WEAK SIGNAL IDENTIFICATION AND INFERENCE IN PENALIZED MODEL SELECTION

BY

PEIBEI SHI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Statistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Professor Annie Qu, Chair
Professor Douglas Simpson
Associate Professor Xiaofeng Shao
Assistant Professor Xiaohui Chen

Abstract

Weak signal identification and inference are very important in the area of penalized model selection, yet they are under-developed and not well-studied. Existing inference procedures for penalized estimators are mainly focused on strong signals. This thesis propose an identification procedure for weak signals in finite samples, and provide a transition phase in-between noise and strong signal strengths. A new two-step inferential method is introduced to construct better confidence intervals for the identified weak signals. Both theory and numerical studies indicate that the proposed method leads to better confidence coverage for weak signals, compared with those using asymptotic inference. In addition, the proposed method outperforms the perturbation and bootstrap resampling approaches. The method is illustrated for HIV antiretroviral drug susceptibility data to identify genetic mutations associated with HIV drug resistance.

We also provide signal's inference method based on the exact distribution of penalized estimator. The finite sample distribution is quite different from its asymptotic counterpart, which can be highly non-normal with a point mass at zero. Numerical studies indicate that the density-based approach works well when true parameter is moderately large. However, it cannot provide accurate inference when signal is weak.

To My Beloved Family.

Acknowledgments

I would never have been able to finish my dissertation without the support of many people. First of all, I would like to show my greatest gratitude to my advisor Professor Annie Qu. I am thankful for her patience, motivation, enthusiasm, and immense knowledge over the past few years. I benefit a lot from her patient advising, frequent meetings and enlightening discussions. She was always patient to help preparing my future career by teaching me how to write scientific papers, and by mentoring me how to present my works. Also, her positive attitude and determination encouraged me whenever I had a hard time for my research.

Special thanks to my committee members: Professors Xiaohui Chen, Xiaofeng Shao and Douglas Simpson, who have generously given their time and expertise to better my work. I would like to express my sincere appreciation to them for their insightful suggestions and constant support. Besides, Professors Georgios Fellouris has generously offered his kind suggestions and helpful discussions on my work. Many thanks also goes to my internship advisor Dr. Colin Wu at NHLBI, his dedication to the biostatistics research has always been an inspiration for me.

I would also like to thank all the faculties, staff and aluminies in the Statistics Department. My thanks must also go to Professor Jeffrey Douglas, Professor John Marden and Professor Adam Martinsek. I also thank our wonderful staff Melissa Banks for always being so helpful and friendly. Special thanks to Dr. Peng Wang, Dr. Yinxiao Huang and Dr. Yunwen Yang, for their kind help on my research and career development.

I am also very grateful to the fellow students in the Statistics Department, who have shared their memories and experiences. Special thanks to Shiyu Wang, Xiwei Tang, Xuan Bi, Jin Wang, Jianjun Hu, Srijan Sengupta, Dan Sewell, Xiaolu Zhu, Xueying Zheng, Hyun Keun Cho, Xinxin Shu, Bin Li, Xianyang Zhang, Na Cui, Yeonjoo Park, Yifeng He, Robert Eisinger and Christopher Kinson, for their support and friendship.

I owe my greatest gratitude to Richard E. Smith and Trevanion E. Smith. They always provided generous and unconditional support for me since I was in high school, especially during those difficult times. I think of them as my grandparents in the United States. I also owe special thanks to my best friends Jiaying Gu and Qiaomin Xie at UIUC. They were always supportive both for my researches and personal life. I also want to thank many of my other friends whom I met in Champaign and who shared a memorable life with me, including Yunwen Xu, Fanglin Lu, Chang Bao, Bo Liu, Siting Chang, Tianjian Lu, Xin Zhang, Xiaowan Zhang and Bin Lian. Finally and most importantly, I would like to thank my parents for their unconditional love and care during my long years of education. You are always my greatest inspiration.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction to penalized model selection

Penalized model selection methods are developed to select variables and estimate coefficients simultaneously, which is extremely useful in variable selection when the dimension of predictors is very large. It has many applications in statistical modeling such as biomedical studies, public health, econometric study, social and environmental sciences. For example, the nationwide Framingham heart study and MESA study are interested in identifying risk factors which are associated with cardiovascular disease including heart attack and stroke. The penalized model selection is able to identify a subgroup of risk factors which are associated with specific diseases, among hundreds of candidate variables. The identified risk factors are incorporated to build more efficient models to predict the outcome of the disease. This is very important and beneficial for patients in clinical studies.

One of the main purpose of variable selection is to avoid model over-fitting as the noise variables increase the dimension of parameter space, and are likely to lead less precision of parameter estimations. Some most well-known penalized model selection methods include LASSO (Tibshirani, 1996), SCAD (Fan and Li, 2001), adaptive LASSO (Zou, 2006), MCP (Zhang, 2010), and truncated-L_1 penalty method (Shen et al., 2012).

The basic framework for regularized model selection can be formulated as follows under a linear regression model framework:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y} = (y_1, \cdots, y_n)^T$ is a response variable, $\boldsymbol{X} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_p)$ is a $n \times p$ design matrix with $p < n$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_p)^T$, and $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I_n})$. Throughout the entire chapter we assume that all covariates are standardized with $\boldsymbol{X_j^T X_j} = n$ for $j = 1, \cdots, p$.

The penalized least square estimator is the minimizer of the penalized least square function:

$$L(\boldsymbol{\theta}) = \frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|^2 + \sum_{j=1}^{p} p_\lambda(|\theta_j|), \tag{1.1}$$

where $\|\cdot\|$ is the Euclidean norm and $p_\lambda(\cdot)$ is a penalty function controlled by a tuning parameter λ. For example, the adaptive LASSO penalty proposed by Zou (2006) has the following form:

$$p_{ALASSO,\lambda}(\theta) = \lambda\frac{|\theta|}{|\hat{\theta}^{LS}|},$$

where θ is a component of $\boldsymbol{\theta}$, and $\hat{\theta}^{LS}$ is the least-square estimator of θ. Similarly, the SCAD penalty proposed by Fan and Li (2001) has the following form:

$$p_{SCAD,\lambda}(\theta) = \begin{cases} \lambda|\theta|, & \text{when } |\theta| \leq \lambda; \\ -\frac{|\theta|^2 - 2a\lambda|\theta| + \lambda^2}{2(a-1)}, & \text{when } \lambda < |\theta| \leq a\lambda; \\ \frac{(a+1)\lambda^2}{2}, & \text{when } |\theta| > a\lambda, \end{cases}$$

where a usually takes the value of 3.7.

The penalized least square estimator $\widehat{\boldsymbol{\theta}}$ is obtained by minimizing (1.1) given a λ, where the best λ can be selected through k-fold cross validation, generalized cross-validation (GCV) (Fan and Li 2001) or the Bayesian information criterion (BIC) (Wang et al., 2007).

A desirable penalized estimators have the following properties: unbiasedness, sparsity and the oracle property. Unbiasedness indicates that the model parameters can be estimated nearly unbiasedly, especially when the true coefficient is sufficiently large. Sparsity refers that the model selection procedure can automatically set noise coefficients to be zero, and reduces model complexity. The oracle property consists of two parts: model selection consistency and asymptotic normality (Fan and Li, 2001; Huang and Xie, 2007). Model selection consistency implies that the selection procedure can successfully distinguish useful features from noises, where the true zero coefficients are estimated to zero with probability approaching 1 when the sample size is large. The asymptotic normality indicates that the estimators of true non-zero coefficients are normally distributed asymptotically. That is, the oracle estimators have the same asymptotic distribution as they would have, if the zero coefficients are eliminated in advance.

Under the orthogonal designed matrix $\boldsymbol{X}$, the adaptive LASSO estimator has an explicit expression:

$$\hat{\theta}_{ALASSO} = \left(|\hat{\theta}^{LS}| - \frac{\lambda}{|\hat{\theta}^{LS}|} \right)_+ \text{sgn}(\hat{\theta}^{LS}), \tag{1.2}$$

where $_+$ takes a positive value of a function. The SCAD estimator has a slightly more complicated explicit expression:

$$\hat{\theta}_{SCAD} = \begin{cases} (|\hat{\theta}^{LS}| - \lambda)_+ sgn(\hat{\theta}^{LS}) & |\hat{\theta}^{LS}| \leq 2\lambda \\ \frac{(a-1)\hat{\theta}^{LS} - a\lambda \text{sgn}(\hat{\theta}^{LS})}{a-2} & 2\lambda \leq |\hat{\theta}^{LS}| \leq a\lambda \\ \hat{\theta}^{LS} & |\hat{\theta}^{LS}| \geq a\lambda. \end{cases} \tag{1.3}$$

In this thesis, we mainly focus on the properties of the adaptive LASSO estimator, although our method can be extended to other penalized estimators.

Define the active and null sets as $\mathcal{A} = \{j : \theta_j \neq 0\}$, $\mathcal{A}^c = \{j : \theta_j = 0\}$, and the estimated active and null sets as $\mathcal{A}_n = \{j : \hat{\theta}_j \neq 0\}$, $\mathcal{A}_n^c = \{j : \hat{\theta}_j = 0\}$. The regularized estimates $\widehat{\boldsymbol{\theta}}$ enjoys oracle properties if we choose tuning parameter λ under certain conditions. That is, if the tuning parameter λ_n satisfies conditions of $\sqrt{n}\lambda_n \to 0, n\lambda_n \to \infty$, then the adaptive LASSO estimator has oracle properties such that $\mathcal{A}_n = \mathcal{A}$ with probability tending to 1 as n goes to infinity. This indicates that the adaptive LASSO is able to successfully classify model parameters into two groups, $\mathcal{A}$ and $\mathcal{A}^c$, if the sample size is large enough.

In addition, the adaptive LASSO has the following asymptotic normality:

$$\sqrt{n}(I(\boldsymbol{\theta}_{\mathcal{A}}) + \boldsymbol{\Omega}) \left\{ \hat{\boldsymbol{\theta}}_{\mathcal{A}} - \boldsymbol{\theta}_{\mathcal{A}} + (I(\boldsymbol{\theta}_{\mathcal{A}}) + \boldsymbol{\Omega})^{-1} \boldsymbol{b} \right\} \to N\left\{ \boldsymbol{0}, I(\boldsymbol{\theta}_{\mathcal{A}}) \right\}.$$

where $\boldsymbol{\Omega} = \text{diag}\left\{ p_\lambda^{''}(|\theta_1|), \cdots, p_\lambda^{''}(|\theta_q|) \right\}$, $\boldsymbol{b} = (p_\lambda^{'}(|\theta_1|)sgn(\theta_1), \cdots, p_\lambda^{'}(|\theta_1|)sgn(\theta_q))^T$ and the term $\boldsymbol{\Omega}$ converges to 0 if the sample size goes to infinity.

Specifically, the estimated bias and covariance to nonzero components of $\widehat{\boldsymbol{\theta}}_{\mathcal{A}_n}$ are given by:

$$\hat{\boldsymbol{b}}(\widehat{\boldsymbol{\theta}}_{\mathcal{A}_n}) = (\frac{1}{n} \boldsymbol{X}_{\mathcal{A}_n}^T \boldsymbol{X}_{\mathcal{A}_n} + \boldsymbol{\Omega})^{-1} (p_\lambda^{'}(|\hat{\theta}_1|)sgn(|\hat{\theta}_1|), \cdots, p_\lambda^{'}(|\hat{\theta}_q|)sgn(|\hat{\theta}_q|))^T, \tag{1.4}$$

and

$$\widehat{Cov}(\widehat{\boldsymbol{\theta}}_{\mathcal{A}_n}) = \left\{ \boldsymbol{X}_{\mathcal{A}_n}^T \boldsymbol{X}_{\mathcal{A}_n} + n\boldsymbol{\Omega} \right\}^{-1} \boldsymbol{X}_{\mathcal{A}_n}^T \boldsymbol{X}_{\mathcal{A}_n} \left\{ \boldsymbol{X}_{\mathcal{A}_n}^T \boldsymbol{X}_{\mathcal{A}_n} + n\boldsymbol{\Omega} \right\}^{-1} \widehat{\sigma}^2, \tag{1.5}$$

where the covariance term has a sandwich form. The bias formula and the sandwich formula are effective to compute the bias and covariance for all nonzero components of penalized estimator $\widehat{\boldsymbol{\theta}}$, especially for large coefficients.

1.2 Review on signal's inference in penalized model selection

In this section, we review several approaches of parameter inference under the penalized model selection framework.

Section 1.1 indicates that we can construct a bias-corrected confidence interval based on (1.4) and (1.5). The finite sample performance of the asymptotic-based inference method relies on the magnitude of coefficient. It performs well only when the model coefficient are large since formula (1.5) might underestimate the estimator's standard deviation when the coefficient is small. Therefore, there exists a gap between finite sample behavior and the asymptotic properties in section 1.1.

We use the following simulation study to illustrate this point. We generate 1000 dataset from a model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \boldsymbol{\theta} = (1, 0.2, 0, 0, 0, 0)^T$, where $n = 100, \sigma = 1$, $\boldsymbol{X}_i \sim N(0, 1)$ for $i = 1, \cdots, 6$, and the covariates are orthogonal, i.e., $cor(\boldsymbol{X}_i, \boldsymbol{X}_j) = 0$. For each simulation, the adaptive LASSO estimator is used to select variables and estimate parameters, and the tuning parameter λ is selected based on the BIC. The simulation results are summarized in Table 1.1. When the coefficient is sufficiently large such as $\theta_1 = 1$, the finite sample behavior of the adaptive LASSO estimator is consistent with the estimator based on the asymptotic property. This is reflected by that the detection power of selecting a true signal is 100%, the standard error calculated based on sandwich formula (0.094) is quite close to the empirical standard error (0.103), and the coverage rate based on the asymptotic confidence interval is also close to 95%. However, these desirable features no longer hold for a small coefficient such as $\theta_2 = 0.2$. Specifically, the adaptive LASSO shrinks θ_2 to be 0 with 61% of time. In addition, the parameter estimation for θ_2 is biased and the standard error calculated by the asymptotic formula is underestimated. This motivating example indicates that the finite sample behavior of the penalized estimator cannot be captured by the asymptotic theory fully, especially when the magnitude of model parameter is small.

Table 1.1: The adaptive LASSO estimator, standard errors (SE), true positive rate and confidence interval coverage rate(CR)

	$\hat{\theta}$	$SE_{True}(\hat{\theta})$	$SE_{Asym}(\hat{\theta})$	CR (95% CI_{Asym})	TPR
$\theta_1 = 1$	1.000	0.113	0.094	94.0%	1
$\theta_2 = 0.2$	0.161	0.103	0.045	73.5%	0.389

SE_{Asym} and CR are calculated based on averaging over a subset when the coefficient is selected

Another common approach for parameter inference is the bootstrap method. The regular bootstrap quantile confidence interval with $1-\alpha$ coverage is:

$$\left[\hat{\theta}^*_{\alpha/2}, \hat{\theta}^*_{1-\alpha/2}\right],$$

and the standard $1-\alpha$-level confidence interval is:

$$\left[\hat{\theta} - z_{\alpha/2}\hat{\sigma}^*, \hat{\theta} + z_{\alpha/2}\hat{\sigma}^*\right],$$

where $\hat{\theta}^*_{\alpha/2}$, $\hat{\theta}^*_{1-\alpha/2}$ and $\hat{\sigma}^*$ are the $\alpha/2$ quantile, $1-\alpha/2$ quantile of the bootstrap estimators and standard deviation based on bootstrap resampling, respectively. However, the penalized model introduces a shrinkage effect and leads to a bias estimator of θ. Therefore, these two types of confidence intervals can be unreliable, especially when the coefficient is small.

The m out of n subsampling bootstrap method has been proposed as a remedy when standard bootstrap fails (Politis and Romero, 1994). It applies bootstrap samples with size m, where $m \to \infty$ and $m/n \to 0$. Bickel et al. (1997) show that m out of n bootstrap method can perform well when standard bootstrap methods fail. Another advantage of the subsampling technique is that it reduces computational cost, especially when the sample size is large. Recently, Efron (2013) proposes a smooth bootstrap procedure which takes model selection into account. The new bootstrap confidence interval is centered at the smoothing average based on the resampling with a new standard error calculation for the smooth estimator. The confidence interval constructed based on the smoothed estimator has better asymptotic properties.

A more recent work by Minnier et al. (2012) introduces a perturbation resampling method which draws inference for the regularized estimators. In each resampling, the perturbation estimator is a minimizer of a perturbed version of objective function. More specifically, suppose $\mathcal{G} = \{G_1, \cdots, G_n\}$ is a set of i.i.d positive random variables, where G_i satisfies $E(G_i) = 1, Var(G_i) = 1$ (e.g.: $G_i \sim Exp(1)$), then the perturbed objective function equals

$$L^*(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \boldsymbol{x_i}\boldsymbol{\theta})^2 G_i + \sum_{j=1}^{p} p'_\lambda(|\hat{\theta}_j^{*,LS}|)|\theta_j|, \tag{1.6}$$

where $\boldsymbol{x_i}$ is the ith row of $\boldsymbol{X}$, $\hat{\theta}_j^{*,LS}$ is the jth component of $\widehat{\boldsymbol{\theta}}^{*,LS}$, and $\widehat{\boldsymbol{\theta}}^{*,LS}$ is obtained based on

$$\widehat{\boldsymbol{\theta}}^{*,LS} = \text{argmin}_{\boldsymbol{\theta}}\left\{\frac{1}{2n}\sum_{i=1}^{n}(y_i - \boldsymbol{x_i}\boldsymbol{\theta})^2 G_i\right\}.$$

The perturbation estimator is the minimizer of (1.6), and is denoted as $\widetilde{\boldsymbol{\theta}}^*$. Minnier et al. (2012) show that

$P(\widetilde{\boldsymbol{\theta}}^*_{\mathcal{A}^c} = 0) \to 1$, and the limiting distribution of $\sqrt{n}(\widetilde{\boldsymbol{\theta}}^*_{\mathcal{A}} - \widehat{\boldsymbol{\theta}}_{\mathcal{A}})$ is the same as $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\mathcal{A}} - \boldsymbol{\theta}_{\mathcal{A}})$ given the data. Here $\mathcal{A}$ is the active set of non-zero coefficients of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$ is the minimizer of (1.1), and $\widehat{\boldsymbol{\theta}}_{\mathcal{A}}$ is the penalized estimator for non-zero coefficients. Based on the fact that $\widehat{\boldsymbol{\theta}}_{\mathcal{A}}$ follows a normal distribution, and the non-active set estimator tends to zero with the probability approaching to 1. Combining both, $\sqrt{n}(\widetilde{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}})$ and $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ have the same limiting distribution. This provides the construction of confidence regions for $\boldsymbol{\theta}$. In addition, Minnier et al. (2012) show that the perturbation method can provide more accurate confidence region, compared with the inference based on the asymptotic distribution. In addition, the perturbation method performs well even when the covariates are moderately correlated.

Recently, Potscher and Leeb (2009), Potscher and Schneider (2009) and Potscher and Schneider (2010) develop an alternative method through density estimation of the penalized estimator for confidence intervals construction when covariates are orthogonal-designed. They show that the distributions of the LASSO-type estimators, such as LASSO, adaptive LASSO and SCAD estimator, are highly non-normal under finite samples, and estimating their distributions becomes nonfeasible when the true coefficient is in magnitude of $n^{-1/2}$. Their theory development for the distribution of the penalized estimator relies on the true parameter, and therefore is difficult to apply in practice since true information is unknown.

1.3 Review on weak signal in penalized model selection

In general, identification and inference for weak signal coefficients play an important role in scientific discovery. A extreme argument is that all useful signals are weak (Donoho and Jin, 2008), where each individual weak signal might not contribute significantly to a model's prediction, but the weak signals combined together could have significant influence to predict a model. In addition, even though some variables do not have strong signal strength, they might still need to be included in the model by design or by scientific importance.

An underlying sufficient condition for successful model selection is that all nonzero parameters should be greater than a uniform signal strength, which is proportional to $\sigma/\sqrt{n}$ (Fan and Li, 2001). In other words, signal strength within a noise level $C\sigma/\sqrt{n}$ should not be detected through a regularized procedure. However, due to an uncertain scale for the constant C, the absolute boundary between noise and signal level is unclear. Therefore, it is important to define a more informative signal magnitude which is applicable in finite samples.

Established asymptotic theory mainly targets strong-signal coefficient estimators. When signal strength is weak, existing penalized methods are more likely to shrink the coefficient estimator to be 0, and no inference can be provided based on asymptotics. Moreover, even if a weak signal is selected in the model selection procedure, inference of weak-signal parameters in finite samples is not valid based on the asymptotic theory. For instance, the diagonal term of (1.5) underestimates the estimator's true standard error, which leads to an under-coverage confidence interval for small coefficient. For finite samples, the inference of the weak signals is still lacking in the current literature.

Standard bootstrap methods are not applicable when the parameter is close to zero (Knight and Fu, 2000). Andrews (2000) shows that standard bootstrap is inconsistency in some situations, especially when the true parameter is close to zero. When the true parameter is zero, the bootstrap is not asymptotically correct to the first order and it provides over-coverage confidence interval. More generally, the standard bootstrap method is not applicable to boundary problems, and the boundary is in order of $n^{-1/2}$ in some standard hypothesis testing problem. Even if m out of n bootstrap provides better inference compared with its standard counterpart, its performance remains poor for small model parameter.

Studies on weak signal identification and inference are quite limited. Among these few studies, Jin et al. (2014) propose a two-step graphlet screening method in high-dimensional variable selection, where all the useful features are assumed to be rare and weak. Their variable selection criterion is measured by Hamming distance, which is quite different from the measurement criterion for strong signal detection in a regular model selection framework. Therefore, whether the selection procedure possesses the oracle property or not is not a concern in their study. In addition, their work mainly focuses on signal detection, not on parameter inference.

Zhang and Zhang (2014) develop a projection approach which projects a high-dimensional model to a low-dimensional problem and construct confidence intervals for regression coefficients. The low dimensional projection approach is quite different from the existing variable selection, thus their inference method is not applicable for the penalized estimator. The recent perturbation resampling method (Minnier, Tian and Cai, 2012) is proposed to draw inference for regularized estimators. However, their approach is more suitable for relatively strong signal rather than weak signal inference. In addition, the perturbation method tends to produce an under-coverage confidence interval when a coefficient is small.

Chapter 2

Weak Signal Identification and Inference in Penalized Model Selection

In this chapter, we propose an identification procedure for weak signals, and provide a weak signal interval in-between noise and strong signal strengths, where the weak signal range is defined based on the signal's detectability under the penalized model selection framework. In addition, we propose a new two-step inferential method to construct better inference for the weak signals. In theory, we investigate finite sample behavior for weak signal inference, and show that our two-step procedure guarantees that the confidence interval reaches an accurate coverage rate under regularity conditions. Our numerical studies also confirm that the proposed method leads to better confidence coverage for weak signals, compared to existing methods based on asymptotic inference, perturbation methods and bootstrap resampling approaches (Efron and Tibshirani, 1993; Efron, 2013). We mainly focus on the adaptive LASSO estimator as an illustration for penalized estimators. Our method, however, is also applicable for other appropriate penalized estimators.

The chapter is organized as follows. In Section 2.1.1, we propose weak signal definition and identification. The two-step inference procedure and its theoretical property for finite samples are illustrated in Section 2.1.3. In Section 2.3, we evaluate finite sample performance of the proposed method and compare it to other available approaches, and apply these methods for an HIV drug resistance data example. The last section provides a brief summary and discussion.

2.1 Methodology

2.1.1 Weak signal definition

The uniform signal strength required for good asymptotic property does not provide much information in finite sample practice. Therefore, it is more important to define a more informative signal magnitude which is applicable in finite samples. This motivates us to define a transition phase in-between noise level and strong-signal level. In the following, we propose three phases corresponding to noise, weak signal and strong signal, where three different levels are defined based on low, moderate and high detectability of signals, respectively.

Suppose a model contains both strong and weak signals. Without loss of generality, the parameter vector $\boldsymbol{\theta}$ consists of three components: $\boldsymbol{\theta} = (\boldsymbol{\Theta^{(S)}}, \boldsymbol{\Theta^{(W)}}, \boldsymbol{\Theta^{(N)}})^T$, where $\boldsymbol{\Theta^{(S)}}, \boldsymbol{\Theta^{(W)}}$ and $\boldsymbol{\Theta^{(N)}}$ represent strong-signal, weak-signal and noise coefficients. We introduce a degree of detectability to measure different signal strength levels as follows.

For any given penalized model selection method, we define P_d as a probability of selecting an individual variable. For example, for the LASSO approach in (1.2), P_d has an explicit form of θ function given n, σ and λ:

$$P_d(\theta) = P(\hat{\theta}_{ALASSO} \neq 0|\theta) = \Phi(\frac{\theta - \sqrt{\lambda}}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta - \sqrt{\lambda}}{\sigma/\sqrt{n}}). \tag{2.1}$$

Clearly, $P_d(\theta)$ is a symmetric function, and $P_d(\theta) \to 0$ for $\theta = 0$, $P_d(\theta) \to 1$ for any $\theta \neq 0$, as $n \to \infty$. For finite samples, $P_d(\theta)$ is an increasing function of $|\theta|$, and measures the detectability of a signal coefficient, which serves as a good indicator of signal strength such that a stronger signal leads to a larger P_d and vice versa.

In the following, we define a strong signal if P_d is close to 1, a noise variable if P_d is close to 0, and a weak signal if a signal strength is in-between strong and noise levels. Specifically, suppose there are two threshold probabilities, γ^s and γ^w derived from P_d, the three signal-strength levels are defined as:

$$\begin{cases} \theta \in \boldsymbol{\Theta^{(S)}} & \text{if } P_d > \gamma^s \\ \theta \in \boldsymbol{\Theta^{(W)}} & \text{if } \gamma^w < P_d \leq \gamma^s \\ \theta \in \boldsymbol{\Theta^{(N)}} & \text{if } P_d \leq \gamma^w, \end{cases} \tag{2.2}$$

where $\tau_0 \preceq \gamma^w < \gamma^s \preceq 1$, and $\tau_0 = \min_\theta P_d(\theta) = 2\Phi(-\frac{\sqrt{n\lambda}}{\sigma})$ can be viewed as a false-positive rate of model selection. Theoretically $\tau_0 \to 0$ when $n \to \infty$ for consistent model selection. In finite samples, τ_0 does not need to be 0, but close to 0.

To see the connection between signal detectability P_d and signal strength, we let ν^γ be a positive solution of $P_d = \gamma$ in (2.1):

$$\gamma = \Phi(\frac{\nu^\gamma - \sqrt{\lambda}}{\sigma/\sqrt{n}}) + \Phi(\frac{-\nu^\gamma - \sqrt{\lambda}}{\sigma/\sqrt{n}}). \tag{2.3}$$

It can be shown that ν^γ is an increasing function of γ. In addition, if the two positive threshold values ν^s and ν^w are solutions of equation (2.3) corresponding to $\gamma = \gamma^s$ and γ^w, then the definition in (2.2) is equivalent to

$$\begin{cases} \theta \in \mathbf{\Theta^{(S)}} & \text{if } |\theta| > \nu^s \\ \theta \in \mathbf{\Theta^{(W)}} & \text{if } \nu^w < |\theta| \le \nu^s \\ \theta \in \mathbf{\Theta^{(N)}} & \text{if } |\theta| \le \nu^w. \end{cases} \tag{2.4}$$

Figure 2.1 also illustrates a connection between definition (2.2) and definition (2.4).

The following lemma provides selections of γ^s and γ^w, which is useful to differentiate weak signals from noise variables. Lemma 1 also infers the order of weak signals, given both γ^s and γ^w are bounded away from 0 and 1.

Lemma 2.1.1 *(Selection of γ^s and γ^w) If assumptions of $\sqrt{n}\lambda_n \to 0, n\lambda_n \to \infty$ are satisfied, and if the threshold values of detectability γ^w and γ^s corresponding to the lower bounds of weak and strong signals satisfy:*

$$\max\left\{\epsilon, 2\Phi(-\frac{\sqrt{n\lambda_n}}{\sigma})\right\} < \gamma^w < \gamma^s < 1-\epsilon,$$

where ϵ is a small positive value; then for any γ in the weak signal range (γ^w, γ^s), we have $\nu^\gamma/\sqrt{\lambda_n} \to 1$.

Although Lemma 2.1.1 implies that ν within the weak signal range converges to zero asymptotically, the weak signal and noise variables have different orders. Specifically, Lemma 1 indicates that if the regularity condition $n\lambda_n \to \infty$ is satisfied, then a weak signal goes to zero more slowly than a noise variable. This is due to the fact that the weak signal has the same order as $\sqrt{\lambda_n}$, which goes to zero more slowly than the order of noise level $n^{-1/2}$. To simplify notation, the tuning parameter λ_n is denoted as λ throughout the rest of the paper.

The definitions in (2.2) and (2.4) are particularly meaningful in finite samples since ν^γ depends on n, λ, σ and γ. That is, the weak signals are relative and depend on the sample size, the signal to noise ratio, and the tuning parameter selection. In other words, weak signals $\mathbf{\Theta^{(W)}}$ might be asymptotically trivial since the three levels automatically degenerate into two levels: zero and nonzero coefficients. However, weak signals should not be ignored in finite samples and serve as a transition phase between noise variables $\mathbf{\Theta^{(N)}}$ and strong signals $\mathbf{\Theta^{(S)}}$.

2.1.2 Weak signal identification

In this section we discuss how to identify weak signals more specifically. We propose a two-step procedure to recover possible weak signals which might be missed in a standard model selection procedure, and distinguish weak signals from strong signals.

