||\boldsymbol{\tilde{z}_i}||^4 &\leq ||\boldsymbol{L}||^4 \sum_{i=1}^{n} ||\boldsymbol{z_i}||^4 \\
&= O(n^{-1}) \quad w.p.1 \qquad (1.3.3)
\end{aligned}
$$

Therefore, our claim follows from (1.3.2) and (1.3.3).

Assumption (A.3) is the moment condition on the error variables through the score function $\psi(\cdot)$. (A.3)(i) is generally assumed to establish asymptotic normality. Assumption (A.4) is inherent in the present setup, since G_i^*'s are introduced by us to define the bootstrapped estimator whereas ϵ_i's are already present in the process of data generation. The conditions present in Assumption (A.5) are moment and smoothness conditions on the perturbing quantities G_i^*'s, required for the valid two term EE in bootstrap setup. The Cramer's condition is very common in the literature of edgeworth expansions. Cramer's condition is satisfied when the distribution of $(G_1^* - \mu_{G^*})$ or $((G_1^* - \mu_G^*), (G_1^* - EG_1^*)^2)$ has a non-degenerate component which is absolutely continuous with respect to Lebesgue measure [cf. Hall (1992)]. An immediate choice of the distribution of G_1^* is $Beta(\gamma, \delta)$ where $3\gamma = \delta = 3/2$. Also one can investigate *Generalized Beta* family of distributions for more choices of the distribution of G_1^*. Assumption (A.6) is the Cramer's condition on the errors. Although this assumption is not needed for obtaining EE of the bootstrapped estimators, it is needed for obtaining EE for the original M-estimator.

Note that the condition (A.7) is somewhat abstract. Hence as pointed out by a referee, some clarification would be helpful. To this end, it is worth mentioning that to find formal EE for the standardized bootstrapped pivot (see section 1.4.1), the most difficult step is to show

$$\max_{|\boldsymbol{\alpha}| \leq p+q+4} \int_{C_1 \leq \gamma_n ||\boldsymbol{t}|| \leq C_2} |D^{\boldsymbol{\alpha}} \mathbf{E}_* e^{i \boldsymbol{t}' \boldsymbol{T}_n^*}| d\boldsymbol{t} = o_p\big(n^{-1/2}\big) \tag{1.3.4}$$

where C_1, C_2 are non-negative constants and $\boldsymbol{T}_n^* = \sum_{i=1}^n \big(\breve{\boldsymbol{X}}_{\boldsymbol{i}}^* - \mathbf{E}_*(\breve{\boldsymbol{X}}_{\boldsymbol{i}}^*)\big)$, with $\breve{\boldsymbol{X}}_{\boldsymbol{i}}^* = \breve{\boldsymbol{v}}_{\boldsymbol{i}}(G_i - \mu_{G^*})\mathbf{1}\big(||\breve{\boldsymbol{v}}_{\boldsymbol{i}}(G_i - \mu_{G^*})|| \leq 1\big)$. Now it is easy to see that for any $|\boldsymbol{\alpha}| \leq p+q+4$, $|D^{\boldsymbol{\alpha}} \mathbf{E}_* e^{i \boldsymbol{t}' \boldsymbol{T}_n^*}|$ is bounded above by a sum of $n^{|\alpha|}$-terms, each of which is bounded above by

$$C(\alpha) \cdot \max\{\mathbf{E}_* ||\breve{\boldsymbol{X}}_{\boldsymbol{i}}^* - \mathbf{E}_*(\breve{\boldsymbol{X}}_{\boldsymbol{i}}^*)||^{|\boldsymbol{\alpha}|} : i \in \boldsymbol{I}_{\boldsymbol{n}}^*\} \cdot \prod_{i \in \boldsymbol{I}_{\boldsymbol{n}}^{*c}} |\mathbf{E}_* e^{i \boldsymbol{t}' \breve{\boldsymbol{X}}_{\boldsymbol{i}}^*}|$$

where $\boldsymbol{I}_{\boldsymbol{n}}^* \subset \{1, \ldots, n\}$ is of size $|\boldsymbol{\alpha}|$ and $\boldsymbol{I}_{\boldsymbol{n}}^{*c} = \{1, \ldots, n\} \backslash \boldsymbol{I}_{\boldsymbol{n}}^*$ and $C(\boldsymbol{\alpha})$ is a constant which depends only on $\boldsymbol{\alpha}$.

Now note that for all $i \in \{1, \ldots, n\}$,

$$\mathbf{E}_* ||\check{\boldsymbol{X}}_{\boldsymbol{i}}^* - \mathbf{E}_*(\check{\boldsymbol{X}}_{\boldsymbol{i}}^*)||^{|\boldsymbol{\alpha}|} \leq 2^{|\boldsymbol{\alpha}|}$$

$$\text{and} \qquad |\mathbf{E}_* e^{i\boldsymbol{t}'\check{\boldsymbol{X}}_{\boldsymbol{i}}^*}| \leq |\mathbf{E}_* e^{i\boldsymbol{t}'\check{\boldsymbol{v}}_{\boldsymbol{i}}(G_i - \mu_{G^*})}| + 2\mathbf{P}_*\big(||\check{\boldsymbol{v}}_{\boldsymbol{i}}(G_i - \mu_{G^*})|| > 1\big)$$

Hence, in view of Cramer's condition (A.5) (iii) and Lemma 1.6.4, if there exists a sequence of sets $\{\boldsymbol{J_n}\}_{n\geq 1}$ such that $\boldsymbol{J_n} \subset \{1, \ldots, n\}$ and for all $i \in \boldsymbol{J_n}$, $\gamma_n^{-1}|\boldsymbol{t}'\check{\boldsymbol{v}}_{\boldsymbol{i}}| > \xi$ for some $\xi > 0$, then for some $0 < \theta < 1$ we have

$$\begin{aligned}
&\sup\Big\{\prod_{i \in \boldsymbol{I_n^{*c}}} |\mathbf{E}_* e^{i\boldsymbol{t}'\check{\boldsymbol{X}}_{\boldsymbol{i}}^*}| : C_1 \leq \gamma_n ||\boldsymbol{t}|| \leq C_2\Big\} \\
&\leq \sup\Big\{\prod_{i \in \boldsymbol{I_n^{*c}} \cap \boldsymbol{J_n}} |\mathbf{E}_* e^{i\gamma_n^{-1}\boldsymbol{t}'\check{\boldsymbol{X}}_{\boldsymbol{i}}^*}| : C_1 \leq ||\boldsymbol{t}|| \leq C_2\Big\} \\
&\leq \theta^{|\boldsymbol{I_n^{*c}} \cap \boldsymbol{J_n}|} \qquad (1.3.5)
\end{aligned}$$

Again, $|\boldsymbol{I_n^{*c}} \cap \boldsymbol{J_n}| \geq |\boldsymbol{J_n}| - |\boldsymbol{\alpha}|$ and $\gamma_n \geq kn^{-1}$. Therefore, to achieve (1.3.4), it is enough to have

$$n^{2(p+q)+4} \cdot \theta^{|\boldsymbol{J_n}| - (p+q+4)} = o(n^{-1/2})$$

Hence due to Lemma 1.6.4, it is enough to have $|\boldsymbol{J_n}| \geq a_n - C \cdot \log \gamma_n$ for some positive constant C and a sequence of constants $\{a_n\}$ increasing to ∞. This observation together with (1.3.5) justifies condition (A.7).

We will denote the assumptions (A.1)-(A.5) by (A.1)$'$-(A.5)$'$ when (A.2) and (A.5) are respectively defined with (ii)$'$ and (iii)$'$ instead of (ii) and (iii).

1.4 Main Results

1.4.1 Rate of Perturbation Bootstrap Approximation

Here we will state the approximation results both in standardized and studentized setup. It is well known that $\sqrt{n}\bar{\boldsymbol{\beta}}_n$ has asymptotic variance $\sigma^2 \boldsymbol{A}_n^{-1}$. So, the standardized version of the M-estimator $\bar{\boldsymbol{\beta}}_n$ is defined as $\boldsymbol{F}_n = \sqrt{n}\sigma^{-1}\boldsymbol{A}_n^{1/2}(\bar{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$. Now to define the standardized version of the corresponding bootstrapped statistic $\boldsymbol{\beta}_n^*$, we need its conditional asymptotic variance, given the data. Using Taylor's expansion, it is quite easy to get the conditional asymptotic variance of $\sqrt{n}\boldsymbol{\beta}_n^*$ as $\bar{\boldsymbol{A}}_{1n}^{-1}\bar{\boldsymbol{A}}_{2n}\bar{\boldsymbol{A}}_{1n}^{-1}$. Note that inverse of the matrices $\bar{\boldsymbol{A}}_{1n}^{-1}$ and $\bar{\boldsymbol{A}}_{2n}^{-1}$ are well defined for sufficiently large sample size n due to the assumption (A.2)(i) and (A.3)(ii). Hence, the standardized bootstrapped M-estimator $\boldsymbol{F}_n^*$ can be defined as

$$\boldsymbol{F}_n^* = \sqrt{n}\bar{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^* - \bar{\boldsymbol{\beta}}_n)$$

where $\bar{\boldsymbol{\Sigma}}_n^{-1/2} = \bar{\boldsymbol{A}}_{2n}^{-1/2}\bar{\boldsymbol{A}}_{1n}$, $\bar{\boldsymbol{A}}_{2n}^{1/2}$ being defined in terms of the spectral decomposition of $\bar{\boldsymbol{A}}_{2n}$; although it can be defined in many different ways [cf. Lahiri (1994)]. Under some regularity conditions, both the distribution of $\boldsymbol{F}_n$ and the conditional distribution of $\boldsymbol{F}_n^*$ can be shown to be approximated asymptotically by a Normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{I}_p$. Hence, it is straightforward that perturbation bootstrap approximation to the distribution of the M-estimator is first order correct. The second order result in standardized case is formally stated in Theorem 1.4.1.

Proposition 1.4.1 *Suppose, the assumptions* (A.1)-(A4), (A.5)(i) *hold. Then there exist constant $C_1 > 0$ and a sequence of Borel sets $\boldsymbol{Q}_{\mathbf{1}n} \subseteq \mathscr{R}^n$, such that $\mathbf{P}((\epsilon_1, \ldots, \epsilon_n) \in \boldsymbol{Q}_{\mathbf{1}n}) \to 1$ as $n \to \infty$, and given $(\epsilon_1, \ldots, \epsilon_n) \in \boldsymbol{Q}_{\mathbf{1}n}$, $n \geq C_1$ such that there exists a sequence of* *statistics* $\{\boldsymbol{\beta_n^*}\}_{n\geq 1}$ *such that*

$$\mathbf{P}_*\big(\boldsymbol{\beta_n^*} \text{ solves } (1.2.1) \text{ and } ||\boldsymbol{\beta_n^*} - \boldsymbol{\bar{\beta}_n}|| \leq C_1.n^{-1/2}.(logn)^{1/2}\big) \geq 1 - \delta_n n^{-1/2}$$

where $\delta_n \equiv \delta_n(\epsilon_1, \ldots, \epsilon_n)$ *tends to 0.*

Theorem 1.4.1 *Let* $\{\boldsymbol{\beta_n^*}\}_{n\geq 1}$ *be a sequence of statistics satisfying Proposition 1.4.1 depending on* $(\epsilon_1, \ldots, \epsilon_n)$*. Assume, the assumptions* (A.1)-(A.5) *hold.*

(a) *Then there exist constant* $C_2 > 0$ *and a sequence of Borel sets* $\boldsymbol{Q_{2n}} \subseteq \mathscr{R}^n$ *and polynomial* $a_n^*(\cdot, \psi, G^*)$ *depending on first three moments of* G_1^* *and on* $\psi(\cdot)$, $\psi'(\cdot)$ *&* $\psi''(\cdot)$ *through the residuals* $\{\bar{\epsilon}_1, \ldots, \bar{\epsilon}_n\}$ *such that given* $(\epsilon_1, \ldots, \epsilon_n) \in \boldsymbol{Q_{2n}}$*, with* $\mathbf{P}((\epsilon_1, \ldots, \epsilon_n) \in \boldsymbol{Q_{2n}}) \to 1$*, we have for* $n \geq C_2$*,*

$$\sup_{B\in\mathscr{B}} |\mathbf{P}_*(\boldsymbol{F_n^*} \in B) - \int_B \xi_n^*(\boldsymbol{x})d\boldsymbol{x}| \leq \delta_n n^{-1/2}$$

where $\xi_n^*(\boldsymbol{x}) = (1 + n^{-1/2}a_n^*(\boldsymbol{x}, \psi, G^*))\phi(\boldsymbol{x})$ *and* $\delta_n \equiv \delta_n(\epsilon_1, \ldots, \epsilon_n)$ *tends to 0.*

(b) *Suppose in addition assumption* (A.6)(i) *holds. Then we have,*

$$\sup_{B\in\mathscr{B}} \big|\mathbf{P}_*(\boldsymbol{F_n^*} \in B) - \mathbf{P}(\boldsymbol{F_n} \in B)\big| = o_p(n^{-1/2})$$

Now, the quantity σ^2 is mostly unavailable in practical circumstances. Hence, the non-pivotal quantity like $\boldsymbol{F_n}$ is very rare in use in providing valid inferences. It is more reasonable to explore the asymptotic properties of a pivotal quantity, like the studentized version of the M-estimator $\boldsymbol{\bar{\beta}_n}$. Depending on the observed residuals $\bar{\epsilon}_i = y_i - \boldsymbol{x_i}'\boldsymbol{\bar{\beta}_n},\ i \in \{1, \ldots, n\}$, the natural way to define an estimator of σ^2 is $\hat{\sigma}_n^2$ where $\hat{\sigma}_n = s_n\tau_n^{-1}$, $\tau_n = n^{-1}\sum_{i=1}^n \psi'(\bar{\epsilon}_i)$ and $s_n^2 = n^{-1}\sum_{i=1}^n \psi^2(\bar{\epsilon}_i)$. Hence, the studentized M-estimator in regression setup may be defined as $\boldsymbol{H_n} = \sqrt{n}\hat{\sigma}_n^{-1}\boldsymbol{A_n^{1/2}}(\boldsymbol{\bar{\beta}_n} - \boldsymbol{\beta})$.

Similarly, studentized version of the corresponding bootstrapped estimator can be construed as $\sqrt{n}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{*-1/2}(\boldsymbol{\beta}_{\boldsymbol{n}}^{*} - \bar{\boldsymbol{\beta}}_{\boldsymbol{n}})$ where $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{*-1/2}$ is defined in the same way as $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{-1/2}$ after replacing $\bar{\epsilon}_i$ by ϵ_i^*, $\epsilon_i^* = y_i - \boldsymbol{x}_{\boldsymbol{i}}'\boldsymbol{\beta}_{\boldsymbol{n}}^{*}$, for each $i \in \{1, \ldots, n\}$. But this definition requires to obtain $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{*-1/2}$ at each bootstrap sample $(G_1^{*b}, \ldots, G_n^{*b})$, $b = 1, \ldots, B$. This is not computationally feasible since computation of $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{*-1/2}$ involves matrix inversion. So, let us define the studentized bootstrapped M-estimator in more computationally feasible way as

$$\boldsymbol{H}_{\boldsymbol{n}}^{*} = \sqrt{n}\sigma_n^{*-1}\hat{\sigma}_n\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{-1/2}(\boldsymbol{\beta}_{\boldsymbol{n}}^{*} - \bar{\boldsymbol{\beta}}_{\boldsymbol{n}})$$

where $\sigma_n^* = s_n^*\tau_n^{*-1}$, $\tau_n^* = n^{-1}\sum_{i=1}^{n}\psi'(\epsilon_i^*)$, $s_n^{*2} = n^{-1}\sum_{i=1}^{n}\psi^2(\epsilon_i^*)$ and $\hat{\sigma}_n^2$ and $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{-1/2}$ are as defined earlier.

Theorem 1.4.2 *Suppose, the assumptions* (A.1)-(A.5) *hold.*

(a) *Then there exist constant $C_3 > 0$ and a sequence of Borel sets $\boldsymbol{Q}_{\boldsymbol{3n}} \subseteq \mathscr{R}^n$ and polynomial $\tilde{a}_n^*(\cdot, \psi, G^*)$ depending on first three moments of G_1^* and on $\psi(\cdot)$, $\psi'(\cdot)$ & $\psi''(\cdot)$ through the residuals $\{\bar{\epsilon}_1, \ldots, \bar{\epsilon}_n\}$, such that given $(\epsilon_1, \ldots, \epsilon_n) \in \boldsymbol{Q}_{\boldsymbol{3n}}$, with $\mathbf{P}((\epsilon_1, \ldots, \epsilon_n) \in \boldsymbol{Q}_{\boldsymbol{3n}}) \to 1$, we have for $n \geq C_3$,*

$$\sup_{B \in \mathscr{B}} |\mathbf{P}_*(\boldsymbol{H}_{\boldsymbol{n}}^{*} \in B) - \int_B \tilde{\xi}_n^*(\boldsymbol{x})d\boldsymbol{x}| \leq \delta_n n^{-1/2}$$

where $\tilde{\xi}_n^*(\boldsymbol{x}) = (1 + n^{-1/2}\tilde{a}_n^*(\boldsymbol{x}, \psi, G^*))\phi(\boldsymbol{x})$ and $\delta_n \equiv \delta_n(\epsilon_1, \ldots, \epsilon_n)$ tends to 0.

Suppose in addition assumption (A.6)(i)$'$ *holds. Then*

(b) *for the collection of Borel sets $\mathscr{B}$ defined by (1.3.1),*

$$\sup_{B \in \mathscr{B}} \left|\mathbf{P}_*(\boldsymbol{H}_{\boldsymbol{n}}^{*} \in B) - \mathbf{P}(\boldsymbol{H}_{\boldsymbol{n}} \in B)\right| = O_p(n^{-1/2})$$

(c) *if* $2E\psi^2(\epsilon_1)E\psi(\epsilon_1)\psi'(\epsilon_1) \neq E\psi'(\epsilon_1)E\psi^3(\epsilon_1)$, *then there exists* $\epsilon > 0$ *such that,*

$$\mathbf{P}\Big(\liminf_{n\to\infty}\sqrt{n}\Big[\sup_{B\in\mathscr{B}}\big|\mathbf{P}_*(\boldsymbol{H_n^*}\in B)-\mathbf{P}(\boldsymbol{H_n}\in B)\big|\Big] > \epsilon\Big) = 1$$

Remark 1.4.1 *Proposition 1.4.1 states that there exists a sequence of perturbation bootstrapped estimator* $\boldsymbol{\beta_n^*}$ *within a neighborhood of length* $C.n^{-1/2}(logn)^{1/2}$ *around the original M-estimator* $\boldsymbol{\bar{\beta}_n}$ *outside a set of bootstrap probability* $o_p(n^{-1/2})$*. This existence result is essential in finding valid EEs in bootstrap regime. This can be compared with Theorem 2.3 (a) of Lahiri (1992), where similar kind of result was shown in case of residual and generalized bootstrap.*

Remark 1.4.2 *Note that, where as the error term in approximating the distribution of M-estimator by perturbation bootstrap is of order* $O_p(n^{-1/2})$ *in the prevalent studentize setup, it reduces the order of the error of approximation to* $o_p(n^{-1/2})$ *in simple standardized setup. This means that the difference between coefficients corresponding to the term* $n^{-1/2}$ *in the EEs of original and bootstrapped estimator can be made arbitrarily small in standardized setup, but not in usual studentized setup.*

Remark 1.4.3 *To understand part (c) of Theorem 1.4.2, consider the usual least square estimator. In least square setup, the condition in the Theorem 1.4.2 (c) reduces to* $E\epsilon^3 \neq 0$*. This simply means that if the studentization in perturbation bootstrapped version is performed analogously as in case of original least square estimator, then the bootstrap distribution can not correct the original distribution upto second order. If this is investigated more deeply, then it can be observed that the usual studentized perturbation bootstrap approximation can not correct for the skewness of the error distribution F.*

1.4.2 Examples

Theorem 1.4.2 concludes that the standard way of performing studentization of the bootstrapped estimator is first order correct. In order to show that the usual studentized setup is not second order correct, we consider following two important special cases with $\psi(x) = x$.

Example 1.4.2.1

Consider the observations $\{y_1, \ldots, y_n\}$ are coming from the distribution F with a location shift μ. This in terms of regression model becomes

$$y_i = \mu + \epsilon_i$$

Hence, in this setup $p = 1$, $\boldsymbol{\beta} = \mu$ and $x_i = 1$ for all $i \in \{1, \ldots, n\}$.

It can be shown that in this setup, $\tilde{\xi}_n(\cdot)$ and $\tilde{\xi}_n^*(\cdot)$, the EE of $\boldsymbol{H_n}$ and $\boldsymbol{H_n^*}$ respectively, turn out to be

$$\tilde{\xi}_n(x) = \Big[1 - n^{-1/2}\big\{\tilde{b}_{11}\frac{d}{dx} + 6^{-1}\tilde{b}_{31}\frac{d^3}{dx^3}\big\}\Big]\phi(x)$$

$$\tilde{\xi}_n^*(x) = \Big[1 - n^{-1/2}\big\{\tilde{b}_{11}^*\frac{d}{dx} + 6^{-1}\tilde{b}_{31}^*\frac{d^3}{dx^3}\big\}\Big]\phi(x)$$

where

$\tilde{b}_{11} = -2^{-1}\sigma^{-3}\mathbf{E}\epsilon_1^3$, $\tilde{b}_{31} = -2\sigma^{-3}\mathbf{E}\epsilon_1^3$

$\tilde{b}_{11}^* = -2\sigma_n^{-1}n^{-1}\sum_{i=1}^n \bar{\epsilon}_i$, $\tilde{b}_{31}^* = \sigma_n^{-3} - 12\sigma_n^{-1}n^{-1}\sum_{i=1}^n \bar{\epsilon}_i$.

It is clear that $\tilde{b}_{11}^*$ as well as $\tilde{b}_{31}^*$ are not converging respectively to $\tilde{b}_{11}$ and $\tilde{b}_{31}$ in probability and hence the perturbation bootstrap method is not second order correct in the above setup when the bootstrapped estimator is studentized in the usual manner.

Example 1.4.2.2

Consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where β_0 and β_1 are parameters of interest and ϵ_i's are iid errors. This model, in terms of our multivariate linear regression structure, can be written as $y_i = \tilde{\boldsymbol{x}}_{\boldsymbol{i}}{}'\boldsymbol{\beta} + \epsilon_i$ where $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ and $\tilde{\boldsymbol{x}}_{\boldsymbol{i}} = (1, x_i)'$. Hence, the EEs of the original and bootstrapped estimators upto the order $o(n^{-1/2})$, after usual studentization, respectively becomes

$$\tilde{\xi}_n(y_1, y_2) = \left[1 - n^{-1/2}\Big\{\sum_{j=1}^{2} \tilde{b}_{11}^{*(j)} \frac{\partial}{\partial y_j} + \sum_{j=0}^{3} \frac{\tilde{b}_{31}^{(j,3-j)}}{j!(3-j)!} D^{(j,3-j)}\Big\}\right]\phi(y_1, y_2)$$

$$\tilde{\xi}_n^*(y_1, y_2) = \left[1 - n^{-1/2}\Big\{\sum_{j=1}^{2} \tilde{b}_{11}^{*(j)} \frac{\partial}{\partial y_j} + \sum_{j=0}^{3} \frac{\tilde{b}_{31}^{*(j,3-j)}}{j!(3-j)!} D^{(j,3-j)}\Big\}\right]\phi(y_1, y_2)$$

where

$$\tilde{b}_{11}^{(j)} = -2^{-1}\left[n^{-1}\textstyle\sum_{i=1}^{n} \mathbf{e_j'}\mathbf{A_n^{-1/2}}\mathbf{\tilde{x}_i}\right]\gamma_1$$

$$\tilde{b}_{31}^{(j_1,j_2)} = -2\left[n^{-1}\textstyle\sum_{i=1}^{n}(\mathbf{e_1'}\mathbf{A_n^{-1/2}}\mathbf{\tilde{x}_i})^{j_1}(\mathbf{e_2'}\mathbf{A_n^{-1/2}}\mathbf{\tilde{x}_i})^{j_2}\right]\gamma_1$$

$$\tilde{b}_{11}^{*(j)} = o_p(1)$$

$$\tilde{b}_{31}^{*(j_1,j_2)} = \left[n^{-1}\textstyle\sum_{i=1}^{n}(\mathbf{e_1'}\mathbf{\bar{A}_{2n}^{-1/2}}\mathbf{\tilde{x}_i})^{j_1}(\mathbf{e_2'}\mathbf{A_{2n}^{-1/2}}\mathbf{\tilde{x}_i})^{j_2}\psi^3(\bar{\epsilon}_i)\right] + o_p(1)$$

where $j = 1$ or 2, $j_1, j_2 \in \{0, 1, 2, 3\}$ such that $j_1 + j_2 = 3$, γ_1 is the coefficient of skewness of ϵ_1, $\mathbf{\bar{A}_{2n}}$ is as defined in general setup and $\mathbf{A_n} = n^{-1}\sum_{i=1}^{n}\mathbf{\tilde{x}_i}\mathbf{\tilde{x}_i'} = \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \bar{x^2} \end{bmatrix}$

where $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$ and $\bar{x^2} = n^{-1}\sum_{i=1}^{n} x_i^2$.

Therefore, it is clear that the naive studentized bootstrapped estimator does not correct for the skewness corresponding to the original version in simple linear regression setup. The usual studentization in perturbation bootstrap becomes second order correct only if $E\epsilon_1^3 = 0$.

Note that, the coefficients $\tilde{b}_{11}^{(j)}, 1 \leq j \leq p$, all can not vanish together unless $\gamma_1 = 0$ and hence $\tilde{b}_{11}^{*(j)}$ can not converge to $\tilde{b}_{11}^{(j)}$ unless $\gamma_1 = 0$. Similarly, it can be shown that same condition is required to have the closeness of the coefficients $\tilde{b}_{31}^{(j,3-j)}$ and $\tilde{b}_{31}^{*(j,3-j)}$. Hence, the two EEs can not get closer unless $\gamma_1 = 0$, similar to the Example 1.4.2.1. This is exactly what is stated in the part (c) of Theorem 1.4.2 in most general form.

1.4.3 Modification to the bootstrapped pivot

As it has been seen that $\boldsymbol{H}_n^*$, the usual studentized version of the perturbation bootstrapped estimator is not attending the desired optimal rate $o_p(n^{-1/2})$, so in the perspective of statistical inference, perturbation bootstrap is not advantageous over asymptotic normal approximation. For the sake of obtaining second order correctness, define the modified studentized $\boldsymbol{\beta}_n^*$ as

$$\tilde{\boldsymbol{H}}_n^* = \sqrt{n}(\tilde{\sigma}_n^*)^{-1}\hat{\sigma}_n \bar{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^* - \bar{\boldsymbol{\beta}}_n) \tag{1.4.1}$$

where

$\tilde{\sigma}_n^* = \tilde{s}_n^* \tilde{\tau}_n^{*-1}$, $\tilde{\tau}_n^* = n^{-1}\sum_{i=1}^{n} \psi'(\epsilon_i^*) G_i^*$, $\tilde{s}_n^{*2} = n^{-1}\sum_{i=1}^{n} \psi^2(\epsilon_i^*)(G_i^* - \mu_{G^*})^2$.

The bootstrapped statistic $\tilde{\boldsymbol{H}}_n^*$ can be seen to be achieving the optimal rate, namely $o_p(n^{-1/2})$, in approximating the original studentized M-estimator $\boldsymbol{H}_n$, which is formally stated in the following theorem:

Theorem 1.4.3 *Suppose, the assumptions* (A.1)$'$-(A.5)$'$ *hold. Also assume* $EG_1^{*4} < \infty$*.*

(a) *Then there exist constant* $C_4 > 0$ *and a sequence of Borel sets* $\boldsymbol{Q_{4n}} \subseteq \mathscr{R}^n$ *and polynomial* $\bar{a}_n^*(\cdot, \psi, G^*)$ *depending on first three moments of* G_1^* *and on* $\psi(\cdot)$, $\psi'(\cdot)$ *&* $\psi''(\cdot)$ *through the residuals* $\{\bar{\epsilon}_1, \ldots, \bar{\epsilon}_n\}$*, such that given* $(\epsilon_1, \ldots, \epsilon_n) \in \boldsymbol{Q_{4n}}$*, with* $\mathbf{P}((\epsilon_1, \ldots, \epsilon_n) \in \boldsymbol{Q_{4n}}) \to 1$*, we have for* $n \geq C_4$*,*

$$\sup_{B \in \mathscr{B}} |\mathbf{P}_*(\tilde{\boldsymbol{H}}_{\boldsymbol{n}}^* \in B) - \int_B \bar{\xi}_n^*(\boldsymbol{x}) d\boldsymbol{x}| \leq \delta_n n^{-1/2}$$

where $\bar{\xi}_n^*(\boldsymbol{x}) = (1 + n^{-1/2} \bar{a}_n^*(\boldsymbol{x}, \psi, G^*))\phi(\boldsymbol{x})$ *and* $\delta_n \equiv \delta_n(\epsilon_1, \ldots, \epsilon_n)$ *tends to 0.*

(b) *Suppose, in addition* (A.6)(i)$'$ *holds. Then, for the collection of Borel sets defined by (1.3.1),*

$$\sup_{B \in \mathscr{B}} \left|\mathbf{P}_*(\tilde{\boldsymbol{H}}_{\boldsymbol{n}}^* \in B) - \mathbf{P}(\boldsymbol{H_n} \in B)\right| = o_p(n^{-1/2})$$

Remark 1.4.4 *The modification that is needed to make the perturbation bootstrap method correct upto second order, suggests that besides incorporating the effect of bootstrap randomization through* $\psi(\cdot)$ *and* $\psi'(\cdot)$ *in the studentization factor of the bootstrap estimator, it is also essential to blend properly the effect of randomization that is coming directly from the perturbing quantities* G_i^**s.*

Remark 1.4.5 *As pointed out by a referee, the usefulness of the above results depend critically on the rate of the probability* $\mathbf{P}\big((\epsilon_1, \ldots, \epsilon_n \in \boldsymbol{Q_{in}})\big)$*,* $i = 1, 2, 3, 4$*. Following the steps of the proofs, it can be shown that* $\mathbf{P}\big((\epsilon_1, \ldots, \epsilon_n \in \boldsymbol{Q_n})\big) = 1 - O\big(n^{-1/2}(\log n)^{-2+\gamma_2}\big)$ *where* $\boldsymbol{Q_n} = \cap_{i=1}^4 \boldsymbol{Q_{in}}$*, for some* $\gamma_2 \in (0, 2)$*, although the rate can be improved under moment condition stronger than (A.3) (ii). In general, if* $\mathbf{E}|\psi(\epsilon_1)|^{2\gamma_3} + \mathbf{E}|\psi'(\epsilon_1)|^{2\gamma_3} + \mathbf{E}|\psi''(\epsilon_1)|^{\gamma_3} < \infty$ *for some natural number* $\gamma_3 \geq 2$*, then analogously it can be shown that* $\mathbf{P}\big((\epsilon_1, \ldots, \epsilon_n \in \boldsymbol{Q_n})\big) = 1 - O\big(n^{-(2\gamma_3 - 3)/2}(\log n)^{-\gamma_3 + \gamma_2}\big)$ *for some* $\gamma_2 \in (0, \gamma_3)$*. This implies that second* *order correctness of perturbation bootstrap can be established in almost sure sense under higher moment condition.*

Remark 1.4.6 *The condition (1.3.1) on the collection of Borel subsets $\mathscr{B}$ of $\mathscr{R}^p$, that is considered in the above theorems, is somewhat abstract. This condition is needed for achieving two goals. One is to obtain valid EE for the normalized part of the underlying pivot [cf. Corollary 20.4 of Bhattacharya and Rao (1986)] and the other one is to bound the remainder term with an order $o(n^{-1/2})$ with probability (or bootstrap probability) $1 - o(n^{-1/2})$. These two together allow us to get EE for the underlying pivots. A natural choice for $\mathscr{B}$ is the collection of all Borel measurable convex subsets of $\mathscr{R}^p$.*

1.5 Extension to independent and non-identically distributed errors

In this section, we will extend second order results of perturbation bootstrap to the model (1.1.1) with independent and non-identically distributed [hereafter referred to as non-iid] errors. 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 heteroscedasticity. We will show that perturbation bootstrap can approximate the exact distribution of the regression M-estimator $\bar{\boldsymbol{\beta}}_{\boldsymbol{n}}$ upto second order even when the errors are non-iid.

Before going to define the original and bootstrap pivots, we state some additional assumptions needed to establish second order correctness. Define, $\boldsymbol{A_{1n}} = n^{-1}\sum_{i=1}^{n} \boldsymbol{x_i x_i'}$ $\mathbf{E}\psi'(\epsilon_i)$ and $\boldsymbol{A_{2n}} = n^{-1}\sum_{i=1}^{n} \boldsymbol{x_i x_i'}\mathbf{E}\psi^2(\epsilon_i)$.

(A.2)(iii)$''$ $n^{-2}\sum_{i=1}^{n}||\boldsymbol{x_i}||^{12}+\sum_{i=1}^{n}\big[||\boldsymbol{\tilde{z}_i}||^4\max\{1,\mathbf{E}|\psi'(\epsilon_i)|^4\}\big]=O(n^{-1})$.

(A.3)(i)$''$ $\mathbf{E}\psi(\epsilon_i)=0$ for all $i\in\{1,\ldots,n\}$.

(A.3)(ii)$''$ $n^{-1}\sum_{i=1}^{n}\big[\mathbf{E}|\psi(\epsilon_i)|^{6+v}+\mathbf{E}|\psi'(\epsilon_i)|^{6+v}+\mathbf{E}|\psi''(\epsilon_i)|^{4+v}\big]=O(1)$ for some $v>0$.

(A.6)(i)$''$ $\big(\psi(\epsilon_n),\psi'(\epsilon_n),\psi^2(\epsilon_n)\big)_{n=1}^{\infty}$ satisfies Cramer's condition in a uniform sense i.e. for any positive b,

$$\limsup_{n\to\infty}\sup_{||(t_1,t_2,t_3)||>b}\Big|\mathbf{E}\big(exp\big(it_1\psi(\epsilon_n)+it_2\psi'(\epsilon_n)+it_3\psi^2(\epsilon_n)\big)\big)\Big|<1.$$

(A.8) $\boldsymbol{A_{1n}}$ and $\boldsymbol{A_{2n}}$ both converge to non-singular matrices as $n\to\infty$.

We will denote the assumptions (A.1)-(A.4) by (A.1)$''$-(A.4)$''$ when (A.2) is defined with (iii)$''$ instead of (iii) and (A.3) is defined with (i)$''$, (ii)$''$ in place of (i) and (ii) respectively.

1.5.1 Rate of Perturbation Bootstrap Approximation

Note that when the regression errors are non-identically distributed, $\sqrt{n}\boldsymbol{\bar{\beta}_n}$ has asymptotic variance $\boldsymbol{A_{1n}^{-1}A_{2n}A_{1n}^{-1}}$. Hence, the natural way of defining studentized pivot corresponding to $\boldsymbol{\bar{\beta}_n}$ is

$$\boldsymbol{\breve{H}_n}=\sqrt{n}\boldsymbol{\bar{\Sigma}_n^{-1/2}}(\boldsymbol{\bar{\beta}_n}-\boldsymbol{\beta})$$

where $\boldsymbol{\bar{\Sigma}_n^{-1/2}}=\boldsymbol{\bar{A}_{2n}^{-1/2}\bar{A}_{1n}}$ with $\boldsymbol{\bar{A}_{1n}}=n^{-1}\sum_{i=1}^{n}\boldsymbol{x_i x_i'}\psi'(\bar{\epsilon}_i)$, $\boldsymbol{\bar{A}_{2n}}=n^{-1}\sum_{i=1}^{n}\boldsymbol{x_i x_i'}\psi^2(\bar{\epsilon}_i)$ and $\bar{\epsilon}_i=y_i-\boldsymbol{x_i'\bar{\beta}_n}$, $i\in\{1,\ldots,n\}$. Define the corresponding bootstrap pivot as

$$\boldsymbol{\breve{H}_n^*}=\sqrt{n}\boldsymbol{\Sigma_n^{*-1/2}}(\boldsymbol{\beta_n^*}-\boldsymbol{\bar{\beta}_n})$$

where $\boldsymbol{\Sigma}_{\boldsymbol{n}}^{*-\mathbf{1/2}} = \boldsymbol{A}_{\mathbf{2n}}^{*-\mathbf{1/2}}\boldsymbol{A}_{\mathbf{1n}}^{*}$ with $\epsilon_i^* = y_i - \boldsymbol{x_i'}\boldsymbol{\beta_n^*}$, $\boldsymbol{A}_{\mathbf{1n}}^{*} = n^{-1}\sum_{i=1}^{n}\boldsymbol{x_i}\boldsymbol{x_i'}\psi'(\epsilon_i^*)G_i^*$ and $\boldsymbol{A}_{\mathbf{2n}}^{*} = n^{-1}\sum_{i=1}^{n}\boldsymbol{x_i}\boldsymbol{x_i'}\psi^2(\epsilon_i^*)(G_i - \mu_{G^*})^2$, $i \in \{1, \ldots, n\}$.

Theorem 1.5.1 *Suppose, the assumptions* (A.1)$''$-(A.4)$''$ *and* (A.5)(i) *hold.*

(a) *Then there exist constant $C_5 > 0$ and a sequence of Borel sets $\boldsymbol{Q_{5n}} \subseteq \mathscr{R}^n$, such that $\mathbf{P}((\epsilon_1, \ldots, \epsilon_n) \in \boldsymbol{Q_{5n}}) \to 1$ as $n \to \infty$, and given $(\epsilon_1, \ldots, \epsilon_n) \in \boldsymbol{Q_{5n}}$, $n \geq C_5$ such that there exists a sequence of statistics $\{\boldsymbol{\beta_n^*}\}_{n\geq 1}$ such that*

$$\mathbf{P}_*\big(\boldsymbol{\beta_n^*} \text{ solves } (1.2.1) \text{ and } ||\boldsymbol{\beta_n^*} - \boldsymbol{\bar{\beta}_n}|| \leq C_5.n^{-1/2}.(logn)^{1/2}\big) \geq 1 - o\big(n^{-1/2}\big)$$

(b) *Suppose in addition* (A.5)(ii),(iii)$'$ *and* (A.8) *hold. Then there exist polynomial $\breve{a}_n^*(\cdot, \psi, G^*)$ depending on first three moments of G_1^* and on $\psi(\cdot)$, $\psi'(\cdot)$ & $\psi''(\cdot)$ through the residuals $\{\bar{\epsilon}_1, \ldots, \bar{\epsilon}_n\}$, such that given $(\epsilon_1, \ldots.., \epsilon_n) \in \boldsymbol{Q_{5n}}$, we have for $n \geq C_5$,*

$$\sup_{B\in\mathscr{B}} |\mathbf{P}_*(\boldsymbol{\breve{H}_n^*} \in B) - \int_B \breve{\xi}_n^*(\boldsymbol{x})d\boldsymbol{x}| \leq \delta_n n^{-1/2}$$

where $\breve{\xi}_n^*(\boldsymbol{x}) = (1 + n^{-1/2}\breve{a}_n^*(\boldsymbol{x}, \psi, G^*))\phi(\boldsymbol{x})$ and $\delta_n \equiv \delta_n(\epsilon_1, \ldots, \epsilon_n)$ tends to 0.

(c) *Suppose, in addition to the assumptions* (A.1)$''$-(A.4)$''$, (A.5)(i),(ii),(iii)$'$ *and* (A.8), (A.6)(i)$''$ *holds. Then, for the collection of Borel sets defined by (1.3.1),*

$$\sup_{B\in\mathscr{B}} \big|\mathbf{P}_*(\boldsymbol{\breve{H}_n^*} \in B) - \mathbf{P}(\boldsymbol{\breve{H}_n} \in B)\big| = o_p(n^{-1/2})$$

Remark 1.5.1 *The form of the studentized pivot $\boldsymbol{\breve{H}_n^*}$, defined for achieving second order correctness in non-iid case is different from $\boldsymbol{\tilde{H}_n^*}$, due to the difference in asymptotic variances of $\boldsymbol{\bar{\beta}_n}$ in two setups. In non-iid case, one cannot ignore computation of the negative square root of a matrix at each bootstrap iteration. But Theorem 1.5.1 is more general than Theorem 1.4.3 in the sense that it also includes the case when errors are iid. Note that $\boldsymbol{\bar{\Sigma}_n^*} = \boldsymbol{\bar{A}}_{\mathbf{1n}}^{*-\mathbf{1}}\boldsymbol{\bar{A}}_{\mathbf{2n}}^{*}\boldsymbol{\bar{A}}_{\mathbf{1n}}^{*-\mathbf{1}}$*

where $\bar{\boldsymbol{A}}^*_{\boldsymbol{1n}} = n^{-1}\sum_{i=1}^n \boldsymbol{x_i x_i'}\psi'(\epsilon_i^*)$ *and* $\bar{\boldsymbol{A}}^*_{\boldsymbol{2n}} = n^{-1}\sum_{i=1}^n \boldsymbol{x_i x_i'}\psi^2(\epsilon_i^*)$ *and* $\sigma_n^* = s_n^*\tau_n^{*-1}$ *where* $\tau_n^* = n^{-1}\sum_{i=1}^n \psi'(\epsilon_i^*)$, $s_n^{*2} = n^{-1}\sum_{i=1}^n \psi^2(\epsilon_i^*)$. *We need to modify* $\bar{\boldsymbol{\Sigma}}^*_{\boldsymbol{n}}$ *and* σ_n^* *to* $\boldsymbol{\Sigma}^*_{\boldsymbol{n}}$ *and* $\tilde{\sigma}_n^*$ *respectively to achieve second order correctness.*

Remark 1.5.2 *There is no difference in employing perturbation bootstrap and the usual residual bootstrap with respect to the accuracy of inference. Under some mild conditions, both are second order correct. But in view Theorem 1.5.1, the advantage of employing perturbation bootstrap instead of residual counterpart is evident when the errors are no longer identically distributed. Perturbation bootstrap continues to be S.O.C. in non-iid case without any modification, whereas a modification in the resampling stage is required for residual bootstrap to achieve the same. To see this consider the heteroscedastic simple linear regression model*

$$y_i = \beta x_i + \epsilon_i \tag{1.5.1}$$

where ϵ_i*'s are independent,* $\mathbf{E}\epsilon_i = 0$ *and* $\mathbf{E}\epsilon_i^2 = \sigma_i^2$. *The least square estimator of* β *is* $\hat{\beta} = \sum_{i=1}^n x_i y_i / \sum_{i=1}^n x_i^2$ *and hence* $\mathbf{Var}(\hat{\beta}) = \sum_{i=1}^n x_i^2\sigma_i^2/(\sum_{i=1}^n x_i^2)^2$. *The residual bootstrap samples are* $y_i^{**} = x_i\hat{\beta} + e_i^*$ *where* $\{e_1^*, \ldots, e_n^*\}$ *is a random sample from* $\{(e_1 - \bar{e}), \ldots, (e_n - \bar{e})\}$ *and* $e_i = y_i - x_i\hat{\beta}$, $i \in \{1, \ldots, n\}$ *are least square residuals. The residual bootstrapped least square estimator is* $\hat{\beta}^{**} = \sum_{i=1}^n x_i y_i^{**}/\sum_{i=1}^n x_i^2$. *Hence,* $\mathbf{Var}(\hat{\beta}^{**}|\epsilon_1, \ldots, \epsilon_n) = \sum_{i=1}^n (e_i - \bar{e})^2/\sum_{i=1}^n x_i^2$ *where* $n^{-1}\sum_{i=1}^n [(e_i - \bar{e})^2 - \sigma_i^2] \to 0$ *as* $n \to \infty$. *Thus* $\mathbf{Var}(\hat{\beta}^{**}|\epsilon_1, \ldots, \epsilon_n)$ *is not a consistent estimator of* $\mathbf{Var}(\hat{\beta})$*and hence residual bootstrap is not second order correct in approximating the distribution of* $\hat{\beta}$. *For details see Liu (1988). Liu (1988) also proposed a weighted bootstrap technique to achieve second order correctness for the model (1.6.28). On the other hand, if* $\hat{\beta}^*$ *is the perturbation bootstrapped least square estimator, then it is easy to show* $\mathbf{Var}(\hat{\beta}^*|\epsilon_1, \ldots, \epsilon_n) = \sum_{i=1}^n x_i^2\sigma_i^2/(\sum_{i=1}^n x_i^2)^2 + O_p(n^{-1})$. *Additionally, a centering adjustment is required in the definition of residual bootstrapped version in general*

M-estimation to achieve second order correctness even when the regression errors are iid [cf. Lahiri (1992)]; whereas in the perturbation bootstrap no adjustment is needed.

1.6 Proofs

First we define some notations. Throughout this section, $C, C_1, C_2, \ldots$ will denote generic constants that do not depend on the variables like n, x, and so on. For a non-negative integral 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 $|\boldsymbol{\alpha}| = \alpha_1 + \ldots + \alpha_l$, $\boldsymbol{\alpha}! = \alpha_1! \ldots \alpha_l!$, $f^{\boldsymbol{\alpha}} = (f_1^{\alpha_1}) \ldots (f_l^{\alpha_l})$. For $\boldsymbol{t} = (t_1, \ldots t_l)' \in \mathscr{R}^l$ and $\boldsymbol{\alpha}$ as above, define $t^{\boldsymbol{\alpha}} = t_1^{\alpha_1} \ldots t_l^{\alpha_l}$. The collection $\mathscr{B}$ will always be used to denote the collection of Borel subsets of $\mathscr{R}^p$ which satisfy (3.1). μ_{G^*} and $\sigma^2_{G^*}$ will respectively denote mean and variance of G_1^*. We want to mention here that only the important steps are presented in the proofs of the proposition and the theorems. For further details see the supplementary material Das and Lahiri. Although the proofs for second order results of perturbation bootstrap go through more or less same line as that for residual bootstrap in Lahiri (1992), the advantage in perturbation bootstrap is that the perturbing quantities are independent of the regression errors and hence it is much easier to obtain suitable stochastic approximation to the bootstrapped pivot and finally the EE than the same in case of residual bootstrap. On the negative side, in our proofs atleast we need Cramer's condition separately on regression errors and on the perturbing quantities [see assumptions (A.5) and (A.6)], whereas for residual bootstrap, one can derive a restricted Cramer's condition on resampled residuals from the Cramer's condition on regression errors to obtain second order correctness. Moreover, second order results can be established for residual bootstrap, after a modification, without any Cramer type condition in the case $p = 1$ [cf. Karabulut and

Lahiri (1997)]. It is of future research to see if similar conclusion can be drawn in case of perturbation bootstrap.

Before coming to the proofs we state some lemmas:

Lemma 1.6.1 *Suppose* Y_i, $i = 1, ..., n$ *are independent r.v's with* $\mathbf{E}(Y_1) = 0$, $\mathbf{E}(|Y_i|^t) < \infty$, $i = 1, ..., n$ *and* $\sum_{i=1}^{n} \mathbf{E}(|Y_i|^t) = \sigma_t$; $S_n = \sum_{i=1}^{n} Y_i$. *Then, for any* $t \geq 2$ *and* $x > 0$

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

proof :

The above inequality was proved in Fuk and Nagaev (1971).

