View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Collection Of Biostatistics Research Archive

Harvard University

Harvard University Biostatistics Working Paper Series

Year 2010 *Paper* 119

A Perturbation Method for Inference on Regularized Regression Estimates

Jessica Minnier[*] Lu Tian[†]

Tianxi Cai[‡]

[*] Harvard University, jminnier@hsph.harvard.edu

[†] Stanford University School of Medicine, lutian@stanford.edu

[‡] Harvard University, tcai@hsph.harvard.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

http://biostats.bepress.com/harvardbiostat/paper119

Copyright ©2010 by the authors.

A Perturbation Method for Inference on Regularized Regression Estimates

Jessica Minnier, Lu Tian and Tianxi Cai*

June 9, 2010

Abstract

Analysis of high dimensional data often seeks to identify a subset of important features and assess their effects on the outcome. Traditional statistical inference procedures based on standard regression methods often fail in the presence of high-dimensional features. In recent years, regularization methods have emerged as promising tools for analyzing high dimensional data. These methods simultaneously select important features and provide stable estimation of their effects. Adaptive LASSO and SCAD for instance, give consistent and asymptotically normal estimates with oracle properties. However, in finite samples, it remains difficult to obtain interval estimators for the regression parameters. In this paper, we propose perturbation resampling based procedures to approximate the distribution of a general class of penalized parameter estimates. Our proposal, justified by asymptotic theory, provides a simple way to estimate the covariance matrix and confidence regions. Through finite sample simulations, we verify the ability of this method to give accurate inference and compare it to other widely used standard deviation and confidence interval estimates. We also illustrate our proposals with a data set used to study the association of HIV drug resistance and a large number of genetic mutations.

KEY WORDS: High dimensional regression; Interval estimation; Oracle property; Regularized estimation; Resampling methods.

*Jessica Minnier is Ph.D. candidate, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115 (E-mail: *jminnier@hsph.harvard.edu*). Lu Tian is Assistant Professor, Department of Health Research & Policy, Stanford University School of Medicine, Palo Alto, CA 94304 (E-mail: *lutian@stanford.edu*). Tianxi Cai is Associate Professor, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115 (E-mail: *tcai@hsph.harvard.edu*). This research was supported by National Institutes of Health Grants T32 AI007358, R01 GM079330, R01 HL089778 and DMS 0854970.

Hosted by The Berkeley Electronic Press

1. INTRODUCTION

Accurate prediction of disease outcomes is fundamental for successful disease prevention and treatment selection. Recent advancement in biological and genomic research has led to the discovery of a vast number of new markers that can potentially be used to develop molecular disease prevention and intervention strategies. For example, gene expression analyses have identified molecular subtypes that are associated with differential prognosis and response to treatment for breast cancer patients (Perou et al. 2000; Dent et al. 2007). For non-small cell lung cancer patients, a composite score consisting of several biological markers including cyclin E and Ki-67 was shown to be highly predictive of patient survival (Dosaka-Akita et al. 2001). However, construction of accurate prediction models with a panel of markers is a difficult task in general. For example, statistical models for calculating individual cancer risk have been developed for a few types of cancer in the past two decades (Gail et al. 1989; Thompson et al. 2006; Cassidy et al. 2008; Freedman et al. 2009). However, much refinement is needed even for the best of these models due to their limited discriminatory accuracy (Spiegelman et al. 1994; Gail and Costantino 2001).

The increasing availability of new potential markers, while holding great promises for better prediction of disease outcomes, imposes challenges to model development due to the high dimensionality in the feature space and the relatively small sample size. To improve prediction with a large number of promising genomic or biological markers, an important step is to build a parsimonious model that only includes important markers. Such a model could reduce the cost associated with unnecessary marker measurements and improve the prediction precision for future patients. For such purposes, various regularization procedures such as the LASSO (Tibshirani 1996; Knight and Fu 2000), the SCAD (Fan and Li 2001, 2002, 2004; Zhang et al. 2006), the adaptive LASSO (ALASSO; Zou 2006; Wang and Leng 2007), the Elastic Net

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

(Zou and Hastie 2005; Zou and Zhang 2009), and one-step local linear approximation (LLA; Zou and Li 2008) have been developed in recent years. These procedures simultaneously identify non-informative variables and produce coefficient estimates for the selected variables to induce a model for prediction.

These regularization procedures, while effective for variable selection and stable estimation, yield estimators whose distributions are difficult to approximate. LASSO type estimators have a non-standard limiting distribution that depends on which components of the coefficient vector are zero. Since the LASSO type estimator is not consistent in variable selection, the limiting distribution cannot be estimated directly. Furthermore, standard bootstrap methods fail when the true coefficient vector is sparse (Knight and Fu 2000). Recently, Chatterjee and Lahiri (2010) proposed a truncated LASSO estimator whose distribution can be approximated using a residual bootstrap procedure. To overcome the difficulties in LASSO estimators, other regularized procedures such as the SCAD and ALASSO have been proposed. These estimators possess asymptotic *oracle* properties including perfect variable selection and super efficiency. However, our simulation results suggest that in finite samples, such oracle properties are far from being true and inference procedures based on asymptotic properties such as those given in Zou (2006) perform poorly especially when the signal to noise ratio (SNR) is high and the between covariate correlations are not low. Recently, Pötscher and Schneider (2009, 2010) developed theory on the coverage probabilities of the confidence intervals for ALASSO type estimators under the orthogonal design. It was shown that estimating the distribution function of the ALASSO estimator is not feasible when the true parameter is of similar magnitude to $n^{-\frac{1}{2}}$, where n is the sample size. It is thus generally difficult to develop well performed confidence regions (CRs) and hypothesis testing procedures based on these regularized estimators. Such difficulties limit their applicability to clinical studies

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

where confidence in statistical evidence is crucial for clinical decision making.

In this paper, we propose resampling methods to derive CR and testing procedures for marker effects estimated from regularized procedures such as the ALASSO and one-step SCAD estimator when the true parameter is fixed. Our preliminary studies suggest that CRs constructed from such resampling procedures perform much better than their asymptotic based counterparts. When the fitted model is merely a *working model*, many frequently used estimation procedures may fail to produce stable parameter estimates. Procedures that can provide stable parameter estimates and valid interval estimates under a possibly misspecified working model are highly valuable when building a prediction model with high dimensional data. Our proposed procedures remain valid even if the fitted model fails to hold, provided that the employed objective function satisfies mild regularity conditions. The rest of the paper is organized as follows. In Section 2, we introduce the proposed perturbation resampling procedures and describe various methods for constructing confidence regions. In Section 3, we demonstrate the validity of the proposed procedures in finite samples via simulation studies. In Section 4, we illustrate our proposed procedure with an HIV drug resistance study where the goal is to predict phenotypic drug resistance levels using genotypic viral mutations.

2. RESAMPLING PROCEDURES

Suppose that $\mathbf{y} = (y_1, \ldots y_n)^{\mathsf{T}}$ is the $n \times 1$ vector of response variables and $\mathbf{x}_j = (x_{1i}, \ldots, x_{pi})^{\mathsf{T}}, i = 1 \ldots n$, are the predictors. Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^{\mathsf{T}}$ be the $n \times p$ matrix of these covariates. Assume that the effect of $\mathbf{x}$ on y is determined via an objective function $\mathcal{L}(\boldsymbol{\theta}; \mathcal{D}) = \ell(y, \alpha + \boldsymbol{\beta}^{\mathsf{T}}\mathbf{x})$, where $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^{\mathsf{T}})^{\mathsf{T}}$, α is an unknown location parameter, $\boldsymbol{\beta}$ is an unknown $p \times 1$ vector of covariate effects, and $\mathcal{D} = (y, \mathbf{x}^{\mathsf{T}})^{\mathsf{T}}$. To assess the association between $\mathbf{x}$ and y, let $\widetilde{\mathcal{L}}(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \mathcal{L}(\boldsymbol{\theta}; \mathcal{D}_i)$ be the objective function used to fit a regression model and $\widetilde{\boldsymbol{\theta}} = (\widetilde{\alpha}, \widetilde{\boldsymbol{\beta}}^{\mathsf{T}})^{\mathsf{T}} = \operatorname{argmin}_{\boldsymbol{\theta}} \widetilde{\mathcal{L}}(\boldsymbol{\theta})$. To obtain

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

a regularized estimator for $\boldsymbol{\theta}_0$, we minimize the regularized objective function

$$\widehat{\mathcal{L}}(\boldsymbol{\theta}) = \widetilde{\mathcal{L}}(\boldsymbol{\theta}) + \sum_{j=1}^{p} p'_{\lambda_n j}(|\widetilde{\beta}_j|)|\beta_j| \tag{1}$$

where $p'_{\lambda_n j}(|\widetilde{\beta}_j|)$ is the derivative of a penalty $p_{\lambda_n j}(|\beta_j|)$ evaluated at the initial estimate of β_{0j}. We consider the cases where $p_{\lambda_n j}(|\beta_j|)$ is the concave SCAD penalty or the L_q penalty for $0 < q < 1$, and utilize a one-step estimator of these penalties with the local linear approximation (LLA) method proposed by Zou and Li (2008). Additionally, we consider the ALASSO penalty of Zou (2006) that arises when $p'_{\lambda_n j}(|\widetilde{\beta}_j|) = n^{-\frac{1}{2}}\lambda_n|\widetilde{\beta}_j|^{-1}$.

2.1 Regularity Conditions

To ensure the asymptotic oracle properties of the regularized estimators and the validity of the proposed resampling procedures, we require the following set of conditions:

C1. $\mathbb{P}\{\mathcal{L}(\boldsymbol{\theta}; \mathcal{D})\}$ has a unique minimum at $\boldsymbol{\theta}_0$ and a continuous secondary derivative with a positive definite $\mathbb{A} = \partial^2\mathbb{P}\{\mathcal{L}(\boldsymbol{\theta}; \mathcal{D})\}/\partial\boldsymbol{\theta}\boldsymbol{\theta}^{\mathsf{T}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} > 0$, where $\mathbb{P}$ is the probability measure generated by the data $\mathcal{X} = \{\mathcal{D}_i, i = 1, ..., n\}$.

C2. The class of functions indexed by $\boldsymbol{\theta}$, $\{\mathcal{L}(\boldsymbol{\theta}; \mathcal{D}) \mid \boldsymbol{\theta} \in \Omega\}$, is Glivenko-Cantelli (Kosorok 2008), where Ω is the compact parameter space containing $\boldsymbol{\theta}_0$.

C3. There exists a "qausi-derivative" function $\mathcal{U}(\boldsymbol{\theta}; \mathcal{D})$ for $\mathcal{L}(\boldsymbol{\theta}; \mathcal{D})$ such that for any positive sequence $\delta_n \to 0$

(a) $\mathbb{P}\{\mathcal{U}^2(\boldsymbol{\theta}_0; \mathcal{D})\} = \mathbb{B}$, a positive definite matrix.

(b) $\mathbb{P}\{\mathcal{L}(\boldsymbol{\theta}; \mathcal{D}) - \mathcal{L}(\boldsymbol{\theta}_0; \mathcal{D}) - \mathcal{U}(\boldsymbol{\theta}_0; \mathcal{D})(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\} = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T\mathbb{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2)$, where $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta_n$.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

(c) For $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta_n = o(1)$,

$$\sup_{\boldsymbol{\theta}} \mathbb{P} \left\{ \frac{\mathcal{L}(\boldsymbol{\theta}; \mathcal{D}) - \mathcal{L}(\boldsymbol{\theta}_0; \mathcal{D}) - \mathcal{U}(\boldsymbol{\theta}_0; \mathcal{D})(\boldsymbol{\theta} - \boldsymbol{\theta}_0)}{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|} \right\}^2 = o(1)$$

(d) The class of functions

$$D_n = \left\{ \frac{\mathcal{L}(\boldsymbol{\theta}; \mathcal{D}) - \mathcal{L}(\boldsymbol{\theta}_0; \mathcal{D}) - \mathcal{U}(\boldsymbol{\theta}_0; \mathcal{D})(\boldsymbol{\theta} - \boldsymbol{\theta}_0)}{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|} \;\middle|\; \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta_n \right\}$$

is Donsker (Kosorok 2008, p11).