The key component of the proposed procedure is to utilize the estimated probability of detection $\widehat{P}_d$. Since the true information of parameter θ is unknown, P_d cannot be calculated directly using (2.1). We propose to estimate P_d by plugging in the least-square estimator $\hat{\theta}_{LS}$ in (2.1). The expectation of the estimator $\widehat{P}_d$ remains an increasing function of $|\theta|$. That is,

$$\widehat{P}_d = \Phi(\frac{\hat{\theta}_{LS} - \sqrt{\lambda}}{\sigma/\sqrt{n}}) + \Phi(\frac{-\hat{\theta}_{LS} - \sqrt{\lambda}}{\sigma/\sqrt{n}}), \tag{2.5}$$

and

$$E(\widehat{P}_d) = \Phi(\frac{\sqrt{n}}{\sqrt{2}\sigma}(\theta - \sqrt{\lambda})) - \Phi(-\frac{\sqrt{n}}{\sqrt{2}\sigma}(\theta + \sqrt{\lambda})).$$

In the following, the weak signal is identified through replacing $P_d, (\gamma^w, \nu^w)$ and (γ^s, ν^s) in (2.2) by $\widehat{P}_d, (\gamma_1, \nu_1)$ and (γ_2, ν_2), where (γ_1, ν_1) and (γ_2, ν_2) satisfy equation (2.3). We denote the identified noise, weak and strong parameter spaces as $\left(\widehat{\boldsymbol{\Theta}}^{(\boldsymbol{N})}, \widehat{\boldsymbol{\Theta}}^{(\boldsymbol{W})}, \widehat{\boldsymbol{\Theta}}^{(\boldsymbol{S})}\right)$, where $\widehat{\boldsymbol{\Theta}}^{(\boldsymbol{N})} = \left\{\theta : |\hat{\theta}_{LS}| \leq \nu_1\right\}$, $\widehat{\boldsymbol{\Theta}}^{(\boldsymbol{W})} = \left\{\theta : \nu_1 < |\hat{\theta}_{LS}| \leq \nu_2\right\}$, and $\widehat{\boldsymbol{\Theta}}^{(\boldsymbol{S})} = \left\{\theta : |\hat{\theta}_{LS}| > \nu_2\right\}$. The details of selecting ν_1 and ν_2 are given below.

Note that in finite samples, there is no ideal threshold value ν_1 which can separate signal variables and noise variables perfectly, as there is a trade-off between recovering weak signals and including noise variables. Here ν_1 is selected to control a signal's false-positive rate τ. Specifically, $\nu_1 = z_{\tau/2}\frac{\sigma}{\sqrt{n}}$ for any given tolerant false-positive rate τ since it can be shown that $P(\theta \notin \widehat{\boldsymbol{\Theta}}^{(\boldsymbol{N})} | \theta = 0) = \tau$, see Lemma 2.7.1 in the supplement. Here we choose the false-positive rate τ to be larger than the τ_0, since we intend to recover most of the weak signals. This is very different from standard model selection which mainly focuses on model selection consistency, but neglects detection of weak signals.

The low threshold value ν_2 for strong signals is selected to ensure that a strong signal can be identified with high probability. We choose $\nu_2 = \sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, and it can be verified that the estimated detection rate $\widehat{P}_d$ for any identified strong signal stays above $1 - \alpha$. In fact, based on (2.5), $\widehat{P}_d$ satisfies the inequality $P_d > E(\widehat{P}_d)$ when the true signal is strong. Figure 2.2 illustrates the relationship between P_d and $E(\widehat{P}_d)$. Therefore there is a high probability that the true detection rate P_d is larger than $1 - \alpha$ when $\widehat{P}_d > 1 - \alpha$.

In summary, the main focus of weak signal identification is to recover weak signals as much as possible, at the cost of having a false-positive rate τ in finite samples. This is in contrast to standard model selection procedures which emphasize consistent model selection with a close-to-zero false-positive rate, but at the cost of not selecting most weak signals.

To better understand the difference and connection between the proposed weak signal identification procedure and the standard model selection procedure, we provide Figure 2.3 for illustration. Let $P_{d,0}(\theta)$

(dashed line) and $P_{d,1}(\theta)$ (dotted line) denote the probabilities of selecting θ in the standard model selection and the proposed weak signal identification, respectively, where $P_{d,0}(\theta) = P(|\hat{\theta}_{LS}| > \sqrt{\lambda})$, and $P_{d,1}(\theta) = P(\nu_1 < |\hat{\theta}_{LS}| < \sqrt{\lambda})$. Then the total selection probability $P_{d,2}(\theta)$ (solid line) for the proposed method is $P_{d,2}(\theta) = P_{d,0}(\theta) + P_{d,1}(\theta) = P(|\hat{\theta}_{LS}| > \nu_1)$. Figure 2.3 indicates that the proposed procedure recovers weak signals better than the standard model selection procedure, but at a cost of a small false-positive rate of including some noise variables. These two procedures have similar detection power for strong signals.

2.1.3 Weak signal inference

In this section, we propose a two-step inference procedure which consists of an asymptotic-based confidence interval for strong signals, and a least-square confidence interval for the identified weak signals.

The asymptotic-based inference method has been developed for the SCAD estimator (Fan and Li, 2001). Zou (2006) also provides the asymptotic distribution of the adaptive LASSO estimator $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\mathcal{A}}_n}$ for nonzero parameters, where $\mathcal{A}_n = \{1, 2, \cdots, q\}$. In finite samples, the adaptive LASSO estimator $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\mathcal{A}}_n}$ is biased due to the shrinkage estimation. The bias term of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\mathcal{A}}_n}$ and the covariance matrix estimator of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\mathcal{A}}_n}$ are given by

$$\hat{\boldsymbol{b}}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\mathcal{A}}_n}) = (\frac{1}{n}\boldsymbol{X}_{\boldsymbol{\mathcal{A}}_n}^T\boldsymbol{X}_{\boldsymbol{\mathcal{A}}_n} + \boldsymbol{\Omega})^{-1}(p_\lambda^{'}(|\hat{\theta}_1|)sgn(|\hat{\theta}_1|), \cdots, p_\lambda^{'}(|\hat{\theta}_q|)sgn(|\hat{\theta}_q|))^T, \tag{2.6}$$

and

$$\widehat{Cov}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\mathcal{A}}_n}) = \left\{\boldsymbol{X}_{\boldsymbol{\mathcal{A}}_n}^T\boldsymbol{X}_{\boldsymbol{\mathcal{A}}_n} + n\boldsymbol{\Omega}\right\}^{-1}\boldsymbol{X}_{\boldsymbol{\mathcal{A}}_n}^T\boldsymbol{X}_{\boldsymbol{\mathcal{A}}_n}\ \left\{\boldsymbol{X}_{\boldsymbol{\mathcal{A}}_n}^T\boldsymbol{X}_{\boldsymbol{\mathcal{A}}_n} + n\boldsymbol{\Omega}\right\}^{-1}\widehat{\sigma}^2, \tag{2.7}$$

where $\boldsymbol{\Omega} = \text{diag}\left\{\frac{\hat{w}_1}{|\hat{\theta}_1|}, \cdots, \frac{\hat{w}_q}{|\hat{\theta}_q|}\right\}$, and $\hat{w}_i = 1/|\hat{\theta}_{LS,i}|$. Although the bias term is asymptotically negligible, it is important to correct the biased term to get more accurate confidence intervals in finite samples. That is, for any strong signal in $\widehat{\boldsymbol{\Theta}}^{(\boldsymbol{S})}$, a $100(1-\alpha)\%$ confidence interval can be constructed as

$$\hat{\theta} + \hat{b}_{AL} \pm z_{\alpha/2}\hat{\sigma}_{AL}, \tag{2.8}$$

where $\hat{b}_{AL}$ and $\hat{\sigma}_{AL}$ are the corresponding biased component in (2.6) and the square root of the diagonal variance component in (2.7).

The above inference procedure performs well for strong signals (Fan and Li, 2001; Zou, 2006; Huang and Xie, 2007). However, this procedure does not apply well to weak signals. This is because weak signals are often missed in standard model selection procedures, and therefore there is no confidence interval constructed for any estimator shrunk to 0. Moreover, even if a weak signal is selected, the variance estimator in (2.7) tends to underestimate its true standard error, and consequently the confidence interval based on (2.8) is under-covered. Here we propose an alternative confidence interval for a weak signal in $\widehat{\boldsymbol{\Theta}}^{(\boldsymbol{W})}$ by utilizing the least-square information as follows.

The proposed inference for weak signals is motivated in that the bias-corrected confidence interval in (2.8) is close to the least-square confidence interval when a signal is strong. Therefore it is natural to construct a least-square confidence interval for a weak signal to solve the problem of excessive shrinkage for weak signal estimators. We construct a $100(1-\alpha)\%$ least-square confidence interval for an identified weak signal as

$$\hat{\theta}_{LS} \pm z_{\alpha/2}\hat{\sigma}_{LS}, \tag{2.9}$$

where $\hat{\theta}_{LS}$ and $\hat{\sigma}_{LS}$ are the components of the least-square estimator and the square root of the diagonal component of the covariance matrix estimator:

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{LS}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y},$$
$$\widehat{Cov}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{LS}}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\widehat{\sigma}^2.$$

Combining (2.8) and (2.9), we provide a two-step confidence interval for any detected signal coefficient as the following:

$$\left\{\hat{\theta} + \hat{b}_{LS} \pm z_{\alpha/2}\hat{\sigma}_{LS}\right\}\mathbf{1}_{\left\{\theta\in\widehat{\boldsymbol{\Theta}}^{(\boldsymbol{W})}\right\}} + \left\{\hat{\theta} + \hat{b}_{AL} \pm z_{\alpha/2}\hat{\sigma}_{AL}\right\}\mathbf{1}_{\left\{\theta\in\widehat{\boldsymbol{\Theta}}^{(\boldsymbol{S})}\right\}}. \tag{2.10}$$

Here we propose different confidence interval constructions for weak and strong signals, and the proposed inference is a mixed procedure combining (2.8) and (2.9). Our inference procedure performs similarly to the asymptotic inference for strong signals, but outperforms the existing inference procedures in that the proposed confidence interval provides more accurate coverage for weak signals. Note that if a signal strength is too weak, neither existing methods nor our method can provide reasonably good inferences. Nevertheless, our method still provides a better inference than the asymptotic-based method across all signal levels.

2.2 Theoretical results

In this section, we establish finite sample theory on coverage rate for the proposed two-step inference method, and compare it with the coverage rate of the asymptotic-based inference method. The asymptotic properties for penalized estimators have been investigated by Fan and Li (2001), Fan and Peng (2004), Zou (2006), Zou and Li (2008) and many others. When the sample size is sufficiently large and the signal strength is strong, the asymptotic inference is quite accurate in capturing the information of the penalized estimators. For instance, the covariance estimator of the penalized estimates in (2.7) is a consistent estimator (Fan and Peng, 2004). However, the sandwich estimator of the covariance only performs well for strong signals, not for weak signals in finite samples. Therefore, it is important to investigate the finite sample property of the penalized estimator, and especially the weak signal estimators for the proposed method.

We construct the exact coverage rates of the $100(1-\alpha)\%$ confidence intervals for the proposed method and the asymptotic method when the sample size is finite. The derivation for finite sample theory is very different from the asymptotic theory. In addition, since the coverage rate function is not monotonic, we need to compare the difference of the two coverage rates piece-wisely.

Given a confidence level parameter α, the following regularity conditions are required for selecting the false-positive rate τ:

(C1) $\tau \geq \alpha$,

(C2) $\frac{\alpha+\tau}{2} < \Phi(-\frac{1}{2}z_{\alpha/2})$, which is equivalent to $\tau < 2\Phi(-\frac{1}{2}z_{\alpha/2}) - \alpha$.

Condition (C1) is to ensure that the second step of the proposed method is able to identify weak signals. Condition (C2) provides a range of τ, so the false positive-rate is not too large. In addition, we also assume that λ satisfies the criterion:

$$\sqrt{\lambda} \geq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}. \tag{2.11}$$

The criterion in (2.11) implies that our focus is the case when $\sqrt{\lambda} \geq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, where excessive shrinkage might affect weak signal selection. It can be verified that $\alpha \geq \tau_0$ if λ is in this range, and this guarantees that $\tau > \tau_0$.

In the following, for any parameter θ and parameter ν associated with a different level of tuning, we introduce three probability functions, P_s, CR_a and CR_b as follows. Let P_s be the detection power of θ:

$$P_s(\theta, \nu) = \Phi(\frac{\theta - \nu}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta - \nu}{\sigma/\sqrt{n}}).$$

We define CR_a as the coverage probability based on the asymptotic inference approach when $|\hat{\theta}_{LS}|$ is larger than ν:

$$CR_a(\theta,\nu) = \begin{cases} \Big\{P_s(\theta,\nu) - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})\Big\} & \\ \cdot I_{\Big\{\nu \le z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}\Big\}} & \text{if } |\theta| \le |\nu - z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}| \\ \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\sqrt{n}(\nu-\theta)}{\sigma}) & \text{if } |\nu - z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}| \le |\theta| \le \nu + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}} \\ 1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) & \text{if } |\theta| > \nu + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}, \end{cases}$$

where $\tilde{\sigma}(\theta) = (1+\frac{\lambda}{\theta^2})^{-1}\sigma$; and CR_b is the coverage probability based on the least-square confidence interval when $|\hat{\theta}_{LS}|$ is larger than ν:

$$CR_b(\theta,\nu) = \begin{cases} \{P_s(\theta,\nu) - \alpha\} \cdot I_{\Big\{\nu \le z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Big\}} & \text{if } |\theta| \le |\nu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}| \\ 1 - \frac{\alpha}{2} - \Phi(\frac{\sqrt{n}(\nu-\theta)}{\sigma}) & \text{if } |\nu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}| \le |\theta| \le \nu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \\ 1-\alpha & \text{if } |\theta| > \nu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}. \end{cases}$$

The explicit expressions of coverage rates based on the asymptotic and the proposed two-step methods are provided in the following lemma.

Lemma 2.2.1 *Suppose n, σ and tuning parameter λ are given, the coverage rate $CR_1(\theta)$ of the $100(1-\alpha)\%$ confidence interval for any coefficient θ based on the asymptotic inference is*

$$CR_1(\theta) = \frac{CR_a(\theta,\nu_0)}{P_s(\theta,\nu_0)}, \tag{2.12}$$

where $\nu_0 = \sqrt{\lambda}$. Given any τ, the coverage rate $CR(\theta)$ of the $100(1-\alpha)\%$ confidence interval for any coefficient θ using the proposed two-step inference method is given by:

$$CR(\theta) = \frac{CR_b(\theta,\nu_1) + CR_a(\theta,\nu_2) - CR_b(\theta,\nu_2)}{P_s(\theta,\nu_1)}, \tag{2.13}$$

where $\nu_0 = \sqrt{\lambda}, \nu_1 = z_{\tau/2}\frac{\sigma}{\sqrt{n}}$, and $\nu_2 = \sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

The derivations of $CR_1(\theta)$ and $CR(\theta)$ are provided in the proof of Lemma 2.2.1 in the supplement. In fact, $CR_1(\theta)$ is the conditional coverage probability based on the asymptotic confidence interval, given that θ is selected using tuning parameter λ. Similarly, $CR(\theta)$ is the conditional coverage probability of the proposed confidence interval in (2.10), given that θ is selected based on the two-step procedure. The expression of $CR(\theta)$ in (2.13) can be interpreted as the summation of two sub-components, where the first component

$\frac{CR_b(\theta,\nu_1)-CR_b(\theta,\nu_2)}{P_s(\theta,\nu_1)}$, corresponds to the conditional coverage probability of the least-square confidence interval when $\nu_1 < |\hat{\theta}_{LS}| < \nu_2$, and the second component $\frac{CR_a(\theta,\nu_2)}{P_s(\theta,\nu_1)}$, is the conditional coverage probability of the asymptotic-based confidence interval when $|\hat{\theta}_{LS}| > \nu_2$.

In addition, we show in the supplement that both $CR_1(\theta)$ and $CR(\theta)$ are piece-wise smooth functions, and require one to compare two coverage rates at each interval separately. We introduce the boundary points associated with $CR_1(\theta)$ and $CR(\theta)$ as follows. Let c_1, c_2, c_3 and c_4 be the solutions of $\theta = \sqrt{\lambda} - z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$, $\theta = \sqrt{\lambda} + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$, $\theta = \sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$, and $\theta = \sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$, respectively. Here c_1 and c_2 are the boundary points of piece-wise intervals for $CR_1(\theta)$ in (2.12), and c_3 and c_4 are the boundary points of piece-wise intervals for $CR(\theta)$ in (2.13). It can be shown that the orders of c_1, c_2, c_3 and c_4 satisfy $c_1 < c_3 < c_2 < c_4$. More specific ranges for c_1, c_2, c_3 and c_4 are provided in Lemma 2.7.2 of the supplement. Since there are no explicit solutions for these boundary points, we rely on the orders of these boundary points to examine the difference between $CR_1(\theta)$ and $CR(\theta)$.

In the following, we define $\Delta(\theta) = CR(\theta) - CR_1(\theta)$ as a difference function between $CR(\theta)$ and $CR_1(\theta)$. Theorem 2.2.1 and Theorem 2.2.2 provide the uniform low bounds of $\Delta(\theta)$ for different ranges of λ when $z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \sqrt{\lambda} < (z_{\alpha/2} + z_{\tau/2})\frac{\sigma}{\sqrt{n}}$ and $\sqrt{\lambda} \ge (z_{\alpha/2} + z_{\tau/2})\frac{\sigma}{\sqrt{n}}$. The mathematical details of the proofs are provided in section 2.7.

Theorem 2.2.1 *Under conditions (C1)-(C2), if λ satisfies $z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \sqrt{\lambda} < (z_{\alpha/2} + z_{\tau/2})\frac{\sigma}{\sqrt{n}}$, the piece-wise lower bounds for $\Delta(\theta)$ are provided as follows:*

(a) when $\theta \in [0, c_1]$, $\Delta(\theta) \ge 1 - \frac{\alpha}{\tau} > 0$;

(b) when $\theta \in [c_1, \nu_0]$, $\Delta(\theta) \ge \frac{2}{1+\alpha} - 2\Phi(\frac{1}{2}z_{\alpha/2}) > 0$;

(c) when $\theta \in [\nu_0, +\infty)$, $\Delta(\theta)$ satisfies either $\Delta(\theta) \ge 0$ or $-\frac{\alpha}{2} < \Delta(\theta) < 0$.

In addition, a more specific lower bound for $\Delta(\theta)$ on $[\nu_0, +\infty)$ is given by:

$$\Delta(\theta) \ge \begin{cases} -4(1-\frac{\alpha}{2})\Phi(-\frac{3}{2}z_{\alpha/2}) & \text{if } \theta \in [\nu_0, \min\{\nu_3, c_3\}] \\ \text{See Table 2.1} & \text{if } \theta \in [\min\{\nu_3, c_3\}, \max\{\nu_3, c_2\}] \\ -\frac{4(1-\alpha)}{(2-\alpha)^2}\Phi(-2z_{\alpha/2}) - \frac{\alpha(1-\alpha)}{2-\alpha} & \text{if } \theta \in [\max\{\nu_3, c_2\}, c_4] \\ -(1-\alpha)\frac{\Phi(-\frac{3}{2}z_{\alpha/2})}{\Phi(\frac{3}{2}z_{\alpha/2})^2} & \text{if } \theta \in [c_4, \nu_4] \\ -(1-\alpha)\frac{\Phi(-2z_{\alpha/2})}{\Phi(2z_{\alpha/2})^2} & \text{if } \theta \in [\nu_4, \infty), \end{cases}$$

where $\nu_3 = (z_{\alpha/2} + z_{\tau/2})\frac{\sigma}{\sqrt{n}}$, and $\nu_4 = \sqrt{\lambda} + 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Table 2.1 provides the lower bounds for $\Delta(\theta)$ on interval $[\min\{\nu_3, c_3\}, \max\{\nu_3, c_2\}]$ under three cases.

Table 2.1: Specific bounds of $\Delta(\theta)$ on interval $[\min\{\nu_3, c_3\}, \max\{\nu_3, c_2\}]$

case 1: $c_3 < \nu_3 < c_2$	$\theta \in [c_3, \nu_3]$ $-2\Phi(-\frac{3}{2}z_{\alpha/2})$	$\theta \in [\nu_3, c_2]$ $-\frac{4(1-\alpha)}{(2-\alpha)^2}\Phi(-2z_{\alpha/2}) - \frac{\alpha(1-\alpha)}{2-\alpha}$
case 2: $c_3 < c_2 < \nu_3$	$\theta \in [c_3, c_2]$ $-2(1-\alpha)\Phi(-\frac{3}{2}z_{\alpha/2})$	$\theta \in [c_2, \nu_3]$ $-\frac{1-\alpha}{[\Phi(\frac{1}{2}z_{\alpha/2})]^2}\Phi(-2z_{\alpha/2})$
case 3: $\nu_3 < c_3 < c_2$	$\theta \in [\nu_3, c_2]$ $-\frac{\alpha}{2}$	

Theorem 2.2.2 *Under conditions (C1)-(C2), if λ satisfies $\sqrt{\lambda} \geq (z_{\alpha/2} + z_{\tau/2})\frac{\sigma}{\sqrt{n}}$, the lower bounds for $\Delta(\theta)$ are provided as follows:*

(a) when $\theta \in [0, \min\{\nu_3, c_1\}]$, $\Delta(\theta) \geq 1 - \frac{\alpha}{\tau} > 0$;

(b) when $\theta \in [\min\{\nu_3, c_1\}, \nu_0]$, see Table 2.2;

(c) when $\theta \in [\nu_0, +\infty)$, $\Delta(\theta) \geq 0$ or $-\frac{\alpha}{2} < \Delta(\theta) < 0$.

Table 2.2: Specific bounds of $\Delta(\theta)$ on interval $[\min\{\nu_3, c_1\}, \nu_0]$

case 4: $\nu_3 < c_1$	$\theta \in [\nu_3, \nu_0]$ $2 - \alpha - 2\Phi(\frac{1}{2}z_{\alpha/2})$	
case 5: $c_1 < \nu_3$	$\theta \in [c_1, \nu_3]$ $\Phi(-\frac{1}{2}z_{\alpha/2}) - \frac{\alpha}{2}$	$\theta \in [\nu_3, \nu_0]$ $2 - \alpha - 2\Phi(\frac{1}{2}z_{\alpha/2})$

Theorem 2.2.1 and Theorem 2.2.2 indicate that the proposed method outperforms the asymptotic-based method, with a uniform lower bound for $\Delta(\theta)$ when $\theta \in [0, \nu_0]$. More specifically, the lower bound of $\Delta(\theta)$ depends on α and τ for case (i) ($\theta \in [0, c_1]$) in Theorem 2.2.1 and case (i) ($\theta \in [0, \min\{\nu_3, c_1\}]$) in Theorem 2.2.2. Since we select τ to be larger than α, it is clear that $\Delta(\theta)$ is bounded above zero. For case (ii) ($\theta \in [c_1, \nu_0]$) in Theorem 2.2.1 and case (ii) ($\theta \in [\min\{\nu_3, c_1\}, \nu_0]$) in Theorem 2.2.2, the lower bound of $\Delta(\theta)$ only depends on α. In fact, the minimum value of $\frac{2}{1+\alpha} - 2\Phi(\frac{1}{2}z_{\alpha/2})$ is larger than 0.22 if $\alpha \in [0.05, 0.1]$, based on Theorem 2.2.1. This also confirms that the proposed method provides a confidence region with at least 22% improvement in coverage rate than the one based on the asymptotic method. The lower bounds of case (ii) in Theorem 2.2.2 can be interpreted in a similar way.

In addition, both Theorem 2.2.1 and Theorem 2.2.2 imply that when $\theta \in (\nu_0, +\infty)$ with a moderately large coefficient, the proposed method performs better than, or close to, the asymptotic method. In summary, the two-step inference method provides more accurate coverage than the one based on the asymptotic inference, and is also more effective for the weak signal region.

In Theorem 2.2.1, since the order relationships among c_2, c_3 and ν_3 change for different ranges of tuning parameters and choices of false positive rate τ, it leads to the three cases in Table 2.1. Figure 2.4, Figure

2.5 and Figure 2.6 illustrate examples for the three cases. Similarly, the order relationships among ν_3 and c_1 also change for different choices of λ and τ in Theorem 2.2.2, leading to the two cases in Table 2.2. Figure 2.7 and Figure 2.8 graph these two cases.

We also measure the difference between $CR(\theta)$ and the nominal confidence level $1-\alpha$. Theorem 2.2.3 provide an upper bound between $CR(\theta)$ and $1-\alpha$ when $\theta \in [\nu_1, +\infty)$. The mathematical details of the proof is provided in section 2.7.

Theorem 2.2.3 *(Difference between $CR(\theta)$ and $1-\alpha$) Under conditions (C1)-(C2), for any given n, σ, tuning parameter λ and proper selection of τ, the difference between $CR(\theta)$ and confidence level $1-\alpha$ when $\theta \geq \nu_1$ can be bounded by:*

$$|(1-\alpha) - CR(\theta)| < \begin{cases} \frac{\Phi(-\frac{1}{2}z_{\alpha/2})}{\Phi(\frac{1}{2}z_{\alpha/2})} - \alpha + 4(1-\alpha)\Phi(-\frac{3}{2}z_{\alpha/2}), & \textit{if } c_3 < \nu_3; \\ \Phi(-\frac{1}{2}z_{\alpha/2}) - \frac{\alpha}{2} + 4(1-\alpha)\Phi(-\frac{3}{2}z_{\alpha/2}), & \textit{if } c_3 \geq \nu_3. \end{cases}$$

Theorem 2.2.3 provide an upper bound between $CR(\theta)$ and $1-\alpha$ when $\theta \in [\nu_1, +\infty)$. According to the details of proof in section 2.7, the worst case of the two-step inference performance for $\theta \geq \nu_1$ happens when θ is close to c_3. The upper bound for the difference between the actual coverage rate and nominal level relies on α, which is provided in the above theorem. For example, the upper bound is roughly around 15% when $\alpha = 0.05$, and around 16% when $\alpha = 0.1$. In general, the difference between $CR(\theta)$ and $1-\alpha$ gradually decreases as signal gets stronger, and the difference becomes marginal when signal turns sufficiently strong.

2.3 Numerical studies

To examine the empirical performance of the proposed inference procedure, we conduct simulation studies to evaluate the accuracy of the confidence intervals described in section 2.1.3. We generate 400 simulated data with a sample size of n under the linear model $y = \boldsymbol{X\theta} + \mathcal{N}(0, \sigma^2)$, where $\boldsymbol{X} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_p)$ and $\boldsymbol{X}_j \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I_n})$. We allow covariates $\boldsymbol{X}$ to be correlated with an AR(1) correlation structure, and the pairwise correlation $cor(\boldsymbol{X_i}, \boldsymbol{X_j}) = \rho^{|i-j|}$. We choose $(n, p, \sigma) = (100, 20, 2)$ and $(400, 50, 2)$, and $\rho = 0, 0.2$ or 0.5 for each setting. In addition, the p-dimensional coefficient vector $\boldsymbol{\theta} = (1, 1, 0.5, \theta, 0, \cdots, 0)$, which consists of two strong signals of 1's, one moderate strong signal of 0.5, one varying-signal θ, and $(p-4)$ null variables. We let the coefficient θ vary between 0 (null) to 1 (strong signal) to examine the confidence coverages across different signal strength levels.

We construct 95% confidence intervals for an identified signal based on (2.10). We implement the glmnet package in R (Friedman et al., 2010) to obtain the adaptive LASSO estimator, where λ is tuned by the

Bayesian information criterion (BIC). For comparison, we also construct standard confidence intervals based on the asymptotic formula in (2.8), along with the bootstrap method (Efron and Tibshirani, 1993), the smoothed bootstrap method (Efron, 2013) and the perturbation method (Minnier et al., 2012). For both regular bootstrap and smoothed bootstrap methods, the number of bootstrap sampling is set to be 4000 (Efron, 2013). For the perturbation method, the resampling time is set to be 500 according to Minnier et al. (2012). In addition, the coverage rate for the OLS estimator is included as a benchmark since there is no shrinkage in OLS estimation and the confidence interval is the most accurate.

Figure 2.9 illustrates the relationship between τ_0 and τ when $\rho = 0.2$ for two model settings $(n, p, \sigma) = (100, 20, 2)$ and $(400, 50, 2)$, where $\tau_0 = 2\Phi(-\frac{\sqrt{n}\lambda}{\sigma})$ based on Section 2.1.1. We choose τ larger than τ_0 according to Section 3.2, that is, the false-positive rate in the weak signal recovery procedure is slightly larger than the false-positive rate in the model selection procedure. In these two model settings, τ_0 are around 0.1 and 0.03 respectively; here we select τ as 0.2. In practice, the selection of τ is flexible, and can be determined by a tolerance level for including noise variables.

Figure 2.10 and Figure 2.11 provide the coverage probabilities for θ varying between 0 and 1 when $\rho = 0.2$ in two model settings. In each figure, ν^s and ν^w are the average threshold coefficients corresponding to the detection powers $P_d = 0.9$ and 0.1, respectively. When the signal strength is close to zero, neither of the coverage rates using our method and the asymptotic method are accurate. However, the proposed method is still better than the asymptotic one, since the asymptotic coverage rate is close to zero; while the bootstrap and perturbation methods tend to provide over-coverage confidence intervals. The proposed method becomes more accurate as the magnitude of signal θ increases, and also outperforms all the other methods especially in the weak signal region. For example, in setting 1, the coverage rate of the proposed method is quite close to 95% when the signal strength is larger than 0.4. On the other hand, the resampling methods and asymptotic inference provide low coverage rates for signal strength below 0.8. When signal strength is above 0.8, the coverages from all methods are accurate and close to 95%.

The results for correlated covariates settings are provided in Table 2.4. For each setting, we select two different values of θ whose detection probabilities P_d are between 0.1 and 0.9. For all these settings, the asymptotic inference, bootstrap and perturbation methods provide confidence intervals far below 95% when signals are weak. In general, our method provides a stable inference even when the correlation coefficient increases, and the coverage rate for weak signals is between 91-97% when $\rho = 0.5$. The asymptotic-based inference has the lowest coverage rates among all, and performs extremely poorly when ρ is larger. The coverage rates based on the perturbation method are all below 75% for weak signals. Note that the coverage rate improvement using the smoothed bootstrap method is not significant compared to the standard bootstrap method. In addition, for $n = 100, p = 50$, the bootstrap and smooth bootstrap methods encounter a singular-designed matrix problem due to a small sample size, which do not produce any simulation results 7-10% times. The average coverage rates provided in Table 2.4 might not be valid and are marked with $*$. In general, our two-step method is computationally more efficient than the resampling methods. Table 2.5 also provides the computation time for a single simulation run for each of these methods.