Lemma 1.6.2 *Let,* $\{\boldsymbol{Y_i} = (Y_{i1}, Y_{i2})', 1 \leq i \leq n\}$ *be a collection of mean zero independent random vectors. Define, for some non random vectors* $\boldsymbol{l_{1i}}$ *and* $\boldsymbol{l_{2i}}$ *of dimensions* p_1 *and* p_2 *respectively with* $\sum_{i=1}^{n} \boldsymbol{l_{ji}} \boldsymbol{l'_{ji}} = \boldsymbol{I_{p_j}}$ *and* $\tilde{\gamma}_n = (\sum_{j=1}^{2} \sum_{i=1}^{n} ||\boldsymbol{l_{ji}}||^4)^{1/2} = O(n^{-1/2})$,

$$\boldsymbol{U_i} = (\boldsymbol{l'_{1i}} Y_{i1}, \boldsymbol{l'_{2i}} Y_{i2})', \qquad \boldsymbol{V_n} = \boldsymbol{Cov}\Big(\sum_{i=1}^{n} \boldsymbol{U_i}\Big), \qquad \boldsymbol{\tilde{U}_i} = \boldsymbol{V_n^{-1/2}} \boldsymbol{U_i}$$

for $1 \leq i \leq n$, *and* $\boldsymbol{S_n} = \sum_{i=1}^{n} \boldsymbol{\tilde{U}_i}$. *Let* $\boldsymbol{\tilde{\alpha}_n} = n^{-1} \sum_{i=1}^{n} \mathbf{E}||\boldsymbol{Y_i}||^3 I(||\boldsymbol{Y_i}||^2 > \lambda \tilde{\gamma}_n^{-1})$, *where* $I(\cdot)$ *is the indicator function and* λ *satisfies* $0 < \lambda < \liminf_{n\to\infty} \lambda_n$, $\lambda_i =$ *the smallest eigen value of* $\boldsymbol{\Sigma_i}$, $\boldsymbol{\Sigma_i} = \mathbf{Cov}(\boldsymbol{Y_i})$. *Assume that*

(a) *there exists a constant* k *such that* $n^{-1} \sum_{i=1}^{n} \mathbf{E}||\boldsymbol{Y_i}||^3 < k$ *for all* $n \geq 1$.

(b) $\tilde{\alpha}_n = o(1)$.

(c) *the characteristic function* g_n *of* $\boldsymbol{Y_n}$ *satisfies* $\limsup_{n\to\infty} \sup_{||(\boldsymbol{t})||>b} |g_n(\boldsymbol{t})| < 1$ *for all* $b > 0$.

Then for the class $\mathscr{B}$ of Borel sets satisfying (1.3.1),

$$\left|\mathbf{P}(\boldsymbol{S_n} \in B) - \int_B \check{\check{\xi}}_n(\boldsymbol{y})d\boldsymbol{y}\right| = o(\tilde{\gamma}_n)$$

where $k = p_1 + p_2$, $\check{\check{\xi}}_n(\boldsymbol{y})$ is the two term EE of the density of S_n.

proof :

The above Lemma follows from Theorem 20.6 of Bhattacharya and Rao (1986) and retracting the proofs of Lemma 3.1 of Lahiri (1992).

Lemma 1.6.3 *Suppose, $\{\boldsymbol{M_{0n}}\}_{n\geq 1}$, $\{\boldsymbol{M_{in}}\}_{n\geq 1}, i = 1, \ldots, p$ be $(p+1)$ sequence of matrices such that for each $n \geq 1$, $\boldsymbol{M_{0n}}$ is of order $p \times (p+r)$. and $\boldsymbol{M_{in}}, 1 \leq i \leq p$, are of order $(p+r) \times (p+r)$, $p \geq 1$, $r \geq 1$. Let, $k = p + r$, $\bar{\boldsymbol{M}}_{\boldsymbol{0n}} = [\boldsymbol{0} : \boldsymbol{I_r}]_{r\times k}$ and $\tilde{\boldsymbol{M}}_{\boldsymbol{0n}} = [\boldsymbol{M_{0n}}' : \bar{\boldsymbol{M}}_{\boldsymbol{0n}}{}']'$. Define the functions $g_n : \mathscr{R}^k \to \mathscr{R}^p$ by $g_n(\boldsymbol{x}) = \boldsymbol{M_{0n}x} + (\boldsymbol{x}'\boldsymbol{M_{1n}x}, \ldots, \boldsymbol{x}'\boldsymbol{M_{pn}x})'$, $\boldsymbol{x} \in \mathscr{R}^k$, $n \geq 1$. Assume that*

(a) *the hypothesis of Result 1.6.2 holds.*

(b) *$max\{||\boldsymbol{M_{in}}|| : 1 \leq i \leq p\} = O(\tilde{\gamma}_n)$ where $\tilde{\gamma}_n$ is as defined in Result 8.2.*

(c) *$||\boldsymbol{M_{0n}}|| = O(1)$, $\liminf_{n\uparrow\infty} \inf\{||\tilde{\boldsymbol{M}}_{\boldsymbol{0n}}\boldsymbol{u}|| : ||\boldsymbol{u}|| = 1, \boldsymbol{u} \in \mathscr{R}^k\} \geq \delta$ for some constant $\delta > 0$.*

Then for the class $\mathscr{B}$ of Borel sets satisfying (1.3.1),

$$\sup_{B\in\mathscr{B}} \left|\mathbf{P}(g_n(\boldsymbol{S_n}) \in B) - \int_B \mathring{\xi}_n(\boldsymbol{x})d\boldsymbol{x}\right| = o(\tilde{\gamma}_n) \quad as\ n \to \infty$$

where $\mathring{\xi}_n(.) = (1 + n^{-1/2}\mathring{a}(\cdot))\phi_{\mathring{\boldsymbol{D}}_{\boldsymbol{n}}}(\cdot)$, $\mathring{\boldsymbol{D}}_{\boldsymbol{n}} = \boldsymbol{M_{0n}M'_{0n}}$ and $\mathring{a}(\cdot)$ is a polynomial whose coefficients are continuous functions of $\mathbf{E}(\boldsymbol{Y_i})^{\boldsymbol{\alpha}}$, $|\boldsymbol{\alpha}| \leq 3$ and $i \in \{1, \ldots, n\}$.

proof :

The above result follows from the Lemma 1.6.2 and Lemma 3.2 of Lahiri (1992).

Lemma 1.6.4 *Under the assumptions* (A.1)-(A.3) *or* (A.1)$''$-(A.3)$''$, *it follows that* $\left(\sum_{i=1}^{n} ||\breve{\boldsymbol{v}}_{\boldsymbol{i}}||^4\right)^{1/2} = O_p(n^{-1/2})$.

proof :

Since, $\breve{\boldsymbol{v}}_{\boldsymbol{i}} = \boldsymbol{B}\bar{\boldsymbol{v}}_{\boldsymbol{i}}$ for each $i \in \{1, \ldots, n\}$, hence we have,

$$\sum_{i=1}^{n} ||\breve{\boldsymbol{v}}_{\boldsymbol{i}}||^4 = \sum_{i=1}^{n} ||\boldsymbol{B}\bar{\boldsymbol{v}}_{\boldsymbol{i}}||^4 \leq ||\boldsymbol{B}||^4 \sum_{i=1}^{n} ||\bar{\boldsymbol{v}}_{\boldsymbol{i}}||^4$$

Now if the canonical decomposition of $\sum_{i=1}^{n} \bar{\boldsymbol{v}}_{\boldsymbol{i}}\bar{\boldsymbol{v}}_{\boldsymbol{i}}'$ is compared with the spectral decomposition of the same, then it is evident that $\boldsymbol{B} = \boldsymbol{C}\boldsymbol{\Lambda}^{-1/2}$ where $\boldsymbol{\Lambda}$ is the $k \times k$ diagonal matrix with eigenvalues of $\sum_{i=1}^{n} \bar{\boldsymbol{v}}_{\boldsymbol{i}}\bar{\boldsymbol{v}}_{\boldsymbol{i}}'$ as the diagonal entries and $\boldsymbol{C}$ is the matrix of the orthonormal eigen-vectors of $n^{-1}\sum_{i=1}^{n} \bar{\boldsymbol{v}}_{\boldsymbol{i}}\bar{\boldsymbol{v}}_{\boldsymbol{i}}'$. Hence, due to assumption (A.2)(ii), for sufficiently large n, $||\boldsymbol{B}||^2 = O_p(n^{-1})$.

Again, if $\tilde{\boldsymbol{L}}$ is the first r columns of $\boldsymbol{L}^{-1}$,

$$\sum_{i=1}^{n} ||\bar{\boldsymbol{v}}_{\boldsymbol{i}}||^4 \leq C.\Big[\sum_{i=1}^{n} ||\boldsymbol{x}_{\boldsymbol{i}}||^4 \psi^4(\bar{\epsilon}_i) + \sum_{i=1}^{n} ||\boldsymbol{z}_{\boldsymbol{i}}||^4 \psi'^4(\bar{\epsilon}_i)\Big]$$

$$= O_p(n) + C.||\tilde{\boldsymbol{L}}||^4 \sum_{i=1}^{n} ||\tilde{\boldsymbol{z}}_{\boldsymbol{i}}||^4 \psi'^4(\bar{\epsilon}_i)$$

Now note that, $||\tilde{\boldsymbol{L}}(\sum_{i=1}^{n} \tilde{\boldsymbol{z}}_{\boldsymbol{i}}\tilde{\boldsymbol{z}}_{\boldsymbol{i}}')\tilde{\boldsymbol{L}}'|| = ||\sum_{i=1}^{n} \boldsymbol{z}_{\boldsymbol{i}}\boldsymbol{z}_{\boldsymbol{i}}'||$ and hence $||\tilde{\boldsymbol{L}}||^2 = O(\sum_{i=1}^{n} ||\boldsymbol{x}_{\boldsymbol{i}}||^4) = O(n)$

So, by assumption (A.2)(iii) or (A.2)(iii)$''$, we have $\sum_{i=1}^{n} ||\bar{\boldsymbol{v}}_{\boldsymbol{i}}||^4 = O_p(n)$.

Hence the lemma follows.

Lemma 1.6.5 *Under the assumptions* (A.2) (i) *and* (A.2) (iii) *or* (A.2) (iii)$''$, *the following is true.*

(a) $\left(\sum_{i=1}^{n}||\boldsymbol{d_i}||^6\right)^{1/4} + \left(\sum_{i=1}^{n}||\boldsymbol{d_i}||^4\right)^{1/2} = O(n^{-1/2})$.

(b) $\sum_{i=1}^{n}||\boldsymbol{x_i}||^j = O(n)$ *for* $j = 3, 4, 5, 6, 6+2\alpha$ *when the errors are iid and for* $j = 6+2\alpha, 3, \ldots, 12$ *when the errors are non-iid.*

proof :

This lemma follows from assumption (A.2) and by applying Hölders inequality.

Proof of Proposition 1.4.1. Suppose,

$$\sum_{i=1}^{n}\boldsymbol{x_i}\psi(y_i - \boldsymbol{x_i'}\boldsymbol{t_n^*})G_i^* = \boldsymbol{0}$$

Now, $\bar{\epsilon}_i = y_i - \boldsymbol{x_i'}\bar{\boldsymbol{\beta}}_n$ and $\sum_{i=1}^{n}\boldsymbol{x_i}\psi(\bar{\epsilon}_i) = 0$. Now writing $\epsilon_i^* = y_i - \boldsymbol{x_i'}\boldsymbol{t_n^*}$, by the Taylor's expansion of $\psi(\cdot)$ in the above equation we have

$$\sum_{i=1}^{n}\boldsymbol{x_i}\psi(\bar{\epsilon}_i)G_i^* + \sum_{i=1}^{n}\boldsymbol{x_i}\boldsymbol{x_i'}(\bar{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{t_n^*})\psi'(\bar{\epsilon}_i)G_i^* + \sum_{i=1}^{n}\boldsymbol{x_i}\frac{[\boldsymbol{x_i'}(\bar{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{t_n^*})]^2}{2}\psi''(u_i)G_i^* = 0 \quad (1.6.1)$$

where $|u_i - \bar{\epsilon}_i| \leq |\epsilon_i^* - \bar{\epsilon}_i|$ for each $i \in \{1, \ldots, n\}$.

Now (1.6.1) can be written as

$$L_n^*(\boldsymbol{t_n^*} - \bar{\boldsymbol{\beta}}_{\boldsymbol{n}}) = \Delta_n^* + R_n^* \quad (1.6.2)$$

where

$\Delta_n^* = n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i\psi(\bar{\epsilon}_i)(G_i^* - \mu_{G^*})$

$L_n^* = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i'\psi'(\bar{\epsilon}_i)G_i^*$

$\mathbf{E}_* L_n^* = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i'\psi'(\bar{\epsilon}_i)\mu_{G^*}$

$R_n^* = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\dfrac{[\boldsymbol{x}_i'(\bar{\boldsymbol{\beta}}_n - \boldsymbol{t}_n^*)]^2}{2}\psi''(u_i)G_i^*$

Since ψ'' is Lipschitz of order α, so we have

$$|\psi''(u_i) - \psi''(\bar{\epsilon}_i)| \leq |u_i - \bar{\epsilon}_i|^\alpha \leq |\epsilon_i^* - \bar{\epsilon}_i|^\alpha \leq ||x_i||^\alpha||t_n^* - \bar{\beta}_n||^\alpha$$

and hence,

$$2||R_n^*|| \leq \frac{1}{n}\sum_{i=1}^n[||\boldsymbol{x_i}||^3|\psi''(\bar{\epsilon}_i)|||\boldsymbol{t_n^*} - \bar{\boldsymbol{\beta}}_{\boldsymbol{n}}||^2G_i^* + ||\boldsymbol{x_i}||^{3+\alpha}||\boldsymbol{t_n^*} - \bar{\boldsymbol{\beta}}_{\boldsymbol{n}}||^{2+\alpha}G_i^*]$$

Therefore by Lemma 1.6.1, it can be said that there exist a constant $C_5 > 0$ and a sequence of Borel sets $\{\boldsymbol{Q_{5n}}\}$ such that $\mathbf{P}(\boldsymbol{Q_{5n}}) \to 1$ and for $(\epsilon_1, ..., \epsilon_n) \in \boldsymbol{Q_{5n}}$ and $n \geq C_5$

$$n^{-1}\sum_{i=1}^n ||\boldsymbol{x_i}||^3|\psi''(\bar{\epsilon}_i)|\mu_{G^*} \leq C_1$$

and for any fixed $0 < \epsilon < 1$,

$$\mathbf{P}_*\Big(\big|\sum_{i=1}^n ||\boldsymbol{x_i}||^3|\psi''(\bar{\epsilon}_i)|(G_i^* - EG_i^*)\big| + \big|\sum_{i=1}^n ||\boldsymbol{x_i}||^{3+\alpha}(G_i^* - EG_i^*)\big| > n\epsilon\Big) = o(n^{-1/2}) \tag{1.6.3}$$

Again note that for any $\epsilon > 0$,

$$\mathbf{P}_*\Big(||L_n^* - \mathbf{E}_* L_n^*|| > \epsilon\Big)$$
$$\leq \sum_{j,k=1}^p \mathbf{P}_*\Big(\big|\sum_{i=1}^n x_{ij}x_{ik}\psi'(\bar{\epsilon}_i)(G_i^* - EG_i^*)\big| > n\epsilon\Big)$$

Now, using Taylor expansion of ψ', Lipschitz property of $\psi''(\cdot)$ and Lemma 1.6.1, we have as $n \to \infty$,

$$||\mathbf{E}_* L_n^*|| = ||\boldsymbol{A_1}||.|E\psi'(\epsilon)|\mu_{G^*} + o_p(1)$$

where $\boldsymbol{A_1}$ is a $p \times p$ matrix as defined in the assumptions and for any $0 < \epsilon < 1$ and any $j, k \in \{1, \ldots, p\}$, by Lemma 1.6.1 and 1.6.5 imply,

$$\mathbf{P}_*\Big(\big|\sum_{i=1}^{n} x_{ij}x_{ik}\psi'(\bar{\epsilon}_i)(G_i^* - EG_i^*)\big| > n\epsilon\Big) = o_p(n^{-1/2})$$

Hence, there exist constant $C_6 > 0$ and a sequence of Borel sets $\boldsymbol{Q_{6n}} \subseteq \mathscr{R}^n$, such that $\mathbf{P}((\epsilon_1, ..., \epsilon_n) \in \boldsymbol{Q_{6n}}) \to 1$ as $n \to \infty$, and given $(\epsilon_1, ..., \epsilon_n) \in \boldsymbol{Q_{6n}}$ and $n \geq C_6$,

$$\mathbf{P}_*\big(L_n^* \text{ is invertible}\big) = 1 - o(n^{-1/2}) \tag{1.6.4}$$

Also, by Lemma 1.6.1 and 1.6.5, using Taylor expansion of ψ and the fact that $\psi''(\cdot)$ is Lipschitz, we can say that there exist constant $C_7 > 0$ and a sequence of Borel sets $\boldsymbol{Q_{7n}} \subseteq \mathscr{R}^n$, such that $\mathbf{P}((\epsilon_1, ..., \epsilon_n) \in \boldsymbol{Q_{7n}}) \to 1$ as $n \to \infty$, and given $(\epsilon_1, ..., \epsilon_n) \in \boldsymbol{Q_{7n}}$ and $n \geq C_7$,

$$\mathbf{P}_*(||\Delta_n^*|| > C.n^{-1/2}(logn)^{1/2}) = o(n^{-1/2}) \tag{1.6.5}$$

Therefore, (1.6.3), (1.6.4) and (1.6.5) imply that on the set $\boldsymbol{Q_{1n}} = \boldsymbol{Q_{5n}} \cap \boldsymbol{Q_{6n}} \cap \boldsymbol{Q_{7n}}$ there exists constant $C_1 > 0$ such that for $n \geq C_1$, (8.3) can be written as

$$(\boldsymbol{t_n^*} - \boldsymbol{\bar{\beta}_n}) = f_n(\boldsymbol{t_n^*} - \boldsymbol{\bar{\beta}_n})$$

where f_n is a continuous function from $\mathscr{R}^p$ to R^p satisfying $\mathbf{P}_*(||f_n(\boldsymbol{t_n^*} - \boldsymbol{\bar{\beta}_n})|| \leq C_1.n^{-1/2}\,(logn)^{1/2}) = 1 - o(n^{-1/2})$ as $n \to \infty$ whenever $||\boldsymbol{t_n^*} - \boldsymbol{\bar{\beta}_n}|| \leq C_1.n^{-1/2}(logn)^{1/2}$.

Hence, proposition 1.4.1 follows by Browder's fixed point theorem.

Proof of Theorem 1.4.1. Suppose, the sequence of statistics $\{\boldsymbol{\beta_n^*}\}_{n\geq 1}$ satisfies Proposition 1.4.1. Then there exists constant $C_8 > 0$ such that for $n \geq C_8$, on a set $\boldsymbol{Q_{8n}}$ of probability converging to 1, (1.6.2) can be written as

$$\sqrt{n}(\boldsymbol{\beta_n^*} - \boldsymbol{\bar{\beta}_n}) = L_n^{*-1}\sqrt{n}[\Delta_n^* + \tilde{\xi}_n^* + R_{1n}^*] \tag{1.6.6}$$

where Δ_n^* and L_n^* are as defined in the proof of the proposition
and $\tilde{\xi}_n^* = n^{-1}\sum_{i=1}^n \boldsymbol{x_i}\dfrac{[\boldsymbol{x_i'}(\boldsymbol{\beta_n^*} - \boldsymbol{\bar{\beta}_n})]^2}{2}\psi''(\bar{\epsilon}_i)G_i^*$ and
$||R_{1n}^*|| \leq ||\boldsymbol{\beta_n^*} - \boldsymbol{\bar{\beta}_n}||^{2+\alpha} n^{-1}\sum_{i=1}^n ||\boldsymbol{x_i}||^{3+\alpha}G_i^*$

Hence, by Markov inequality as $n \to \infty$,

$$\mathbf{P}_*(||R_{1n}^*|| > C.n^{-(2+\alpha)/2}(logn)^{(2+\alpha)/2}) = o_p(n^{-1/2}) \tag{1.6.7}$$

Also, (1.6.6) can be written as

$$\sqrt{n}(\boldsymbol{\beta_n^*} - \boldsymbol{\bar{\beta}_n}) = L_n^{*-1}\sqrt{n}\Delta_n^* + R_{2n}^* \tag{1.6.8}$$

where from (1.6.7) and the proposition, using Lemma 1.6.1, we can say that as $n \to \infty$,

$$\mathbf{P}_*(||R_{2n}^*|| > C.n^{-1/2}(logn)) = o_p(n^{-1/2}) \tag{1.6.9}$$

Again from 1.6.6,

$$\sqrt{n}(\boldsymbol{\beta_n^*} - \boldsymbol{\bar{\beta}_n}) = L_n^{*-1}\Big[\sqrt{n}\Delta_n^* + \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{x_i}\frac{[\boldsymbol{x_i'}(L_n^{*-1}\Delta_n^*)]^2}{2}\psi''(\bar{\epsilon}_i)G_i^*\Big] + R_{3n}^* \tag{1.6.10}$$

where

$$R_{3n}^* = L_n^{*-1}\sqrt{n}R_{1n}^* + n^{-3/2}\sum_{i=1}^n \frac{[\boldsymbol{x_i'}(R_{2n}^*)]^2}{2}\psi''(\bar{\epsilon}_i)G_i^* + \\ n^{-1}\sum_{i=1}^n \boldsymbol{x_i}[(\boldsymbol{x_i'}(L_n^{*-1}\Delta_n^*))(\boldsymbol{x_i'}R_{2n}^*)]\psi''(\bar{\epsilon}_i)G_i^*$$

By (1.6.7) and using Lemma 1.6.1, it follows that

$$\mathbf{P}_*(||R^*_{3n}|| = o(n^{-1/2})) = 1 - o_p(n^{-1/2}) \tag{1.6.11}$$

Again it is straightforward that

$$L_n^{*-1} = (\mathbf{E}_* L_n^*)^{-1} + W_n^* + \tilde{Z}_n^* \tag{1.6.12}$$

Where

$W_n^* = (\mathbf{E}_* L_n^*)^{-1}(\mathbf{E}_* L_n^* - L_n^*)(\mathbf{E}_* L_n^*)^{-1}$

$\tilde{Z}_n^* = (\mathbf{E}_* L_n^*)^{-1}(\mathbf{E}_* L_n^* - L_n^*)(\mathbf{E}_* L_n^*)^{-1}(\mathbf{E}_* L_n^* - L_n^*)L_n^{*-1}$

$\mathbf{E}_* L_n^* = n^{-1}\sum_{i=1}^n \boldsymbol{x_i x_i'}\psi'(\bar{\epsilon}_i)\mu_{G^*} = \boldsymbol{\bar{A}_{1n}}\mu_{G^*}$.

By Lemma 1.6.1 and the Libschitz property of $\psi''(\cdot)$, it can be shown that

$$\mathbf{P}_*(||L_n^* - \mathbf{E}_* L_n^*|| > C.n^{-1/4}(logn)^{-1/2}) = o_p(n^{-1/2})$$

and hence,

$$\mathbf{P}_*(||\tilde{Z}_n^*|| > C.n^{-1/2}(logn)^{-1}) \leq \mathbf{P}_*(||W_n^*|| > C.n^{-1/4}(logn)^{-1/2}) = o_p(n^{-1/2}) \tag{1.6.13}$$

Since, as $n \to \infty$, $||\mathbf{E}_* L_n^*|| = ||\boldsymbol{A_1}||.|E\psi'(\epsilon)|\mu_{G^*} + o_p(1)$, as mentioned earlier.

Again $\boldsymbol{\bar{\Sigma}_n} = O_p(1)$. So, from (1.6.6)-(1.6.13), it follows that there exist constant $C_9 > 0$ and a sequence of Borel sets $\boldsymbol{Q_{9n}} \subseteq \boldsymbol{Q_{8n}}$, such that $\mathbf{P}((\epsilon_1, \ldots\ldots, \epsilon_n) \in \boldsymbol{Q_{9n}}) \to 1$ as $n \to \infty$, and given $(\epsilon_1, \ldots.., \epsilon_n) \in \boldsymbol{Q_{9n}}$, for $n \geq C_9$,

$$\boldsymbol{F_n^*} = \sqrt{n}\boldsymbol{\bar{\Sigma}_n}\left[(\mathbf{E}_* L_n^*)^{-1}\Delta_n^* + W_n^*\Delta_n^* + (\mathbf{E}_* L_n^*)^{-1}\chi_n^*\right] + R_{4n}^*$$

where $\chi_n^* = n^{-1}\sum_{i=1}^n \left[\boldsymbol{x_i}\frac{[\boldsymbol{x_i'}((\mathbf{E}_* L_n^*)^{-1}\Delta_n^*)]^2}{2}\psi''(\bar{\epsilon}_i)\right]\mu_{G^*}$ and

$$\mathbf{P}_*(||R_{4n}^*|| = o(n^{-1/2})) = 1 - o(n^{-1/2})$$

Therefore, by arguments similar to (4.12) of Qumsiyeh (1990a), we have

$$\sup_{B\in\mathscr{B}} |\mathbf{P}_*(\boldsymbol{F_n^*} \in B) - \mathbf{P}_*(\boldsymbol{U_n^*} \in B)| = o_p(n^{-1/2}) \tag{1.6.14}$$

where $\boldsymbol{U_n^*} = \sqrt{n}\boldsymbol{\bar{\Sigma}_n}\left[(\mathbf{E}_* L_n^*)^{-1}\Delta_n^* + W_n^*\Delta_n^* + (\mathbf{E}_* L_n^*)^{-1}\chi_n^*\right]$.

Now, for all $1 \le i \le n$, define,

$Y_i^* = (G_i^* - \mu_{G^*})$, $\boldsymbol{X_i^*} = \boldsymbol{\breve{v}_i} Y_i^*$, $\boldsymbol{V_n^*} = \sum_{i=1}^n \mathbf{Cov}_*(\boldsymbol{X_i^*})$, $\boldsymbol{\tilde{X}_i^*} = \boldsymbol{V_n^{*-1/2}}\boldsymbol{X_i^*}$ and $\boldsymbol{S_n^*} = \sum_{i=1}^n \boldsymbol{\tilde{X}_i^*}$.

Partition $\boldsymbol{B}^{-1}$ as $\boldsymbol{B}^{-1} = (\boldsymbol{\bar{B}_1'}, \boldsymbol{\bar{B}_2'})'$ where $\boldsymbol{\bar{B}}_1$ is of order $p \times k$.

Hence, $\boldsymbol{\bar{B}_1 V_n^{*1/2} \tilde{X}_i^*} = \boldsymbol{\bar{x}_i} Y_i^*$ and $\boldsymbol{\bar{B}_2 V_n^{*1/2} \tilde{X}_i^*} = \boldsymbol{\bar{z}_i} Y_i^*$.

Note that

$$\sqrt{n}\boldsymbol{\bar{\Sigma}_n}(\mathbf{E}_* L_n^*)^{-1}\Delta_n^* = \frac{\boldsymbol{\bar{A}_{2n}}^{-1/2}\mu_{G^*}^{-1}}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{x_i}\psi(\bar{\epsilon}_i)(G_i^* - \mu_{G^*}) = \boldsymbol{T_n^* S_n^*} \tag{1.6.15}$$

where $\boldsymbol{T_n^*} = \dfrac{\boldsymbol{\bar{A}_{2n}}^{-1/2}\mu_{G^*}^{-1}\boldsymbol{\bar{B}_1 V_n}^{*1/2}}{\sqrt{n}}$

For any $j \in \{1, \ldots, p\}$, jth row of $(\mathbf{E}_* L_n^* - L_n^*)$ is

$$-n^{-1}\sum_1^n (x_{ij}x_{i1}, \ldots, x_{ij}x_{ip})\psi'(\bar{\epsilon})Y_i^* = (n^{-1}\sum_1^n \boldsymbol{\bar{z}_i'} Y_i^*)\boldsymbol{E_{jn}^*} = n^{-1}\boldsymbol{S_n^{*\prime} V_n^{*1/2}\bar{B}_2' E_{jn}^*}$$

for some matrices $\boldsymbol{E_{jn}^*}$ each of order $q \times p$, $j \in \{1, \ldots, p\}$ and $||\boldsymbol{E_{jn}^*}|| \le q$.

Hence,

$$W_n^* \sqrt{n}\Delta_n^* = n^{-1}\bar{\boldsymbol{A}}_{\boldsymbol{2n}}{}^{-1/2}\mu_{G^*}^{-1}(\boldsymbol{S_n^{*\prime}}\boldsymbol{V_n^{*1/2}}\bar{\boldsymbol{B}}_2'\boldsymbol{E}_{\boldsymbol{1n}}^*, \ldots, \boldsymbol{S_n^{*\prime}}\boldsymbol{V_n^{*1/2}}\bar{\boldsymbol{B}}_{\boldsymbol{2}}'\boldsymbol{E}_{\boldsymbol{pn}}^*)'\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}{}^{1/2}\boldsymbol{T_n^*}\boldsymbol{S_n^*}$$

$$= \mu_{G^*}^{-1}\sum_{j=1}^{p} h_{jn}\boldsymbol{S_n^{*\prime}}\tilde{\boldsymbol{E}}_{\boldsymbol{jn}}^*\boldsymbol{S_n^*} \tag{1.6.16}$$

for some matrices $\tilde{\boldsymbol{E}}_{\boldsymbol{jn}}^*$, each of order $k \times k$, where

$\tilde{\boldsymbol{E}}_{\boldsymbol{jn}}^* = n^{-1}\boldsymbol{V_n^{*1/2}}\bar{\boldsymbol{B}}_{\boldsymbol{2}}'\boldsymbol{E}_{\boldsymbol{jn}}^*\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{\boldsymbol{1/2}}\boldsymbol{T_n^*}$ and $\bar{\boldsymbol{A}}_{\boldsymbol{2n}}{}^{-1/2} = (\boldsymbol{h_{1n}}, \ldots, \boldsymbol{h_{pn}}), \boldsymbol{h_{jn}} = (h_{1jn}, \cdots, h_{pjn})$.

Again,

$$\sqrt{n}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}{}^{1/2}(\mathbf{E}_* L_n^*)^{-1}\xi_n^* = n^{-1/2}\mu_{G^*}^{-1}\bar{\boldsymbol{A}}_{\boldsymbol{2n}}{}^{-1/2}\sum_{i=1}^{n}\boldsymbol{x_i}\frac{[\boldsymbol{x_i'}((\mathbf{E}_* L_n^*)^{-1}\Delta_n^*)]^2}{2}\psi''(\bar{\epsilon}_i)\mu_{G^*}$$

$$= 2^{-1}\bar{\boldsymbol{A}}_{\boldsymbol{2n}}{}^{-1/2}\sum_{i=1}^{n}\boldsymbol{x_i}\psi''(\bar{\epsilon}_i)\boldsymbol{S_n^{*\prime}}\boldsymbol{P_{in}^*}\boldsymbol{S_n^*}$$

where $\boldsymbol{P_{in}^*} = \dfrac{\boldsymbol{T_n^{*\prime}}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{\boldsymbol{1/2}}\boldsymbol{x_i}\boldsymbol{x_i'}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{\boldsymbol{1/2}}\boldsymbol{T_n^*}}{n\sqrt{n}}$

Therefore,

$$\boldsymbol{U_n^*} = \boldsymbol{M_{0n}^*}\boldsymbol{S_n^*} + (\boldsymbol{S_n^{*\prime}}\boldsymbol{M_{1n}^*}\boldsymbol{S_n^*}, \ldots, \boldsymbol{S_n^{*\prime}}\boldsymbol{M_{pn}^*}\boldsymbol{S_n^*})' \tag{1.6.17}$$

where

$\boldsymbol{M_{0n}^*} = \boldsymbol{T_n^*}$ and for each $j \in \{1, \ldots, p\}$,

$\boldsymbol{M_{jn}^*} = \mu_{G^*}^{-1}\sum_{k=1}^{p} h_{jkn}\tilde{\boldsymbol{E}}_{\boldsymbol{kn}}^* + 2^{-1}\sum_{i=1}^{n}\boldsymbol{h_{jn}'}\boldsymbol{x_i}\psi''(\bar{\epsilon}_i)\boldsymbol{P_{in}^*}$.

Now note that $||\boldsymbol{M_{0n}^*}|| = O_p(1)$ since $\bar{\boldsymbol{A}}_{\boldsymbol{2n}} \asymp_p \sigma_{G^*}^2\boldsymbol{A_2}$, $\boldsymbol{V_n^*} = \sum_{i=1}^{n}\mathbf{Cov}_*(\boldsymbol{X_i^*}) = \sigma_{G^*}^2\sum_{i=1}^{n}\breve{\boldsymbol{v}}_{\boldsymbol{i}}\tilde{\bar{\boldsymbol{v}}}_{\boldsymbol{i}}' = \sigma_{G^*}^2\boldsymbol{I_k}$ and

$||\bar{\boldsymbol{B}}_{\boldsymbol{1}}||^2 \leq ||\boldsymbol{B^{-1}}||^2 \leq \sum_{i=1}^{n}||\bar{\boldsymbol{v}}_{\boldsymbol{i}}||^2 = \sum_{i=1}^{n}||\boldsymbol{x_i}||^2\psi^2(\bar{\epsilon}_i) + \sum_{i=1}^{n}||\boldsymbol{z_i}||^2\psi'^2(\bar{\epsilon}_i) = O_p(n)$.

Also by Holder's inequality and the fact that $||\boldsymbol{M}_{\boldsymbol{0n}}^*|| = O_p(1)$ and $\sum_{i=1}^n ||\boldsymbol{d_i}||^4 = O(n^{-1})$, it follows that $||\sum_{i=1}^n \boldsymbol{x_i}\psi''(\bar{\epsilon}_i)\boldsymbol{P}_{\boldsymbol{in}}^*|| \leq n^{-3/2}\sum_{i=1}^n |\psi''(\bar{\epsilon}_i)|||\boldsymbol{x_i}||^3||\boldsymbol{T}_{\boldsymbol{n}}^*||^2 = O_p(n^{-1/2})$. Again, $||\sum_{j=1}^p \boldsymbol{h_{jn}}\tilde{\boldsymbol{E}}_{\boldsymbol{jn}}^*|| = O_p(n^{-1/2})$ since $||\bar{\boldsymbol{B}}_{\boldsymbol{2}}|| \leq ||\boldsymbol{B}^{-1}||^2 = O_p(n)$. All of These together imply that $||\boldsymbol{M}_{\boldsymbol{jn}}^*|| = O_p(n^{1/2})$ for all $j \in \{1, \ldots, p\}$.

Therefore, by Lemma 1.6.2 and Lemma 1.6.3,

$$\sup_{B\in\mathscr{B}} |\mathbf{P}_*(\boldsymbol{U}_{\boldsymbol{n}}^* \in B) - \int_B \xi_n^*(\boldsymbol{x})d\boldsymbol{x}| = o_p(n^{-1/2}) \quad as\ n \to \infty \tag{1.6.18}$$

The first three conditional cumulants of $\boldsymbol{t'U_n^*}$ are given by

$\kappa_1(\boldsymbol{t'U_n^*}) = \mathbf{E}_*(\boldsymbol{t'U_n^*}) = n^{-1/2}\sum_{|\boldsymbol{\nu}|=1} b_{11}^{*(\boldsymbol{\nu})}\boldsymbol{t}^{\boldsymbol{\nu}}$

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

$\kappa_3(\boldsymbol{t'U_n^*}) = \mathbf{E}_*(\boldsymbol{t'U_n^*})^3 - 3\mathbf{E}_*(\boldsymbol{t'U_n^*})^2.\mathbf{E}_*(\boldsymbol{t'U_n^*}) + 2(\mathbf{E}_*(\boldsymbol{t'U_n^*}))^3 = n^{-1/2}\sum_{|\boldsymbol{\nu}|=3}\dfrac{3!}{\boldsymbol{\nu}!}b_{31}^{*(\boldsymbol{\nu})}$
$\boldsymbol{t}^{\boldsymbol{\nu}} + o_p(n^{-1/2})$.

Hence, using the transformation techniques of Bhattacharya and Ghosh (1978), we have,

$$\xi_n^*(\boldsymbol{x}) = \left[1 - n^{-1/2}\Big\{\sum_{|\boldsymbol{\nu}|=1} b_{11}^{*(\boldsymbol{\nu})}D^{\boldsymbol{\nu}} + \sum_{|\boldsymbol{\nu}|=3}\frac{b_{31}^{*(\boldsymbol{\nu})}}{\boldsymbol{\nu}!}D^{\boldsymbol{\nu}}\Big\}\right]\phi(\boldsymbol{x})$$

where if $\boldsymbol{\nu}_1$ is a $p \times 1$ vector with all the elements being 0, except the jth one and $\boldsymbol{\nu}_2$ is a $p \times 1$ vector with all the elements being 0, except the j_1, j_2 and j_3 positions then

$$\begin{aligned} b_{11}^{*(\boldsymbol{\nu}_1)} &= \sum_{k=1}^p h_{jkn}\Big(n^{-1}\sum_{i=1}^n \big[\boldsymbol{z_i'E_{kn}^*\bar{A}_{1n}^{-1}x_i}\psi(\bar{\epsilon}_i)\psi'(\bar{\epsilon}_i)\big]\Big) \\ &\quad + (2n)^{-1}\sum_{i=1}^n a_{jin}^*\boldsymbol{x_i'\bar{A}_{1n}^{-1}\bar{A}_{2n}\bar{A}_{1n}^{-1}x_i}\psi''(\bar{\epsilon}_i) \end{aligned} \tag{1.6.19}$$

$$
\begin{aligned}
b_{31}^{*(\boldsymbol{\nu}_2)} =& n^{-1}\sum_{i=1}^{n}\left[\Big(\prod_{m=1}^{3}a^*_{j_m in}\Big)\psi^3(\bar{\epsilon}_i)\right] \\
&+2n^{-2}\sum_{i,j=1}^{n}\Big[a^*_{j_1 in}a^*_{j_2 in}\big(\sum_{k=1}^{p}h_{j_3 kn}\boldsymbol{z_i'E^*_{kn}\bar{A}^{-1}_{1n}x_j}\big)\psi^2(\bar{\epsilon}_i)\psi(\bar{\epsilon}_j)\psi'(\bar{\epsilon}_j)\Big] \\
&+2n^{-2}\sum_{i,j=1}^{n}\Big[a^*_{j_1 in}a^*_{j_3 in}\big(\sum_{k=1}^{p}h_{j_2 kn}\boldsymbol{z_i'E^*_{kn}\bar{A}^{-1}_{1n}x_j}\big)\psi^2(\bar{\epsilon}_i)\psi(\bar{\epsilon}_j)\psi'(\bar{\epsilon}_j)\Big] \\
&+2n^{-2}\sum_{i,j=1}^{n}\Big[a^*_{j_2 in}a^*_{j_3 in}\big(\sum_{k=1}^{p}h_{j_1 kn}\boldsymbol{z_i'E^*_{kn}\bar{A}^{-1}_{1n}x_j}\big)\psi^2(\bar{\epsilon}_i)\psi(\bar{\epsilon}_j)\psi'(\bar{\epsilon}_j)\Big] \\
&+3n^{-3}\sum_{i,j,l=1}^{n}a^*_{j_1 in}a^*_{j_2 in}a^*_{j_3 in}\big(\boldsymbol{x_j'\bar{A}^{-1}_{1n}x_l x_l'\bar{A}^{-1}_{1n}x_i}\big)\psi''(\bar{\epsilon}_l)\psi^2(\bar{\epsilon}_i)\psi^2(\bar{\epsilon}_j) \qquad (1.6.20)
\end{aligned}
$$

where $\boldsymbol{h'_{jn}x_i} = a^*_{jin}, j \in \{1,\cdots,p\}, i \in \{1,\ldots,n\}$.

Now, for the original standardized M-estimator $\boldsymbol{F_n}$, it can be shown that there exists a linear stochastic approximation $\boldsymbol{U_n}$ such that $\boldsymbol{F_n} = \boldsymbol{U_n} + R_n$ where as $n \to \infty$,

$$
\mathbf{P}(||R_n|| = o(n^{-1/2})) = 1 - o(n^{-1/2})
$$

and

$$
\sup_{B\in\mathscr{B}} |\mathbf{P}(\boldsymbol{U_n} \in B) - \int_B \xi_n(\boldsymbol{x})d\boldsymbol{x}| = o(n^{-1/2})
$$

Similar to the bootstrapped standardized case, it can be shown that

$$
\xi_n(\boldsymbol{x}) = \left[1 - n^{-1/2}\Big\{\sum_{|\boldsymbol{\nu}|=1} b_{11}^{(\boldsymbol{\nu})}D^{\boldsymbol{\nu}} + \sum_{|\boldsymbol{\nu}|=3}\frac{b_{31}^{(\boldsymbol{\nu})}}{\boldsymbol{\nu}!}D^{\boldsymbol{\nu}}\Big\}\right]\phi(\boldsymbol{x})
$$

where

$$b_{11}^{(\boldsymbol{\nu}_1)} = s^{-1}\tau^{-1}\sum_{k=1}^{p} A_{jkn}^{-1/2}\Big(n^{-1}\sum_{i=1}^{n}\big[\boldsymbol{z_i' E_{kn}^* A_n^{-1} x_i}\big]\Big)\mathbf{E}\psi(\epsilon_1)\psi'(\epsilon_1)$$

$$+ s\tau^{-2}\Big[(2n)^{-1}\sum_{i=1}^{n} a_{jin}\boldsymbol{x_i' A_n^{-1} x_i}\Big]\mathbf{E}\psi''(\epsilon_1) \qquad (1.6.21)$$

$$b_{31}^{(\boldsymbol{\nu}_2)} = s^{-3}\Big[n^{-1}\sum_{i=1}^{n}\Big[\prod_{m=1}^{3} a_{j_m in}\Big]\Big]\mathbf{E}\psi^3(\epsilon_1)$$

$$+ s^{-1}\tau^{-1}\Big[2n^{-2}\sum_{i,j=1}^{n} a_{j_1 in}a_{j_2 in}\big(\sum_{k=1}^{p} A_{jkn}^{-1/2}\boldsymbol{z_i' E_{kn}^* A_n^{-1} x_j}\big)\Big]\mathbf{E}\psi(\epsilon_1)\psi'(\epsilon_1)$$

$$+ s^{-1}\tau^{-1}\Big[2n^{-2}\sum_{i,j=1}^{n} a_{j_1 in}a_{j_3 in}\big(\sum_{k=1}^{p} A_{jkn}^{-1/2}\boldsymbol{z_i' E_{kn}^* A_n^{-1} x_j}\big)\Big]\mathbf{E}\psi(\epsilon_1)\psi'(\epsilon_1)$$

$$+ s^{-1}\tau^{-1}\Big[2n^{-2}\sum_{i,j=1}^{n} a_{j_2 in}a_{j_3 in}\big(\sum_{k=1}^{p} A_{jkn}^{-1/2}\boldsymbol{z_i' E_{kn}^* A_n^{-1} x_j}\big)\Big]\mathbf{E}\psi(\epsilon_1)\psi'(\epsilon_1)$$

$$+ s\tau^{-2}\Big[3n^{-3}\sum_{i,j,l=1}^{n} a_{j_1 in}a_{j_2 in}a_{j_3 in}\big(\boldsymbol{x_j' A_n^{-1} x_l x_l' A_n^{-1} x_i}\big)\Big]\mathbf{E}\psi''(\epsilon_1) \qquad (1.6.22)$$

where $\boldsymbol{A_n^{-1/2}} = \Big(\boldsymbol{A_{1n}^{-1/2}}, \ldots, \boldsymbol{A_{pn}^{-1/2}}\Big)$, $\boldsymbol{A_{jn}^{-1/2}} = \Big(A_{j1n}^{-1/2}, \ldots, A_{jpn}^{-1/2}\Big)'$, $(\boldsymbol{A_{jn}^{-1/2}})'\boldsymbol{x_i} = \boldsymbol{e_j' A_n^{-1/2} x_i} = a_{jin}$, $j \in \{1, \cdots, p\}, i \in \{1, \ldots, n\}$.

Since $\big|b_{11}^{*(\boldsymbol{\nu}_1)} - b_{11}^{(\boldsymbol{\nu}_1)}\big| + \big|b_{31}^{*(\boldsymbol{\nu}_2)} - b_{31}^{(\boldsymbol{\nu}_2)}\big| \to 0$ in probability as $n \to \infty$ for all $p \times 1$ vectors $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ with $|\boldsymbol{\nu}_1| = 1$ and $|\boldsymbol{\nu}_2| = 3$, hence, for $n > N$, N being a fixed natural number, we have for any $\eta > 0$

$$\mathbf{P}\Big(\sup_{B\in\mathscr{B}}\big|\mathbf{P}_*(\boldsymbol{F_n^*} \in B) - \mathbf{P}(\boldsymbol{F_n} \in B)\big| > \eta . n^{-1/2}\Big)$$

$$\leq \mathbf{P}\Big(\sup_{B\in\mathscr{B}}\Big|\int_B \xi_n^*(\boldsymbol{x})d\boldsymbol{x} - \int_B \xi_n(\boldsymbol{x})d\boldsymbol{x}\Big| > \frac{\eta}{2}.n^{-1/2}\Big) + o(1)$$

$$\leq \mathbf{P}\Big(n^{-1/2}\sum_{|\boldsymbol{\nu}|=1}|b_{11}^{*(\boldsymbol{\nu})} - b_{11}^{(\boldsymbol{\nu})}| + n^{-1/2}\sum_{|\boldsymbol{\nu}|=3}|b_{31}^{*(\boldsymbol{\nu})} - b_{31}^{(\boldsymbol{\nu})}| > \frac{\eta}{2}.n^{-1/2}\Big) + o(1)$$

$$\to 0 \quad \text{as } n \to \infty$$

Therefore, Theorem 1.4.1 follows.

Proof of Theorem 1.4.2. Usual studentized bootstrapped estimator is

$$\boldsymbol{H_n^*} = \sqrt{n}\sigma_n^{*-1}\hat{\sigma}_n\boldsymbol{\bar{\Sigma}_n^{-1/2}}(\boldsymbol{\beta_n^*} - \boldsymbol{\bar{\beta}_n}) = \sigma_n^{*-1}\hat{\sigma}_n\boldsymbol{F_n^*}$$

where

$\sigma_n^* = s_n^*\tau_n^{*-1}$, $\hat{\sigma}_n = s_n\tau_n^{-1}$, $\tau_n^* = n^{-1}\sum_{i=1}^n \psi'(\epsilon_i^*)$, $s_n^{*2} = n^{-1}\sum_{i=1}^n \psi^2(\epsilon_i^*)$,
$\tau_n = n^{-1}\sum_{i=1}^n \psi'(\bar{\epsilon}_i)$, $s_n^2 = n^{-1}\sum_{i=1}^n \psi^2(\bar{\epsilon}_i)$ and $\boldsymbol{\bar{\Sigma}_n^{-1/2}} = \boldsymbol{\bar{A}_{2n}^{-1/2}}\boldsymbol{\bar{A}_{1n}}$.