These conditions are parallel to the conditions required in Proposition A1-A3 in Jin et al. (2001). These conditions also guarantee that $\widetilde{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$ and $n^{\frac{1}{2}}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to $N(\mathbf{0}, \mathbb{A}^{-1}\mathbb{B}\mathbb{A}^{-1})$. Let $\mathcal{A} = \{j : \beta_{0j} \neq 0\}$ of size $p_{\neq 0}$ and $\mathcal{A}^c = \{j : \beta_{0j} = 0\}$, where a_j denotes the jth component of a vector $\boldsymbol{a}$.

Following similar arguments to those given in Zou (2006), Zou and Li (2008) and the unconditional arguments given in the Appendix, $\widehat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}} \widehat{\mathcal{L}}(\boldsymbol{\theta})$ has 'good' properties for certain choices of λ_n, including the oracle property,

Lemma 1: (Oracle properties) Suppose that $\lambda_n \to 0$ and $\lambda_n n^{\frac{1}{2}} \to \infty$. Then the regularized estimates must satisfy the following:

1. Consistency in variable selection: $\lim_n \text{pr}(\widehat{\mathcal{A}} = \mathcal{A}) = 1$, where $\widehat{\mathcal{A}} = \{j : \hat{\beta}_j \neq 0\}$
2. Asymptotic normality: $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{\mathcal{A}} - \boldsymbol{\theta}_{0\widehat{\mathcal{A}}}) \to_d \mathcal{N}(\mathbf{0}, \mathbb{A}_{11}^{-1}\mathbb{B}_{11}\mathbb{A}_{11}^{-1})$, where $\mathbb{A}_{11}$ and $\mathbb{B}_{11}$ are the respective $p_{\neq 0} \times p_{\neq 0}$ submatrices of $\mathbb{A}$ and $\mathbb{B}$ corresponding to $\mathcal{A}$.

This lemma guarantees that the regularized estimate asymptotically chooses the correct model and has the optimal estimation rate. However, estimating the distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ in finite samples remains difficult. To estimate the standard errors of the SCAD estimates $\widehat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}} \{\widetilde{\mathcal{L}}(\boldsymbol{\theta}) + \sum_{j=1}^{p} p_{\lambda_n j}(|\beta_j|)\}$ when

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

$\widetilde{\mathcal{L}}(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\mathcal{L}(\boldsymbol{\theta};\mathcal{D}_i)$ is smooth in $\boldsymbol{\theta}$, Fan and Li (2001) proposed a local quadratic approximation (LQA) method. This gives a sandwich estimator for the covariance matrix of the estimated nonzero parameters:

$$\widehat{\mathrm{cov}}\left(\widehat{\boldsymbol{\theta}}_{\widehat{\mathcal{A}}}\right) = \{\nabla^2\widetilde{\mathcal{L}}(\widehat{\boldsymbol{\theta}}_{\widehat{\mathcal{A}}}) + \Sigma_\lambda(\widehat{\boldsymbol{\theta}}_{\widehat{\mathcal{A}}})\}^{-1}\widehat{\mathrm{cov}}\{\nabla\widetilde{\mathcal{L}}(\widehat{\boldsymbol{\theta}}_{\widehat{\mathcal{A}}})\}\{\nabla^2\widetilde{\mathcal{L}}(\widehat{\boldsymbol{\theta}}_{\widehat{\mathcal{A}}}) + \Sigma_\lambda(\widehat{\boldsymbol{\theta}}_{\widehat{\mathcal{A}}})\}^{-1} \qquad (2)$$

where $\nabla\widetilde{\mathcal{L}}(\widehat{\boldsymbol{\theta}}_{\widehat{\mathcal{A}}}) = \partial\widetilde{\mathcal{L}}(\widehat{\boldsymbol{\theta}}_{\widehat{\mathcal{A}}})/\partial\boldsymbol{\theta}$, $\nabla^2\widetilde{\mathcal{L}}(\widehat{\boldsymbol{\theta}}_{\widehat{\mathcal{A}}}) = \partial^2\widetilde{\mathcal{L}}(\widehat{\boldsymbol{\theta}}_{\widehat{\mathcal{A}}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$, and $\Sigma_\lambda(\widehat{\boldsymbol{\theta}}_{\widehat{\mathcal{A}}})$ is a diagonal matrix with the (j,j)th element being $I(\widehat{\beta}_j \neq 0)p'_{\lambda_n 1}(|\widehat{\beta}_j|)/|\widehat{\beta}_j|$. The LQA approach can also be used to construct a covariance estimate for the ALASSO estimates where $p'_{\lambda_n j}(|\widetilde{\beta}_j|) = n^{-\frac{1}{2}}\lambda_n|\widetilde{\beta}_j|^{-1}$. Similar to covariance estimates in Tibshirani (1996) and Fan and Li (2001) for penalized estimates, this procedure estimates the standard errors for variables with $\hat{\beta}_j = 0$ as 0. Although this sandwich estimator has been proven to be consistent (Fan and Peng 2004) under the linear regression model, it tends to underestimate the standard errors, and normal confidence regions (CRs) using this estimate often do not provide acceptable coverage in finite sample.

To approximate the covariance of $\widehat{\boldsymbol{\theta}}$ more accurately, we propose a perturbation method to estimate the distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ for a general class of objective functions and penalties. Let $\mathcal{G} = \{G_i, i = 1, \ldots, n\}$ be a set of independent and identically distributed $(i.i.d)$ positive random variables with mean and variance equal to one. We first perturb the initial objective function and obtain

$$\widetilde{\mathcal{L}}^*(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\mathcal{L}(\boldsymbol{\theta},\mathcal{D}_i)G_i, \qquad \text{and} \qquad \widetilde{\boldsymbol{\theta}}^* = \underset{\boldsymbol{\theta}}{\mathrm{argmin}}\,\widetilde{\mathcal{L}}^*(\boldsymbol{\theta}).$$

Then with the same set $\mathcal{G}$, we obtain the minimizer of a stochastically perturbed version of the regularized objective function:

$$\widehat{\mathcal{L}}^*(\boldsymbol{\theta}) = \widetilde{\mathcal{L}}^*(\boldsymbol{\theta}) + \sum_{j=1}^{p} p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|)|\beta_j| \qquad (3)$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

where λ_n^* satisfies the same order constraints as λ_n as discussed in the Lemma 1. In practice, one may select λ_n and λ_n^* based on the BIC criterion with the corresponding objective functions. In the appendix we first show that $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}} - \boldsymbol{\theta}_{0\mathcal{A}})$ converges in distribution to $N(\mathbf{0}, \mathbb{A}_{11}^{-1}\mathbb{B}_{11}\mathbb{A}_{11}^{-1})$, the same limiting distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. Furthermore, $\mathbb{P}^*(\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}^c} = 0) \rightarrow 1$, where $\mathbb{P}^*$ is the probability measure generated by both $\mathcal{X}$ and $\mathcal{G}$. In addition, we show that the distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}} - \widehat{\boldsymbol{\theta}}_{\mathcal{A}})$ conditional on the data can be used to approximate the unconditional distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{\mathcal{A}} - \boldsymbol{\theta}_{\mathcal{A}0})$ and that $\mathbb{P}^*(\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}^c} = 0 \mid \mathcal{X}) \rightarrow 1$. In practice, these results allows us to estimate the distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ by generating a large number, M, say, of random samples $\mathcal{G}$. We obtain $\widehat{\boldsymbol{\theta}}^*_m$ by minimizing the perturbed objective function for each sample $m = 1, \ldots M$, and then approximate the theoretical distribution of $\widehat{\boldsymbol{\theta}}$ by the empirical distribution $\{\widehat{\boldsymbol{\theta}}^*_m, m = 1, \ldots M\}$. Specifically, the covariance matrix of $\widehat{\boldsymbol{\theta}}$ can be estimated by the sample covariance matrix constructed from $\{\widehat{\boldsymbol{\theta}}^*_m, m = 1, \ldots M\}$.

Estimating the distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ based on the distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}}) \mid \mathcal{X}$ leads to the construction of three possible $(1-\alpha)100\%$ confidence regions for $\boldsymbol{\theta}_0$. For the first, let $\widehat{\sigma}_j^2 = M^{-1}\sum_{m=1}^{M}(\widehat{\beta}^*_{mj} - \widehat{\beta}_j)^2$. We construct a normal CR for β_{0j}, $\mathrm{CR}_j^{*\mathrm{N}}$, centered at $\widehat{\beta}_j$ with standard deviation $\widehat{\sigma}_j^*$. This method is in contrast to $\mathrm{CR}^{\mathrm{Asym}}$ obtained with standard deviations $\widehat{\sigma}_j^{\mathrm{Asym}}$ estimated with the asymptotically consistent LQA sandwich estimator in Fan and Li (2001) and Zou (2006). In contrast to setting the standard error to 0 when $\widehat{\beta}_j = 0$, we set $\mathrm{CR}_j^{*\mathrm{N}} = \{0\}$ if the proportion of $\widehat{\beta}_j^*$ being 0 is larger than a threshold $\widehat{p}_{high}$, such that $\widehat{p}_{high} \rightarrow p_{high} < 1$. This method accounts for the superefficiency due to the oracle property and results in a shorter interval with valid coverage. Secondly, we simply take the $(\alpha/2)100$th and $(1-\alpha/2)100$th quantiles of $\widehat{\beta}_j^*$ as the upper and lower bounds of $\mathrm{CR}_j^{*\mathrm{Q}}$. For the third, we estimate the density of $\widehat{\beta}_j^*$ with a kernel density estimator and choose the $(1-\alpha)100\%$ highest density region, $\mathrm{CR}_j^{*\mathrm{HDR}}$. We estimate the density of $\widehat{\beta}_j^* \mid \mathcal{X}$ as a mixed density with

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

distribution $f_j^*(\beta) = \widehat{\mathcal{P}}_{0j} I(\beta = 0) + (1 - \widehat{\mathcal{P}}_{0j}) f_j^*(\beta)$,where $\widehat{\mathcal{P}}_{0j}$ is the proportion of $\widehat{\beta}_j^*$ set to 0, and $f_j^*(\beta)$ is the unknown distribution of $\widehat{\beta}_j^*$ given that it is not set to 0. Thus, our highest density confidence region $\mathrm{CR}_j^{*\mathrm{HDR}}$ is defined as

$$\mathrm{CR}_j^{*\mathrm{HDR}} = \begin{cases} \{0\} & \text{if } \widehat{\mathcal{P}}_{0j} \geq \widehat{p}_{high} \\ \{\beta : f_j^*(\beta) \geq \widehat{c}_1\} \cup \{0\} & \text{if } \widehat{p}_{low} \leq \widehat{\mathcal{P}}_{0j} < \widehat{p}_{high} \\ \{\beta : f_j^*(\beta) \geq \widehat{c}_2\} \cup \{0\} & \text{if } \alpha \leq \widehat{\mathcal{P}}_{0j} < \max(\alpha, \widehat{p}_{low}) \\ \{\beta : f_j^*(\beta) \geq \widehat{c}_3\} & \text{if } \widehat{\mathcal{P}}_{0j} < \alpha \end{cases}$$

where $\widehat{c}_1$, $\widehat{c}_2$, and $\widehat{c}_3$ are chosen such that for $H(c) = \int I\{f_j^*(\beta) \geq c\} f_j^*(\beta) d\beta$, we have $H(\widehat{c}_1) = (1 - \alpha - \widehat{\mathcal{P}}_{0j})/(1 - \widehat{\mathcal{P}}_{0j}), H(\widehat{c}_2) = 1 - \alpha + \alpha(\widehat{\mathcal{P}}_{0j} + \widehat{p}_{low}), H(\widehat{c}_3) = 1 - \alpha$, while $\widehat{p}_{low} \to 0$ and $\widehat{p}_{high} \to p_{high} < 1$. The details of this method are relegated to the Appendix. Note that $\{0\}$ is included in our confidence region when $\widehat{\mathcal{P}}_0$ is sufficiently large, or when $f_j^*(\beta)$ is sufficiently large around a neighborhood of 0.