Figure 2.12 also presents the probabilities of assigning each signal category for a given θ value, where the probabilities for identified strong signal ($P(\theta \in \widehat{\mathbf{\Theta}}^{(\boldsymbol{S})})$), weak signal ($P(\theta \in \widehat{\mathbf{\Theta}}^{(\boldsymbol{W})})$) and null variable ($P(\theta \in \widehat{\mathbf{\Theta}}^{(\boldsymbol{N})})$) are denoted as solid, dotted and dashed lines respectively. The probability of each identified signal category relies on signal strength. Specifically, when a signal is close to zero, it is likely to be identified as zero most of the time, with the highest $P(\theta \in \widehat{\mathbf{\Theta}}^{(\boldsymbol{N})})$); when a signal falls into the weak signal region, $P(\theta \in \widehat{\mathbf{\Theta}}^{(\boldsymbol{W})})$ becomes dominant; and when θ increases to be a strong signal, $P(\theta \in \widehat{\mathbf{\Theta}}^{(\boldsymbol{S})})$ also gradually increases and reaches to 1.

2.4 Applications

In this section, we apply HIV drug resistance data
(http://hivdb.stanford.edu/) to illustrate the proposed method. The HIV drug resistance study aims to identify the association of protease mutations with susceptibility to the antiretroviral drug. Since antiretroviral drug resistance is a major obstacle to the successful treatment of HIV-1 infection, studying the generic basis of HIV-1 drug resistance is crucial for developing new drugs and designing an optimal therapy for patients. The study was conducted on 702 HIV-infected patients, where 79 out of 99 protease codons in the viral genome have mutations. Here the drug resistance is measured in units of IC_{50}.

We consider a linear model:

$$\boldsymbol{y} = \sum_{i=1}^{p} \boldsymbol{X_i}\theta_i + \boldsymbol{\varepsilon}, \tag{2.14}$$

where the response variable $\boldsymbol{y}$ is the log-transformation of nonnegative IC_{50}, and the model predictors $\boldsymbol{X_i}$ are binary variables indicating the mutation presence for each codon. The total number of candidate codons $p = 79$. We are interested in examining which codon mutations have effect on drug resistance.

We apply the proposed two-step inference method to identify codons' mutation presence which have strong or mild effects on HIV drug resistance. We use the GLMNET in R to obtain the adaptive LASSO estimator for the linear model in (2.14), where the initial weight of each coefficient is based on the OLS estimator. The tuning parameter λ is selected by the Bayesian information criterion, and σ is estimated similarly as in Zou (2006). To control the noise variable selection, we choose $\tau = 0.05$. According to the proposed identification procedure in Section 2.1.2, we calculate two threshold values ν_1 and ν_2 as 0.061 and 0.136, which correspond to two threshold probabilities, $\gamma_1 = 0.327$ and $\gamma_2 = 0.975$, for identifying weak and strong signals, respectively.

We constructed 95% confidence intervals using the proposed method and the perturbation approach (Minnier et al. 2011) for the chosen variables. Since all the predictors are binary variables, the bootstrap resampling method is not applicable, as it leads to a singular design matrix.

In the first step, we apply the adaptive LASSO procedure which selects 17 codons; in the second step, our method identifies additional 11 codons associated with drug resistance. Among 28 codons we identified, 13 of them are identified as strong signals and 15 of them as weak signals. Minnier et al.'s (2011) approach identified 18 codons, where the 13 signals (codon 10, 30, 32, 33, 46, 47, 48, 50, 54, 76, 84, 88 and 90) are the same as our strong-signal codons, and their remaining 5 signals (codon 37, 64, 71, 89 and 93) are among our 15 identified weak signals. In previous studies, Wu (2009) identifies all 13 strong signals using a permutation test for the regression coefficients obtained from LASSO; while Johnson et al. (2005) collect drug resistance mutation information based on multiple research studies, and discover 9 strong-signal codons (10, 32, 46, 47, 50, 54, 76, 84, 90) which are relevant to drug resistance. Neither of these approaches distinguish between strong-signal and weak-signal codons.

Figure 2.13 presents a graphical summary showing the half-width of the constructed confidence intervals based on our method and the perturbation approach, where strong signals are labeled in blue, and weak signals are labeled in red. To make full comparisons for both strong and weak signals, Figure 2.13 includes confidence intervals for all selected variables based on our method, even if some of them are not selected by Minnier et al.'s (2011) approach. Table 2.3 also provides the average half-widths of confidence intervals in each signal category. In general, our method provides shorter lengths of confidence intervals for all strong signals, and longer lengths of confidence intervals for weak signals compared to the perturbation approach. This is not surprising, since the variables with weak coefficients associated with the response variable are relatively weaker, and likely result in wider confidence intervals to ensure a more accurate coverage.

Table 2.3: Average half width of the CIs

	All Selected Variables	Strong Signals	Weak Signals
CI^{Our}	0.147	0.118	0.173
CI^{Ptb}	0.171	0.197	0.148

In summary, our approach recovers more codons than other existing approaches. One significance of our method lies in its capability of identifying a pool of strong signals which have strong evidence association with HIV drug resistance, and a pool of possible weak signals which might be mildly associated with drug resistance. In many medical studies, it is important not to miss statistically weak signals, which could be clinically valuable predictors.

2.5 Summary and discussion

In this chapter, we propose weak signal identification under the penalized model selection framework, and develop a new two-step inferential method which is more accurate in providing confidence coverage for weak signal parameters in finite samples. The proposed method is applicable for true models involving both strong and weak signals. One advantage of the new method is that it is able to recover possible weak signals which are missed due to excessive shrinkage in model selection, in addition to distinguishing weak signals from strong signals.

The two-step inference procedure imposes different selection criteria and confidence interval construction for strong and weak signals. Both theory and numerical studies indicate that the combined approach is more effective compared to the asymptotic inference approach, and bootstrap sampling and other resampling methods. Moreover, the computation cost of the proposed method is much lower than resampling procedures.

The primary concern for the existing model selection procedure is that it applies excessive shrinkage in order to achieve model selection consistency. However, this results in low detection power for weak signals in finite samples. The essence of the proposed method is to apply a mild tuning in identifying weak signals. Therefore, there is always a trade-off between a signal's detection power and false-discovery rate. In our approach, we intend to recover weak signals as much as possible, without sacrificing too much model selection consistency by including too many noise variables.

In our numerical studies, we notice that the resampling methods do not provide good inference for weak signals. Specifically, the coverage probability of bootstrap confidence interval is over-covered and exceeds the $(1-\alpha)100\%$ confidence level when the true parameter is close to 0. This is not surprising, as Andrew (2000) shows that the bootstrap procedure is inconsistent for boundary problems, such as in our case where the boundary parameters are in the order of $1/\sqrt{n}$.

In the proposed two-step inference procedure, although we utilize information from the least-square estimators, our approach is very different from applying the least-square inference directly without a model selection step. The non-penalization method is not feasible when the dimension of covariates is very large, e.g., to examine or visualize thousands of conference intervals without model selection. Therefore it is essential to make a sound statistical inference in conjunction with the variable selection, simultaneously. In addition, our method is also different from drawing inference after model selection, which could be problematic; as once we fail to select some important variables, no inference can be made for the parameters associated with these variables.

In this paper, we develop our method and theory under the orthogonal design assumption. However, our numerical studies indicate that the proposed method is still valid when correlations among covariates are moderate. It would be interesting to extend the current method to non-orthogonal designed covariates problems. In addition, it is important to explore weak-signal inference for high-dimensional model settings when the dimension of covariates exceeds the sample size. For high-dimensional settings, one possible direction is to scale down the model size using variable screening methods (Fan and Lv, 2007; Li et al., 2012). This enables us to have a smaller number of features than the sample size, and we can apply the weak signal identification procedure for the downsized model. These problems are quite challenging and deserve further investigation.

2.6 A tuning parameter adjustment approach

In section 2.1.2, we identify weak signal and select tuning parameter under the principle of controlling signal selection's type I error. In this section, we propose a new method of selecting tuning parameter, by maximizing the summation of type I error and detection power of a predetermined weak signal's lower bound. We define the lower bound of weak signal as $\theta_{\gamma_w} = \sqrt{\lambda}$, whose detection power is around 0.5 if we only use model selection. In the weak signal identification, a mild tuning parameter λ_{new} is selected to better identify signals around θ_{γ_w}. The basic rule of selecting λ_{new} is to increase the detection probability for signals around θ_{γ_w} as much as possible, while controlling for the false positive rate of selecting variables at the same time.

When tuning parameter equals θ_{γ_w}, the type I error equals $2\Phi(-\frac{\sqrt{n\lambda_{new}}}{\sigma})$, and the selection power for $\theta = \theta_{\gamma_w}$ equals $\Phi(\frac{\sqrt{n\lambda}-\sqrt{n\lambda_{new}}}{\sigma}) + \Phi(\frac{-\sqrt{n\lambda}-\sqrt{n\lambda_{new}}}{\sigma})$. In this way, we select λ_{new} based on:

$$\begin{aligned}\lambda_{new} &= \arg\max\{\text{ new type I error + new detection power for } \theta_{\gamma_w}\} \\ &= \arg\max\left\{1 - 2\Phi(-\frac{\sqrt{n\lambda_{new}}}{\sigma}) + \Phi(\frac{\sqrt{n\lambda}-\sqrt{n\lambda_{new}}}{\sigma}) + \Phi(\frac{-\sqrt{n\lambda}-\sqrt{n\lambda_{new}}}{\sigma})\right\}.\end{aligned}$$

Equivalently, λ_{new} is the solution to the equation below:

$$\phi(\frac{\sqrt{n}(\sqrt{\lambda}-\sqrt{\lambda_{new}})}{\sigma}) + \phi(\frac{\sqrt{n}(-\sqrt{\lambda}-\sqrt{\lambda_{new}})}{\sigma}) - 2\phi(-\frac{\sqrt{n\lambda_{new}}}{\sigma}) = 0, \tag{2.15}$$

and it can be verified from (2.15) that the following inequalities hold for λ_{new}:

$$\lambda/4 < \lambda_{new} < \lambda. \tag{2.16}$$

We conduct simulations to verify the performance of the new tuning parameter selection approach. Our model is generated in the same way with that in section 2.3, and $n = 100, p = 50, \sigma = 2$, together with $\boldsymbol{\theta} = (1, 1, 0.5, 0.5, 0, \cdots, 0)$. We conduct 400 simulation runs, and we observe the the relationship between λ and λ_{new}, the detection power of weak signal lower bound θ_{γ_w}, and the false positive rate under λ and λ_{new}. The simulation results are displayed in Figure 2.19. Simulation shows that the new tuning parameter is always smaller than the original tuning parameter, which is consistent with the conclusion in (2.16). The new tuning parameter is more stable over different simulations, compared with the original tuning parameter λ. Weak signal lower bound θ_{γ_w} varies in different simulation, its detection power becomes much higher than 0.5 when applying the new tuning parameter, compared with that using original λ. In addition, the type I error using λ_{new} is slightly larger than that using λ. However, the new type I error is still under control and is around 0.2 for different simulations.

There exists several advantages of the new tuning parameter adjustment approach. First, it connects weak signal lower bound with λ. The weak signal lower bound no longer only relies on signal selection's type I error. Second, weak signal identification step is based on the model selection step, thus the model selection procedure and weak signal identification procedure are closely connected. In this way, one can also check that the new weak signal lower bound θ_{γ_w} falls outside the doughnut hole asymptotically.

2.7 Proofs of the main results

Notations

$\nu_0 = \sqrt{\lambda}, \nu_1 = z_{\tau/2}\frac{\sigma}{\sqrt{n}}, \nu_2 = \sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, $\nu_3 = (z_{\alpha/2} + z_{\tau/2})\frac{\sigma}{\sqrt{n}}$, $\nu_4 = \sqrt{\lambda} + 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Tuning Parameter Selection

BIC criteria function is:

$$BIC(\lambda) = log(\hat{\sigma}^2_\lambda) + \hat{q}_\lambda \frac{log(n)}{n},$$

where $\hat{\sigma}_\lambda$ is the estimated standard deviation based on λ, and $\hat{q}_\lambda$ is the number of covariates in the model.

Proof of Lemma 2.1.1

For any γ satisfies $\epsilon < \gamma < 1 - \epsilon$, we show that ν^γ that solves $P_d = \gamma$ follows $\frac{\nu^\gamma}{\sqrt{\lambda_n}} \to 1$, as $n \to \infty$.

P_d can be rewritten as $P_d = \Phi(\frac{\sqrt{n\lambda_n}}{\sigma}(\frac{\nu^\gamma}{\sqrt{\lambda_n}} - 1)) - \Phi(-\frac{\sqrt{n\lambda_n}}{\sigma}(1 + \frac{\nu^\gamma}{\sqrt{\lambda_n}}))$. Given $n\lambda_n \to \infty$, if $\lim_{n\to+\infty} \frac{\nu^\gamma}{\sqrt{\lambda_n}} >$

1, then $P_d(\nu^\gamma) \to 1$, as $n \to \infty$; else if $\lim_{n\to+\infty} \frac{\nu^\gamma}{\sqrt{\lambda_n}} < 1$, then $P_d(\nu^\gamma) \to 0$, as $n \to \infty$. Since $P_d(\nu^\gamma) = \gamma \in (\epsilon, 1-\epsilon)$, we have $\lim_{n\to+\infty} \frac{\nu^\gamma}{\sqrt{\lambda_n}} = 1$, as $n \to \infty$.

Therefore, both ν^s and ν^w satisfies $\frac{\nu^s}{\sqrt{\lambda_n}} \to 1$ and $\frac{\nu^w}{\sqrt{\lambda_n}} \to 1$.

Expectation of $\widehat{P}_d$

Since $\sqrt{n}/\sigma\hat{\theta}_{LS} \sim N(0,1)$, then

$$\begin{aligned} E(\widehat{P}_d) &= E\left\{\Phi(\frac{\sqrt{n}}{\sigma}(\hat{\theta}_{LS} - \sqrt{\lambda})) - \Phi(-\frac{\sqrt{n}}{\sigma}(\sqrt{\lambda} + \hat{\theta}_{LS}))\right\} \\ &= \Phi(\frac{\sqrt{n}}{\sqrt{2}\sigma}(\theta - \sqrt{\lambda})) - \Phi(-\frac{\sqrt{n}}{\sqrt{2}\sigma}(\theta + \sqrt{\lambda})). \end{aligned}$$

Proof of Lemma 2.2.1

Define $CI_a : \left\{\theta : |\hat{\theta}_{LS} - \theta| < z_{\alpha/2}\tilde{\sigma}(\theta)/\sqrt{n}\right\}$, $CI_b : \left\{\theta : |\hat{\theta}_{LS} - \theta| < z_{\alpha/2}\sigma/\sqrt{n}\right\}$. The confidence interval in (2.10) can be rewritten as:

$$CI_a \cdot I_{\left\{|\hat{\theta}_{LS}| \geq \nu_2\right\}} + CI_b \cdot I_{\left\{\nu_1 < |\hat{\theta}_{LS}| < \nu_2\right\}}.$$

Based on CI_a, CI_b, we define functions $CR_a(\theta,\nu), CR_b(\theta,\nu)$ in the following manners:

$$CR_a(\theta,\nu) = P(\theta \in CI_a, |\hat{\theta}_{LS}| > \nu), \tag{2.17}$$

$$CR_b(\theta,\nu) = P(\theta \in CI_b, |\hat{\theta}_{LS}| > \nu)\sigma/\sqrt{n}). \tag{2.18}$$

Besides, define $P_s(\theta,\nu)$ as $P(|\hat{\theta}_{LS}| > \nu)$, which equals to

$$P_s(\theta,\nu) = \Phi(\frac{\theta - \nu}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta - \nu}{\sigma/\sqrt{n}}). \tag{2.19}$$

The explicit expression of $CR_a(\theta,\nu)$ is derived based on three cases:

(i). If $\nu < \theta - z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$, then

$$CR_a(\theta,\nu) = P\left(|\hat{\theta}^{LS} - \theta| \leq z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}\right) = 1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}).$$

(ii). If $|\theta - z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}| < \nu < \theta + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$, then

$$CR_a(\theta,\nu) = P\left(\nu < \hat{\theta}^{LS} < \theta + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}\right) = \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu - \theta}{\sigma/\sqrt{n}}).$$

(iii). If $\nu < -\theta + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$, then

$$\begin{aligned} CR_a(\theta,\nu) &= P\left(\theta - z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}} < \hat{\theta}^{LS} < -\nu\right) + P\left(\nu < \hat{\theta}^{LS} < \theta + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}\right) \\ &= P_s(\theta,\nu) - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}). \end{aligned}$$

The expression of $CR_b(\theta,\nu)$ can be derived in a similar way. Therefore, $CR_a(\theta,\nu)$ and $CR_b(\theta,\nu)$ have the explicit expressions as:

$$CR_a(\theta,\nu) = \begin{cases} \left(P_s(\theta,\nu) - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})\right)\cdot & \\ I_{\left\{\nu < z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}\right\}} & \text{if } |\theta| < |\nu - z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}|, \\ \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu-\theta}{\sigma/\sqrt{n}}) & \text{if } |\nu - z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}| \le |\theta| \le \nu + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}, \\ 1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) & \text{if } |\theta| > \nu + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}, \end{cases}$$

and

$$CR_b(\theta,\nu) = \begin{cases} (P_s(\theta,\nu) - \alpha)\, 1_{\left\{\nu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}} & \text{if } |\theta| < |\nu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}|, \\ 1 - \frac{\alpha}{2} - \Phi(\frac{\nu-\theta}{\sigma/\sqrt{n}}) & \text{if } |\nu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}| < |\theta| < \nu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \\ 1 - \alpha & \text{if } |\theta| > \nu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}. \end{cases}$$

The equations in (2.17)-(2.19) are used to provide explicit expressions for $CR_1(\theta)$ and $CR(\theta)$ in Lemma 2.2.1. According to the definition of CR_1,

$$\begin{aligned} CR_1(\theta) &= P(\theta \text{ in asymptotic-based CI}|\theta \text{ is selected in model selection}) \\ &= P\left(|\hat{\theta}_{LS} - \theta| < z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}} \,\middle|\, |\hat{\theta}_{LS}| > \sqrt{\lambda}\right) \\ &= \frac{P(|\hat{\theta}_{LS} - \theta| < z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}, |\hat{\theta}_{LS}| > \sqrt{\lambda})}{P(|\hat{\theta}_{LS}| > \sqrt{\lambda})} \\ &= \frac{CR_a(\theta,\nu_0)}{P_s(\theta,\nu_0)}, \end{aligned}$$

where $\nu_0 = \sqrt{\lambda}$.

Similarly, $CR(\theta)$ can be expressed as

$$\begin{aligned}
CR(\theta) &= P(\theta \in \text{ CI as in (2.10) } |\theta \text{ is selected by the two-step procedure}) \\
&= \frac{P(\theta \in CI_a, |\hat{\theta}_{LS}| \geq \nu_2) + P(\theta \in CI_b, \nu_1 < |\hat{\theta}_{LS}| < \nu_2)}{P(|\hat{\theta}_{LS}| > \nu_1)} \\
&= \frac{P(\theta \in CI_a, |\hat{\theta}_{LS}| \geq \nu_2) + P(\theta \in CI_b, |\hat{\theta}_{LS}| > \nu_2) - P(\theta \in CI_b, |\hat{\theta}_{LS}| > \nu_1)}{P(|\hat{\theta}_{LS}| > \nu_1)} \\
&= \frac{CR_a(\theta, \nu_2) + CR_b(\theta, \nu_1) - CR_b(\theta, \nu_2)}{P_s(\theta, \nu_1)}.
\end{aligned}$$

Lemma 2.7.1 *If we select $\nu_1 = z_{\tau/2}\frac{\sigma}{\sqrt{n}}$, then the false positive rate of weak signal identification procedure equals τ.*

Proof of Lemma 2.7.1

By definition, the false positive rate equals $P(\theta \in \widehat{\boldsymbol{\Theta}}^{(\boldsymbol{w})} \cup \widehat{\boldsymbol{\Theta}}^{(\boldsymbol{s})}|\theta = 0) = P(|\hat{\theta}_{LS}| > \nu_1|\theta = 0) = 2\Phi(-\frac{\sqrt{n}}{\sigma}\nu_1) = \tau$.

Lemma 2.7.2 *Under conditions (C1)-(C2), when λ satisfies conditions $\sqrt{\lambda} > z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$,*

(a) if c_1 is the solution to $\theta = \sqrt{\lambda} - z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$, then $c_1 \in \left((z_{\alpha/2} - z_{\tau/2})\frac{\sigma}{\sqrt{n}}, \sqrt{\lambda}\right)$;

(b) if c_2 is the solution to $\theta = \sqrt{\lambda} + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$, then $c_2 \in \left(\sqrt{\lambda} + \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$;

(c) if c_3 is the solution to $\theta = \sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$, then $c_3 \in \left(\sqrt{\lambda}, \sqrt{\lambda} + \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$;

(d) if c_4 is the solution to $\theta = \sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$, then
$c_4 \in \left(\sqrt{\lambda} + \frac{3}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \sqrt{\lambda} + 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$.

In addition, the order relationships of c_1, c_2, c_3 and c_4 follow: $c_1 < c_3 < c_2 < c_4$.

Proof of Lemma 2.7.2

Define functions $K_1(\theta)$, $K_2(\theta)$, $K_3(\theta)$ and $K_4(\theta)$ as follows:

$$\begin{aligned}
K_1(\theta) &= \theta - \sqrt{\lambda} + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}, \\
K_2(\theta) &= \theta - (\sqrt{\lambda} + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}), \\
K_3(\theta) &= \theta - (\sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}), \\
K_4(\theta) &= \theta - (\sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}).
\end{aligned}$$

The function $K_1(\theta)$ is an increasing function of θ, and there exists a unique solution to $K_1(\theta) = 0$. Since $K_1(\sqrt{\lambda}) > 0$, so $c_1 < \sqrt{\lambda}$. Moreover, we have $K_1((z_{\alpha/2} - z_{\tau/2})\frac{\sigma}{\sqrt{n}}) = (z_{\alpha/2} - z_{\tau/2})\frac{\sigma}{\sqrt{n}} - \sqrt{\lambda} + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}} < z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}} - z_{\tau/2}\frac{\sigma}{\sqrt{n}} < (\frac{1}{2}z_{\alpha/2} - z_{\tau/2})\frac{\sigma}{\sqrt{n}} < 0$, where we use that $\tilde{\sigma}(\theta) < \frac{\sigma}{2}$ when $\theta < \sqrt{\lambda}$, and $\frac{1}{2}z_{\alpha/2} < z_{\tau/2}$ followed by condition (C1) and (C2). Combining $K_1(\sqrt{\lambda}) > 0$ and $K_1((z_{\alpha/2} - z_{\tau/2})\frac{\sigma}{\sqrt{n}}) < 0$, we have $c_1 \in \left((z_{\alpha/2} - z_{\tau/2})\frac{\sigma}{\sqrt{n}}, \sqrt{\lambda}\right)$.

Based on the definition of c_2, $c_2 = \sqrt{\lambda} + z_{\alpha/2}\frac{\tilde{\sigma}(c_2)}{\sqrt{n}}$, it is obvious that $c_2 > \sqrt{\lambda}$. Further we have $\sigma/2 < \tilde{\sigma}(c_2) < \sigma$, thus $c_2 \in \left(\sqrt{\lambda} + \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$. In the following, we show that there exists a unique solution to the function $K_2(\theta)$ in $\left(\sqrt{\lambda} + \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \sqrt{\lambda} + 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$, through showing that $K_2'(\theta) > 0$ in this interval. Note that $K_2'(\theta) = 1 - z_{\alpha/2}/\sqrt{n}\tilde{\sigma}'(\theta)$, where $\tilde{\sigma}'(\theta) = \frac{2\sigma\lambda}{\theta^3}(1 + \frac{\lambda}{\theta^2})^{-2}$, it is sufficient to show $\theta^3(1 + \frac{\lambda}{\theta^2})^2 > 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\lambda$. The proof can be done in two steps. First, given that $\theta < \sqrt{\lambda} + 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ and condition (C2), we have $\theta < 3\sqrt{\lambda}$ and consequently $(1 + \frac{\lambda}{\theta^2})^2 > \frac{100}{81}$. Second, we can show that $\frac{100}{81}\theta^3 > \frac{100}{81}(\sqrt{\lambda} + \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}})^3 > \frac{100}{81} \cdot \left[(\sqrt{\lambda})^3 + 3(\sqrt{\lambda})^2 \cdot \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = \frac{250}{81}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\lambda > 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\lambda$. Combining these two steps, it holds that $\theta^3(1 + \frac{\lambda}{\theta^2})^2 > 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\lambda$. Equivalently, $K_2'(\theta) > 0$ on $\left(\sqrt{\lambda} + \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \sqrt{\lambda} + 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$.

Similar to c_1, $c_3 = \sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - z_{\alpha/2}\frac{\tilde{\sigma}(c_3)}{\sqrt{n}}$. Based on the fact that $\sigma/2 < \tilde{\sigma}(c_3) < \sigma$, together with that $K_3(\theta)$ is an increasing function, it is straightforward to show that $c_3 \in \left(\sqrt{\lambda}, \sqrt{\lambda} + \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$.

In addition, c_4 satisfies $c_4 = \sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + z_{\alpha/2}\frac{\tilde{\sigma}(c_4)}{\sqrt{n}}$. First, it can be shown that $c_4 \in \left(\sqrt{\lambda} + \frac{3}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \sqrt{\lambda} + 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$. The function $K_4(\theta)$ has the same derivative with $K_2(\theta)$, thus $K_4'(\theta) > 0$ in $\left(\sqrt{\lambda} + \frac{3}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \sqrt{\lambda} + 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$. Therefore, there exists a unique root of $K_4(\theta)$ in the above interval.

Lemma 2.7.3 *Given* $\theta \in \left(\sqrt{\lambda} + \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \sqrt{\lambda} + 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$, *then* $\theta > c_2$ *if and only if* $\theta > \sqrt{\lambda} + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$, *and* $\theta > c_4$ *if and only if* $\theta > \sqrt{\lambda} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$.

Proof of Lemma 2.7.3 The conclusions directly come from the fact that $K_2'(\theta) = K_4'(\theta) > 0$ on $\left(\sqrt{\lambda} + \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \sqrt{\lambda} + 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$, as is shown in Lemma 2.7.2.

The conclusions in Lemma 2.7.2 and Lemma 2.7.3 are critical in our proof of Theorem 2.2.1 and Theorem 2.2.2.

Lemma 2.7.4 *(Monotonicity of $CR_1(\theta)$)*

Suppose $\theta > 0$, $CR_1(\theta)$ *is a piece-wise monotonic function on* $[0, c_2]$. *More specifically,* $CR_1(\theta)$ *is a non-decreasing function on* $[0, c_1]$, *an increasing function on* $[c_1, c_2]$.

Proof of Lemma 2.7.4

When $\theta \in [0, c_1]$, both $-\Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)$ and $P_s(\theta, \nu_0)$ are increasing functions of θ. In addition, $\Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) < 0$ and $P_s(\theta, \nu_0) > 0$. It is straightforward to show that the derivative of $\frac{-\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_d^{(1)}(\theta)}$ is always positive, so CR_1 is non-decreasing when $\theta \in [0, c_1]$.

When $\theta \in [c_1, c_2]$, CR_1 can be written as

$$CR_1(\theta) = \frac{\Phi\left(\frac{\sigma}{\sqrt{n}}(\theta - \sqrt{\lambda})\right) - \Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)}{P_s(\theta, \nu_0)}.$$

It is sufficient to show that $\frac{\Phi\left(\frac{\sigma}{\sqrt{n}}(\theta-\sqrt{\lambda})\right)}{\Phi\left(\frac{\sigma}{\sqrt{n}}(\theta-\sqrt{\lambda})\right)+\Phi\left(-\frac{\sigma}{\sqrt{n}}(\theta+\sqrt{\lambda})\right)}$ is increasing w.r.t θ, which is equivalent to show that its derivative w.r.t θ is always positive. This holds true because $\Phi\left(\frac{\sigma}{\sqrt{n}}(\theta - \sqrt{\lambda})\right)$ is an increasing function of θ, $\Phi\left(-\frac{\sigma}{\sqrt{n}}(\theta + \sqrt{\lambda})\right)$ is a decreasing function of θ, and both of them are positive.

Lemma 2.7.5 *For any fixed parameter value $\nu > 0$, the function*

$$\frac{CR_b(\theta, \nu)}{P_s(\theta, \nu)} = \begin{cases} \left(1 - \frac{\alpha}{P_s(\theta,\nu)}\right) 1_{\left\{\nu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}} & \text{if } |\theta| < |\nu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}| \\ \frac{1-\frac{\alpha}{2}-\Phi\left(\frac{\sqrt{n}}{\sigma}(\nu-\theta)\right)}{P_s(\theta,\nu)} & \text{if } |\nu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}| < |\theta| < \nu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \\ \frac{1-\alpha}{P_s(\theta,\nu)} & \text{if } |\theta| > \nu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \end{cases}$$

is

(a) non-decreasing, when $|\theta| \leq |\nu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}|$;

(b) increasing, when $|\nu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}| < |\theta| < \nu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$;

(c) decreasing, when $|\theta| \geq \nu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Proof of Lemma 2.7.5

By definition, for any positive ν,

$$P_s(\theta, \nu) = \Phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta + \nu)\right),$$
$$\text{and } \frac{\partial P_s(\theta, \nu)}{\partial \theta} = \frac{\sqrt{n}}{\sigma}\left[\phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu)\right) - \phi\left(-\frac{\sqrt{n}}{\sigma}(\theta + \nu)\right)\right] > 0.$$

Then $P_s(\theta, \nu)$ is an increasing function w.r.t $|\theta|$.