Now, by Taylor expansion

$$\tau_n^* = \tau_n + n^{-1}\sum_{i=1}^n [\boldsymbol{x_i'}(\boldsymbol{\beta_n^*} - \boldsymbol{\bar{\beta}_n})]\psi''(u_i) \tag{1.6.23}$$

and

$$s_n^{*2} = n^{-1}\sum_{i=1}^n \Big[\psi(\bar{\epsilon}_i) + \psi'(\bar{\epsilon}_i)[\boldsymbol{x_i'}(\boldsymbol{\beta_n^*} - \boldsymbol{\bar{\beta}_n})] + \frac{\big[\boldsymbol{x_i'}(\boldsymbol{\beta_n^*} - \boldsymbol{\bar{\beta}_n})\big]^2}{2}\psi''(v_i)\Big]^2 \tag{1.6.24}$$

where $|u_i - \bar{\epsilon}_i|$, $|v_i - \bar{\epsilon}_i| \leq |\epsilon_i^* - \bar{\epsilon}_i|$ for each $i \in \{1, \ldots, n\}$.

Hence, it follows that

$$\tau_n^* = \tau_n + n^{-1}\sum_{i=1}^{n}\psi''(\bar{\epsilon}_i)[\boldsymbol{x_i'}((\mathbf{E}_* L_n^*)^{-1}\Delta_n^*)] + \tilde{R}_{5n}^* = \tau_n + R_{5n}^* \text{ (say)} \tag{1.6.25}$$

$$s_n^{*2} = s_n^2 + 2n^{-1}\sum_{i=1}^{n}\psi(\bar{\epsilon}_i)\psi'(\bar{\epsilon}_i)[\boldsymbol{x_i'}((\mathbf{E}_* L_n^*)^{-1}\Delta_n^*)] + \tilde{R}_{6n}^* = s_n^2 + R_{6n}^* \text{ (say)} \tag{1.6.26}$$

where there exists constant $C_{10} > 0$ such that for $n \geq C_{10}$ by Lemma 1.6.1,

$$\mathbf{P}_*(|R_{5n}^*| > C_{10}.n^{-1/2}(logn)^{1/2}) + \mathbf{P}_*(|R_{6n}^*| > C_{10}.n^{-1/2}(logn)^{1/2}) = o_p(n^{-1/2})$$

$$\mathbf{P}_*(|\tilde{R}_{5n}^*| > C_{10}.n^{-1/2}(logn)^{-1}) + \mathbf{P}_*(|\tilde{R}_{6n}^*| > C_{10}.n^{-1/2}(logn)^{-1}) = o_p(n^{-1/2})$$

Now note that

$$\begin{aligned}
\sigma_n^{*-1} - \hat{\sigma}_n^{-1} =& \frac{\tau_n s_n^*(\tau_n^* s_n - \tau_n s_n^*)}{s_n^{*2} s_n \tau_n} \\
=& \frac{(\tau_n^{*2} - \tau_n^2)\bar{s}_n^2 + \tau_n^2(s_n^2 - s_n^{*2})}{2s_n^{*2} s_n \tau_n} \\
&- \frac{s_n^{*2}(\tau_n^* - \tau_n)^2 + \tau_n^{*2}(s_n - s_n^*)^2 + 2s_n^*\tau_n^*(\tau_n^* - \tau_n)(s_n - s_n^*)}{2s_n^{*2} s_n \tau_n}
\end{aligned}$$

and

$$(s_n^* - s_n)^2 \leq |s^{*2} - s_n^2|$$

and

$$s_n^{*-2} = s_n^{-2} + s_n^{-2}(s_n^2 - s_n^{*2})s_n^{-2} + s_n^{-2}(s_n^2 - s_n^{*2})s_n^{-2}(s_n^2 - s_n^{*2})s_n^{*-2}$$

$$\begin{aligned}
(s_n^* + s_n)^{-1} =& (2s_n)^{-1} + (2s_n)^{-1}(s_n - s_n^*)(2s_n)^{-1} \\
&+ (2s_n)^{-1}(s_n - s_n^*)(2s_n)^{-1}(s_n - s_n^*)(s_n^* + s_n)^{-1}
\end{aligned}$$

Now, it follows from the proof of Theorem 1.4.1 that

$$\boldsymbol{F_n^*} = \bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{\boldsymbol{-1/2}}(\mathbf{E}_* L_n^*)^{-1}\sqrt{n}\Delta_n^* + R_{0n}^* \tag{1.6.27}$$

and

$$\mathbf{P}_*(||R_{0n}^*|| = o(n^{-1/4})) = 1 - o_p(n^{-1/2})$$

Again, from (1.6.25) and (1.6.26) we have,

$$\begin{aligned}
\hat{\sigma}_n^{-1} - \sigma_n^{*-1} &= -\frac{2\tau_n s_n^2(\tau_n^* - \tau_n) + \tau_n^2(s_n^2 - s_n^{*2})}{2\bar{s}_n^3\tau_n} + R_{7n}^* \\
&= (2s_n^3\tau_n)^{-1}\Big[2\tau_n s_n^2\Big(n^{-1}\sum_{i=1}^{n}\psi''(\bar{\epsilon}_i)[\boldsymbol{x_i'}((\mathbf{E}_* L_n^*)^{-1}\Delta_n^*)]\Big) \\
&\quad - \tau_n^2\Big(2n^{-1}\sum_{i=1}^{n}\psi(\bar{\epsilon}_i)\psi'(\bar{\epsilon}_i)[\boldsymbol{x_i'}((\mathbf{E}_* L_n^*)^{-1}\Delta_n^*)]\Big)\Big] + \tilde{R}_{7n}^* \\
&= Z_n^* + \tilde{R}_{7n}^* \ \text{ (say)}
\end{aligned} \tag{1.6.28}$$

where there exists constant $C_{11} > 0$ such that for $n \geq C_{11}$,

$$\begin{aligned}
&\mathbf{P}_*(|R_{7n}^*| > C_{11}.n^{-1/2}(logn)^{-1}) + \mathbf{P}_*(|\tilde{R}_{7n}^*| > C_{11}.n^{-1/2}(logn)^{-1}) \\
&\qquad + \mathbf{P}_*(|Z_n^*| > C_{11}.n^{-1/2}logn) = o_p(n^{-1/2})
\end{aligned}$$

Therefore, (1.6.27) and (1.6.28) jointly imply

$$\boldsymbol{H_n^*} = \boldsymbol{U_n^*} - \sqrt{n}\hat{\sigma}_n\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{\boldsymbol{-1/2}}Z_n^*((\mathbf{E}_* L_n^*)^{-1}\Delta_n^*) + R_{8n}^* = \boldsymbol{\tilde{U}_n^*} + R_{8n}^* \ \text{ (say)} \tag{1.6.29}$$

where there exists constant $C_{12} > 0$ and a sequence of Borel sets $\boldsymbol{Q_{10n}}$ such that

$\mathbf{P}(\boldsymbol{Q_{10n}}) \uparrow 1$ and given $(\epsilon_1, \ldots, \epsilon_n) \in \boldsymbol{Q_{10n}}$ and $n \geq C_{12}$,

$$\mathbf{P}_*(||R^*_{8n}|| = o(n^{-1/2})) = 1 - o(n^{-1/2})$$

Now, we have seen that

$$\boldsymbol{U^*_n} = \boldsymbol{M^*_{0n}S^*_n} + (\boldsymbol{S^*_n}'\boldsymbol{M^*_{1n}S^*_n}, \ldots, \boldsymbol{S^*_n}'\boldsymbol{M^*_{pn}S^*_n})'$$

where $\boldsymbol{M^*_{0n}} = O_p(1)$ and $\boldsymbol{M^*_{jn}} = O_p(n^{-1/2})$ for all $j \in \{1, \ldots, p\}$ and $\boldsymbol{S^*_n}$ is as defined in the proof of Theorem 1.4.1.

Again,

$$\begin{aligned}
&\sqrt{n}\hat{\sigma}_n \bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{-\mathbf{1/2}} Z^*_n((\mathbf{E}_* L^*_n)^{-1}\Delta^*_n) \\
&= (2s_n\tau_n)^{-2}\Big[\frac{2\tau_n s_n^2}{n\sqrt{n}}\boldsymbol{S^*_n}'\boldsymbol{T^*_n}'\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{\mathbf{1/2}}\sum_{i=1}^{n}\psi''(\bar{\epsilon}_i)\boldsymbol{x_i} - \frac{2\tau_n^2}{n\sqrt{n}}\boldsymbol{S^*_n}'\boldsymbol{T^*_n}'\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{\mathbf{1/2}}\sum_{i=1}^{n}\psi(\bar{\epsilon}_i)\psi'(\bar{\epsilon}_i)\boldsymbol{x_i}\Big]\boldsymbol{T^*_n S^*_n} \\
&= (\boldsymbol{S^*_n}'\check{\boldsymbol{E}}^*_{\boldsymbol{1n}}\boldsymbol{S^*_n}, \ldots, \boldsymbol{S^*_n}'\check{\boldsymbol{E}}^*_{\boldsymbol{1n}}\boldsymbol{S^*_n})' \quad \text{(say)}
\end{aligned}$$

for some $k \times k$ matrices $\check{\boldsymbol{E}}^*_{\boldsymbol{jn}}$ with
$\check{\boldsymbol{E}}^*_{\boldsymbol{jn}} = \frac{(s_n\tau_n)^{-2}}{n\sqrt{n}}\sum_{i=1}^n\left[\tau_n s_n^2\psi''(\bar{\epsilon}_i) - \tau_n^2\psi(\bar{\epsilon}_i)\psi'(\bar{\epsilon}_i)\right]\boldsymbol{T^*_n}'\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{\mathbf{1/2}}\boldsymbol{x_i}\boldsymbol{t^*_{jn}}'$ where $\boldsymbol{T^*_n} = (\boldsymbol{t^*_{1n}}, \ldots, \boldsymbol{t^*_{pn}})'$. Note that, $||\check{\boldsymbol{E}}^*_{\boldsymbol{jn}}|| = O_p(n^{-1/2})$.

Therefore, we have

$$\tilde{\boldsymbol{U}}^*_{\boldsymbol{n}} = \tilde{\boldsymbol{M}}^*_{\boldsymbol{0n}}\boldsymbol{S^*_n} + (\boldsymbol{S^*_n}'\tilde{\boldsymbol{M}}^*_{\boldsymbol{1n}}\boldsymbol{S^*_n}, \ldots, \boldsymbol{S^*_n}'\tilde{\boldsymbol{M}}^*_{\boldsymbol{pn}}\boldsymbol{S^*_n})' \tag{1.6.30}$$

where $\tilde{\boldsymbol{M}}^*_{\boldsymbol{0n}} = \boldsymbol{M^*_{0n}} = O_p(1)$ and $\tilde{\boldsymbol{M}}^*_{\boldsymbol{jn}} = \boldsymbol{M^*_{jn}} - \check{\boldsymbol{E}}^*_{\boldsymbol{jn}} = O_p(n^{-1/2})$ for all $j \in \{1, \ldots, p\}$.

Hence, by Result 1.6.2 and 1.6.3, in the similar way as in the proof of Theorem 1.4.1, we have,

$$\sup_{B\in\mathscr{B}}\left|\mathbf{P}_*(\tilde{\boldsymbol{U}}_{\boldsymbol{n}}^* \in B) - \int_B \tilde{\xi}_n^*(\boldsymbol{x})d\boldsymbol{x}\right| = o_p(n^{-1/2}) \quad \text{as } n \to \infty \tag{1.6.31}$$

where

$$\tilde{\xi}_n^*(\boldsymbol{x}) = \left[1 - n^{-1/2}\Big\{\sum_{|\boldsymbol{\nu}|=1} \tilde{b}_{11}^{*(\boldsymbol{\nu})} D^{\boldsymbol{\nu}} + \sum_{|\boldsymbol{\nu}|=3} \frac{\tilde{b}_{31}^{*(\boldsymbol{\nu})}}{\boldsymbol{\nu}!} D^{\boldsymbol{\nu}}\Big\}\right]\phi(\boldsymbol{x}) \tag{1.6.32}$$

since, the second cumulant of $\boldsymbol{t'}\tilde{\boldsymbol{U}}_{\boldsymbol{n}}^*$ is $\kappa_2(\boldsymbol{t'}\tilde{\boldsymbol{U}}_{\boldsymbol{n}}^*) = \boldsymbol{t'}\boldsymbol{t} + o_p(n^{-1/2})$.

Hence part (a) of Theorem 1.4.2 follows by (4.12) of Qumsiyeh (1990a).

Again the two term EE of the studentized regression M-estimator $\boldsymbol{H_n}$ can be found in the similar way as

$$\tilde{\xi}_n(\boldsymbol{x}) = \left[1 - n^{-1/2}\Big\{\sum_{|\boldsymbol{\nu}|=1} \tilde{b}_{11}^{(\boldsymbol{\nu})} D^{\boldsymbol{\nu}} + \sum_{|\boldsymbol{\nu}|=3} \frac{\tilde{b}_{31}^{(\boldsymbol{\nu})}}{\boldsymbol{\nu}!} D^{\boldsymbol{\nu}}\Big\}\right]\phi(\boldsymbol{x}) \tag{1.6.33}$$

Hence, for $n > N$, N being a fixed natural number, we have for $k > 0$

$$\mathbf{P}\Big(\sup_{B\in\mathscr{B}}\left|\mathbf{P}_*(\boldsymbol{H}_{\boldsymbol{n}}^* \in B) - \mathbf{P}(\boldsymbol{H_n} \in B)\right| > k.n^{-1/2}\Big)$$

$$\leq \boldsymbol{P}\Big(\sup_{B\in\mathscr{B}}\Big|\int_B \tilde{\xi}_n^*(\boldsymbol{x})d\boldsymbol{x} - \int_B \tilde{\xi}_n(\boldsymbol{x})d\boldsymbol{x}\Big| > \frac{k}{2}.n^{-1/2}\Big) + o(1)$$

$$\leq \mathbf{P}\Big(n^{-1/2}\sum_{|\boldsymbol{\nu}|=1}|\tilde{b}_{11}^{*(\boldsymbol{\nu})} - \tilde{b}_{11}^{(\boldsymbol{\nu})}| + n^{-1/2}\sum_{|\boldsymbol{\nu}|=3}|\tilde{b}_{31}^{*(\boldsymbol{\nu})} - \tilde{b}_{31}^{(\boldsymbol{\nu})}| > \frac{k}{2}.n^{-1/2}\Big) + o(1)$$

which can be made less than any positive number by considering k large enough, since for any $p \times 1$ vectors $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ with $|\boldsymbol{\nu}_1| = 1$ and $|\boldsymbol{\nu}_2| = 3$, it can be shown that $\Big(|\tilde{b}_{11}^{(\boldsymbol{\nu}_1)}| + |\tilde{b}_{31}^{(\boldsymbol{\nu}_2)}|\Big) = O(1)$ and $\Big(|\tilde{b}_{11}^{*(\boldsymbol{\nu}_1)}| + |\tilde{b}_{31}^{*(\boldsymbol{\nu}_2)}|\Big) = O_p(1)$. Therefore, part (b) follows.

Now to prove part (c), we need to investigate more deeply the coefficients $\tilde{b}_{11}^{*(\boldsymbol{\nu})}$, $\tilde{b}_{31}^{*(\boldsymbol{\nu})}$,

$\tilde{b}_{11}^{(\boldsymbol{\nu})}, \tilde{b}_{31}^{(\boldsymbol{\nu})}$ that are present in (1.6.32) and (1.6.33). The first and third cumulants of $\tilde{\boldsymbol{U}}_{\boldsymbol{n}}^{*}$ and $\tilde{\boldsymbol{U}}_{\boldsymbol{n}}$ are given by

$$\kappa_1(\boldsymbol{t'}\tilde{\boldsymbol{U}}_{\boldsymbol{n}}^{*}) = \textstyle\sum_{|\boldsymbol{\nu}|=1} \tilde{b}_{11}^{*(\boldsymbol{\nu})} \boldsymbol{t}^{\boldsymbol{\nu}}, \quad \kappa_1(\boldsymbol{t'}\tilde{\boldsymbol{U}}_{\boldsymbol{n}}) = \textstyle\sum_{|\boldsymbol{\nu}|=1} \tilde{b}_{11}^{(\boldsymbol{\nu})} \boldsymbol{t}^{\boldsymbol{\nu}}.$$

$$\kappa_3(\boldsymbol{t'}\tilde{\boldsymbol{U}}_{\boldsymbol{n}}^{*}) = \textstyle\sum_{|\boldsymbol{\nu}|=3} \frac{3!}{\boldsymbol{\nu}!} \tilde{b}_{31}^{*(\boldsymbol{\nu})} \boldsymbol{t}^{\boldsymbol{\nu}} + o_p(n^{-1/2}), \quad \kappa_3(\boldsymbol{t'}\tilde{\boldsymbol{U}}_{\boldsymbol{n}}) = \textstyle\sum_{|\boldsymbol{\nu}|=3} \frac{3!}{\boldsymbol{\nu}!} \tilde{b}_{31}^{(\boldsymbol{\nu})} \boldsymbol{t}^{\boldsymbol{\nu}} + o(n^{-1/2}).$$

where if $\boldsymbol{\nu}_1$ is a $p \times 1$ vector with all the elements being 0, except the jth one and $\boldsymbol{\nu}_2$ is a $p \times 1$ vector with all the elements being 0, except the j_1, j_2 and j_3 positions, then

$$\tilde{b}_{11}^{*(\boldsymbol{\nu_1})} = b_{11}^{*(\boldsymbol{\nu_1})} + (s_n \tau_n)^{-2} n^{-1} \sum_{i=1}^{n} \left[\left(\tau_n^2 \psi(\bar{\epsilon}_i) \psi'(\bar{\epsilon}_i) - \tau_n s_n^2 \psi''(\bar{\epsilon}_i) \right) \boldsymbol{h'_{jn}} \boldsymbol{\bar{A}_{2n}} \boldsymbol{\bar{A}_{1n}^{-1}} \boldsymbol{x_i} \right]$$

$$\tilde{b}_{11}^{(\boldsymbol{\nu_1})} = b_{11}^{(\boldsymbol{\nu_1})} + \tau^{-2} s^{-1} \left(\tau \mathbf{E} \psi(\epsilon_1) \psi'(\epsilon_1) - s^2 \mathbf{E} \psi''(\epsilon_1) \right) n^{-1} \sum_{i=1}^{n} a_{jin}$$

$$+ s^{-1} \Big(\tau^{-1} \mathbf{E} \psi(\epsilon_1) \psi'(\epsilon_1) - 2^{-1} s^{-2} \mathbf{E} \psi^3(\epsilon_1) \Big) n^{-1} \sum_{i=1}^{n} a_{jin}$$

$$\tilde{b}_{31}^{*(\boldsymbol{\nu_2})} = b_{31}^{*(\boldsymbol{\nu_2})}$$

$$+ 6n^{-3} \sum_{i,j,k=1}^{n} \left[\Big(s_n^{-2} \psi(\bar{\epsilon}_l) \psi'(\bar{\epsilon}_l) - \tau_n^{-1} \psi''(\bar{\epsilon}_l) \Big) (\boldsymbol{x'_i} \boldsymbol{\bar{A}_{1n}^{-1}} \boldsymbol{x_k}) a_{j_1 in}^{*} a_{j_2 jn}^{*} a_{j_3 jn}^{*} \psi^2(\bar{\epsilon}_i) \psi^2(\bar{\epsilon}_j) \right]$$

$$\tilde{b}_{31}^{(\boldsymbol{\nu_2})} = b_{31}^{(\boldsymbol{\nu_2})}$$

$$+ \tau^{-1} s \left(s^{-2} E \psi(\epsilon_1) \psi'(\epsilon_1) - \tau^{-1} \mathbf{E} \psi''(\epsilon_1) \right) 6n^{-3} \sum_{i,j,k=1}^{n} \left[(\boldsymbol{x'_i} \boldsymbol{A_n^{-1}} \boldsymbol{x_k}) a_{j_1 in} a_{j_2 jn} a_{j_3 jn} \right]$$

$$+ s^{-1} \left(\tau^{-1} \mathbf{E} \psi(\epsilon_1) \psi'(\epsilon_1) - 2^{-1} s^{-2} \mathbf{E} \psi^3(\epsilon_1) \right) 6n^{-2} \sum_{i,j=1}^{n} \left[a_{j_1 in} a_{j_2 jn} a_{j_3 jn} \right]$$

Suppose, $Z_1^{\boldsymbol{\nu}_1} = n^{-1} \sum_{i=1}^{n} \boldsymbol{e'_j} \boldsymbol{A_n^{-1/2}} \boldsymbol{x_i}$, $Z_2^{\boldsymbol{\nu}_2} = 6n^{-2} \sum_{i,j=1}^{n} \left[a_{j_1 in} a_{j_2 jn} a_{j_3 jn} \right]$ and $L =$

$\mathbf{E}\psi^2(\epsilon_1)\ \mathbf{E}\psi(\epsilon_1)\psi'(\epsilon_1) - 2^{-1}\mathbf{E}\psi'(\epsilon_1)\mathbf{E}\psi^3(\epsilon_1)$. Note that if $L \neq 0$ then for any $p \times 1$ vectors $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ with $|\boldsymbol{\nu}_1| = 1$ and $|\boldsymbol{\nu}_2| = 3$, $|\tilde{b}_{11}^{*(\boldsymbol{\nu_1})} - \tilde{b}_{11}^{(\boldsymbol{\nu_1})}| \neq o_p(1)$ and $|\tilde{b}_{31}^{*(\boldsymbol{\nu_2})} - \tilde{b}_{31}^{(\boldsymbol{\nu_2})}| \neq o_p(1)$. Hence, for some non-null set $\tilde{B} \in \mathscr{B}$ with $\tilde{B} \subseteq (-\infty, 0)^p$,

$$\begin{aligned}
&\mathbf{P}\Big(\liminf_{n\to\infty} \sqrt{n}\Big[\sup_{B\in\mathscr{B}} \big|\mathbf{P}_*(\boldsymbol{H}_n^* \in B) - \mathbf{P}(\boldsymbol{H}_n \in B)\big|\Big] > \epsilon\Big) \\
&\geq \mathbf{P}\Big(\liminf_{n\to\infty} \sqrt{n}\big|\mathbf{P}_*(\boldsymbol{H}_n^* \in \tilde{B}) - \mathbf{P}(\boldsymbol{H}_n \in \tilde{B})\big| > \epsilon\Big) \\
&\geq \mathbf{P}\Bigg(\liminf_{n\to\infty} \sqrt{n}\Bigg| \int_{\tilde{B}} \tilde{\xi}^*(\boldsymbol{x})d\boldsymbol{x} - \int_{\tilde{B}} \tilde{\xi}(\boldsymbol{x})d\boldsymbol{x} + o_p(n^{-1/2})\Bigg| > \epsilon\Bigg) \\
&\geq \mathbf{P}\Bigg(\liminf_{n\to\infty} \sqrt{n}\Bigg| \int_{\tilde{B}} \tilde{\xi}^*(\boldsymbol{x})d\boldsymbol{x} - \int_{\tilde{B}} \tilde{\xi}(\boldsymbol{x})d\boldsymbol{x}\Bigg| > \epsilon/2\Bigg) \\
&\geq \mathbf{P}\Bigg(\liminf_{n\to\infty} \Bigg| \sum_{|\boldsymbol{\nu}|=1} Z_1^{\boldsymbol{\nu}} \int_{\tilde{B}} D^{\boldsymbol{\nu}}\phi(\boldsymbol{x})d\boldsymbol{x} + \sum_{|\boldsymbol{\nu}|=3} Z_2^{\boldsymbol{\nu}} \int_{\tilde{B}} D^{\boldsymbol{\nu}}\phi(\boldsymbol{x})d\boldsymbol{x} + o_p(1)\Bigg| > \epsilon/2|L|\Bigg) \\
&= 1
\end{aligned}$$

for some $\epsilon > 0$, since $\left|\sum_{|\boldsymbol{\nu}|=1} Z_1^{\boldsymbol{\nu}} \int_{\tilde{B}} D^{\boldsymbol{\nu}}\phi(\boldsymbol{x})d\boldsymbol{x} + \sum_{|\boldsymbol{\nu}|=3} Z_2^{\boldsymbol{\nu}} \int_{\tilde{B}} D^{\boldsymbol{\nu}}\phi(\boldsymbol{x})d\boldsymbol{x}\right|$ is independent of n and also non vanishing for some choice of $\tilde{B}$ with $\tilde{B} \subseteq (-\infty, 0)^p$.

Therefore, part (c) of Theorem 1.4.2 follows.

Proof of Theorem 1.4.3. The modified studentized bootstrapped M-estimator is

$$\tilde{\boldsymbol{H}}_n^* = \sqrt{n}(\tilde{\sigma}_n^*)^{-1}\hat{\sigma}_n \bar{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^* - \bar{\boldsymbol{\beta}}_n) \tag{1.6.34}$$

where $\tilde{\sigma}_n^* = \tilde{s}_n^* \tilde{\tau}_n^{*-1}$, $\tilde{\tau}_n^* = n^{-1}\sum_{i=1}^n \psi'(\epsilon_i^*)G_i^*$ and $\tilde{s}_n^{*2} = n^{-1}\sum_{i=1}^n \psi^2(\epsilon_i^*)(G_i^* - \mu_{G^*})^2$.

Define, $\bar{s}_n = s_n \sigma_{G^*}$, $\bar{\tau}_n = \tau_n \mu_{G^*}$.

Now, using (1.6.23) and (1.6.24) it follows,

$$\tilde{\tau}_n^* - \bar{\tau}_n = n^{-1}\sum_{i=1}^{n}\psi'(\bar{\epsilon}_i)(G_i^* - \mu_{G^*}) - n^{-1}\sum_{i=1}^{n}\psi''(\bar{\epsilon}_i)[\boldsymbol{x_i'}((\mathbf{E}_* L_n^*)^{-1}\Delta_n^*)]\mu_{G^*} + \tilde{R}_{9n}^*$$

$$= n^{-1}\sum_{i=1}^{n}\psi'(\bar{\epsilon}_i)(G_i^* - \mu_{G^*}) + R_{9n}^* \text{ (say)} \tag{1.6.35}$$

$$\tilde{s}_n^{*2} - \bar{s}_n^2 = n^{-1}\sum_{i=1}^{n}\psi^2(\bar{\epsilon}_i)\big[(G_i^* - \mu_{G^*}) - \sigma_{G^*}^2\big]$$

$$- 2n^{-1}\sum_{i=1}^{n}\psi(\bar{\epsilon}_i)\psi'(\bar{\epsilon}_i)[\boldsymbol{x_i'}((\mathbf{E}_* L_n^*)^{-1}\Delta_n^*)]\sigma_{G^*}^2 + \tilde{R}_{10n}^*$$

$$= n^{-1}\sum_{i=1}^{n}\psi^2(\bar{\epsilon}_i)\big[(G_i^* - \mu_{G^*}) - \sigma_{G^*}^2\big] + R_{10n}^* \text{ (say)} \tag{1.6.36}$$

where there exists constant $C_{13} > 0$ such that for $n \geq C_{13}$ by Lemma 1.6.1,

$$\mathbf{P}_*(|R_{9n}^*| > C_{13}.n^{-1/2}(logn)^{1/2}) + \mathbf{P}_*(|R_{10n}^*| > C_{13}.n^{-1/2}(logn)^{1/2}) = o_p(n^{-1/2})$$

$$\mathbf{P}_*(|\tilde{R}_{9n}^*| > C_{13}.n^{-1/2}(logn)^{-1}) + \mathbf{P}_*(|\tilde{R}_{10n}^*| > C_{13}.n^{-1/2}(logn)^{-1}) = o_p(n^{-1/2})$$

Hence, similar to (1.6.28), it can be shown in the same way that

$$\hat{\sigma}_n^{-1} - \tilde{\sigma}_n^{*-1} = -\frac{2\tau_n s_n^2(\tau_n^* - \tau_n) + \tau_n^2(s_n^2 - s_n^{*2})}{2\bar{s}_n^3\tau_n} - \bar{Z}_n^* + R_{11n}^*$$

$$= Z_n^* - \bar{Z}_n^* + \tilde{R}_{11n}^* \text{ (say)} \tag{1.6.37}$$

where Z_n^* is as defined in the proof of Theorem 1.4.2 and

$$\bar{Z}_n^* = 2^{-1}\left(\bar{\tau}_n\bar{s}_n\right)^{-2}\Bigg[2\bar{\tau}_n\bar{s}_n^2\Big(n^{-1}\sum_{i=1}^{n}\psi'(\bar{\epsilon}_i)(G_i^*-\mu_{G^*})\Big)$$

$$-\bar{\tau}_n^2\Big(n^{-1}\sum_{i=1}^{n}\psi^2(\bar{\epsilon}_i)\big[(G_i^*-\mu_{G^*})^2-\sigma_{G^*}^2\big]\Big)\Bigg]$$

and there exists constant $C_{14}>0$ such that for $n\geq C_{14}$,

$$\mathbf{P}_*(|R_{11n}^*|>C_{14}.n^{-1/2}(logn)^{-1})+\mathbf{P}_*(|\tilde{R}_{11n}^*|>C_{14}.n^{-1/2}(logn)^{-1})$$

$$+\mathbf{P}_*(|Z_n^*|>C_{14}.n^{-1/2}logn)+\mathbf{P}_*(|\bar{Z}_n^*|>C_{14}.n^{-1/2}logn)=o_p(n^{-1/2})$$

Therefore, from (1.6.29) and (1.6.37) we have,

$$\tilde{\boldsymbol{H}}_{\boldsymbol{n}}^* = \tilde{\boldsymbol{U}}_{\boldsymbol{n}}^* + \sqrt{n}\hat{\sigma}_n\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{-\mathbf{1/2}}\bar{Z}_n^*((\mathbf{E}_*L_n^*)^{-1}\Delta_n^*)+R_{12n}^* = \bar{\boldsymbol{U}}_{\boldsymbol{n}}^*+R_{12n}^*\ \text{(say)} \tag{1.6.38}$$

where there exists constant $C_{15}>0$ such that for $n\geq C_{15}$,

$$\mathbf{P}_*(||R_{12n}^*||=o(n^{-1/2}))=1-o_p(n^{-1/2})$$

Now, for all $i\in\{1,\ldots,n\}$, define,

$Y_{1i}^*=(G_i^*-\mu_{G^*})$, $Y_{2i}^*=(G_i^*-\mu_{G^*})^2-\sigma_{G^*}^2$

$\boldsymbol{X_i^*}=\big(\breve{\boldsymbol{v}}'_i Y_{1i}^*, n^{-1/2}\psi^2(\bar{\epsilon}_i)Y_{2i}^*\big)'$ where $\breve{\boldsymbol{v}}_{\boldsymbol{i}}$ is defined with $\breve{\boldsymbol{z}}_{\boldsymbol{i}}$ in place of $\boldsymbol{z_i}$.

$\boldsymbol{V_n^*}=\sum_{i=1}^{n}\mathbf{Cov}_*(\boldsymbol{X_i^*})$, $\tilde{\boldsymbol{X}}_{\boldsymbol{i}}^*=\boldsymbol{V_n^{*-1/2}}\boldsymbol{X_i}$ and $\bar{\boldsymbol{S}}_{\boldsymbol{n}}^*=\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_{\boldsymbol{i}}^*$.

Partition $\boldsymbol{B^{-1}}$ as $\boldsymbol{B^{-1}}=(\bar{\boldsymbol{B}}_{\mathbf{1}}',\bar{\boldsymbol{B}}_{\mathbf{2}}')'$ where $\bar{\boldsymbol{B}}_{\mathbf{1}}$ is of order $p\times k$. Here $k=p+q+1$ where as in the previous two theorems $k=p+q$. Also partition $\boldsymbol{V_n^{*1/2}}$ as $\boldsymbol{V_n^{*1/2}}=(\boldsymbol{V_{1n}^{*\prime}},\boldsymbol{V_{2n}^*})'$ where $\boldsymbol{V_{1n}^*}$ is of order $k\times(k+1)$. Also, suppose $\bar{\boldsymbol{B}}_{\mathbf{2}}=(\bar{\boldsymbol{B}}_{\mathbf{2}}^{(\mathbf{1})},\ldots,\bar{\boldsymbol{B}}_{\mathbf{2}}^{(\boldsymbol{q}+\mathbf{1})})'$.

Therefore, we have

$$\bar{\boldsymbol{U}}_n^* = \bar{\boldsymbol{M}}_{0n}^* \bar{\boldsymbol{S}}_n^* + (\bar{\boldsymbol{S}}_n^{*\prime} \bar{\boldsymbol{M}}_{1n}^* \bar{\boldsymbol{S}}_n^*, \ldots, \bar{\boldsymbol{S}}_n^{*\prime} \bar{\boldsymbol{M}}_{pn}^* \bar{\boldsymbol{S}}_n^*)' \tag{1.6.39}$$

where $\bar{\boldsymbol{M}}_{0n}^* = \tilde{\boldsymbol{M}}_{0n}^*$, $\bar{\boldsymbol{M}}_{jn}^* = \tilde{\boldsymbol{M}}_{jn}^* + \bar{\boldsymbol{E}}_{jn}^*$ and

$$\bar{\boldsymbol{E}}_{jn}^* = 2^{-1}(\bar{\tau}_n \bar{s}_n)^{-2}\left[\left(n^{-1/2} 2\bar{\tau}_n \bar{s}_n\right) \boldsymbol{V}_{1n}^{*\prime} \bar{\boldsymbol{B}}_{2}^{(q+1)} - \left(n^{-1}\bar{\tau}_n^2\right)\boldsymbol{V}_{2n}^*\right] \boldsymbol{h}_{jn}' \bar{\boldsymbol{B}}_1 \mu_{G^*}^{-1} \boldsymbol{V}_{1n}^*$$

As in the proof of the Theorem 4.1, here also $||\bar{\boldsymbol{B}}_1||, ||\bar{\boldsymbol{B}}_2|| = O_p(n)$. Also in the current setup $||\boldsymbol{V}_{1n}^*||, ||\boldsymbol{V}_{2n}^*|| = O_p(1)$. Hence we have $||\bar{\boldsymbol{M}}_{jn}^*|| = O_p(n^{1/2})$. Therefore, by Lemma 1.6.2, 1.6.3 and (4.12) of Qumsiyeh (1990a), it follows that

$$\sup_{B \in \mathscr{B}} |\mathbf{P}_*(\tilde{\boldsymbol{H}}_n^* \in B) - \int_B \bar{\xi}_n^*(\boldsymbol{x}) d\boldsymbol{x}| = o_p(n^{-1/2})$$

where

$$\bar{\xi}_n^*(\boldsymbol{x}) = \left[1 - n^{-1/2}\left\{\sum_{|\boldsymbol{\nu}|=1} \bar{b}_{11}^{*(\boldsymbol{\nu})} D^{\boldsymbol{\nu}} + \sum_{|\boldsymbol{\nu}|=3} \frac{\bar{b}_{31}^{*(\boldsymbol{\nu})}}{\boldsymbol{\nu}!} D^{\boldsymbol{\nu}}\right\}\right]\phi(\boldsymbol{x}) \tag{1.6.40}$$

since the first three cumulants of $\boldsymbol{t}'\bar{U}_n^*$ are given by

$\kappa_1(\boldsymbol{t}'\bar{\boldsymbol{U}}_n^*) = n^{-1/2}\sum_{|\boldsymbol{\nu}|=1} \bar{b}_{11}^*(\boldsymbol{\nu})\boldsymbol{t}^{\boldsymbol{\nu}}$

$\kappa_2(\boldsymbol{t}'\bar{\boldsymbol{U}}_n^*) = \boldsymbol{t}'\boldsymbol{t} + o_p(n^{-1/2})$

$\kappa_3(\boldsymbol{t}'\bar{\boldsymbol{U}}_n^*) = n^{-1/2}\sum_{|\boldsymbol{\nu}|=3} \frac{3!}{\boldsymbol{\nu}!}\bar{b}_{31}^{*(\boldsymbol{\nu})}\boldsymbol{t}^{\boldsymbol{\nu}} + o_p(n^{-1/2})$

where if $\boldsymbol{\nu}_1$ is a $p \times 1$ vector with all the elements being 0, except the jth one and $\boldsymbol{\nu}_2$ is a $p \times 1$ vector with all the elements being 0, except the j_1, j_2 and j_3 positions, then

$$\bar{b}_{11}^{*(\boldsymbol{\nu}_1)} = \tilde{b}_{11}^{*(\boldsymbol{\nu}_1)} + \tau_n^{-1} n^{-1}\sum_{i=1}^n a_{jin}^* \psi(\bar{\epsilon}_i)\psi'(\bar{\epsilon}_i) - 2^{-1}s_n^{-2}n^{-1}\sum_{i=1}^n a_{jin}^*\psi^3(\bar{\epsilon}_i)$$

and

$$\bar{b}_{31}^{*(\boldsymbol{\nu_2})} = \tilde{b}_{31}^{*(\boldsymbol{\nu_2})} + 6n^{-2} \sum_{i,j=1}^{n} a^*_{j_1 in} a^*_{j_1 jn} a^*_{j_1 jn} \Big[\tau_n^{-1} \psi(\bar{\epsilon}_i)\psi'(\bar{\epsilon}_i)\psi^2(\bar{\epsilon}_j) - 2^{-1} s_n^{-2} \psi^3(\bar{\epsilon}_i)\psi^2(\bar{\epsilon}_j)\Big]$$

Note that, $\big|\bar{b}_{11}^{*(\boldsymbol{\nu}_1)} - \tilde{b}_{11}^{(\boldsymbol{\nu}_1)}\big| + \big|\bar{b}_{31}^{*(\boldsymbol{\nu}_2)} - \tilde{b}_{31}^{(\boldsymbol{\nu}_2)}\big| \to 0$ in probability as $n \to \infty$ for all $p \times 1$ vectors $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ with $|\boldsymbol{\nu}_1| = 1$ and $|\boldsymbol{\nu}_2| = 3$. Hence, for $n > C$, for some constant C, we have for any $\zeta > 0$

$$\begin{aligned}
&\mathbf{P}\Big(\sup_{B \in \mathscr{B}} \big|\mathbf{P}_*(\tilde{\boldsymbol{H}}_{\boldsymbol{n}}^* \in B) - \mathbf{P}(\boldsymbol{H}_{\boldsymbol{n}} \in B)\big| > \zeta . n^{-1/2}\Big) \\
&\leq \mathbf{P}\Big(\sup_{B \in \mathscr{B}} \Big| \int_B \bar{\xi}_n^*(\boldsymbol{x}) d\boldsymbol{x} - \int_B \tilde{\xi}_n(\boldsymbol{x}) d\boldsymbol{x}\Big| > \frac{\zeta}{2} . n^{-1/2}\Big) + o(1) \\
&\leq \mathbf{P}\Big(n^{-1/2} \sum_{|\boldsymbol{\nu}|=1} |\bar{b}_{11}^{*(\boldsymbol{\nu})} - \tilde{b}_{11}^{(\boldsymbol{\nu})}| + n^{-1/2} \sum_{|\boldsymbol{\nu}|=3} |\bar{b}_{31}^{*(\boldsymbol{\nu})} - \tilde{b}_{31}^{(\boldsymbol{\nu})}| > \frac{\zeta}{2} . n^{-1/2}\Big) + o(1) \\
&\to 0 \quad \text{as } n \to \infty
\end{aligned}$$

Therefore, Theorem 1.4.3 follows.

Proof of Theorem 1.5.1 In the non-iid set-up, the original studentized pivot and the studentized bootstrapped pivot are defined as

$$\breve{\boldsymbol{H}}_{\boldsymbol{n}} = \sqrt{n} \bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{-1/2} (\bar{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}) \text{ and } \breve{\boldsymbol{H}}_{\boldsymbol{n}}^* = \sqrt{n} \boldsymbol{\Sigma}_{\boldsymbol{n}}^{*-1/2} (\boldsymbol{\beta}_{\boldsymbol{n}}^* - \bar{\boldsymbol{\beta}}_{\boldsymbol{n}})$$

where $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{-1/2} = \bar{\boldsymbol{A}}_{\boldsymbol{2n}}^{-1/2} \bar{\boldsymbol{A}}_{\boldsymbol{1n}}$ with $\bar{\boldsymbol{A}}_{\boldsymbol{1n}} = n^{-1} \sum_{i=1}^n \boldsymbol{x_i x_i'} \psi'(\bar{\epsilon}_i)$, $\bar{\boldsymbol{A}}_{\boldsymbol{2n}} = n^{-1} \sum_{i=1}^n \boldsymbol{x_i x_i'} \psi^2(\bar{\epsilon}_i)$ and $\boldsymbol{\Sigma}_{\boldsymbol{n}}^{*-1/2} = \boldsymbol{A}_{\boldsymbol{2n}}^{*-1/2} \boldsymbol{A}_{\boldsymbol{1n}}^*$ with $\boldsymbol{A}_{\boldsymbol{1n}}^* = n^{-1} \sum_{i=1}^n \boldsymbol{x_i x_i'} \psi'(\epsilon_i^*) G_i^*$, $\boldsymbol{A}_{\boldsymbol{2n}}^* = n^{-1} \sum_{i=1}^n \boldsymbol{x_i x_i'} \psi^2(\epsilon_i^*)(G_i^* - \mu_{G^*})^2$. Here, $\bar{\epsilon}_i = y_i - \boldsymbol{x_i'} \bar{\boldsymbol{\beta}}_{\boldsymbol{n}}$ and $\epsilon_i^* = y_i - \boldsymbol{x_i'} \boldsymbol{\beta}_{\boldsymbol{n}}^*$, $i \in \{1, \ldots, n\}$.

Part (a) follows through the same line of arguments as in proof of proposition 1.4.1.

To prove part (b) and (c), note that through the same line of arguments Theorem 4.1 goes through in non-iid setup also when the bootstrapped and unbootstrapped pivots are respectively defined as $\boldsymbol{F}_n^* = \sqrt{n}\bar{\boldsymbol{A}}_{2n}^{-1/2}\bar{\boldsymbol{A}}_{1n}(\boldsymbol{\beta}_n^* - \bar{\boldsymbol{\beta}}_n)$ and $\tilde{\boldsymbol{F}}_n = \sqrt{n}\boldsymbol{A}_{2n}^{-1/2}\boldsymbol{A}_{1n}(\boldsymbol{\beta}_n^* - \bar{\boldsymbol{\beta}}_n)$ where $\bar{\boldsymbol{A}}_{1n}$ and $\bar{\boldsymbol{A}}_{2n}$ are defined in the previous paragraph and $\boldsymbol{A}_{1n} = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i' \mathbf{E}\psi'(\epsilon_i)$ and $\boldsymbol{A}_{2n} = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i'\mathbf{E}\psi^2(\epsilon_i)$. Hence, from the proofs of Theorem 4.2 and 4.3, it follows that it is enough to investigate the quantities $(\bar{\boldsymbol{A}}_{1n} - \boldsymbol{A}_{1n})$, $(\bar{\boldsymbol{A}}_{2n} - \boldsymbol{A}_{2n})$, $(\boldsymbol{A}_{1n}^*\mu_{G^*}^{-1} - \bar{\boldsymbol{A}}_{1n})$ and $(\boldsymbol{A}_{2n}^*\sigma_{G^*}^{-2} - \bar{\boldsymbol{A}}_{2n})$. Using Taylor's theorem and the fact that $\psi''(\cdot)$ is Lipschitz, it can be shown that

$$\begin{aligned}\bar{\boldsymbol{A}}_{1n} - \boldsymbol{A}_{1n} =& - n^{-1}\sum_{i=1}^n \Big[\boldsymbol{x}_i\boldsymbol{x}_i'\{\boldsymbol{x}_i'(\bar{\boldsymbol{\beta}}_n - \boldsymbol{\beta})\}\mathbf{E}\psi''(\epsilon_i)\Big] \\ &+ n^{-1}\sum_{i=1}^n \Big[\boldsymbol{x}_i\boldsymbol{x}_i'\big(\psi'(\epsilon_i) - \mathbf{E}\psi'(\epsilon_i)\big)\Big] + R_{6n} \qquad (1.6.41)\end{aligned}$$

$$\begin{aligned}\bar{\boldsymbol{A}}_{2n} - \boldsymbol{A}_{2n} =& - n^{-1}\sum_{i=1}^n \Big[\boldsymbol{x}_i\boldsymbol{x}_i'\{\boldsymbol{x}_i'(\bar{\boldsymbol{\beta}}_n - \boldsymbol{\beta})\}\mathbf{E}\psi(\epsilon_i)\psi'(\epsilon_i)\Big] \\ &+ n^{-1}\sum_{i=1}^n \Big[\boldsymbol{x}_i\boldsymbol{x}_i'\big(\psi^2(\epsilon_i) - \mathbf{E}\psi^2(\epsilon_i)\big)\Big] + R_{7n} \qquad (1.6.42)\end{aligned}$$

$$\begin{aligned}\boldsymbol{A}_{1n}^*\mu_{G^*}^{-1} - \bar{\boldsymbol{A}}_{1n} =& - n^{-1}\sum_{i=1}^n \Big[\boldsymbol{x}_i\boldsymbol{x}_i'\{\boldsymbol{x}_i'(\boldsymbol{\beta}_n^* - \bar{\boldsymbol{\beta}}_n)\}\psi''(\bar{\epsilon}_i)\Big] \\ &+ n^{-1}\sum_{i=1}^n \Big[\boldsymbol{x}_i\boldsymbol{x}_i'\psi'(\bar{\epsilon}_i)(G_i^* - \mu_{G^*})\mu_{G^*}^{-1}\Big] + R_{6n}^* \qquad (1.6.43)\end{aligned}$$

$$\begin{aligned}\boldsymbol{A}_{2n}^*\sigma_{G^*}^{-2} - \bar{\boldsymbol{A}}_{2n} =& - n^{-1}\sum_{i=1}^n \Big[\boldsymbol{x}_i\boldsymbol{x}_i'\{\boldsymbol{x}_i'(\boldsymbol{\beta}_n^* - \bar{\boldsymbol{\beta}}_n)\}\psi(\bar{\epsilon}_i)\psi'(\bar{\epsilon}_i)\Big] \\ &+ n^{-1}\sum_{i=1}^n \Big[\boldsymbol{x}_i\boldsymbol{x}_i'\psi^2(\bar{\epsilon}_i)\big[(G_i^* - \mu_{G^*})^2 - \sigma_{G^*}^2\big]\sigma_{G^*}^{-2}\Big] + R_{7n}^* \qquad (1.6.44)\end{aligned}$$

where

$$\mathbf{P}\big(||R_{6n}|| + ||R_{7n}|| = o(n^{-1/2})\big) = 1 - o(n^{-1/2}) \tag{1.6.45}$$

$$\mathbf{P}_*\big(||R^*_{6n}|| + ||R^*_{7n}|| = o(n^{-1/2})\big) = 1 - o_p(n^{-1/2}) \tag{1.6.46}$$

It is easy to observe, from the differences in the studentization terms in proofs of Theorem 1.4.2 and 1.4.3, that (1.6.41)-(1.6.44) are the equations which are exactly what we expect in achieving second order correctness [see (1.6.23), (1.6.24), (1.6.35), (1.6.36) for comparison]. Therefore, the rest of the proof follows in the same line as that of Theorem 1.4.3.

1.7 Conclusion

Second order results of perturbation bootstrap method in regression M-estimation are established. It is shown that the classical way of studentization in bootstrap setup is not sufficient for correcting the distribution of the regression M-estimator upto second order. This is a general statement corresponding to the fact that the usual studentized bootstrapped estimator is not capable of correcting the effect of skewness of the error distribution in least square regression. Novel modification is proposed in general setup by properly incorporating the effect of the randomization of the random perturbing quantities in the prevalent studentization factor and is shown as second order correct in both iid and non-iid error setup. Thus, in a way the results in this chapter establish perturbation bootstrap method as a refinement of the approximation of the exact distribution of the regression M-estimator over asymptotic normality. The second order result in non-iid case establishes robustness of the perturbation bootstrap towards the presence of heteroscedasticity, similar to wild bootstrap, but in more general setup of M-estimation. This is an important finding from the perspective of valid inferences regarding the regression parameters.