In practice, when assessing the effects of multiple features, it is often important to adjust for multiple comparisons. For interval estimation, we may construct a $(1 - \alpha)100\%$ simultaneous confidence region to cover the entire parameter vector $\boldsymbol{\theta}_0$. We may then make statements about the importance of each of the covariates in the presence of other covariates while maintaining a type I error of α. For the regularized estimator, we define the simultaneous region as $\mathrm{CR}^{*\mathrm{Sim}} = \prod_{j \notin \widehat{\mathcal{A}}^*} \{0\} \times \prod_{j \in \widehat{\mathcal{A}}^*} (\hat{\beta}_j - \gamma_\alpha \hat{\sigma}_j^*, \hat{\beta}_j + \gamma_\alpha \hat{\sigma}_j^*)$ where $\widehat{\mathcal{A}}^* = \{j : \widehat{\mathcal{P}}_{0j} < \widehat{p}_{high}\}$ and γ_α is the $(1 - \alpha)100\%$ quantile of $\left\{ \max \left\{ |\hat{\beta}_{jm}^* - \hat{\beta}_j| / \widehat{\sigma}_j^* \right\}_{j \in \widehat{\mathcal{A}}^*} \right\}_{m=1}^{M}$. We compare the performance of these confidence regions with numerical examples in Sections 3 and 4.

3. SIMULATION STUDIES

To examine the validity of our procedures in finite samples, we performed simulation studies to assess the performance of the corresponding confidence regions. For

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

each setting, we simulated 1500 data sets with n observations generated under the linear model, $y = \mathbf{X}\boldsymbol{\beta} + \epsilon$, where $x_{ij} \sim \mathcal{N}(0,1)$, the pairwise correlation between $\mathbf{x}_i$ and $\mathbf{x}_j$ was set to $cor(\mathbf{x}_i, \mathbf{x}_j) = \rho$, $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, and $\boldsymbol{\beta}, \rho$, and σ were varied between settings. In each setting, $\boldsymbol{\beta}$ was sparse and included medium and high signals. We obtained ALASSO estimators for each simulated data set with λ chosen by the BIC and then $M = 500$ perturbed samples using our proposed method with $\mathcal{G}$ generated from a mean 1 exponential distribution. The sample size n was set to 100, 200, 400, or 1000, while ρ was 0, 0.2, or 0.5, and σ was 1 or 2. To compute the highest density regions $\text{CR}^{*\text{HDR}}$ we utilized the `hdrcde` package in `R` with the "ndr" bandwidth estimator as presented in Scott (1992) based on Silverman's rule of thumb (Silverman 1986). We chose $\widehat{p}_{low} = \min\{(\log(n)/n)^{(1-p/n)/4}, 0.49\}$ and $\widehat{p}_{high} = \min\{1 - n^{-(1+p/n)/4}, 0.95\}$. We substituted the σ used in the standard deviation estimate from Zou (2006) analogous to equation (2) with the known σ from the simulations.

We present the results for simulations with $n = 100$, 200 and 400 when $\sigma = 1$ or 2 and $p = 10$ or 20. In these cases, the true $\boldsymbol{\beta}_0$ contains two large effects of $\beta_{0j} = 1$, two moderate effects of $\beta_{0j} = 0.5$, and six (for $p = 10$) or sixteen (for $p = 20$) noise parameters where $\beta_{0j} = 0$. To examine the effect of regularization we compare our CRs for the regularized estimators to CR^{OLS}, the normal CR based on the empirical standard error of the perturbed ordinary least squares (OLS) estimates.

[Table 1 and 2 about here.]

In Tables 1 and 2 we see that $\text{CR}^{*\text{N}}$ usually has better coverage than CR^{Asym} and CR^{OLS} and sacrifices very little in length. The asymmetric $\text{CR}^{*\text{HDR}}$ has higher coverage than CR^{Asym} and CR^{OLS} for the moderate signal $\beta_{0j} = 0.5$ and has the shortest length when $\beta_{0j} = 0$. The other perturbation based confidence region $\text{CR}^{*\text{Q}}$ performs similarly to CR^{OLS} when β_{0j} is nonzero. The standard deviation estimate from Zou (2006), $\widehat{\sigma}^{\text{Asym}}$ (also see Table 3), is not large enough to cover β_{0j} sufficiently,

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

and while the coverage probability of the CR^{OLS} is not extremely low, it is notably outperformed by the other confidence regions when $\beta_{0j} = 0$. We omit the results from the settings where $n = 1000$ because the results have similar patterns as those with $n = 400$. For these large sample cases with n greater than or equal to 400 we saw convergence to 95% coverage for the normal CRs, highest density regions, and OLS CRs in all settings when the true parameter was nonzero. For true zero parameters, the coverage probabilities of our confidence regions converged to 1, while the OLS CR converged to 0.95. A tradeoff associated with our method is that while the coverage of our perturbation confidence regions tends to be higher than CR^{OLS} and $\mathrm{CR}^{\text{Asym}}$, some power is sacrificed for moderate signals of $\beta_{0j} = 0.5$. This loss is minimal, however, and only appears in difficult cases when sample size is low and ρ and σ are high. We see that when $\sigma = 1$ all CRs except CR^{OLS} perform well, though $\mathrm{CR}^{\text{Asym}}$ still tends to have slightly lower coverage than perturbation based regions and often does not reach 95% coverage. Also, when $\beta_{0j} = 0$, CR^{OLS} has coverage lower than 95% for small samples while our methods produce regions with coverage probability near 1 and very short lengths reflecting the oracle properties. Overall, the most disparity between our methods and previous methods is seen when the SNR is low.

The coverage probabilities and lengths of our simultaneous confidence regions are also displayed in Tables 1 and 2. We compared our method to $\mathrm{CR}^{*\text{SimOLS}}$, constructed analogously to $\mathrm{CR}^{*\text{Sim}}$ except $\widehat{\mathcal{A}}^* = \{j | j = 1, \ldots, p\}$ and $\mathrm{CR}^{*\text{SimOLS}}$ is centered at the OLS estimates and the standard error is the sample standard deviation of the perturbed OLS estimates. Our regularized $\mathrm{CR}^{*\text{Sim}}$ has the advantage of shrinking the dimension of the region by reducing some CRs to the point $\{0\}$ when $\widehat{\mathcal{P}}_{0j}$ is large. We see that our $\mathrm{CR}^{*\text{Sim}}$ outperforms $\mathrm{CR}^{*\text{SimOLS}}$ in coverage and has shorter length for $\sigma = 1$ and comparable length for $\sigma = 2$. For large sample settings when $n = 1000$, $\mathrm{CR}^{*\text{SimOLS}}$ converges further to 95% coverage with levels around 94% and $\mathrm{CR}^{*\text{Sim}}$ has

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

coverage almost always over 95%.

[Table 3 about here.]

In Table 3 we also present the standard error estimates when $\sigma = 2$. For notation, let the empirical standard deviations of the estimators $\widehat{\beta}_j$ and $\widetilde{\beta}_j$ be denoted as $\widetilde{\sigma}_j$ and $\widetilde{\sigma}_j^{\mathrm{OLS}}$, respectively. We see that our estimate of the standard error from the perturbed samples, $\widehat{\sigma}_j^*$, does well in estimating $\widetilde{\sigma}_j$. However, the standard error proposed by Zou (2006) underestimates the true standard error of the parameter estimates, especially when $\sigma = 2$ and $\beta_{0j} = 0.5$ or 0. When the SNR is higher, $\widetilde{\sigma}_j^{\mathrm{Asym}}$ estimates $\widetilde{\sigma}_j$ well except when $\beta_{0j} = 0$ because $\hat{\sigma}_j^{\mathrm{Asym}} = 0$ whereas $\widetilde{\sigma}_j$ and $\widehat{\sigma}_j^*$ are clearly nonzero.

4. EXAMPLE: HIV DRUG RESISTANCE

We illustrate our methods in a real example using the HIV antiretroviral drug susceptibility data described in Rhee et al. (2003). This dataset was refined from the Stanford HIV Drug Resistance Database (available at *http://hivdb.stanford.edu/*), and is used to study the association of protease mutations with susceptibility to the protease inhibitor anti-retroviral (ARV) drug amprenavir. The data consist of mutation information at 99 protease codons in the viral genome, of which 79 contain mutations, and ARV drug resistance assays for $n = 702$ HIV infected patients. Drug resistance was measured in units of IC_{50}, the amount of drug needed to inhibit viral replication by 50% in units of fold increase compared to drug-sensitive wildtype virus. Researchers are interested in determining which protease mutations are associated with ARV resistance so that they may develop a genotype test for resistance that looks for these mutations in the patient's infecting HIV strain. Therefore, we aim to examine the effect of the presence of any of the mutations at 79 codons on IC_{50}, where higher IC_{50} measurements indicate higher levels of drug resistance. We chose to log-transform the non-negative IC_{50} outcome and represented the presence of each

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

of the mutations as a binary predictor in our regression model. We removed the fifteen mutations that occurred less than 0.5% in the data set. Recently, Wu (2009) analyzed these data with a permutation test for regression coefficients of LASSO. In this paper, we will analyze the data using ALASSO and gain inference by using our perturbation methods to construct CRs and standard errors.

For this analysis, we fit an ALASSO linear model with λ chosen to minimize the BIC. We generated M=500 perturbation variable sets $\mathcal{G}$, consisting of $n = 702$ *i.i.d.* variables from an exponential distribution with mean and variance equal to 1, and for each $\mathcal{G}$ we minimized the perturbed objective function to obtain $\widehat{\boldsymbol{\beta}}^*_m$. We constructed 95% CRs using our perturbation method and compared inference gained from CR$^{*\mathrm{N}}$ and CR$^{*\mathrm{HDR}}$ to the inference from CR$^{\mathrm{Asym}}$ and CR$^{\mathrm{OLS}}$. We estimated the σ used in the standard deviation estimate from Zou (2006) analogous to equation (2) with the known σ from the simulations.

[Figure 1 about here.]

We present a graphical summary of the analysis results in Figure 1. Previous studies by Prado et al. (2002) and results collected by Johnson et al. (2005) found that mutations at codons 10, 32, 46, 47, 50, 54, 73, 82, 84 and 90 emerge in amprenavir resistant viral genomes. Using a permutation based p-value adjusted for multiple testing, Wu (2009) determined these mutations (except 73 and 82) as well as additional codon mutations to be significantly associated with amprenavir susceptibility at the $\alpha = 0.05$ level for a total of thirteen significant associations. The ALASSO estimator estimated the same set of thirteen coefficients as nonzero. The confidence region from nonregularized estimates CR$^{\mathrm{OLS}}$ was significant for all sixteen mutations including these thirteen. However our perturbation based CR$^{*\mathrm{N}}$ and CR$^{*\mathrm{HDR}}$ for three of these mutations did include zero. We see in Figure 2 that parameters for codons 48 and 50 have clearly nonsignificant CRs, suggesting that the parameters were esti-

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

mated as zero in the perturbation samples too often to be deemed significant. The parameter for codon 32 has marginally nonsignificant confidence regions and $\widehat{\mathcal{P}}_{0j}$ is marginally close to 0.05.

[Figure 2 about here.]

Our use of ALASSO provides estimates of the effects of each mutation while adjusting for the presence of other mutations. Several studies have shown that mutations associated with resistance to protease inhibitors can have varying effects when combined with other mutations (Schumi and DeGruttola 2008; Van Marck et al. 2009). For instance, the mutation at codon 32 has been found to have no effect on resistance of the protease inhibitor drug darunavir when a mutation at codon 84 is present (Van Marck et al. 2009). Our method allows us to determine the size of associations without orthogonalizing predictors and we adjust for multiple testing with the simultaneous confidence region $\mathrm{CR}^{*\mathrm{Sim}}$. Results could be impacted by studies summarized in Johnson et al. (2005) that may not have adjusted for other mutations, and the use of LASSO estimators that do not have oracle properties in Wu (2009). Furthermore, our methods produce CRs for the coefficients of mutations that were estimated as zero. These CRs quantify the uncertainty in our estimation and can aid scientists who wish to conduct future drug therapy studies involving the codons.