When $\theta > 0$, it can be shown that $\frac{CR_b(\theta,\nu)}{P_s(\theta,\nu)}$ is non-decreasing on $\left[0, |\nu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}|\right]$, and decreasing on $\left[\nu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, +\infty\right)$.

When $\theta \in \left[|\nu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}|, \nu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$,

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\left(\frac{CR_b(\theta,\nu)}{P_s(\theta,\nu)}\right) \\
&= \frac{\sqrt{n}}{\sigma}\cdot\frac{1}{P_s(\theta,\nu)^2}\cdot\left[\phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu)\right)\Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu)\right)\right. \\
&\left.+\phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu)\right)\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu)\right)+\frac{\alpha}{2}\phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu)\right)-\frac{\alpha}{2}\phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu)\right)\right] > 0,
\end{aligned}$$

therefore $\frac{CR_b(\theta,\nu)}{P_s(\theta,\nu)}$ is an increasing function w.r.t to θ on $\left[|\nu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}|, \nu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$.

Lemma 2.7.6 *The formulas for $CR_1(\theta)$ and $CR(\theta)$ in Lemma 2.2.1 can also be expressed as:*

$$CR_1(\theta) = \frac{CR_a(\theta,\nu_0)}{P_s(\theta,\nu_0)}, \tag{2.20}$$

$$CR(\theta) = \begin{cases} \frac{CR_b(\theta,\nu_1)}{P_s(\theta,\nu_1)} & \text{if } |\theta| < \nu_0 \\ \frac{CR_b(\theta,\nu_1)}{P_s(\theta,\nu_1)} - \frac{CR_b(\theta,\nu_2)}{P_s(\theta,\nu_1)} & \text{if } \nu_0 \le |\theta| \le c_3 \\ \frac{CR_b(\theta,\nu_1)}{P_s(\theta,\nu_1)} + \frac{CR_a(\theta,\nu_2)}{P_s(\theta,\nu_1)} - \frac{CR_b(\theta,\nu_2)}{P_s(\theta,\nu_1)} & \text{if } |\theta| > c_3. \end{cases} \tag{2.21}$$

Then $CR_1(\theta) = J_1(\theta), CR(\theta) = J_2(\theta) - J_3(\theta) + J_4(\theta)$, where the four functions $J_1(\theta), J_2(\theta), J_3(\theta)$ and $J_4(\theta)$ are defined as:

$$\begin{aligned}
J_1(\theta) &= \frac{CR_a(\theta,\nu_0)}{P_s(\theta,\nu_0)}, J_2(\theta) = \frac{CR_b(\theta,\nu_1)}{P_s(\theta,\nu_1)}, \\
J_3(\theta) &= \frac{CR_b(\theta,\nu_2)}{P_s(\theta,\nu_1)}, J_4(\theta) = \frac{CR_a(\theta,\nu_2)}{P_s(\theta,\nu_1)}.
\end{aligned}$$

Proof of Lemma 2.7.6

The Lemma follows immediately from Lemma 2.2.1, $J_3(\theta) = 0$ when $\theta \le \nu_0$, and $J_4(\theta) = 0$ when $\theta \le c_3$.

In fact, J_1 represents CR_1 in (2.20), and J_2, J_3 and J_4 represent different components of CR in (2.21), which are critical to the following derivations of the difference between the two coverage functions, as in Theorem 2.2.1 and Theorem 2.2.2.

Proof of Theorem 2.2.1

(i). When $\theta \in [0, c_1]$, we have $\Delta(\theta) \ge 1 - \frac{\alpha}{\tau} > 0$.

When $\theta \in [0, c_1]$, it is obvious that $CR_1(\theta) = 0$. By Lemma 2.7.5, $CR(\theta)$ is increasing on $[0, \nu_0]$, and $CR(\theta) = 1 - \frac{\alpha}{\tau}$ when $\theta = 0$. Thus $CR(\theta) - CR_1(\theta) \ge 1 - \frac{\alpha}{\tau}$ for $\theta \in [0, c_1]$, which provides the first lower bound in Theorem 2.2.1. Note that here we also use $c_1 < \nu_0$ by Lemma 2.7.2.

(ii). When $\theta \in [c_1, \nu_0]$, we have $\Delta(\theta) \ge \frac{2}{1+\alpha} - 2\Phi(\frac{1}{2}z_{\alpha/2}) > 0$.

When $\theta \in [c_1, \nu_0]$, by definition

$$
\begin{aligned}
CR_1(\theta) &= \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_0 - \theta)\right)}{P_s(\theta, \nu_0)}, \\
CR(\theta) &= \frac{1 - \frac{\alpha}{2} - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_1 - \theta)\right)}{P_s(\theta, \nu_1)}.
\end{aligned}
$$

In the following, we show that $\frac{\partial CR_1(\theta)}{\partial\theta} > \frac{\partial CR(\theta)}{\partial\theta}$, so $CR(\theta) - CR_1(\theta)$ is decreasing when $\theta \in [c_1, \nu_0]$. The first order derivatives of $CR_1(\theta)$ and $CR(\theta)$ are:

$$
\frac{\partial CR_1(\theta)}{\partial\theta} = \left[\frac{z_{\alpha/2}}{\sigma}\phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)\tilde{\sigma}(\theta)' + \frac{\sqrt{n}}{\sigma}\phi\left(\frac{\sqrt{n}}{\sigma}(\nu_0 - \theta)\right)\right] P_s(\theta, \nu_0)^{-1} \tag{2.22}
$$

$$
-\left[\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_0 - \theta)\right)\right] P_s(\theta, \nu_0)^{-2} \tag{2.23}
$$

$$
\frac{\partial CR(\theta)}{\partial\theta} = \frac{\sqrt{n}}{\sigma}\phi\left(\frac{\sqrt{n}}{\sigma}(\nu_1 - \theta)\right) P_s(\theta, \nu_1)^{-1} \tag{2.24}
$$

$$
-\left[1 - \frac{\alpha}{2} - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_1 - \theta)\right)\right] P_s(\theta, \nu_1)^{-2}, \tag{2.25}
$$

where each first-order derivative is composed of two parts. We show the inequality of each part separately. First we show that (2.22) > (2.24), which is sufficient by showing

$$
\phi\left(\frac{\sqrt{n}}{\sigma}(\nu_0 - \theta)\right) P_s(\theta, \nu_0)^{-1} > \phi\left(\frac{\sqrt{n}}{\sigma}(\nu_1 - \theta)\right) P_s(\theta, \nu_1)^{-1}.
$$

This is equivalent to show

$$
\frac{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta + \nu_1)\right)}{\phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_1)\right)} > \frac{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta + \nu_0)\right)}{\phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0)\right)}. \tag{2.26}
$$

The inequality in (2.26) can be proved based on monotinicity of two functions $\frac{\Phi(x)}{\phi(x)}$ and $\frac{\Phi(-x-y)}{\phi(x-y)}$. Specifically, it can be shown that $\frac{\Phi(x)}{\phi(x)}$ is an increasing function of $x \in \mathbb{R}$, and $\frac{\Phi(-x-y)}{\phi(x-y)}$ is a decreasing function of $y \in \mathbb{R}^+$, for any fixed value of $x > 0$. More specifically, since $\nu_1 < \nu_0$, we have

$$
\frac{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_1)\right)}{\phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_1)\right)} > \frac{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0)\right)}{\phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0)\right)} \text{ and } \frac{\Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta + \nu_1)\right)}{\phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_1)\right)} > \frac{\Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta + \nu_0)\right)}{\phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0)\right)},
$$

based on which the inequality in (2.26) holds.

Next we show that (2.24) < (2.25), which is equivalent with

$$\left[1-\frac{\alpha}{2}-\Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_1-\theta)\right)\right]P_s(\theta,\nu_1)^{-2} > \left[\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)-\Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_1-\theta)\right)\right]P_s(\theta,\nu_0)^{-2}.$$

It can be shown by

$$\frac{1-\frac{\alpha}{2}-\Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_1-\theta)\right)}{P_s(\theta,\nu_1)^2} > \frac{1-\frac{\alpha}{2}-\Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_0-\theta)\right)}{P_s(\theta,\nu_0)^2} > \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)-\Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_1-\theta)\right)}{P_s(\theta,\nu_0)^2}.$$

Based on the above arguments, we can conclude that for $\theta\in[c_1,\nu_0]$:

$$\frac{\partial CR_1(\theta)}{\partial\theta} > \frac{\partial CR(\theta)}{\partial\theta}.$$

Therefore, $\min_{\theta\in[c_1,\nu_0]}\Delta(\theta)=CR(\nu_0)-CR_1(\nu_0)$. More specifically,

$$\begin{aligned} CR(\nu_0) &= \frac{\Phi(\frac{\nu_0-\nu_1}{\sigma/\sqrt{n}})-\frac{\alpha}{2}}{\Phi(\frac{\nu_0-\nu_1}{\sigma/\sqrt{n}})+\Phi(\frac{-\nu_0-\nu_1}{\sigma/\sqrt{n}})} > \frac{1-\alpha}{1+\alpha}, \\ CR_1(\nu_0) &= \frac{\Phi(\frac{1}{2}z_{\alpha/2})-\frac{\alpha}{2}}{\frac{1}{2}+\Phi(-\frac{2\nu_0}{\sigma/\sqrt{n}})} < 2\Phi(\frac{1}{2}z_{\alpha/2})-1, \end{aligned}$$

thus

$$CR(\nu_0) \quad - \quad CR_1(\nu_0) > \frac{2}{1+\alpha}-2\Phi(\frac{1}{2}z_{\alpha/2}),$$

which provides the second lower bound in Theorem 2.2.1.

(iii). When $\theta\in[\nu_0,+\infty)$, we have $\Delta(\theta)$ satisfies either $\Delta(\theta)\geq 0$ or $-\frac{\alpha}{2}<\Delta(\theta)<0$. We discuss over three different cases.

Case 1: $c_3<\nu_3<c_2$. The case is illustrated in Figure 2.4 and Figure 2.14.

When $\theta\in[\nu_0,c_3]$, we have

$$\begin{aligned} J_1(\theta) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})-\Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})+\Phi(\frac{-\theta-\nu_0}{\sigma/\sqrt{n}})}, J_2(\theta)=\frac{1-\frac{\alpha}{2}-\Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})+\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}, \\ \text{and } J_3(\theta) &= \frac{1-\frac{\alpha}{2}-\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})+\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}, \end{aligned}$$

thus

$$\begin{aligned}
\Delta(\theta) &= J_2(\theta) - J_1(\theta) - J_3(\theta) \\
&= \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_0}{\sigma/\sqrt{n}})}.
\end{aligned}$$

Further,

$$\begin{aligned}
\Delta(\theta) &= \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_0}{\sigma/\sqrt{n}})} \\
&> \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})} \\
&> \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})} \\
&= \frac{\left[\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})\right]\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) - \left[\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})\right]\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})}{\left[\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})\right]\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})} \\
&\quad - \frac{\left[\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})\right]\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}{\left[\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})\right]\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})} \\
&= \Delta_1(\theta) - \Delta_2(\theta),
\end{aligned}$$

where the second inequality uses that $z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}} \leq \nu_2 - \theta$ when $\theta \leq c_3$ by Lemma 2.7.3. Here $\Delta_1(\theta)$ and $\Delta_2(\theta)$ are defined as:

$$\begin{aligned}
\Delta_1(\theta) &= \frac{\left[\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})\right]\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) - \left[\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})\right]\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})}{\left[\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})\right]\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})}, \\
\Delta_2(\theta) &= \frac{\left[\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})\right]}{\left[\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})\right]\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})}\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}}).
\end{aligned}$$

First, it is straightforward to show that $\Delta_1(\theta) > 0$. Second, $\Delta_2(\theta)$ can be bounded from above by some small value. In fact,

$$\begin{aligned}
\Delta_2(\theta) &< \frac{\left[\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})\right]}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})}\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}}) \\
&< 4\left[\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})\right]\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}}) \\
&< 4\left[1 - \frac{\alpha}{2} - \Phi(-\frac{1}{2}z_{\alpha/2})\right]\Phi(-\frac{3}{2}z_{\alpha/2}),
\end{aligned}$$

where we use that $\nu_2 - \theta < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, $-\frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \nu_0 - c_3 < \nu_0 - \theta$ and $\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}}) < \Phi(\frac{-\nu_0-\nu_1}{\sigma/\sqrt{n}}) < \Phi(-\frac{3}{2}z_{\alpha/2})$ when $\nu_0 < \theta < c_3$. Combining the lower bounds for $\Delta_1(\theta)$ and $\Delta_2(\theta)$, we have:

$$\Delta(\theta) > -4\left[1 - \frac{\alpha}{2} - \Phi(-\frac{1}{2}z_{\alpha/2})\right]\Phi(-\frac{3}{2}z_{\alpha/2}).$$

In fact, the lower bound on the right hand side is quite close to zero.

When $\theta \in [c_3, \nu_3]$,

$$\Delta(\theta) = J_2(\theta) - J_1(\theta) - J_3(\theta) + J_4(\theta),$$

where

$$\begin{aligned}
J_1(\theta) &= \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_0 - \theta)\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta + \nu_0)\right)}, \\
J_2(\theta) &= \frac{1 - \frac{\alpha}{2} - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_1 - \theta)\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta + \nu_1)\right)}, \\
J_3(\theta) &= \frac{1 - \frac{\alpha}{2} - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_3 - \theta)\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta + \nu_1)\right)}, \\
J_4(\theta) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_3 - \theta)\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta - \nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta + \nu_1)\right)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Delta(\theta) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_0}{\sigma/\sqrt{n}})} \\
&= \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_0}{\sigma/\sqrt{n}})} \\
&> \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} \\
&\quad + \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})} \\
&= \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\left[\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})\right]\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} \cdot \left(-\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})\right) \\
&\quad + \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})\left[\frac{1}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})} - \frac{1}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})}\right] \\
&> \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\left[\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})\right]\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} \cdot \left(-\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})\right) \\
&> -2\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}}) > -2\Phi(-\frac{3}{2}z_{\alpha/2}),
\end{aligned}$$

the second inequality holds since

$$\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) > \frac{1}{2}, 0 < \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\left[\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})\right]} < 1,$$

and the last inequality holds since $-\theta-\nu_1 < -(z_{\alpha/2}+z_{\tau/2})\frac{\sigma}{\sqrt{n}} < -\frac{3}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, when $\theta \geq c_3 > \sqrt{\lambda} \geq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. When $\theta \in [\nu_3, c_2]$, $\Delta(\theta) = J_2(\theta) - J_1(\theta) + J_4(\theta) - J_3(\theta)$, where

$$\begin{aligned}
J_1(\theta) &= \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_0-\theta)\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_0)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_0)\right)}, \\
J_2(\theta) &= \frac{1-\alpha}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)}, \\
J_3(\theta) &= \frac{1-\frac{\alpha}{2}-\Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_2-\theta)\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)}, \\
J_4(\theta) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_2-\theta)\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)}.
\end{aligned}$$

Therefore,

$$\Delta(\theta) \quad = \quad \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \frac{\alpha}{2}}{P_s(\theta,\nu_1)} - \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_0-\theta)\right)}{P_s(\theta,\nu_0)}.$$

Further we have $\Delta(\theta) > \Delta_1(\theta) + \Delta_2(\theta)$, where

$$\begin{aligned}\Delta_1(\theta) \quad &= \quad \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \frac{\alpha}{2}}{P_s(\theta,\nu_1)} - \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \frac{\alpha}{2}}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})},\\ \text{and } \Delta_2(\theta) \quad &= \quad \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \frac{\alpha}{2}}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})}.\end{aligned}$$

It is straightforward to get a bound for $\Delta_1(\theta)$. In fact,

$$\Delta_1(\theta) = \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \frac{\alpha}{2}}{P_s(\theta,\nu_1)\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} \cdot \left[-\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})\right],$$

here $\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \frac{\alpha}{2} < 1-\alpha, P_s(\theta,\nu_1) > \Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) > 1-\frac{\alpha}{2}$, and $-\theta-\nu_1 < -\nu_3-\nu_1 < -2z_{\alpha/2}\sigma/\sqrt{n}$. Therefore, $\Delta_1(\theta) < 0$ and $|\Delta_1(\theta)| < \frac{4(1-\alpha)}{(2-\alpha)^2}\Phi(-2z_{\alpha/2})$.

It takes a few more steps to bound $\Delta_2(\theta)$. In fact,

$$\begin{aligned}\Delta_2(\theta) \quad &= \quad \frac{1-\frac{\alpha}{2} - \Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) - \Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})}\\ &= \quad \left[\frac{1-\frac{\alpha}{2}}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} - 1\right] + \Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)\left[\frac{1}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})} - \frac{1}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})}\right]\\ &> \quad \left[\frac{1-\frac{\alpha}{2}}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} - 1\right] + \frac{\alpha}{2}\left[\frac{1}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})} - \frac{1}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})}\right],\end{aligned}$$

the inequality holds since $\Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) > \frac{\alpha}{2}$. It can also be shown that both $\frac{1}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})} - \frac{1}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})}$ and $\frac{1-\frac{\alpha}{2}}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} - 1$ are decreasing functions of θ, given $\theta > \nu_0 > \nu_1$. Therefore,

$$\begin{aligned}\Delta_2(\theta) \quad &> \quad \left[\frac{1-\frac{\alpha}{2}}{\Phi(\frac{\nu_2-\nu_1}{\sigma/\sqrt{n}})} - 1\right] + \frac{\alpha}{2}\left[\frac{1}{\Phi(\frac{\nu_2-\nu_0}{\sigma/\sqrt{n}})} - \frac{1}{\Phi(\frac{\nu_2-\nu_1}{\sigma/\sqrt{n}})}\right]\\ &= \quad \frac{1-\alpha}{\Phi(\frac{\nu_2-\nu_1}{\sigma/\sqrt{n}})} - \frac{1-\alpha}{1-\frac{\alpha}{2}} > -\frac{\alpha(1-\alpha)}{2-\alpha}.\end{aligned}$$

Combining the lower bounds of $\Delta_1(\theta)$ and $\Delta_2(\theta)$, the lower bound for $\Delta(\theta)$ is provided by

$$\Delta(\theta) > -\frac{4(1-\alpha)}{(2-\alpha)^2}\Phi(-2z_{\alpha/2}) - \frac{\alpha(1-\alpha)}{2-\alpha} > -\frac{\alpha}{2}.$$

When $\theta \in [c_2, c_4]$,

$$\begin{aligned}
J_1(\theta) &= \frac{1 - 2\Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_0)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_0)\right)}, \\
J_2(\theta) &= \frac{1-\alpha}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)}, \\
J_3(\theta) &= \frac{1 - \frac{\alpha}{2} - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_2-\theta)\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)}, \\
J_4(\theta) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_2-\theta)\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\Delta(\theta) &= J_2(\theta) - J_1(\theta) + J_4(\theta) - J_3(\theta) \\
&= \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \frac{\alpha}{2}}{P_s(\theta,\nu_1)} - \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)}{P_s(\theta,\nu_0)}.
\end{aligned}$$

Again, $\Delta(\theta) > \Delta_1(\theta) + \Delta_2(\theta)$, where

$$\begin{aligned}
\Delta_1(\theta) &= \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \frac{\alpha}{2}}{P_s(\theta,\nu_1)\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right)} \cdot \left[-\Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)\right], \\
\text{and } \Delta_2(\theta) &= \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \frac{\alpha}{2}}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right)} - \frac{2\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - 1}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_0)\right)}.
\end{aligned}$$

Firstly,

$$\Delta_1(\theta) < 0 \text{ and } |\Delta_1(\theta)| < \frac{4(1-\alpha)}{(2-\alpha)^2}\Phi(-2z_{\alpha/2}),$$

which holds true because $\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) - \frac{\alpha}{2} < 1-\alpha$, $P_s(\theta,\nu_1) > \Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right) > 1-\frac{\alpha}{2}$, $\frac{1}{2}z_{\alpha/2} < z_{\tau/2}$, and $-\nu_3 - \nu_1 = -(z_{\alpha/2} + 2z_{\tau/2})\frac{\sigma}{\sqrt{n}} < -2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Secondly, when $\theta > c_2$, it holds that $\theta > \nu_0 + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$ according to Lemma 2.7.3, and further $\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_0)\right) >$

$\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)$. Therefore,

$$\begin{aligned}
\Delta_2(\theta) &> \frac{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)-\frac{\alpha}{2}}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right)} - \frac{2\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)-1}{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)} \\
&> \Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)-\frac{\alpha}{2} - \frac{2\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)-1}{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)} \\
&= \Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) + \frac{1}{\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)} - \frac{\alpha}{2} - 2.
\end{aligned}$$

The function on the right hand side is a decreasing function of $\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)$. Given that $\Phi\left(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right) < 1-\frac{\alpha}{2}$, we have

$$\Delta_2(\theta) > 1-\frac{\alpha}{2}-\frac{1}{1-\frac{\alpha}{2}}-\frac{\alpha}{2}-2 = -\frac{\alpha(1-\alpha)}{2-\alpha}.$$

Combining the lower bounds for $\Delta_1(\theta)$ and $\Delta_2(\theta)$, we have

$$\Delta(\theta) > -\frac{4(1-\alpha)}{(2-\alpha)^2}\Phi(-2z_{\alpha/2}) - \frac{\alpha(1-\alpha)}{2-\alpha}. \tag{2.27}$$

This lower bound for $\Delta(\theta)$ is exactly the same with that in the interval $[\nu_3, c_2]$.

When $\theta \in [c_4, \nu_4]$,

$$\begin{aligned}
J_1(\theta) &= \frac{1-2\Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_0)\right)+\Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_0)\right)}, \\
J_2(\theta) &= \frac{1-\alpha}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right)+\Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)}, \\
J_3(\theta) &= \frac{1-\frac{\alpha}{2}-\Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_2-\theta)\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right)+\Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)}, \\
J_4(\theta) &= \frac{1-2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right)+\Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\Delta(\theta) &= \frac{\Phi\left(\frac{\sqrt{n}}{\sigma}(\nu_2-\theta)\right)-\frac{\alpha}{2}}{P_s(\theta,\nu_1)} + \left[1-2\Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)\right]\left[\frac{1}{P_s(\theta,\nu_1)}-\frac{1}{P_s(\theta,\nu_0)}\right] \\
&> \left[1-2\Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)\right]\left[\frac{1}{P_s(\theta,\nu_1)}-\frac{1}{P_s(\theta,\nu_0)}\right],
\end{aligned}$$

the inequality holds since $\nu_2 - \theta \geq \nu_2 - \nu_4 = -z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ when $\theta \leq \nu_4$.

Let $\Delta_1(\theta) = \left[1 - 2\Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)\right]\left[\frac{1}{P_s(\theta,\nu_1)} - \frac{1}{P_s(\theta,\nu_0)}\right]$. We show that $\Delta_1(\theta)$ is negative but quite close to zero. When $\theta > c_4$, $P_s(\theta,\nu_1) > P_s(\theta,\nu_0) > \Phi(\frac{3}{2}z_{\alpha/2})$, and further $P_s(\theta,\nu_1) - P_s(\theta,\nu_0) \in \left(0, \Phi(-\frac{3}{2}z_{\alpha/2})\right)$. Therefore,

$$0 < \frac{1}{P_s(\theta,\nu_0)} - \frac{1}{P_s(\theta,\nu_1)} = \frac{P_s(\theta,\nu_1) - P_s(\theta,\nu_0)}{P_s(\theta,\nu_1)P_s(\theta,\nu_0)} < \frac{\Phi(-\frac{3}{2}z_{\alpha/2})}{\Phi(\frac{3}{2}z_{\alpha/2})^2},$$

together with $\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) < 1-\alpha$, we have

$$\Delta_1(\theta) < 0 \text{ and } |\Delta_1(\theta)| < (1-\alpha)\frac{\Phi(-\frac{3}{2}z_{\alpha/2})}{\Phi(\frac{3}{2}z_{\alpha/2})^2}.$$

Therefore,

$$\Delta(\theta) > -(1-\alpha)\frac{\Phi(-\frac{3}{2}z_{\alpha/2})}{\Phi(\frac{3}{2}z_{\alpha/2})^2}.$$

When $\theta \in [\nu_4, +\infty)$,

$$\begin{aligned}
J_1(\theta) &= \frac{1 - 2\Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_0)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_0)\right)}, \\
J_2(\theta) &= \frac{1-\alpha}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)}, \\
J_3(\theta) &= \frac{1-\alpha}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)}, \\
J_4(\theta) &= \frac{1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi\left(\frac{\sqrt{n}}{\sigma}(\theta-\nu_1)\right) + \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta+\nu_1)\right)}.
\end{aligned}$$

Therefore,

$$\Delta(\theta) = \left[1 - 2\Phi\left(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}\right)\right]\left[\frac{1}{P_s(\theta,\nu_1)} - \frac{1}{P_s(\theta,\nu_0)}\right].$$

Here $\Delta(\theta) < 0$, and

$$\Delta(\theta) > \frac{1}{P_s(\theta,\nu_1)} - \frac{1}{P_s(\theta,\nu_0)} > -\frac{\Phi(-2z_{\alpha/2})}{\Phi(2z_{\alpha/2})},$$

where we use that $\theta - \nu_0 > 2z_\alpha\frac{\sigma}{\sqrt{n}}$ when $\theta > \nu_4$. In fact, $P_s(\theta,\nu_1) \approx P_s(\theta,\nu_0)$ when θ gets quite large, thus

$\Delta(\theta) \approx 0$.

We complete the proof for case 1 in Theorem 2.2.1.

Case 2: $c_3 < c_2 < \nu_3$. The case is illustrated in Figure 2.5 and Figure 2.15.

When $\theta \in [\nu_0, c_3]$,

$$\begin{aligned} J_1(\theta) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_0)}, \\ J_2(\theta) &= \frac{1-\frac{\alpha}{2} - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_1)}, J_3(\theta) = \frac{1-\frac{\alpha}{2} - \Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_1)}, \\ \text{and } \Delta(\theta) &= J_2(\theta) - J_1(\theta) - J_3(\theta) \\ &= \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_1)} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_0)}, \end{aligned}$$

the expression of $\Delta(\theta)$ is exactly the same with that of case 1 on $[\nu_0, c_3]$, thus a same lower bound can be provided in a similar fashion with case 1.

When $\theta \in [c_3, c_2]$, we have

$$\begin{aligned} J_1(\theta) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_0}{\sigma/\sqrt{n}})}, J_2(\theta) = \frac{1-\frac{\alpha}{2} - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}, \\ J_3(\theta) &= \frac{1-\frac{\alpha}{2} - \Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}, J_4(\theta) = \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}. \end{aligned}$$

Therefore,

$$\begin{aligned} \Delta(\theta) &= J_2(\theta) - J_1(\theta) + J_4(\theta) - J_3(\theta) \\ &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_1)} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_0)} \\ &= \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_1)} - \frac{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_0)}. \end{aligned}$$

Define

$$\begin{aligned} \Delta_1(\theta) &= \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_1)} - \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})}, \\ \Delta_2(\theta) &= \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})}, \end{aligned}$$

then $\Delta(\theta) > \Delta_1(\theta) + \Delta_2(\theta)$.

Here $\Delta_1(\theta)$ can also be expressed as

$$\Delta_1(\theta) = \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_1)\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} \cdot \left[-\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})\right]. \tag{2.28}$$

We know $\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) < 1-\alpha$ since $\theta < c_2 < \nu_3$, and $P_s(\theta,\nu_1) > \Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) > \frac{1}{2}$ since $\theta > \nu_1$. Besides, the condition that $\nu_3 > c_2$ provides $\nu_1 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} > \sqrt{\lambda} + \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, so $\nu_1 > \sqrt{\lambda} - \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ and further $\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}}) < \Phi(-\frac{3}{2}z_{\alpha/2})$. Based on these conclusions, we have $|\Delta_1(\theta)| < 4(1-\alpha)\Phi(-\frac{3}{2}z_{\alpha/2})$.

It is straightforward to show that $\Delta_2(\theta) > 0$. Together we know $\Delta(\theta) > -4(1-\alpha)\Phi(-\frac{3}{2}z_{\alpha/2})$.

When $\theta \in [c_2, \nu_3]$,

$$\begin{aligned} J_1(\theta) &= \frac{1-2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_0}{\sigma/\sqrt{n}})}, J_2(\theta) = \frac{1-\frac{\alpha}{2} - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}, \\ J_3(\theta) &= \frac{1-\frac{\alpha}{2} - \Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}, J_4(\theta) = \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}. \end{aligned}$$

Therefore,

$$\begin{aligned} \Delta(\theta) &= J_2(\theta) - J_1(\theta) + J_4(\theta) - J_3(\theta) \\ &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_1)} - \frac{1-2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_0)} \\ &= \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_1)} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_0)}. \end{aligned}$$

Define

$$\begin{aligned} \Delta_1(\theta) &= \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_1)} - \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})}, \\ \Delta_2(\theta) &= \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})}, \end{aligned}$$

then $\Delta(\theta) > \Delta_1(\theta) + \Delta_2(\theta)$.

In fact, $\Delta_1(\theta)$ can be written exactly as that in (2.28). Here $\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}}) < \Phi(-2z_{\alpha/2})$ since $\frac{-\theta-\nu_1}{\sigma/\sqrt{n}} \le \frac{-c_2-\nu_1}{\sigma/\sqrt{n}} < \frac{\sqrt{n}}{\sigma}\left[-(\sqrt{\lambda} + \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) - (\sqrt{\lambda} + \frac{1}{2}z_{\alpha/2})\frac{\sigma}{\sqrt{n}}\right] \le -2z_{\alpha/2}$, $\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) < 1-\alpha$, and $\Phi(\frac{1}{2}z_{\alpha/2}) < \Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) < P_s(\theta,\nu_1)$. Combining these conclusions we have $\Delta_1(\theta) < 0$ and $|\Delta_1(\theta)| < \frac{1-\alpha}{\left[\Phi(\frac{1}{2}z_{\alpha/2})\right]^2}\Phi(-2z_{\alpha/2})$.