Chapter 2

Perturbation Bootstrap in Lasso

2.1 Introduction

Consider the multiple linear regression model

$$y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.1.1}$$

where $y_1, \ldots, y_n$ are responses, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are design vectors, $\epsilon_1, \ldots, \epsilon_n$ are errors, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is the p-dimensional vector of regression parameters. We will consider both the cases when the design vectors are fixed and when they are random, separately. Additionally, we assume that ϵ_i are independent (but possibly depends on $\boldsymbol{X}_i$) when design is fixed and $(\epsilon_i, \boldsymbol{X}_i)$ are independent when the design is random. When p is sufficiently large, it is common to approach regression model (2.1.1) with the assumption that the vector $\boldsymbol{\beta}$ is sparse, that is the set $\mathcal{A} = \{j : \beta_j \neq 0\}$ has cardinality $p_0 = |\mathcal{A}|$ that is much smaller than p, meaning that only a few of the covariates are "active". One of the widely used methods under the sparsity assumption is the least absolute shrinkage and selection operator or the Lasso, introduced by Tibshirani (1996).

It is defined as the minimizer of l_1-penalized least-square criterion function,

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{t}} \left[\sum_{i=1}^{n} (y_i - \boldsymbol{X}_i' \boldsymbol{t})^2 + \lambda_n \sum_{j=1}^{p} |t_j| \right], \tag{2.1.2}$$

where $\lambda_n > 0$ is the penalty parameter. 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 parsimonious model section [cf. Zhao and Yu (2006), Zou (2006) and Wainwright (2009)]. Another important aspect of Lasso is its computational feasibility in high dimensional regression problems [cf. Efron et al. (2004), Friedman et al. (2007), Fu (1998), Osborne et al. (2000)].

Asymptotic properties of the Lasso estimator was first investigated by Knight and Fu (2000) [hereafter referred to as KF(00)] for the model (2.1.1) with non-random design and homoscedastic error, when p is assumed fixed. Subsequently, in fixed p setting, Wagener and Dette (2012) and Camponovo (2015) extended results obtained by KF(00) to the cases when errors are heteroscedastic and the design is random. Although it follows that the Lasso estimator is $\sqrt{n}$-consistent for the finite dimensional model (2.1.1), corresponding asymptotic distribution is complicated for constructing confidence regions and perform tests on regression parameter. Hence, a practically reasonable approximation to the distribution of the Lasso estimator is necessary for the purpose of inference. A general alternative approach of approximation besides the asymptotic distribution is the corresponding bootstrap distribution. When the design is non-random and errors are homoscedastic, KF(00) considered the residual bootstrap [cf. Freedman (1981)] in Lasso regression. Chatterjee and Lahiri (2010) investigated the asymptotic properties of usual residual bootstrap in Lasso and found that it fails drastically in approximating the distribution of Lasso when the regression parameter vector is sparse. More precisely, they found that the asymptotic distribution of the residual bootstrapped Lasso estimator is a non-degenerate random measure on $\mathcal{R}^p$ when one or more components of the regression parameter is zero.

To make residual bootstrap work, Chatterjee and Lahiri (2011) proposed a modification in residual bootstrap method and showed that it is consistent in approximating the distribution of the Lasso estimator when the design is non-random and errors are homoscedastic. Recently, Camponovo (2015) investigated paired bootstrap [cf. Freedman (1981)] in Lasso when the design is random and the errors are heteroscedastic. He pointed out two reasons which are resulting in the failure of the paired bootstrap method and introduced a modification analogous to the modification constructed by Chatterjee and Lahiri (2011) and established its consistency in estimating the distribution of the Lasso estimator.

In this chapter, we construct a generalized bootstrap method based on the perturbation approach [cf. Jing et. al (2001), Das and Lahiri (2016)] in both random and non-random design case and when the errors are allowed to be heteroscedastic. See Section 2 for details. We show that the conditional distribution of the corresponding bootstrapped Lasso estimator is consistent in estimating the complicated distribution of the original Lasso estimator, even in sparse set-up. To prove our results, first we show strong consistency of Lasso, extending the result of Chatterjee and Lahiri (2010) for the case of non-random design and homoscedastic errors, and then we employ the techniques of finding asymptotic distribution of the M-estimators with non-differentiable convex objective functions [cf. Pollard (1991), Hjort and Pollard (1993), Geyer (1994, 1996), Kato (2009)], similar to what was employed by KF(00), to find the asymptotic distribution of the original Lasso estimator.

We conclude this section with a brief literature review. The perturbation bootstrap was introduced by Jin, Ying, and Wei (2001) as a resampling procedure where the objective function has a U-process structure. Work on the perturbation bootstrap in the linear regression setup is limited. Some work has been carried out by Chatterjee and Bose (2005), Minnier et al. (2011), Zhou, Song and Thompson (2012), Das and Lahiri (2016), Das et al. (2017). As a variable selection procedure, Tibshirani (1996) introduced the Lasso. Literature on bootstrap methods in Lasso is very limited. Asymptotic properties of standard residual bootstrap in Lasso has been investigated by Knight and Fu (2000) and Chatterjee and Lahiri (2010). Chatterjee and Lahiri (2010) considered regression model (2.1.1) with non-random covariates and homoscedastic errors and showed that residual bootstrap fails drastically in approximating distribution of Lasso when one or more regression coefficients are zero. Subsequently Chatterjee and Lahiri (2011) developed a thresholding approach which rectifies this problem. Recently Camponovo (2015) developed a modified pairs bootstrap for Lasso in regression model (2.1.1) with random covariates and heteroscedastic errors and established its distributional consistency.

The rest of the chapter is organized as follows. Our proposed perturbation bootstrap method for Lasso is described in Section 2.2. The naive perturbation bootstrap and its shortcomings are also discussed in this Section. Main results concerning the estimation properties of the perturbation bootstrap are given in Section 2.3. The proofs are provided in Section 2.4. A moderate simulation study is presented in Section 2.5. Section 2.6 states concluding remarks.

2.2 Description of the Bootstrap Method

This section is divided into two parts. The first sub-section describes the naive way of defining perturbation bootstrapped Lasso estimator and subsequently points out its shortcomings. The second sub-section describes our proposed modification in defining the perturbation bootstrap version of the Lasso estimator.

2.2.1 Naive Perturbation Bootstrap

The perturbation bootstrap version of the least square estimator is defined as the minimizer of the objective function $\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{t}^*)^2 G_i^*$ [cf. Jin et al. (2001), Das and Lahiri (2016)]. Thus the natural way of defining corresponding Lasso estimator is

$$\check{\boldsymbol{\beta}}_n^* = \arg\min_{\boldsymbol{t}^*} \Big[\sum_{i=1}^{n} (y_i - \boldsymbol{X}_i'\boldsymbol{t}^*)^2 G_i^* + \lambda_n^* \sum_{j=1}^{p} |t_j^*| \Big], \tag{2.2.1}$$

similar to how perturbation bootstrap version is defined for adaptive lasso and SCAD estimators in Minnier et al. (2011). Here λ_n^* is a penalty parameter having same asymptotic property as λ_n. Again $G_1^*, \ldots, G_n^*$ are n independent copies of a non-degenerate random variable $G^* \in [0, \infty)$ having expectation μ_{G^*} and variance $\sigma_{G^*}^2$ with $\sigma_{G^*}^2 = \mu_{G^*}^2$. Note that the restriction $\sigma_{G^*}^2 = \mu_{G^*}^2$, that is on the distribution of G^*, is more general than the assumptions $\mu_{G^*} = 1$ and $\sigma_{G^*}^2 = 1$, considered in Minnier et al. (2011). Two immediate choices of the family of distribution of G^* are Exponential(λ) and Beta(α, β) with $\alpha = (\beta - \alpha)/(\alpha + \beta)$. Other choices can be found easily by investigating the generalized beta family of distributions.

Now consider a sequence of constants $\{a_n\}_{n \geq 1}$ such that $a_n + (n^{-1/2} \log n) a_n^{-1} \to 0$ as $n \to \infty$ and a $\sqrt{n}$-consistent estimator $\check{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$. Define the modified estimator

$\tilde{\boldsymbol{\beta}}_n = (\tilde{\beta}_{n,1}, \ldots, \tilde{\beta}_{n,p})'$ with $\tilde{\beta}_{n,j} = \breve{\beta}_{n,j}\mathbf{1}(|\breve{\beta}_{n,j}| > a_n)$, $\mathbf{1}(\cdot)$ being the indicator function. It follows from the results of KF(00) and Camponovo (2015) that the Lasso estimator $\hat{\boldsymbol{\beta}}_n$ is $\sqrt{n}$-consistent and hence for the sake of definiteness we take $\breve{\boldsymbol{\beta}}_n = \hat{\boldsymbol{\beta}}_n$. We need to use $\sqrt{n}(\breve{\boldsymbol{\beta}}_n^* - \tilde{\boldsymbol{\beta}}_n)$, in place of $\sqrt{n}(\breve{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)$, to approximate the distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$. The reason is that $\tilde{\boldsymbol{\beta}}_n$ can capture the signs of the zero components with probability close to 1 for sufficiently large n whereas the original Lasso estimator $\hat{\boldsymbol{\beta}}_n$ can not, as pointed out in Section 2 of Chatterjee and Lahiri (2011).

Now to point out the shortcomings of the naive perturbation bootstrapped Lasso estimator $\breve{\boldsymbol{\beta}}_n^*$, define $\breve{\boldsymbol{u}}_n^* = \sqrt{n}(\breve{\boldsymbol{\beta}}_n^* - \tilde{\boldsymbol{\beta}}_n)$. Then we have

$$\breve{\boldsymbol{u}}_n^* = \underset{\boldsymbol{v}^*}{\arg\min} \left[\boldsymbol{v}^{*\prime} \boldsymbol{C}_n^* \boldsymbol{v}^* - 2\boldsymbol{v}^{*\prime} \boldsymbol{W}_n^* + \lambda_n^* \sum_{j=1}^{p} \left(|\tilde{\beta}_{j,n} + \frac{v_j^*}{\sqrt{n}}| - |\tilde{\beta}_{j,n}| \right) \right] \tag{2.2.2}$$

where $\boldsymbol{C}_n^* = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i' G_i^*$, $\boldsymbol{W}_n^* = n^{-1/2}\sum_{i=1}^n \tilde{\epsilon}_i \boldsymbol{X}_i G_i^*$ and $\tilde{\epsilon}_i = y_i - \boldsymbol{x}_i'\tilde{\boldsymbol{\beta}}_n$, $i \in \{1, \ldots, n\}$. Clearly $\boldsymbol{W}_n^*$ is a sequence of non-centered random vectors and hence the asymptotic mean of that quantity is not necessarily $\mathbf{0}$, needed for consistency in estimating the distribution of the Lasso estimator [see in (2.4.8) that $\tilde{\boldsymbol{W}}_n^*$ is properly centered in $\boldsymbol{Z}_n^*(\boldsymbol{v}^*)$]. Additionally from computational perspective, it would be plausible if $\boldsymbol{C}_n^*$ can be replaced by $\mu_{G^*}\boldsymbol{C}_n$, since $\boldsymbol{C}_n^*$ needs to be computed at each bootstrap iteration. If we implement the modification described in the next sub-section, then both the theoretical and computational shortcomings of the naive method become resolved and distributional consistency is achieved by perturbation bootstrap.

2.2.2 Modified Perturbation Bootstrap

In order to rectify the naive method, we need to incorporate a perturbed least-squares criterion involving predicted values $\tilde{y}_i = \boldsymbol{x}_i'\tilde{\boldsymbol{\beta}}_n$, $i = 1, \ldots, n$, besides the the same involving the observed values $y_1, \ldots, y_n$ in the objective function. Formally, the modified perturbation bootstrap version of the Lasso estimator is defined as

$$\hat{\boldsymbol{\beta}}_n^* = \arg\min_{\boldsymbol{t}^*} \Bigg[\sum_{i=1}^{n} (y_i - \boldsymbol{X}_i'\boldsymbol{t}^*)^2 (G_i^* - \mu_{G^*}) + \sum_{i=1}^{n} (\tilde{y}_i - \boldsymbol{X}_i'\boldsymbol{t}^*)^2 (2\mu_{G^*} - G_i^*) + \mu_{G^*}\lambda_n \sum_{j=1}^{p} |t_j^*| \Bigg]. \tag{2.2.3}$$

Now we point out an important computational characteristic of our proposed bootstrap method. For $i = 1, \ldots, n$, set $z_i = \tilde{y}_i + \tilde{\epsilon}_i \mu_{G^*}^{-1}(G_i^* - \mu_{G^*})$, where $\tilde{\epsilon}_i = y_i - \tilde{y}_i$. Then the perturbation bootstrap version of the Lasso estimator, $\hat{\boldsymbol{\beta}}_n^*$, can be expressed as

$$\hat{\boldsymbol{\beta}}_n^* = \arg\min_{\boldsymbol{t}} \Big[\sum_{i=1}^{n} \big(z_i - \boldsymbol{X}_i'\boldsymbol{t}\big)^2 + \lambda_n \sum_{j=1}^{p} |t_j| \Big] \tag{2.2.4}$$

This representation indicates that $\hat{\boldsymbol{\beta}}_n^*$ is a Lasso estimator corresponding to the pseudo observations $\{z_1, \ldots, z_n\}$ and the design vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Hence we can employ existing computationally fast algorithms, eg. LARS [cf. Efron et al. (2004)], "one-at-a-time" coordinate-wise descent [cf. Friedman et al. (2007)], even for finding the perturbation bootstrap version of the Lasso estimator. In the next section, we present the theoretical properties of the proposed perturbation bootstrap method.

2.3 Main Results

2.3.1 Notations

Suppose, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ denotes the true value of the regression parameter vector. Let, $\mathcal{B}(\mathcal{R}^p)$ denotes the Borel sigma-field defined on $\mathcal{R}^p$ and $\rho(\cdot, \cdot)$ denotes the Prokhorov metric on the collection of all probability measures on $\big(\mathcal{R}^p, \mathcal{B}(\mathcal{R}^p)\big)$. Set $\mathcal{A} = \{j : \beta_j \neq 0\}$ and $p_0 = |\mathcal{A}|$, supposing, without loss of generality, that $\mathcal{A}_n = \{1, \ldots, p_0\}$. $\mathscr{E}$ is the sigma-field generated by $\{\epsilon_i : i \geq 1\}$ when the design is non-random whereas when the design is non-random, $\mathscr{E}$ is the sigma-field generated by $\{(\epsilon_i, \boldsymbol{X}_i) : i \geq 1\}$. Further assume that $\mathbf{F}_n$ is the distribution of $\boldsymbol{T}_n = \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$. The bootstrap version of $\boldsymbol{T}_n$ is $\boldsymbol{T}_n^* = \sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \tilde{\boldsymbol{\beta}}_n)$ and $\hat{\mathbf{F}}_n$ is the conditional distribution of $\boldsymbol{T}_n^*$ given $\mathscr{E}$. Define $\mathbf{F}_\infty$ to be the limit distribution of $\boldsymbol{T}_n$. Suppose, $\mathbf{P}_*$ and $\mathbf{E}_*$ respectively denote the bootstrap probability and bootstrap expectation conditional on data.

2.3.2 Results in the case of non-random designs

In this subsection, we consider the linear regression model (2.1.1) when the design vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are fixed. The errors $\epsilon_1, \ldots, \epsilon_n$ are independent; but may not be identically distributed. Thus ϵ_i may be assumed to depend on $\boldsymbol{X}_i$, $i \in \{1, \ldots, n\}$. In this setup, we show that $\hat{\mathbf{F}}_n$ is a valid approximation of $\mathbf{F}_n$, as formally stated in the next theorem.

Theorem 2.3.1 *Suppose the following regularity conditions hold:*

(A.1) $n^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \to \boldsymbol{C}$ *as* $n \to \infty$, *for some positive definite matrix* $\boldsymbol{C}$.

(A.2) $n^{-1}\sum_{i=1}^{n}||\boldsymbol{X}_i||^{4+\delta} = O(1)$ *for some* $\delta > 0$.

(A.3) $\mathbf{E}\epsilon_i = 0$ *for all* $i \in \{1,\ldots,n\}$ *and* $n^{-1}\sum_{i=1}^{n}\mathbf{E}|\epsilon_i|^{4+\delta} = O(1)$.

(A.4) $n^{-1}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i'\mathbf{E}\epsilon_i^2 \rightarrow \boldsymbol{\Sigma}$ *as* $n \rightarrow \infty$, *for some positive definite matrix* $\boldsymbol{\Sigma}$.

(A.5) $\lambda_n/\sqrt{n} \rightarrow \lambda_0 \in [0,\infty)$.

(A.6) G_i^* *and* ϵ_i *are independent for all* $i \in \{1,\ldots,n\}$ *and* $\mathbf{E}G_1^{*3} < \infty$.

Then it follows that $\rho(\hat{\mathbf{F}}_n, \mathbf{F}_n) \rightarrow 0$ *as* $n \rightarrow \infty$, *with probability* 1.

2.3.2.1 Incompetency of Residual Bootstrap when Errors are Heteroscedastic

First, let us briefly describe the residual bootstrap method in Lasso, developed in Chatterjee and Lahiri (2011). The modified residuals $\{\tilde{\epsilon}_1,\ldots,\tilde{\epsilon}_n\}$ by $\tilde{\epsilon}_i = y_i - \boldsymbol{X}_i'\tilde{\beta}_n$. Suppose $\bar{\epsilon}_n$ is the mean of the residuals. Then select a random sample $\{r_1^*,\ldots,r_n^*\}$ from $\{(\tilde{\epsilon}_1 - \bar{\epsilon}_n),\ldots,(\tilde{\epsilon}_n - \bar{\epsilon}_n)\}$ and define

$$y_i^* = \boldsymbol{X}_i'\tilde{\beta}_n + r_i^* \qquad, i = 1,\ldots,n$$

Then the residual bootstrapped Lasso estimator is defined as

$$\breve{\boldsymbol{\beta}}_n^* = \arg\min_{\boldsymbol{t}^*}\Big[\sum_{i=1}^{n}(y_i^* - \boldsymbol{X}_i'\boldsymbol{t}^*)^2 + \lambda_n\sum_{j=1}^{p}|t_j^*|\Big]. \tag{2.3.1}$$

To prove incompetency of residual bootstrap in Lasso, our argument is similar to as in Liu (1988) for least square estimator (LSE). When $p = 1$, Liu (1988) showed that the conditional variance of residual bootstrapped LSE is not a consistent estimator of the variance of LSE. Wagener and Dette (2011) showed that under the conditions (A.1), (A.4), (A.5), $\mathbf{E}\epsilon_i = 0$ for all $i \in \{1,\ldots,n\}$ and $n^{-1}\max_i||\boldsymbol{X}_i||^2\sigma_i^2 \rightarrow 0$ [Note that this condition follows from (A.2) and (A.3)]

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \arg\min_{\boldsymbol{v}} Z(\boldsymbol{v})$$

where

$$Z(\boldsymbol{v}) = \left[\boldsymbol{v}'\boldsymbol{C}\boldsymbol{v} - 2\boldsymbol{v}'\boldsymbol{W} + \lambda_0\Big(\sum_{j=1}^{p_0} sgn(\beta_j)v_j + \sum_{j=p_0+1}^{p} |v_j|\Big)\right]$$

where $\boldsymbol{W}$ follows $N(\boldsymbol{0}, \boldsymbol{\Sigma})$, when $\mathcal{A} = \{1, \ldots, p_0\}$. The following proposition implies the incompetency of residual bootstrap in case of Lasso when errors are heteroscedastic.

Proposition 2.3.1 *Suppose, conditions (A.1)-(A.3) and (A.5) hold. Also assume that* $n^{-1}\sum_{i=1}^{n} \mathbf{E}\epsilon_i^2 \to s^2\ (> 0)$. *Then the following is true.*

$$\sqrt{n}(\breve{\boldsymbol{\beta}}_n^* - \tilde{\boldsymbol{\beta}}) \xrightarrow{d} \arg\min_{\boldsymbol{v}} \tilde{Z}(\boldsymbol{v})$$

$$\textit{with}\ \ \tilde{Z}(\boldsymbol{v}) = \left[\boldsymbol{v}'\boldsymbol{C}\boldsymbol{v} - 2\boldsymbol{v}'\tilde{\boldsymbol{W}} + \lambda_0\Big(\sum_{j=1}^{p_0} sgn(\beta_j)v_j + \sum_{j=p_0+1}^{p} |v_j|\Big)\right]$$

where $\tilde{\boldsymbol{W}}$ *follows* $N(\boldsymbol{0}, s^2\boldsymbol{C})$.

Remark 2.3.1 *Note that* $\boldsymbol{\Sigma} = s^2\boldsymbol{C}$ *when the errors are homoscedastic and hence* $\tilde{Z}(\boldsymbol{v})$ *becomes* $Z(\boldsymbol{v})$. *Therefore residual bootstrap works in homoscedastic setting, as found in Chatterjee and Lahiri (2011). But when the errors become heteroscedastic, then residual bootstrap is no longer valid in approximating the distribution of Lasso as long as* $\boldsymbol{\Sigma} \neq s^2\boldsymbol{C}$. *This is indeed the case in most of the situations whenever the variance of the errors depend on the covariates. On the other hand, perturbation bootstrap can consistently approximate even when the regression errors are no longer homoscedastic. This is an advantage of perturbation bootstrap over the* *residual bootstrap.*

2.3.3 The Result in the case of random designs

In this subsection, we establish that $\hat{\mathbf{F}}_n$ approximates $\mathbf{F}_n$ consistently when the design vectors are random and errors $\epsilon_1, \ldots, \epsilon_n$ are independent, but not necessarily identically distributed. The result is stated in the theorem 2.3.2 below.

Theorem 2.3.2 *Suppose the following regularity conditions hold:*

(B.1) $\mathbf{E}\Big[n^{-1}\sum_{i=1}^{n} \boldsymbol{X}_i\boldsymbol{X}_i'\Big] \rightarrow \boldsymbol{C}$ *as* $n \rightarrow \infty$, *for some positive definite matrix* $\boldsymbol{C}$.

(B.2) $n^{-1}\sum_{i=1}^{n} \mathbf{E}||\boldsymbol{X}_i||^{4+\delta} = O(1)$ *for some* $\delta > 0$.

(B.3) $\mathbf{E}(\epsilon_i|\boldsymbol{X}_i) = 0$ *for all* $i \in \{1, \ldots, n\}$ *and* $n^{-1}\sum_{i=1}^{n} \mathbf{E}|\epsilon_i|^{4+\delta} = O(1)$.

(B.4) $\mathbf{E}\Big(n^{-1}\sum_{i=1}^{n} \boldsymbol{X}_i\boldsymbol{X}_i'\epsilon_i^2\Big) \rightarrow \boldsymbol{\Sigma}$ *as* $n \rightarrow \infty$, *for some positive definite matrix* $\boldsymbol{\Sigma}$.

(B.5) $\lambda_n/\sqrt{n} \rightarrow \lambda_0 \in [0, \infty)$.

(B.6) G_i^* *and* $(\epsilon_i, \boldsymbol{X}_i)$ *are independent for all* $i \in \{1, \ldots, n\}$ *and* $\mathbf{E}G_1^{*3} < \infty$.

Then it follows that $\rho(\hat{\mathbf{F}}_n, \mathbf{F}_n) \rightarrow 0$ *as* $n \rightarrow \infty$, *with probability* 1.

Remark 2.3.2 *Theorem 2.3.1 and 2.3.2 establishes distributional consistency of our proposed bootstrap method in approximating the distribution of Lasso. Formally, it follows that on a set of probability* 1, $\hat{\mathbf{F}}_n \rightarrow \mathbf{F}_\infty$ *in distribution and hence* $\mathbf{F}(B) - \hat{\mathbf{F}}(B) \rightarrow 0$ *as* $n \rightarrow \infty$ *for all* $B \in \mathcal{R}^p$ *with* $\boldsymbol{F}_\infty(\partial B) = 0$. *Since the form of the asymptotic distribution of* $\hat{\boldsymbol{\beta}}$ *is quite complicated [cf. Chatterjee and Lahiri (2011), Camponovo (2015)], validity of perturbation bootstrap approximation, as stated in the theorems, is essential if one wants to infer about* $\boldsymbol{\beta}$ *based on the Lasso estimator* $\hat{\boldsymbol{\beta}}_n$.

Remark 2.3.3 *One needs to implement pairs or residual bootstrap depending on whether the covariates are random or non-random and also there is significant difference in implementation between these two procedures. On the other hand Theorem 2.3.1 and 2.3.2 implies that one can implement perturbation bootstrap without thinking about the nature of the covariates. This means that the perturbation bootstrap is more general than the resample-based bootstrap (residual and pairs) of Freedman (1981) in case of Lasso.*

2.4 Proofs

Suppose, K, K_1, K_2 denote generic constants which do not depend on n. Let, $(C.1) - (C.6)$ refers to $(A.1) - (A.6)$ when the design matrix is non-random and to $(B.1) - (B.6)$ when the design matrix is random. Recall that $\mathbf{P}_*$ denotes the conditional probability given $\mathscr{E}$ and $\mathbf{E}_*(\cdot) = \mathbf{E}(\cdot|\mathscr{E})$. Define $\boldsymbol{C}_n = n^{-1}\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i'$ and $\tilde{\boldsymbol{W}}_n^* = n^{-1/2}\sum_{i=1}^n \tilde{\epsilon}_i\boldsymbol{X}_i(G_i^* - \mu_{G^*})$. For a random vector $\boldsymbol{Z}$ and a sigma-field $\mathcal{C}$, suppose $\mathcal{L}(\boldsymbol{Z})$ denotes the distribution of $\boldsymbol{Z}$ and $\mathcal{L}(\boldsymbol{Z}|\mathcal{C})$ denotes the conditional distribution of $\boldsymbol{Z}$ given $\mathcal{C}$. For simplicity set $\mathcal{L}(\boldsymbol{Z}|\sigma(\boldsymbol{W})) = \mathcal{L}(\boldsymbol{Z}|\boldsymbol{W})$. Also suppose that $||\cdot||$ is the euclidean norm. We will write $w.p.$ to denote "with probability". " $\xrightarrow{d}$ " denotes the convergence in distribution.

Now we state some lemmas before proving theorem 2.3.1 and 2.3.2.

Lemma 2.4.1 *Let $\{Z_n\}$ be a sequence of independent random variables such that $n^{-\alpha_n}\sum_{i=1}^n$ $\mathbf{E}|X_n|^{\alpha_n} = O(1)$ as $n \to \infty$ for some $\alpha_n \in [1, 2]$. Then as $n \to \infty$,*

$$n^{-1}\sum_{i=1}^n (Z_j - \mathbf{E}Z_j) = o(1), \quad w.p.\ 1.$$

The above Lemma is stated as Theorem 8.4.6 in Atreya and Lahiri (2006).

Lemma 2.4.2 *Suppose conditions* $(C.2)-(C.5)$ *hold. Then the following holds:*

$$\Big|\Big|n^{-1/2}\sum_{i=1}^{n}\epsilon_i \boldsymbol{X}_i\Big|\Big| = o(\log n), \ \ w.p.\ 1.$$

Proof of Lemma 2.4.2. Assume $s_n^2 = \max_{j\in\{1,\ldots,p\}}\sum_{i=1}^{n}\mathbf{E}X_{ij}^2\epsilon_i^2$. Note that by assumption (C.4), $s_n^2 = O(n)$. Rest of the proof now follows in the same line as in the Lemma 4.1 of Chatterjee and Lahiri (2010).

Lemma 2.4.3 *Suppose conditions* $(C.1)-(C.5)$ *hold. Then*

$$\big|\big|\boldsymbol{T}_n\big|\big| = O(\log n), \ \ w.p.\ 1. \tag{2.4.1}$$

Proof of Lemma 2.4.3. lemma 2.4.3 can be proved using same line of arguments as in the proof of the Lemma 4.2 in Chatterjee and Lahiri (2010). We omit the details to save space.

Lemma 2.4.4 *Suppose conditions* $(C.1)-(C.3)$ *hold. Then the following is true:*

$$n^{-1}\Big|\Big|\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\tilde{\epsilon}_i^2 - \mathbf{E}\Big(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\epsilon_i^2\Big)\Big|\Big| + n^{-3/2}\sum_{i=1}^{n}||\boldsymbol{X}_i||^3|\tilde{\epsilon}_i|^3 = o(1), \ w.p.\ 1. \tag{2.4.2}$$

Proof of Lemma 2.4.4. First note that

$$\begin{aligned}
&n^{-1}\Big|\Big|\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\tilde{\epsilon}_i^2 - \mathbf{E}\Big(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\epsilon_i^2\Big)\Big|\Big| \\
\le&\Big|\Big|n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\big(\tilde{\epsilon}_i^2-\epsilon_i^2\big)\Big|\Big| + \Big|\Big|n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\big(\epsilon_i^2-\mathbf{E}\epsilon_i^2\big)\Big|\Big| \\
\le&\sum_{j,k=1}^{p}\Big|n^{-1}\sum_{i=1}^{n}X_{ij}X_{ik}\big(\tilde{\epsilon}_i^2-\epsilon_i^2\big)\Big| + \sum_{j,k=1}^{p}\Big|n^{-1}\sum_{i=1}^{n}X_{ij}X_{ik}\big(\epsilon_i^2-\mathbf{E}\epsilon_i^2\big)\Big|.
\end{aligned} \tag{2.4.3}$$

Again since $(\tilde{\epsilon}_i - \epsilon_i)^2 = \left[\boldsymbol{X}_i'(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_n)\right]\left[\boldsymbol{X}_i'(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_n) + 2\epsilon_i\right]$, hence it follows that

$$\begin{aligned}\big|n^{-1}\sum_{i=1}^{n} X_{ij}X_{ik}\big(\tilde{\epsilon}_i^2 - \epsilon_i^2\big)\big| \leq & n^{-1}\Big(\sum_{i=1}^{n}||\boldsymbol{X}_i||^4\Big)||\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}||^2 \\ & + 2n^{-1}\Big(\sum_{i=1}^{n}||\boldsymbol{X}_i||^3|\epsilon_i|\Big)||\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}||.\end{aligned}$$

Now note that by Lemma 2.4.3 $||\sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta})|| = O(\log n)$ $w.p.$ 1. Using this, Lemma 2.4.1 and the fact that

$$\Big[\max_{i\in\{1,\ldots,n\}}||\boldsymbol{X}_i||\big(1+|\epsilon_i|\big)\Big] \leq \Big(\sum_{i=1}^{n}||\boldsymbol{X}_i||^2\big(1+|\epsilon_i|^2\big)\Big)^{1/2} = O(n^{1/2}), \qquad w.p.\ 1, \tag{2.4.4}$$

we have

$$\big|n^{-1}\sum_{i=1}^{n} X_{ij}X_{ik}\big(\tilde{\epsilon}_i^2 - \epsilon_i^2\big)\big| = o(1), \qquad w.p.\ 1.$$

On the other hand, following is the direct implication of Lemma 2.4.1

$$\big|n^{-1}\sum_{i=1}^{n} X_{ij}X_{ik}\big(\epsilon_i^2 - \mathbf{E}\epsilon_i^2\big)\big| = o(1), \qquad w.p.\ 1.$$

This shows that the first term on the LHS of 2.4.3 is $o(1)$, $w.p.$ 1. Next consider the second term on the LHS of 2.4.3. It is easy to see that

$$\begin{aligned}\sum_{i=1}^{n}||\boldsymbol{X}_i||^3\Big(|\tilde{\epsilon}_i|^3 - |\epsilon_i|^3\Big) \leq & \Big(\sum_{i=1}^{n}||\boldsymbol{X}_i||^6\Big)||\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}||^3 \\ & + 3\Big(\sum_{i=1}^{n}||\boldsymbol{X}_i||^5|\epsilon_i|\Big)||\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}||^2 + 3\Big(\sum_{i=1}^{n}||\boldsymbol{X}_i||^4|\epsilon_i|^2\Big)||\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}||\end{aligned} \tag{2.4.5}$$

Using (2.4.4), Hölder's inequality, Lemma 2.4.1 and Lemma 2.4.3, one can show from (2.4.5) that

$$n^{-3/2}\sum_{i=1}^{n}||\boldsymbol{X}_i||^3\Big(|\tilde{\epsilon}_i|^3-|\epsilon_i|^3\Big)=o(1), \qquad w.p.\ 1$$

Again it is easy to show that $n^{-3/2}\sum_{i=1}^{n}||\boldsymbol{X}_i||^3|\epsilon_i|^3=o(1)$ $w.p.$ 1.

Hence Lemma 2.4.4 follows.

Lemma 2.4.5 *Suppose conditions* $(C.1)-(C.6)$ *hold. Then*

$$\mathcal{L}\Big(\mu_{G^*}^{-1}\tilde{\boldsymbol{W}}_n^*\Big|\mathscr{E}\Big)\xrightarrow{d}\mathbf{N}\big(\mathbf{0},\boldsymbol{\Sigma}\big),\ \ w.p.\ 1, \tag{2.4.6}$$

Proof of Lemma 2.4.5. Consider $A\in\mathscr{E}$ such that $\mathbf{P}(A)=1$ and on the the set A the following holds:

$$n^{-1}\Big|\Big|\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\tilde{\epsilon}_i^2-\mathbf{E}\Big(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\epsilon_i^2\Big)\Big|\Big|+n^{-3/2}\sum_{i=1}^{n}||\boldsymbol{X}_i||^3|\tilde{\epsilon}_i|^3=o(1).$$

Hence, using Cramer-Wold device, it is enough to show that, on A,

$$\mathcal{L}\Big(\boldsymbol{t}'\tilde{\boldsymbol{W}}_n^*\le x\mu_{G^*}\Big|\mathscr{E}\Big)\xrightarrow{d}\mathbf{N}\big(\mathbf{0},\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}\big), \tag{2.4.7}$$

for any $\boldsymbol{t}\in\mathcal{R}^p$ and $\boldsymbol{t}\neq\mathbf{0}$. Now due to Lemma 2.4.4 and assumption (C.4), (2.4.7) follows if we can show that on the set A

$$\sup_{x\in\mathcal{R}^p}\Big|\mathbf{P}_*\Big(\boldsymbol{t}'\tilde{\boldsymbol{W}}_n^*\le x\mu_{G^*}\Big)-\Phi\big(xs_n^{-1}(\boldsymbol{t})\big)\Big|=o(1)$$

where $s_n^2(\boldsymbol{t})=\boldsymbol{t}'\Big[n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\tilde{\epsilon}_i^2\Big]\boldsymbol{t}$. Now due to Berry-Essen Theorem [cf. Bhattacharya and Rao (1986)], on the set A

$$\sup_{x\in\mathcal{R}^p}\Big|\mathbf{P}_*\Big(\boldsymbol{t}'\tilde{\boldsymbol{W}}_n^*\le x\mu_{G^*}\Big)-\Phi(xs_n^{-1}(\boldsymbol{t}))\Big|$$

$$\leq (2.75) \frac{\sum_{i=1}^n \mathbf{E}_* \left| n^{-1/2} \boldsymbol{t}' \boldsymbol{X}_i \tilde{\epsilon}_i (G_i^* - \mu_{G^*}) \right|^3}{\mu_{G^*}^3 \left\{ \boldsymbol{t}' \left[n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i' \tilde{\epsilon}_i^2 \right] \boldsymbol{t} \right\}^{3/2}}$$

$$\leq (2.75) \frac{||\boldsymbol{t}||^3 n^{-3/2} \sum_{i=1}^n ||\boldsymbol{X}_i||^3 |\tilde{\epsilon}_i|^3 \mathbf{E}_* |G_i^* - \mu_{G^*}|^3}{\mu_{G^*}^3 s_n^3(\boldsymbol{t})}$$

$$= o(1).$$

The last equality follows from Lemma 2.4.4. Therefore Lemma 2.4.5 follows.

Proof of Theorem 2.3.1. The Perturbation bootstrap version of the Lasso estimator is given by

$$\hat{\boldsymbol{\beta}}_n^* = \arg\min_{\boldsymbol{t}^*} \Bigg[\sum_{i=1}^n (y_i - \boldsymbol{x}_i' \boldsymbol{t}^*)^2 (G_i^* - \mu_{G^*}) + \sum_{i=1}^n [\boldsymbol{x}_i' (\boldsymbol{t}^* - \tilde{\boldsymbol{\beta}}_n)]^2 (2\mu_{G^*} - G_i^*) + \mu_{G^*} \lambda_n \sum_{j=1}^p |t_j^*| \Bigg].$$

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

$$\hat{\boldsymbol{u}}_n^* = \arg\min_{\boldsymbol{v}^*} \left[\boldsymbol{v}^{*\prime} \boldsymbol{C}_n \boldsymbol{v}^* - 2\mu_{G^*}^{-1} \boldsymbol{v}^{*\prime} \tilde{\boldsymbol{W}}_n^* + \lambda_n \sum_{j=1}^p \left(|\tilde{\beta}_{j,n} + \frac{v_j^*}{\sqrt{n}}| - |\tilde{\beta}_{j,n}| \right) \right]$$

$$= \arg\min_{\boldsymbol{v}^*} \boldsymbol{Z}_n^*(\boldsymbol{v}^*) \quad \text{(say)}. \tag{2.4.8}$$

Define the set $B \in \mathscr{E}$ such that $\mathbf{P}(B) = 1$ and on the the set B, (2.4.1), (2.4.2) and (2.4.6) hold. Note that due to the definition of $\tilde{\boldsymbol{\beta}}_n$, there exists $N(\omega)$ for each $\omega \in B$ such that for $n > N(\omega)$,

$$\begin{cases} \tilde{\beta}_{j,n} = \tilde{\beta}_{j,n} \ \text{ and } \ sgn(\tilde{\beta}_{j,n}) = sgn(\beta_j) \text{ for } j \in \mathcal{A} \\ \tilde{\beta}_{j,n} = 0 \text{ for } j \in \{1, \ldots, p\} \setminus \mathcal{A} \end{cases}$$

Therefore, on the set B,

$$\boldsymbol{Z}_n^*(\boldsymbol{v}) \xrightarrow{d} Z(\boldsymbol{v})$$

where $Z(\boldsymbol{v}) = \left[\boldsymbol{v}'\boldsymbol{C}\boldsymbol{v} - 2\boldsymbol{v}'\boldsymbol{W} + \lambda_0\Big(\sum_{j=1}^{p_0} sgn(\beta_j)v_j + \sum_{j=p_0+1}^{p}|v_j|\Big)\right]$ and $\boldsymbol{W}$ follows $N(\boldsymbol{0}, \boldsymbol{\Sigma})$ distribution. Hence it follows, by the results of Geyer(1996), that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \tilde{\boldsymbol{\beta}}_n) \xrightarrow{d} \arg\min_{\boldsymbol{v}} Z(\boldsymbol{v})$$

Again it is shown in Lemma 3.1 of Wagener and Dette (2012) that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \arg\min_{\boldsymbol{v}} Z(\boldsymbol{v})$$

Therefore Theorem 2.3.1 follows.

Proof of Proposition 2.3.1. Writing $\tilde{\boldsymbol{u}}_n^* = \sqrt{n}(\breve{\boldsymbol{\beta}}_n^* - \tilde{\boldsymbol{\beta}}_n)$, we have from (2.3.1)

$$\tilde{\boldsymbol{u}}_n^* = \arg\min_{\boldsymbol{v}} \left[\boldsymbol{v}'\boldsymbol{C}_n\boldsymbol{v} - 2\boldsymbol{v}'\left[n^{-1/2}\sum_{i=1}^{n}\boldsymbol{X}_i r_i^*\right] + \lambda_n \sum_{j=1}^{p}\left(|\tilde{\beta}_{j,n} + \frac{v_j}{\sqrt{n}}| - |\tilde{\beta}_{j,n}|\right)\right]$$

Note that $Var\Big(n^{-1/2}\sum_{i=1}^{n}\boldsymbol{X}_i r_i^*|\mathscr{E}\Big) = \Big(n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\Big)\Big(n^{-1}\sum_{i=1}^{n}(\tilde{\epsilon}_i - \bar{\epsilon}_n)^2\Big)$ which converges to $s^2\boldsymbol{C}$ as $n \to \infty$. Therefore Proposition 2.3.1 follows through the same line of arguments, as in Theorem 2.3.1.

Proof of Theorem 2.3.2. Theorem 2.3.2 follows by retracting the steps of Theorem 2.3.1 and using Lemma 1 of Camponovo (2015).

2.5 Simulation results

We study through simulation the coverage of one-sided and two-sided 90% confidence intervals for individual regression coefficients separately for both the situations when design is non-random and random, constructed using Theorem 2.3.1 and 2.3.2. We compare empirical coverages of perturbation bootstrap (PB) with that of residual (RB) and paired bootstrap (PaB). We have also compared the empirical coverage of confidence region obtained by perturbation bootstrap with other two methods.

Under the settings

$$(n, p, p_0) \in \{(100, 10, 6), (500, 10, 6), (1000, 10, 6)\},$$

we generate n independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of $(X, Y) \in \mathbb{R}^p \times \mathbb{R}$ from the model $Y = X'\boldsymbol{\beta} + \epsilon$, where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$. We have considered two cases in both non-random and random design setup. Case I is when ϵ_i/s_i is generated from $(\chi^2_2 - 2)$ distribution and case II is when ϵ_i/s_i is generated from standard normal distribution with $s_i^2 = p^{-1}\sum_{j=1}^{p} |X_{ij}|^5$ where X_{ij} is the (i, j)th element of design matrix $\boldsymbol{X}$. The design vectors $X_i = (X_{i1}, \ldots, X_{ip})'$, $i \in \{1, \ldots, n\}$, are independent and identically distributed mean-zero multivariate normal random vector such that

$$\mathrm{Cov}(X_{1j}, X_{1k}) = \mathbf{1}(j = k) + 0.3^{|j-k|}\mathbf{1}(j \le p_0)\mathbf{1}(k \le p_0)\mathbf{1}(j \ne k)$$

for $1 \le j, k \le p$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ with β_j defined as $\beta_j = (3/4) + (1/4)j$ for $j = 1, \ldots, p_0$ and $\beta_j = 0$ for $j = p_0 + 1, \ldots, p$.

We compute the empirical coverage over 1000 simulated data sets of one- and two-sided confidence intervals for each regression coefficient under crossvalidation-selected values of λ_n. One sided intervals are right sided. To construct the bootstrap intervals for each of the 1000 simulated data sets, we generate 1200 independent samples from $Exp(1)$. The values of λ_n are thereafter held fixed for all bootstrap simulations on the same dataset.

Table 2.1: Empirical coverage of 90% confidence intervals for regression coefficients over 1000 simulations under $(n, p, p_0) = (100, 10, 6)$ by Lasso when the design is non-random.

Coverage and *(avg. width)* of two-sided 90% CIs

β_j	Case I		Case II	
	PB	*RB*	*PB*	*RB*
1	0.598	0.518	0.789	0.712
	(1.654)	*(1.283)*	*(1.061)*	*(0.850)*
1.25	0.710	0.473	0.784	0.602
	(1.981)	*(1.215)*	*(1.187)*	*(0.786)*
1.50	0.714	0.647	0.834	0.771
	(2.018)	*(1.747)*	*(1.228)*	*(1.052)*
1.75	0.763	0.733	0.856	0.829
	(1.743)	*(1.650)*	*(1.006)*	*(0.933)*
2.00	0.760	0.764	0.893	0.900
	(1.716)	*(1.785)*	*(0.986)*	*(1.024)*
2.25	0.805	0.781	0.885	0.858
	(1.829)	*(1.833)*	*(1.037)*	*(1.034)*
0	0.800	0.783	0.863	0.814
	(1.066)	*(0.927)*	*(0.629)*	*(0.547)*
0	0.845	0.782	0.873	0.805
	(1.087)	*(0.891)*	*(0.627)*	*(0.525)*
0	0.843	0.755	0.855	0.743
	(1.157)	*(0.877)*	*(0.691)*	*(0.526)*
0	0.861	0.685	0.852	0.706
	(1.381)	*(0.901)*	*(0.802)*	*(0.525)*
Coverage of one-sided 90% CIs				
1	0.629	0.626	0.824	0.796
1.25	0.701	0.671	0.822	0.751
1.50	0.769	0.763	0.863	0.827
1.75	0.801	0.807	0.876	0.859
2.00	0.822	0.826	0.893	0.900
2.25	0.837	0.847	0.864	0.870
0	0.873	0.852	0.878	0.856
0	0.733	0.697	0.847	0.794
0	0.832	0.753	0.840	0.798
0	0.865	0.731	0.854	0.771

Tables 2.1 and 2.2 are displaying the empirical coverage probabilities of 90% CIs for each of the regression coefficients under settings $(n, p, p_0) \in \{(100, 10, 6), (1000, 10, 6)\}$,

when the design is non-random. For the random design case, Tables 2.3 and 2.4 are displaying the empirical coverages when $(n, p, p_0) \in \{(100, 10, 6), (1000, 10, 6)\}$. Tables 2.1 and 2.2 are comparing our perturbation bootstrap method with residual bootstrap

Table 2.2: Empirical coverage of 90% confidence intervals for regression coefficients over 1000 simulations under $(n, p, p_0) = (1000, 10, 6)$ by Lasso when the design is non-random.

Coverage and *(avg. width)* of two-sided 90% CIs				
β_j	Case I		Case II	
	PB	*RB*	*PB*	*RB*
1	0.881	0.840	0.890	0.801
	(3.124)	*(2.565)*	*(1.847)*	*(1.502)*
1.25	0.884	0.828	0.893	0.818
	(3.562)	*(2.616)*	*(1.957)*	*(1.484)*
1.50	0.914	0.841	0.899	0.848
	(3.665)	*(3.133)*	*(2.004)*	*(1.741)*
1.75	0.895	0.858	0.904	0.852
	(3.331)	*(3.054)*	*(1.770)*	*(1.620)*
2.00	0.907	0.828	0.914	0.861
	(3.300)	*(3.172)*	*(1.795)*	*(1.719)*
2.25	0.893	0.760	0.895	0.845
	(3.473)	*(3.182)*	*(1.781)*	*(1.702)*
0	0.908	0.825	0.901	0.847
	(2.122)	*(1.798)*	*(1.149)*	*(0.969)*
0	0.892	0.815	0.893	0.806
	(2.238)	*(1.742)*	*(1.189)*	*(0.942)*
0	0.909	0.782	0.916	0.845
	(2.386)	*(1.753)*	*(1.246)*	*(0.975)*
0	0.920	0.747	0.886	0.821
	(2.535)	*(1.764)*	*(1.391)*	*(0.955)*
Coverage of one-sided 90% CIs				
1	0.875	0.848	0.882	0.838
1.25	0.889	0.854	0.893	0.848
1.50	0.895	0.859	0.891	0.865
1.75	0.891	0.874	0.892	0.862
2.00	0.904	0.862	0.910	0.876
2.25	0.908	0.825	0.900	0.863
0	0.881	0.848	0.915	0.877
0	0.901	0.836	0.888	0.831
0	0.882	0.839	0.912	0.849
0	0.895	0.803	0.882	0.850

and Tables 2.3 and 2.4 are comparing our perturbation bootstrap method with the paired bootstrap.