5. DISCUSSION

In this paper, we address the problem of constructing a covariance estimate for parameter estimates obtained with a general objective function and concave penalty functions including adaptive LASSO and SCAD. The proposed methods for covariance estimates are simple to implement and possess the attractive property that parameters estimated as zero have nonzero standard errors. We may then construct confidence regions for each parameter estimate and obtain more meaningful inference.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

We have shown through extensive simulation studies using the ALASSO penalty that our perturbation method results in confidence regions with accurate coverage probability. The perturbation based normal CR does not sacrifice much in length and has reasonable coverage for small sample sizes. We set the CR to $\{0\}$ when the proportion of perturbed estimates set to 0 is higher than a threshold, and therefore shorten the length by utilizing the oracle property. The perturbation based highest density region has even shorter length and good coverage probability, especially for the moderate signal $\beta_{0j} = 0.5$ in comparison to all other confidence regions. The asymptotic based normal interval that uses the standard error estimate presented in Zou (2006) fails to reach nominal coverage levels due to the underestimation of the standard error, most notably when the standard error is estimated as 0 when $\widehat{\boldsymbol{\beta}} = 0$. However, our estimate of the standard error of the parameter estimates based on our perturbation samples is close to the empirical standard error of the ALASSO estimates, even for parameters estimated as 0. Additionally, we propose a simultaneous CR that adjusts for multiple comparisons. We again utilize the oracle property and reduce the dimension of our region by setting intervals to $\{0\}$ when the proportion of zero perturbed parameter estimates is high. Therefore, the average length of our region will often be shorter than the simultaneous OLS region. For instance, when all covariates are independent, the OLS length is asymptotically proportional to $\gamma_{\text{OLS}} = \max\left\{\left|(\widetilde{\beta}_j - \beta_{0j})/\sigma\right|\right\}_{j=1}^{p}$ whereas the perturbation region length is asymptotically proportional to $(p_{\neq 0}/p)\gamma$ where $\gamma = \max\left\{\left|(\widehat{\beta}_j - \beta_{0j})/\sigma\right|\right\}_{\beta_{0j} \neq 0}$. Note that $\gamma \leq \gamma_{\text{OLS}}$ and so the length of the perturbation region will be shorter than the OLS length when the true model is sparse. Similarly, when the covariates are not independent, $\left\{(\widetilde{\beta}_j - \beta_{0j})/\sigma\right\}_{j=1}^{p} \sim \mathcal{N}(\mathbf{0}, \mathrm{Corr}(\widetilde{\boldsymbol{\beta}}))$ and the perturbation region generally has shorter average length than the OLS region. Simple simulations show that when $p_{\neq 0}$ parameters are estimated as nonzero, we expect the perturbation region length

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

to be approximately 0.36 times the OLS region length when $p = 10$ and $p_{\neq 0} = 4$ and approximately 0.16 times the OLS region length when $p = 20$ and $p_{\neq 0} = 4$ for both the independent case and the compound symmetry case when $\rho = 0.5$ and $\sigma = 1$.

Additionally, it is well known that regularized estimators, while possessing asymptotic oracle properties, are prone to bias in finite samples. Bias correction for the ALASSO estimator can be achieved based on our perturbation samples. We present the technical details of the estimation of the bias in the appendix. We find that this bias correction works well in practice, especially when the signal is small or moderate, as when $\beta_{0j} = 0.5$. For example, in our simulations when $p = 20, n = 100, \rho = 0.5, \sigma = 1$, and $\beta_{0j} = 0.5$, the bias of $\widehat{\beta}_j$ is -0.071 while the bias of $\widehat{\beta}_{0j}^{\mathrm{BC}}$ is -0.042. Similar gains are seen for most settings. The bias corrected estimator has empirical standard error similar to that of the original ALASSO estimator but with smaller bias. We could construct analogous bias-corrected estimators based on other penalties and objective functions. The model size with the ALASSO and bias-corrected ALASSO estimator in our simulations is very close to 4 when $\sigma = 1$, except for the difficult cases when $n = 100$ and $p = 20$ for which the average model size is closer to 4.5. For the settings where the SNR is low with $\sigma = 2$, the oracle property is weak in finite samples and so the model size is between 5 and 6 when $p = 10$ and between 7 and 10 when $p = 20$. We note that when p is large relative to n, initial parameter estimates obtained with ridge regression can produce more stable results. Also, our methods are robust to misspecification of the model and are valid even when the true model is not sparse.

APPENDIX

A.1 Justification for the Resampling Method

To show that the distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ can be estimated by that of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}}) \mid \mathcal{X}$ under conditions C1-C3, we first consider the distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0)$ under

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

the product probability measure $\mathbb{P}^*$ generated by the data, $\mathcal{X}$, and the perturbation variables $\mathcal{G} = \{G_i, i = 1, \dots, n\}$. Throughout, we assume that the parameter space for $\boldsymbol{\theta}$, denoted by Ω, is a compact set and $\boldsymbol{\theta}_0$ is an interior point of Ω. We let $\mathbb{P}_n$ denote the empirical measure generated by $\mathcal{X}$ and $\mathbb{G}_n = n^{-\frac{1}{2}}(\mathbb{P}_n - \mathbb{P})$. We use notation $\rightarrow_p$ to denote convergence in probability.

We first show that $\widehat{\boldsymbol{\theta}}^* \rightarrow_p \boldsymbol{\theta}_0$. First note that $\sum_{j=1}^p p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|)|\beta_j| \rightarrow 0$ in probability. When the penalty is L_q, $p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|) = \lambda_n|\beta_j|^q$, $p'(|\widetilde{\beta}^*_j|) \rightarrow_p p'(|\beta_{0j}|)$ by the continuous mapping theorem and $\lambda_n \rightarrow 0$. For the SCAD penalty, $p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|) = \lambda_n I(|\widetilde{\beta}^*_j| \leq \lambda_n) + (a\lambda_n - |\widetilde{\beta}^*_j|)_+ I(|\widetilde{\beta}^*_j| > \lambda_n)/(a-1)$. We consider two cases: (i) $\beta_{0j} \neq 0$, and (ii) $\beta_{0j} = 0$. For case (i), $\lambda_n \rightarrow 0$ and $|\widetilde{\beta}^*_j| \rightarrow_p |\beta_{0j}|$. Thus, $I(|\widetilde{\beta}^*_j| \leq \lambda_n) \rightarrow_p 0$ and $(a\lambda_n - |\widetilde{\beta}^*_j|)_+ \rightarrow_p 0$. For case (ii), $\lambda_n \rightarrow 0$ and $(a\lambda_n - |\widetilde{\beta}^*_j|)_+ \rightarrow_p 0$. Finally, for the ALASSO penalty, $p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|) = \lambda_n|n^{\frac{1}{2}}\widetilde{\beta}^*_j|^{-1}$, $|n^{\frac{1}{2}}\widetilde{\beta}^*_j| = O_{\mathbb{P}}(1)$, and $\lambda_n \rightarrow 0$. Now, since the class of functions indexed by $\boldsymbol{\theta}$, $\{\mathcal{L}(\boldsymbol{\theta}; \mathcal{D})G : \boldsymbol{\theta} \in \Omega\}$, is Glivenko-Cantelli, $|\widehat{\mathcal{L}}^*(\boldsymbol{\theta}) - \mathbb{P}\{\mathcal{L}(\boldsymbol{\theta}; \mathcal{D})\}| \leq |(\mathbb{P}_n - \mathbb{P})\{\mathcal{L}(\boldsymbol{\theta}; \mathcal{D})G\}| + \sum_{j=1}^p p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|)|\beta_j|$ uniformly converges to zero. This implies the convergence of $\widehat{\boldsymbol{\theta}}^* \rightarrow_p \boldsymbol{\theta}_0$.

We next show that $\|\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0\| = O_{\mathbb{P}^*}(n^{-\frac{1}{2}})$. It is sufficient to show that for any $\epsilon > 0$, there exits $C > 0$ such that

$$\mathbb{P}^*\left(\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \geq Cn^{-\frac{1}{2}}} \widehat{\mathcal{L}}^*(\boldsymbol{\theta}) > \widehat{\mathcal{L}}^*(\boldsymbol{\theta}_0)\right) > 1 - \epsilon \tag{A.1}$$

Since D_n is Donsker, the class of functions

$$\left\{\frac{\mathcal{L}(\boldsymbol{\theta}; \mathcal{D})G - \mathcal{L}(\boldsymbol{\theta}_0; \mathcal{D})G - \mathcal{U}(\boldsymbol{\theta}_0; \mathcal{D})G(\boldsymbol{\theta} - \boldsymbol{\theta}_0)}{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta_n\right\}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

is $\mathbb{P}^*$-Donsker as well and

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\delta_n}\left|\,\mathbb{G}_n\left\{\frac{\mathcal{L}(\boldsymbol{\theta};\mathcal{D})G-\mathcal{L}(\boldsymbol{\theta}_0;\mathcal{D})G-\mathcal{U}(\boldsymbol{\theta}_0;\mathcal{D})G(\boldsymbol{\theta}-\boldsymbol{\theta}_0)}{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|}\right\}\right|=o_{\mathbb{P}^*}(1).$$

This implies uniformly for $\boldsymbol{\theta}\in\{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\delta_n\}$

$$\widetilde{\mathcal{L}}^*(\boldsymbol{\theta})=\widetilde{\mathcal{L}}^*(\boldsymbol{\theta}_0)+\mathbb{P}_n\{\mathcal{U}(\boldsymbol{\theta}_0;\mathcal{D})G\}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^{\mathsf{T}}\mathbb{A}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+o_{\mathbb{P}^*}(\|\delta_n^2+n^{-\frac{1}{2}}\delta_n\|). \tag{A.2}$$

Letting $\boldsymbol{\theta}=\boldsymbol{\theta}_0+n^{-\frac{1}{2}}\mathbf{u}$, it follows from (A.2) that we may approximate $n\{\widehat{\mathcal{L}}^*(\boldsymbol{\theta}_0+n^{-\frac{1}{2}}\mathbf{u})-\widehat{\mathcal{L}}^*(\boldsymbol{\theta}_0)\}$ with $\mathbb{G}_n\{\mathcal{U}(\boldsymbol{\theta}_0;\mathcal{D})G\}\mathbf{u}+\frac{1}{2}\mathbf{u}^{\mathsf{T}}\mathbb{A}\mathbf{u}+n\sum_{j=1}^{p}p'_{\lambda_n^*j}(|\widetilde{\beta}_j^*|)\left(\left|\beta_{0j}+n^{-\frac{1}{2}}u_j\right|-|\beta_{0j}|\right)+o_{\mathbb{P}^*}(\|\mathbf{u}\|^2+\|\mathbf{u}\|)$. Therefore, $\frac{1}{2}\mathbf{u}^{\mathsf{T}}\mathbb{A}\mathbf{u}$ is the dominating term for the difference and one may select sufficiently big C such that (A.1) holds.

Now we show the "consistency" of variable selection, i.e., $\mathbb{P}^*(\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}^c}=0)\to 1$ as $n\to\infty$. It suffices to to show that for any constant C and given $\tilde{\boldsymbol{\theta}}_{\mathcal{A}}$ such that $\|\tilde{\boldsymbol{\theta}}_{\mathcal{A}}-\boldsymbol{\theta}_{0\mathcal{A}}\|=O_{\mathbb{P}^*}(n^{-\frac{1}{2}})$

$$\mathbb{P}^*\left[\text{argmin}_{\|\boldsymbol{\theta}_{\mathcal{A}^c}\|\le Cn^{-\frac{1}{2}}}\widehat{\mathcal{L}}^*\left\{\left(\tilde{\boldsymbol{\theta}}_{\mathcal{A}}^{\mathsf{T}},\boldsymbol{\theta}_{\mathcal{A}^c}^{\mathsf{T}}\right)^{\mathsf{T}}\right\}=0\right]\to 1. \tag{A.3}$$