In addition,

$$\Delta_2(\theta) = \left[1 - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})}\right] + \left[\frac{1}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})} - \frac{1}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})}\right]\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) > 0,$$

since $\frac{\theta-\nu_0}{\sigma/\sqrt{n}} > z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}$ when $\theta > c_2$, according to Lemma 2.7.3.

Therefore, $\Delta(\theta) > -\frac{1-\alpha}{\left[\Phi(\frac{1}{2}z_{\alpha/2})\right]^2}\Phi(-2z_{\alpha/2})$. The right hand side is quite close to zero. When $\alpha \leq 0.1, \Phi(-2z_{\alpha/2}) < 5 \cdot 10^{-4}$.

When $\theta \in [\nu_3, c_4], [c_4, \nu_4]$, and $[\nu_4, \infty)$. the expressions of $\Delta(\theta)$ are exactly the same with those for case 1 in $[c_2, c_4]$, thus the same lower bounds can be provided.

Case 3: $\nu_3 < c_3 < c_2$. The case is illustrated in Figure 2.6 and Figure 2.16.

When $\theta \in [\nu_0, \nu_3]$,

$$\begin{aligned}
J_1(\theta) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_0}{\sigma/\sqrt{n}})}, \\
J_2(\theta) &= \frac{1-\frac{\alpha}{2} - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}, \\
J_3(\theta) &= \frac{1-\frac{\alpha}{2} - \Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Delta(\theta) &= J_2(\theta) - J_1(\theta) - J_3(\theta) \\
&= \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_0}{\sigma/\sqrt{n}})},
\end{aligned}$$

a same lower bound can be provided in the same way with case 1 when $\theta \in [\nu_0, c_3]$.

When $\theta \in [\nu_3, c_3]$,

$$\begin{aligned}
J_1(\theta) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_0}{\sigma/\sqrt{n}})}, \\
J_2(\theta) &= \frac{1-\alpha}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}, \\
J_3(\theta) &= \frac{1-\frac{\alpha}{2} - \Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}.
\end{aligned}$$

Thus

$$\begin{aligned}\Delta(\theta) &= J_2(\theta) - J_1(\theta) - J_3(\theta)\\ &= \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \frac{\alpha}{2}}{P_s(\theta,\nu_1)} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_0)}.\end{aligned}$$

Further,

$$\begin{aligned}\Delta(\theta) &> \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \frac{\alpha}{2}}{P_s(\theta,\nu_1)} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})}\\ &> \Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}}) - \frac{\alpha}{2} - 1 + \frac{\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})}\\ &> \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \frac{\alpha}{2} - 1 + \frac{\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})},\end{aligned}$$

where the third inequality holds since $\nu_2 - \theta \geq z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{n}$ when $\theta \leq c_3$ by Lemma 2.7.2. Define the right hand side of above inequality as $\Delta_1(\theta)$, which equals:

$$\Delta_1(\theta) = \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \frac{\alpha}{2} - 1 + \frac{\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})}. \tag{2.29}$$

Then

$$\begin{aligned}\Delta_1(\theta) &> \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \frac{\alpha}{2} - 1 + \frac{\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}\\ &= \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) + \frac{1}{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})} - 2 - \frac{\alpha}{2}\\ &> -\frac{\alpha(1-\alpha)}{2-\alpha} > -\frac{\alpha}{2},\end{aligned}$$

the first inequality holds since $\theta - \nu_0 < z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{n}$ when $\theta < c_3 < c_2$, based on Lemma 2.7.3, and the second inequality holds since $\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) < 1 - \frac{\alpha}{2}$. Therefore $\Delta(\theta) > -\frac{\alpha}{2}$ when $\theta \in [\nu_3, c_3]$.

When $\theta \in [c_3, c_2]$,

$$
\begin{aligned}
J_1(\theta) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_0}{\sigma/\sqrt{n}})}, \\
J_2(\theta) &= \frac{1-\alpha}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}, \\
J_3(\theta) &= \frac{1-\frac{\alpha}{2} - \Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}, \\
J_4(\theta) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu\nu_2-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) + \Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\Delta(\theta) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \frac{\alpha}{2}}{P_s(\theta,\nu_1)} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_0)}, \\
\text{and } \Delta(\theta) &> \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \frac{\alpha}{2} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\nu_0-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_0}{\sigma/\sqrt{n}})}.
\end{aligned}
$$

Note that the right hand side of the above inequality is exact the same with $\Delta_1(\theta)$ in (2.29), thus we have a lower bound of $\Delta(\theta)$ on $[c_3, c_2]$ given by $\Delta(\theta) > -\frac{\alpha(1-\alpha)}{2-\alpha} > -\frac{\alpha}{2}$.

When $\theta \in [c_2, c_4]$, $[c_4, \nu_4]$, $[\nu_4, +\infty)$, the same lower bounds can be provided with those for case 1 and case 2.

Proof of Theorem 2.2.2

(i). When $\theta \in [0, \min\{\nu_3, c_1\}]$, we know $CR(\theta)$ is increasing based on Lemma 2.7.5, together with $CR_1(\theta) = 0$, we have $\Delta(\theta) \geq \Delta(\theta = 0) = 1 - \frac{\alpha}{\tau}$.

(ii). When $\theta \in [\min\{\nu_3, c_1\}, \nu_0]$, we discuss over two different cases: $\nu_3 < c_1$ and $\nu_3 \geq c_1$ separately.

Case 4: $\nu_3 < c_1$. The case is illustrated in Figure 2.7 and Figure 2.17.

When $\theta \in [\nu_3, \nu_0]$,

$$
\Delta(\theta) = J_2(\theta) - J_1(\theta).
$$

Here $J_2(\theta)$ is decreasing according to Lemma 2.7.5, and $J_1(\theta)$ is non-decreasing according to Lemma 2.7.4.

Therefore,

$$\begin{aligned}
\Delta(\theta) &\geq J_2(\theta = \nu_0) - J_1(\theta = \nu_0),\\
\text{where } J_2(\theta = \nu_0) &= \frac{1-\alpha}{P_s(\nu_0, \nu_1)},\\
\text{and } J_1(\theta = \nu_0) &= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\nu_0)}{\sigma}) - \Phi(\frac{\sqrt{n}}{\sigma}(\nu_0 - \nu_0))}{P_s(\nu_0, \nu_0)} = \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\nu_0)}{\sigma}) - \frac{1}{2}}{P_s(\nu_0, \nu_0)}.
\end{aligned}$$

Since $P_s(\nu_0, \nu_1) < 1$, it holds true that $J_2(\theta = \nu_0) > 1 - \alpha$. Besides, $J_1(\theta = \nu_0) = \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\nu_0)}{\sigma}) - \frac{1}{2}}{P_s(\nu_0, \nu_0)} < 2\Phi(\frac{1}{2}z_{\alpha/2}) - 1$, where we use that $\frac{\tilde{\sigma}(\nu_0)}{\sigma} = \frac{1}{2}$ and $P_s(\nu_0, \nu_0) > \frac{1}{2}$. Therefore $\Delta(\theta) \geq 1 - \alpha - (2\Phi(\frac{1}{2}z_{\alpha/2}) - 1) = 2 - \alpha - 2\Phi(\frac{1}{2}z_{\alpha/2})$.

Case 5: $\nu_3 \geq c_1$. The case is illustrated in Figure 2.8 and Figure 2.18.

When $\theta \in [c_1, \nu_3]$, we have

$$\begin{aligned}
\Delta(\theta) &= \frac{1 - \frac{\alpha}{2} - \Phi(\frac{\sqrt{n}}{\sigma}(\nu_1 - \theta))}{P_s(\theta, \nu_1)} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\sqrt{n}}{\sigma}(\nu_0 - \theta))}{P_s(\theta, \nu_0)}\\
&= \frac{1 - \frac{\alpha}{2} - \Phi(\frac{\sqrt{n}}{\sigma}(\nu_1 - \theta))}{P_s(\theta, \nu_1)} - \frac{\Phi(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0)) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta, \nu_0)}.
\end{aligned}$$

The first component of $\Delta(\theta)$ follows:

$$\begin{aligned}
&\frac{1 - \frac{\alpha}{2} - \Phi(\frac{\sqrt{n}}{\sigma}(\nu_1 - \theta))}{P_s(\theta, \nu_1)} > 1 - \frac{\alpha}{2} - \Phi(\frac{\sqrt{n}}{\sigma}(\nu_1 - \theta))\\
&> 1 - \frac{\alpha}{2} - \Phi(\frac{\sqrt{n}}{\sigma}(\nu_1 - c_1)) > 1 - \frac{\alpha}{2} - \Phi(-\frac{1}{2}z_{\alpha/2}),
\end{aligned}$$

the third inequality holds since $\sqrt{\lambda} \geq (z_{\alpha/2} + z_{\tau/2})\frac{\sigma}{\sqrt{n}}$, and further $\nu_1 - c_1 = \nu_1 - (\sqrt{\lambda} - z_{\alpha/2}\frac{\tilde{\sigma}(c_1)}{\sqrt{n}}) < \nu_1 - \sqrt{\lambda} + \frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < -\frac{1}{2}z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

The second component of $\Delta(\theta)$ follows:

$$\begin{aligned}
&\frac{\Phi(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0)) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta, \nu_0)} < \frac{\Phi(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0)) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0))}\\
= \;& 1 - \frac{\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0))} < 1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) < 1 - 2\Phi(-\frac{1}{2}z_{\alpha/2}),
\end{aligned}$$

since $\Phi(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0)) < \frac{1}{2}$ and $\tilde{\sigma}(\theta) < \frac{\sigma}{2}$ when $\theta < \nu_0$.

Combining the bounds for the first and second component of $\Delta(\theta)$, we have $\Delta(\theta) > 1 - \frac{\alpha}{2} - \Phi(-\frac{1}{2}z_{\alpha/2}) - (1 - 2\Phi(-\frac{1}{2}z_{\alpha/2})) > \Phi(-\frac{1}{2}z_{\alpha/2}) - \frac{\alpha}{2}$.

When $\theta \in [\nu_3, \nu_0]$,

$$\Delta(\theta) = \frac{1-\alpha}{P_s(\theta, \nu_1)} - \frac{\Phi(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0)) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta, \nu_0)},$$

with the same argument as in case 4 when $\theta \in [\nu_3, \nu_0]$, it holds true that $\Delta(\theta) > 1 - \frac{\alpha}{2} - \Phi(-\frac{1}{2}z_{\alpha/2})$.

When $\theta \in [\nu_0, c_3]$,

$$\begin{aligned} CR(\theta) &= J_2(\theta) - J_3(\theta), \\ CR_1(\theta) &= J_1(\theta), \end{aligned}$$

and thus

$$\Delta(\theta) = \frac{\Phi(\frac{\sqrt{n}}{\sigma}(\nu_2 - \theta)) - \frac{\alpha}{2}}{P_s(\theta, \nu_1)} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\sqrt{n}}{\sigma}(\nu_0 - \theta))}{P_s(\theta, \nu_0)}.$$

Since $J_1(\theta)$ increases according to Lemma 2.7.4, $J_2(\theta)$ decreases, and $J_3(\theta)$ increases on $[\nu_0, c_3]$ according to Lemma 2.7.5, $\Delta(\theta)$ is a decreasing function within the above interval. Therefore,

$$\Delta(\theta) \geq \Delta(\theta = c_3),$$

and

$$\begin{aligned} \Delta(\theta = c_3) &> \Phi(\frac{\sqrt{n}}{\sigma}(\nu_2 - c_3)) - \frac{\alpha}{2} - \frac{\Phi(\frac{\sqrt{n}}{\sigma}(c_3 - \nu_0)) - \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(c_3)}{\sigma})}{\Phi(\frac{\sqrt{n}}{\sigma}(c_3 - \nu_0))} \\ &= \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(c_3)}{\sigma}) - \frac{\alpha}{2} - 1 + \frac{\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(c_3)}{\sigma})}{\Phi(\frac{\sqrt{n}}{\sigma}(c_3 - \nu_0))} \\ &> \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(c_3)}{\sigma}) - \frac{\alpha}{2} - 1 + \frac{\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(c_3)}{\sigma})}{\Phi(\frac{1}{2}z_{\alpha/2})} \\ &= \Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(c_3)}{\sigma})\left[\frac{1}{\Phi(\frac{1}{2}z_{\alpha/2})} - 1\right] - \frac{\alpha}{2} > -\frac{\alpha}{2}, \end{aligned}$$

where we use that $\nu_2 - c_3 = z_{\alpha/2}\frac{\tilde{\sigma}(c_3)}{\sqrt{n}}$ based on the definition of c_3, and $\frac{\sqrt{n}}{\sigma}(c_3 - \nu_0) < \frac{1}{2}z_{\alpha/2}$ based on Lemma 2.7.2. Therefore $\Delta(\theta) > -\frac{\alpha}{2}$ when $\theta \in [\nu_0, c_3]$.

When $\theta \in [c_3, c_2]$,

$$\begin{aligned}
\Delta(\theta) &= J_2(\theta) - J_1(\theta) + J_4(\theta) - J_3(\theta) \\
&= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \frac{\alpha}{2}}{P_s(\theta,\nu_1)} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\sqrt{n}}{\sigma}(\nu_0-\theta))}{P_s(\theta,\nu_0)} \\
&> \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \frac{\alpha}{2} - \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \Phi(\frac{\sqrt{n}}{\sigma}(\nu_0-\theta))}{\Phi(\frac{\sqrt{n}}{\sigma}(\theta-\nu_0))} \\
&= \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) + \frac{\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\sqrt{n}}{\sigma}(\theta-\nu_0))} - \frac{\alpha}{2} - 1 \\
&> \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) + \frac{1}{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})} - 2 - \frac{\alpha}{2} \\
&> -\frac{\alpha(1-\alpha)}{2-\alpha} > -\frac{\alpha}{2},
\end{aligned}$$

where the second inequality holds since $\theta < \nu_0 + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$ when $\theta < c_2$, based on Lemma 2.7.3. Therefore $\Delta(\theta) > -\frac{\alpha}{2}$ when $\theta \in [c_3, c_2]$.

When $\theta \in [c_2, c_4]$,

$$\begin{aligned}
\Delta(\theta) &= J_2(\theta) - J_1(\theta) + J_4(\theta) - J_3(\theta) \\
&= \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \frac{\alpha}{2}}{P_s(\theta,\nu_1)} - \frac{1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_0)} \\
&> \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \frac{\alpha}{2} - \frac{1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\sqrt{n}}{\sigma}(\theta-\nu_0))} \\
&> \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) - \frac{\alpha}{2} - \frac{1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})} \\
&= \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma}) + \frac{1}{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})} - 2 - \frac{\alpha}{2} \\
&> -\frac{\alpha(1-\alpha)}{2-\alpha} > -\frac{\alpha}{2},
\end{aligned}$$

where the second inequality holds since $\theta > \nu_0 + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}}$ when $\theta > c_2$ by Lemma 2.7.3. Therefore $\Delta(\theta) > -\frac{\alpha}{2}$ when $\theta \in [c_2, c_4]$.

When $\theta \in [c_4, \nu_4]$,

$$\begin{aligned}\Delta(\theta) &= J_2(\theta) - J_1(\theta) + J_4(\theta) - J_3(\theta) \\ &= \frac{1-\alpha}{P_s(\theta,\nu_1)} - \frac{1 - \frac{\alpha}{2} - \Phi(\frac{\sqrt{n}}{\sigma}(\nu_2 - \theta))}{P_s(\theta,\nu_1)} + \frac{1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_1)} - \frac{1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_0)} \\ &= \frac{\Phi(\frac{\sqrt{n}}{\sigma}(\nu_2 - \theta)) - \frac{\alpha}{2}}{P_s(\theta,\nu_1)} + \left[1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})\right]\left[\frac{1}{P_s(\theta,\nu_1)} - \frac{1}{P_s(\theta,\nu_0)}\right] \\ &> \frac{1}{P_s(\theta,\nu_1)} - \frac{1}{P_s(\theta,\nu_0)} > 1 - \frac{1}{P_s(\theta,\nu_0)} \\ &> 1 - \frac{1}{\Phi(\frac{3}{2}z_{\alpha/2})} = -\frac{\Phi(-\frac{3}{2}z_{\alpha/2})}{\Phi(\frac{3}{2}z_{\alpha/2})},\end{aligned}$$

where the first inequality holds since $\Phi(\frac{\sqrt{n}}{\sigma}(\nu_2 - \theta)) - \frac{\alpha}{2} > 0$, when $\theta < \nu_4$; the third inequality holds since $\theta - \nu_0 > z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sqrt{n}} > \frac{3}{2}z_{\alpha}\frac{\sigma}{\sqrt{n}}$ when $\theta > c_4$, and thus $P_s(\theta,\nu_0) > \Phi(\frac{3}{2}z_{\alpha/2})$. Therefore, $\Delta(\theta) > -\frac{\Phi(-\frac{3}{2}z_{\alpha/2})}{\Phi(\frac{3}{2}z_{\alpha/2})}$ when $\theta \in [c_4, \nu_4]$.

When $\theta \in [\nu_4, +\infty)$, we have

$$\begin{aligned}\Delta(\theta) &= \frac{1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_1)} - \frac{1 - 2\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{P_s(\theta,\nu_0)} \\ &> \frac{1}{P_s(\theta,\nu_1)} - \frac{1}{P_s(\theta,\nu_0)} \\ &> 1 - \frac{1}{\Phi(\frac{\sqrt{n}}{\sigma}(\theta - \nu_0))} > -\frac{\Phi(-2z_{\alpha/2})}{\Phi(2z_{\alpha/2})},\end{aligned}$$

where the last inequality holds since $\theta - \nu_0 > 2z_{\alpha}\frac{\sigma}{\sqrt{n}}$ when $\theta > \nu_4$. In fact, $P_s(\theta,\nu_1) \approx P_s(\theta,\nu_0)$ when θ becomes quite large, and further $\Delta(\theta) \approx 0$.

The Bias-corrected confidence interval

The penalized likelihood function can be written as

$$L(\boldsymbol{\theta}) = \frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|^2 + \sum_{j=1}^{p} p_\lambda(|\theta_j|) = \frac{1}{2n}\|\boldsymbol{y} - \sum_{j=1}^{p}\boldsymbol{X_j}\theta_j\|^2 + \lambda_n\sum_{j=1}^{p}\hat{w}_j|\theta_j|,$$

and its first derivative equals:

$$\frac{\partial L(\boldsymbol{\theta})}{\partial\theta_j} = -\frac{1}{n}\boldsymbol{X_j}^T(\boldsymbol{y} - \sum_{j=1}^{p}\boldsymbol{X_j}\theta_j) + \lambda_n\hat{w}_j sgn(\theta_j), j = 1, \cdots, p.$$

We know $\hat{\boldsymbol{\theta}}^{\boldsymbol{AL}}$ satisfies

$$\left.\frac{\partial L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}^{\boldsymbol{AL}}} = 0,$$

which can be rewritten as

$$\frac{1}{n}\boldsymbol{X}^T(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\theta}}^{\boldsymbol{AL}})-\boldsymbol{\Gamma}^{\boldsymbol{AL}}=0, \tag{2.30}$$

where $\boldsymbol{\Gamma}^{\boldsymbol{AL}}=(\lambda_n\hat{w}_1 sgn(\hat{\theta}_1^{AL}),\cdots,\lambda_n\hat{w}_p sgn(\hat{\theta}_p^{AL}))^T$. When n is large enough, it satisfies $\hat{\theta}_j^{AL}=0$ for $j\in\mathcal{A}^c$, we can rewrite (2.30) for $\theta_j, j\in\mathcal{A}$ as

$$\frac{1}{n}\boldsymbol{X}_{\boldsymbol{\mathcal{A}}}^T(\boldsymbol{y}-\boldsymbol{X}_{\boldsymbol{\mathcal{A}}}\hat{\boldsymbol{\theta}}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{AL}})-\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{AL}}=0. \tag{2.31}$$

A Taylor expansion of (2.31) at $\boldsymbol{\theta}^*_{\boldsymbol{\mathcal{A}}}$ provides

$$\frac{1}{n}\boldsymbol{X}_{\boldsymbol{\mathcal{A}}}^T(\boldsymbol{y}-\boldsymbol{X}_{\boldsymbol{\mathcal{A}}}\boldsymbol{\theta}^*_{\boldsymbol{\mathcal{A}}})-\frac{1}{n}(\boldsymbol{X}_{\boldsymbol{\mathcal{A}}}^T\boldsymbol{X}_{\boldsymbol{\mathcal{A}}})(\hat{\boldsymbol{\theta}}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{AL}}-\boldsymbol{\theta}^*_{\boldsymbol{\mathcal{A}}})-\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{AL}}=0,$$

and further it holds that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\mathcal{A}}}^{\boldsymbol{AL}}-\boldsymbol{\theta}^*_{\boldsymbol{\mathcal{A}}})+(\frac{1}{n}\boldsymbol{X}_{\boldsymbol{\mathcal{A}}}^T\boldsymbol{X}_{\boldsymbol{\mathcal{A}}})^{-1}\cdot\sqrt{n}\boldsymbol{\Gamma}^{\boldsymbol{AL}}\to N(0,\boldsymbol{C}_{\boldsymbol{11}}^{-1}\sigma^2). \tag{2.32}$$

Based on (2.32), the bias term $\widehat{\boldsymbol{b}}_{\mathcal{A}}^{AL}$ is given by

$$\widehat{\boldsymbol{bias}}_{\mathcal{A}}^{AL}=(\frac{1}{n}\boldsymbol{X}_{\boldsymbol{\mathcal{A}}}^T\boldsymbol{X}_{\boldsymbol{\mathcal{A}}})^{-1}(\lambda_n\hat{w}_1 sgn(\hat{\theta}_1^{AL}),\cdots,\lambda_n\hat{w}_d sgn(\hat{\theta}_q^{AL}))^T,$$

and the bias term $\widehat{\boldsymbol{b}}_{\mathcal{A}}^{SCAD}$ is given by

$$\widehat{\boldsymbol{bias}}_{\mathcal{A}}^{SCAD}=(\frac{1}{n}\boldsymbol{X}_{\boldsymbol{\mathcal{A}}}^T\boldsymbol{X}_{\boldsymbol{\mathcal{A}}}+\boldsymbol{\Omega})^{-1}(p_\lambda'(|\hat{\theta}_1|)sgn(|\hat{\theta}_1|),\cdots,p_\lambda'(|\hat{\theta}_d|)sgn(|\hat{\theta}_q|))^T.$$

We also compare the coverage rate obtained from the proposed method with the nominal level $1-\alpha$. We first provide some properties about $CR(\theta)$ in the following lemma.

Lemma 2.7.7 *Under conditions (C1)-(C2), given n,σ, tuning parameter λ and proper selection of τ, when $\theta\in[0,\nu_0]$, $CR(\theta)$ is an increasing function. In addition, when $\theta\in[\nu_0,\max\{c_3,\nu_3\}]$, $CR(\theta)$ can be approximated by a piece-wise monotonic function $\widetilde{CR}(\theta)$, and it satisfies that*

$$|\widetilde{CR}(\theta)-CR(\theta)|<4(1-\alpha)\Phi(-\frac{3}{2}z_{\alpha/2}).$$

The monotonicity of $\widetilde{CR}(\theta)$ can be displayed in two cases:

a) if $c_3 < \nu_3$, $\widetilde{CR}(\theta)$ *is decreasing on* $[\nu_0, c_3]$, *and increasing on* $[c_3, \nu_3]$*;*

b) if $c_3 \geq \nu_3$, $\widetilde{CR}(\theta)$ *is decreasing on* $[\nu_0, c_3]$.

The explicit expression for $\widetilde{CR}(\theta)$ is provided in the proof of Lemma 2.7.7. In view of lemma 2.7.7, $CR(\theta)$ is roughly piece-wise monotonic up to a small uniformly bound $4(1-\alpha)\Phi(-\frac{3}{2}z_{\alpha/2})$. We will utilize its approximate piece-wise monotonicity to quantify the difference between $CR(\theta)$ and the nominal confidence level $1-\alpha$. Theorem 2.2.3 provides the upper bound for their difference when $\theta \geq c_1$.

Proof of Lemma 2.7.7

When $\theta \in [0, \nu_0]$, $CR(\theta) = J_2(\theta) = \frac{CR_b(\theta,\nu_1)}{P_s(\theta,\nu_1)}$, $CR(\theta)$ is increasing immediately from Lemma 2.7.5.

When $\theta \in [\nu_0, c_4]$, based on Lemma 2.7.6, we write the expression of $CR(\theta)$ explicitly according to the following two cases that $c_3 < \nu_3$ and $c_3 \geq \nu_3$.

i). if $c_3 < \nu_3$, then

$$CR(\theta) = \begin{cases} \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})-\Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_1)} & \text{if } \theta \in [\nu_0, c_3]\,, \\ \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})-\Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_1)} & \text{if } \theta \in [c_3, \nu_3]\,, \\ \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})-\frac{\alpha}{2}}{P_s(\theta,\nu_1)} & \text{if } \theta \in [\nu_3, c_4]\,. \end{cases}$$

Define $\widetilde{CR}(\theta)$ as:

$$\widetilde{CR}(\theta) = \begin{cases} \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})-\Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} & \text{if } \theta \in [\nu_0, c_3]\,, \\ \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})-\Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} & \text{if } \theta \in [c_3, \nu_3]\,, \\ \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})-\frac{\alpha}{2}}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} & \text{if } \theta \in [\nu_3, c_4]\,. \end{cases} \tag{2.33}$$

We show that $\widetilde{CR}(\theta)$ is a piece-wise monotonic function, and the difference between $\widetilde{CR}(\theta)$ and $CR(\theta)$ satisfies $|\widetilde{CR}(\theta) - CR(\theta)| < 4(1-\alpha)\Phi(-\frac{3}{2}z_{\alpha/2})$. Firstly, when $\theta \in [\nu_0, c_3]$,

$$\widetilde{CR}(\theta) = 1 - \frac{\Phi(\frac{\theta-\nu_2}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})}.$$

For any fixed $y > 0$, $\frac{\Phi(x-y)}{\Phi(x)}$ is an increasing function of x, where $x \in R$, thus $\widetilde{CR}(\theta)$ is decreasing of θ. Moreover, $CR(\theta) - \widetilde{CR}(\theta) = \frac{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})-\Phi(\frac{\theta-\nu_2}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_1)\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} \cdot \left[-\Phi(-\frac{\theta+\nu_1}{\sigma/\sqrt{n}})\right]$, it can be easily shown that $\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) - \Phi(\frac{\theta-\nu_2}{\sigma/\sqrt{n}})$ is increasing on $\big(-\infty, \frac{1}{2}(\nu_1+\nu_2)\big)$, and decreasing on $\big(\frac{1}{2}(\nu_1+\nu_2), +\infty\big)$, thus $\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) -$

$\Phi(\frac{\theta-\nu_2}{\sigma/\sqrt{n}}) \leq 2\Phi(\frac{\frac{1}{2}(\nu_1+\nu_2)}{\sigma/\sqrt{n}})-1 < 1-\alpha$. Together with $P_s(\theta,\nu_1) > \Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) > \frac{1}{2}$, $\Phi(-\frac{\theta+\nu_1}{\sigma/\sqrt{n}}) < \Phi(-\frac{3}{2}z_{\alpha/2})$ and $\theta > \nu_0$, it holds true that $|CR(\theta) - \widetilde{CR}(\theta)| < 4(1-\alpha)\Phi(-\frac{3}{2}z_{\alpha/2})$.

Secondly, when $\theta \in [c_3, \nu_3]$,

$$\widetilde{CR}(\theta) = 1 - \frac{\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})}.$$

Since $\Phi(-z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})$ is decreasing and $\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})$ is increasing w.r.t θ, $\widetilde{CR}(\theta)$ is increasing on $[c_3, \nu_3]$. The upper bound for $|CR(\theta) - \widetilde{CR}(\theta)|$ can be proved in a similar fashion.

ii). if $c_3 \geq \nu_3$, then

$$CR(\theta) = \begin{cases} \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})-\Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{P_s(\theta,\nu_1)} & \text{if } \theta \in [\nu_0, \nu_3], \\ \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})-\frac{\alpha}{2}}{P_s(\theta,\nu_1)} & \text{if } \theta \in [\nu_3, c_3], \\ \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})-\frac{\alpha}{2}}{P_s(\theta,\nu_1)} & \text{if } \theta \in [c_3, c_4]. \end{cases}$$

We define

$$\widetilde{CR}(\theta) = \begin{cases} \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})-\Phi(\frac{\nu_1-\theta}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} & \text{if } \theta \in [\nu_0, \nu_3], \\ \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})-\frac{\alpha}{2}}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} & \text{if } \theta \in [\nu_3, c_3], \\ \frac{\Phi(z_{\alpha/2}\frac{\tilde{\sigma}(\theta)}{\sigma})-\frac{\alpha}{2}}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} & \text{if } \theta \in [c_3, c_4]. \end{cases} \tag{2.34}$$

Similar to case i), we show that $\widetilde{CR}(\theta)$ is a piece-wise monotonic function, and the difference between $\widetilde{CR}(\theta)$ and $CR(\theta)$ satisfies $|\widetilde{CR}(\theta) - CR(\theta)| < 4(1-\alpha)\Phi(-\frac{3}{2}z_{\alpha/2})$. When $\theta \in [\nu_0, \nu_3]$,

$$\widetilde{CR}(\theta) = 1 - \frac{\Phi(\frac{\theta-\nu_2}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})},$$

for any fixed $y > 0$, $\frac{\Phi(x-y)}{\Phi(x)}$ is an increasing function of $x \in R$, thus $\widetilde{CR}(\theta)$ is decreasing of θ. The upper bound for $|CR(\theta) - \widetilde{CR}(\theta)|$ can be shown in the same fashion with that in case i) when $\theta \in [\nu_0, c_3]$.