In the non-random design case, for which Table 2.1 and 2.2 display the empirical coverages of 90% CIs, the perturbation bootstrap intervals based on $\boldsymbol{T}_n^*$ perform much better than residual bootstrap intervals for both zero and non-zero regression coefficients. When $n = 1000$, perturbation bootstrap two-sided and one-sided intervals

Table 2.3: Empirical coverage of 90% confidence intervals for regression coefficients over 1000 simulations under $(n, p, p_0) = (100, 10, 6)$ by Lasso when the design is random.

Coverage and *(avg. width)* of two-sided 90% CIs				
β_j	Case I		Case II	
	PB	*PaB*	*PB*	*PaB*
1	0.642	0.683	0.827	0.834
	(1.408)	*(1.580)*	*(0.883)*	*(0.944)*
1.25	0.704	0.723	0.874	0.875
	(1.562)	*(1.723)*	*(0.940)*	*(0.995)*
1.50	0.747	0.774	0.873	0.873
	(1.684)	*(1.823)*	*(0.954)*	*(1.004)*
1.75	0.781	0.800	0.852	0.856
	(1.709)	*(1.845)*	*(0.952)*	*(1.003)*
2.00	0.815	0.840	0.852	0.851
	(1.749)	*(1.888)*	*(0.956)*	*(1.005)*
2.25	0.831	0.856	0.843	0.845
	(1.696)	*(1.826)*	*(0.932)*	*(0.979)*
0	0.861	0.852	0.866	0.810
	(1.137)	*(1.219)*	*(0.641)*	*(0.675)*
0	0.854	0.820	0.865	0.790
	(1.157)	*(1.251)*	*(0.645)*	*(0.678)*
0	0.871	0.849	0.869	0.815
	(1.107)	*(1.196)*	*(0.630)*	*(0.663)*
0	0.876	0.847	0.867	0.787
	(1.161)	*(1.241)*	*(0.642)*	*(0.672)*
Coverage of one-sided 90% CIs				
1	0.679	0.755	0.832	0.854
1.25	0.749	0.776	0.870	0.886
1.50	0.806	0.831	0.869	0.890
1.75	0.828	0.852	0.867	0.875
2.00	0.824	0.861	0.867	0.882
2.25	0.828	0.860	0.849	0.869
0	0.828	0.817	0.878	0.798
0	0.827	0.830	0.860	0.790
0	0.848	0.843	0.873	0.800
0	0.819	0.832	0.870	0.785

achieve closest-to-nominal coverage in both the cases. The empirical coverages of

90% residual bootstrap CIs, specially the two-sided ones, continue to perform poorly even when the sample size increases from 100 to 1000. These finite sample results are justifying the incompetency of the residual bootstrap when errors are heteroscedastic,

Table 2.4: Empirical coverage of 90% confidence intervals for regression coefficients over 1000 simulations under $(n, p, p_0) = (1000, 10, 6)$ by Lasso when the design is random.

Coverage and *(avg. width)* of two-sided 90% CIs				
β_j	Case I		Case II	
	PB	*PaB*	*PB*	*PaB*
1	0.910	0.907	0.900	0.890
	(2.911)	*(3.106)*	*(1.668)*	*(1.733)*
1.25	0.895	0.893	0.910	0.903
	(3.131)	*(3.307)*	*(1.746)*	*(1.804)*
1.50	0.903	0.903	0.892	0.889
	(3.251)	*(3.404)*	*(1.762)*	*(1.815)*
1.75	0.909	0.906	0.899	0.898
	(3.289)	*(3.437)*	*(1.758)*	*(1.812)*
2.00	0.891	0.894	0.904	0.903
	(3.333)	*(3.487)*	*(1.762)*	*(1.814)*
2.25	0.903	0.901	0.907	0.901
	(3.227)	*(3.369)*	*(1.716)*	*(1.766)*
0	0.910	0.906	0.900	0.784
	(2.207)	*(2.322)*	*(1.189)*	*(1.233)*
0	0.908	0.902	0.917	0.776
	(2.235)	*(2.359)*	*(1.195)*	*(1.238)*
0	0.897	0.892	0.892	0.755
	(2.172)	*(2.295)*	*(1.181)*	*(1.223)*
0	0.907	0.898	0.890	0.752
	(2.238)	*(2.351)*	*(1.193)*	*(1.233)*
Coverage of one-sided 90% CIs				
1	0.884	0.886	0.904	0.901
1.25	0.898	0.902	0.891	0.893
1.50	0.896	0.899	0.894	0.900
1.75	0.901	0.913	0.897	0.897
2.00	0.895	0.898	0.906	0.911
2.25	0.897	0.900	0.900	0.899
0	0.897	0.886	0.907	0.768
0	0.904	0.891	0.920	0.790
0	0.892	0.884	0.890	0.763
0	0.899	0.889	0.911	0.765

as stated in proposition 2.3.1. When $n = 100$ and errors are heteroscedastic normal, the one- and two-sided perturbation bootstrap intervals achieve sub-nominal coverage for the smaller non-zero regression coefficients given that those coefficients were occasionally estimated to be zero, but achieves close-to-nominal coverage for the larger non-zero regression coefficients. For the zero coefficients, the difference in empirical coverages are higher than the non-zero ones in both case I and II, showing that the thresholding, performed on original Lasso estimator to get proper centering in bootstrap setup, has significant influence in achieving nominal coverage in case of perturbation bootstrap than the same in case of residual bootstrap, when the errors are heteroscedastic.

Table 2.3 and 2.4 display the empirical coverages of 90% CIs in the random design case. In case I, that is when the errors are heteroscedastic centered χ^2_2, the perturbation bootstrap intervals based on $\boldsymbol{T}_n^*$ perform similarly to paired bootstrap intervals for both zero and non-zero regression coefficients. In case II, although the performance of perturbation and paired bootstrap are equivalent for the non-zero regression coefficie-

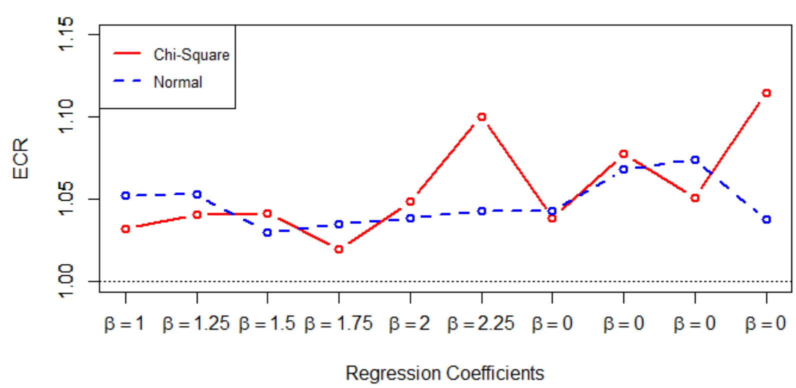

Figure 2.1: Ratio of Empirical Coverage of 90% CIs constructed by Perturbation and Residual Bootstrap in Lasso for Regression Coefficients

nts, the empirical coverages of the paired bootstrap are worse than that of perturbation bootstrap intervals even when $n = 1000$. Perturbation and paired bootstrap are performing almost similarly in random design since both can consistently approximate the distribution of Lasso even when the errors are heteroscedastic.

To compare our perturbation bootstrap method with the existing bootstrap methods, we define the quantity "Empirical Coverage Ratio" (ECR) as

$$ECR = \frac{\text{Empirical Coverage of a Perturbation Bootstrap CI}}{\text{Empirical Coverage of corresponding CI by other bootstrap method}}$$

Figures 2.1 and 2.2 are respectively displaying ECR corresponding to the one-sided 90% CIs constructed using the residual and the paired bootstrap when $n = 1000$. Note that ECR values displayed in the Figure 2.1 are greater than 1 for each of the regression coefficients in both the cases and hence the perturbation bootstrap method is perform-

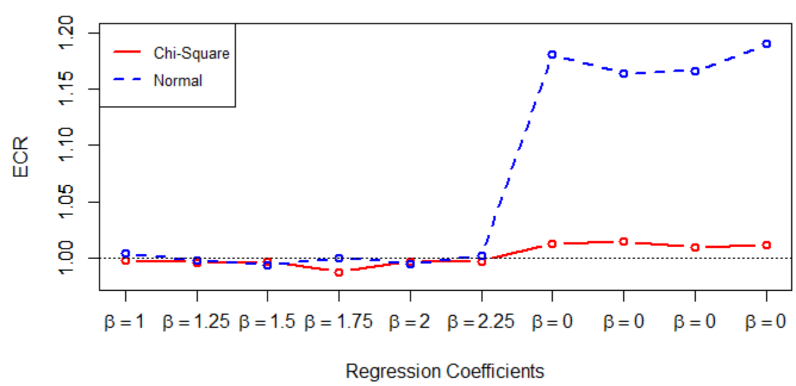

Figure 2.2: Ratio of Empirical Coverage of 90% CIs constructed by Perturbation and Paired Bootstrap in Lasso for Regression Coefficients

ing better than its residual counterpart in terms of coverages of 90% one-sided CIs when $n = 1000$. Conclusions will be similar for other two sample sizes also, for $n = 100$ see Table 2.1. Figure 2.2 is indicating that the perturbation and the paired bootstrap are equivalent in our finite sample setting for non zero regression coefficients in both the cases. But for zero regression coefficients, ECR values are greater than 1 when errors are heteroscedastic mean zero Normal and and greater than 1.15 when errors are heteroscedastic centered χ^2_2. Therefore perturbation bootstrap is performing better than paired bootstrap uniformly for all the zero coefficients when $n = 1000$. Same is true for $n = 100$ also, as can be seen in Table 2.3.

Table 2.5: Comparison of empirical coverages of 90% confidence region for regression parameter vector by Perturbation, Residual and Paired Bootstrap in Lasso.

Empirical Coverage of 90% Coverage Region

n	Case I		Case II	
	PB	RB	PB	RB
100	0.893	0.683	0.916	0.663
500	0.932	0.773	0.942	0.806
1000	0.948	0.754	0.932	0.795
n	Case I		Case II	
	PB	PaB	PB	PaB
100	0.894	0.875	0.924	0.880
500	0.946	0.896	0.928	0.881
1000	0.943	0.895	0.938	0.879

Table 2.5 is displaying the empirical coverages of 90% two-sided confidence regions, constructed using perturbation, residual and paired bootstrap, under $(n, p, p_0) \in \{(100, 10, 6), (500, 10, 6), (1000, 10, 6)\}$ in both the cases. Whereas perturbation bootstrap empirical coverages are little higher than 0.90 for $n = 100$ and 500, residual bootstrap is performing poorly for all the three sample sizes and in both choices of error distribution. On the other hand paired bootstrap is performing nominally.

We see that the perturbation bootstrap is able to produce reliable confidence intervals for regression coefficients and that it is able to do so under cross validation choice of penalty parameter.

2.6 Conclusion

Perturbation Bootstrap method is proposed in case of Lasso. It is shown that the perturbation bootstrap is capable of consistently approximating the distribution of Lasso in both the cases when the covariates are random and non-random and even when the errors are heteroscedastic. Residual bootstrap of Chattaerjee and Lahiri (2011) is shown to fail when the errors are heteroscedastic and covariates are non-random. Thus perturbation bootstrap is better than residual bootstrap in the sense of validity of bootstrap in heteroscedasticity. When the covariates are random then on needs to implement pairs bootstrap, the validity of which is shown by Camponovo (2015). The implementation of pairs bootstrap is significantly different from residual bootstrap. Thus one needs to select the residual and the pair bootstrap depending on the nature of the covariates, where as the perturbation bootstrap works irrespective of the nature of the covariates. Therefore the results in this chapter establish the proposed perturbation bootstrap method as a more general bootstrap method than the resample-based bootstrap methods (residual and pairs) as a tool of distributional approximation of Lasso.

Chapter 3

Perturbation Bootstrap in Adaptive Lasso

3.1 Introduction

Consider the multiple linear regression model

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n, \tag{3.1.1}$$

where $y_1, \ldots, y_n$ are responses, $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed (iid) random variables, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are known non-random design vectors, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is the p-dimensional vector of regression parameters. When the dimension p is large, it is common to approach regression model (3.1.1) with the assumption that the vector $\boldsymbol{\beta}$ is sparse, that is that the set $\mathcal{A} = \{j : \beta_j \neq 0\}$ has cardinality $p_0 = |\mathcal{A}|$ much smaller than p, meaning that only a few of the covariates are "active". The Lasso estimator introduced by Tibshirani (1996) is well suited to the sparse setting because of its property that it sets some regression coefficients exactly equal to 0. One disadvantage of the Lasso, however, is that it produces non-trivial asymptotic bias for the non-zero regression parameters, primarily because it shrinks all estimators toward zero [cf. Knight and Fu (2000)].

Building on the Lasso, Zou (2006) proposed the Adaptive Lasso [hereafter referred to as Alasso] estimator $\hat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ in the regression problem (3.1.1) as

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{t}} \left[\sum_{i=1}^{n} (y_i - \boldsymbol{x}_i' \boldsymbol{t})^2 + \lambda_n \sum_{j=1}^{p} |\tilde{\beta}_{j,n}|^{-\gamma} |t_j| \right], \tag{3.1.2}$$

where $\tilde{\beta}_{j,n}$ is the jth component of a root-n-consistent estimator $\tilde{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$, such as the ordinary least squares (OLS) estimator when $p \leq n$ or the Lasso or ridge regression estimator when $p > n$, $\lambda_n > 0$ is the penalty parameter, and $\gamma > 0$ is a constant governing the influence of the preliminary estimator $\tilde{\boldsymbol{\beta}}_n$ on the Alasso fit. Zou (2006) showed in the fixed-p setting that under some regularity conditions and with the right choice of λ_n, the Alasso estimator enjoys the so-called oracle property [cf. Fan and Li (2001)]; that is, it is variable-selection consistent and it estimates the non-zero regression parameters with the same precision as the OLS estimator which one would compute if the set of active covariates were known.

In an important recent work, Minnier, Tian and Cai (2011) introduced the perturbation bootstrap in the Alasso setup. To state their main results, let $\boldsymbol{\beta}_n^{*N} = (\beta_{1,n}^{*N}, \ldots, \beta_{p,n}^{*N})'$ be the naive perturbation bootstrap Alasso estimator prescribed by Minnier, Tian and Cai (2011) and define $\hat{\mathcal{A}}_n = \{j : \hat{\beta}_{j,n} \neq 0\}$ and $\mathcal{A}_n^{*N} = \{j : \beta_{j,n}^{*N} \neq 0\}$. These authors showed that under some regularity conditions and with p fixed as $n \to \infty$

$$\mathbf{P}_*(\mathcal{A}_n^{*N} = \hat{\mathcal{A}}_n) \to 1 \;\text{ and }\; \sqrt{n}(\boldsymbol{\beta}_n^{*N(1)} - \hat{\boldsymbol{\beta}}_n^{(1)})|\boldsymbol{\varepsilon} \asymp_d \sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}^{(1)}),$$

where $\boldsymbol{\varepsilon}_n = (\epsilon_1, \ldots, \epsilon_n)$, $\boldsymbol{z}^{(1)}$ denotes the sub-vector of $\boldsymbol{z} \in \mathcal{R}^p$ corresponding to the co-ordinates in $\mathcal{A} = \{j : \beta_j \neq 0\}$, "$\asymp_d$" denotes asymptotic equivalence in distribution, and $\mathbf{P}_*$ denotes bootstrap probability conditional on the data. Thus Minnier, Tian and Cai (2011) [hereafter referred to as MTC(11)] showed that, in the fixed-p setting and conditionally on the data, the naive perturbation bootstrap version of the Alasso estimator is variable-selection consistent in the sense that it recovers the support of the Alasso estimator with probability tending to one and that its distribution conditional on the data converges at the same time to that of the Alasso estimator for the non-zero regression parameters. But the accuracy of inference for non-zero regression parameters relies on the rate of convergence of the bootstrap distribution of $\sqrt{n}(\boldsymbol{\beta}_n^{*N(1)} - \hat{\boldsymbol{\beta}}_n^{(1)})|\boldsymbol{\varepsilon}$ to the distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}^{(1)})$ after proper studentization. Furthermore, Chatterjee and Lahiri (2013) showed that the convergence of the Alasso estimators of the nonzero regression coefficients to their oracle Normal distribution is quite slow, owing to the bias induced by the penalty term in (3.1.2). Thus, it would be important for the accuracy of inference if second-order correctness can be achieved in approximating the distribution of the Alasso estimator by the perturbation bootstrap. Second-order correctness implies that the distributional approximation has a uniform error rate of $o_p(n^{-1/2})$. We show in this chapter, however, that the distribution of the naive perturbation bootstrap version of the Alasso estimator, as defined by MTC(11), cannot be second order correct even in fixed dimension. For more details, see Section 3.4.

We introduce a modified perturbation bootstrap for the Alasso estimator for which second order correctness does hold, even when the number of regression parameters $p = p_n$ is allowed to increase with the sample size n. We also show in Proposition 3.2.1 that the modified perturbation bootstrap version of the Alasso estimator (defined in Section 3.2) can be computed by minimizing simple criterion functions. This makes our bootstrap procedure computationally simple and inexpensive.

In this chapter, we consider some pivotal quantities based on Alasso estimators and establish that the modified perturbation bootstrap estimates the distribution of these pivotal quantities up to second order, i.e. with an error that is of much smaller magnitude than what we would obtain by using the Normal approximation under the knowledge of the true active set of covariates. We will refer to the Normal approximation which uses knowledge of the true set of active covariates as the oracle Normal approximation. Our main results show that the modified perturbation bootstrap method enables, for example, the construction of confidence intervals for the nonzero regression coefficients with smaller coverage error than those based on the oracle Normal approximation.

More precisely, we consider pivots which are studentizations of the quantities

$$\sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \quad \text{and} \quad \sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) + \breve{\boldsymbol{b}}_n,$$

where $\boldsymbol{D}_n$ is a $q \times p$ matrix (q fixed) producing q linear combinations of interest of $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}$ and where $\breve{\boldsymbol{b}}_n$ is a bias correction term which we will define in section 3.5. We find that in the $p \leq n$ case, the modified perturbation bootstrap can estimate the distribution of the first pivot with an error of order $o_p(n^{-1/2})$ (see Theorem 3.5.1). This is much smaller than the error of the oracle Normal approximation, which was shown in Theorem 3.1 of Chatterjee and Lahiri (2013) to be of the order $O_p(n^{-1/2} + ||\boldsymbol{b}_n|| + c_n)$, where $\boldsymbol{b}_n$ is the bias targeted by $\breve{\boldsymbol{b}}_n$ and $c_n > 0$ is determined by the initial estimator $\tilde{\boldsymbol{\beta}}_n$ and the tuning parameters λ_n and γ; both $||\boldsymbol{b}_n||$ and c_n are typically greater in magnitude than $n^{-1/2}$ and hence determine the rate of the oracle Normal approximation. We also discover that the bias correction in the second pivot improves the error rate so that the modified perturbation bootstrap estimator achieves the rate

$O_p(n^{-1})$ (see Theorem 3.5.2), which is a significant improvement over the best possible rate of oracle Normal approximation, namely $O(n^{-1/2})$. In the $p > n$ case, we find that the modified perturbation bootstrap estimates the distributions of studentized versions of both the bias-corrected and un-bias-corrected pivots with the rate $o_p(n^{-1/2})$ (see Theorem **??**), establishing the second-order correctness of our modified perturbation bootstrap in the high-dimensional setting.

We show that the naive perturbation bootstrap of MTC(11) is not second-order correct (see Theorem 3.4.1) by investigating the Karush-Kuhn-Tucker (KKT) condition [cf. Boyd and Lieven (2004)] corresponding to their minimization problem. It is shown that second order correctness is not attainable by the naive version of the perturbation bootstrap, primarily due to lack of proper centering of the naive bootstrapped Alasso criterion function. We derive the form of the centering constant by analyzing the corresponding approximation errors using the theory of Edgeworth expansion. To accommodate the centering correction, we modify the perturbation bootstrap criterion function for the Alasso; see Section 3.2 for details. In addition, we also find out that it is beneficial, from both theoretical and computational perspectives, to modify the perturbation bootstrap version of the initial estimators in a similar way. To prove second order correctness of the modified perturbation bootstrap Alasso, the key steps are to find an Edgeworth expansion of the bootstrap pivotal quantities based on the modified criterion function and to compare it with the Edgeworth expansion of the sample pivots. We want to mention that in our setting, the dimension p of the regression parameter vector can grow polynomially in the sample size n at a rate depending on the number of finite moments of the error distribution. Extension to the case in which p grows exponentially with n would be possible under some strong assumptions, e.g. under finiteness of the moment generating function of the regression errors.

We conclude this section with a brief literature review. The perturbation bootstrap was introduced by Jin, Ying, and Wei (2001) as a resampling procedure where the objective function has a U-process structure. Work on the perturbation bootstrap in the linear regression setup is limited. Some work has been carried out by Chatterjee and Bose (2005), MTC(11), Zhou, Song and Thompson (2012), and Das and Lahiri (2016). As a variable selection procedure, Tibshirani (1996) introduced the Lasso. Zou (2006) proposed the Alasso as an improvement over the Lasso. For the Alasso and related popular penalized estimation and variable selection procedures, the residual bootstrap has been investigated by Knight and Fu (2000), Hall, Lee and Park (2009), Chatterjee and Lahiri (2010, 2011, 2013), Wang and Song (2011), MTC(11), Van De Geer et al. (2014), and Camponovo (2015), among others.

The rest of the chapter is organized as follows. The modified perturbation bootstrap for the Alasso is introduced and discussed in Section 3.2. Assumptions and explanations of those are presented in Section 3.3. Main results concerning the estimation properties of the studentized modified perturbation bootstrap pivotal quantities are given in Section 3.5. Negative results on the naive perturbation bootstrap approximation proposed by MTC(11) are discussed in 3.4. Intuitions and explanations behind the modification of the modified perturbation bootstrap and Proofs of the main results are provided in Section 3.6. Section 3.7 presents simulation results exploring the finite-sample performance of the modified perturbation bootstrap in comparison with other methods for constructing confidence intervals based on Alasso estimators and Section 3.8 gives an illustration on real data.

3.2 The modified perturbation bootstrap for Alasso

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 perturbation bootstrap Alasso estimator. We define our bootstrap version of the Alasso estimator as the minimizer of a carefully constructed penalized objective function which involves the Alasso predicted values $\hat{y}_i = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n$, $i = 1, \ldots, n$ as well as the observed values $y_i, \ldots, y_n$. These sets of values appear in the objective function in two perturbed least-squares criteria. Similar modification is also needed in defining the bootstrap versions of the Alasso initial estimators, see (3.2.2). The motivation behind this construction is detailed in Section 3.4. We point out in Section 3.5 why the naive perturbation bootstrap formulation of MTC(11) fails to achieve second order correctness.

We formally define the modified perturbation bootstrap version $\hat{\boldsymbol{\beta}}_n^*$ of the Alasso estimator $\hat{\boldsymbol{\beta}}_n$ as

$$\hat{\boldsymbol{\beta}}_n^* = \arg\min_{\boldsymbol{t}^*} \Bigg[\sum_{i=1}^{n} (y_i - \boldsymbol{x}_i'\boldsymbol{t}^*)^2 (G_i^* - \mu_{G^*}) + \sum_{i=1}^{n} (\hat{y}_i - \boldsymbol{x}_i'\boldsymbol{t}^*)^2 (2\mu_{G^*} - G_i^*) + \mu_{G^*}\lambda_n \sum_{j=1}^{p} |\tilde{\beta}_{j,n}^*|^{-\gamma} |t_j^*| \Bigg], \qquad (3.2.1)$$

where $\tilde{\beta}_{j,n}^*$ is the jth component of $\tilde{\boldsymbol{\beta}}_n^*$, the modified perturbation bootstrap version of the Alasso initial estimator $\tilde{\boldsymbol{\beta}}_n$. We construct $\tilde{\boldsymbol{\beta}}_n^*$ as

$$\tilde{\boldsymbol{\beta}}_n^* = \arg\min_{\boldsymbol{t}^*} \Bigg[\sum_{i=1}^{n} (y_i - \boldsymbol{x}_i'\boldsymbol{t}^*)^2 (G_i^* - \mu_{G^*})$$

$$+\sum_{i=1}^{n}(\hat{y}_i - \boldsymbol{x}_i'\boldsymbol{t}^*)^2(2\mu_{G^*} - G_i^*) + \mu_{G^*}\tilde{\lambda}_n \sum_{j=1}^{p} |t_j^*|^l \Bigg], \qquad (3.2.2)$$

where $\tilde{\lambda}_n = 0$ when $\tilde{\boldsymbol{\beta}}_n$ is taken as the OLS, which we use when $p \leq n$, and $l = 1$ or 2 according as the initial estimator $\tilde{\boldsymbol{\beta}}_n$ is taken as the Lasso or ridge regression estimator when $p > n$. Note that $\tilde{\lambda}_n$ may be different from λ_n.

We point out that the modified perturbation bootstrap estimators can be computed using existing algorithms. Define $\boldsymbol{L}_1(\boldsymbol{t}) = \sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{t})^2(G_i^* - \mu_{G^*}) + \sum_{i=1}^{n}(\hat{y}_i - \boldsymbol{x}_i'\boldsymbol{t})^2(2\mu_{G^*} - G_i^*) + \mu_{G^*}\tilde{\lambda}_n \sum_{j=1}^{p} c_j|t_j|^l$ for some non-negative constants c_j, $j = 1, \cdots, p$, independent of $\boldsymbol{t} = (t_1, \cdots, t_n)'$. Now set $z_i = \hat{y}_i + \hat{\epsilon}_i \mu_{G^*}^{-1}(G_i^* - \mu_{G^*})$, where $\hat{\epsilon}_i = y_i - \hat{y}_i$ for $i = 1, \ldots, n$ and let $\boldsymbol{L}_2(\boldsymbol{t}) = \sum_{i=1}^{n}\left(z_i - \boldsymbol{x}_i'\boldsymbol{t}\right)^2 + \tilde{\lambda}_n \sum_{j=1}^{p} c_j|t_j|^l$. Then we have the following proposition.

Proposition 3.2.1 $\arg\min_{\boldsymbol{t}} \boldsymbol{L}_1(\boldsymbol{t}) = \arg\min_{\boldsymbol{t}} \boldsymbol{L}_2(\boldsymbol{t})$.

This proposition allows us to compute $\tilde{\boldsymbol{\beta}}_n^*$ as well as $\hat{\boldsymbol{\beta}}_n^*$ by minimizing standard objective functions on some pseudo-values. Note that the modified perturbation bootstrap versions of the Alasso estimator as well as of the Alasso initial estimator can be obtained simply by properly perturbing the Alasso residuals in the decomposition $y_i = \hat{y}_i + \hat{\epsilon}_i$, $i = 1, \ldots, n$.

3.3 Assumptions

We first introduce some notations required for stating our assumptions and useful for the proofs later. 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. Set $\mathcal{A}_n = \{j : \beta_{j,n} \neq 0\}$ and $p_0 := p_{0,n} = |\mathcal{A}_n|$. For simplicity, we shall suppress the subscript n in the notations p_n and p_{0n}. Without loss of generality, we shall assume that $\mathcal{A}_n = \{1, \ldots, p_0\}$. Let $\boldsymbol{C}_n = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i'$ and partition it according to $\mathcal{A}_n = \{1, \ldots, p_0\}$ 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 $p_0 \times p_0$. Define $\tilde{\boldsymbol{x}}_i = \boldsymbol{C}_n^{-1}\boldsymbol{x}_i$ (when $p \le n$) and $\mathrm{sgn}(x) = -1, 0, 1$ according as $x < 0$, $x = 0$, $x > 0$, respectively. Suppose $\boldsymbol{D}_n$ is a known $q \times p$ matrix with $\mathrm{tr}(\boldsymbol{D}_n\boldsymbol{D}_n') = O(1)$ and q is not dependent on n. Let $\boldsymbol{D}_n^{(1)}$ contain the first p_0 columns of $\boldsymbol{D}_n$. Define

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

where $\bar{\boldsymbol{x}}_n = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i = (\bar{\boldsymbol{x}}_n^{(1)\prime}, \bar{\boldsymbol{x}}_n^{(1)\prime})'$, $\sigma^2 = \mathbf{Var}(\epsilon_1) = \mathbf{E}(\epsilon_1^2)$, and where μ_3 and μ_4 are, respectively, the third and fourth central moments of ϵ_1. 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 $\check{\boldsymbol{x}}_i^{(1)} = \boldsymbol{C}_{11,n}^{-1/2}\boldsymbol{x}_i^{(1)}$. Let K be a positive constant and r be a positive integer ≥ 3 unless otherwise specified. By $\mathbf{P}_*$ and $\mathbf{E}_*$ we denote, respectively, probability and expectation with respect to the distribution of G^* conditional upon the observed data.

We now introduce our assumptions.

(A.1) Let $\eta_{11,n}$ denote the smallest eigenvalue of the matrix $\boldsymbol{C}_{11,n}$.

- (i) $\eta_{11,n} > Kn^{-a}$ for some $a \in [0, 1)$.
- (ii) $\max\{n^{-1}\sum_{i=1}^n |x_{i,j}|^{2r} : 1 \le j \le p\} + \{n^{-1}\sum_{i=1}^n \left|(\boldsymbol{C}_{11,n}^{-1})_{j\cdot}\boldsymbol{x}_i^{(1)}\right|^{2r} : 1 \le j \le p_0\} = O(1)$.
- (iii) $\max\{n^{-1}\sum_{i=1}^n |\tilde{x}_{i,j}|^{2r} : 1 \le j \le p\} = O(1)$, where $\tilde{x}_{i,j}$ is the jth element of $\tilde{\boldsymbol{x}}_i$. (when $p \le n$)

(iii)′ $\max\{c_{11,n}^{j,j} : 1 \leq j \leq p_0\} = O(1)$, where $c_{11,n}^{j,j}$ is the (j,j)th element of $C_{11,n}^{-1}$. (when $p > n$)

(A.2) There exists a $\delta \in (0,1)$ such that for all $n > \delta^{-1}$,

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

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

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

(A.3) $\max\{|\beta_{j,n}| : j \in \mathcal{A}_n\} = O(1)$ and $\min\{|\beta_{j,n}| : j \in \mathcal{A}_n\} \geq Kn^{-b}$ for some $b \geq 0$ such that $4b < 1$ and $a + 2b \leq 1$, where a is defined as in (A.1)(i).

(A.4) (i) $\mathbf{E}|\epsilon_1|^r < \infty$. $\mathbf{E}\epsilon_1 = 0$.

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

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

(A.5) (i) $\mathbf{E}_*(G_1^*)^r < \infty$. $\mathbf{Var}(G_1^*) = \sigma_{G^*}^2 = \mu_{G^*}^2$, $\mathbf{E}_*(G_1^* - \mu_{G^*})^3 = \mu_{G^*}^3$.

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

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

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

(A.6) There exists $\delta_1 \in (0,1)$ such that for all $n > \delta_1^{-1}$,

(i) $\dfrac{\lambda_n}{\sqrt{n}} \leq \delta_1^{-1}n^{-\delta_1}\min\Big\{\dfrac{n^{-b\gamma}}{p_0}, \dfrac{n^{-b\gamma-a/2}}{\sqrt{p_0}}\Big\}$.

(ii) $\dfrac{\lambda_n}{\sqrt{n}}n^{\gamma/2} \geq \delta_1 n^{\delta_1} p_0$

(iii) $p_0 = o\big(n^{1/2}(\log n)^{-3/2}\big)$.

(A.7) There exists $C \in (0, \infty)$ and $\delta_2 \in (0, \gamma^{-1}\delta_1)$, δ_1 being defined in the assumption (A.6), such that

$$\mathbf{P}\Big(\max\{\big|\sqrt{n}(\tilde{\beta}_{j,n} - \beta_{j,n})\big| : 1 \le j \le p\} > C.n^{\delta_2}\Big) = o(n^{-1/2})$$

$$\mathbf{P}_*\Big(\max\{\big|\sqrt{n}(\tilde{\beta}^*_{j,n} - \hat{\beta}_{j,n})\big| : 1 \le j \le p\} > C.n^{\delta_2}\Big) = o_p(n^{-1/2})$$

Now we explain the assumptions briefly. Assumption (A.1) describes the regularity conditions needed on the growth of the design vectors. Assumption (A.1)(i) is a restriction on the smallest eigenvalue of $\boldsymbol{C}_{11,n}$. Assumption (A.1)(i) is a weaker condition than assuming that $\boldsymbol{C}_{11,n}$ converges to a positive definite matrix. (A.1)(ii) and (iii) are needed to bound the weighted sums of types $\big[\sum_{i=1}^{n} \boldsymbol{x}_i \epsilon_i\big]$, $\big[\sum_{i=1}^{n} \tilde{\boldsymbol{x}}_i \epsilon_i\big]$, $\big[\boldsymbol{C}_{11,n}^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i^{(1)} \epsilon_i\big]$ (second one only when $p \le n$). For $r = 2$ (A.1)(iii) is equivalent to the condition that the diagonal elements of the matrix $\boldsymbol{C}_n^{-1}$ are uniformly bounded. Also for general value of r, (A.1)(ii) and (iii) are much weaker than conditioning on l_r-norms of the design vectors. Here the value of r is specified by the underlying Edgeworth expansion. Assumption (A.1)(iii) requires $p \le n$ and hence is not defined when $p > n$. Note that the condition (A.1)(iii)$'$ needs $p_0 \le n$ which is true in our setup due to assumption (A.6)(iii).

Assumptions (A.2)(i) bounds the eigenvalues of the matrix $\boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{D}_n^{(1)'}$ away from infinity. It is necessary to obtain bounds needed in the studentized setup. Assumption (A.2)(ii) is a condition similar to the conditions in (A.1)(ii) and (iii); but involving the $q \times p$ matrix $\boldsymbol{D}_n$. This condition is needed for showing necessary closeness of the covariance matrix estimators $\breve{\boldsymbol{\Sigma}}_n, \tilde{\boldsymbol{\Sigma}}_n$ [defined in Section 3.5] to their population counterparts (for details see Lemma 3.6.5). Assumption (A.2)(iii) bounds the minimum eigen value of the matrix S_n away from 0. This condition along with the Cramer conditions given in (A.4) and (A.5) enable certain Edgeworth expansions.

Assumption (A.3) separates the relevant covariates from the non-relevant ones. The condition on the minimum is needed to ensure that the non-zero regression coefficients cannot converge to zero faster than the error rate, that is not faster than $O(n^{-1/2})$. We mention that one can assume $b < 1/2$ instead of assuming $b < 1/4$, but with the price of putting another restriction on the penalty parameter λ_n. We do not consider such a setting here.

Assumption (A.4)(i) is a moment condition on the error term needed for valid Edgeworth expansion. Assumption (A.4)(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. Hall (1992)]. Assumption (A.4)(ii) is only needed to get a valid Edgeworth expansion for the original Alasso estimator in the studentized setup. Assumptions (A.5)(i) and (iii) are the analogous conditions that are needed on the perturbing random quantities to get a valid Edgeworth expansion in the bootstrap setting. Assumption (A.5)(ii) is natural, since the ϵ_i are present already in the data generating process, whereas G_i^* are introduced by the user. One can look for Generalized Beta and Generalized Gamma families for suitable choices of the distribution of G^*. The pdf of Generalized Beta family of distributions is

$$GB(y; f, g, h, \omega, \rho) = \begin{cases} \dfrac{|f| y^{f\omega - 1}\Big(1 - (1-c)(y/g)^f\Big)^{\rho - 1}}{g^{f\omega} B(\omega, \rho)\Big(1 + c(y/g)^f\Big)^{\omega + \rho}} & \text{for } 0 < y^f < \dfrac{g^f}{1-h} \\ 0 & \text{otherwise} \end{cases}$$

where $0 \leq h \leq 1$ and other parameters are all positive. We interpret $1/0$ as ∞. The function $B(\omega, \rho)$ is the beta function. Choices of the distribution of G^* can be obtained by finding solution of (f, g, h, ω, ρ) from the following two equations

$$\frac{B(\omega+2/f,\rho)}{B(\omega,\rho)} {}_2F_1\left[\omega+2/f, 2/f; h; \omega+\rho+2/f\right]$$

$$= 2\left[\frac{B(\omega+1/f,\rho)}{B(\omega,\rho)} {}_2F_1\left[\omega+1/f, 1/f; h; \omega+\rho+1/f\right]\right]^2$$

and $$\frac{B(\omega+3/f,\rho)}{B(\omega,\rho)} {}_2F_1\left[\omega+3/f, 3/f; h; \omega+\rho+3/f\right]$$

$$= 5\left[\frac{B(\omega+1/f,\rho)}{B(\omega,\rho)} {}_2F_1\left[\omega+1/f, 1/f; h; \omega+\rho+1/f\right]\right]^3$$

where ${}_2F_1$ denotes hypergeometric series. The pdf of Generalized Gamma family of distributions is given by

$$GG(y;\omega,\rho,\nu) = \begin{cases} \dfrac{(\nu/\omega^{\rho}) y^{\rho-1} e^{(y/\omega)^{\nu}}}{\Gamma(\rho/\nu)} & \text{for } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

where all the parameters are positive and $\Gamma(\cdot)$ denotes the gamma function. For this family, the suitable choices of the distribution of G^* can be obtained by considering any positive value of the parameter ω and solving the following two equations for (ρ, ν),

$$\left[\Gamma((\rho+2)/\nu)\right] * \Gamma(\rho/\nu) = 2\left[\Gamma((\rho+1)/\nu)\right]^2$$

$$\text{and} \quad \left[\Gamma((\rho+3)/\nu)\right] * \left[\Gamma(\rho/\nu)\right]^2 = 5\left[\Gamma((\rho+1)/\nu)\right]^3.$$

One immediate choice of the distribution of G^* from Generalized Beta family is the Beta(α, β) distribution with $3\alpha = \beta = 3/2$. We have utilized this distribution as the distribution of the perturbing quantities G_i^*'s in our simulations, presented in Section

3.7. Outside these two generalized family of distributions, one possible choice is the distribution of $(M_1 + M_2)$ where M_1 and M_2 are independent and M_1 is a Gamma random variable with shape and scale parameters 0.008652 and 2 respectively and M_2 is a Beta random variable with both the parameters 0.036490. Another possible choice is the distribution of $(M_3 + M_4)$ where M_3 and M_4 are independent and M_3 is an Exponential random variable with mean $\left(79 - 15\sqrt{33}\right)/16$ and M_4 is an Inverse Gamma random variable with both shape and scale parameters $\left(4 + \sqrt{11/3}\right)$.

Assumptions (A.6)(i) and (ii) can be compared with the condition (c) $\lambda_n/\sqrt{n} \to 0$ and $n^{\gamma/2}\lambda_n/\sqrt{n} \to \infty$ [cf. Zou (2006), Caner and Fan (2010)]. Whereas (c) is ensuring the oracle normal approximation, (A.6)(i) and (ii) are required for obtaining Edgeworth expansions. Lastly, (A.6)(iii) limits how quickly the number of non-zero regression coefficients may grow. Though it would seem that $p_0 = O(n)$ with $p_0 \leq n$ should be a sufficient restriction on the growth rate of p_0 for approximating the distribution of the Alasso estimator, a careful analysis reveals that further reduction in the growth rate of p_0 is necessary for accommodating the studentization. Clearly it is difficult to comprehend what are the possible choices of $p_0, \lambda_n, \gamma, a, b$ would be to satisfy the assumptions presented in (A.6). Thus it is better to present some possible choices of those parameters.

First consider $a = 0$ and $b = 0$, that is assume that the design matrix corresponding to non-zero regression coefficients, $\boldsymbol{C}_{11,n}$, is non-singular and the magnitudes of non-zero regression coefficients are bounded from both above and below. In that case it is easy to check that one set of possible choices are $p_0 = O(n^{\gamma/5})$ and $\lambda_n = C.n^{1/2-\gamma/4}$ for some constant $C > 0$, provided $\gamma \in (0, 2)$. In particular if $\gamma = 1$ then the choices of p_0 and λ_n maybe respectively $p_0 = O(n^{1/5})$ and $\lambda_n = C.n^{1/4}$ when $a = b = 0$. Again

p_0 can grow with n at the rate $o(n^{1/2}(\log n)^{-3/2}$, when $\gamma > 2$ and $\lambda_n = C.n^{(2-\gamma)/6}$ for some constant $C > 0$ whenever $a = b = 0$.

In general if $a \in [0, 1/2)$ and $b < 1/4$, then it can be shown that the possible choices of γ, p_0 and λ_n are respectively $4a/(1-2b) < \gamma < 2/(1+2b)$, $p_0 = O(n^{[(1-2b)\gamma]/5})$ and $\lambda_n = C.n^{1/2-\gamma/4-b\gamma/2}$ for some constant $C > 0$. On the other hand if $a \in [1/2, 1)$ and $a + 2b < 1$, one set of possible choices would be $\gamma \geq 2$, $p_0 = O(n^{2/3-(a+2b\gamma+4c)/3})$ and $\lambda_n = C.n^{1/6-(a+2b\gamma+c)/3}$ for some constants $c, C > 0$. With $a = 1/2$ and $b = 0$, clearly the choices of p_0 and λ_n reduce to $p_0 = O(n^{1/2-\delta})$ and $\lambda_n = C.n^{-\delta/4}$ for some $\delta, C > 0$.

Assumption (A.7) places deviation bounds on both the sample and bootstrap initial estimators which are needed to get valid Edgeworth expansions. These conditions are satisfied by OLS estimator in $p \leq n$ case [cf. Lemma 3.6.2]. Note that non-bootstrap part of (A.7) is satisfied if there exists a linear approximation of the type $\sum_{i=1}^{n} a_{i,j}\epsilon_i$ of $\sqrt{n}(\tilde{\beta}_{j,n} - \beta_{j,n})$, where $\max\left\{ \sum_{i=1}^{n} |a_{i,j}|^r : 1 \leq j \leq p \right\} = o\left(p^{-1} n^{-1/2+r\delta_2}\right)$ and $\mathbf{E}(|\epsilon_1|^r) < \infty$ for some $r \geq 3$. The bootstrap deviation bound corresponding to (A.7) holds provided similar approximation exits with $(G_1^* - \mu_{G^*})$ in place of ϵ_1. More precisely, for the ridge estimator and for its perturbation bootstrap version defined in Sec 3.2, if for some $r \geq 4$, the conditions

(a) $\mathbf{E}|\epsilon_1|^r + \mathbf{E}_*(G_1^*)^r < \infty$.

(b) $\max\{n^{-1}\sum_{i=1}^{n}(|\boldsymbol{x}_i|^{2r} + |\check{\boldsymbol{x}}_i|^{2r}) : 1 \leq j \leq p\} = O(n^{\delta_2/2})$ for all $i \in \{1, \cdots, n\}$.

(c) $\max\left\{\boldsymbol{e}_j'(\boldsymbol{C}_n + \tilde{\lambda}_n n^{-1}\boldsymbol{I}_p)^{-1}\boldsymbol{\beta}_n : 1 \leq j \leq p\right\} = O(n^{(1+\delta_2)/2}\tilde{\lambda}_n^{-1})$.

(d) $\sup\left\{\boldsymbol{e}_j'(\boldsymbol{C}_n + \tilde{\lambda}_n n^{-1}\boldsymbol{I}_p)^{-1}\boldsymbol{z}_n : ||\boldsymbol{z}_n|| \leq 1\right\} = O(n^{(1+\delta_2)/2}\tilde{\lambda}_n^{-1})$ for all $j \in \{1, \cdots, p\}$.

are satisfied, then the assumption (A.7) holds. Here $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_p\}$ is the standard basis of $\mathcal{R}^p$, $\breve{\boldsymbol{x}}_i = (\boldsymbol{C}_n + \tilde{\lambda}_n n^{-1}\boldsymbol{I}_p)^{-1}\boldsymbol{x}_i$ and $\tilde{\lambda}_n$ is the penalty parameter corresponding to the ridge estimator [cf. Sec 3.2]. This follows analogously to proposition 8.4 of Chatterjee and Lahiri (2013) after applying Lemma 3.6.1, stated in Section 3.6.

3.4 Impossibility of Second-order correctness of the naive perturbation bootstrap

In this section we describe the naive perturbation bootstrap as defined by MTC(11) for the Alasso and show that second-order correctness can not be achievable by their naive perturbation bootstrap method. When the objective function is the usual least squares criterion function the naive perturbation bootstrap Alasso estimator $\boldsymbol{\beta}_n^{*N}$ is defined in MTC(11) as

$$\boldsymbol{\beta}_n^{*N} = \arg\min_{\boldsymbol{v}_n^*} \left[\sum_{i=1}^{n} (y_i - \boldsymbol{x}_i' \boldsymbol{v}_n^*)^2 G_i^* + \lambda_n^* \sum_{j=1}^{p} |\tilde{\beta}_{j,n}^{*N}|^{-\gamma} |v_{j,n}^*| \right], \tag{3.4.1}$$

where

(i) $\lambda_n^* > 0$ is such that $\lambda_n^* n^{-1/2} \to 0$ and $\lambda_n^* \to \infty$ as $n \to \infty$.

(ii) the initial naive bootstrap estimator is defined as

$$\tilde{\boldsymbol{\beta}}_n^{*N} = \arg\min_{\boldsymbol{v}_n^*} \Big[\sum_{i=1}^{n} (y_i - \boldsymbol{x}_i' \boldsymbol{v}_n^*)^2 G_i^* \Big]$$

and $\tilde{\beta}_{j,n}^{*N}$ is the jth component of $\tilde{\boldsymbol{\beta}}_n^{*N}$.

(iii) $\{G_1^*, \ldots, G_n^*\}$ is a set of iid non-negative random quantities with mean and variance both equal to 1.