Let $\tilde{\mathbf{u}}_{\mathcal{A}}$ and $\mathbf{u}_{\mathcal{A}^c}$ denote $n^{\frac{1}{2}}(\tilde{\boldsymbol{\theta}}_{\mathcal{A}}-\boldsymbol{\theta}_{0\mathcal{A}})$ and $n^{\frac{1}{2}}\boldsymbol{\theta}_{\mathcal{A}^c}$, respectively. It follows from (A.2)

$$\begin{aligned}
&n\left[\widehat{\mathcal{L}}^*\left\{\left(\boldsymbol{\theta}_{0\mathcal{A}}^{\mathsf{T}}+n^{-\frac{1}{2}}\tilde{\mathbf{u}}_{\mathcal{A}}^{\mathsf{T}},n^{-\frac{1}{2}}\mathbf{u}_{\mathcal{A}^c}^{\mathsf{T}}\right)^{\mathsf{T}}\right\}-\widehat{\mathcal{L}}^*\left\{\left(\boldsymbol{\theta}_{0\mathcal{A}}^{\mathsf{T}}+n^{-\frac{1}{2}}\tilde{\mathbf{u}}_{\mathcal{A}}^{\mathsf{T}},0^{\mathsf{T}}\right)^{\mathsf{T}}\right\}\right]\\
&=\left[\mathbb{G}_n\{\mathcal{U}(\boldsymbol{\theta}_0;\mathcal{D})_{\mathcal{A}^c}^{\mathsf{T}}G\}+\tilde{\mathbf{u}}_{\mathcal{A}}^{\mathsf{T}}\mathbb{A}_{12}\right]\mathbf{u}_{\mathcal{A}^c}+\frac{1}{2}\mathbf{u}_{\mathcal{A}^c}^{\mathsf{T}}\mathbb{A}_{22}\mathbf{u}_{\mathcal{A}^c}+n\sum_{j\in\mathcal{A}^c}p'_{\lambda_n^*j}(|\widetilde{\beta}_j^*|)|n^{-\frac{1}{2}}u_j|\\
&\quad+o_{\mathbb{P}^*}(\|\mathbf{u}_{\mathcal{A}^c}\|^2+\|\mathbf{u}_{\mathcal{A}^c}\|)=\sum_{j\in\mathcal{A}^c}n^{\frac{1}{2}}p'_{\lambda_n^*j}(|\widetilde{\beta}_j^*|)\,|u_j|+R_n(\mathbf{u}_{\mathcal{A}^c}).
\end{aligned}$$

where $\sup_{\|\mathbf{u}_{\mathcal{A}^c}\|\le C}R_n(\mathbf{u}_{\mathcal{A}^c})/\left(\|\mathbf{u}_{\mathcal{A}^c}\|^2+\|\mathbf{u}_{\mathcal{A}^c}\|\right)=o_{\mathbb{P}^*}(1)$. Zou and Li (2008) consider the limiting behavior of $n^{\frac{1}{2}}p'_{\lambda_n^*j}(|\widetilde{\beta}_j^*|)$ for SCAD and L_q penalties in their proof of the

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

oracle properties of the one-step LLA estimator. They show that for both cases, when $j \in \mathcal{A}^c$, $n^{\frac{1}{2}} p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|) \to_p \infty$. Additionally, for the ALASSO penalty, $n^{\frac{1}{2}} p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|) = n^{-\frac{1}{2}}\lambda_n |n^{\frac{1}{2}}\widetilde{\beta}^*_j|^{-1}$, when $j \in \mathcal{A}^c$, we have $n^{-\frac{1}{2}}\lambda_n \to \infty$ and $|n^{\frac{1}{2}}\widetilde{\beta}^*_j| = O_{\mathbb{P}^*}(1)$. Hence, for all three types of penalties, $n^{\frac{1}{2}} p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|) \to_p \infty$. Thus, for any $\epsilon > 0$, there exist $C_1 > C_0 > 0$ and N_0 such that $\mathbb{P}^*\left\{\sum_{j\in\mathcal{A}^c} n^{\frac{1}{2}} p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|)\,|u_j| \geq C_1 \sum_{j\in\mathcal{A}^c} |u_j|\right\} \geq 1-\epsilon$ and $\mathbb{P}^*\left\{C_0 \sum_{j\in\mathcal{A}^c} |u_j| \geq |R_n(\mathbf{u}_{\mathcal{A}^c})|\right\} \geq 1-\epsilon$ for $\|\mathbf{u}_{\mathcal{A}^c}\| \leq C$ and $n \geq N_0$. This implies that with probability greater than $1-2\epsilon$, $n\left[\widehat{\mathcal{L}}^*\left\{\left(\tilde{\boldsymbol{\theta}}^{\intercal}_{\mathcal{A}}, n^{-\frac{1}{2}}\mathbf{u}^{\intercal}_{\mathcal{A}^c}\right)^{\intercal}\right\} - \widehat{\mathcal{L}}^*\left\{\left(\tilde{\boldsymbol{\theta}}^{\intercal}_{\mathcal{A}}, 0^{\intercal}\right)^{\intercal}\right\}\right] \geq (C_1 - C_0)\sum_{j\in\mathcal{A}^c} |u_j| \geq 0$, which implies (A.3).

Lastly, we will justify the oracle property of $\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}}$. Since $\mathbb{P}^*(\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}^c} = 0) \to 1$, $\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}}$ can be considered as the minimizer of $\widehat{\mathcal{L}}^*_{\mathcal{A}}(\boldsymbol{\theta}_{\mathcal{A}}) = \widehat{\mathcal{L}}^*\{(\boldsymbol{\theta}^{\intercal}_{\mathcal{A}}, 0^{\intercal})^{\intercal}\}$. Following the approach of Zou (2006), we consider the reparametrization

$$\widehat{\mathcal{L}}^*_{\mathcal{A}}(\boldsymbol{\theta}_{0\mathcal{A}} + n^{-\frac{1}{2}}\mathbf{u}_{\mathcal{A}}) = \mathbb{P}_n \mathcal{L}\left\{\left(\boldsymbol{\theta}^{\intercal}_{0\mathcal{A}} + n^{-\frac{1}{2}}\mathbf{u}^{\intercal}_{\mathcal{A}}, 0^{\intercal}\right)^{\intercal}, \mathcal{D}_i\right\} G_i + \sum_{j\in\mathcal{A}} p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|)\left|\beta_{0j} + n^{-\frac{1}{2}}u_j\right|. \tag{A.4}$$

Let $\hat{\mathbf{u}}^{(n)}_{\mathcal{A}} = \arg\min_{\mathbf{u}_{\mathcal{A}}} \widehat{\mathcal{L}}^*_{\mathcal{A}}(\boldsymbol{\theta}_{0\mathcal{A}} + n^{-\frac{1}{2}}\mathbf{u}_{\mathcal{A}})$. Note $\hat{\mathbf{u}}^{(n)}_{\mathcal{A}} = n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}} - \boldsymbol{\theta}_{0\mathcal{A}})$ is also the minimizer of $V^*_n(\mathbf{u}_{\mathcal{A}}) \equiv \widehat{\mathcal{L}}^*_{\mathcal{A}}(\boldsymbol{\theta}_{0\mathcal{A}} + n^{-\frac{1}{2}}\mathbf{u}_{\mathcal{A}}) - \widehat{\mathcal{L}}^*(\boldsymbol{\theta}_0)$, as $\widehat{\mathcal{L}}^*(\boldsymbol{\theta}_0)$ is a constant. Again, it follows from (A.2)

$$\begin{aligned} V^*_n(\mathbf{u}_{\mathcal{A}}) &= n^{\frac{1}{2}}\mathbf{u}^{\intercal}_{\mathcal{A}}\mathbb{P}_n\{\mathcal{U}_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{D})G\} + \frac{1}{2}\mathbf{u}^{\intercal}_{\mathcal{A}}\mathbb{A}_{11}\mathbf{u}_{\mathcal{A}} \\ &\quad + n\sum_{j\in\mathcal{A}} p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|)\left(\left|\beta_{0j} + n^{-\frac{1}{2}}u_j\right| - |\beta_{0j}|\right) + o_{\mathbb{P}^*}(\|\mathbf{u}_{\mathcal{A}}\|^2 + \|\mathbf{u}_{\mathcal{A}}\|) \end{aligned}$$

To examine the limiting behavior of the third term of $V^*_n(\mathbf{u})$, we have $\beta_{0j} \neq 0$, $n^{\frac{1}{2}}(|\beta_{j0} + n^{-\frac{1}{2}}u_j| - |\beta_{j0}|) \to_p u_j \operatorname{sgn}(\beta_{0j})$, since $j \in \mathcal{A}$. Also, as Zou and Li (2008) proved in their appendix, $n^{\frac{1}{2}} p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|) \to_p 0$ when $j \in \mathcal{A}$ for the SCAD and L_q penalties. For the ALASSO penalty, $n^{\frac{1}{2}} p'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|) = \lambda_n |\widetilde{\beta}_j|^{-1}$, $\lambda_n \to 0$, and $|\widetilde{\beta}_j|^{-1} \to_p |\beta_{0j}|^{-1}$ for $\beta_{0j} \neq 0$. Therefore, by Slutsky's theorem, we have

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

$np'_{\lambda^*_n j}(|\widetilde{\beta}^*_j|)\left(\left|\beta_{0j} + n^{-\frac{1}{2}}u_j\right| - |\beta_{0j}|\right) = o_{\mathbb{P}^*}(1)$ and

$$V^*_n(\mathbf{u}_{\mathcal{A}}) = \mathbf{u}^{\intercal}_{\mathcal{A}}\mathbb{G}_n\{\mathcal{U}_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{D})G\} + \frac{1}{2}\mathbf{u}^{\intercal}_{\mathcal{A}}\mathbb{A}_{11}\mathbf{u}_{\mathcal{A}} + o_{\mathbb{P}^*}(1 + \|\mathbf{u}_{\mathcal{A}}\|^2 + \|\mathbf{u}_{\mathcal{A}}\|).$$

Thus, $\hat{\mathbf{u}}^{(n)}_{\mathcal{A}} = -\mathbb{A}^{-1}_{11}\mathbb{G}_n\{\mathcal{U}_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{D})G\} + o_{\mathbb{P}^*}(1)$. Since $\mathbb{G}_n\{\mathcal{U}_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{D})G\}$ converges to $N(\mathbf{0}, \mathbb{B}_{11})$ in distribution, $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}} - \boldsymbol{\theta}_{0\mathcal{A}}) \rightarrow_d N(\mathbf{0}, \mathbb{A}^{-1}_{11}\mathbb{B}_{11}\mathbb{A}^{-1}_{11})$ and $\mathbb{P}^*(\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}^C} = 0) \rightarrow 1$. Then the perturbed regularized estimator $\widehat{\boldsymbol{\theta}}^*$ is asymptotically normal in the true nonzero parameter set.

Similar arguments as given above can be used to justify that the distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}}) \mid \mathcal{X}$ approximates that of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. Specifically, we can similarly obtain $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{\mathcal{A}} - \boldsymbol{\theta}_{0\mathcal{A}}) = -\mathbb{A}^{-1}_{11}\mathbb{G}_n\{\mathcal{U}_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{D})\} + o_{\mathbb{P}}(1)$ and $\mathbb{P}^*(\widehat{\boldsymbol{\theta}}_{\mathcal{A}^C} = 0) \rightarrow 1$. Therefore, $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}} - \widehat{\boldsymbol{\theta}}_{\mathcal{A}}) = -\mathbb{A}^{-1}_{11}\mathbb{G}_n\{\mathcal{U}_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{D})(G-1)\} + o_{\mathbb{P}^*}(1)$. Since $-\mathbb{A}^{-1}_{11}\mathbb{G}_n\{\mathcal{U}_{\mathcal{A}}(\boldsymbol{\theta}_0, \mathcal{D})(G-1)\} \mid \mathcal{X} \rightarrow_d N(\mathbf{0}, \mathbb{A}^{-1}_{11}\widehat{\mathbb{B}}_{11}\mathbb{A}^{-1}_{11})$ and $\widehat{\mathbb{B}}_{11} \rightarrow_{\mathbb{P}} \mathbb{B}_{11}$, $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}} - \widehat{\boldsymbol{\theta}}_{\mathcal{A}}) \mid \mathcal{X}$ and $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{\mathcal{A}} - \boldsymbol{\theta}_{0\mathcal{A}})$ converge in distribution to the same limit. Furthermore, $\mathbb{P}^*(\widehat{\boldsymbol{\theta}}^*_{\mathcal{A}^C} = 0|\mathcal{X}) \rightarrow 1$.