When $\theta \in [\nu_3, c_3]$,

$$\widetilde{CR}(\theta) = \frac{1-\frac{\alpha}{2}}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} - \frac{\Phi(\frac{\theta-\nu_2}{\sigma/\sqrt{n}})}{\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})},$$

it can be checked easily that $\widetilde{CR}(\theta)$ is a decreasing function on this interval. Moreover, $CR(\theta) - \widetilde{CR}(\theta) = \frac{\Phi(\frac{\nu_2-\theta}{\sigma/\sqrt{n}})-\frac{\alpha}{2}}{P_s(\theta,\nu_1)\Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}})} \cdot \left[-\Phi(\frac{-\theta-\nu_1}{\sigma/\sqrt{n}})\right]$. Since $\frac{\nu_2-\theta}{\sigma/\sqrt{n}} < \frac{\nu_2-\nu_3}{\sigma/\sqrt{n}} < z_{\alpha/2}$, $\frac{-\nu_1-\theta}{\sigma/\sqrt{n}} < \frac{-\nu_1-\nu_3}{\sigma/\sqrt{n}} < -2z_{\alpha/2}$, together with $P_s(\theta,\nu_1) > \Phi(\frac{\theta-\nu_1}{\sigma/\sqrt{n}}) > 1-\frac{\alpha}{2}$, we have $|CR(\theta)-\widetilde{CR}(\theta)| < \frac{4(1-\alpha)}{(2-\alpha)^2}\Phi(-2z_{\alpha/2}) < 4(1-\alpha)\Phi(-\frac{3}{2}z_{\alpha/2})$.

Combining the conclusions of case i) and case ii), it holds true that $\widetilde{CR}(\theta)$ is a piece-wise monotonic function, and the difference between $\widetilde{CR}(\theta)$ and $CR(\theta)$ can be bounded from above by a small value.

Proof of Theorem 2.2.3

Based on lemma 2.7.7, $\min_{\theta\geq\nu_1}\widetilde{CR}(\theta) = \widetilde{CR}(c_3)$. We discuss the bound for $1-\alpha-\widetilde{CR}(c_3)$ separately for two cases: $c_3 < \nu_3$ and $c_3 > \nu_3$.

If $c_3 < \nu_3$,

$$\widetilde{CR}(c_3) = 1 - \frac{\Phi(\frac{c_3-\nu_2}{\sigma/\sqrt{n}})}{\Phi(\frac{c_3-\nu_1}{\sigma/\sqrt{n}})}, \text{ thus } (1-\alpha)-\widetilde{CR}(c_3) = \frac{\Phi(\frac{c_3-\nu_2}{\sigma/\sqrt{n}})}{\Phi(\frac{c_3-\nu_1}{\sigma/\sqrt{n}})} - \alpha.$$

For any $a, b > 0, a > b$, the function $\frac{\Phi(x-a)}{\Phi(x-b)}$ is an increasing function of x, since $c_3 \in \left(\sqrt{\lambda}, \sqrt{\lambda}+\frac{1}{2}z_{\alpha/2}\sigma/\sqrt{n}\right)$, it holds true that

$$\frac{\Phi(\frac{c_3-\nu_2}{\sigma/\sqrt{n}})}{\Phi(\frac{c_3-\nu_1}{\sigma/\sqrt{n}})} < \frac{\Phi(\frac{\sqrt{\lambda}+\frac{1}{2}z_{\alpha/2}\sigma/\sqrt{n}-\nu_2}{\sigma/\sqrt{n}})}{\Phi(\frac{\sqrt{\lambda}+\frac{1}{2}z_{\alpha/2}\sigma/\sqrt{n}-\nu_1}{\sigma/\sqrt{n}})} = \frac{\Phi(-\frac{1}{2}z_{\alpha/2})}{\Phi(\frac{\sqrt{n\lambda}}{\sigma}+\frac{1}{2}z_{\alpha/2}-z_{\tau/2})} < \frac{\Phi(-\frac{1}{2}z_{\alpha/2})}{\Phi(\frac{1}{2}z_{\alpha/2})},$$

where in the last inequality we have used that $\frac{\sqrt{n\lambda}}{\sigma}+\frac{1}{2}z_{\alpha/2}-z_{\tau/2} \geq \frac{3}{2}z_{\alpha/2}-z_{\tau/2} \geq \frac{1}{2}z_{\alpha/2}$. Similarly, it holds true that

$$\frac{\Phi(\frac{c_3-\nu_2}{\sigma/\sqrt{n}})}{\Phi(\frac{c_3-\nu_1}{\sigma/\sqrt{n}})} > \frac{\Phi(\frac{\sqrt{\lambda}-\nu_2}{\sigma/\sqrt{n}})}{\Phi(\frac{\sqrt{\lambda}-\nu_1}{\sigma/\sqrt{n}})} = \frac{\frac{\alpha}{2}}{\Phi(\frac{\sqrt{\lambda}-\nu_1}{\sigma/\sqrt{n}})} \geq \frac{\alpha}{2-\alpha},$$

where in the last inequality we have used that $\frac{\sqrt{\lambda}-\nu_1}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$. Therefore,

$$(1-\alpha)-\widetilde{CR}(c_3) \in \left(\frac{\alpha(\alpha-1)}{2-\alpha}, \frac{\Phi(-\frac{1}{2}z_{\alpha/2})}{\Phi(\frac{1}{2}z_{\alpha/2})}-\alpha\right).$$

If $c_3 > \nu_3$,

$$\widetilde{CR}(c_3) = \frac{\Phi(\frac{\nu_2-c_3}{\sigma/\sqrt{n}})-\frac{\alpha}{2}}{\Phi(\frac{c_3-\nu_1}{\sigma/\sqrt{n}})}.$$

We know $\Phi(\frac{\nu_2-c_3}{\sigma/\sqrt{n}}) = \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(c_3)}{\sigma}) < \Phi(\frac{9}{13}z_{\alpha/2})$, which is because $\tilde{\sigma}(c_3) < \frac{9}{13}\sigma$ given $c_3 < \sqrt{\lambda}+\frac{1}{2}z_{\alpha/2}\sigma/\sqrt{n} <$

$\frac{3}{2}\sqrt{\lambda}$. Moreover, $\Phi(\frac{c_3-\nu_1}{\sigma/\sqrt{n}}) > \Phi(\frac{\nu_3-\nu_1}{\sigma/\sqrt{n}}) = 1-\frac{\alpha}{2}$, together we have

$$\widetilde{CR}(c_3) \leq \frac{2\Phi(\frac{9}{13}z_{\alpha/2}) - \alpha}{2-\alpha}.$$

On the other hand,

$$\frac{\Phi(\frac{\nu_2-c_3}{\sigma/\sqrt{n}}) - \frac{\alpha}{2}}{\Phi(\frac{c_3-\nu_1}{\sigma/\sqrt{n}})} > \Phi(\frac{\nu_2-c_3}{\sigma/\sqrt{n}}) = \Phi(z_{\alpha/2}\frac{\tilde{\sigma}(c_3)}{\sigma}) - \frac{\alpha}{2} > \Phi(\frac{1}{2}z_{\alpha/2}) - \frac{\alpha}{2},$$

where the last inequality we have used that $\tilde{\sigma}(c_3) > \frac{\sigma}{2}$ given $c_3 > \sqrt{\lambda}$. Therefore,

$$(1-\alpha) - \widetilde{CR}(c_3) \in \left(1-\alpha - \frac{2\Phi(\frac{9}{13}z_{\alpha/2}) - \alpha}{2-\alpha}, \Phi(-\frac{1}{2}z_{\alpha/2}) - \frac{\alpha}{2}\right).$$

Combining the bound for $1-\alpha-\widetilde{CR}(\theta)$ and the conclusions in Lemma 2.7.7, we have the upper bound for $1-\alpha-CR(\theta)$.

Table 2.4: Coverage probabilities of confidence regions when $\sigma = 2$

n	θ		p=20			p=50		
			$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$
100	0.3	CR^{Our}	90.0	96.1	91.4	92.2	93.3	97.0
		CR^{Asym}	66.0	48.6	36.5	26.3	18.9	11.7
		CR^{Ptb}	67.3	68.6	74.2	68.6	64.8	58.6
		CR^{Bs}	74.5	77.0	88.7	100.0*	100.0*	100.0*
		CR^{smBS}	68.1	65.4	74.3	95.1*	93.5*	92.3*
		CR^{OLS}	95.0	95.3	94.3	92.0	93.5	95.0
	0.75	CR^{Our}	92.9	91.4	96.0	92.3	96.5	94.2
		CR^{Asym}	89.0	85.3	75.8	92.3	90.3	74.5
		CR^{Ptb}	87.6	90.9	86.4	90.0	93.8	78.9
		CR^{Bs}	91.4	90.7	88.0	98.9*	98.8*	100.0*
		CR^{smBS}	89.3	89.1	89.2	91.3*	95.3*	91.1*
		CR^{OLS}	93.8	94.8	95.0	94.3	94.8	95.3
200	0.2	CR^{Our}	91.0	93.6	93.2	88.6	94.2	91.2
		CR^{Asym}	48.2	35.1	25.0	12.8	5.0	11.9
		CR^{Ptb}	61.4	65.3	69.4	48.1	44.2	49.3
		CR^{Bs}	58.5	58.1	72.8	56.1	61.0	63.5
		CR^{smBS}	54.7	50.5	62.6	46.0	48.8	46.4
		CR^{OLS}	94.5	95.3	95.3	94.5	95.0	95.0
	0.6	CR^{Our}	94.7	95.5	94.2	95.4	96.7	95.7
		CR^{Asym}	90.1	91.5	74.9	89.0	80.4	69.6
		CR^{Ptb}	90.2	92.6	86.1	84.2	88.0	89.6
		CR^{Bs}	90.7	91.7	88.4	86.9	89.4	82.7
		CR^{smBS}	88.4	89.5	91.2	80.7	84.2	81.4
		CR^{OLS}	95.8	95.8	93.5	96.3	93.3	95.8
400	0.15	CR^{Our}	90.2	95.7	92.4	96.7	95.9	92.6
		CR^{Asym}	10.3	40.0	31.0	10.3	2.6	6.4
		CR^{Ptb}	33.6	51.0	60.3	33.6	39.3	44.7
		CR^{Bs}	35.3	54.2	57.9	35.3	39.1	38.6
		CR^{smBS}	27.4	49.9	52.2	27.4	32.2	31.3
		CR^{OLS}	94.3	96.5	93.0	94.3	93.5	93.0
	0.4	CR^{Our}	96.2	95.7	94.8	93.6	97.0	96.7
		CR^{Asym}	89.6	92.0	83.6	89.6	91.0	80.5
		CR^{Ptb}	79.5	89.3	89.2	79.5	75.8	70.7
		CR^{Bs}	87.5	89.3	80.3	87.5	82.4	70.3
		CR^{smBS}	79.8	87.2	84.3	79.8	76.6	71.0
		CR^{OLS}	95.8	94.3	95.0	95.8	94.5	93.8

$*$ indicates that the bootstrap and smooth bootstrap methods encounter a singular-designed matrix probelm (7-10% times), and only partial simulation results are used for calculation.

Table 2.5: Computation time for one simulation run

Two-step Method	Bootstrap	Perturbation
0.86s	127.61s	262.07s

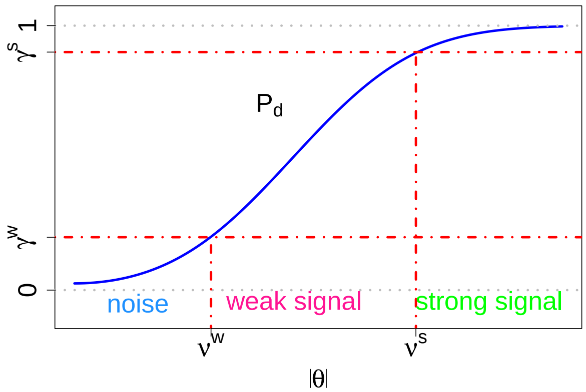

Figure 2.1: Define signal level based on P_d

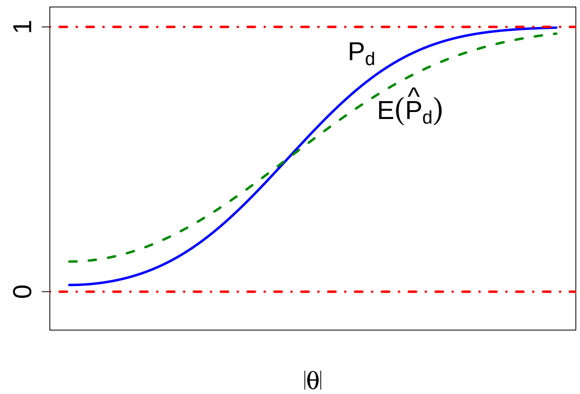

Figure 2.2: P_d and $E(\widehat{P}_d)$

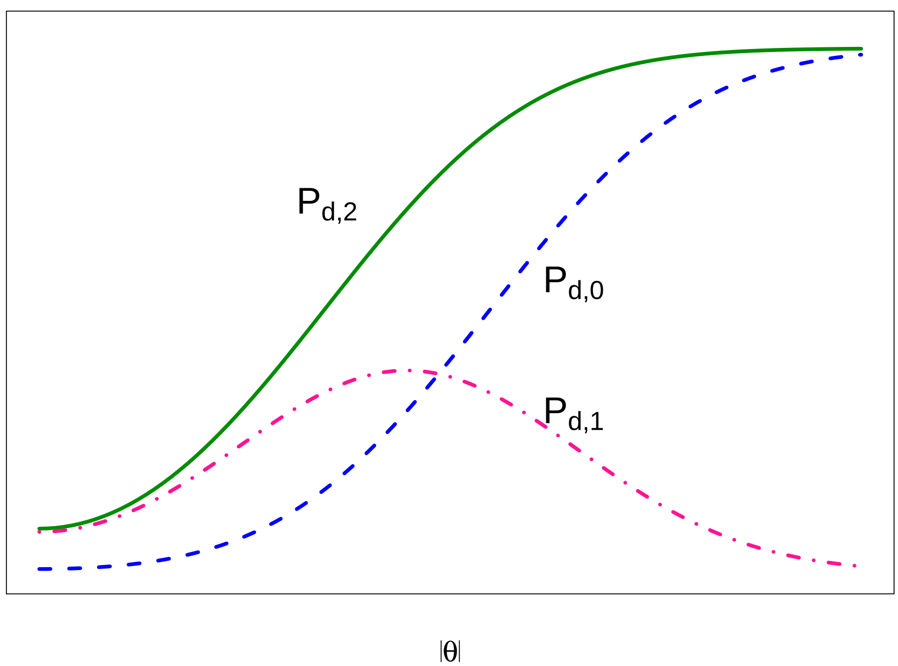

Figure 2.3: Signal's detectability in two-step procedure

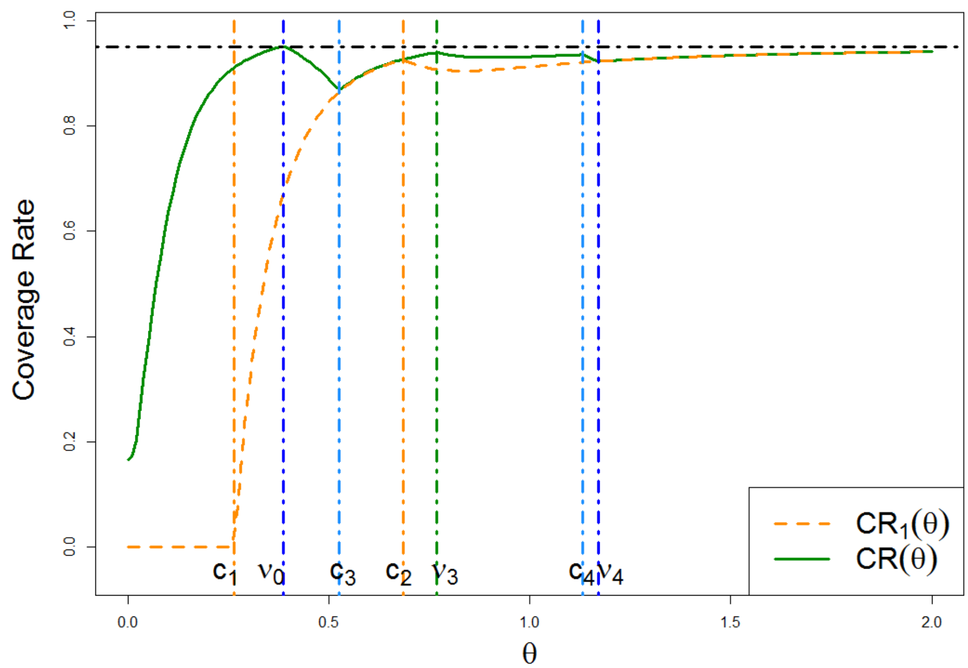

Figure 2.4: $CR(\theta)$ versus $CR_1(\theta)$ (An example: Case 1)

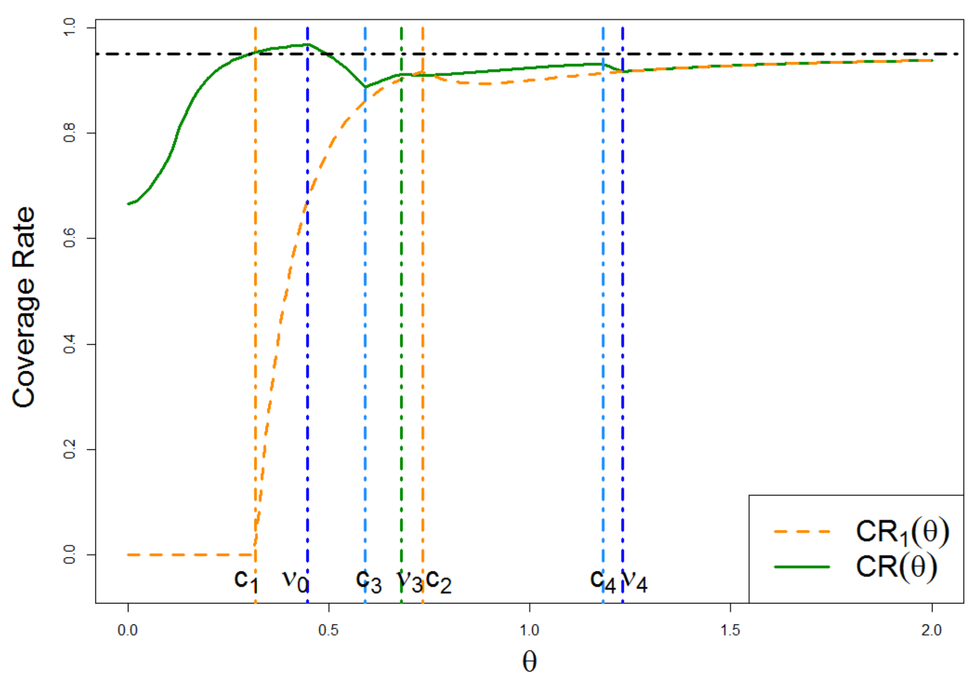

Figure 2.5: $CR(\theta)$ versus $CR_1(\theta)$ (An example: Case 2)

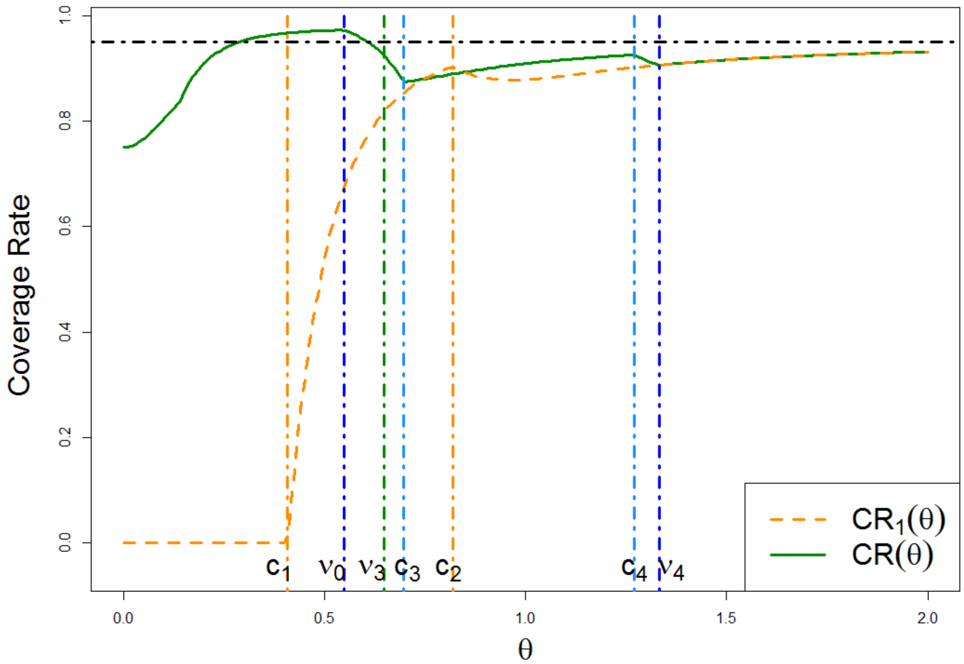

Figure 2.6: $CR(\theta)$ versus $CR_1(\theta)$ (An example: Case 3)

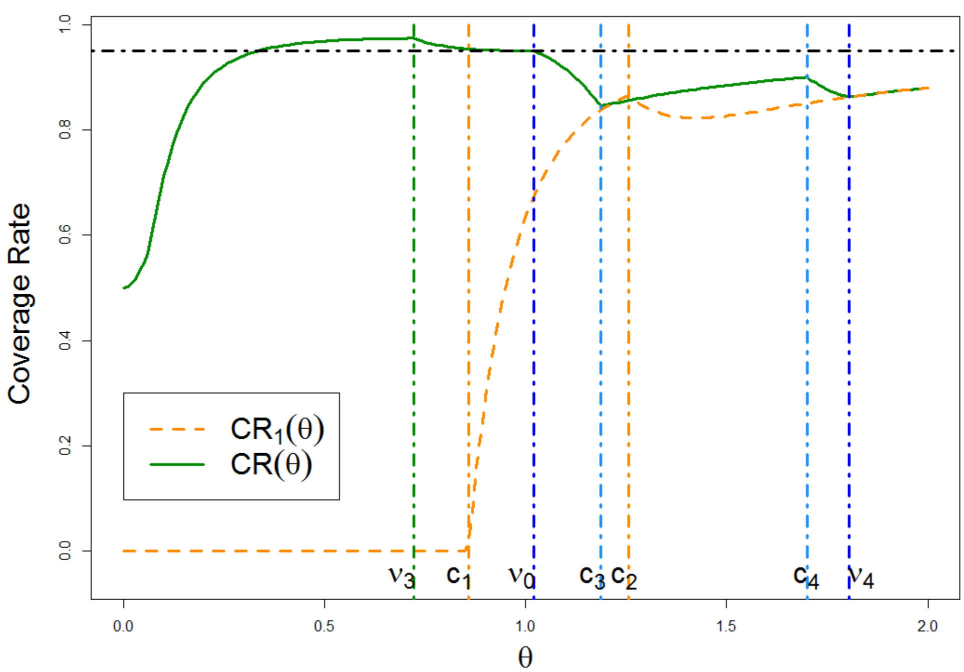

Figure 2.7: $CR(\theta)$ versus $CR_1(\theta)$ (An example: Case 4)

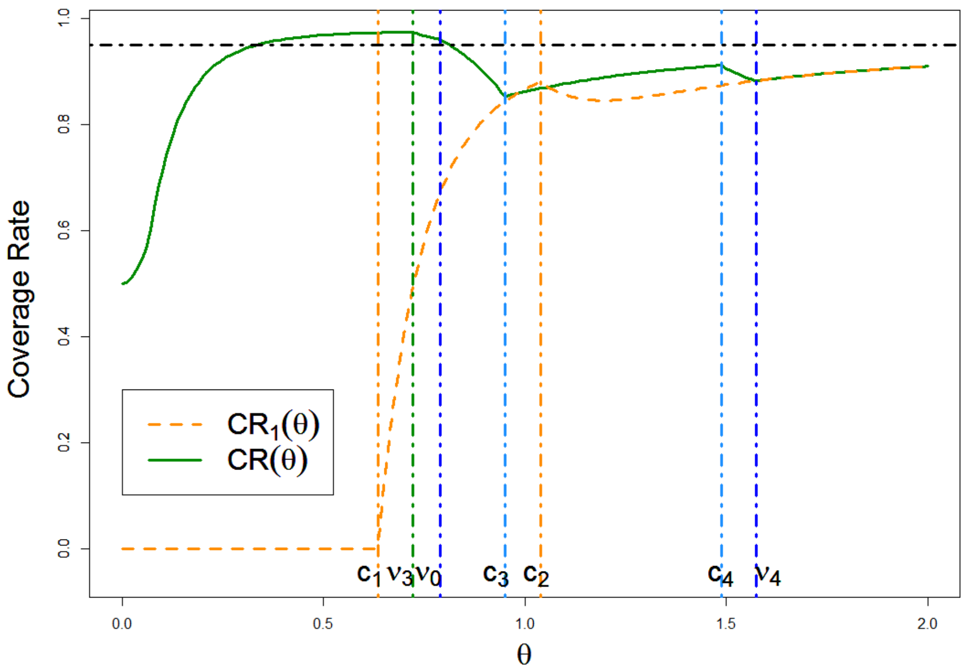

Figure 2.8: $CR(\theta)$ versus $CR_1(\theta)$ (An example: Case 5)

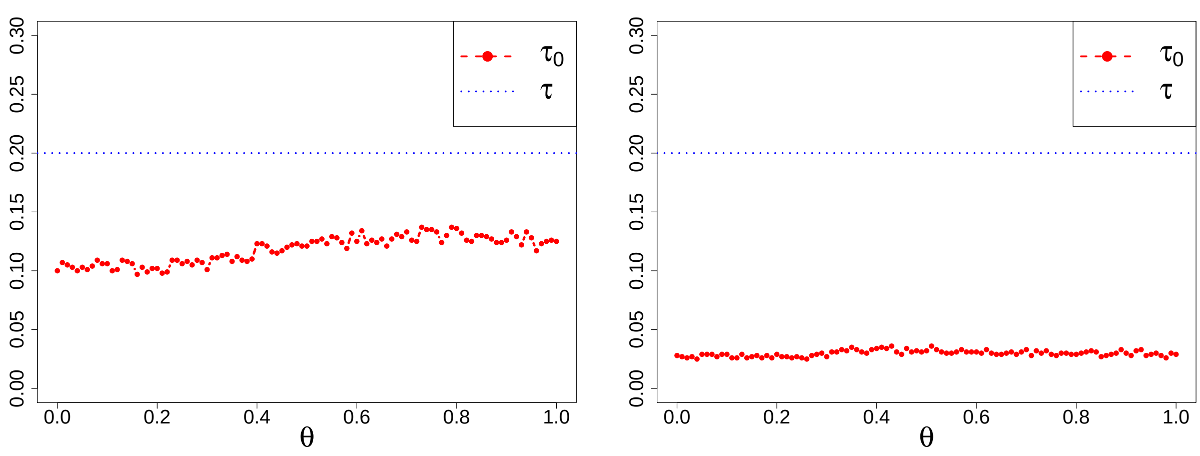

Figure 2.9: False positive rate for simulation setting 1 & 2

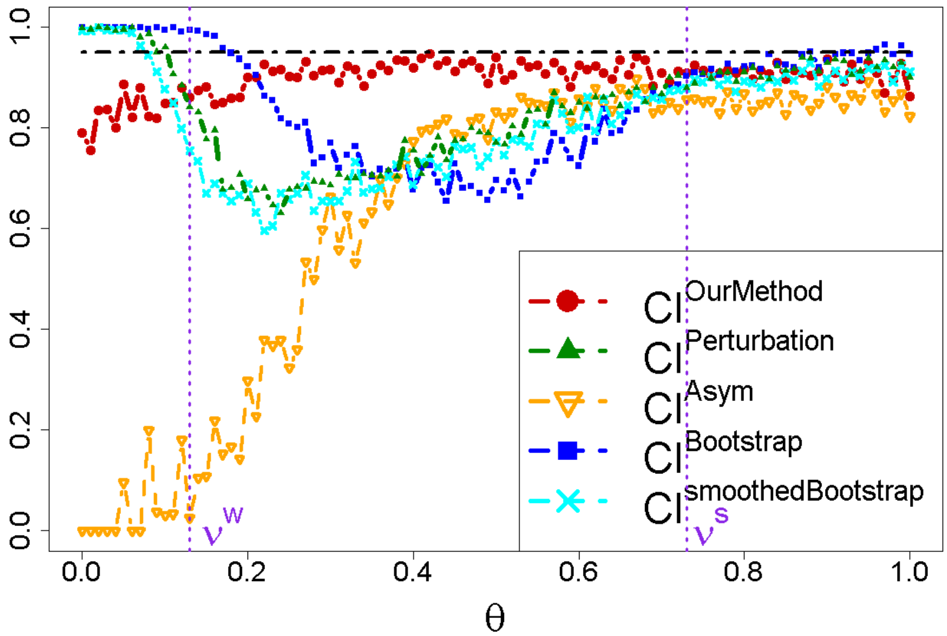

Figure 2.10: 95% confidence interval's coverage rates for simulation setting 1 when $\rho = 0.2$

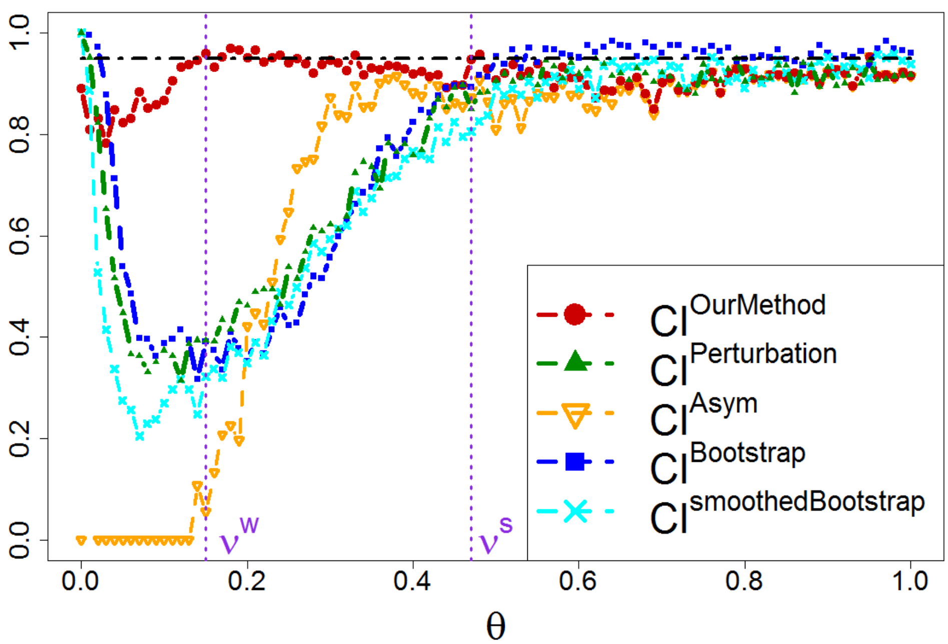

Figure 2.11: 95% confidence interval's coverage rates for simulation setting 2 when $\rho = 0.2$