Note that the initial estimator $\tilde{\boldsymbol{\beta}}_n^{*N}$ is unique only when p is less than or equal to n. We now consider the quantity $\boldsymbol{u}_n^{*N} = \sqrt{n}(\boldsymbol{\beta}_n^{*N} - \hat{\boldsymbol{\beta}}_n)$, which we can show from (3.4.1) to be the minimizer

$$\boldsymbol{u}_n^{*N} = \underset{\boldsymbol{w}_n^*}{\arg\min} \left[\boldsymbol{w}_n^{*'} \boldsymbol{C}_n^* \boldsymbol{w}_n^* - 2\boldsymbol{w}_n^* \boldsymbol{W}_n^* + \lambda_n^* \sum_{j=1}^{p} |\tilde{\beta}_{j,n}^{*N}|^{-\gamma} \Big(|\hat{\beta}_{j,n} + \frac{w_{j,n}^*}{\sqrt{n}}| - |\hat{\beta}_{j,n}| \Big) \right], \tag{3.4.2}$$

where $\hat{\beta}_{j,n}$ is the jth component of the Alasso estimator $\hat{\boldsymbol{\beta}}_n$, $\boldsymbol{C}_n^* = n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' G_i^*$, and $\boldsymbol{W}_n^* = n^{-1/2} \sum_{i=1}^{n} \hat{\epsilon}_i \boldsymbol{x}_i G_i^*$. To describe the solution of MTC(11), assume $\mathcal{A} = \{j : \beta_j \neq 0\} = \{1, \ldots, p_0\}$. MTC(11) claimed that when $\gamma = 1$ and p is fixed, $\big((\boldsymbol{u}_{n1}^{*N})', \mathbf{0}\big)'$ is a solution of (3.4.2) for sufficiently large n, where

$$\boldsymbol{u}_{n1}^{*N} = \boldsymbol{C}_{11,n}^{-1} n^{-1/2} \sum_{i=1}^{n} \epsilon_i \boldsymbol{x}_i^{(1)} (G_i^* - 1) \text{ and } ||\boldsymbol{u}_n^{*N} - \Big((\boldsymbol{u}_{n1}^{*N})', \mathbf{0}\Big)'||_\infty = o_{p_*}(1).$$

However, to achieve second order correctness, we need to obtain a solution $\big((\boldsymbol{u}_{n2}^{*N})', \mathbf{0}\big)'$ of (3.4.2) such that $||\boldsymbol{u}_n^{*N} - \big((\boldsymbol{u}_{n2}^{*N})', \mathbf{0}\big)'||_\infty = o_{p_*}(n^{-1/2})$. We show that such an $\boldsymbol{u}_{n2}^{*N}$ has the form

$$\boldsymbol{u}_{n2}^{*N} = \boldsymbol{C}_{11,n}^{*-1} \Big[\boldsymbol{W}_n^{*(1)} - \frac{\lambda_n^*}{\sqrt{n}} \tilde{\boldsymbol{s}}_n^{*(1)} \Big]$$

for sufficiently large n, where $\boldsymbol{W}_n^{*(1)}$ is the first p_0 components of $\boldsymbol{W}_n^*$ and the jth component of $\tilde{\boldsymbol{s}}_n^{*(1)}$ equals to $\mathrm{sgn}(\hat{\beta}_{j,n})||\tilde{\beta}_{jn}^{*N}|^{-\gamma}$, $j \in \mathcal{A}$ (Here we drop the subscript n from the notations of true parameter values since we are considering p to be fixed in this section). We establish this fact by exploring the KKT condition corresponding to (3.4.2), which is given by

$$2\boldsymbol{C}_n^* \boldsymbol{w}_n^* - 2\boldsymbol{W}_n^* + \frac{\lambda_n^*}{\sqrt{n}} \boldsymbol{\Gamma}_n^* \boldsymbol{l}_n = \mathbf{0}, \tag{3.4.3}$$

for some $\boldsymbol{l}_n = (l_{1n}, \ldots, l_{pn})'$ with $l_{j,n} \in [-1, 1]$ for $j = 1, \ldots, p$ and $\boldsymbol{\Gamma}_n^* = \mathrm{diag}\big(|\tilde{\beta}_{1n}^{*N}|^{-\gamma},$

$\ldots, |\tilde{\beta}_{pn}^{*N}|^{-\gamma}\big)$. Since $\boldsymbol{C}_n^*$ is a non-negative definite matrix, (3.4.2) is a convex optimization problem; hence (3.4.3) is both necessary and sufficient in solving (3.4.2).

Note that $\boldsymbol{W}_n^*$ is not centered and hence we need to adjust the solution $\big((\boldsymbol{u}_{n2}^{*N})', \mathbf{0}\big)'$ for centering before investigating if the naive perturbation bootstrap can asymptotically correct the distribution of Alasso up to second order. Clearly, the centering adjustment term is $\boldsymbol{Ad}_n^* = \big(\boldsymbol{Ad}_n^{*(1)\prime}, \mathbf{0}'\big)'$ where $\boldsymbol{Ad}_n^{*(1)} = \boldsymbol{C}_{11,n}^{*-1} n^{-1/2} \sum_{i=1}^n \hat{\epsilon}_i \boldsymbol{x}_i^{(1)}$. It follows from the steps of the proofs of the results of Section 3.5 that we need $||\boldsymbol{Ad}_n^*|| = o_{p_*}(n^{-1/2})$ to achieve second-order correctness. We show that this is indeed not the case even in the fixed p setting.

More precisely, we negate the second-order correctness of the naive perturbation bootstrap of MTC(11) by first showing that $\big((\boldsymbol{u}_{n2}^{*N})', \mathbf{0}'\big)'$ satisfies the KKT condition (3.4.3) exactly with bootstrap probability converging to 1. Then we show that $\sqrt{n}||\boldsymbol{Ad}_n^*||$ diverges in bootstrap probability to ∞, which in turn implies that the conditional cdf of $\boldsymbol{F}_n^{*N} = \sqrt{n}\big(\boldsymbol{\beta}_n^{*N} - \hat{\boldsymbol{\beta}}_n\big)$ can not approximate the cdf of $\boldsymbol{F}_n = \sqrt{n}\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\big)$ with the uniform accuracy $O_p(n^{-1/2})$, needed for the validity of second-order correctness. We formalize these arguments in the following theorem.

Theorem 3.4.1 *Let p be fixed and $\boldsymbol{C}_n \to \boldsymbol{C}$, a positive definite matrix. Define $Z_n^{*-1} = \sqrt{n}||\boldsymbol{Ad}_n^*||$. Suppose, $(\log n/n)^{1/2}.\max\{\lambda_n, \lambda_n^*\} \to 0$ and $(\log n)^{-(\gamma+1)/2}.\min\{\lambda_n, \lambda_n^*\}$ $.\min\{1, n^{(\gamma-1)/2}\} \to \infty$ as $n \to \infty$. Also assume that (A.1)(i), (ii) and (A.4)(i) hold with $r = 4$. Then there exists a sequence of borel sets $\{\boldsymbol{A}_n\}_{n\geq 1}$ with $\mathbf{P}(\boldsymbol{\varepsilon}_n \in \boldsymbol{A}_n) \to 1$ and given $\boldsymbol{\varepsilon}_n = (\epsilon_1, \ldots, \epsilon_n)' \in \boldsymbol{A}_n$, the following conclusions hold.*

(a) $\mathbf{P}_*\left(\boldsymbol{u}_n^{*N} = \big((\boldsymbol{u}_{n2}^{*N})', \mathbf{0}'\big)'\right) = 1 - o(n^{-1/2})$.

(b) $\mathbf{P}_*\left(Z_n^* > \epsilon\right) = o(n^{-1/2})$ *for any* $\epsilon > 0$.

(c) $\sup_{\boldsymbol{x}\in\mathcal{R}^p}\Big|\mathbf{P}_*\big(\boldsymbol{F}_n^{*N}\le\boldsymbol{x}\big)-\mathbf{P}\big(\boldsymbol{F}_n\le\boldsymbol{x}\big)\Big|\ge K.\dfrac{\lambda_n}{\sqrt{n}}$ *for some* $K>0$.

Remark 3.4.1 *Theorem 3.4.1 (a), (b) state that the naive perturbation bootstrap is incompetent in approximating the distribution of Alasso up to second order. The fundamental reason behind second order incorrectness is the inadequate centering in the form of* $\sqrt{n}(\boldsymbol{\beta}_n^{*N}-\hat{\boldsymbol{\beta}}_n)$*. Although the adjustment term necessary for centering is* $o_{p_*}(1)$*, which essentially helps to establish distributional consistency in MTC(11), the term is coarser than* $n^{-1/2}$*, leading to second order incorrectness. Additionally, it is worth mentioning that studentization will also not help in achieving second order correctness by naive perturbation bootstrap of MTC(11), since the necessary centering correction can not be accomplished by any sort of studentization. Part (c) conveys uniformly how far the naive bootstrap cdf is from the original cdf.*

3.5 Modified Perturbation Bootstrap and its Higher Order Properties

This section is divided into two sub-sections. The first one describes briefly the motivation behind considering the perturbation bootstrap modification in Alasso. The second sub-section describes higher order asymptotic properties of our modified perturbation bootstrap method.

3.5.1 Motivation for the modified perturbation bootstrap

Theorem 3.4.1 establishes that the naive perturbation bootstrap of MTC(11) does not provide a solution for approximating the distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_n)$ up to second order. As it is mentioned earlier, the problem occurs because $\boldsymbol{W}_n^*$ is not centered. Let $\breve{\boldsymbol{W}}_n^*$ denotes the centered version of $\boldsymbol{W}_n^*$, that is $\breve{\boldsymbol{W}}_n^*=n^{-1/2}\sum_{i=1}^n\hat{\epsilon}_i\boldsymbol{x}_i(G_i^*-\mu_{G^*})$, and consider the vector equation

$$2\boldsymbol{C}_n^* \boldsymbol{w}_n^* - 2\breve{\boldsymbol{W}}_n^* + \frac{\lambda_n^*}{\sqrt{n}} \boldsymbol{\Gamma}_n^* \boldsymbol{l}_n = \boldsymbol{0}, \tag{3.5.1}$$

which is same as (3.4.3) after replacing $\boldsymbol{W}_n^*$ with $\breve{\boldsymbol{W}}_n^*$. Note that the solution to (3.5.1) is of the form $((\boldsymbol{u}_{n3}^{*(1)})', \boldsymbol{0}')'$, where $\boldsymbol{u}_{n3}^{*(1)} = \boldsymbol{C}_{11,n}^{*-1}\Big[\breve{\boldsymbol{W}}_n^{*(1)} - \frac{\lambda_n^*}{\sqrt{n}} \tilde{\mathbf{s}}_n^{*(1)}\Big]$. Although this form is adequate for achieving second-order correctness in fixed dimension, there are some computational and higher-dimensional issues that we now address.

Note that $\boldsymbol{C}_{11,n}^*$ is a matrix involving random quantities $\{G_1^*, \ldots, G_n^*\}$. Thus $\boldsymbol{C}_{11,n}^*$ will not remain same for each bootstrap iteration and hence each bootstrap iteration will require computing the inverse of $\boldsymbol{C}_{11,n}^*$ afresh. This is computationally expensive and the expense increases as the number of non-zero regression parameters increases. Therefore it will be computationally advantageous if we can replace $\boldsymbol{C}_{11,n}^*$ by $\boldsymbol{C}_{11,n}$ in the form of $\boldsymbol{u}_{n3}^{*(1)}$.

Now define, $\boldsymbol{u}_{n4}^{*(1)} = \boldsymbol{C}_{11,n}^{-1}\Big[\breve{\boldsymbol{W}}_n^{*(1)} - \frac{\lambda_n^*}{\sqrt{n}} \tilde{\mathbf{s}}_n^{*(1)}\Big]$. If we look closely at the bias term $-\frac{\lambda_n^*}{\sqrt{n}} \boldsymbol{C}_{11,n}^{-1} \tilde{\mathbf{s}}_n^{*(1)}$, then it is clear that the primary contribution of the bias towards $\boldsymbol{u}_{n4}^{*(1)}$ is $-\frac{\lambda_n^*}{\sqrt{n}} \boldsymbol{C}_{11,n}^{-1} \tilde{\mathbf{s}}_n^{(1)}$, where jth component of $\tilde{\mathbf{s}}_n^{(1)}$ is equal to $\mathrm{sgn}(\hat{\beta}_{j,n})||\tilde{\beta}_{jn}|^{-\gamma}$, $j \in \mathcal{A}$, where $\tilde{\beta}_{j,n}$ is the jth component of the OLS estimator $\tilde{\boldsymbol{\beta}}_n$. By Taylor's expansion, $\big(\tilde{\mathbf{s}}_n^{*(1)} - \tilde{\mathbf{s}}_n^{(1)}\big)$ depends on the OLS residuals. The OLS residuals again depend on all p estimated regression parameters, unlike Alasso residuals which depend only on the estimates of the p_0 non-zero components. Since it is needed to bound $||\tilde{\mathbf{s}}_n^{*(1)} - \tilde{\mathbf{s}}_n^{(1)}||_\infty$ for achieving valid edgeworth expansion, we will come up with an implicit bound on the dimension p, which we do not want to impose. On the other hand, if the difference depends on Alasso residuals instead of OLS ones, then the implicit condition will be on p_0 and this is reasonable as p_0 can be much smaller than p. Additionally, $\tilde{\boldsymbol{\beta}}_n^{*N}$ involves inversion of the random matrix $\boldsymbol{C}_n^*$ and hence it is computationally expensive. Thus if $\boldsymbol{C}_n^*$ can be replaced by some fixed matrix, say $\boldsymbol{C}_n$, then the bootstrap will be computationally advantageous.

However, if we implement the modification described in Section 3.2, then both the theoretical and computational shortcomings of the perturbation bootstrap method become resolved and the second-order correctness is achieved even in increasing dimension under some mild regularity conditions. Additionally, we also have the nice structure due to the modification, which enables us to employ existing computational algorithms, as pointed out in Proposition 3.2.1.

3.5.2 Higher Order Results

Define, $\boldsymbol{T}_n = \sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)$. Without loss of generality we have assumed that $\mathcal{A}_n = \{j : \beta_{j,n} \neq 0\} = \{1, \dots, p_0\}$. Hence, by Taylor's expansion it is immediate from the form of Alasso estimator that $\boldsymbol{\Sigma}_n = n^{-1}\sum_{i=1}^{n}\big(\boldsymbol{\xi}_i^{(0)} + \boldsymbol{\eta}_i^{(0)}\big)\big(\boldsymbol{\xi}_i^{(0)} + \boldsymbol{\eta}_i^{(0)}\big)'$ or $\bar{\boldsymbol{\Sigma}}_n = n^{-1}\sum_{i=1}^{n}\boldsymbol{\xi}_i^{(0)}\boldsymbol{\xi}_i'^{(0)}$ can be considered as the asymptotic variance of $\boldsymbol{T}_n/\sigma$ at sample size n. Here $\boldsymbol{\xi_i}^{(0)} = \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}$, $\boldsymbol{\eta}_i^{(0)} = \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{\eta}_i$. For each $i \in \{1, \dots, n\}$, $\boldsymbol{\eta}_i$ is a $p_0 \times 1$ vector with jth element $\Big(\dfrac{\lambda_n}{2n}\tilde{x}_{i,j}\dfrac{\gamma}{|\beta_{j,n}|^{\gamma+1}}sgn(\beta_{j,n})\Big)$ where $\tilde{\boldsymbol{x}}_i = \boldsymbol{C}_n^{-1}\boldsymbol{x}_i$ (when $p \leq n$) and $\mathrm{sgn}(x) = -1, 0, 1$ according as $x < 0$, $x = 0$, $x > 0$, respectively, as defined earlier. The bias corresponding to $\boldsymbol{T}_n$ is $-\boldsymbol{b}_n = -\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\,\boldsymbol{s}_n^{(1)}\dfrac{\lambda_n}{2\sqrt{n}}$, where $\boldsymbol{D}_n^{(1)}$ and $\boldsymbol{C}_{11,n}$ are as defined earlier and $\boldsymbol{s}_n^{(1)}$ is a $p_0 \times 1$ vector with jth element $\mathrm{sgn}(\beta_{j,n})|\beta_{j,n}|^{-\gamma}$. Although $\bar{\boldsymbol{\Sigma}}_n$ is defined for all p, $\boldsymbol{\Sigma}_n$ is only defined when $p \leq n$. Also $\bar{\boldsymbol{\Sigma}}_n$ is the asymptotic variance of $[\boldsymbol{T}_n + \boldsymbol{b}_n]/\sigma$.

Define the set $\hat{\mathcal{A}}_n = \{j : \hat{\beta}_{j,n} \neq 0\}$ and $\hat{p}_{0,n} = |\hat{\mathcal{A}}_n|$, supposing, without loss of generality, 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 $\boldsymbol{x}_i$. Hence, the bias-correction term $\check{\boldsymbol{b}}_n$ corresponding to $\boldsymbol{T}_n$ can be defined as

$$\check{\boldsymbol{b}}_n = \hat{\boldsymbol{D}}_n^{(1)}\hat{\boldsymbol{C}}_{11,n}^{-1}\hat{\mathbf{s}}_n^{(1)}\frac{\lambda_n}{2\sqrt{n}},$$

where $\hat{\mathbf{s}}_n^{(1)}$ is the $\hat{p}_{0,n} \times 1$ vector with jth entry equal to $\operatorname{sgn}(\hat{\beta}_{j,n})|\tilde{\beta}_{j,n}|^{-\gamma}$, $j \in \hat{\mathcal{A}}_n$.

Therefore, the studentized pivots can be constructed as

$$R_n = \begin{cases} \hat{\sigma}_n^{-1}\hat{\boldsymbol{\Sigma}}_n^{-1/2}\boldsymbol{T}_n & \text{for } p \le n \\ \hat{\sigma}_n^{-1}\check{\boldsymbol{\Sigma}}_n^{-1/2}\boldsymbol{T}_n & \text{for } p > n \end{cases} \quad \text{and} \quad \check{R}_n = \check{\sigma}_n^{-1}\check{\boldsymbol{\Sigma}}_n^{-1/2}[\boldsymbol{T}_n + \check{\boldsymbol{b}}_n],$$

where the matrices $\hat{\boldsymbol{\Sigma}}_n$ and $\check{\boldsymbol{\Sigma}}_n$ have the form

$$\hat{\boldsymbol{\Sigma}}_n = n^{-1}\sum_{i=1}^n \left(\hat{\boldsymbol{\xi}}_i^{(0)} + \hat{\boldsymbol{\eta}}_i^{(0)}\right)\left(\hat{\boldsymbol{\xi}}_i^{(0)} + \hat{\boldsymbol{\eta}}_i^{(0)}\right)' \quad \text{and} \quad \check{\boldsymbol{\Sigma}}_n = n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\xi}}_i^{(0)}\hat{\boldsymbol{\xi}}_i^{(0)'}, \tag{3.5.2}$$

and

$$\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^2 \quad \text{and} \quad \check{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \tilde{\epsilon}_i^2,$$

where $\hat{\epsilon}_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n$, $\tilde{\epsilon}_i = y_i - \sum_{j\in\hat{\mathcal{A}}_n} x_{ij}\tilde{\beta}_{j,n}$, $\hat{\boldsymbol{\xi}}_i^{(0)} = \hat{\boldsymbol{D}}_n^{(1)}\hat{\boldsymbol{C}}_{11,n}^{-1}\hat{\boldsymbol{x}}_i^{(1)}$ and $\hat{\boldsymbol{\eta}}_i^{(0)} = \hat{\boldsymbol{D}}_n^{(1)}\hat{\boldsymbol{C}}_{11,n}^{-1}$ $\hat{\boldsymbol{\eta}}_i$, with

$$\hat{\boldsymbol{\eta}}_i = \left(\frac{\lambda_n}{2n}\tilde{x}_{i,j}\frac{\gamma}{|\hat{\beta}_{j,n}|^{\gamma+1}}\operatorname{sgn}(\hat{\beta}_{j,n})\right)_{j\in\hat{\mathcal{A}}_n}.$$

We construct perturbation bootstrap versions $\boldsymbol{R}_n^*$ and $\check{\boldsymbol{R}}_n^*$ of $\boldsymbol{R}_n$ and $\check{\boldsymbol{R}}_n$ first by replacing $\boldsymbol{T}_n$ with $\boldsymbol{T}_n^* = \sqrt{n}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)$. We then replace $\hat{\boldsymbol{\Sigma}}_n$ and $\check{\boldsymbol{\Sigma}}_n$ with $\check{\boldsymbol{\Sigma}}_n$ and $\tilde{\boldsymbol{\Sigma}}_n$, respectively, which we define by replacing $\hat{\boldsymbol{\xi}}_i^{(0)}$ with $\check{\boldsymbol{\xi}}_i^{(0)} = \hat{\boldsymbol{\xi}}_i^{(0)}\hat{\epsilon}_i$ and $\hat{\boldsymbol{\eta}}_i^{(0)}$ with

$\breve{\boldsymbol{\eta}}_i^{(0)} = \hat{\boldsymbol{\eta}}_i^{(0)} \hat{\epsilon}_i$ in (3.5.2). We replace $\breve{\boldsymbol{b}}_n$ with $\breve{\boldsymbol{b}}_n^* = \hat{\boldsymbol{D}}_n^{*(1)} \hat{\boldsymbol{C}}_{11,n}^{*-1} \hat{\mathbf{s}}_n^{*(1)} \lambda_n / (2\sqrt{n})$, where $\hat{\mathbf{s}}_n^{*(1)}$ is the $|\hat{\mathcal{A}}_n^*| \times 1$ vector with jth entry equal to $\text{sgn}(\hat{\beta}_{j,n}^*)|\tilde{\beta}_{j,n}^*|^{-\gamma}$, $j \in \hat{\mathcal{A}}_n^* = \{j : \hat{\boldsymbol{\beta}}_{j,n}^* \neq 0\}$. The matrix $\hat{\boldsymbol{C}}_{11,n}^*$ is the $|\hat{\mathcal{A}}_n^*| \times |\hat{\mathcal{A}}_n^*|$ sub-matrix of $\boldsymbol{C}_n$ with rows and columns in $\hat{\mathcal{A}}_n^*$ and $\hat{\boldsymbol{D}}_n^{*(1)}$ is the $q \times |\hat{\mathcal{A}}_n^*|$ sub-matrix of $\boldsymbol{D}_n$ with columns in $\hat{\mathcal{A}}_n^*$. Lastly, we need

$$\hat{\sigma}_n^{*2} = n^{-1} \mu_{G^*}^{-2} \sum_{i=1}^{n} \hat{\epsilon}_i^{*2} \left(G_i^* - \mu_{G^*} \right)^2 \quad \text{and} \quad \breve{\sigma}_n^{*2} = n^{-1} \mu_{G^*}^{-2} \sum_{i=1}^{n} \tilde{\epsilon}_i^{*2} \left(G_i^* - \mu_{G^*} \right)^2,$$

where $\hat{\epsilon}_i^* = y_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}_n^*$, $\tilde{\epsilon}_i^* = y_i - \sum_{j \in \hat{\mathcal{A}}_n^*} x_{ij} \tilde{\beta}_{j,n}^*$. With these we construct R_n^* and $\breve{R}_n^*$ as

$$R_n^* = \begin{cases} \hat{\sigma}_n^{*-1} \hat{\sigma}_n \breve{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{T}_n^* & \text{for } p \leq n \\ \hat{\sigma}_n^{*-1} \hat{\sigma}_n \tilde{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{T}_n^* & \text{for } p > n \end{cases} \quad \text{and} \quad \breve{R}_n^* = \breve{\sigma}_n^{*-1} \breve{\sigma}_n \tilde{\boldsymbol{\Sigma}}_n^{-1/2} [\boldsymbol{T}_n^* + \breve{\boldsymbol{b}}_n^*].$$

We are motivated to look at these studentized or pivot quantities by the fact that studentization improves the rate of convergence of bootstrap estimators in many settings [cf. Hall (1992)].

3.5.2.1 Results for $p \leq n$.

Theorem 3.5.1 *Let (A.1)–(A.6) hold with $r = 6$. Then*

$$\sup_{B \in \mathcal{C}_q} \left| \mathbf{P}_* \left(\mathbf{R}_n^* \in B \right) - \mathbf{P} \left(\mathbf{R}_n \in B \right) \right| = o_p(n^{-1/2})$$

Theorem 3.5.1 shows that after proper studentization, the modified perturbation bootstrap approximation of the distribution of the Alasso estimator is second-order correct. The error rate reduces to $o_p(n^{-1/2})$ from $O(n^{-1/2})$, the best possible rate obtained by the oracle Normal approximation. This is a significant improvement from the perspective of inference. As a consequence, the precision of the percentile confidence intervals based on $\mathbf{R}_n^*$ will be greater than that of confidence intervals based on the oracle Normal approximation.

We point out that the error rate in Theorem 3.5.1 cannot be reduced to the optimal rate of $O_p(n^{-1})$, unlike in the fixed-dimension case. To achieve this optimal rate by our modified bootstrap method, we now consider a bias corrected pivot $\check{\mathbf{R}}_n$ and its modified perturbation bootstrap version $\check{\mathbf{R}}_n^*$. The following theorem states that it achieves the optimal rate.

Theorem 3.5.2 *Let (A.1)–(A.6) hold with $r = 8$. Then*

$$\sup_{B \in \mathcal{C}_q} \left| \mathbf{P}_* \big(\check{\mathbf{R}}_n^* \in B \big) - \mathbf{P} \big(\check{\mathbf{R}}_n \in B \big) \right| = O_p(n^{-1})$$

Theorem 3.5.2 suggests that the modified perturbation bootstrap achieves notable improvement in the error rate over the oracle Normal approximation irrespective of the order of the bias term. Thus Theorem 3.5.2 establishes the perturbation bootstrap method as an effective method for approximating the distribution of the Alasso estimator when $p \leq n$.

3.5.2.2 Results for $p > n$

We now present results for the quality of perturbation bootstrap approximation when the dimension p of the regression parameter can be much larger than the sample size n. We consider the initial estimator $\tilde{\boldsymbol{\beta}}_n$ to be some $\sqrt{n}$-consistent bridge estimator, for example Lasso or Ridge estimator, in defining the Alasso estimator by (3.1.2). The bootstrap version of Lasso or Ridge is defined by (3.2.2). Higher order results are presented separately for two cases based on growth of p with sample size n. First we consider the case when p can grow polynomially and then we move to the situation when p can grow exponentially.

3.5.2.2.1 p grows polynomially.

Theorem 3.5.3 *Let (A.1)(i), (ii), (iii)′ and (A.2)–(A.6) and (A.7) hold and $p = O(n^{(r-3)/2})$ for some positive integer $r \geq 3$. Now if $b = 0$ [cf. condition (A.3) in Section 3.3] and $r \geq 8$, then we have*

$$\sup_{B \in \mathcal{C}_q} \left|\mathbf{P}_*\big(\mathbf{R}_n^* \in B\big) - \mathbf{P}\big(\mathbf{R}_n \in B\big)\right| = o_p(n^{-1/2})$$

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

Theorem 3.5.3 states that our proposed modified perturbation bootstrap approximation is second-order correct, even when p grows polynomially with n. The error rate obtained by our proposed method is significantly better than $O(n^{-1/2})$, which is the best-attainable rate of the oracle Normal approximation. When p can grow at a polynomial rate with n, the validity of our method depends on the existence of some polynomial moment of the error distribution. To see why, note that it is essential to have

$$\mathbf{P}\Big(\max_{1\leq j\leq p} |\check{W}_{j,n}| > K.\sqrt{\log n}\Big) = o(n^{-1/2}) \qquad \text{and}$$

$$\mathbf{P}_*\Big(\max_{1\leq j\leq p} |\check{W}_{j,n}^*| > K.\sqrt{\log n}\Big) = o_p(n^{-1/2}) \tag{3.5.3}$$

to obtain second order correctness, as presented in Theorem 3.5.3. Here K is a constant > 1, $\check{W}_{j,n} = n^{-1/2}\sum_{i=1}^{n} \epsilon_i x_{i,j}$ and $\check{W}_{j,n}^* = n^{-1/2}\sum_{i=1}^{n} \hat{\epsilon}_i x_{i,j}(G_i^* - \mu_{G^*})$. In view of Lemma 3.6.1, the following bound is needed to conclude (3.5.3)

$$p.\Big(\max_{1\leq j\leq p} \big[\sum_{i=1}^{n} |x_{i,j}|^{2r}\big]\Big)\Big(\mathbf{E}|\epsilon_1|^r\Big)^2 = o\Big(n^{(r-1)/2}(\log n)^{r/2}\Big)$$

Clearly under the assumption $\max\{n^{-1}\sum_{i=1}^{n}|x_{i,j}|^{2r} : 1 \leq j \leq p\} = O(1)$ [cf. condition (A.1) (ii)], we must have $p = o\big(n^{(r-3)/2}(\log n)^{r/2}\big)$ provided $\mathbf{E}|\epsilon_1|^r < \infty$. Therefore in view of condition (A.1) (ii), p can grow like $\Big(a_n.n^l.(\log n)^{l+3/2}\Big)$ where $a_n \to 0$ as $n \to \infty$, provided $\mathbf{E}|\epsilon_1|^{2l+3} < \infty$. This implies that p can grow polynomially with n under the assumption that some polynomial moment of the error distribution exists.

3.5.2.2.2 ***p* grows exponentially.** When p grows exponentially with some fractional power of n, existence of polynomial moment of some order of regression errors ϵ_i's [cf. condition (A.4) (i)] is not enough to achieve higher order accuracy. Indeed, we need to have some control over the moment generating function of the error variable. Following two important cases are considered in this setting.

Errors are Sub-Gaussian: Suppose error ϵ_1 is sub-gaussian. This means that there exists $d > 0$ such that

$$\mathbf{E}[e^{\kappa\epsilon_1}] \leq e^{\kappa^2 d^2/2} \text{ for all } \kappa \in \mathcal{R}. \tag{3.5.4}$$

When the regression errors have sub-gaussian tails, we need to choose the perturbing quantities G_i^*'s effectively to have sub-gaussian tails, that is there exists $d^* > 0$ such that

$$\mathbf{E}_*[e^{\kappa(G_1^*-\mu_{G^*})}] \leq e^{\kappa^2 d^{*2}/2} \text{ for all } \kappa \in \mathcal{R}. \tag{3.5.5}$$

Theorem 3.5.4 *Let (A.1)(i), (ii), (iii)′ and (A.2)–(A.6) and (A.7) hold with $r = 8$ and $b = 0$. Also assume that (3.5.4) & (3.5.5) hold and $p = O\Big(\exp\big(n^{(\delta_1-\gamma\delta_2)}\big)\Big)$ where δ_1 and δ_2 are defined in assumptions (A.6) and (A.7) in Section 3.3. Then the conclusions of Theorem 3.5.3 hold.*

Errors are Sub-Exponential: Consider the regression errors to be sub-exponential, that is there exist positive parameters d, h such that

$$\mathbf{E}[e^{\kappa\epsilon_1}] \le e^{\kappa^2 d^2/2} \;\text{ for all } |\kappa| < 1/h. \tag{3.5.6}$$

Similar to sub-gaussian case, we need to choose the perturbing quantities G_i^*'s to be sub-exponential besides the errors being sub-exponential, that is there exist positive parameters d^*, h^* such that

$$\mathbf{E}_*[e^{\kappa(G_1^* - \mu_{G^*})}] \le e^{\kappa^2 d^{*2}/2} \;\text{ for all } |\kappa| < 1/h^*. \tag{3.5.7}$$

Theorem 3.5.5 *Let (A.1)(i), (ii), (iii)′ and (A.2)–(A.5), (A.6)(i), (ii) and (A.7) hold with $r = 8$ and $b = 0$. Also assume that (3.5.6) and (3.5.7) hold.*

(a) *If $p = O\Big(\exp\big(n^{(\delta_1 - \gamma\delta_2)}\big)\Big)$ and $p_0 = O\Big(n^{(1-\delta_1+\gamma\delta_2)/2}\Big)$ are satisfied where δ_1 and δ_2 are defined in assumptions (A.6) and (A.7) in Section 3.3, then the conclusions of Theorem 3.5.3 hold.*

(b) *If $p = O\Big(\exp\big(n\big)\Big)$, $n^{(-\delta_1+\gamma\delta_2)} = o\big(p_0^2/n\big)$ and $p_0/\sqrt{n} = o\big((\log n)^{-3/2}\big)$ are satisfied where δ_1 and δ_2 are defined in assumptions (A.6) and (A.7) in Section 3.3, then the conclusions of Theorem 3.5.3 hold.*

Theorem 3.5.4 and 3.5.5 show that our perturbation bootstrap method remains valid as a second order correct method even when the dimension p grows exponentially with some fractional power of n. Moreover, we can achieve exponential growth of p in some situations when errors are sub-exponential, as stated in part (b) of Theorem 3.5.5. To obtain higher order results stated in Theorem 3.5.4 and Theorem 3.5.5, we need to relax (3.5.3) a bit for $j = p_0 + 1, \ldots, p$. It follows from the proofs and condition (A.6)(ii) that we can relax (3.5.3) for $j = p_0 + 1, \ldots, p$, to the following

$$\mathbf{P}\Big(\max_{p_0+1 \le j \le p} |\breve{W}_{j,n}| > K.n^{(\delta_1 - \gamma\delta_2)}.p_0 \Big) = o(n^{-1/2}) \qquad \text{and}$$

$$\mathbf{P}_*\Big(\max_{p_0+1 \le j \le p} |\breve{W}^*_{j,n}| > K.n^{(\delta_1 - \gamma\delta_2)}.p_0 \Big) = o_p(n^{-1/2}), \tag{3.5.8}$$

keeping higher order results valid. Now consider using Hoeffding's inequality in sub-gaussian case and Bernstein's inequality in sub-exponential case. As a result, the following two bounds are needed respectively in sub-gaussian and sub-exponential case to conclude (3.5.8)

$$p.\exp\left(- \frac{C_1.n^{1+2(\delta_1-\gamma\delta_2)}.p_0^2}{2.\max_{1\le j\le p}\Big[\sum_{i=1}^n \big(|x_{i,j}|^2 + |x_{i,j}|^4\big)\Big]} \right) = o\Big(n^{-1/2}\Big) \qquad \text{and}$$

$$p.\exp\left(- \frac{C_2.n^{1+2(\delta_1-\gamma\delta_2)}.p_0^2}{2\Big(\max_{1\le j\le p}\Big[\sum_{i=1}^n \big(|x_{i,j}|^2 + |x_{i,j}|^4\big)\Big] + C_3.n^{1/2+(\delta_1-\gamma\delta_2)}.p_0\Big)} \right) = o\Big(n^{-1/2}\Big).$$

C_1, C_2, C_3 are some positive constants. In view of the assumption $\max\{n^{-1}\sum_{i=1}^n |x_{i,j}|^{2r} : 1 \le j \le p\} = O(1)$ [cf. condition (A.1) (ii)], the first bound is implied by $p = o\Big(\exp\big(C.n^{2(\delta_1-\gamma\delta_2)}.p_0^2\big)\,.n^{-1/2}\Big)$, whereas $p = o\Big(\exp\Big(\dfrac{C.n^{2(\delta_1-\gamma\delta_2)}.p_0^2}{1+p_0.n^{-1/2+(\delta_1-\gamma\delta_2)}}\Big).n^{-1/2}\Big)$ is required to obtain the second bound. Here C is some positive constant. These requirements on the growth of p are implying the growth conditions stated in Theorem 3.5.4 and Theorem 3.5.5.

Remark 3.5.1 *Note that the matrices $\breve{\mathbf{\Sigma}}_n$ and $\tilde{\mathbf{\Sigma}}_n$ used in defining the bootstrap pivots do not depend on $G_1^*, \ldots, G_n^*$. Hence it is not required to compute the negative square roots of these matrices for each Monte Carlo bootstrap iteration; these must only be computed once. This is a notable feature of our modified perturbation bootstrap method from the perspective of* *computational complexity.*

Remark 3.5.2 *When the dimension p is increasing exponentially, then it is important to choose the distribution of G_i^*'s appropriately depending on whether the regression errors are sub-gaussian or sub-exponential. Note that if a random variable W_1 has distribution Beta(a_1, b_1), then by Hoeffding's inequality,*

$$\mathbf{E}\left[e^{\kappa(W_1 - \mathbf{E}W_1)}\right] \leq e^{\kappa^2/8} \text{ for all } \kappa \in \mathcal{R}$$

and hence W_1 is sub-gaussian with parameter value $1/4$, for any choice of (a_1, b_1). On the other hand, if W_2 has Gamma distribution with shape parameter a_2 and scale parameter b_2 then

$$\begin{aligned}\log \mathbf{E}\left[e^{\kappa(W_2 - \mathbf{E}W_2)}\right] &= -a_2 b_2 \kappa - a_2 \log(1 - b_2 \kappa), \text{ for } |\kappa| < 1/b_2 \\ &\leq \frac{a_2 b_2^2 \kappa^2}{2(1 - b_2 \kappa)}, \text{ for } |\kappa| < 1/b_2 \\ &\leq a_2 b_2^2 \kappa^2, \text{ for } |\kappa| < 1/2b_2\end{aligned}$$

where the first inequality follows from the fact that $-\log(1-u) \leq u + \dfrac{u^2}{2(1-u)}$ for $0 \leq u < 1$. Therefore W_2 is sub-exponential with parameters $(b_2\sqrt{2a_2}, 2b_2)$ and hence $W_1 + W_2$ is also sub-exponential with parameters $(\sqrt{1/4 + 2a_2 b_2^2}, 2b_2)$ when W_1 and W_2 are independent. These observations imply that Beta$(1/2, 3/2)$ is an appropriate choice for the distribution of G_i^*'s when the errors are sub-gaussian and the distribution of $(M_1 + M_2)$ is an appropriate choice for the distribution of G_i^*'s when the errors are sub-exponential where M_1 and M_2 are independent and M_1 is a Gamma random variable with shape and scale parameters 0.008652 and 2 respectively and M_2 is a Beta random variable with both the parameters 0.036490.

3.6 Proofs

First we define some additional notations and recall some notations that are defined earlier. Define. $\breve{\boldsymbol{W}}_n = n^{-1/2}\sum_{i=1}^n \epsilon_i \boldsymbol{x}_i$ and $\breve{\boldsymbol{W}}_n^* = n^{-1/2}\sum_{i=1}^n \hat{\epsilon}_i \boldsymbol{x}_i (G_i^* - \mu_{G^*})$. Write $\breve{\boldsymbol{W}}_n^{(0)} = \breve{\boldsymbol{W}}_n$, $\breve{\boldsymbol{W}}_n^{*(0)} = \breve{\boldsymbol{W}}_n^*$, $p^{(0)} = p$, $p^{(1)} = p_0$, and $p^{(2)} = p - p_0$. Define, $\tilde{b}_n = \sigma^{-1}\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{b}_n$ when $p \le n$ and $\tilde{b}_n = \sigma^{-1}\bar{\boldsymbol{\Sigma}}_n^{-1/2}\boldsymbol{b}_n$ when $p > n$. Recall that $\boldsymbol{b}_n = \boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{s}_n^{(1)}\dfrac{\lambda_n}{2\sqrt{n}}$, where $\boldsymbol{D}_n^{(1)}$ and $\boldsymbol{C}_{11,n}$ are as defined earlier and $\boldsymbol{s}_n^{(1)}$ is a $p_0 \times 1$ vector with jth element $\mathrm{sgn}(\beta_{j,n})|\beta_{j,n}|^{-\gamma}$. Note that under the conditions (A.2)(i), (A.3), and (A.6)(i), $||\boldsymbol{\Sigma}_n|| = O(1)$, $||\bar{\boldsymbol{\Sigma}}_n|| = O(1)$, $||\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1/2}|| = O(1)$ and $||\boldsymbol{s}_n^{(1)}|| \le K\sqrt{p_0}.n^{b\gamma}$. Hence, $||\tilde{\boldsymbol{b}}_n|| = O(n^{-\delta_1})$. Define $\breve{\boldsymbol{\xi}}_i^{(0)} = \breve{\boldsymbol{\Sigma}}_n^{-1/2}\hat{\boldsymbol{\xi}}_i^{(0)}$ and $\breve{\boldsymbol{\eta}}_i^{(0)} = \breve{\boldsymbol{\Sigma}}_n^{-1/2}\hat{\boldsymbol{\eta}}_i^{(0)}$ or $\breve{\boldsymbol{\xi}}_i^{(0)} = \tilde{\boldsymbol{\Sigma}}_n^{-1/2}\hat{\boldsymbol{\xi}}_i^{(0)}$ and $\breve{\boldsymbol{\eta}}_i^{(0)} = \tilde{\boldsymbol{\Sigma}}_n^{-1/2}\hat{\boldsymbol{\eta}}_i^{(0)}$, $i = 1, \ldots, n$, according as $p \le n$ or $p > n$. Here $\breve{\boldsymbol{\Sigma}}_n$, $\tilde{\boldsymbol{\Sigma}}_n$, $\hat{\boldsymbol{\xi}}_i^{(0)}$ and $\hat{\boldsymbol{\eta}}_i^{(0)}$ are as defined in Section 3.5. Also define $\breve{\boldsymbol{b}}_n = \breve{\boldsymbol{\Sigma}}_n^{-1/2}\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\breve{\boldsymbol{s}}_n^{(1)}\dfrac{\lambda_n}{2\sqrt{n}}$ when $p \le n$ and $\breve{\boldsymbol{b}}_n = \tilde{\boldsymbol{\Sigma}}_n^{-1/2}\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\breve{\boldsymbol{s}}_n^{(1)}\dfrac{\lambda_n}{2\sqrt{n}}$ when $p > n$, where $\breve{\boldsymbol{s}}_n^{(1)} = (\breve{s}_{1n}, \ldots, \breve{s}_{p_0 n})'$ and $\breve{s}_{j,n} = \mathrm{sgn}(\hat{\beta}_{j,n})|\hat{\beta}_{j,n}|^{-\gamma}$.

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 $f = (f_1, f_2, \ldots, f_l) : \mathscr{R}^l \to \mathscr{R}^l$, $l \ge 1$, 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, $1 \le j \le l$. 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 $\boldsymbol{0}$ and dispersion matrix $\boldsymbol{V}$ having jth row $\boldsymbol{V}_{j.}$ and let $\boldsymbol{\phi}_V$ denote the density of $\boldsymbol{\Phi}_V$. We write $\boldsymbol{\Phi}_V = \boldsymbol{\Phi}$ and $\boldsymbol{\phi}_V = \boldsymbol{\phi}$ when $\boldsymbol{V}$ is the identity matrix. Define 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, $\boldsymbol{A}_{1n} = \Big\{\{||\breve{\boldsymbol{W}}_n^{(1)}||_\infty \le K\sqrt{\log n}\} \cap \{||\breve{\boldsymbol{W}}_n^{(2)}||_\infty \le K\sqrt{\log n}\} \cap \{||\sqrt{n}(\tilde{\boldsymbol{\beta}} -$

$\boldsymbol{\beta})||_\infty \leq K\sqrt{\log n}\Big\}$ for $p \leq n$ and $\boldsymbol{A}_{1n} = \Big\{\{||\check{\boldsymbol{W}}_n^{(1)}||_\infty \leq K\sqrt{\log n}\} \cap \{||\check{\boldsymbol{W}}_n^{(2)}||_\infty \leq K\sqrt{\log n}\} \cap \{||\sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})||_\infty \leq C.n^{\delta_2}\}\Big\}$ for $p > n$. We have assumed $\mathcal{A}_n = \{1, \ldots, p_0\}$. $\check{\boldsymbol{W}}_n^{(1)}$ and $\check{\boldsymbol{W}}_n^{(2)}$ are respectively first p_0 and last $(p - p_0)$ components of $\check{\boldsymbol{W}}_n$. Note that, $\mathbf{P}(\boldsymbol{A}_{1n}) \geq 1 - O(p.n^{-(r-2)/2})$ for $p \leq n$ and $\mathbf{P}(\boldsymbol{A}_{1n}) \geq 1 - o(n^{-1/2})$ for $p > n$ [cf.Lemma 8.1 of Chatterjee and Lahiri (2013)].

Note that, $\check{\boldsymbol{b}}_n = O_p(n^{-\delta_1})$, by Lemma 3.6.4 and 3.6.5, described below. Suppose, $r_1 = \min\{a \in \mathcal{N} : ||\check{\boldsymbol{b}}_n||^{a+1} = o_p(n^{-1/2})\}$, $\mathcal{N}$ being the set of natural numbers. Define the conditional Lebesgue density of two-term Edgeworth expansion of $\boldsymbol{R}_n^*$ as

$$\begin{aligned}\zeta_n^*(\boldsymbol{x}) =& \phi(\boldsymbol{x})\Bigg[1 + \sum_{k=1}^{r_1} \frac{1}{k!}\Big\{\sum_{|\boldsymbol{\alpha}|=k} \check{\boldsymbol{b}}_n^{\alpha} H_{\boldsymbol{\alpha}}(\boldsymbol{x})\Big\} + \frac{1}{\sqrt{n}}\Bigg[\frac{1}{6}\sum_{|\boldsymbol{\alpha}|=3} \boldsymbol{t}^{\boldsymbol{\alpha}} \bar{\boldsymbol{\xi}}_n^{*(1)}(\boldsymbol{\alpha}) H_{\boldsymbol{\alpha}}(\boldsymbol{x}) \\ & - \frac{1}{2\hat{\sigma}_n^2}\Big\{\sum_{|\boldsymbol{\alpha}|=1} \boldsymbol{t}^{\boldsymbol{\alpha}} \bar{\boldsymbol{\xi}}_n^{*(3)}(\boldsymbol{\alpha}) H_{\boldsymbol{\alpha}}(\boldsymbol{x}) + \sum_{|\boldsymbol{\alpha}|=1}\sum_{|\boldsymbol{\zeta}|=2} \boldsymbol{t}^{\boldsymbol{\alpha}+\boldsymbol{\zeta}} \bar{\boldsymbol{\xi}}_n^{*(3)}(\boldsymbol{\alpha}) \bar{\boldsymbol{\xi}}_n^{*(1)}(\boldsymbol{\zeta}) H_{\boldsymbol{\alpha}+\boldsymbol{\zeta}}(\boldsymbol{x})\Big\}\Bigg]\Bigg],\end{aligned}$$

where $x \in \mathcal{R}^q$, $\bar{\boldsymbol{\xi}}_n^{*(j)}(\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^n \left(\check{\boldsymbol{\xi}}_i^{(0)} \hat{\epsilon}_i^j\right)^{\boldsymbol{\alpha}}$, $j = 0, 1, \ldots$ and $H_{\boldsymbol{\alpha}}(\boldsymbol{x}) = (-D)^{\boldsymbol{\alpha}}\phi(\boldsymbol{x})$, where $\phi(\cdot)$ is the standard normal density on $\mathcal{R}^q$.

Lemma 3.6.1 *Suppose $Y_1, \ldots, Y_n$ are zero mean independent r.v.s and $\mathbf{E}(|Y_i|^t) < \infty$ for $i = 1, \ldots, n$ and $\sum_{i=1}^n \mathbf{E}(|Y_i|^t) = \sigma_t$; $S_n = \sum_{i=1}^n Y_i$. Then, for any $t \geq 2$ and $x > 0$*

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

Proof of Lemma 3.6.1. This inequality was proved in Fuk and Nagaev (1971).

Lemma 3.6.2 *Under assumptions* (A.1), (A.3), (A.4)(i) *and* (A.5)(i), (ii) *with $r = 3$,*

(i) $\mathbf{P}_*\big(||\check{\boldsymbol{W}}_n^{*(1)}|| > K\sqrt{p_0 \log n}\big) = O_p(p_0.n^{-(r-2)/2})$.

(ii) $\mathbf{P}_*\big(||\check{\boldsymbol{W}}_n^{*(l)}||_\infty > K\sqrt{\log n}\big) = O_p(p^{(l)}.n^{-(r-2)/2})$, *for $l = 0, 1, 2$.*

(iii) $\mathbf{P}_*\big(||\sqrt{n}\big(\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n\big)||_\infty > K\sqrt{\log n}\big) = O_p(p.n^{-(r-2)/2})$*, when* $p \le n$*.*

Proof of Lemma 3.6.2. This lemma follows through the same line of Lemma 8.1 of Chatterjee and Lahiri (2013) and employing Lemma 3.6.1, stated above.

Lemma 3.6.3 *Suppose p is fixed. Then under condition* (A.1)(ii) *and* (A.4)(i) *with r=2,*

$$\mathbf{P}_*\Big(||\tilde{\boldsymbol{\beta}}_n^{*N} - \tilde{\boldsymbol{\beta}}_n|| = o\big(n^{-1/2}(\log n)^{1/2}\big)\Big) \ge 1 - o_p\big(n^{-1/2}\big)$$

Proof of Lemma 3.6.3. This lemma is proved in Proposition 1.4.1.