A.2 Justification of highest density region and bias estimate

For $j \in \mathcal{A}^C$, $\mathbb{P}^*(\widehat{\beta}^*_j = 0) \rightarrow 1$ and thus for any $\alpha > 0$, $\mathbb{P}^*(\widehat{\mathcal{P}}_{0j} > \alpha) \rightarrow 1$, and $\mathbb{P}(\widehat{\mathcal{P}}_{0j} < \widehat{p}_{high}) + \mathbb{P}(\widehat{\mathcal{P}}_{0j} < \widehat{p}_{low}) \rightarrow 0$. Hence, $\mathbb{P}^*(0 \in \text{CR}^{*\text{HDR}}_j) \rightarrow 1$. For $j \in \mathcal{A}$, $n^{\frac{1}{2}}(\widehat{\beta}^*_j - \widehat{\beta}_j) \mid \mathcal{X} \rightarrow_d N(0, \sigma^2_j)$ and $\widehat{\mathcal{P}}_{0j} \rightarrow_p 0$, where σ^2_j is the asymptotic variance of $n^{\frac{1}{2}}(\widehat{\beta}_j - \beta_{0j})$. It follows that $\sup_x |n^{-\frac{1}{2}}f^*_j(\widehat{\beta}_{0j} + n^{-\frac{1}{2}}x) - \phi_{\sigma_j}(x)| \rightarrow_p 0$ where $\phi_\sigma(x) = \phi(x/\sigma)/\sigma$ and $\phi(\cdot)$ is the density function of the standard normal. Therefore, $\sup_\beta |n^{-\frac{1}{2}}f^*_j(\beta) - \phi_{\sigma_j}\{n^{\frac{1}{2}}(\beta - \widehat{\beta}_{0j})\}| \rightarrow_p 0$ and $n^{-\frac{1}{2}}\widehat{c}_3 \rightarrow_p c_{30}$, where c_{30} is the solution to $\int I\{\phi_{\sigma_j}(x) > c_{30}\}\phi_{\sigma_j}(x)dx = 1 - \alpha$. It follows that with respect to probability measure $\mathbb{P}^*$, $\text{pr}(\beta_{0j} \in \text{CR}^{*\text{HDR}}_j) = \text{pr}\left\{f^*_j(\beta_{0j}) \geq \widehat{c}_3\right\} + o_{\mathbb{P}^*}(1) = \text{pr}\left\{n^{-\frac{1}{2}}f^*_j(\beta_{0j}) \geq n^{-\frac{1}{2}}\widehat{c}_3\right\} + o_{\mathbb{P}^*}(1) = \text{pr}\left[\phi_{\sigma_j}\{n^{\frac{1}{2}}(\beta_{0j} - \widehat{\beta}_{0j})\} \geq c_{30}\right] + o_{\mathbb{P}^*}(1) \rightarrow 1 - \alpha$.

Here we define our bias corrected estimator for β_{0j}, $\widehat{\beta}^{BC}_j = \widehat{\beta}_j + I(\widehat{\beta}_j \neq 0)\widehat{\text{bias}}_j$, where $\widehat{\text{bias}}_j = \left(\frac{1}{M}\sum^M_{m=1}\widehat{\beta}^*_{j,m}\right)(-1)^{I\left[\sum^M_{m=1}\{I(\widehat{\beta}^*_{j,m}>0) - I(\widehat{\beta}^*_{j,m}<0)\}<0\right]}\left(\widehat{\mathbb{A}}^{-1}_{\lambda}\right)_{jj}/\{n\max(|\widehat{\xi}_{7.5}|,$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

$|\widehat{\xi}_{97.5}|)\}$, $\widehat{\mathbb{A}}_\lambda = n^{-1}\left(\mathbf{X}_{\widehat{\mathcal{A}}}^{\intercal}\mathbf{X}_{\widehat{\mathcal{A}}} + n^{-\frac{1}{2}}\lambda_n \mathrm{diag}\left\{1/\widetilde{\beta}_j^2\right\}_{j=1}^{p}\right)$ and $\widehat{\xi}_r$ is the r percentile of $\{\widetilde{\beta}_{j,m}^{*}, m = 1, \ldots M\}$. We estimate $\mathbb{A}$ for ALASSO with $\widehat{\mathbb{A}}_\lambda$ following the methods of Cai et al. (2009) where a stabilized estimate of the covariance of coefficients from an accelerated failure time model is used.

REFERENCES

Cai, T., Huang, J., and Tian, L. (2009), "Regularized estimation for the accelerated failure time model," *Biometrics*, 65, 394–404.

Cassidy, A., Myles, J., van Tongeren, M., Page, R., Liloglou, T., Duffy, S., and Field, J. (2008), "The LLP risk model: an individual risk prediction model for lung cancer," *British Journal of Cancer*, 98, 270.

Chatterjee, A. and Lahiri, S. (2010), "Asymptotic properties of the residual bootstrap for lasso estimators," *Proceedings of the American Mathematical Society*, (accepted).

Dent, R., Trudeau, M., Pritchard, K., Hanna, W., Kahn, H., Sawka, C., Lickley, L., Rawlinson, E., Sun, P., and Narod, S. (2007), "Triple-negative breast cancer: clinical features and patterns of recurrence," *Clinical Cancer Research*, 13, 4429.

Dosaka-Akita, H., Hommura, F., Mishina, T., Ogura, S., Shimizu, M., Katoh, H., and Kawakami, Y. (2001), "A risk-stratification model of non-small cell lung cancers using cyclin E, Ki-67, and ras p21: different roles of G1 cyclins in cell proliferation and prognosis," *Cancer Research*, 61, 2500.

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.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

— (2002), “Variable Selection for Cox’s Proportional Hazards Model and Frailty Model,” *The Annals of Statistics*, 30, 74–99.

— (2004), “New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis,” *Journal of the American Statistical Association*, 99, 710–723.

Fan, J. and Peng, H. (2004), “On Nonconcave Penalized Likelihood With Diverging Number of Parameters,” *The Annals of Statistics*, 32, 928–961.

Freedman, A., Slattery, M., Ballard-Barbash, R., Willis, G., Cann, B., Pee, D., Gail, M., and Pfeiffer, R. (2009), “Colorectal cancer risk prediction tool for white men and women without known susceptibility,” *Journal of Clinical Oncology*, 27, 686.

Gail, M., Brinton, L., Byar, D., Corle, D., Green, S., Schairer, C., and Mulvihill, J. (1989), “Projecting individualized probabilities of developing breast cancer for white females who are being examined annually,” *Journal of the National Cancer Institute*, 81, 1879.

Gail, M. and Costantino, J. (2001), “Validating and improving models for projecting the absolute risk of breast cancer,” *Journal of the National Cancer Institute*, 93, 334.

Jin, Z., Ying, Z., and Wei, L. (2001), “A simple resampling method by perturbing the minimand,” *Biometrika*, 88, 381–390.

Johnson, V., Brun-Vézinet, F., Clotet, B., Conway, B., Kuritzkes, D., Pillay, D., Schapiro, J., Telenti, A., and Richman, D. (2005), “Update of the drug resistance mutations in HIV-1: Fall 2005,” *Top HIV Med*, 13, 125–131.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Knight, K. and Fu, W. (2000), "Asymptotics for Lasso-Type Estimators," *The Annals of Statistics*, 28, 1356–1378.

Kosorok, M. (2008), *Introduction to empirical processes and semiparametric inference*, New York: Springer Verlag.

Perou, C., Sørlie, T., Eisen, M., van de Rijn, M., Jeffrey, S., Rees, C., Pollack, J., Ross, D., Johnsen, H., Akslen, L., et al. (2000), "Molecular portraits of human breast tumours," *Nature*, 406, 747–752.

Pötscher, B. M. and Schneider, U. (2009), "On the distribution of the adaptive LASSO estimator," *Journal of Statistical Planning and Inference*, 139, 2775–2790.

— (2010), "Confidence Sets Based on Penalized Maximum Likelihood Estimators in Gaussian Regression," *Electronic Journal of Statistics*, 4, 334–360.

Prado, J., Wrin, T., Beauchaine, J., Ruiz, L., Petropoulos, C., Frost, S., Clotet, B., D'Aquila, R., and Martinez-Picado, J. (2002), "Amprenavir-resistant HIV-1 exhibits lopinavir cross-resistance and reduced replication capacity," *Aids*, 16, 1009.

Rhee, S., Gonzales, M., Kantor, R., Betts, B., Ravela, J., and Shafer, R. (2003), "HIV reverse transcriptase and sequence database," *Nucleic Acids Res*, 31, 298–303.

Schumi, J. and DeGruttola, V. (2008), "Resampling-based analyses of the effects of combinations of HIV genetic mutations on drug susceptibility," *Statistics in Medicine*, 27.

Scott, D. (1992), *Multivariate density estimation: theory, practice, and visualization*, Wiley-Interscience.

Silverman, B. (1986), *Density estimation for statistics and data analysis*, Chapman & Hall/CRC.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Spiegelman, D., Colditz, G., Hunter, D., and Hertzmark, E. (1994), "Validation of the Gail et al. model for predicting individual breast cancer risk," *Journal of the National Cancer Institute*, 86, 600.

Thompson, I., Ankerst, D., Chi, C., Goodman, P., Tangen, C., Lucia, M., Feng, Z., Parnes, H., and Coltman Jr, C. (2006), "Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial," *Journal of the National Cancer Institute*, 98, 529.

Tibshirani, R. (1996), "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society, Series B*, 58, 267–288.

Van Marck, H., Dierynck, I., Kraus, G., Hallenberger, S., Pattery, T., Muyldermans, G., Geeraert, L., Borozdina, L., Bonesteel, R., Aston, C., et al. (2009), "The Impact of Individual Human Immunodeficiency Virus Type 1 Protease Mutations on Drug Susceptibility Is Highly Influenced by Complex Interactions with the Background Protease Sequence," *Journal of Virology*, 83, 9512.

Wang, H. and Leng, C. (2007), "Unified LASSO estimation via least squares approximation," *Journal of the American Statistical Association*, 102, 1039–1048.

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

Zhang, H., Ahn, J., Lin, X., and Park, C. (2006), "Gene Selection using Support Vector Machines with Non-convex Penalty," *Bioinformatics*, 22, 88–95.

Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Zou, H. and Hastie, T. (2005), “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society* B, 67, 301–320.

Zou, H. and Li, R. (2008), “One-step sparse estimates in nonconcave penalized likelihood models,” *The Annals of Statistics*, 36, 1509–1533.

Zou, H. and Zhang, H. (2009), “On the adaptive elastic-net with a diverging number of parameters,” *The Annals of Statistics*, 37, 1733–1751.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Table 1: Coverage probabilities (lengths) of confidence regions when $\sigma = 1$.