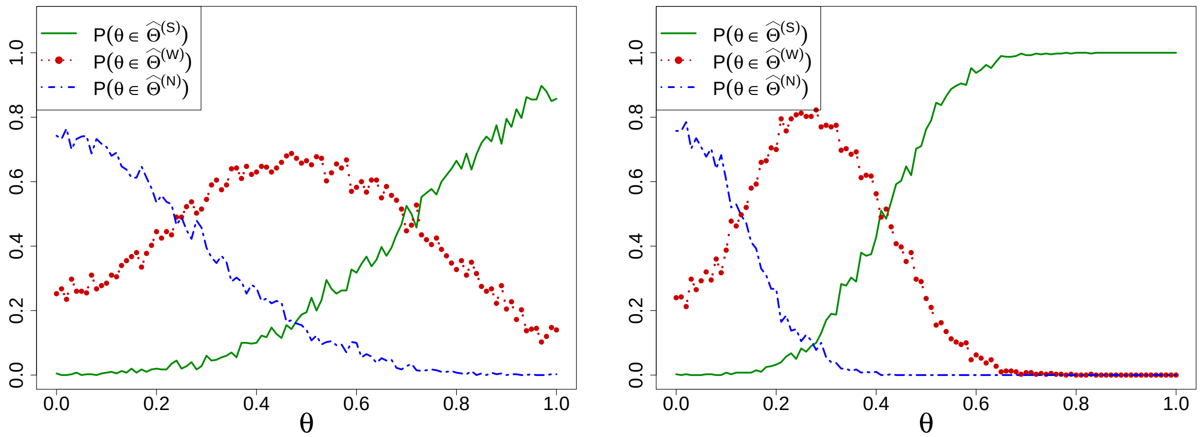

Figure 2.12: Empirical probabilities of identifying each signal level for simulation setting 1 & 2

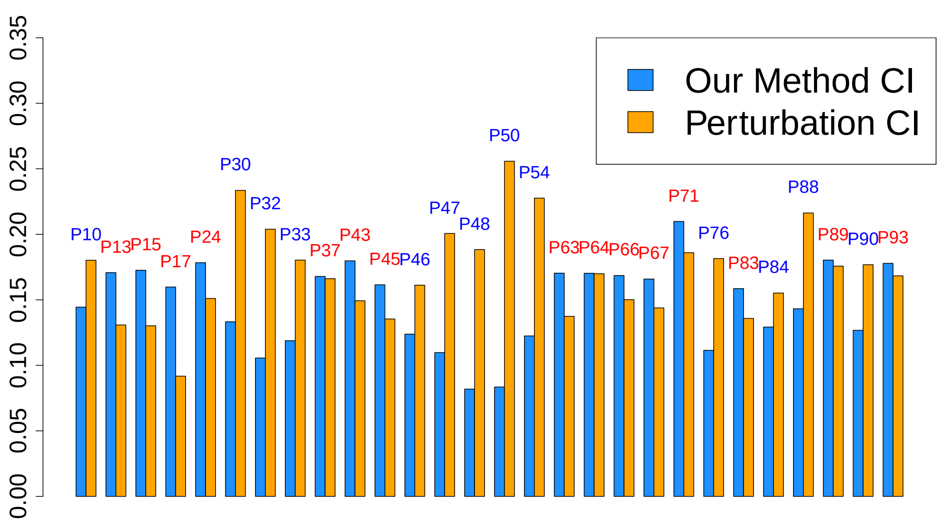

Figure 2.13: Half width of confidence intervals of selected signals for HIV data

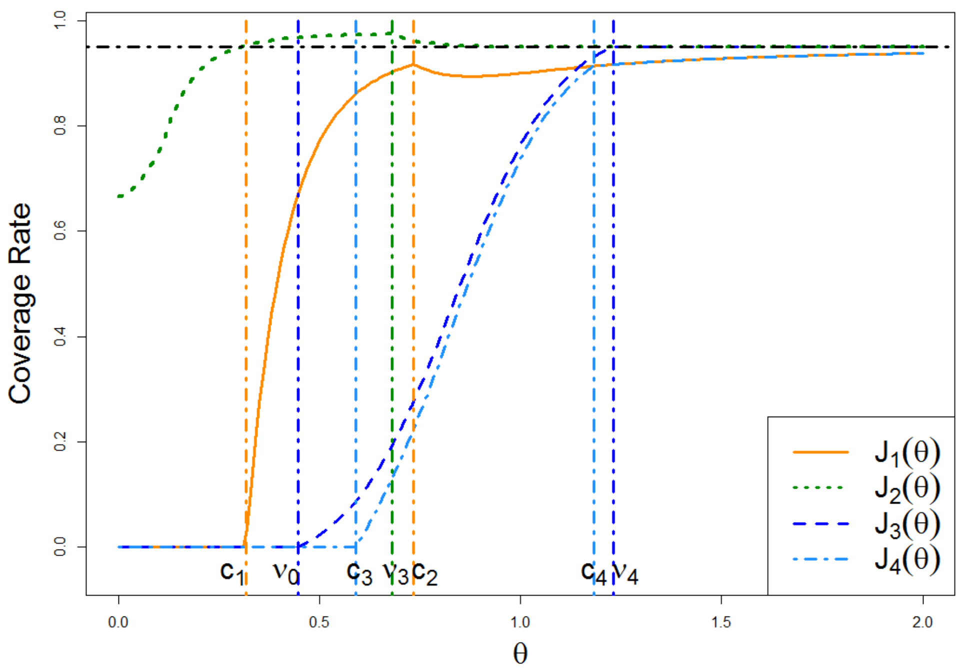

Figure 2.14: $J_1(\theta) - J_4(\theta)$ (Case 1)

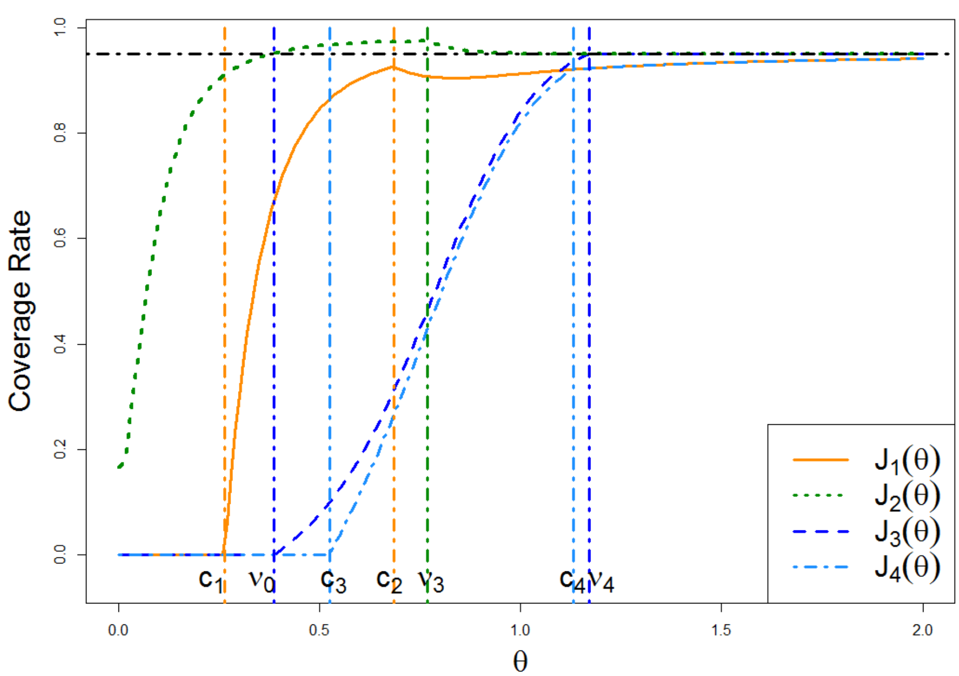

Figure 2.15: $J_1(\theta) - J_4(\theta)$ (Case 2)

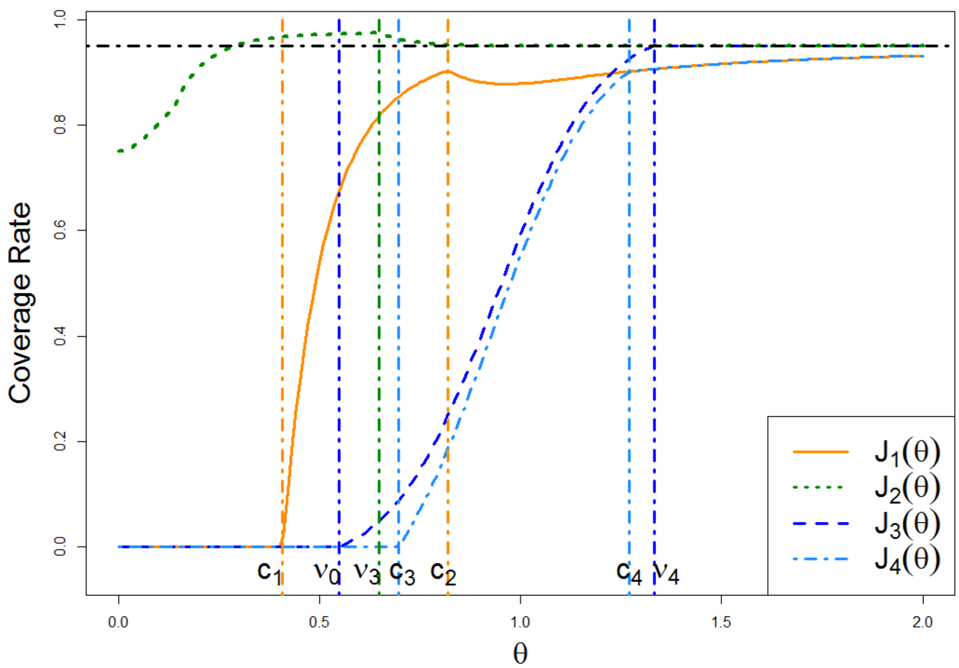

Figure 2.16: $J_1(\theta) - J_4(\theta)$ (Case 3)

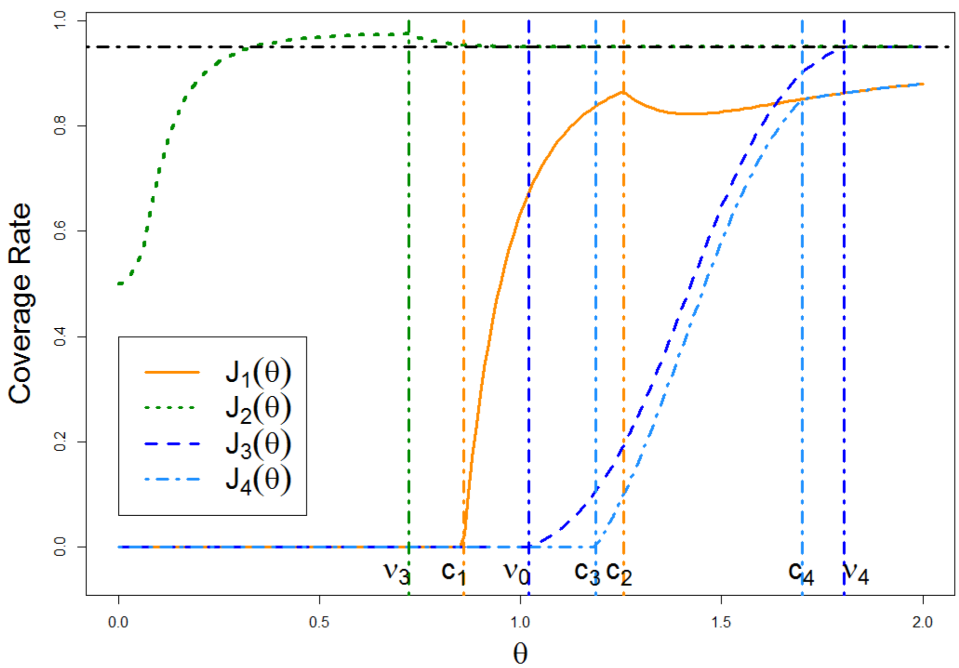

Figure 2.17: $J_1(\theta) - J_4(\theta)$ (Case 4)

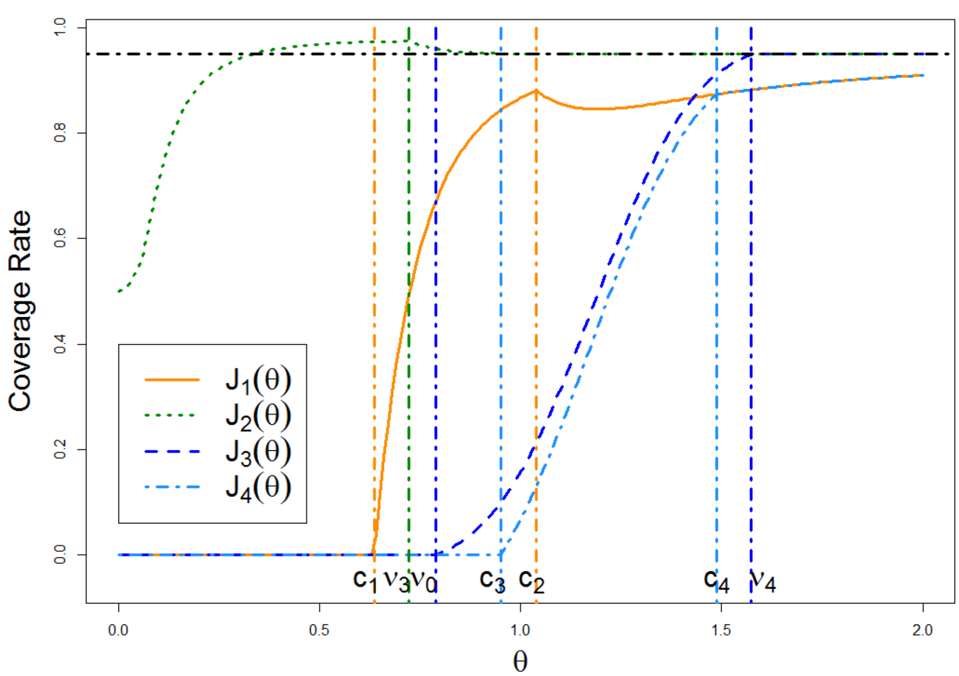

Figure 2.18: $J_1(\theta) - J_4(\theta)$ (Case 5)

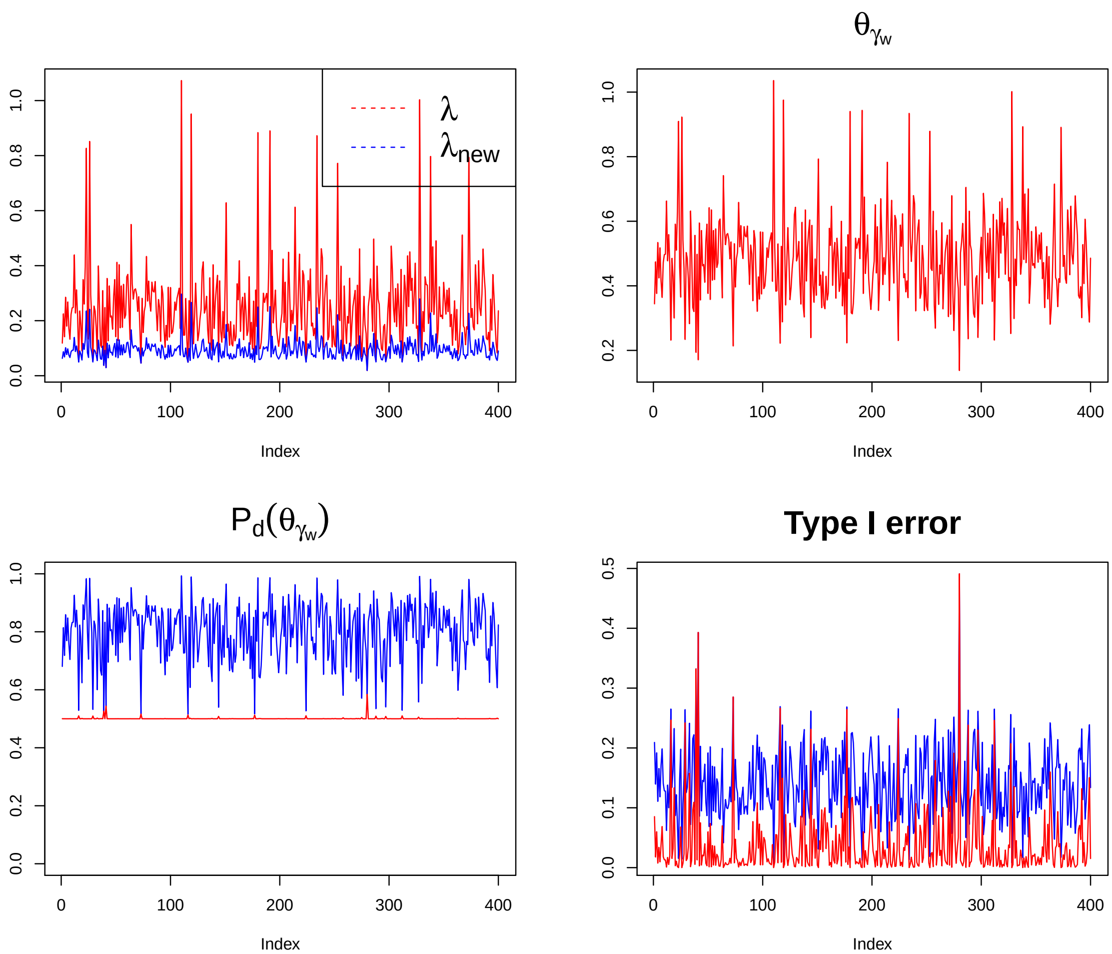

Figure 2.19: Simulation results of the tuning parameter adjustment approach

Chapter 3

A Density Approach of Signal's Inference for Penalized Model Selection

In this chapter, we provide signal's inference method based on the exact distribution of the penalized estimator. This follows the theory developed by Potscher and Schneider (2008) and Potscher and Leeb (2009). Under the orthogonal design assumption, the common penalized estimators have explicit expression for the penalized estimators. For example, the SCAD and adaptive LASSO estimator are provided in (1.2) and (1.3).

In addition, the density functions for these estimators can be directly obtained based on their explicit forms. This provides exact finite sample distributions, where the density formula is a function of sample size, standard deviation of noise term, the tuning parameter and the true parameters. The finite sample distribution is quite different from the asymptotic distribution described in section 1.2 when the true coefficient is small, which is reflected by that the finite sample distribution can be highly non-normal, and consists of a point mass at zero and several piece-wise continuous mixture components.

In theory, the exact finite sample distribution can provide a more accurate inference than the approximate asymptotic-based inference. However, the exact distribution relies on the true parameter, which is unknown in practice. Therefore, here we replace the true parameter by its estimated value for constructing confidence intervals, and we provide the numerical performance of the estimated confidence interval to assess the feasibility of the density approach.

This chapter is organized as follows. In section 3.1, we provide explicit expressions of the density functions for the adaptive LASSO and the SCAD estimator. The standard confidence interval and quantile confidence interval are constructed based on their density formulas. Section 3.2 evaluates the performance of the proposed inference procedures. Section 3.3 discusses the estimation of non-detection probability, and compares to several different estimators. Summary and discussion are provided in Section 3.4.

3.1 Penalized estimator: finite sample distribution

We follow similar arguments by Potscher and Schneider (2008) and Potscher and Leeb (2009). Given true parameter θ, the sample size n, tuning parameter λ, the standard error σ and θ corresponding to the covariate $\boldsymbol{x}$, the finite sample cumulative distribution of $\xi = \sqrt{n}(\hat{\theta}_{SCAD} - \theta)$ is

$$
\begin{aligned}
dF_{n,\theta,\lambda}(\xi) = & \left\{ \Phi(\frac{n\lambda - \boldsymbol{x}^T\boldsymbol{x}\theta}{\sqrt{\boldsymbol{x}^T\boldsymbol{x}}\sigma}) - \Phi(-\frac{n\lambda + \boldsymbol{x}^T\boldsymbol{x}\theta}{\sqrt{\boldsymbol{x}^T\boldsymbol{x}}\sigma}) \right\} d\delta_{-n^{1/2}\theta}(\xi) && (3.1) \\
& + \{f_1(\xi) + f_2(\xi) + f_3(\xi) + f_4(\xi) + f_5(\xi) + f_6(\xi)\} \, d\xi, && (3.2)
\end{aligned}
$$

where

$$
\begin{aligned}
f_1(\xi) &= \frac{1}{\sigma}\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}}\phi\left(\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}}\frac{\xi}{\sigma}\right)\mathbf{1}\{\xi < -\sqrt{n}(a\lambda + \theta)\}, \\
f_2(\xi) &= \frac{1}{\sigma}(\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}} - \frac{1}{a-1}\sqrt{\frac{n}{\boldsymbol{x}^T\boldsymbol{x}}})\phi\left((\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}} - \frac{1}{a-1}\sqrt{\frac{n}{\boldsymbol{x}^T\boldsymbol{x}}}) \cdot \frac{\xi}{\sigma} - \frac{n(a\lambda + \theta)}{(a-1)\sqrt{\boldsymbol{x}^T\boldsymbol{x}}\sigma}\right) \\
& \quad \mathbf{1}\{-\sqrt{n}(a\lambda + \theta) \le \xi < -\sqrt{n}(\lambda + \theta)\}, \\
f_3(\xi) &= \frac{1}{\sigma}\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}}\phi(\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}} \cdot \frac{\xi}{\sigma} - \frac{n\lambda}{\sqrt{\boldsymbol{x}^T\boldsymbol{x}}\sigma})\mathbf{1}\{-\sqrt{n}(\theta + \lambda) \le \xi < -\sqrt{n}\theta\}, \\
f_4(\xi) &= \frac{1}{\sigma}\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}}\phi(\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}} \cdot \frac{\xi}{\sigma} + \frac{n\lambda}{\sqrt{\boldsymbol{x}^T\boldsymbol{x}}\sigma})\mathbf{1}\{-\sqrt{n}\theta \le \xi < \sqrt{n}(\lambda - \theta)\},
\end{aligned}
$$

$$
\begin{aligned}
f_5(\xi) &= \frac{1}{\sigma}(\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}} - \frac{1}{a-1}\sqrt{\frac{n}{\boldsymbol{x}^T\boldsymbol{x}}})\phi\left((\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}} - \frac{1}{a-1}\sqrt{\frac{n}{\boldsymbol{x}^T\boldsymbol{x}}}) \cdot \frac{\xi}{\sigma} + \frac{n(a\lambda - \theta)}{(a-1)\sqrt{\boldsymbol{x}^T\boldsymbol{x}}\sigma}\right) \\
& \quad \mathbf{1}\{\sqrt{n}(\lambda - \theta) \le \xi < \sqrt{n}(a\lambda - \theta)\}, \\
f_6(\xi) &= \frac{1}{\sigma}\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}}\phi\left(\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}}\frac{\xi}{\sigma}\right)\mathbf{1}\{\xi \ge \sqrt{n}(a\lambda - \theta)\}.
\end{aligned}
$$

Note that $\hat{\theta}_{SCAD}$ follows a mixture distribution, which consists of a point mass at 0 and several piecewise continuous distributions. It can be verified that for a fixed parameter θ, the above mixture distribution coverages to normal distribution, as sample size n goes to infinity and the tuning parameter λ is properly selected. This indicates that the finite sample distribution and the asymptotic distribution for the penalized estimator agree with each other if the sample size is large enough. However, the finite sample distribution should be more accurate for the penalized estimator when the sample size is small.

We provide the cumulative distribution function of ξ as follows:

$$F_{n,\theta,\lambda}(\xi) = \begin{cases} F_1(\xi), & \text{if } \xi < -\sqrt{n}(a\lambda+\theta) \\ F_2(\xi) + P_2, & \text{if } -\sqrt{n}(a\lambda+\theta) \le \xi < -\sqrt{n}(\lambda+\theta) \\ F_3(\xi) + P_2 + P_3, & \text{if } -\sqrt{n}(\theta+\lambda) \le \xi < -\sqrt{n}\theta \\ F_4(\xi) + P_2 + P_3 + P_4 + P_0, & \text{if } -\sqrt{n}\theta \le \xi < \sqrt{n}(\lambda-\theta) \\ F_5(\xi) + P_2 + P_3 + P_4 + P_5 + P_0, & \text{if } \sqrt{n}(\lambda-\theta) \le \xi < \sqrt{n}(a\lambda-\theta) \\ F_1(\xi) + P_2 + P_3 + P_4 + P_5 + P_6 + P_0, & \text{if } \xi \ge \sqrt{n}(a\lambda-\theta), \end{cases}$$

where

$$\begin{aligned} F_1(\xi) &= \Phi(\frac{\xi\sqrt{\boldsymbol{x}^T\boldsymbol{x}}}{\sqrt{n}\sigma}); \\ F_2(\xi) &= \Phi(\frac{\xi(\frac{\boldsymbol{x}^T\boldsymbol{x}}{\sqrt{n}} - \frac{\sqrt{n}}{a-1}) - n(\frac{a\lambda+\theta}{a-1})}{\sqrt{\boldsymbol{x}^T\boldsymbol{x}}\sigma}); \\ F_3(\xi) &= \Phi(\frac{\xi\sqrt{\boldsymbol{x}^T\boldsymbol{x}}}{\sqrt{n}\sigma} - \frac{n\lambda}{\sqrt{\boldsymbol{x}^T\boldsymbol{x}}\sigma}); \\ F_4(\xi) &= \Phi(\frac{\xi\sqrt{\boldsymbol{x}^T\boldsymbol{x}}}{\sqrt{n}\sigma} + \frac{n\lambda}{\sqrt{\boldsymbol{x}^T\boldsymbol{x}}\sigma}); \\ F_5(\xi) &= \Phi(\frac{\xi(\frac{\boldsymbol{x}^T\boldsymbol{x}}{\sqrt{n}} - \frac{\sqrt{n}}{a-1}) + n(\frac{a\lambda-\theta}{a-1})}{\sqrt{\boldsymbol{x}^T\boldsymbol{x}}\sigma}), \end{aligned}$$

and

$$\begin{aligned} P_0 &= \Phi(\frac{n\lambda - \boldsymbol{x}^T\boldsymbol{x}\theta}{\sqrt{\boldsymbol{x}^T\boldsymbol{x}}\sigma}) - \Phi(-\frac{n\lambda + \boldsymbol{x}^T\boldsymbol{x}\theta}{\sqrt{\boldsymbol{x}^T\boldsymbol{x}}\sigma}), \\ P_2 &= F_1(-\sqrt{n}(a\lambda+\theta)) - F_2(-\sqrt{n}(a\lambda+\theta)), \\ P_3 &= F_2(-\sqrt{n}(\lambda+\theta)) - F_3(-\sqrt{n}(\lambda+\theta)), \\ P_4 &= F_3(-\sqrt{n}\theta) - F_4(-\sqrt{n}\theta), \\ P_5 &= F_4(\sqrt{n}(\lambda-\theta)) - F_5(\sqrt{n}(\lambda-\theta)), \\ P_6 &= F_5(\sqrt{n}(a\lambda-\theta)) - F_1(\sqrt{n}(a\lambda-\theta)). \end{aligned}$$

Based on the above formulation of $F_{n,\theta,\lambda}(\xi)$, suppose $\xi_{\alpha/2}$ and $\xi_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of $F_{n,\theta,\lambda}(\xi)$ respectively, a $100(1-\alpha)\%$ level quantile confidence interval CI^Q for θ can be constructed by:

$$CI^Q = \left(\hat{\theta}_{SCAD} - \frac{\xi_{1-\alpha/2}}{\sqrt{n}}, \hat{\theta}_{SCAD} - \frac{\xi_{\alpha/2}}{\sqrt{n}}\right). \tag{3.3}$$

In addition, from the density function in (3.1), suppose $E(\xi)$ and $\sigma(\xi)$ are the mean and standard deviation of ξ, we can construct a standard confidence interval CI^S as:

$$CI^S = \left(\hat{\theta}_{SCAD} - \frac{E(\xi) + z_{\alpha/2}\sigma(\xi)}{\sqrt{n}}, \hat{\theta}_{SCAD} - \frac{E(\xi) - z_{\alpha/2}\sigma(\xi)}{\sqrt{n}}\right). \tag{3.4}$$

Note that calculations of $\xi_{\alpha/2}$, $\xi_{1-\alpha/2}$, $E(\xi)$ and $\sigma(\xi)$ rely on the knowledge of the true parameter θ, which is unknown in practice. Computationally, we can substitute θ by a least square estimator or other regularized estimators.