Lemma 3.6.4 *Suppose assumptions* (A.1)-(A.3), (A.4)(i), (A.5)(i), (ii) *and* (A.6) *hold with* $r = 4$*. Then*

$$||\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n||_\infty = O_p(n^{-1/2}) \ \ \textit{and on the set}\, \boldsymbol{A}_{1n}, \ \ ||\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n||_\infty = O_{p_*}(n^{-1/2})$$

Proof of Lemma 3.6.4. This lemma follows from Markov inequality and using the condition $\{n^{-1}\sum_{i=1}^n \big|(\boldsymbol{C}_{11,n}^{-1})_{j\cdot}\boldsymbol{x}_i^{(1)}\big|^{2r} : 1 \le j \le p_0\} = O(1)$ [stated in assumption (A.1)(ii)], after observing the form of the Alasso estimator obtained in (8.5) of Theorem 8.2 (a) of Chatterjee and Lahiri (2013) and the solution $\hat{\boldsymbol{u}}_n^*$ of the equation 3.6.2 obtained in the proof of part (a) of Lemma 3.6.6.

Lemma 3.6.5 *Under the assumptions* (A.1)-(A.3), (A.4)(i) *and* (A.6)(i) *and* (iii) *with* $r = 6$*, we have*

$$||\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n|| = o_p(n^{-(1+\delta_1)/2}),\ ||\check{\boldsymbol{\Sigma}}_n - \bar{\boldsymbol{\Sigma}}_n|| = o_p(n^{-1})$$

$$||\check{\boldsymbol{\Sigma}}_n - \sigma^2\boldsymbol{\Sigma}_n||, ||\tilde{\boldsymbol{\Sigma}}_n - \sigma^2\bar{\boldsymbol{\Sigma}}_n|| = O_p(n^{-1/2}),$$

where δ_1 *is as defined in assumption* (A.6).

Proof of Lemma 3.6.5. First we show that $||\hat{\Sigma}_n - \Sigma_n|| = o_p(n^{-(1+\delta_1)/2})$. Note that by Lemma 3.6.4, for $n \geq n_0$ (for some n_0),

$$\hat{\mathbf{\Sigma}}_n - \mathbf{\Sigma}_n = n^{-1}\sum_{i=1}^{n}(\hat{\boldsymbol{\eta}}_i^{(0)} - \boldsymbol{\eta}_i^{(0)})(\boldsymbol{\xi}_i^{(0)} + \boldsymbol{\eta}_i^{(0)})' + n^{-1}\sum_{i=1}^{n}(\boldsymbol{\xi}_i^{(0)} + \hat{\boldsymbol{\eta}}_i^{(0)})'(\hat{\boldsymbol{\eta}}_i^{(0)} - \boldsymbol{\eta}_i^{(0)})'$$

where on the set on the set $\boldsymbol{A}_{1n}$ we have

$$\sum_{i=1}^{n}||\boldsymbol{\eta}_i^{(0)} - \boldsymbol{\eta}_i^{(0)}||^2 \leq K^2(\gamma)||\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}||^2.\frac{\lambda_n^2}{n^2}\Big(\max_{1\leq j\leq p}\sum_{i=1}^{n}|\tilde{x}_{i,j}|^2\Big)||\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n^{(1)}||^2.n^{2b(\gamma+2)}$$

$$\leq K(\gamma,\delta_1)n^{-(1+2\delta_1)}$$

and

$$n^{-1}\sum_{i=1}^{n}||\boldsymbol{\xi}_i^{(0)}||^2 + n^{-1}\sum_{i=1}^{n}||\boldsymbol{\eta}_i^{(0)}||^2 + n^{-1}\sum_{i=1}^{n}||\hat{\boldsymbol{\eta}}_i^{(0)}||^2$$

$$\leq tr\big(\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{D}_n^{(1)}\big) + K||\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}||^2.\frac{\lambda_n^2}{n^2}\Big(\max_{1\leq j\leq p} n^{-1}\sum_{i=1}^{n}|\tilde{x}_{i,j}|^2\Big)||.p_0 n^{2b(\gamma+1)}$$

$$= O(1),$$

since $||\boldsymbol{C}_{11,n}^{-1/2}||^2 \leq K\min\{p_0, n^a\}$ and $||\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1/2}||^2 \leq q||\boldsymbol{D}_n^{(1)}\boldsymbol{C}_{11,n}^{-1}\boldsymbol{D}_n^{(1)'}||$.

Therefore, by the Cauchy-Schwarz inequality, we have $||\hat{\mathbf{\Sigma}}_n - \mathbf{\Sigma}_n|| = o_p(n^{-(1+\delta_1)/2})$. It follows directly from Lemma 3.6.4 that $\check{\mathbf{\Sigma}}_n = \bar{\mathbf{\Sigma}}_n$ for sufficiently large n. Hence $||\check{\mathbf{\Sigma}}_n - \bar{\mathbf{\Sigma}}_n|| = o_p(n^{-1})$.

Now to prove the second part, note that for $n \geq n_1$ (for some n_1),

$$\tilde{\mathbf{\Sigma}}_n - \sigma^2\bar{\mathbf{\Sigma}}_n = n^{-1}\sum_{i=1}^{n}\boldsymbol{\xi}_i^{(0)}\boldsymbol{\xi}_i^{(0)'}(\hat{\epsilon}_i^2 - \sigma^2)$$

and

$$
\begin{aligned}
\check{\boldsymbol{\Sigma}}_n - \sigma^2 \boldsymbol{\Sigma}_n = & n^{-1} \sum_{i=1}^{n} \boldsymbol{\xi}_i^{(0)} \boldsymbol{\xi}_i^{(0)'} (\hat{\epsilon}_i^2 - \sigma^2) + n^{-1} \sum_{i=1}^{n} \left(\hat{\boldsymbol{\eta}}_i^{(0)} - \boldsymbol{\eta}_i^{(0)}\right) \boldsymbol{\xi}_i^{(0)'} \hat{\epsilon}_i^2 \\
& + n^{-1} \sum_{i=1}^{n} \boldsymbol{\eta}_i^{(0)} \boldsymbol{\xi}_i^{(0)'} (\hat{\epsilon}_i^2 - \sigma^2) + n^{-1} \sum_{i=1}^{n} \boldsymbol{\xi}_i^{(0)} \left(\hat{\boldsymbol{\eta}}_i^{(0)} - \boldsymbol{\eta}_i^{(0)}\right)' \hat{\epsilon}_i^2 \\
& + n^{-1} \sum_{i=1}^{n} \boldsymbol{\xi}_i^{(0)} \boldsymbol{\eta}_i^{(0)'} (\hat{\epsilon}_i^2 - \sigma^2) + n^{-1} \sum_{i=1}^{n} \hat{\boldsymbol{\eta}}_i^{(0)} \left(\hat{\boldsymbol{\eta}}_i^{(0)} - \boldsymbol{\eta}_i^{(0)}\right)' \hat{\epsilon}_i^2 \\
& + n^{-1} \sum_{i=1}^{n} \left(\hat{\boldsymbol{\eta}}_i^{(0)} - \boldsymbol{\eta}_i^{(0)}\right) \boldsymbol{\eta}_i^{(0)'} \hat{\epsilon}_i^2 + n^{-1} \sum_{i=1}^{n} \boldsymbol{\eta}_i^{(0)} \boldsymbol{\eta}_i^{(0)'} (\hat{\epsilon}_i^2 - \sigma^2).
\end{aligned}
$$

Now we need to find the order of the term $||n^{-1} \sum_{i=1}^{n} \boldsymbol{\xi}_i^{(0)} \boldsymbol{\xi}_i^{(0)'} (\hat{\epsilon}_i^2 - \sigma^2)||$ to find the order of $||\check{\boldsymbol{\Sigma}}_n - \sigma^2 \boldsymbol{\Sigma}_n||$, since other terms can be shown to be of smaller order by using Hölder's inequality. Note that by Lemma 3.6.1, (A.1)(ii),

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

and due to Lemma 3.6.4, (A.1)(ii) and (A.2)(i) and (ii),

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

Hence, the second part of Lemma 3.6.5 follows.

Lemma 3.6.6 *Let $p \le n$ and suppose that* (A.1)–(A.6) *hold with $r = 6$. Then on a set A_{2n} with $\mathbf{P}(\boldsymbol{\varepsilon} \in \boldsymbol{A}_{2n}) \to 1$, when $\boldsymbol{\varepsilon} \in \boldsymbol{A}_{2n}$, we have*

(a) *if $p \le n$, then*

$$
\sup_{B \in \mathcal{C}_q} \left| \mathbf{P}_*(\boldsymbol{R}_n^* \in B) - \int_B \xi_n^*(\boldsymbol{x}) d\boldsymbol{x} \right| = o(n^{-1/2}),
$$

(b) *if $p > n$, $b = 0$ and additionally conditions* (A.7) *and* (A.1)(iii)′ *(in place of* (A.1)(iii)*)* *hold, then*

$$\sup_{B\in\mathcal{C}_q}\left|\mathbf{P}_*(\boldsymbol{R}_n^*\in B)-\int_B \xi_n^*(\boldsymbol{x})d\boldsymbol{x}\right|=o(n^{-1/2}).$$

Proof of Lemma 3.6.6. The modified perturbation bootstrap Alasso estimator is given by

$$\hat{\boldsymbol{\beta}}_n^*=\arg\min_{\boldsymbol{t}^*}\Bigg[\sum_{i=1}^n(y_i-\boldsymbol{x}_i'\boldsymbol{t}^*)^2(G_i^*-\mu_{G^*})$$
$$+\sum_{i=1}^n[\boldsymbol{x}_i'(\boldsymbol{t}^*-\hat{\boldsymbol{\beta}}_n)]^2(2\mu_{G^*}-G_i^*)+\mu_{G^*}\lambda_n\sum_{j=1}^p|\tilde{\beta}_{j,n}^*|^{-\gamma}|t_{j,n}^*|\Bigg].$$

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

$$\hat{\boldsymbol{u}}_n^*=\arg\min_{\boldsymbol{v}^*}\left[\boldsymbol{v}^{*\prime}\boldsymbol{C}_n\boldsymbol{v}^*-2\boldsymbol{v}^*\mu_{G^*}^{-1}\breve{\boldsymbol{W}}_n^*+\lambda_n\sum_{j=1}^p|\tilde{\beta}_{j,n}^*|^{-\gamma}\Big(|\hat{\beta}_{j,n}+\frac{v_j^*}{\sqrt{n}}|-|\hat{\beta}_{j,n}|\Big)\right]$$
$$=\arg\min_{\boldsymbol{v}^*}\boldsymbol{Z}_n(\boldsymbol{v}^*)\qquad\text{(say).}\tag{3.6.1}$$

Note that $\boldsymbol{Z}_n(\boldsymbol{v}^*)$ is convex in $\boldsymbol{v}^*$. Hence, the KKT condition is necessary and sufficient. The KKT condition corresponding to (3.6.1) is given by

$$2\boldsymbol{C}_n\boldsymbol{v}^*-2\mu_{G^*}^{-1}\breve{\boldsymbol{W}}_n^*+\frac{\lambda_n}{\sqrt{n}}\breve{\boldsymbol{\Gamma}}_n^*\breve{\boldsymbol{l}}_n=\boldsymbol{0}\tag{3.6.2}$$

for some $\breve{l}_{j,n}\in[-1,1]$ for all $j\in\{1,\dots,p\}$, where $\breve{\boldsymbol{l}}_n=(\breve{l}_{1n},\dots,\breve{l}_{pn})'$ and $\breve{\boldsymbol{\Gamma}}_n^*=diag\big(|\tilde{\beta}_{1n}^*|^{-\gamma},\dots,|\tilde{\beta}_{pn}^*|^{-\gamma}\big)$. It is easy to show that on the set $\boldsymbol{A}_{1n}$, $\Big(\big(\hat{\boldsymbol{u}}_n^{*(1)}\big)',\boldsymbol{0}'\Big)'$, where $\hat{\boldsymbol{u}}_n^{*(1)}=\boldsymbol{C}_{11,n}^{-1}\Big[\mu_{G^*}^{-1}\breve{\boldsymbol{W}}_n^{*(1)}-\frac{\lambda_n}{2\sqrt{n}}\tilde{\boldsymbol{s}}_n^{*(1)}\Big]$ is the unique solution of (3.6.2) and hence $\hat{\boldsymbol{u}}_n^*=\Big(\big(\hat{\boldsymbol{u}}_n^{*(1)}\big)',\boldsymbol{0}'\Big)'$, is the unique solution of the minimization problem (3.6.1), where $\tilde{\boldsymbol{s}}_n^{*(1)}=(\tilde{s}_{1n}^*,\dots,\tilde{s}_{p_0n}^*)$ and $\tilde{s}_{j,n}^*=sgn(\hat{\beta}_{j,n})|\tilde{\beta}_{j,n}^*|^{-\gamma}$.

To prove part (a) note that

$$\begin{aligned}
\hat{\sigma}_n^* \hat{\sigma}_n^{-1} \boldsymbol{R}_n^* &= \breve{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{T}_n^* \qquad (3.6.3)\\
&= \breve{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{D}_n^{(1)} \hat{\boldsymbol{u}}_n^{*(1)} \\
&= \breve{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \Big[\mu_{G^*}^{-1} \breve{\boldsymbol{W}}_n^{*(1)} - \frac{\lambda_n}{2\sqrt{n}} \tilde{\boldsymbol{s}}_n^{*(1)} \Big] \\
&= \mu_{G^*}^{-1} n^{-1/2} \sum_{i=1}^{n} (\breve{\boldsymbol{\xi}}_i^{(0)} + \breve{\boldsymbol{\eta}}_i^{(0)}) \hat{\epsilon}_i (G_i^* - \mu_{G^*}) - \breve{\boldsymbol{b}}_n + Q_{1n}^* \\
&= \boldsymbol{T}_{1n}^* - \breve{\boldsymbol{b}}_n + Q_{1n}^* \quad (say) \qquad (3.6.4)
\end{aligned}$$

Again note that

$$\begin{aligned}
&\mu_{G^*}^2 \left[\hat{\sigma}_n^{*2} - \hat{\sigma}_n^2 \right] \\
=& n^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^{*2} (G_i^* - \mu_{G^*})^2 - n^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^2 \sigma_{G^*}^2 \\
=& n^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^2 \Big[(G_i^* - \mu_{G^*})^2 - \sigma_{G^*}^2 \Big] + 2n^{-1} \sum_{i=1}^{n} (\hat{\epsilon}_i^* - \hat{\epsilon}_i) \hat{\epsilon}_i \Big[(G_i^* - \mu_{G^*})^2 - \sigma_{G^*}^2 \Big] \\
&+ 2n^{-1} \sum_{i=1}^{n} (\hat{\epsilon}_i^* - \hat{\epsilon}_i) \hat{\epsilon}_i \sigma_{G^*}^2 + n^{-1} \sum_{i=1}^{n} (\hat{\epsilon}_i^* - \hat{\epsilon}_i)^2 (G_i^* - \mu_{G^*})^2, \qquad (3.6.5)
\end{aligned}$$

where under condition (A.1), (A.5)(i) and (A.6)(iii) we have that the order of the last three terms in the expression of $\mu_{G^*}^2 \left[\hat{\sigma}_n^{*2} - \hat{\sigma}_n^2 \right]$ is $o_{p_*}\big(n^{-1/2} (\log n)^{-1/2}\big)$ on the set $\boldsymbol{A}_{1n}$, whereas

$$\mathbf{P}_* \Big(n^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^2 \Big[(G_i^* - \mu_{G^*})^2 - \sigma_{G^*}^2 \Big] > K n^{-1/2} (\log n)^{1/2} \Big) = O(p_0 n^{-(r-2)/2}),$$

by Lemma 3.6.1.

Therefore, considering Taylor's expansion of $\hat{\sigma}_n^{*-1}$ around $\hat{\sigma}_n^{-1}$, we have

$$\begin{aligned}
\boldsymbol{R}_n^* &= \boldsymbol{T}_{1n}^* - \check{\boldsymbol{b}}_n - \left(2\hat{\sigma}_n^2\right)^{-1}\mu_{G^*}^{-3} Z_{1n}^* \check{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \check{\boldsymbol{W}}_n^{*(1)} + Q_{2n}^* \\
&= \boldsymbol{R}_{1n}^* - \check{\boldsymbol{b}}_n + Q_{2n}^*, \quad \text{(say)} \qquad (3.6.6)
\end{aligned}$$

where $Z_{1n}^* = n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^2 \Big[(G_i^* - \mu_{G^*})^2 - \sigma_{G^*}^2\Big]$ and on the set $\boldsymbol{A}_{1n}$ we have $||Q_{2n}^*|| = o_{p_*}(n^{-1/2})$.

The first three cumulants of $\boldsymbol{t}'\boldsymbol{R}_{1n}^*$ are given by

$$\begin{aligned}
\kappa_1\big(\boldsymbol{t}'\boldsymbol{R}_{1n}^*\big) &= -\frac{1}{\sqrt{n}}.\frac{1}{2\hat{\sigma}_n^2}\textstyle\sum_{|\boldsymbol{\alpha}|=1} \boldsymbol{t}^{\boldsymbol{\alpha}} \bar{\boldsymbol{\xi}}_n^{*(3)}(\boldsymbol{\alpha}) + o_p(n^{1/2}) \\
\kappa_2\big(\boldsymbol{t}'\boldsymbol{R}_{1n}^*\big) &= \mathbf{Var}_*\big(\boldsymbol{t}'\boldsymbol{R}_{1n}^*\big) = \boldsymbol{t}'\boldsymbol{t} + o_p(n^{-1/2}) \\
\kappa_3\big(\boldsymbol{t}'\boldsymbol{R}_{1n}^*\big) &= \mathbf{E}_*\big(\boldsymbol{t}'\boldsymbol{R}_{1n}^*\big)^3 - 3\mathbf{E}_*\big(\boldsymbol{t}'\boldsymbol{R}_{1n}^*\big)^2.\mathbf{E}_*\big(\boldsymbol{t}'\boldsymbol{R}_{1n}^*\big) + 2\Big(\mathbf{E}_*\big(\boldsymbol{t}'\boldsymbol{R}_{1n}^*\big)\Big)^3 \\
&= \frac{1}{\sqrt{n}}\left[\textstyle\sum_{|\boldsymbol{\alpha}|=3} \boldsymbol{t}^{\boldsymbol{\alpha}} \bar{\boldsymbol{\xi}}_n^{*(1)}(\boldsymbol{\alpha}) - \frac{3}{\hat{\sigma}_n^2}\sum_{|\boldsymbol{\alpha}|=1}\sum_{|\boldsymbol{\zeta}|=2} \boldsymbol{t}^{\boldsymbol{\alpha}+\boldsymbol{\zeta}} \bar{\boldsymbol{\xi}}_n^{*(3)}(\boldsymbol{\alpha})\bar{\boldsymbol{\xi}}_n^{*(1)}(\boldsymbol{\zeta})\right] + o_p(n^{-1/2}).
\end{aligned}$$

Now, using the quadratic form technique of Das and Lahiri (2016), we have on the set $\boldsymbol{A}_{1n}$

$$\sup_{B\in\mathcal{C}_q}\left|\mathbf{P}_*(\boldsymbol{R}_{1n}^* \in B) - \int_B \xi_{1n}^*(\boldsymbol{x})d\boldsymbol{x}\right| = o(n^{-1/2}),$$

where

$$\begin{aligned}
\xi_{1n}^*(\boldsymbol{x}) =& \phi(\boldsymbol{x})\Bigg[1 + \frac{1}{\sqrt{n}}\Bigg[\frac{1}{6}\sum_{|\boldsymbol{\alpha}|=3} \boldsymbol{t}^{\boldsymbol{\alpha}} \bar{\boldsymbol{\xi}}_n^{*(1)}(\boldsymbol{\alpha}) H_{\boldsymbol{\alpha}}(\boldsymbol{x}) \\
&- \frac{1}{2\hat{\sigma}_n^2}\Big\{\sum_{|\boldsymbol{\alpha}|=1} \boldsymbol{t}^{\boldsymbol{\alpha}} \bar{\boldsymbol{\xi}}_n^{*(3)}(\boldsymbol{\alpha}) H_{\boldsymbol{\alpha}}(\boldsymbol{x}) + \sum_{|\boldsymbol{\alpha}|=1}\sum_{|\boldsymbol{\zeta}|=2} \boldsymbol{t}^{\boldsymbol{\alpha}+\boldsymbol{\zeta}} \bar{\boldsymbol{\xi}}_n^{*(3)}(\boldsymbol{\alpha})\bar{\boldsymbol{\xi}}_n^{*(1)}(\boldsymbol{\zeta}) H_{\boldsymbol{\alpha}+\boldsymbol{\zeta}}(\boldsymbol{x})\Big\}\Bigg]\Bigg].
\end{aligned}$$

Now, Lemma 3.6.6 part (a) follows by Corollary 2.6 of Bhattacharya and Rao (1986) and noting that $\{B + \boldsymbol{b} : B \in \mathcal{C}_q\} = \mathcal{C}_q$ and that

$$\begin{aligned}\mathbf{P}(\boldsymbol{R}_n^* \in B) &= \mathbf{P}(\boldsymbol{R}_{1n}^* \in B + \check{\boldsymbol{b}}_n) + o(n^{-1/2})\\ &= \int_{B+\check{\boldsymbol{b}}_n} \xi_{1n}^*(\boldsymbol{x})d\boldsymbol{x} + o(n^{-1/2})\\ &= \int_B \xi_{1n}^*(\boldsymbol{x} + \check{\boldsymbol{b}}_n)d\boldsymbol{x} + o(n^{-1/2})\\ &= \int_B \xi_n^*(\boldsymbol{x})d\boldsymbol{x} + o(n^{-1/2}).\end{aligned}$$

Now for part (b) note that for $n \geq n_0$, on the set $\boldsymbol{A}_{1n}$ we have

$$\begin{aligned}\hat{\sigma}_n^* \hat{\sigma}_n^{-1} \boldsymbol{R}_n^* &= \tilde{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{T}_n^*\\ &= \tilde{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{D}_n^{(1)} \hat{\boldsymbol{u}}_n^{*(1)}\\ &= \tilde{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \Big[\mu_{G^*}^{-1} \check{\boldsymbol{W}}_n^{*(1)} - \frac{\lambda_n}{2\sqrt{n}} \tilde{\boldsymbol{s}}_n^{*(1)}\Big]\\ &= \mu_{G^*}^{-1} \tilde{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \check{\boldsymbol{W}}_n^{*(1)} - \check{\boldsymbol{b}}_n + Q_{3n}^* \quad \text{(say)}\\ &= \mu_{G^*}^{-1} n^{-1/2} \sum_{i=1}^n \check{\boldsymbol{\xi}}_i^{(0)} \hat{\epsilon}_i (G_i^* - \mu_{G^*}) - \check{\boldsymbol{b}}_n + Q_{3n}^* && (3.6.7)\\ &= \boldsymbol{T}_{2n}^* - \check{\boldsymbol{b}}_n + Q_{3n}^*, \quad \text{(say)} && (3.6.8)\end{aligned}$$

where $Q_{3n}^* = \tilde{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{\Delta}_n^{*(1)} + Q_{1n}^*$, where $\boldsymbol{\Delta}_n^{*(1)}$ is a $p_0 \times 1$ vector with jth component $\lambda_n n^{-1/2} \big(\tilde{\beta}_{j,n}^* - \hat{\beta}_{j,n}\big) \gamma \check{s}_{j,n} |\hat{\beta}_{j,n}|^{-1}$ and Q_{1n}^* is as defined in part (a).

Now since $b = 0$, by (A.1)(iii)′, (A.2)(i), (A.3), Lemma 3.6.4 and the fact that $||Q_{1n}^*|| = o_p(n^{-1/2})$, one can show that on the set $\boldsymbol{A}_{1n}$,

$$\mathbf{P}_*\big(||Q_{3n}^*|| > K(p_0 \lambda_n n^{-1+\delta_2} + o(n^{-1/2})\big) = o(n^{-1/2}).$$

Now since by $(A.6)(i)$, $p_0\lambda_n n^{-1+\delta_2} = o(n^{-1/2})$, similarly to (3.6.6), we have

$$\begin{aligned}
\boldsymbol{R}_n^* &= \boldsymbol{T}_{2n}^* - \check{\boldsymbol{b}}_n - \left(2\hat{\sigma}_n^2\right)^{-1}\mu_{G^*}^{-3} Z_{1n}^* \check{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \check{\boldsymbol{W}}_n^{*(1)} + Q_{4n}^* \\
&= \boldsymbol{R}_{2n}^* - \check{\boldsymbol{b}}_n + Q_{4n}^* \quad \text{(say)} \qquad (3.6.9)
\end{aligned}$$

where on the set $\boldsymbol{A}_{1n}$ we have $||Q_{4n}^*|| = o_{p_*}(n^{-1/2})$. Therefore, two-term Edgeworth expansions of $\boldsymbol{R}_n^*$ and $\boldsymbol{R}_{2n}^* - \check{\boldsymbol{b}}_n$ coincide on the set $\boldsymbol{A}_{1n}$, by Corollary 2.6 of Bhattachary and Rao (1986). Rest of part (b) of Lemma 3.6.6 follows analogously to part (a).

Proof of Proposition 3.2.1. Note that for any $\boldsymbol{t} \in \mathcal{R}^p$ and for each $i \in \{1, \ldots, n\}$, $y_i - \boldsymbol{x}_i'\boldsymbol{t} = \hat{\epsilon}_i + \boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n - \boldsymbol{t})$, and hence

$$\begin{aligned}
\sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i'\boldsymbol{t}^*\right)^2(G_i^* - \mu_{G^*}) &= \sum_{i=1}^{n}\left[\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n - \boldsymbol{t})\right]^2(G_i^* - \mu_{G^*}) \\
&\quad - 2(\boldsymbol{t} - \hat{\boldsymbol{\beta}}_n)'\check{\boldsymbol{W}}_n^* + \sum_{i=1}^{n}\hat{\epsilon}_i^2(G_i^* - \mu_{G^*})^2.
\end{aligned}$$

Therefore,

$$\arg\min_{\boldsymbol{t}} \boldsymbol{L}_1(\boldsymbol{t}) = \arg\min_{\boldsymbol{t}} \left[\sum_{i=1}^{n}\left[\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n - \boldsymbol{t})\right]^2 - 2\mu_{G^*}^{-1}(\boldsymbol{t} - \hat{\boldsymbol{\beta}}_n)'\check{\boldsymbol{W}}_n^* + c\lambda_n \sum_{j=1}^{p} c_j|t_j|^l\right].$$

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

$$\sum_{i=1}^{n}(z_i - \boldsymbol{x}_i'\boldsymbol{t})^2 = \sum_{i=1}^{n}\left[\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}}_n - \boldsymbol{t})\right]^2 - 2\mu_{G^*}^{-1}(\boldsymbol{t} - \hat{\boldsymbol{\beta}}_n)'\check{\boldsymbol{W}}_n^* + \mu_{G^*}^{-2}\sum_{i=1}^{n}\left[\hat{\epsilon}_i(G_i^* - \mu_{G^*})\right]^2.$$

Therefore, Proposition 3.2.1 follows.

Proof of Theorem 3.4.1. The KKT condition (3.4.3) corresponding to the Alasso criterion function, defined in MTC(11), can be rewritten through the vector $\boldsymbol{w}^* =$

$\left(\boldsymbol{w}_n^{*(1)\prime}, \boldsymbol{w}_n^{*(2)\prime}\right)'$ as

$$2\boldsymbol{C}_{11,n}^{*}\boldsymbol{w}_n^{*(1)} + 2\boldsymbol{C}_{12,n}^{*}\boldsymbol{w}_n^{*(2)} - 2\boldsymbol{W}_n^{*(1)} + \frac{\lambda_n^*}{\sqrt{n}}\boldsymbol{\Gamma}_n^{*(1)}\boldsymbol{l}_n^{(1)} = \boldsymbol{0} \tag{3.6.10}$$

and for each $j \in \{p_0+1, \ldots, p\}$

$$-\frac{\lambda_n^*}{2\sqrt{n}}|\tilde{\beta}_{j,n}^{*N}|^{-\gamma} \leq \left[\left(\boldsymbol{C}_{21,n}^{*}\right)_{j.}\boldsymbol{w}_n^{*(1)} + \left(\boldsymbol{C}_{22,n}^{*}\right)_{j.}\boldsymbol{w}_n^{*(2)} - W_{j,n}^{*}\right] \leq \frac{\lambda_n^*}{2\sqrt{n}}|\tilde{\beta}_{j,n}^{*N}|^{-\gamma}. \tag{3.6.11}$$

Here, $\boldsymbol{W}_n^* = n^{-1/2}\sum_{i=1}^n \hat{\epsilon}_i\boldsymbol{x}_i G_i^*$, $\boldsymbol{W}_n^{*(1)}$ is the vector of the first p_0 components of $\boldsymbol{W}_n^*$, $W_{j,n}^*$ is the jth component of $\boldsymbol{W}_n^*$ for $j \in \{1, \ldots, p\}$, $\boldsymbol{l}_n^{(1)} = (l_{1n}, \ldots, l_{p_0 n})'$ with $l_{k,n} \in [-1, 1]$ for $k = 1, \ldots, p_0$ and $\boldsymbol{\Gamma}_n^{*(1)} = \text{diag}\left(|\tilde{\beta}_{1n}^{*N}|^{-\gamma}, \ldots, |\tilde{\beta}_{p_0 n}^{*N}|^{-\gamma}\right)$ and $\boldsymbol{C}_n^* = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i' G_i^* = \begin{bmatrix} \boldsymbol{C}_{11,n}^* & \boldsymbol{C}_{12,n}^* \\ \boldsymbol{C}_{21,n}^* & \boldsymbol{C}_{22,n}^* \end{bmatrix}$ where $\boldsymbol{C}_{11,n}^*$ is of dimension $p_0 \times p_0$. $\left(\boldsymbol{C}_{21,n}^*\right)_{j.}$ is the jth row of $\boldsymbol{C}_{21,n}^*$, $j \in \{p_0+1, \ldots, p\}$.

Now, to prove part (a) of Theorem 3.4.1, it is enough to show that $\left(\boldsymbol{u}_{n2}^{*N\prime}, \boldsymbol{0}'\right)'$ satisfies (3.6.10) and (3.6.11) separately with bootstrap probability $1 - o_p(n^{-1/2})$. The vector $\boldsymbol{u}_{n2}^{*N}$ is defined as $\boldsymbol{u}_{n2}^{*N} = \boldsymbol{C}_{11,n}^{*-1}\left[\boldsymbol{W}_n^{*(1)} - \dfrac{\lambda_n^*}{\sqrt{n}}\tilde{\boldsymbol{s}}_n^{*N(1)}\right]$, where the jth component of $\tilde{\boldsymbol{s}}_n^{*N(1)}$ is equal to $\text{sgn}(\hat{\beta}_{j,n})|\tilde{\beta}_{jn}^{*N}|^{-\gamma}$, $j \in \{1, \ldots, p_0\}$.

Note that $\left(\boldsymbol{u}_{n2}^{*N\prime}, \boldsymbol{0}'\right)'$ exactly satisfies (3.6.10) if $\boldsymbol{l}_n^{(1)} = \left(\text{sgn}(\hat{\beta}_{1,n}), \ldots, \hat{\beta}_{p_0,n})\right)$. Thus we can conclude that $\left(\boldsymbol{u}_{n2}^{*N\prime}, \boldsymbol{0}'\right)'$ satisfies (3.6.10) with bootstrap probability $1 - o_p$ $(n^{-1/2})$, if we can show that $\left|\left|\boldsymbol{C}_{11,n}^{*-1}\left[\boldsymbol{W}_n^{*(1)} - \dfrac{\lambda_n^*}{\sqrt{n}}\boldsymbol{\Gamma}_n^{*(1)}\boldsymbol{l}_n^{(1)}\right]\right|\right| = o(n^{1/2})$ with bootstrap probability $1 - o_p(n^{-1/2})$. Under the assumptions (A.1)(ii) and (A.4)(i) with $r = 4$, we have on the set $\boldsymbol{A}_{1n}$,

$$\begin{aligned}
&\mathbf{P}_*\left(||\boldsymbol{C}_{11,n}^* - \boldsymbol{C}_{11,n}\mu_{G^*}|| > K.p_0.n^{-1/2}.(\log n)^{1/2}\right) \\
&\leq \sum_{j,k=1}^{p_0} \mathbf{P}_*\left(\Big|\sum_{i=1}^n x_{ij}x_{ik}(G_i^* - \mu_{G^*})\Big| > K.n^{-1/2}.(\log n)^{1/2}\right)
\end{aligned}$$

$$= o(n^{-1/2}) \tag{3.6.12}$$

and

$$n^{-1/2}\sum_{i=1}^{n}\boldsymbol{x}_i^{(1)}\hat{\epsilon}_i = \frac{\lambda_n}{2\sqrt{n}}\Big(\text{sgn}(\beta_{1,n})|\tilde{\beta}_{1,n}|^{-\gamma},\ldots,\text{sgn}(\beta_{p_0,n})|\tilde{\beta}_{p_0,n}|^{-\gamma}\Big)' = \frac{\lambda_n}{2\sqrt{n}}\tilde{\mathbf{s}}_n^{(1)}(\text{say}),$$

where on the set $\boldsymbol{A}_{1n}$, $|\tilde{\beta}_{j,n}|^{-\gamma}$ is bounded for all $j \in \{1,\ldots,p_0\}$ and $n^{-1/2}\lambda_n \to 0$. These facts along with Proposition 3.6.3 imply that on the set $\boldsymbol{A}_{1n}$

$$\mathbf{P}_*\bigg(\boldsymbol{C}_{11,n}^{*-1}\Big[\boldsymbol{W}_n^{*(1)} - \frac{\lambda_n^*}{\sqrt{n}}\boldsymbol{\Gamma}_n^{*(1)}\boldsymbol{l}_n^{(1)}\Big] = o(n^{1/2})\bigg) = 1 - o(n^{-1/2}).$$

Now, note that on the set $\boldsymbol{A}_{1n}$

$$\mathbf{P}_*\Big(\max_j\big\{||(\boldsymbol{C}_{21,n}^*)_{j\cdot} - (\boldsymbol{C}_{21,n})_{j\cdot}\mu_{G^*}|| : j \in \{p_0+1,\ldots,p\}\big\} > K.p_0^{1/2}.n^{-1/2}.(\log n)^{1/2}\Big)$$

$$\leq \sum_{k=1}^{p_0}\sum_{j=p_0+1}^{p}\mathbf{P}_*\Big(\big|\sum_{i=1}^{n}x_{ij}x_{ik}(G_i^* - \mu_{G^*})\big| > K.n^{-1/2}.(\log n)^{1/2}\Big)$$

$$= o(n^{-1/2}),$$

and due to Lemma 3.6.3,

$$\mathbf{P}_*\Big(\min_j\big\{|\tilde{\beta}_{j,n}^{*N}|^{-\gamma} : j \in \{p_0+1,\ldots,p\}\big\} > K.n^{\gamma/2}(\log n)^{-\gamma/2}\Big) = 1 - o_p(n^{-1/2}).$$

Again for $j \in \{p_0+1,\ldots,p\}$,

$$\begin{aligned}W_{jn}^* =& n^{-1/2}\sum_{i=1}^{n}\hat{\epsilon}_i x_{ij}(G_i^* - \mu_{G^*}) + \mu_{G^*}.n^{-1/2}\sum_{i=1}^{n}x_{ij}\epsilon_i \\ & - \mu_{G^*}\Big(n^{-1}\sum_{i=1}^{n}x_{ij}\boldsymbol{x}_i^{(1)}\Big)'\boldsymbol{C}_{11,n}^{-1}\Big[n^{-1/2}\sum_{i=1}^{n}\epsilon_i\boldsymbol{x}_i^{(1)} - \frac{\lambda_n}{2\sqrt{n}}\tilde{\mathbf{s}}_n^{(1)}\Big]\end{aligned}$$

Since, $\boldsymbol{C}_n \to \boldsymbol{C}$, a pd matrix, and $\max\{\lambda_n, \lambda_n^*\}.(\log n/n)^{1/2} \to 0$, we have

$$\mathbf{P}_*\Big(|W_{jn}^*| > K.(\log n)^{1/2}\Big) = 1 - o_p(n^{-1/2}).$$

Hence due to $\min\{\lambda_n, \lambda_n^*\}.(\log n)^{-(\gamma+1)/2}.n^{(\gamma-1)/2} \to \infty$, we have on the set $\boldsymbol{A}_{1n}$

$$\mathbf{P}_*\Bigg(\big(\boldsymbol{u}_{n2}^{*N\prime}, \mathbf{0}'\big)' \text{ satisfies (3.6.11)}\Bigg) = 1 - o_p(n^{-1/2}).$$

Therefore part (a) of Theorem 3.4.1 follows.

Now for part (b), note that since $n^{-1/2}\sum_{i=1}^n \boldsymbol{x}_i^{(1)}\hat{\epsilon}_i = \dfrac{\lambda_n}{2\sqrt{n}}\Big(\text{sgn}(\beta_{1,n})|\tilde{\beta}_{1,n}|^{-\gamma}, \ldots,$ $\text{sgn}(\beta_{p_0,n})|\tilde{\beta}_{p_0,n}|^{-\gamma}\Big)' = \dfrac{\lambda_n}{2\sqrt{n}}\tilde{\boldsymbol{s}}_n^{(1)}$ (say), so due to (3.6.12) and the fact that $n^{-1/2}.$ $(\log n)^{1/2}.\lambda_n \to 0$, it follows that on the set $\boldsymbol{A}_{1n}$,

$$\mathbf{P}_*\Bigg(\sqrt{n}\bigg|\bigg|(\boldsymbol{C}_{11,n}^{*-1} - \boldsymbol{C}_{11,n}^{-1}\mu_{G^*}^{-1})n^{-1/2}\sum_{i=1}^n \boldsymbol{x}_i^{(1)}\hat{\epsilon}_i\bigg|\bigg|_\infty = o(1)\Bigg) = 1 - o(n^{-1/2}). \quad (3.6.13)$$

Again, as $\boldsymbol{C}_n \to \boldsymbol{C}$ for some $p \times p$ positive definite matrix $\boldsymbol{C}$ and $\mathbf{P}\Big(||\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}|| = O\big(n^{-1/2}(\log n)^{1/2}\big)\Big) \geq 1 - o\big(n^{-1/2}\big)$, we have $\mathbf{P}\Big(\boldsymbol{A}_{1n} \cap \boldsymbol{A}_{1n}^\epsilon\Big) \to 1$ for $A_{1n}^\epsilon = \{\lambda_n||\boldsymbol{C}_{11,n}^{-1}\tilde{\boldsymbol{s}}_n^{(1)}|| > \epsilon^{-1}\}$ for any $\epsilon > 0$. Hence, on the set $\boldsymbol{A}_{1n} \cap \boldsymbol{A}_{1n}^\epsilon$ we have

$$\mathbf{P}_*\Big(Z_n^* > \epsilon\Big) = o_p\big(n^{-1/2}\big).$$

Therefore part (b) follows.

Now to prove part (c), It is enough to show

$$\sup_{\boldsymbol{x} \in \mathcal{R}^{p_0}} \Big|\mathbf{P}_*\big(\boldsymbol{F}_n^{*(1)} \leq \boldsymbol{x}\big) - \mathbf{P}\big(\boldsymbol{F}_n^{(1)} \leq \boldsymbol{x}\big)\Big| \geq K.\frac{\lambda_n}{\sqrt{n}} \text{ for some } K > 0. \quad (3.6.14)$$

where $\boldsymbol{F}_n^{*(1)}$ and $\boldsymbol{F}_n^{(1)}$ are sub vectors of $\boldsymbol{F}_n^*$ and $\boldsymbol{F}_n$ respectively, comprising of first p_0

components. Note that

$$\begin{aligned}
\boldsymbol{F}_n^{*(1)} &= \boldsymbol{C}_{11,n}^{*-1}\Big[\boldsymbol{W}_n^{*(1)} - \frac{\lambda_n^*}{\sqrt{n}}\tilde{\boldsymbol{s}}_n^{*N(1)}\Big] \\
&= \boldsymbol{C}_{11,n}^{-1}\mu_{G^*}^{-1}\Big[\boldsymbol{W}_n^{*(1)} - \frac{\lambda_n^*}{\sqrt{n}}\tilde{\boldsymbol{s}}_n^{*N(1)}\Big] + \big(\boldsymbol{C}_{11,n}^{*-1} - \boldsymbol{C}_{11,n}^{-1}\mu_{G^*}^{-1}\big)\Big[\boldsymbol{W}_n^{*(1)} - \frac{\lambda_n^*}{\sqrt{n}}\tilde{\boldsymbol{s}}_n^{*N(1)}\Big] \\
&= \breve{\boldsymbol{F}}_n^{*(1)} + \breve{\boldsymbol{R}}_{1n}^* \ \text{ (say)}
\end{aligned}$$

where $\boldsymbol{W}_n^{*(1)} = n^{-1/2}\sum_{i=1}^n \hat{\epsilon}_i \boldsymbol{x}_i(G_i^* - \mu_{G^*}) + n^{-1/2}\sum_{i=1}^n \hat{\epsilon}_i \boldsymbol{x}_i \mu_{G^*}$ with $n^{-1/2}\sum_{i=1}^n \boldsymbol{x}_i^{(1)}\hat{\epsilon}_i = \dfrac{\lambda_n}{2\sqrt{n}}\tilde{\boldsymbol{s}}_n^{(1)}$. Hence due to the fact that $\max\{\lambda_n, \lambda_n^*\}.n^{-1/2} \to 0$ and $\tilde{\boldsymbol{s}}_n^{*N(1)}$ & $\tilde{\boldsymbol{s}}_n^{(1)}$ are bounded in respective probabilities, it follows from Lemma 3.6.1 that

$$\mathbf{P}_*\Big(||\breve{\boldsymbol{R}}_{1n}^*|| \le c_n.n^{-1/2}\Big) = 1 - o_p(1)$$

where $\{c_n\}$ is a sequence of positive constants increasing to ∞ with $c_n = o(\sqrt{\log n})$.

Now write $\breve{\boldsymbol{F}}_n^{*(1)} = \tilde{\boldsymbol{F}}_n^{*(1)} + \tilde{\boldsymbol{A}}\boldsymbol{d}_n^{(1)}$, where $\tilde{\boldsymbol{A}}\boldsymbol{d}_n^{(1)} = \boldsymbol{C}_{11,n}^{-1} n^{-1/2}\sum_{i=1}^n \hat{\epsilon}_i \boldsymbol{x}_i^{(1)}$. Now similar to (3.6.3), it can be shown that for sufficiently large n,

$$\boldsymbol{F}_n^{(1)} = n^{-1/2}\sum_{i=1}^n (\tilde{\boldsymbol{\xi}}_i^{(0)} + \tilde{\boldsymbol{\eta}}_i^{(0)})\epsilon_i + \tilde{\boldsymbol{R}}_{2n}$$

$$\tilde{\boldsymbol{F}}_n^{*(1)} = \mu_{G^*}^{-1} n^{-1/2}\sum_{i=1}^n (\tilde{\boldsymbol{\xi}}_i^{(0)}\hat{\epsilon}_i + \tilde{\boldsymbol{\eta}}_i^{(0)}\bar{\epsilon}_i)(G_i^* - \mu_{G^*}) + \tilde{\boldsymbol{R}}_{2n}^*$$

where $\mathbf{P}\Big(||\tilde{\boldsymbol{R}}_{2n}|| = o(n^{-1/2})\Big) = 1 - o(1)$ and $\mathbf{P}_*\Big(||\tilde{\boldsymbol{R}}_{2n}^*|| = o(n^{-1/2})\Big) = 1 - o_p(1)$. Here, $\tilde{\boldsymbol{\xi}}_i^{(0)} = \boldsymbol{C}_{11,n}^{-1}\boldsymbol{x}_i^{(1)}$, $\tilde{\boldsymbol{\eta}}_i^{(0)} = \boldsymbol{C}_{11,n}^{-1}\tilde{\eta}_i$ with j th component $\big[j \in \mathcal{A} = \{k : \beta_j \neq 0\}\big]$ of $\tilde{\eta}_i$ is $\Big(\dfrac{\lambda_n}{2n}\tilde{x}_{i,j}\dfrac{\gamma}{|\tilde{\beta}_{j,n}|^{\gamma+1}} sgn(\hat{\beta}_{j,n})\Big)$. Here we have assumed without loss of generality that $\mathcal{A} = \{1, \dots, p_0\}$. and $\hat{\epsilon}_i$ and $\bar{\epsilon}_i$ are respectively Alasso and OLS residuals. Then by

Berry-Essen Theorem and Lemma 3.1 of Bhattacharya and Rao (1986), we have

$$\sup_{\boldsymbol{x}\in\mathcal{R}^{p_0}}\Big|\mathbf{P}\big(\boldsymbol{F}_n^{(1)}\leq\boldsymbol{x}\big)-\boldsymbol{\Phi}_{\boldsymbol{V}_n}(\boldsymbol{x})\Big|=O(n^{-1/2})$$

$$\text{and}\quad\sup_{\boldsymbol{x}\in\mathcal{R}^{p_0}}\Big|\mathbf{P}_*\big(\tilde{\boldsymbol{F}}_n^{*(1)}+\breve{\boldsymbol{R}}_{1n}^*\leq\boldsymbol{x}\big)-\boldsymbol{\Phi}_{\tilde{\boldsymbol{V}}_n}(\boldsymbol{x})\Big|=O_p(c_n.n^{-1/2})\tag{3.6.15}$$

where $\boldsymbol{V}_n=n^{-1}\sum_{i=1}^n\big(\tilde{\boldsymbol{\xi}}_i^{(0)}+\tilde{\boldsymbol{\eta}}_i^{(0)}\big)'\big(\tilde{\boldsymbol{\xi}}_i^{(0)}+\tilde{\boldsymbol{\eta}}_i^{(0)}\big)\sigma^2$ and $\tilde{\boldsymbol{V}}_n=n^{-1}\sum_{i=1}^n\big(\tilde{\boldsymbol{\xi}}_i^{(0)}\hat{\epsilon}_i+\tilde{\boldsymbol{\eta}}_i^{(0)}\bar{\epsilon}_i\big)'$ $\big(\tilde{\boldsymbol{\xi}}_i^{(0)}\hat{\epsilon}_i+\tilde{\boldsymbol{\eta}}_i^{(0)}\bar{\epsilon}_i\big)$. Now similar to Lemma 3.6.5, it can be shown that $\Big|\Big|\tilde{\boldsymbol{V}}_n-\boldsymbol{V}_n\Big|\Big|=o_p(c_n.n^{-1/2})$ with c_n, as defined earlier. Hence by Turnbull (1930) and noting (14.66) of Lemma 14.6 of Bhattacharya and Rao (1986) and the facts that $\tilde{\boldsymbol{V}}_n=O_p(1)$ & $\boldsymbol{V}_n=O(1)$, we have

$$\sup_{\boldsymbol{x}\in\mathcal{R}^{p_0}}\Big|\boldsymbol{\Phi}_{\tilde{\boldsymbol{V}}_n}(\boldsymbol{x})-\boldsymbol{\Phi}_{\boldsymbol{V}_n}(\boldsymbol{x}))\Big|\leq\Big|\Big|\tilde{\boldsymbol{V}}_n-\boldsymbol{V}_n\Big|\Big|=o_p(c_n.n^{-1/2})\tag{3.6.16}$$

Therefore by (3.6.15) and (3.6.16) and noting that $c_n=o(\sqrt{\log n})$, we have

$$\sup_{\boldsymbol{x}\in\mathcal{R}^{p_0}}\Big|\mathbf{P}_*\big(\tilde{\boldsymbol{F}}_n^{*(1)}+\breve{\boldsymbol{R}}_{1n}^*\leq\boldsymbol{x}\big)-\mathbf{P}\big(\boldsymbol{F}_n^{(1)}\leq\boldsymbol{x}\big)\Big|=o_p(\lambda_n.n^{-1/2})\tag{3.6.17}$$

Now defining $\boldsymbol{Ad}_n^{(1)}=\boldsymbol{C}_{11,n}^{-1}\dfrac{\lambda_n}{2\sqrt{n}}\boldsymbol{s}_n^{(1)}$, by (3.6.15), (3.6.17) and Taylor expansion, we have for any $\boldsymbol{x}\in\mathcal{R}^{p_0}$,

$$\begin{aligned}\mathbf{P}_*\big(\boldsymbol{F}_n^{*(1)}\leq\boldsymbol{x}\big)&=\mathbf{P}_*\Big(\tilde{\boldsymbol{F}}_n^{*(1)}+\breve{\boldsymbol{R}}_{1n}^*+\tilde{\boldsymbol{A}}\boldsymbol{d}_n^{(1)}\leq\boldsymbol{x}\Big)\\&=\mathbf{P}\Big(\boldsymbol{F}_n^{(1)}\leq\boldsymbol{x}-\boldsymbol{Ad}_n^{(1)}+O(n^{-1/2})\Big)+o_p(\lambda_n.n^{-1/2})\\&=\boldsymbol{\Phi}_{\boldsymbol{V}_n}\Big(\boldsymbol{x}-\boldsymbol{Ad}_n^{(1)}+O(n^{-1/2})\Big)+O(n^{-1/2})+o_p(\lambda_n.n^{-1/2})\\&=\boldsymbol{\Phi}_{\boldsymbol{V}_n}(\boldsymbol{x})-\frac{\lambda_n}{2\sqrt{n}}\Big[\tilde{\boldsymbol{s}}_n^{(1)\prime}\boldsymbol{C}_{11,n}^{-1}(D_1,\ldots,D_p)'\boldsymbol{\Phi}_{\boldsymbol{V}_n}(\tilde{\boldsymbol{x}})\Big]+o_p(\lambda_n.n^{-1/2})\end{aligned}$$

$$= \mathbf{P}\Big(\boldsymbol{F}_n^{(1)} \le \boldsymbol{x}\Big) - \frac{\lambda_n}{2\sqrt{n}}\Big[\tilde{\boldsymbol{s}}_n^{(1)\prime}\boldsymbol{C}_{11,n}^{-1}(D_1,\ldots,D_p)'\Phi_{\boldsymbol{V}_n}(\tilde{\boldsymbol{x}})\Big] + o_p(\lambda_n/\sqrt{n})$$

for some $\tilde{\boldsymbol{x}}$ with $||\tilde{\boldsymbol{x}} - \boldsymbol{x}|| \le ||\boldsymbol{A}\boldsymbol{d}_n^{(1)}||$.