p	β_0		$n =$ **100**			$n =$ **200**			$n =$ **400**		
			$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$
10	1	$\text{CR}^{*\text{N}}$	92.7 (40)	93.5 (42)	93.7 (55)	95.3 (28)	94.9 (30)	95.6 (38)	94.9 (20)	94.6 (21)	95.3 (26)
		$\text{CR}^{*\text{HDR}}$	92.2 (39)	92.3 (42)	92.8 (54)	94.8 (28)	94.5 (29)	94.3 (37)	94.4 (20)	94.2 (21)	94.7 (26)
		$\text{CR}^{*\text{Q}}$	91.5 (39)	92.0 (42)	92.8 (54)	94.2 (28)	94.7 (29)	94.3 (37)	93.8 (19)	94.0 (21)	94.7 (26)
		CR^{Zou}	93.3 (40)	94.5 (42)	92.3 (51)	95.3 (28)	95.1 (29)	94.8 (36)	94.9 (20)	94.6 (21)	94.6 (25)
		CR^{OLS}	91.8 (38)	91.5 (40)	91.7 (51)	94.3 (27)	94.5 (29)	94.4 (36)	94.7 (19)	93.9 (21)	93.7 (26)
	0.5	$\text{CR}^{*\text{N}}$	95.3 (45)	95.1 (49)	92.2 (64)	93.3 (30)	96.1 (32)	96.5 (42)	94.3 (20)	95.7 (22)	95.3 (28)
		$\text{CR}^{*\text{HDR}}$	92.8 (41)	93.9 (43)	92.7 (53)	91.2 (29)	93.8 (31)	93.5 (39)	93.5 (20)	95.1 (21)	94.1 (27)
		$\text{CR}^{*\text{Q}}$	90.9 (44)	90.5 (47)	90.9 (60)	90.7 (29)	93.6 (31)	92.3 (41)	93.3 (20)	94.9 (21)	94.1 (27)
		CR^{Zou}	90.9 (40)	89.7 (42)	88.3 (49)	90.6 (28)	93.7 (29)	90.5 (36)	93.4 (20)	94.4 (21)	92.7 (25)
		CR^{OLS}	92.7 (38)	91.9 (40)	90.9 (51)	92.5 (27)	94.3 (29)	92.1 (36)	93.4 (19)	94.2 (21)	94.3 (26)
	0	$\text{CR}^{*\text{N}}$	99.7 (8)	99.9 (7)	99.9 (9)	99.8 (4)	99.9 (5)	100 (5)	99.9 (2)	100 (2)	100 (3)
		$\text{CR}^{*\text{HDR}}$	99.9 (4)	100 (4)	99.9 (5)	100 (2)	99.9 (2)	100 (2)	100 (1)	100 (1)	100 (1)
		$\text{CR}^{*\text{Q}}$	100 (19)	100 (20)	100 (25)	100 (12)	99.9 (12)	100 (16)	100 (7)	100 (7)	100 (9)
		CR^{OLS}	92.4 (38)	90.9 (41)	92.8 (51)	94.3 (27)	93.3 (29)	93.3 (36)	94.5 (19)	94.7 (21)	93.9 (26)
		$\text{CR}^{*\text{Sim}}$	94.9 (29)	93.7 (30)	90.4 (39)	95.3 (19)	96.0 (20)	97.3 (26)	94.3 (13)	94.9 (13)	96.6 (17)
		$\text{CR}^{*\text{SimOLS}}$	87.7 (54)	85.8 (58)	85.7 (73)	90.7 (39)	91.9 (42)	90.5 (52)	92.5 (28)	93.2 (30)	93.3 (37)
20	1	$\text{CR}^{*\text{N}}$	92.7 (40)	93.5 (43)	93.6 (57)	93.1 (28)	94.2 (30)	94.9 (38)	93.3 (20)	94.7 (21)	94.8 (27)
		$\text{CR}^{*\text{HDR}}$	91.2 (39)	92.6 (42)	92.5 (55)	91.8 (27)	93.5 (30)	93.9 (38)	92.5 (19)	94.6 (21)	94.0 (26)
		$\text{CR}^{*\text{Q}}$	90.7 (39)	92.1 (42)	92.1 (55)	91.7 (27)	92.9 (29)	94.0 (38)	92.3 (19)	94.5 (21)	94.0 (26)
		CR^{Zou}	92.7 (40)	94.1 (42)	90.7 (52)	93.0 (28)	93.9 (29)	93.6 (36)	93.1 (20)	94.1 (21)	93.8 (25)
		CR^{OLS}	90.3 (38)	90.3 (42)	90.3 (52)	91.6 (27)	92.1 (29)	90.9 (37)	93.0 (19)	93.3 (21)	92.7 (26)
	0.5	$\text{CR}^{*\text{N}}$	93.3 (46)	93.4 (50)	90.5 (66)	93.9 (30)	94.7 (33)	96.2 (44)	93.5 (21)	93.2 (22)	95.3 (29)
		$\text{CR}^{*\text{HDR}}$	89.5 (42)	92.1 (45)	91.8 (55)	92.1 (30)	90.4 (32)	93.3 (41)	91.8 (20)	91.1 (22)	92.6 (28)
		$\text{CR}^{*\text{Q}}$	86.1 (44)	86.9 (48)	88.5 (61)	91.1 (30)	89.7 (32)	91.8 (43)	91.3 (20)	90.9 (22)	92.6 (28)
		CR^{Zou}	85.3 (40)	85.9 (42)	84.9 (48)	91.3 (28)	90.2 (29)	88.1 (36)	91.9 (20)	91.4 (21)	91.8 (25)
		CR^{OLS}	90.9 (38)	89.5 (41)	90.6 (52)	92.5 (27)	92.3 (29)	92.0 (37)	93.7 (19)	93.2 (21)	93.5 (26)
	0	$\text{CR}^{*\text{N}}$	100 (6)	100 (5)	99.9 (7)	99.7 (3)	100 (3)	99.9 (4)	100 (2)	99.9 (2)	99.8 (2)
		$\text{CR}^{*\text{HDR}}$	100 (3)	100 (3)	100 (4)	99.9 (1)	100 (1)	100 (2)	100 (1)	100 (1)	100 (1)
		$\text{CR}^{*\text{Q}}$	100 (13)	100 (14)	100 (18)	99.9 (8)	100 (9)	100 (11)	100 (5)	100 (5)	100 (7)
		CR^{OLS}	90.4 (38)	90.9 (41)	89.9 (52)	93.6 (27)	93.6 (29)	91.7 (37)	93.7 (19)	94.0 (21)	92.4 (26)
		$\text{CR}^{*\text{Sim}}$	93.0 (19)	93.6 (19)	88.3 (25)	95.3 (11)	96.6 (12)	97.2 (16)	95.3 (7)	95.7 (8)	96.5 (10)
		$\text{CR}^{*\text{SimOLS}}$	78.3 (58)	78.5 (64)	77.6 (80)	86.9 (41)	85.2 (45)	86.6 (57)	90.9 (29)	90.1 (32)	91.3 (41)

NOTE: We multiply values by 100. The lengths of the simultaneous confidence regions are averaged over the number of parameters.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Table 2: Coverage probabilities (lengths) of confidence regions when $\sigma = 2$.

p	β_0		$n =$ **100**			$n =$ **200**			$n =$ **400**		
			$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$
10	1	$\text{CR}^{*\text{N}}$	93.5 (78)	92.7 (84)	91.4 (109)	93.2 (55)	94.0 (59)	94.2 (75)	93.5 (39)	94.4 (42)	94.9 (53)
		$\text{CR}^{*\text{HDR}}$	92.7 (76)	91.3 (82)	93.9 (103)	93.2 (54)	93.2 (58)	93.0 (74)	92.9 (39)	94.0 (42)	94.5 (52)
		$\text{CR}^{*\text{Q}}$	92.8 (77)	91.0 (83)	90.9 (106)	93.2 (54)	93.1 (58)	93.2 (74)	93.1 (39)	93.7 (41)	94.4 (52)
		CR^{Zou}	82.3 (58)	80.2 (61)	78.7 (74)	81.6 (40)	82.5 (42)	79.3 (52)	82.7 (28)	82.1 (30)	82.7 (36)
		CR^{OLS}	92.0 (75)	91.7 (81)	91.2 (102)	93.2 (54)	93.8 (58)	93.1 (73)	92.8 (39)	94.0 (42)	94.5 (52)
	0.5	$\text{CR}^{*\text{N}}$	88.2 (81)	87.7 (86)	85.9 (107)	93.2 (58)	93.0 (63)	90.3 (79)	94.5 (40)	94.1 (44)	95.4 (56)
		$\text{CR}^{*\text{HDR}}$	92.3 (74)	91.4 (78)	91.7 (97)	95.3 (54)	94.6 (58)	92.4 (70)	93.7 (39)	92.7 (42)	95.9 (52)
		$\text{CR}^{*\text{Q}}$	91.1 (76)	90.7 (80)	91.8 (98)	92.5 (57)	92.3 (61)	92.8 (75)	93.8 (40)	93.0 (43)	94.4 (55)
		CR^{Zou}	79.1 (53)	76.3 (55)	70.7 (60)	79.3 (39)	78.7 (41)	76.1 (49)	81.5 (28)	79.8 (29)	79.9 (36)
		CR^{OLS}	91.3 (75)	91.5 (81)	91.6 (102)	93.0 (54)	92.6 (58)	93.2 (73)	94.1 (39)	93.1 (42)	94.3 (52)
	0	$\text{CR}^{*\text{N}}$	96.3 (69)	97.0 (72)	96.7 (92)	97.9 (51)	98.0 (54)	97.8 (66)	98.3 (36)	98.2 (39)	98.7 (49)
		$\text{CR}^{*\text{HDR}}$	98.1 (60)	98.3 (61)	98.5 (79)	99.3 (38)	98.9 (40)	99.1 (50)	99.1 (25)	99.1 (26)	99.6 (33)
		$\text{CR}^{*\text{Q}}$	99.0 (66)	98.7 (70)	99.1 (88)	99.7 (46)	99.7 (50)	99.6 (62)	99.9 (33)	99.8 (35)	99.7 (44)
		CR^{OLS}	91.5 (75)	90.9 (81)	91.2 (102)	93.7 (54)	92.7 (58)	93.4 (73)	94.0 (39)	94.3 (41)	93.9 (52)
		$\text{CR}^{*\text{Sim}}$	91.7 (110)	91.7 (116)	91.2 (148)	96.0 (82)	95.9 (87)	95.3 (109)	98.5 (59)	98.9 (64)	98.2 (80)
		$\text{CR}^{*\text{SimOLS}}$	85.6 (108)	86.2 (116)	87.1 (146)	90.7 (77)	91.3 (83)	92.2 (104)	91.5 (55)	93.1 (59)	93.1 (74)
20	1	$\text{CR}^{*\text{N}}$	91.0 (79)	91.7 (87)	91.7 (111)	92.6 (54)	92.7 (60)	93.9 (77)	93.3 (39)	94.3 (42)	93.3 (54)
		$\text{CR}^{*\text{HDR}}$	90.3 (77)	90.0 (84)	93.5 (106)	92.4 (54)	92.4 (59)	92.6 (75)	93.4 (38)	93.6 (42)	92.5 (53)
		$\text{CR}^{*\text{Q}}$	90.3 (77)	89.5 (85)	91.2 (108)	91.9 (54)	92.1 (59)	92.0 (75)	93.2 (38)	93.4 (42)	92.5 (53)
		CR^{Zou}	79.5 (59)	80.0 (63)	79.5 (78)	82.0 (40)	81.9 (43)	81.2 (53)	81.5 (28)	81.5 (30)	80.9 (37)
		CR^{OLS}	89.6 (76)	90.3 (83)	90.5 (104)	92.3 (53)	92.2 (59)	92.0 (74)	93.5 (38)	93.4 (42)	92.5 (53)
	0.5	$\text{CR}^{*\text{N}}$	86.0 (80)	84.7 (87)	81.3 (106)	92.5 (58)	91.1 (64)	88.5 (80)	93.7 (40)	94.1 (44)	93.3 (57)
		$\text{CR}^{*\text{HDR}}$	90.7 (75)	89.5 (81)	88.0 (100)	94.4 (55)	94.5 (59)	92.4 (73)	92.3 (39)	92.2 (43)	94.8 (54)
		$\text{CR}^{*\text{Q}}$	89.8 (75)	89.3 (80)	87.1 (96)	91.9 (57)	91.5 (61)	92.7 (75)	92.5 (40)	92.4 (44)	92.9 (56)
		CR^{Zou}	75.5 (53)	73.5 (55)	68.1 (62)	79.0 (40)	77.7 (42)	74.5 (49)	78.7 (28)	79.7 (30)	75.9 (37)
		CR^{OLS}	89.9 (76)	89.0 (83)	87.9 (104)	92.5 (54)	91.9 (59)	92.8 (74)	92.9 (38)	92.7 (42)	93.4 (53)
	0	$\text{CR}^{*\text{N}}$	96.7 (65)	96.4 (71)	97.3 (90)	98.1 (47)	98.2 (52)	98.3 (66)	98.3 (34)	98.8 (37)	98.3 (47)
		$\text{CR}^{*\text{HDR}}$	97.6 (61)	97.9 (66)	98.1 (83)	98.9 (41)	98.7 (45)	99.1 (58)	98.9 (27)	99.3 (30)	98.8 (38)
		$\text{CR}^{*\text{Q}}$	98.8 (61)	98.6 (66)	99.1 (84)	99.5 (44)	99.2 (47)	99.7 (60)	99.4 (31)	99.6 (33)	99.5 (42)
		CR^{OLS}	88.7 (76)	89.4 (83)	89.5 (104)	91.9 (54)	92.9 (59)	93.1 (74)	93.4 (38)	95.3 (42)	92.7 (53)
		$\text{CR}^{*\text{Sim}}$	93.9 (122)	93.4 (133)	91.9 (167)	97.9 (90)	97.7 (98)	97.1 (125)	99.7 (66)	99.6 (72)	99.3 (91)
		$\text{CR}^{*\text{SimOLS}}$	79.3 (117)	80.0 (128)	76.9 (161)	86.9 (82)	84.9 (90)	89.1 (114)	89.7 (59)	90.8 (64)	89.9 (81)

NOTE: We multiply values by 100. The lengths of the simultaneous confidence regions are averaged over the number of parameters.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