Similarly, the finite sample distribution of $\omega = \sqrt{n}(\hat{\theta}_{ALASSO} - \theta)$ is:

$$F_{ALASSO,n,\theta,\lambda}(\omega) = \mathbf{1}\{\sqrt{n}\theta + \omega \geq 0\}\mathbf{\Phi}(z_2(\omega)) + \mathbf{1}\{\sqrt{n}\theta + \omega < 0\}\mathbf{\Phi}(z_1(\omega)), \tag{3.5}$$

where $z_1(\omega)$ and $z_2(\omega)$ are the roots of solving the following equation:

$$z^2 + \frac{\sqrt{n}\theta - \omega}{\sigma}\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}}z - \left(\frac{n\lambda}{\sigma^2} + \frac{\boldsymbol{x}^T\boldsymbol{x}}{\sqrt{n}\sigma^2}\theta\omega\right) = 0. \tag{3.6}$$

It follows immediately from (3.5) that the density function of ω, which we called $f_{ALASSO,n,\theta,\lambda}(\omega)$ is:

$$\begin{aligned} f_{ALASSO,n,\theta,\lambda}(\omega) &= \left\{\Phi\left(\frac{\sqrt{n\lambda} - \sqrt{\boldsymbol{x}^T\boldsymbol{x}}\theta}{\sigma}\right) - \Phi\left(\frac{-\sqrt{n\lambda} - \sqrt{\boldsymbol{x}^T\boldsymbol{x}}\theta}{\sigma}\right)\right\} d\delta_{-\sqrt{n}\theta}(\omega) & (3.7)\\ &+ \{f_1(\omega) + f_2(\omega)\}\, d\omega, & (3.8)\end{aligned}$$

where

$$\begin{aligned} f_1(\omega) &= c_0(1 - s(\omega))\phi(z_1(\omega))\mathbf{1}\{\sqrt{n}\theta + \omega < 0\},\\ f_2(\omega) &= c_0(1 + s(\omega))\phi(z_2(\omega))\mathbf{1}\{\sqrt{n}\theta + \omega \geq 0\},\\ c_0 &= \frac{1}{2\sigma}\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}},\\ s(\omega) &= \frac{c_2(\xi)}{\sqrt{[c_2(\xi)]^2 + c_3}}. \end{aligned}$$

The solutions of (3.6) can be explicitly expressed as

$$
\begin{aligned}
z_1(\omega) &= -c_1(\omega) - \sqrt{[c_2(\omega)]^2 + c_3}, \\
z_2(\omega) &= -c_1(\omega) + \sqrt{[c_2(\omega)]^2 + c_3},
\end{aligned}
$$

where

$$
\begin{aligned}
c_1(\omega) &= \frac{\sqrt{n}\theta - \omega}{2\sigma}\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}}, \\
c_2(\omega) &= \frac{\sqrt{n}\theta + \omega}{2\sigma}\sqrt{\frac{\boldsymbol{x}^T\boldsymbol{x}}{n}}, \\
c_3 &= \frac{n\lambda}{\sigma^2}.
\end{aligned}
$$

Following $F_{n,\theta,\lambda}(\omega)$, suppose $\omega_{\alpha/2}$ and $\omega_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of $F_{n,\theta,\lambda}(\omega)$ respectively, a $100(1-\alpha)$%-level quantile confidence interval CI^Q for θ can be constructed by:

$$
CI^Q = [\hat{\theta}_{ALASSO} - \frac{\omega_{1-\alpha/2}}{\sqrt{n}}, \hat{\theta}_{ALASSO} - \frac{\omega_{\alpha/2}}{\sqrt{n}}]. \tag{3.9}
$$

We utilize the density function in (3.7), and let $E(\omega)$ and $\sigma(\omega)$ be the mean and standard deviation of ω, then construct a standard confidence interval CI^S as:

$$
CI^S = [\hat{\theta}_{ALASSO} - \frac{E(\omega) + z_{\alpha/2}\sigma(\omega)}{\sqrt{n}}, \hat{\theta}_{ALASSO} - \frac{E(\omega) - z_{\alpha/2}\sigma(\omega)}{\sqrt{n}}]. \tag{3.10}
$$

Since the finite sample distribution for the weak signal deviates from a normal distribution, the symmetric standard confidence interval based on the normality might not be the most efficient confidence interval. There are several potential drawbacks of the standard confidence interval. The first one is the estimation of standard deviation. The existing formulas are known to be underestimated for the true standard errors which lead to under-coverage confidence intervals. For example, it is possible that the standard deviation can be estimated to be zero based on the above formula. Secondly, the standard confidence interval becomes unreliable as the distribution is asymmetric, therefore the mean μ is no longer the center of the distribution. Given the actual density is asymmetric and highly-skewed, one extreme case for the ill-behaved standard confidence interval is that the true distribution of the estimator is so skewed that μ lies outside its $100(1-\alpha)$% quantile. In such case, the standard confidence interval can be misleading for the true signal. In another situation, the distribution of the penalized estimator might no longer be continuous. This occurs when the signal is weak, and there is positive point mass at zero. For discontinuous density function, it makes more sense to construct a quantile confidence interval instead. Although there are potential drawbacks for the standard confidence intervals in theory, their performance is not worse than the quantile confidence interval in practice. We compare their finite sample performance in the next section.

3.2 Numerical studies

To examine the empirical performance of the density-based inference approach, we conduct simulation studies in a similar setting as in Section 2.3. We generate 400 simulated data with a sample size of n for the linear model $y = \boldsymbol{X\theta} + \mathcal{N}(0, \sigma^2)$, where $\boldsymbol{X} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_p)$ and $\boldsymbol{X}_j \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I_n})$. We allow covariates $\boldsymbol{X}$ to be correlated with an AR(1) correlation structure, and the pairwise correlation $cor(\boldsymbol{X_i}, \boldsymbol{X_j}) = \rho^{|i-j|}$. We choose $(n, p, \sigma) = (100, 20, 2), (100, 50, 2)$ and $(400, 50, 2)$ for each setting. To assess the performance of the proposed methods when the orthogonal-designed assumption is violated, we let $\rho = 0.2$ for each setting. In addition, the p-dimensional coefficient vector $\boldsymbol{\theta} = (1, 1, 0.5, \theta, 0, \cdots, 0)$, which consists of two strong signals of 1's, one moderate strong signal of 0.5, one varying-signal θ, and $(p-4)$ null variables. We allow the coefficient θ vary between 0 (null) to 2 (strong signal) in order to examine the confidence coverages across different signal strength levels.

We construct 95% confidence intervals based on (3.9) and (3.10). Figures 3.1, 3.3 and 3.5 provide the coverage probabilities for θ varying between 0 and 2. Note that when the signal strength is close to zero, neither of the coverage rates using the standard method or quantile method are accurate. However, the proposed method becomes more accurate as the magnitude of θ increases. For example, when $\rho = 0.2$, in setting 1, the coverage rate of the proposed method is quite close to 95% when the signal strength is larger than 0.8; and in setting 3, the coverage rate of the proposed method is quite close to 95% when the signal strength is larger than 0.5.

The proposed inference method is still accurate for strong signals even when p is close to n if correlation between covariates is low, as is shown in Figure 3.3. In all the settings, the standard confidence intervals perform similarly, if not better than, the quantile confidence intervals. On the other hand, the performance of density-based inferential method is deteriorate as the correlation between covariates increases. For example, the coverage probabilities in setting 1 are less than 95% for any signal strength when $\rho = 0.5$.

To compare the density-based inference approach with the asymptotic-based inference method and the proposed two-step inference method in Chapter 2, we also include the coverage probabilities for all methods in Figures 3.2, 3.4 and 3.6. In general, the density-based standard confidence interval provides more accurate inference than the asymptotic-based method. The density-based quantile confidence interval performs better for the weak signals, but not for the moderately large signals. The two-step inference method provides better confidence intervals than either of these two density approaches. For example, when $\rho = 0.2$, in setting 1, the coverage rate of the density approach is close to 95% when the signal strength is larger than 0.8; in setting 3, the coverage rate of the density approach is close to 95% when the signal strength is larger than 1. Therefore, the density approach works effectively only for strong signals. When signal strength becomes sufficiently large, all these methods can provide accurate inference.

3.3 More estimators for P_d

The estimation of P_d is crucial to the signal's identification. In addition to the proposed estimator of P_d described above, other estimators are provided as follows. We can replace unknown θ by a ridge estimator or a least-square estimator in the expression of P_d. Indeed, there exist a bias-variance trade-off for each of these estimators. We explain the details as following.

To simplify our discussion, we assume that the covariates are standardized such that $\boldsymbol{x}^T\boldsymbol{x} = n$. For a fixed λ, the explicit estimator of adaptive LASSO is:

$$\hat{\theta}_{ALASSO} = (|\hat{\theta}^{LS}| - \frac{\lambda}{|\hat{\theta}^{LS}|})_+ \mathrm{sgn}(\hat{\theta}^{LS}).$$

Therefore we have the detection probability P_d with the following form:

$$P(\hat{\theta}_{ALASSO} \neq 0) = \Phi(\frac{-\sqrt{n\lambda} + \sqrt{n}\theta}{\sigma}) + \Phi(\frac{-\sqrt{n\lambda} - \sqrt{n}\theta}{\sigma}). \tag{3.11}$$

Thus the expected value of $\widehat{P}_d$ by replacing θ with $\hat{\theta}_{LS}$ in P_d becomes

$$\begin{aligned} E(\widehat{P}_d) &= E\left\{\Phi(\frac{\sqrt{n}}{\sigma}(\hat{\theta}_{LS} - \sqrt{\lambda})) + \Phi(-\frac{\sqrt{n}}{\sigma}(\hat{\theta}_{LS} + \sqrt{\lambda}))\right\} \\ &= \Phi(\frac{\sqrt{n}}{\sqrt{2}\sigma}(\theta - \sqrt{\lambda})) + \Phi(-\frac{\sqrt{n}}{\sqrt{2}\sigma}(\theta + \sqrt{\lambda})). \end{aligned}$$

It is obvious that $\widehat{P}_d$ is not an unbiased estimator of P_d. We can modify the estimator $\widehat{P}_d$ by rescaling σ, which is $P_d(n, \hat{\theta}_{LS}, k\sigma, \lambda)$. The expectation of the new estimator becomes:

$$E\{P_d(n, \hat{\theta}_{LS}, k\sigma, \lambda)\} = E\left\{\Phi(\frac{\sqrt{n}}{k\sigma}(\hat{\theta}_{LS} - \sqrt{\lambda})) + \Phi(-\frac{\sqrt{n}}{k\sigma}(\hat{\theta}_{LS} + \sqrt{\lambda}))\right\} \tag{3.12}$$

$$= \Phi(\frac{\sqrt{n}}{\sqrt{1+k^2}\sigma}(\theta - \sqrt{\lambda})) + \Phi(-\frac{\sqrt{n}}{\sqrt{1+k^2}\sigma}(\theta + \sqrt{\lambda})). \tag{3.13}$$

When $k \gg 1$, (3.13) becomes quite close to P_d. Thus we can obtain a nearly-unbiased estimator for P_d, by rescaling σ in $\widehat{P}_d$.

An alternative way of estimating P_d is to use other estimator of θ rather than $\hat{\theta}_{LS}$. For example, the ridge estimator is another option, since $\hat{\theta}_{ridge} = \frac{1}{1+\rho}\hat{\theta}_{LS}$ and it is a rescaled version of $\hat{\theta}_{LS}$. The expected value of $\widehat{P}_d$ by replacing θ with $\hat{\theta}_{ridge}$ becomes:

$$
\begin{aligned}
E(\widehat{P}_d) &= E(P_d(n, \hat{\theta}_{ridge}, k\sigma, \lambda)) \\
&= E\left\{\Phi(\frac{\sqrt{n}}{k\sigma}(\frac{1}{1+\rho}\hat{\theta}_{LS} - \sqrt{\lambda})) + \Phi(-\frac{\sqrt{n}}{k\sigma}(\frac{1}{1+\rho}\hat{\theta}_{LS} + \sqrt{\lambda}))\right\} \\
&= \Phi\left(\frac{\sqrt{n}}{\sqrt{k^2 + (\frac{1}{1+\rho})^2\sigma}}(\frac{1}{1+\rho}\theta - \sqrt{\lambda})\right) + \Phi\left(-\frac{\sqrt{n}}{\sqrt{k^2 + (\frac{1}{1+\rho})^2\sigma}}(\frac{1}{1+\rho}\theta + \sqrt{\lambda})\right),
\end{aligned}
$$

where any given scaling parameters r and k also satisfy:

$$
E\left\{\Phi(\frac{\sqrt{n}(\sqrt{\lambda} - r\hat{\theta}_{LS})}{k\sigma})\right\} = \Phi(\frac{\sqrt{n}(\sqrt{\lambda} - a\theta)}{\sqrt{k^2 + r^2\sigma}}).
$$

If we let $k = \frac{\sqrt{\rho^2+2\rho}}{1+\rho}$, then

$$
E(\widehat{P}_d) = \Phi\left(\frac{\sqrt{n}}{\sigma}(\frac{1}{1+\rho}\theta - \sqrt{\lambda})\right) + \Phi\left(\frac{\sqrt{n}}{\sigma}(\sqrt{\lambda} + \frac{1}{1+\rho}\theta)\right).
$$

The above estimator indicates that there exists a bias-variance trade-off. That is, if we obtain less biased estimator of P_d, then there is more variation for the estimator. In general, the estimator by plugging in $\hat{\theta}_{LS}$ achieves better performance compared with other estimators, as it provides a stable estimation of P_d, although is at the cost of a bit of bias.

3.4 Summary and discussion

In this chapter, we propose signal's inference method using the exact finite sample distribution of the penalized estimator. The finite sample cumulative distribution function and the probability density function for a single model parameter can be obtained marginally if we assume covariates are orthogonally designed. The proposed method is applicable for any model parameter, whether the corresponding covariate is selected or not. However, the finite sample distribution of the penalized estimator requires the knowledge of the true parameter. In practice, the true information is unknown, and therefore we replace the true parameter by the least-square estimator.

Our numerical studies indicate that the proposed inferential method only works well when the signal is strong. When a signal strength is weak, the density-based approach provides an under-coverage confidence interval. In addition, the density-based approach does not perform well when the covariates are correlated, even if the correlation is weak or moderate. This is not surprising since the method is developed based on the marginal distribution of the estimator for each model parameter, and the marginal distribution assuming independence of covariates does not capture all the information for the joint distribution if covariates are correlated. For non-orthogonal designed covariates, it remains an open question of whether a joint distribution of model parameter estimator can be obtained or not. In addition, it is worth investigating to estimate the detection probability jointly.

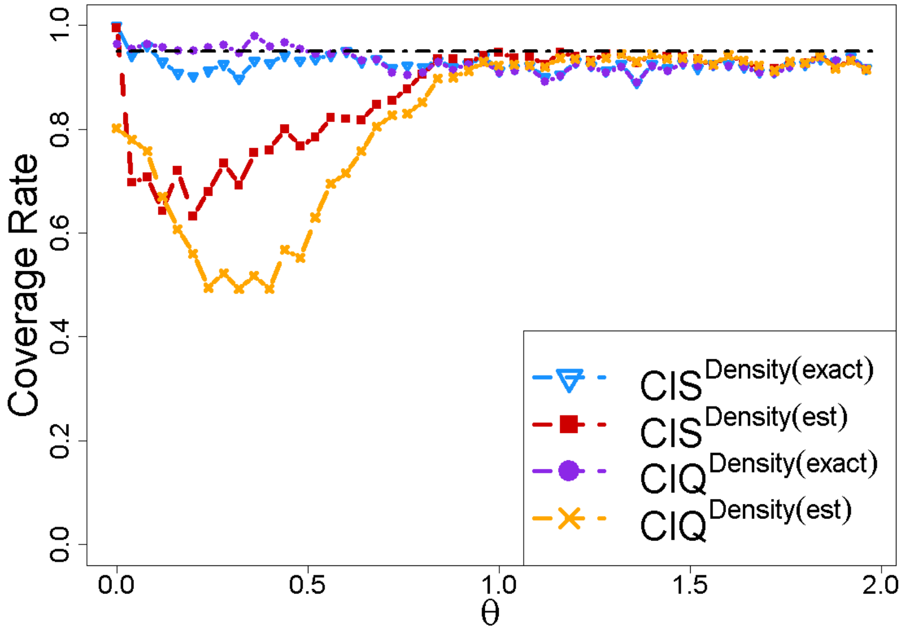

Figure 3.1: 95% confidence interval's coverage rates for n=100, p=20 when $\rho = 0.2$

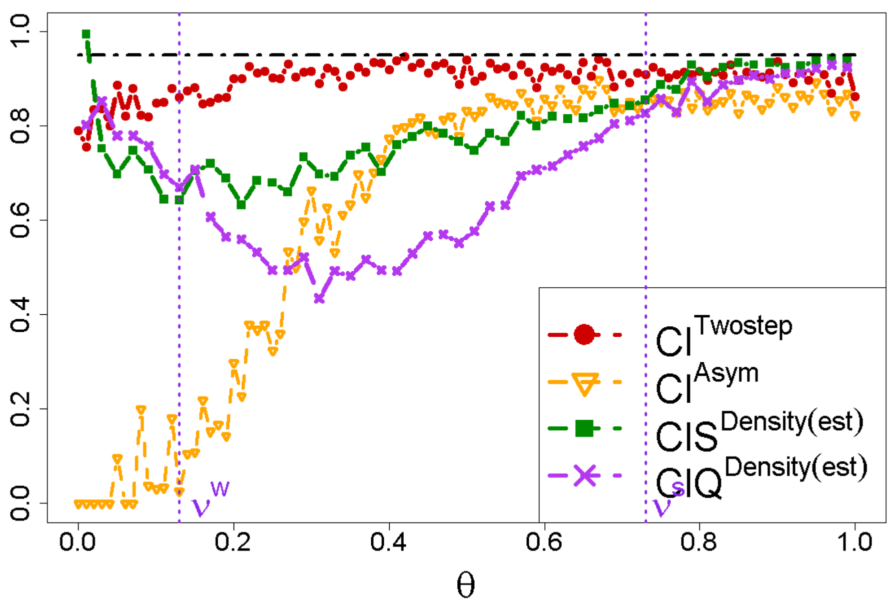

Figure 3.2: 95% confidence interval's coverage rates for n=100, p=20 when $\rho = 0.2$(all methods)

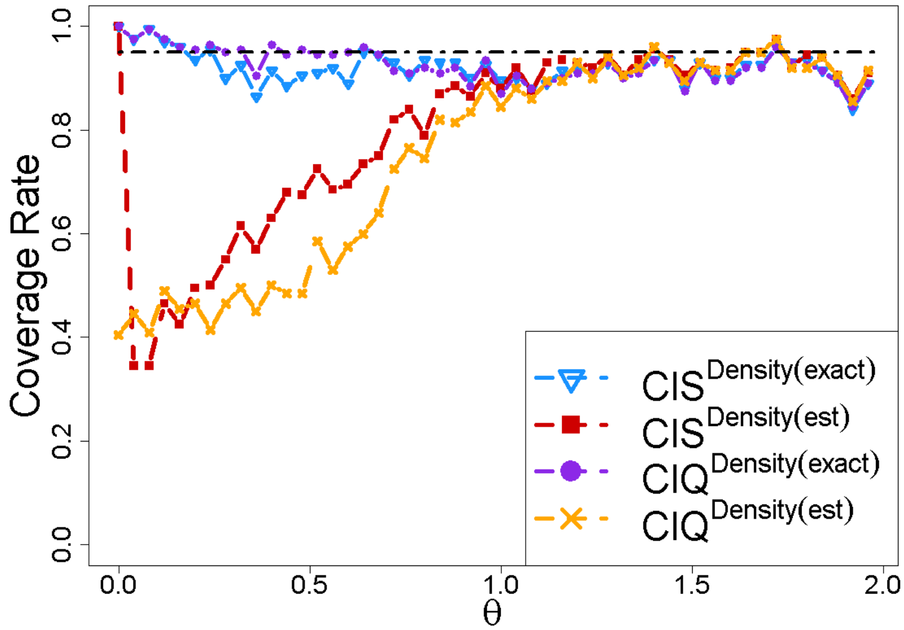

Figure 3.3: 95% confidence interval's coverage rates for n=100, p=50 when $\rho = 0.2$

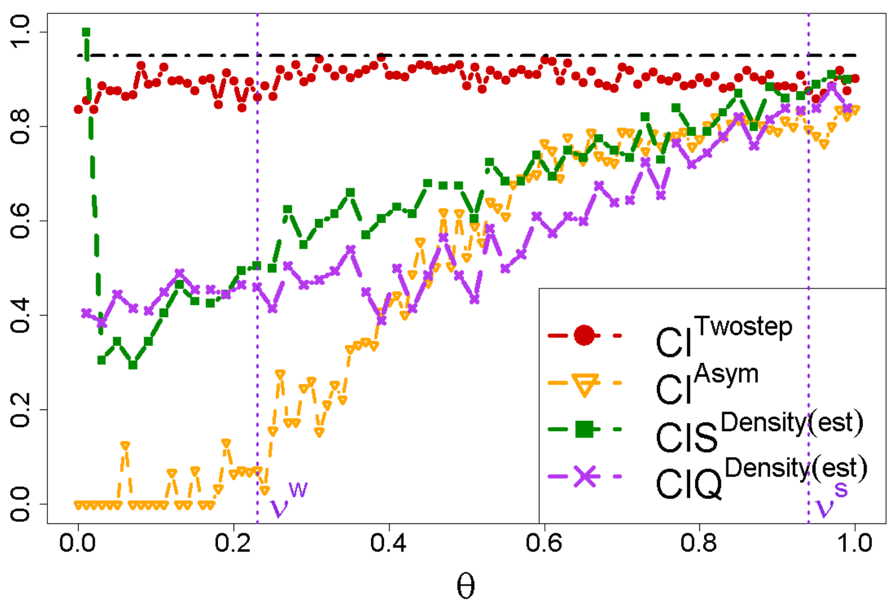

Figure 3.4: 95% confidence interval's coverage rates for n=100, p=50 when $\rho = 0.2$(all methods)

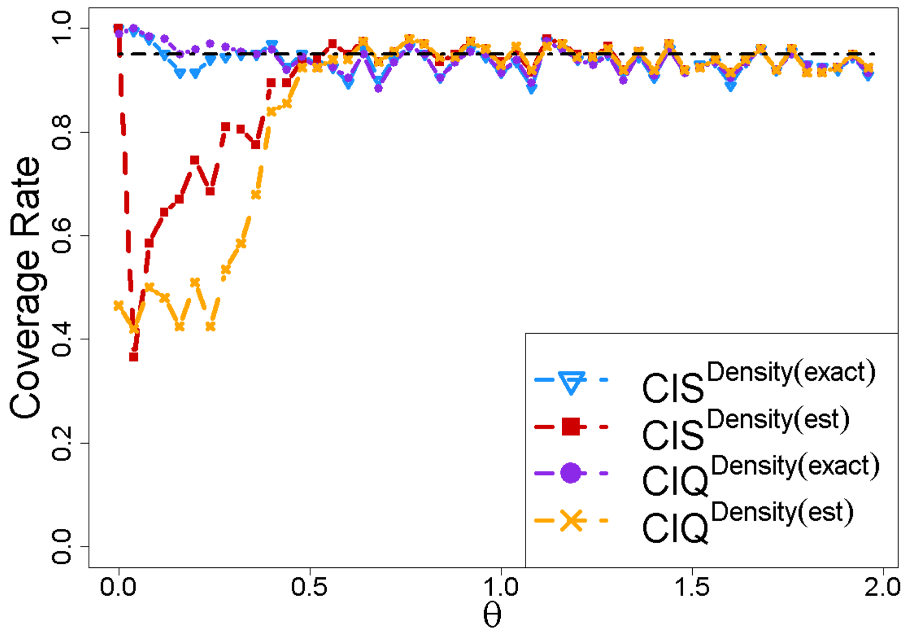

Figure 3.5: 95% confidence interval's coverage rates for n=400, p=50 when $\rho = 0.2$

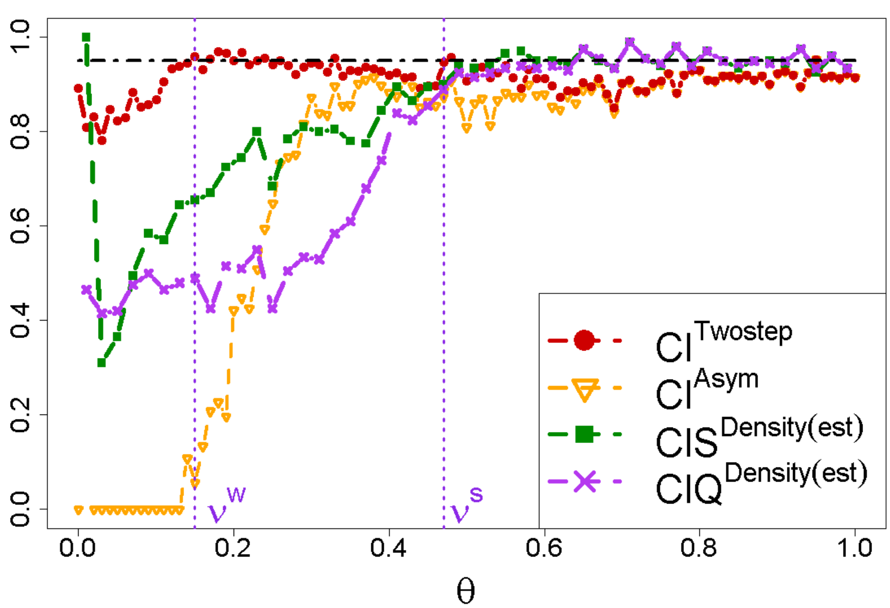

Figure 3.6: 95% confidence interval's coverage rates for n=400, p=50 when $\rho = 0.2$(all methods)

References

[1] Andrews, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 68(2), 399-405.

[2] Berk, R., Brown, L., Buja, A., Zhang, K. and Zhao, L. (2013). Valid post-selection inference. *The Annals of Statistics*, 41, 802-837.

[3] Bickel, P. J., Gtze, F., and van Zwet, W. R. (1997). Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica*, 7, 1-31.

[4] Candes, E., and Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics*, 2313-2351.

[5] 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. *The Annals of Statistics*, 41(3), 1232-1259.

[6] Donoho, D. and Jin, J. (2008). Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. *Proceedings of the National Academy of Sciences*, 105(39), 14790-14795.

[7] Efron, B. (2014). Estimation and accuracy after model selection. *Journal of the American Statistical Association*, 109, 991-1007.

[8] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348-1360.

[9] Fan, J. and Lv, J. (2010). A selected overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20, 101-148 .

[10] Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society, Ser. B*, 70, 849911.

[11] Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*, 32(3), 928-961.

[12] Fan, Y. and Tang, C. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: Ser. B*, 75, pp. 531-552.

[13] Fan, J., Xue, L., and Zou, H. (2014). Strong oracle optimality of folded concave penalized estimation. *The Annals of Statistics*, 42(3), 819.

[14] Farchione, D. and Kabaila P. (2012). Confidence intervals in regression centered on the SCAD estimator. *Statistics and Probability Letters*, 82, 1953-1960.

[15] Fessler, J. A.(1996). Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Application to tomography. *IEEE Transactions on Image Processing*, 5(3), 493-506.

[16] Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1.

[17] Huang, J., and Xie, H. (2007). Asymptotic oracle properties of SCAD-penalized least squares estimators. *Lecture Notes-Monograph Series*, 149-166.

[18] Jin, J. (2009). Impossibility of successful classification when useful features are rare and weak. *Proceedings of the National Academy of Sciences*, 106(22), 8859-8864.

[19] Jin, J., Zhang, C. H., and Zhang, Q. (2014). Optimality of graphlet screening in high dimensional variable selection. *The Journal of Machine Learning Research*, 15(1), 2723-2772.

[20] Johnson, V. A., Brun-Vzinet, F., Clotet, B., Kuritzkes, D. R., Pillay, D., Schapiro, J. M., and Richman, D. D. (2006). Update of the drug resistance mutations in HIV-1: Fall 2006. *Top HIV Med*, 14(3), 125-130.

[21] Kabaila, P., and Leeb, H. (2006). On the large-sample minimal coverage probability of confidence intervals after model selection. *Journal of the American Statistical Association*, 101(474), 619-629.

[22] Kim, Y., Choi, H., and Oh, H. S. (2008). Smoothly clipped absolute deviation on high dimensions. *Journal of the American Statistical Association*, 103(484), 1665-1673.

[23] Knight, K., and Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, 1356-1378.

[24] Leeb, H. and Ptscher, B.M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, 21(1), 21-59.

[25] Li, R., Zhong, W., and Zhu, L. (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association*, 107(499), 1129-1139.

[26] Minnier, J., Tian, L., and Cai, T. (2011). A perturbation method for inference on regularized regression estimates. *Journal of the American Statistical Association*, 106(496).

[27] Politis, D. N., and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, 2031-2050.

[28] Potscher, B.M. and Leeb, H. (2009). On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding. *Journal of Multivariate Analysis*, 100, 2065-2082.

[29] Potscher, B. M., and Schneider, U. (2009). On the distribution of the adaptive LASSO estimator. *Journal of Statistical Planning and Inference*, 139(8), 2775-2790.

[30] Potscher, B.M. and Schneider, U. (2010). Confidence sets based on penalized maximum likelihood estimators in Gaussian regression. *Electronic Journal of Statistics*, 4, 334-360.

[31] Shen, X., Pan, W., and Zhu, Y. (2012). Likelihood-based selection and sharp parameter estimation. *Journal of the American Statistical Association*, 107(497), 223-232.

[32] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Ser. B*, 267-288.

[33] Wang, H., and Leng, C. (2007). Unified LASSO estimation by least squares approximation. *Journal of the American Statistical Association*, 102(479).

[34] Wang, H., Li, B., and Leng, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society: Ser. B*, 71(3), 671-683.

[35] Wang, H., Li, R., and Tsai, C. L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, 94(3), 553-568.

[36] Wu, M. (2009). A parametric permutation test for regression coefficients in LASSO regularized regression. Ph.D thesis, Harvard School of Public Health, Dept. of Biostatistics, Boston, MA.

[37] Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2), 894-942.

[38] Zhang, Y., Li, R., and Tsai, C. L. (2010). Regularization parameter selections via generalized information criterion. *Journal of the American Statistical Association*, 105(489), 312-323.

[39] Zhang, C. H., and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Ser. B*, 76(1), 217-242.

[40] Zhao, P., and Yu, B. (2006). On model selection consistency of Lasso. *The Journal of Machine Learning Research*, 7, 2541-2563.

[41] Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418-1429.

[42] Zou, H., and Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *The Annals of Statistics*, 36(4), 1509.