Therefore (3.6.14) follows from the triangle inequality and the fact that $\sup_{\boldsymbol{x}\in\mathcal{R}^{p_0}} [f(\boldsymbol{x}) + g(\boldsymbol{x})] \le \sup_{\boldsymbol{x}\in\mathcal{R}^{p_0}} f(\boldsymbol{x}) + \sup_{\boldsymbol{x}\in\mathcal{R}^{p_0}} g(\boldsymbol{x})$.

Proof of Theorem 3.5.1. By Lemma 3.6.6 we have

$$\sup_{B\in\mathcal{C}_q} \Big|\mathbf{P}_*(\boldsymbol{R}_n^* \in B) - \int_B \xi_n^*(\boldsymbol{x})d\boldsymbol{x}\Big| = o_p(n^{-1/2}). \tag{3.6.18}$$

Now, retracting the steps of Lemma 3.6.6 and using the fact that $||\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n|| = o_p(n^{-(1+\delta_1)/2})$ [cf. Lemma 3.6.5], it can be shown that

$$\sup_{B\in\mathcal{C}_q} \Big|\mathbf{P}(\boldsymbol{R}_n \in B) - \int_B \xi_n(\boldsymbol{x})d\boldsymbol{x}\Big| = o(n^{-1/2}), \tag{3.6.19}$$

where

$$\begin{aligned}\xi_n(\boldsymbol{x}) =& \phi(\boldsymbol{x})\Bigg[1 + \sum_{k=1}^{r}\frac{1}{k!}\Big\{\sum_{\boldsymbol{\alpha}=k}\tilde{\boldsymbol{b}}_n^{\alpha}H_{\boldsymbol{\alpha}}(\boldsymbol{x})\Big\} + \frac{1}{\sqrt{n}}\Bigg[-\frac{\mu_3}{2\sigma^3}\sum_{|\boldsymbol{\alpha}|=1}\boldsymbol{t}^{\boldsymbol{\alpha}}\bar{\boldsymbol{\xi}}_n(\boldsymbol{\alpha})H_{\boldsymbol{\alpha}}(\boldsymbol{x})\\ &+ \frac{\mu_3}{6\sigma^3}\Big\{\sum_{|\boldsymbol{\alpha}|=3}\boldsymbol{t}^{\boldsymbol{\alpha}}\bar{\boldsymbol{\xi}}_n(\boldsymbol{\alpha})H_{\boldsymbol{\alpha}}(\boldsymbol{x}) - 3\sum_{|\boldsymbol{\alpha}|=3}\sum_{|\boldsymbol{\zeta}|=1}\boldsymbol{t}^{\boldsymbol{\alpha}+\boldsymbol{\zeta}}\bar{\boldsymbol{\xi}}_n(\boldsymbol{\alpha})\bar{\boldsymbol{\xi}}_n(\boldsymbol{\zeta})H_{\boldsymbol{\alpha}+\boldsymbol{\zeta}}(\boldsymbol{x})\Big\}\Bigg]\Bigg],\end{aligned}$$

where $x \in \mathcal{R}^q$, $\bar{\boldsymbol{\xi}}_n(\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^n \Big(\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\xi}_i^{(0)}\Big)^{\boldsymbol{\alpha}}$. For details see the proof of Theorem 8.2 of Chatterjee and Lahiri (2013). Now due to assumption (A.6)(i), Lemma 3.6.4 and Lemma 3.6.5 and the facts that $||\boldsymbol{b}_n|| = O(n^{-\delta_1})$ and $||\check{\boldsymbol{b}}_n|| = O_p(n^{-\delta_1})$, the coefficients of $n^{-1/2}$ in $\xi_n^*(\boldsymbol{x})$ converge to those of $\xi_n(\boldsymbol{x})$ in probability and $||\tilde{\boldsymbol{b}}_n^{\boldsymbol{\alpha}} - \check{\boldsymbol{b}}_n^{\boldsymbol{\alpha}}|| = o(n^{-1/2})$, for all $\boldsymbol{\alpha}$ such that $|\boldsymbol{\alpha}| \le r_1$. Therefore Theorem 3.5.1 follows.

Proof of Theorem 3.5.2. By Lemma 3.6.6, on the set $\boldsymbol{A}_{1n}$, we have for $n > n_1$, $\hat{\mathcal{A}}_n = \mathcal{A}_n^*$ and

$$\begin{aligned}
\boldsymbol{T}_n^* + \check{\boldsymbol{b}}_n^* &= \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \Big[\mu_{G^*}^{-1} \check{\boldsymbol{W}}_n^{*(1)} - \frac{\lambda_n}{2\sqrt{n}} \tilde{\mathbf{s}}_n^{*(1)} \Big] + \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \hat{s}_n^{*(1)} \frac{\lambda_n}{2\sqrt{n}} \\
&= \mu_{G^*}^{-1} \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \check{\boldsymbol{W}}_n^{*(1)} + \frac{\lambda_n}{2\sqrt{n}} \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \big(\tilde{\mathbf{s}}_n^{*(1)} - \hat{\mathbf{s}}_n^{*(1)} \big) \\
&= \mu_{G^*}^{-1} \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \check{\boldsymbol{W}}_n^{*(1)} + Q_{4n}^*, \quad (say) \qquad (3.6.20)
\end{aligned}$$

where the jth element of $\hat{s}_n^{*(1)}$ is $\operatorname{sgn}\big(\hat{\beta}_{j,n}^*\big) |\tilde{\beta}_{j,n}^*|^{-\gamma}$. Now since $||\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n||_\infty = O_{p_*}(n^{-1/2})$ on the set $\boldsymbol{A}_{1n}$, one can conclude that on the set $\boldsymbol{A}_{1n}$, $\mathbf{P}_*\big(\tilde{\mathbf{s}}_n^{*(1)} = \hat{\mathbf{s}}_n^{*(1)}\big) = 1$ for sufficiently large n. Hence we can conclude that $\mathbf{P}_*\big(||Q_{4n}^*|| \neq 0\big) = o(n^{-1})$.

Now expansion and error bounds of the quantity $\big[\check{\sigma}_n^{*2} - \check{\sigma}_n^2\big]$, similar to (3.6.5), hold. Thus by Taylor's expansion of $\check{\sigma}_n^*$ around $\check{\sigma}_n$ and by (3.6.20), one has

$$\begin{aligned}
\check{\boldsymbol{R}}_n^* &= \mu_{G^*}^{-1} \tilde{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{D}_n^{(1)} \boldsymbol{C}_{11,n}^{-1} \check{\boldsymbol{W}}_n^{*(1)} \Big[1 - \frac{1}{2\hat{\sigma}_n^2} (\hat{\sigma}_n^* - \hat{\sigma}_n) + \frac{3}{4\hat{\sigma}_n^4} \frac{(\hat{\sigma}_n^* - \hat{\sigma}_n)^2}{2} \Big] + Q_{5,n}^* \\
&= \boldsymbol{R}_{3n}^* + Q_{5n}^*, \quad \text{(say)} \qquad (3.6.21)
\end{aligned}$$

where on the set $\boldsymbol{A}_{1n}$,

$$\mathbf{P}_*\big(||Q_{5n}^*|| = o(n^{-1})\big) = o(n^{-1}).$$

Thus by Corollary 2.6 of Bhattacharya and Rao (1986), the Edgeworth expansions of $\boldsymbol{R}_{3n}^*$ and $\check{\boldsymbol{R}}_n^*$ agree up to order $o(n^{-1})$. Now, similarly to Lemma 3.6.6, using the transformation technique of Bhattacharya and Ghosh (1978), one can obtain the three-term Edgeworth expansion of $\boldsymbol{R}_{3n}^*$, say $\pi_n^*(\boldsymbol{x})$, which will contain terms involving n^{-1} as well as $n^{-1/2}$. The coefficients in $\pi_n^*(\boldsymbol{x})$ will involve $\check{\sigma}_n^2$, μ_{G^*}, $\mathbf{E}_*(G_1^* - \mu_{G^*})^4$,

$\bar{\boldsymbol{\xi}}_n^{*(j)}(\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^n \left(\check{\boldsymbol{\xi}}_i^{(0)}\hat{\epsilon}_i^j\right)^{\boldsymbol{\alpha}}$ (for $j = 1,3$) and $\bar{\boldsymbol{\eta}}_n^{*(j)}(\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^n \left(\check{\boldsymbol{\eta}}_i^{(0)}\hat{\epsilon}_i^j\right)^{\boldsymbol{\alpha}}$ (for $j = 1,3$), where $\boldsymbol{\alpha} \in \mathcal{N}^q$ such that $|\boldsymbol{\alpha}| = 1,\ldots,4$. Similarly, one can construct a three-term Edgeworth expansion of $\check{\boldsymbol{R}}_n$, say $\pi_n(\boldsymbol{x})$, which will involve σ^2, μ_3, μ_4, $\tilde{\boldsymbol{\xi}}_n(\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^n \left(\bar{\boldsymbol{\Sigma}}_n^{-1/2}\boldsymbol{\xi}_i^{(0)}\right)^{\boldsymbol{\alpha}}$, and $\tilde{\boldsymbol{\eta}}_n(\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^n \left(\bar{\boldsymbol{\Sigma}}_n^{-1/2}\boldsymbol{\eta}_i^{(0)}\right)^{\boldsymbol{\alpha}}$, $j = 1,3$ and $\boldsymbol{\alpha} \in \mathcal{N}^q$ such that $|\boldsymbol{\alpha}| = 1,\ldots,4$, in the coefficients of $n^{-l/2}$, $l = 1,2$. It is easy to see that the coefficient of $n^{-1/2}$ in $\pi_n(\boldsymbol{x})$ and $\pi_n^*(\boldsymbol{x})$ match with that in ξ_{1n} and ξ_{1n}^* respectively, where ξ_{1n}^* is as defined in the proof of Lemma 3.6.6 and $\xi_{1n}(\boldsymbol{x}) = \xi_n(\boldsymbol{x}) - \sum_{k=1}^r \dfrac{1}{k!}\left\{\sum_{\boldsymbol{\alpha}=k}\tilde{\boldsymbol{b}}_n^{\alpha}H_{\boldsymbol{\alpha}}(\boldsymbol{x})\right\}\phi(\boldsymbol{x})$ with $\xi_n(\boldsymbol{x})$ being defined as in Theorem 3.5.1 after replacing $\boldsymbol{\Sigma}_n$ with $\bar{\boldsymbol{\Sigma}}_n$. Now due to the conditions (A.1)–(A.6) with $r = 8$, $(\check{\sigma}_n^2 - \sigma_n^2) = O_p(n^{-1/2})$ and $||\tilde{\Sigma}_n - \sigma^2\Sigma_n|| = O_p(n^{-1/2})$ [Lemma 3.6.5] and the fact that $||\tilde{\Sigma}_n^{-1/2} - \sigma^{-1}\Sigma_n^{-1/2}|| \leq K.||\tilde{\Sigma}_n - \sigma^2\Sigma_n||$ [cf. Turnbull (1930)], the coefficient of $n^{-1/2}$ in $\xi_{1n}^*(\boldsymbol{x})$ converges to that of $\xi_{1n}(\boldsymbol{x})$ [Similarly as in the proof of Theorem 3.5.1], whereas the coefficients of n^{-1} in $\pi_n^*(\boldsymbol{x})$ and $\pi_n(\boldsymbol{x})$ are bounded in respective probabilities. Therefore, theorem 3.5.2 follows.

Proof of Theorem ??. The first part follows by Lemma 3.6.6 (b) and retracing the proof of Theorem 3.5.1. And the second part follows analogously to the proof of Theorem 3.5.2.

3.7 Simulation results

We study through simulation the coverage of one-sided and two-sided 95% confidence intervals for individual nonzero regression coefficients constructed via the pivot quantities $\boldsymbol{R}_n$ and $\check{\boldsymbol{R}}_n$ as well as via their modified perturbation bootstrap versions R_n^* and $\check{\boldsymbol{R}}_n^*$. To make further comparisons, we also construct confidence intervals based on a Normal approximation to the distribution of a local quadratic approximation pivot $\boldsymbol{R}_n^{\mathrm{LQA}}$, which uses the estimator of $\mathbf{Cov}((\beta_j, j \in \hat{A}_n)')$ proposed in the original Alasso paper by Zou (2006). We also consider the confidence interval from the oracle Normal approximation, which is based on the closeness in distribution of $\boldsymbol{T}_n$ to a Normal$(0, \sigma^2 \boldsymbol{D}^{(1)} \boldsymbol{C}_{11,n}^{-1} \boldsymbol{D}^{(1)})$ random variable, where we use the true active set of covariates $\mathcal{A}_n$ to compute $\boldsymbol{C}_{11,n}^{-1}$. We denote this by $\boldsymbol{R}_n^{\mathrm{oracle}}$. For the sake of comparison, we also consider the confidence intervals based on the naive perturbation bootstrap from MTC(11) which in that paper are denoted by CN^{*Q} and CN^{*N}.

Table 3.1: Empirical coverage of 95% confidence intervals for nonzero regression coefficients by Alasso under $(n, p, p_0) = (120, 100, 4)$ using $\tilde{\lambda}_n = 0$ and crossvalidation choice of λ_n. The median λ_n choice was $0.79 \cdot n^{1/4}$. One-sided intervals are bounded in the $\mathrm{sgn}(\beta_j)$ direction.

Coverage and *(avg. width)* of two-sided 95% CIs: $(n, p, p_0) = (120, 100, 4)$

β_j	R_n^{LQA}	R_n^{oracle}	CN^{*Q}	CN^{*N}	R_n	$\check{R}_n$	R_n^*	$\check{R}_n^*$
-0.75	0.42	0.33	0.13	0.53	0.36	0.46	0.69	0.79
	(0.52)	*(0.37)*	*(0.44)*	*(0.47)*	*(1.15)*	*(0.41)*	*(0.39)*	*(0.53)*
1.50	0.53	0.49	0.21	0.76	0.52	0.61	0.90	0.92
	(0.48)	*(0.39)*	*(0.59)*	*(0.61)*	*(0.73)*	*(0.43)*	*(0.47)*	*(0.63)*
-2.25	0.74	0.70	0.43	0.86	0.72	0.78	0.89	0.93
	(0.50)	*(0.39)*	*(0.57)*	*(0.58)*	*(1.03)*	*(0.43)*	*(0.43)*	*(0.57)*
3.00	0.87	0.83	0.64	0.87	0.86	0.87	0.86	0.91
	(0.45)	*(0.37)*	*(0.49)*	*(0.49)*	*(0.50)*	*(0.41)*	*(0.37)*	*(0.48)*
Coverage of one-sided 95% CIs								
-0.75	0.33	0.25	0.10	0.41	0.27	0.37	0.71	0.83
1.50	0.45	0.42	0.16	0.66	0.44	0.53	0.91	0.95
-2.25	0.66	0.64	0.36	0.80	0.66	0.71	0.91	0.96
3.00	0.82	0.79	0.54	0.84	0.80	0.85	0.89	0.94

Under the settings

$$(n, p, p_0) \in \{(120, 100, 4), (200, 80, 4), (150, 250, 6), (150, 500, 8)\},$$

we generate n independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of $(X, Y) \in \mathbb{R}^p \times \mathbb{R}$ from the model $Y = X'\boldsymbol{\beta} + \epsilon$, where ϵ is a standard normal random variable, $X = (X_1, \ldots, X_p)'$

Table 3.2: Empirical coverage of 95% confidence intervals for nonzero regression coefficients by Alasso under $(n, p, p_0) = (200, 80, 4)$ using $\tilde{\lambda}_n = 0$ and crossvalidation choice of λ_n. The median λ_n choice was $1.40 \cdot n^{1/4}$. One-sided intervals are bounded in the $\text{sgn}(\beta_j)$ direction.

Coverage and *(avg. width)* of two-sided 95% CIs: $(n, p, p_0) = (200, 80, 4)$

β_j	R_n^{LQA}	R_n^{oracle}	CN^{*Q}	CN^{*N}	R_n	$\check{R}_n$	R_n^*	$\check{R}_n^*$
-0.75	0.39	0.35	0.12	0.67	0.36	0.45	0.88	0.91
	(0.39)	*(0.30)*	*(0.44)*	*(0.46)*	*(0.30)*	*(0.29)*	*(0.38)*	*(0.45)*
1.50	0.55	0.52	0.19	0.83	0.52	0.60	0.92	0.93
	(0.35)	*(0.31)*	*(0.52)*	*(0.53)*	*(0.31)*	*(0.30)*	*(0.39)*	*(0.47)*
-2.25	0.75	0.74	0.39	0.91	0.74	0.79	0.91	0.94
	(0.35)	*(0.31)*	*(0.47)*	*(0.47)*	*(0.31)*	*(0.30)*	*(0.34)*	*(0.39)*
3.00	0.89	0.89	0.64	0.93	0.89	0.90	0.91	0.91
	(0.32)	*(0.30)*	*(0.39)*	*(0.39)*	*(0.30)*	*(0.29)*	*(0.29)*	*(0.33)*
Coverage of one-sided 95% CIs								
-0.75	0.30	0.28	0.11	0.57	0.28	0.37	0.89	0.95
1.50	0.43	0.40	0.12	0.73	0.40	0.51	0.96	0.97
-2.25	0.65	0.64	0.31	0.82	0.64	0.71	0.94	0.95
3.00	0.82	0.81	0.53	0.86	0.81	0.84	0.92	0.95

is a mean-zero multivariate normal random vector such that

$$\text{Cov}(X_j, X_k) = \mathbf{1}(j = k) + 0.3^{|j-k|}\mathbf{1}(j \leq p_0)\mathbf{1}(k \leq p_0)\mathbf{1}(j \neq k)$$

for $1 \leq j, k \leq p$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ with β_j defined as $\beta_j = (3/4)j(-1)^j\mathbf{1}(j \leq p_0)$ for $j = 1, \ldots, p$.

We compute the empirical coverage over 500 simulated data sets of one- and two-sided confidence intervals for each nonzero regression coefficient under crossvalidation-selected values of $\tilde{\lambda}_n$ and λ_n, where $\tilde{\lambda}_n$ is the value of the tuning parameter used to obtain the preliminary Lasso estimate $\tilde{\boldsymbol{\beta}}_n$ and λ_n is the value of the tuning parameter used to obtain the Alasso estimate $\hat{\boldsymbol{\beta}}_n$. To construct the bootstrap intervals for each of the 500 simulated data sets, we generate 1000 independent samples from $Beta(1/2, 3/2)$.

Table 3.3: Empirical coverage of 95% confidence intervals for nonzero regression coefficients by Alasso under $(n, p, p_0) = (150, 250, 6)$ using crossvalidation choices of $\tilde{\lambda}_n$ and λ_n. The median $\tilde{\lambda}_n$ and λ_n choices were $0.01 \cdot n^{1/2}$ and $0.32 \cdot n^{1/4}$. One-sided intervals are bounded in the $\text{sgn}(\beta_j)$ direction.

Coverage and *(avg. width)* of two-sided 95% CIs: $(n, p, p_0) = (150, 250, 6)$

β_j	R_n^{LQA}	R_n^{oracle}	CN^{*Q}	CN^{*N}	R_n	$\check{R}_n$	R_n^*	$\check{R}_n^*$
-0.75	0.61	0.55	0.41	0.73	0.56	0.72	0.89	0.94
	(0.42)	*(0.32)*	*(0.44)*	*(0.47)*	*(0.33)*	*(0.39)*	*(0.43)*	*(0.56)*
1.50	0.72	0.69	0.52	0.86	0.70	0.85	0.93	0.95
	(0.42)	*(0.34)*	*(0.50)*	*(0.51)*	*(0.34)*	*(0.41)*	*(0.47)*	*(0.65)*
-2.25	0.90	0.88	0.72	0.91	0.88	0.92	0.92	0.95
	(0.43)	*(0.34)*	*(0.46)*	*(0.46)*	*(0.34)*	*(0.41)*	*(0.42)*	*(0.59)*
3.00	0.89	0.86	0.81	0.90	0.87	0.92	0.90	0.95
	(0.48)	*(0.34)*	*(0.42)*	*(0.43)*	*(0.34)*	*(0.41)*	*(0.39)*	*(0.53)*
-3.75	0.88	0.86	0.80	0.88	0.87	0.93	0.87	0.92
	(0.52)	*(0.34)*	*(0.40)*	*(0.41)*	*(0.34)*	*(0.41)*	*(0.37)*	*(0.50)*
4.50	0.91	0.88	0.85	0.89	0.89	0.94	0.89	0.95
	(0.49)	*(0.32)*	*(0.36)*	*(0.36)*	*(0.33)*	*(0.39)*	*(0.33)*	*(0.43)*
Coverage of one-sided 95% CIs								
-0.75	0.54	0.48	0.37	0.67	0.49	0.65	0.88	0.94
1.50	0.63	0.58	0.44	0.78	0.59	0.75	0.93	0.96
-2.25	0.85	0.82	0.65	0.88	0.83	0.89	0.94	0.95
3.00	0.85	0.83	0.74	0.87	0.84	0.89	0.91	0.94
-3.75	0.87	0.83	0.76	0.87	0.84	0.88	0.89	0.94
4.50	0.89	0.86	0.82	0.87	0.86	0.92	0.91	0.94

In the $p \leq n$ setting, we set $\tilde{\lambda}_n = 0$, whereby we use the ordinary least squares estimate for the preliminary estimator $\tilde{\boldsymbol{\beta}}_n$. When $p > n$, the value of $\tilde{\lambda}_n$ is chosen via 10-fold crossvalidation and $\tilde{\boldsymbol{\beta}}_n$ is computed under the selected value of $\tilde{\lambda}_n$. Once $\tilde{\boldsymbol{\beta}}_n$ is obtained, 10-fold crossvalidation is used to select λ_n. The values $\tilde{\lambda}_n$ and λ_n are thereafter held fixed for all bootstrap computations on the same dataset. In each crossvalidation procedure, the largest value of the tuning parameter for which the crossvalidation prediction error lies within one standard error of its minimum is used so that greater penalization is preferred. Tables 3.1 and 3.2 display the coverage results under crossvalidation selection of $\tilde{\lambda}_n$ and λ_n for the $n > p$ cases $(n, p, p_0) \in \{(120, 100, 4), (200, 80, 4)\}$ and Tables 3.3 and 3.4 for the $n \leq p$ cases $(n, p, p_0) \in \{(150, 250, 6), (150, 500, 8)\}$. The median values of the crossvalidation selections of $\tilde{\lambda}_n$ and λ_n under each setting are provided in the table captions in the forms $c_1 \cdot n^{1/2}$ and $c_2 \cdot n^{1/4}$ where c_1 and c_2 are constants. These correspond to the forms of the theoretical choices of $\tilde{\lambda}_n$ and λ_n under the choice of $\gamma = 1$.

In the $(n, p, p_0) = (120, 100, 4)$ case, for which Table 3.1 displays the results, the modified perturbation bootstrap intervals based on R_n^* and $\check{R}_n^*$ achieve the closest-to-nominal coverage, the $\check{R}_n^*$ interval performing somewhat better due to the bias correction. The two-sided $\check{R}_n^*$ interval achieves sub-nominal coverage for the smallest regression coefficient $\beta_j = -0.75$, given that this coefficient was occasionally estimated to be zero, but achieves close-to-nominal coverage for the larger regression coefficients. The coverage of the other intervals is much more dramatically effected by the magnitude of the regression coefficient β_j, a phenomenon which is even more pronounced in the one-sided coverages; for example, the coverage of the $\check{R}_n$ interval rises from 0.37 for $\beta_1 = -0.75$ to 0.85 for $\beta_4 = 3.00$. Given that the modified perturbation bootstrap distributions of R_n^* and $\check{R}_n^*$ result in much closer-to-nominal coverages than the Normal approximations to the distributions of R_n and $\check{R}_n$, we may conclude that the sample size is too small for the asymptotically-Normal pivots to have sufficiently approached their limiting distribution; the second-order correctness of the modified perturbation bootstrap is thus apparent.

For the $(n, p, p_0) = (200, 80, 4)$ case, for which Table 3.2 displays the results, we see again that all the confidence intervals besides the R_n^* and $\check{R}_n^*$ intervals achieve sub-nominal coverage. Given the larger sample size, the coverage of $\check{R}_n^*$ is close-to-nominal even for the smallest regression coefficient $\beta_1 = -0.75$. The bias correction of in the $\check{R}_n^*$ interval makes less of a difference in this case than in the $(n, p, p_0) = (120, 100, 4)$ case due to the larger sample size and smaller p. In spite of the larger sample size, however, the coverages of all the confidence intervals besides the modified perturbation bootstrap intervals are still dramatically affected by the size of the regression coefficient β_j, and none of them achieves close-to-nominal coverage for any of the nonzero regression coefficients.

In the $p > n$ settings, the modified perturbation bootstrap interval based on $\check{R}_n^*$ continues to perform very well. Under the $(n, p, p_0) = (150, 250, 6)$ setting, for which Table 3.3 shows the results, the $\check{R}_n^*$ interval achieves the nominal coverage across all regression coefficients for both two- and one-sided intervals. The confidence intervals based on the asymptotic normality of the respective pivot all have sub-nominal coverage for most of the regression coefficients, and their coverages are dramatically affected by the magnitude of the true regression coefficient.

The results are similar for the $(n, p, p_0) = (150, 500, 8)$ case, for which Table 3.4 shows the results. The only confidence interval which reliably achieves the nominal coverage is the modified perturbation bootstrap interval based on $\check{R}_n^*$. We note that the width of the $\check{R}_n^*$ interval seems to adapt more to the magnitude of the regression coefficient than the widths of the Normal-based confidence intervals, which remain fairly constant across all magnitudes of β_j, resulting in poorer coverage for smaller regression coefficients. In contrast, the $\check{R}_n^*$ interval is able to achieve nominal coverage even for the smallest values of β_j by producing suitably wider confidence intervals.

Table 3.4: Empirical coverage of 95% confidence intervals for nonzero regression coefficients by Alasso under $(n, p, p_0) = (150, 500, 8)$ using crossvalidation choices of $\tilde{\lambda}_n$ and λ_n. The median $\tilde{\lambda}_n$ and λ_n choices were $0.01 \cdot n^{1/2}$ and $0.30 \cdot n^{1/4}$. One-sided intervals are bounded in the $\text{sgn}(\beta_j)$ direction.

Coverage and *(avg. width)* of two-sided 95% CIs: $(n, p, p_0) = (150, 500, 8)$

β_j	R_n^{LQA}	R_n^{oracle}	CN^{*Q}	CN^{*N}	R_n	$\check{R}_n$	R_n^*	$\check{R}_n^*$
-0.75	0.68	0.54	0.42	0.81	0.56	0.77	0.86	0.94
	(0.53)	*(0.31)*	*(0.48)*	*(0.52)*	*(0.32)*	*(0.41)*	*(0.40)*	*(0.60)*
1.50	0.75	0.67	0.50	0.86	0.68	0.81	0.85	0.93
	(0.53)	*(0.32)*	*(0.51)*	*(0.52)*	*(0.33)*	*(0.43)*	*(0.42)*	*(0.65)*
-2.25	0.87	0.79	0.69	0.86	0.80	0.88	0.85	0.94
	(0.50)	*(0.32)*	*(0.46)*	*(0.47)*	*(0.33)*	*(0.43)*	*(0.38)*	*(0.59)*
3.00	0.89	0.82	0.76	0.86	0.83	0.92	0.86	0.94
	(0.49)	*(0.33)*	*(0.43)*	*(0.44)*	*(0.33)*	*(0.43)*	*(0.36)*	*(0.54)*
-3.75	0.89	0.83	0.79	0.86	0.84	0.92	0.87	0.94
	(0.48)	*(0.33)*	*(0.41)*	*(0.42)*	*(0.33)*	*(0.43)*	*(0.35)*	*(0.52)*
4.50	0.92	0.86	0.85	0.88	0.87	0.91	0.86	0.91
	(0.49)	*(0.33)*	*(0.40)*	*(0.41)*	*(0.34)*	*(0.44)*	*(0.34)*	*(0.51)*
-5.25	0.90	0.85	0.83	0.87	0.86	0.93	0.86	0.94
	(0.50)	*(0.32)*	*(0.39)*	*(0.39)*	*(0.33)*	*(0.43)*	*(0.34)*	*(0.50)*
6.00	0.94	0.89	0.88	0.90	0.90	0.93	0.88	0.93
	(0.48)	*(0.31)*	*(0.35)*	*(0.35)*	*(0.32)*	*(0.41)*	*(0.31)*	*(0.46)*
Coverage of one-sided 95% CIs								
-0.75	0.58	0.48	0.36	0.73	0.49	0.69	0.84	0.94
1.50	0.69	0.59	0.44	0.79	0.60	0.75	0.86	0.93
-2.25	0.81	0.73	0.61	0.80	0.74	0.83	0.85	0.93
3.00	0.86	0.79	0.70	0.83	0.80	0.86	0.85	0.94
-3.75	0.85	0.79	0.70	0.82	0.80	0.88	0.86	0.93
4.50	0.90	0.83	0.82	0.85	0.84	0.89	0.87	0.94
-5.25	0.87	0.84	0.80	0.85	0.84	0.90	0.87	0.94
6.00	0.90	0.85	0.83	0.86	0.85	0.91	0.87	0.94

We see that the modified perturbation bootstrap is able to produce reliable confidence intervals for regression coefficients in the high-dimensional setting, and that it is able to do so under data-based selection of the tuning parameters.

3.8 Data analysis

To illustrate the construction of confidence intervals for regression coefficients in the high-dimensional linear regression model using the modified perturbation bootstrap, we present an analysis of the `riboflavin` data set considered in Bühlmann et al. (2014), which those authors make publicly available in their supplementary material. The data contains $n = 71$ independent records consisting of a response variable which is the logarithm of the riboflavin production rate and of 4088 gene expression levels in batches of *Bacillis subtilis* bacteria. Of the 4088, we pre-select 200 genes by sorting them in order of decreasing empirical variance and keeping the first 200. We then fit the linear regression model to the data set with $n = 71$ and $p = 200$ and compute confidence intervals for the regression coefficients selected by the Alasso procedure. We choose $\tilde{\lambda}_n$ and λ_n using 10-fold crossvalidation. Figure 3.1 displays the confidence intervals for the Alasso-selected covariates obtained from the R_n^{LQA}, $\check{R}_n$, CN^{*N}, and $\check{R}_n^*$ pivots, where 1000 bootstrap replicates were used for the bootstrap-based intervals.

The interval based on the R_n^{LQA} pivot (straight line) and the CN^{*N} interval (jagged), are symmetric around the estimated value of the regression coefficient (the CN^{*N} interval is formed by adding and substracting an upper quantile of a Normal distribution with a bootstrap-estimated variance). The intervals based on $\check{R}_n$ are asymmetric owing to the bias correction (which is quite small in this example) and, in the case of the $\check{R}_n^*$ interval, owing to the bias correction and to the asymmetry of the bootstrap distribution of $\check{R}_n^*$. For some of the coefficients, the $\check{R}_n^*$ interval is highly asymmetric, suggesting that the distribution of the pivot $\check{R}_n$ may still be far from Normal.

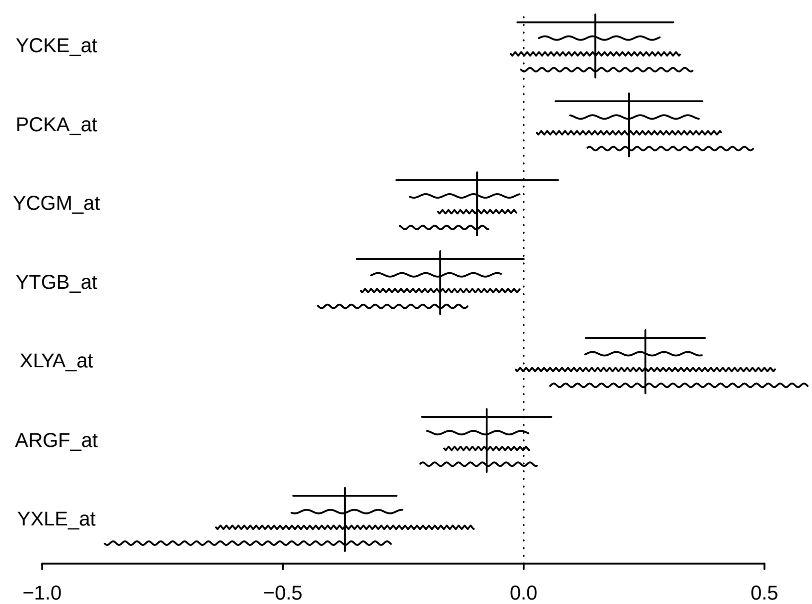

Figure 3.1: Confidence intervals based on R_n^{LQA} (straight), $\check{R}_n$ (wavy), CN^{*N} (jagged), and $\check{R}_n^*$ (wiggly) for each of the Alasso selected genes from the `riboflavin` data set.

3.9 Conclusion

Second order results of Perturbation Bootstrap method in Alasso are established. It is shown that the naive perturbation bootstrap of Minnier et al. (2011) is not sufficient for correcting the distribution of the Alasso estimator upto second order. Novel modification is proposed in bootstrap objective function to achieve second order correctness even in high dimensional setting. The modification is also shown to be computationally efficient. Thus, in a way the results in this chapter establish perturbation bootstrap method as a significant refinement of the approximation of the exact distribution of the regression M-estimator over normal approximation. This is an important finding from the perspective of valid inferences regarding the regression parameters based on Alasso estimator.

Bibliography

[1] ALLEN, M. and DATTA, S. (1999). A Note on Bootstrapping M-Estimators in ARMA Models. *J. Time Series Analysis*. **20** 365–379.

[2] ARCONES, A. M. and GINÉ, E. (1992). On the bootstrap of M-estimators and other statistical functionals. *Exploring the Limits of Bootstrap*. Edited by R. LePage and L. Billard. Wiley, New York. **13-47**.

[3] ATHREYA, K. B. and LAHIRI, S. N. (2006). *Measure theory and probability theory*. Springer Texts in Statistics.

[4] BARBE, P. and BERTAIL, P. (2012). *The weighted bootstrap*. Lecture Notes in Statistics.

[5] BERAN, R. (1986). Discussion: Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *Ann. Statist.* **14** 1295–1298.

[6] BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434-451.

[7] BHATTACHARYA, R. N. and RANGA, R. (1986). *Normal approximation and asymptotic expansions*. John Wiley & Sons.

[8] BICKEL, PETER J. and FREEDMAN, D. A. (1981b). Some Asymptotic Theory for the Bootstrap. *Ann. Statist.* **9.6** 1196–1217.

[9] BOYD, S. and LIEVEN, V. (2004). *Convex Optimization*. Cambridge University Press.

[10] BÜCHER, A. and DETTE, H. (2013). Multiplier bootstrap of tail copulas with applications. *Bernoulli*. **19** , 1655-1687.

[11] BÜCHER, A. and KOJADINOVIC, I. (2016a). Dependent multiplier bootstraps for non-degenerate U-statistics under mixing conditions with applications. *J. Statistical Planning and Inference*. **170** 83-105.

[12] BÜCHER, A. and KOJADINOVIC, I. (2016b). A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *Bernoulli*. **22** 927-968.

[13] BÜHLMANN, P., KALISCH, M. and MEIER, L. (2014). High-dimensional statistics with a view towards applications in biology. *Annual Review of Statistics and its Applications*. **1** 255–278.

[14] CAMPNOVO, L. (2015). On the validity of the pairs bootstrap for lasso estimators. *Biometrika*. **102** 981-987.

[15] CANER M. and FAN Q. (2010). “The Adaptive Lasso Method for Instrumental Variable Selection". Working Paper, North Carolina State University.

[16] CHATTERJEE, S. and BOSE, A. (2005). Generalized bootstrap for estimating equations. *Ann. Statist.* **33** 414-436.

[17] CHATTERJEE, A. and LAHIRI, S. N. (2010). Asymptotic properties of the residual bootstrap for Lasso estimators. *Proc. Amer. Math. Soc.* **138** 4497–4509.

[18] CHATTERJEE, A. and LAHIRI, S. N. (2011). Bootstrapping lasso estimators. *J. Amer. Statist. Assoc.* **106** 608-625.

[19] 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.

[20] CHENG, G. and HUANG, J. Z. (2010). Bootstrap consistency for general semi-parametric M -estimation. *Ann. Statist.* **38** 2884–2915.

[21] CHENG, G. (2015). Moment Consistency of the Exchangeably Weighted Bootstrap for Semiparametric M-estimation. *Scand J Statist.* **42** 665–684.

[22] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013) Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786-2819.

[23] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014) Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787-1818.

[24] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2016) Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Processes and their Applications*. **126** 3632–3651.

[25] DAS, D., KARL, G. and LAHIRI S. N. (2017). Perturbation Bootstrap in Adaptive Lasso. Preprint. Available at arXiv:1703.03165.

[26] DAS, D. and LAHIRI S. N. (2016). Second Order Correctness of Perturbation Bootstrap M-Estimator of Multiple Linear Regression Parameter. Preprint. Available at arXiv:1605.01440.

[27] EFRON, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist.* **7** 1-26.

[28] EFRON, B., HASTIE, T., JONSTON, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499.

[29] EL BANTLI, F. (2004). M-estimation in linear models under nonstandard conditions. *J. Statist. Plann. Inference*. **121** 231-248.

[30] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348-1360.

[31] FENG X., HE X. and HU J. (2011). Wild bootstrap for quantile regression. *Biometrika*. **98** 995–999.

[32] FREEDMAN, D. A. (1981). Bootstrapping Regression Models. *Ann. Statist.* **9** 1218–1228.

[33] FRIEDMAN, J., HASTIE, T., HÖFLING, H. and TIBSHIRANI, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.* **1** (2007) 302–332.

[34] FUK, D. H. and NAGAEV, S. V. (1971). Probabilistic inequalities for sums of independent random variables. *Teor. Verojatnost. i Primenen.* **16** 660-675.

[35] FU, W. (1998). Penalized Regressions: The Bridge versus the Lasso. *Journal of Computational and Graphical Statistics.* **7(3)** 397–416.

[36] GEYER, C. J. (1994). On the asymptotics ofconstrained M-estimation. *Ann. Statist.* **22** 1993–2010.

[37] GEYER, C. J. (1996). On the asymptotics of convex stochastic optimization. Unpublished manuscript.

[38] HALL, P. (1992). *The bootstrap and Edgeworth expansion*. Springer Series in Statistics.

[39] HALL, P., LEE, Y. K., PARK, B. U. and Paul, D. (1992). Tie-respecting bootstrap methods for estimating distributions of sets and functions of eigenvalues. *Bernoulli*. **15** 380-401.

[40] HAEUSLER, E., MASON, D. M. and NEWTON, M.A. (1991). Weighted Bootstrapping of Means. *Centrum voor Wiskunde en Informatica Quarterly*. **4** 213-228.

[41] HJORT, N.L. and POLLARD, D. (1993). Asymptotics for minimizers of convex processes, unpublished manuscript.

[42] HLAVKA, Z. (2003). Asymptotic properties of robust three-stage procedure based on bootstrap for M-estimator. *J. Statistical Planning and Inference*. **115** 637-656.

[43] HU, F. (1996). Efficiency and Robustness of a Resampling M-Estimator in the Linear Model. *J. Multivariate Analysis*. **78** 252-271.

[44] HU, F. and KALBFLElSCH D. J. (2000). The estimating function bootstrap. *The Canadian Journal of Statistics.* **28** 449-499.

[45] HUBER, P. (1981). *Robust Statistics*. Wiley, New York.

[46] JIN, Z. , YING, Z. and WEI, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika*. **88** 381-390 .

[47] KARABULUT, I.K. and LAHIRI, S.N. (1997). Two-term Edgeworth expansion for M-estimators of a linear regression parameter without Cramer-type conditions and an application to the bootstrap. *Proceedings of the Australian Mathematical Society, Ser. A*. **62** 361-370.

[48] KENGO, K. (2009). Asymptotics for argmin processes: Convexity arguments. *Journal of Multivariate Analysis*. **100** **8** 1816-1829.

[49] KLINE, P. and SANTOS A. (2012). A Score Based Approach to Wild Bootstrap Inference. Journal of Econometric Methods. **1** 23–41.

[50] KNIGHT, K. and FU, W. (2000). Asymptotics for lasso-type estimators. *Ann. Statist.* **28** 1356-1378.

[51] LAHIRI, S. N. (1989b). Bootstrap approximation and Edgeworth Expansion for the distributions of the M-estimators of a regression parameter. Preprint 89-36, Dept. Statistics, Iowa State Univ.

[52] LAHIRI, S. N. (1992). Bootstrapping M-estimators of a multiple linear regression parameter. *Ann. Statist.* **20** 1548-1570.

[53] LAHIRI, S. N. (1994). On two-term Edgeworth expansions and bootstrap approximations for Studentized multivariate M-estimators. *Sankhya A*. **56** 201-226.

[54] LAHIRI, S. N. (1996). On Edgeworth Expansion and Moving Block Bootstrap for Studentized M-Estimators in Multiple Linear Regression Models. *J. Multivariate Analysis*. **56** 42-59.

[55] LAHIRI, S. N., and ZHU, J. (2006). Resampling methods for spatial regression models under a class of stochastic designs. *Ann. Statist.* **34** 1774-1813.

[56] LEE, STEPHEN M.S. (2012). General M-estimation and its bootstrap. *J. Korean Statistical Society*. **41** 471-490.

[57] LIU, R. Y. (1988). Bootstrap Procedures under some Non-IID Models. . *Ann. Statist.* **16** 1696–1708.

[58] MA, S. and KOSOROK, M. R (2004). Robust semiparametric M-estimation and the weighted bootstrap. *J. Multivariate Analysis*. **96** 190-217.

[59] MAMMEN, E. (1993). Bootstrap and Wild Bootstrap for High Dimensional Linear Models. *Ann. Statist*. **21** 255–285.

[60] MASON, D. M. and NEWTON, M. A. (1992). A Rank Statistics Approach to the Consistency of a General Bootstrap. *Ann. Statist.* **20** 1611–1624.

[61] MINNIER, J., TIAN, L. and CAI, T (2011). A perturbation method for inference on regularized regression estimates. *J. Amer. Statist. Assoc.* **106** 1371-1382.

[62] NAVIDI, W. (1989). Edgeworth Expansions for Bootstrapping Regression Models. *Ann. Statist*. **17** 1472–1478.

[63] OSBORNE, M., PRESNELL, B. and TURLACH, B. (2000). On the LASSO and Its Dual. *Journal of Computational and Graphical Statistics*. **9(2)** 319-337.

[64] POLLARD, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory.* **7** 186âĂŞ199.

[65] QUMSIYEH, M. B. (1990a). Edgeworth expansion in regression models. *J. Multivariate Analysis*. **35** 86-101.

[66] QUMSIYEH, M. B. (1994). Bootstrapping and empirical edgeworth expansions in multiple linear regression models. *Comm. Statist. Theory Methods*. **23** 3227–3239.

[67] RAO, C. and ZHAO, L. (1992). Approximation to the Distribution of M-Estimates in Linear Models by Randomly Weighted Bootstrap. *Sankhya A*. **54** 323-331.

[68] RUBIN, D. B. (1981). The Bayesian Bootstrap. *Ann. Statist.* **9** 130–134.

[69] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc.* **58** 267-288.

[70] TURNBULL, H. W. (1930). A Matrix Form of Taylor's Theorem. *Proc. Edinburgh Math. Soc.* **33** 33-54.

[71] VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEUR, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166-1202.

[72] WAGENER, J. and DETTE, H. (2012). Bridge estimators and the adaptive Lasso under heteroscedasticity. *Math. Methods Statist.* **21** 109âĂŞ126.

[73] WAINWRIGHT, M. J. (2009). Sharp Thresholds for High-Dimensional and Noisy Sparsity Recovery Using ell_1 -Constrained Quadratic Programming (Lasso). *IEEE Transactions on Information Theory.* **55** 2183-2202.

[74] WANG, X. and SONG, L. (2011). Adaptive lasso variable selection for the accelerated failure models. *Comm. Statist. Theory Methods*. **40** 4372-4386.

[75] WANG, X. M. and ZHOU, W. (2004). Bootstrap Approximation to the Distribution of M-estimates in a Linear Model. *Acta Math Sinica*. **20** 9–104.

[76] WELLNER, J. A. and ZHAN, Y. (1996), "Bootstrapping Z-Estimators". Technical report, University of Washington, Dept. of Statistics.

[77] WU, C. F. J. (1986). Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *Ann. Statist.* **14** 1261–1295.

[78] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541âĂŞ2563.

[79] ZOU. H. (2006). The Adaptive Lasso and Its Oracle Properties. *J. Amer. Statist. Assoc.* **101** 1418-1429.

[80] ZHOU, Q. M., SONG, P. X.-K. and THOMPSON, M. E. (2006). Information ratio test for model misspecification in quasi-likelihood inference. *J. Amer. Statist. Assoc.* **107** 205-213.