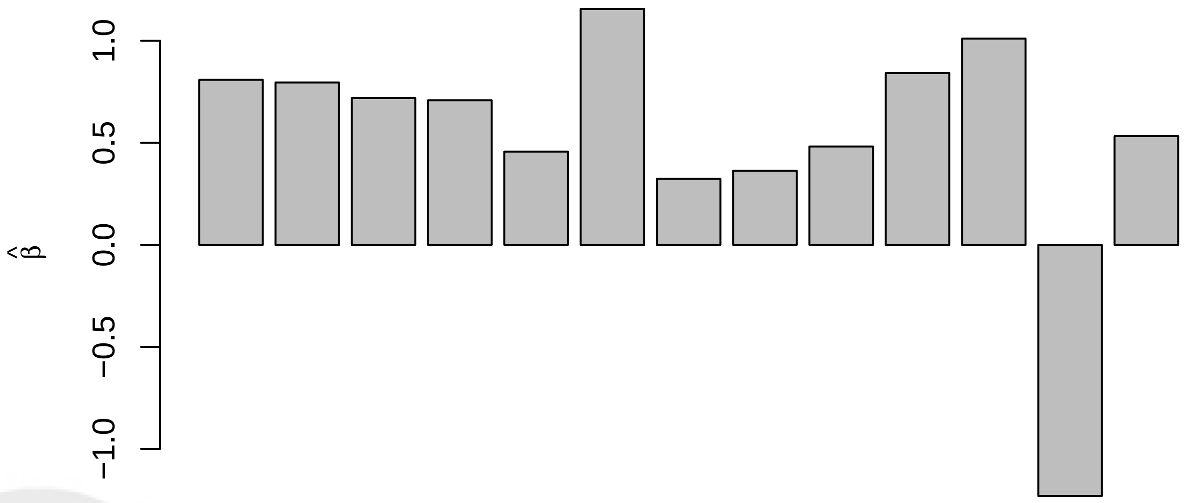
Table 3: Empirical s.d. of the parameter estimates ($\widetilde{\sigma}$) and average s.e. estimates ($\hat{\sigma}$).

p	β_0		$n =$ **100**			$n =$ **200**			$n =$ **400**		
			$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$
10	1	$\widetilde{\boldsymbol{\sigma}}$	21.2	24.0	30.1	14.8	15.6	19.9	10.4	11.0	13.5
		$\widetilde{\boldsymbol{\sigma}}^{\text{OLS}}$	20.9	23.8	28.8	14.7	15.4	19.6	10.4	11.0	13.5
		$\widehat{\boldsymbol{\sigma}}^{*}$	19.9	21.5	27.7	14.0	15.1	19.2	10.0	10.7	13.5
		$\widehat{\boldsymbol{\sigma}}^{\text{OLS}}$	19.2	20.6	25.9	13.8	14.8	18.5	9.9	10.6	13.3
		$\widehat{\boldsymbol{\sigma}}^{\text{Zou}}$	14.8	15.5	19.0	10.2	10.8	13.2	7.1	7.5	9.2
	0.5	$\widetilde{\boldsymbol{\sigma}}$	23.3	25.1	31.1	15.5	16.9	21.9	10.5	11.5	14.3
		$\widetilde{\boldsymbol{\sigma}}^{\text{OLS}}$	21.6	23.2	29.2	14.6	15.9	19.9	10.2	11.1	13.7
		$\widehat{\boldsymbol{\sigma}}^{*}$	20.6	22.1	27.8	14.8	16.0	20.2	10.3	11.1	14.3
		$\widehat{\boldsymbol{\sigma}}^{\text{OLS}}$	19.1	20.5	26.1	13.8	14.8	18.6	9.8	10.6	13.3
		$\widehat{\boldsymbol{\sigma}}^{\text{Zou}}$	13.6	14.0	15.4	10.1	10.6	12.4	7.1	7.5	9.2
	0	$\widetilde{\boldsymbol{\sigma}}$	17.3	18.0	22.8	10.7	11.3	14.3	7.2	7.5	9.4
		$\widetilde{\boldsymbol{\sigma}}^{\text{OLS}}$	21.9	23.1	29.5	14.3	15.7	19.5	10.3	10.8	13.7
		$\widehat{\boldsymbol{\sigma}}^{*}$	18.6	19.8	25.2	13.2	14.3	17.7	9.4	10.0	12.7
		$\widehat{\boldsymbol{\sigma}}^{\text{OLS}}$	19.2	20.6	26.0	13.8	14.8	18.5	9.9	10.6	13.3
		$\widehat{\boldsymbol{\sigma}}^{\text{Zou}}$	5.7	5.5	6.8	3.1	3.2	3.9	1.9	1.8	2.4
20	1	$\widetilde{\boldsymbol{\sigma}}$	22.9	25.0	31.9	14.9	16.1	20.4	10.4	11.3	14.2
		$\widetilde{\boldsymbol{\sigma}}^{\text{OLS}}$	22.9	24.8	31.3	14.9	16.2	20.6	10.5	11.5	14.5
		$\widehat{\boldsymbol{\sigma}}^{*}$	20.1	22.1	28.4	13.9	15.3	19.6	9.9	10.8	13.7
		$\widehat{\boldsymbol{\sigma}}^{\text{OLS}}$	19.3	21.2	26.5	13.6	14.9	18.8	9.8	10.7	13.4
		$\widehat{\boldsymbol{\sigma}}^{\text{Zou}}$	15.0	16.1	19.9	10.3	11.0	13.6	7.1	7.6	9.5
	0.5	$\widetilde{\boldsymbol{\sigma}}$	24.0	26.5	33.1	16.0	17.7	22.9	11.0	11.8	15.4
		$\widetilde{\boldsymbol{\sigma}}^{\text{OLS}}$	22.5	24.9	32.8	14.9	16.5	20.5	10.5	11.5	14.3
		$\widehat{\boldsymbol{\sigma}}^{*}$	20.6	22.4	27.4	14.8	16.2	20.5	10.3	11.3	14.6
		$\widehat{\boldsymbol{\sigma}}^{\text{OLS}}$	19.3	21.1	26.6	13.7	14.9	18.8	9.8	10.7	13.5
		$\widehat{\boldsymbol{\sigma}}^{\text{Zou}}$	13.5	14.1	15.9	10.2	10.8	12.6	7.2	7.6	9.4
	0	$\widetilde{\boldsymbol{\sigma}}$	17.1	17.8	22.5	10.0	10.8	13.2	6.4	6.5	8.7
		$\widetilde{\boldsymbol{\sigma}}^{\text{OLS}}$	23.6	24.9	32.1	14.8	16.1	20.5	10.3	10.8	14.3
		$\widehat{\boldsymbol{\sigma}}^{*}$	17.4	18.9	24.0	12.2	13.4	17.2	8.8	9.6	12.2
		$\widehat{\boldsymbol{\sigma}}^{\text{OLS}}$	19.3	21.1	26.6	13.7	14.9	18.9	9.8	10.7	13.5
		$\widehat{\boldsymbol{\sigma}}^{\text{Zou}}$	5.9	5.8	7.8	3.1	3.3	4.2	1.7	1.7	2.2

NOTE: We present results for settings when σ, the standard deviation of ϵ, is 2. All values are multiplied by 100. Note that $\hat{\sigma}_j^{\text{Zou}} = 0$ when $\hat{\beta}_j = 0$, but $\hat{\beta}_j$ and $\hat{\beta}_j^*$ are not always 0 in the simulations, and therefore the average $\hat{\sigma}_j^{\text{Zou}}$ is nonzero.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Figure 1: Perturbation methods results denoting significant associations between genetic mutations and drug susceptibility.

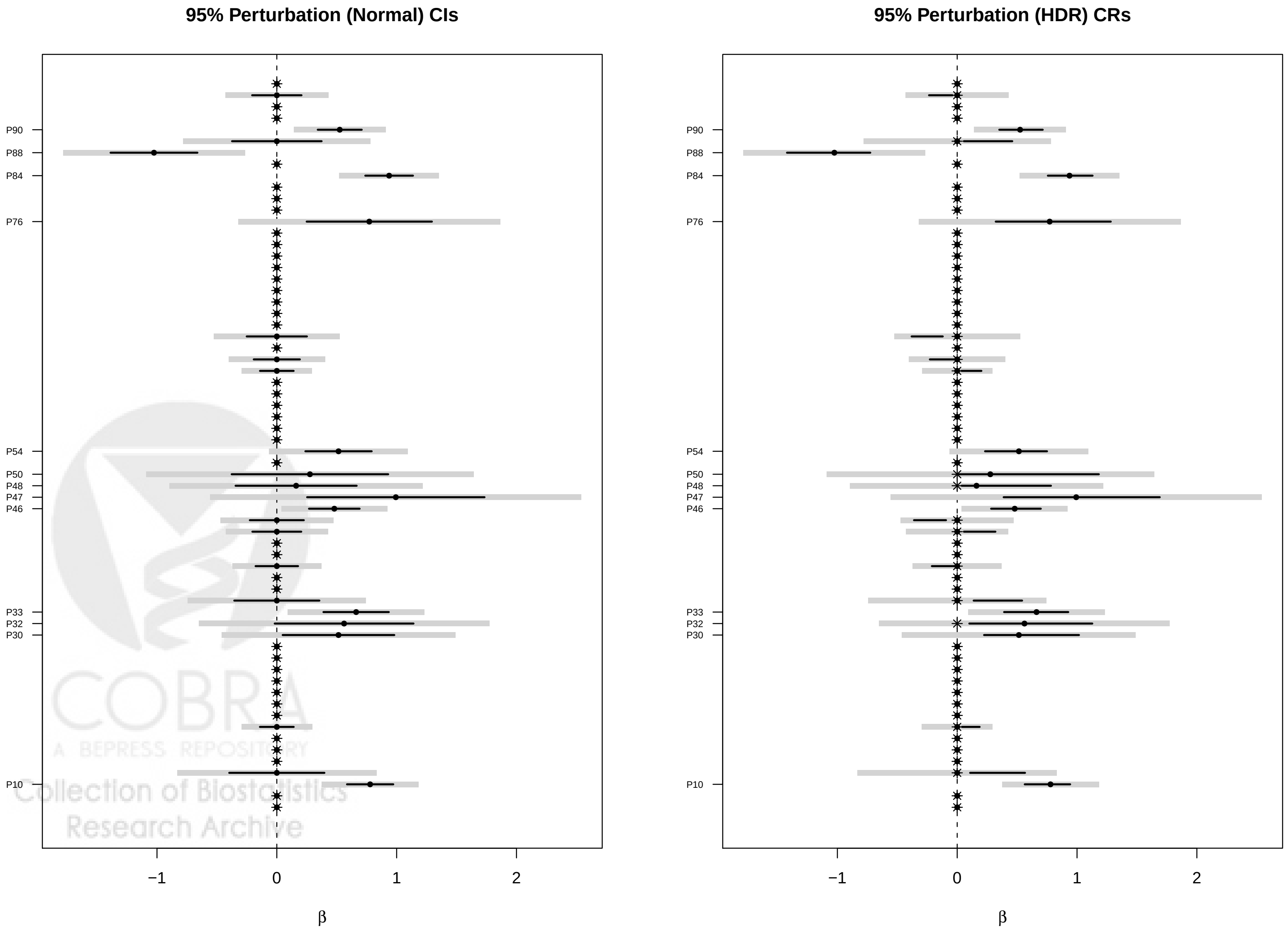


Codon	P10	P30	P32	P33	P46	P47	P48	P50	P54	P76	P84	P88	P90
Johnson et al	*		*		*	*		*	*	*	*		*
Wu Adjusted	*	*	*	*	*	*	*	*	*	*	*	*	*
OLS	*	*	*	*	*	*	*	*	*	*	*	*	*
Asymptotic Based	*	*	*	*	*	*			*	*	*	*	*
Perturbation (Normal)	*	*		*	*	*			*	*	*	*	*
Perturbation (HDR)	*	*		*	*	*			*	*	*	*	*
$\hat{p}_0$	0	0.01	0.08	0	0	0.03	0.29	0.31	0	0.01	0	0	0

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Figure 2: 95% perturbation CRs ($\mathrm{CR}^{*\mathrm{N}}$ and $\mathrm{CR}^{*\mathrm{HDR}}$) for the association between genetic mutations and antiretroviral drug susceptibility. Estimated coefficients $\widehat{\beta}_j$ are represented with a circle on each CR line and a star at zero signifies that the CR includes the point mass at zero. The shaded region denotes the simultaneous confidence region $\mathrm{CR}^{*\mathrm{Sim}}$. Note that even coefficients estimated as zero may have CRs around their estimates and that $\mathrm{CR}^{*\mathrm{HDR}}$ may be asymmetrical and noncontiguous.



COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive