High-dimensional Censored Regression via the Penalized Tobit Likelihood

Tate Jacobson
School of Statistics, University of Minnesota
and
Hui Zou
School of Statistics, University of Minnesota

September 23, 2022

Abstract

High-dimensional regression and regression with a left-censored response are each well-studied topics. In spite of this, few methods have been proposed which deal with both of these complications simultaneously. The Tobit model—long the standard method for censored regression in economics—has not been adapted for high-dimensional regression at all. To fill this gap and bring up-to-date techniques from high-dimensional statistics to the field of high-dimensional left-censored regression, we propose several penalized Tobit models. We develop a fast algorithm which combines quadratic minimization with coordinate descent to compute the penalized Tobit solution path. Theoretically, we analyze the Tobit lasso and Tobit with a folded concave penalty, bounding the ℓ_2 estimation loss for the former and proving that a local linear approximation estimator for the latter possesses the strong oracle property. Through an extensive simulation study, we find that our penalized Tobit models provide more accurate predictions and parameter estimates than other methods on high-dimensional left-censored data. We use a penalized Tobit model to analyze high-dimensional left-censored HIV viral load data from the AIDS Clinical Trials Group and identify potential drug resistance mutations in the HIV genome. A supplementary file contains intermediate theoretical results and technical proofs.

Keywords: censored regression, coordinate descent, folded concave penalty, high dimensions, strong oracle property, Tobit model

1 Introduction

In many regression problems, the dependent variable can only be observed within a restricted range. We say that such a response is *censored* if we retain some information from the observations which fall outside of this range rather than losing them entirely. In particular, we still observe the predictors for these cases and know whether the unobserved response value fell below or above the range. Censored data appear in many disciplines, either as a consequence of the data collection process or due to the nature of the response itself. For instance, biological assays used to measure human immunodeficiency virus (HIV) viral load in plasma cannot detect viral concentrations below certain (known) thresholds. As such, the observed viral load is left-censored. Because censoring violates a key assumption of linear regression, ordinary least squares (OLS) estimates of the regression coefficients will be biased and inconsistent if the response is censored (Amemiya, 1984). Recognizing this, researchers in different disciplines have developed regression techniques to deal with various types of censoring. Among these, the Tobit model has long been the standard method for modeling a left-censored response in economics.

Tobin (1958) originally developed the Tobit model to study how annual expenditures on durable goods relate to household income. Noting that most low-income households spend \$0/year on durable goods, he designed the Tobit likelihood to treat response values at this (known) lower limit differently than those above the limit. He described the Tobit model as a "hybrid of probit analysis and multiple regression," as it models the probability of the response falling at the lower limit using an approach similar to probit analysis while still treating the response as continuous (Tobin, 1958). Because left-censored data are common in household surveys and other micro-sample survey data, the Tobit model has enjoyed lasting popularity in economics and the social sciences. In the half-century since its introduction, it has been extended to handle right-censored and interval-censored data

(Amemiya, 1984) and has been adopted in other disciplines.

In recent years, high-dimensional data have become increasingly common in many fields of study. This presents some researchers with the challenge of analyzing data with both high-dimensional covariates and a left-censored response. Consider the HIV viral load example from earlier. There is now a sizable literature around modeling the relationship between HIV viral load and mutations in the HIV genome. Given the number of mutations that can occur, this is inherently a high-dimensional problem, where the number of predictors p is much larger than the number of observations n. At the same time, the observed viral load is left-censored. Previous studies in this area have avoided the problem of having both high-dimensional covariates and a left-censored response by reducing HIV viral load to a binary response, such as $y = \mathbb{1}_{\text{viral load}>200 \text{ copies/mL}}$. In taking this approach, however, the modelers lose a great deal of information about the response. To directly model HIV viral load in this setting, researchers need techniques designed specifically for high-dimensional left-censored regression.

While high-dimensional regression and left-censored regression have been thoroughly studied as separate topics, few methods have been developed which handle both high-dimensional covariates and a left-censored response simultaneously. Müller and van de Geer (2016) and Zhou and Liu (2016) have extended the least absolute deviation estimator of Powell (1984) for high-dimensional data while Johnson (2009), Li et al. (2014), and Soret et al. (2018) have extended the Buckley-James estimator (Buckley and James, 1979). To our knowledge, no existing methods directly extend the Tobit model. Theoretically, this under-studied area has fallen behind the broader field of high-dimensional statistics, with estimators achieving weaker guarantees and requiring stronger assumptions. Among existing high-dimensional left-censored regression techniques, only Müller and van de Geer's (2016) estimator has any theoretical guarantees in the setting where $p \gg n$. This estimator, however, does not achieve consistent model selection. On the other hand, the estimators of

Johnson (2009), Zhou and Liu (2016), and Li et al. (2014) are shown to possess the weak oracle property, but only in the fixed p case. We aim to improve on these high-dimensional left-censored regression techniques by developing an estimator which possesses the strong oracle property even when $p \gg n$.

In this study, we develop penalized Tobit models for high-dimensional censored regression. The negative log-likelihood in Tobin's (1958) original formulation of the Tobit model is non-convex, creating technical problems for optimization in a high-dimensional setting. We use Olsen's (1978) convex reparameterization of the negative log-likelihood in our penalized Tobit models so that we can solve our problem using convex optimization methods. In particular, we leverage the fact that the negative log-likelihood satisfies the quadratic majorization condition to develop a generalized coordinate descent (GCD) algorithm (Yang and Zou, 2013) for minimizing the penalized negative log-likelihood.

For our theoretical study, we analyze the Tobit lasso and Tobit with a folded concave penalty in a high-dimensional setting with $p \gg n$. We derive a bound for the ℓ_2 estimation loss for the Tobit lasso estimator which holds with high probability. We introduce a local linear approximation (LLA) algorithm for Tobit regression with a folded concave penalty and prove that, when initialized with the Tobit lasso estimator, this algorithm finds the oracle estimator in one step and converges to it in two steps with probability rapidly converging to 1 as n and p diverge. To our knowledge, this makes the two-step LLA estimator the first estimator for high-dimensional left-censored regression to possess the strong oracle property.

We have implemented the GCD algorithm and the LLA algorithm (specifically with the SCAD penalty (Fan and Li, 2001)) in the `tobitnet` package in `R`, which is available at `https://github.com/TateJacobson/tobitnet`.

This paper is organized as follows. In Section 2 we review the Tobit model and its statistical foundations. In Section 3 we introduce penalized Tobit models and develop our

GCD algorithm to fit them. In Section 4 we carry out our theoretical study of the Tobit lasso and our LLA algorithm for Tobit with a folded concave penalty. Section 5 presents the results of an extensive simulation study comparing our penalized Tobit models with penalized least-squares models and Soret et al.'s (2018) high-dimensional Buckley-James estimator (the best available alternative for high-dimensional left-censored regression) in terms of their prediction, estimation, and selection performance. In Section 6 we analyze real high-dimensional left-censored data from the AIDS Clinical Trials Group, modeling the relationship between HIV viral load and HIV genotypic mutations using the two-step Tobit LLA estimator and Soret et al.'s Buckley-James estimator in order to identify potential drug resistance mutations (DRMs). Intermediate theoretical results and technical proofs are provided in the supplementary material for this paper.

2 The Tobit Model

Suppose that we observe a set of predictors, $x_1, \ldots, x_p$, and a response $y \geq c$ where c is a known lower limit (for example, $c = 50$ if our response is HIV viral load and our assays cannot measure concentrations below 50 copies/mL). In Tobit regression we assume that there exists a latent response variable y^* such that $y = \max\{y^*, c\}$ and that y^* comes from a linear model $y^* = \mathbf{x}'\boldsymbol{\beta} + \epsilon$, where $\mathbf{x} = (1, x_1, \ldots, x_p)' \in \mathbb{R}^{p+1}$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)' \in \mathbb{R}^{p+1}$, and $\epsilon \sim N(0, \sigma^2)$. In the following developments we assume that $c = 0$ without loss of generality.

From this latent-variable formulation we can derive a likelihood for the censored response. Let $\{(y_i, \mathbf{x}_i')\}_{i=1}^n$ be i.i.d copies of $(y, \mathbf{x}')$ and define $d_i = \mathbb{1}_{y_i > 0}$. Let $\Phi(\cdot)$ denote the standard normal CDF. The Tobit likelihood is given by

$$L_n(\boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2 \right\} \right]^{d_i} \left[\Phi\left(\frac{-\mathbf{x}_i'\boldsymbol{\beta}}{\sigma} \right) \right]^{1-d_i}.$$

Noting that $P(y_i^* \leq 0) = P(\mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i \leq 0) = \Phi\left(\frac{-\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)$, we see that this likelihood is a mixture of a normal density and a point mass at 0. After dropping an ignorable constant the log-likelihood is given by

$$\log L_n(\boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^{n} d_i \left[-\log(\sigma) - \frac{1}{2\sigma^2}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2 \right] + (1 - d_i) \log\left(\Phi\left(\frac{-\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)\right).$$

3 Penalized Tobit Regression

In high-dimensional regression, the most commonly used approach is to exploit sparsity in the regression coefficient vector. While we might initially consider simply adding a penalty term to the Tobit log-likelihood to create an objective function for penalized Tobit regression, $\log L_n(\boldsymbol{\beta}, \sigma^2)$ is not concave in $(\boldsymbol{\beta}, \sigma^2)$, frustrating this approach. Thankfully, Olsen (1978) found that the reparameterization $\boldsymbol{\delta} = \boldsymbol{\beta}/\sigma$ and $\gamma^2 = \sigma^{-2}$ results in a concave log-likelihood:

$$\log L_n(\boldsymbol{\delta}, \gamma) = \sum_{i=1}^{n} d_i \left[\log(\gamma) - \frac{1}{2}(\gamma y_i - \mathbf{x}_i'\boldsymbol{\delta})^2 \right] + (1 - d_i) \log\left(\Phi\left(-\mathbf{x}_i'\boldsymbol{\delta}\right)\right).$$

Note that $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ must have the same degree of sparsity. We use Olsen's reparameterization to develop our penalized Tobit models. Our objective is to minimize

$$R_n(\boldsymbol{\delta}, \gamma) = \ell_n(\boldsymbol{\delta}, \gamma) + P_\lambda(\boldsymbol{\delta}) \tag{1}$$

with respect to $(\boldsymbol{\delta}, \gamma)$, where $\ell_n(\boldsymbol{\delta}, \gamma) = -\frac{1}{n} \log L_n(\boldsymbol{\delta}, \gamma)$ is our convex loss function (the *Tobit loss* for short) and $P_\lambda(\boldsymbol{\delta})$ is a penalty function. Note that, unlike with other loss functions, we cannot separate out the scale parameter γ in the Tobit loss when estimating the regression coefficients $\boldsymbol{\delta}$.

Coordinate descent (CD) is currently the most popular algorithm for high-dimensional regression in the literature (Friedman et al., 2010). Given the relatively complex form of the Tobit loss, the standard CD algorithm requires solving a nonlinear convex program

repeatedly for p variables for many cycles until convergence. As a result, the computation time will be notably longer than for penalized least squares. Fortunately, another benefit of using Olsen's reparameterization is that the Tobit likelihood can be shown to enjoy a nice *quadratic majorization condition* that serves as a foundation for using the majorization-minimization (MM) principle and coordinate descent to solve penalized Tobit regression. The combination of the MM principle and CD is called generalized coordinate descent (Yang and Zou, 2013). By using GCD, each of the coordinate-wise updates becomes a simple univariate quadratic minimization problem.

For reference, we say that a univariate function $f : \mathbb{R} \to \mathbb{R}$ satisfies the *quadratic majorization condition* if there exists $M \in \mathbb{R}^+$ such that $f(t+a) \leq f(t) + f'(t)a + \frac{M}{2}a^2$ for all $t, a \in \mathbb{R}$. Without loss of generality, we assume that our predictors are standardized—that is, $\frac{1}{n}\sum_{i=1}^n x_{ij} = 0$ and $\frac{1}{n}\sum_{i=1}^n x_{ij}^2 = 1$, for $j = 1, \ldots, p$. Consider coordinate-wise updates of δ_0 and δ_j, $j = 1, \ldots, p$. We treat δ_0 as a special case of δ_j in the following developments, keeping in mind that $x_{i0} = 1$ for all i. For ease of notation, let $\mathbf{x}_{i(-j)} = (x_{i0}, x_{i1}, \ldots, x_{i,j-1}, x_{i,j+1}, \ldots, x_{ip})' \in \mathbb{R}^p$ and $\boldsymbol{\delta}_{(-j)} = (\delta_0, \delta_1, \ldots, \delta_{j-1}, \delta_{j+1}, \delta_p)' \in \mathbb{R}^p$.

Let $\tilde{\boldsymbol{\delta}}$ and $\tilde{\gamma}$ denote the current values for $\boldsymbol{\delta}$ and γ. Let $j \in \{0, 1, \ldots, p\}$ and leave $\tilde{\boldsymbol{\delta}}_{(-j)}$ and $\tilde{\gamma}$ fixed. Then the Tobit loss is viewed as a univariate function of δ_j. After dropping ignorable constants (which have no impact in minimization), we can express the Tobit loss with respect to δ_j as

$$\ell_n(\delta_j|\tilde{\boldsymbol{\delta}}, \tilde{\gamma}) = \frac{1}{n}\sum_{i=1}^n d_i \frac{1}{2}(\tilde{\gamma} y_i - \mathbf{x}'_{i,(-j)}\tilde{\boldsymbol{\delta}}_{(-j)} - x_{ij}\delta_j)^2 - (1 - d_i)\log\Phi(-\mathbf{x}'_{i,(-j)}\tilde{\boldsymbol{\delta}}_{(-j)} - x_{ij}\delta_j).$$

Theorem 1. *$\ell_n(\delta_j|\tilde{\boldsymbol{\delta}}, \tilde{\gamma})$ satisfies the quadratic majorization condition with $M = \frac{1}{n}\sum_{i=1}^n x_{ij}^2$. Under the standardization of predictors, $M = 1$.*

To illustrate the whole process of GCD, we focus on the weighted lasso penalty: $P_\lambda(\boldsymbol{\delta}) = \sum_{j=1}^p \lambda w_j |\delta_j|$. When $w_j = 1$ for all j, this penalty reduces to the lasso penalty. The weighted lasso penalty form will also be used in the computation of the folded-concave-

penalized Tobit estimator (see the next section for details).

The standard coordinate descent algorithm needs to minimize $\ell_n(\delta_j|\tilde{\boldsymbol{\delta}},\tilde{\gamma}) + \lambda w_j|\delta_j|$, which requires another iterative procedure to find the minimizer. By Theorem 1, we have that

$$Q(\delta_j|\tilde{\boldsymbol{\delta}},\tilde{\gamma}) := \ell_n(\tilde{\delta}_j|\tilde{\boldsymbol{\delta}},\tilde{\gamma}) + \ell_n'(\tilde{\delta}_j|\tilde{\boldsymbol{\delta}},\tilde{\gamma})\cdot(\delta_j-\tilde{\delta}_j) + \frac{1}{2}(\delta_j-\tilde{\delta}_j)^2$$

is a quadratic majorization function for $\ell_n(\delta_j|\tilde{\boldsymbol{\delta}},\tilde{\gamma})$. By the MM principle, we can simply minimize $Q(\delta_j|\tilde{\boldsymbol{\delta}},\tilde{\gamma})+\lambda w_j|\delta_j|$ to update δ_j, while leaving the other parameters fixed at their current values. For $j = 1,\ldots,p$, we update δ_j as the minimizer of $Q(\delta_j|\tilde{\boldsymbol{\delta}},\tilde{\gamma}) + \lambda w_j|\delta_j|$, which is given by a soft-thresholding rule (Tibshirani, 1996): $\hat{\delta}_j = S\left(\tilde{\delta}_j - \ell_n'(\tilde{\delta}_j|\tilde{\boldsymbol{\delta}},\tilde{\gamma}), w_j\lambda\right)$, where $S(z,t) = (|z|-t)_+\mathrm{sgn}(z)$. For updating δ_0, we use the minimizer of $Q(\delta_0|\tilde{\boldsymbol{\delta}},\tilde{\gamma})$, which is given by $\hat{\delta}_0 = \tilde{\delta}_0 - \ell_n'(\tilde{\delta}_0|\tilde{\boldsymbol{\delta}},\tilde{\gamma})$. Lastly, we need to update γ. We can show that given $\tilde{\boldsymbol{\delta}}$, $\ell_n(\gamma|\tilde{\boldsymbol{\delta}})$ is minimized by $\hat{\gamma} = \frac{\sum_{i=1}^n d_i y_i(\mathbf{x}_i'\tilde{\boldsymbol{\delta}}) + \sqrt{\left(\sum_{i=1}^n d_i y_i(\mathbf{x}_i'\tilde{\boldsymbol{\delta}})\right)^2 + 4\left(\sum_{i=1}^n d_i y_i^2\right)\sum_{i=1}^n d_i}}{2\sum_{i=1}^n d_i y_i^2}$. For completeness, we show the whole GCD algorithm in Algorithm 1.

Algorithm 1: GCD algorithm for penalized Tobit with the weighted lasso penalty

Initialize $(\tilde{\boldsymbol{\delta}},\tilde{\gamma})$;

repeat

  Compute $\hat{\delta}_0 = \tilde{\delta}_0 - \ell_n'(\tilde{\delta}_0|\tilde{\boldsymbol{\delta}},\tilde{\gamma})$; Set $\tilde{\delta}_0 = \hat{\delta}_0$;

  for $j = 1,\ldots,p$ **do**

    Compute $\hat{\delta}_j = S\left(\tilde{\delta}_j - \ell_n'(\tilde{\delta}_j|\tilde{\boldsymbol{\delta}},\tilde{\gamma}), w_j\lambda\right)$; Set $\tilde{\delta}_j = \hat{\delta}_j$;

  end

  Compute $\hat{\gamma} = \frac{\sum_{i=1}^n d_i y_i(\mathbf{x}_i'\tilde{\boldsymbol{\delta}}) + \sqrt{\left(\sum_{i=1}^n d_i y_i(\mathbf{x}_i'\tilde{\boldsymbol{\delta}})\right)^2 + 4\left(\sum_{i=1}^n d_i y_i^2\right)\sum_{i=1}^n d_i}}{2\sum_{i=1}^n d_i y_i^2}$; Set $\tilde{\gamma} = \hat{\gamma}$;

until *convergence*;

4 Theoretical Results

Our objective function (1) is flexible enough to accommodate a wide variety of penalties. In this section we offer theoretical studies of penalized Tobit estimators. We focus on the

Tobit estimator with the lasso penalty and Tobit estimators with folded concave penalties.

4.1 Setup

Suppose that $\boldsymbol{\delta}^*$ and γ^* are the true parameter values for $\boldsymbol{\delta}$ and γ. For notational convenience, we define $\Theta = (\theta_0, \ldots, \theta_{p+1})' := (\boldsymbol{\delta}', \gamma)'$. We let $\mathcal{A} = \{j : \delta_j^* \neq 0\} \subseteq \{1, \ldots, p\}$ denote the true support set and define $\mathcal{A}' = \mathcal{A} \cup \{0, p+1\}$ and $s = |\mathcal{A}|$. Under the sparsity assumption, $s \ll p$. Note that we continue to index $\boldsymbol{\delta}$, Θ, and the columns of $\mathbf{X}$ from 0 to $p+1$ to accommodate δ_0 in $\boldsymbol{\delta}$.

We adopt the following notation throughout our analysis. For a matrix $\mathbf{A} \in [a_{ij}]_{n \times m}$ and sets of indices $\mathcal{S} \subseteq \{1, \ldots, m\}$ and $\mathcal{T} \subseteq \{1, \ldots, n\}$, we use $\mathbf{A}_{(\mathcal{S})}$ to denote the submatrix consisting of the *columns* of $\mathbf{A}$ with indices in $\mathcal{S}$ and $\mathbf{A}_{\mathcal{T}}$ to denote the submatrix consisting of the *rows* of $\mathbf{A}$ with indices in $\mathcal{T}$. We let $\lambda_{\max}(\mathbf{A})$ denote the largest eigenvalue of $\mathbf{A}$, let $\mathbf{A} \succ 0$ signify that $\mathbf{A}$ is positive definite, and let $\text{vec}(\mathbf{A}) \in \mathbb{R}^{nm}$ denote the vectorization of $\mathbf{A}$. We define several matrix norms: the ℓ_∞-norm $\|\mathbf{A}\|_\infty = \max_i \sum_j |a_{ij}|$, the ℓ_1-norm $\|\mathbf{A}\|_1 = \max_j \sum_i |a_{ij}|$, the ℓ_2-norm $\|\mathbf{A}\|_2 = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$, the entry-wise maximum $\|\mathbf{A}\|_{\max} = \max_{(i,j)} |a_{i,j}|$, and the entry-wise minimum $\|\mathbf{A}\|_{\min} = \min_{(i,j)} |a_{i,j}|$. We let $\nabla_{\mathcal{S}} \log L_n(\Theta)$ and $\nabla^2_{\mathcal{S}} \log L_n(\Theta)$ denote the gradient and Hessian, respectively, of $\log L_n(\Theta)$ with respect to $\Theta_{\mathcal{S}}$.

Because we handle censored and uncensored observations differently in the Tobit likelihood, we introduce notation to clearly differentiate between them. Let n_1 denote the number of observations for which $y_i > 0$ and $n_0 = n - n_1$. Let $\mathbf{X}_1$ be the $n_1 \times (p+1)$ matrix of predictors corresponding to the observations for which $y_i > 0$ and let $\mathbf{X}_0$ be the $n_0 \times (p+1)$ matrix of predictors corresponding to the observations for which $y_i \leq 0$. Define

$\mathbf{y}_0$ and $\mathbf{y}_1$ likewise. We then reorder our observations so that

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \end{bmatrix}.$$

4.2 The lasso-penalized Tobit estimator

Consider the lasso-penalized Tobit estimator found by minimizing

$$R_n(\boldsymbol{\delta}, \gamma) = \ell_n(\boldsymbol{\delta}, \gamma) + \lambda_{\text{lasso}} \sum_{j=1}^{p} |\delta_j|. \tag{2}$$

We assume that the following restricted eigenvalue condition holds:

$$\kappa = \min_{\mathbf{u} \in \mathcal{C}} \frac{\mathrm{E}\left[\left\| \begin{bmatrix} -\mathbf{X}_1 & \mathbf{y}_1 \end{bmatrix} \mathbf{u} \right\|_2^2 \right]}{n \left\| \mathbf{u} \right\|_2^2} \in (0, \infty) \tag{A0}$$

where $\mathcal{C} = \{\mathbf{u} \neq \mathbf{0} : \left\| \mathbf{u}_{\mathcal{A}'^c} \right\|_1 \leq 3 \left\| \mathbf{u}_{\mathcal{A}'} \right\|_1\}$. It is worth pointing out that condition (A0) is similar in spirit to the restricted eigenvalue condition used in lasso-penalized least squares (Bickel et al., 2009) but also has an important technical difference because the response variable appears together with the predictors—even in the fixed design setting, the entire matrix $\begin{bmatrix} -\mathbf{X}_1 & \mathbf{y}_1 \end{bmatrix}$ is random. Thus we must take the expectation in condition (A0) so that κ will be deterministic.

In the following result, we bound the ℓ_2 estimation loss of the Tobit lasso estimator with high probability. Let $g(s) = \phi(s)/\Phi(s)$, where $\phi(\cdot)$ denotes the standard normal density function. We assume the following:

(A1) $\max_j \left\| \mathbf{x}_{(j)} \right\|_2 = O(\sqrt{n})$, $\sum_{i=1}^n (\mathbf{x}_i' \boldsymbol{\delta}^*)^2 = O(n)$, $\max_{j,k} \sum_{i=1}^n x_{ij}^2 x_{ik}^2 = O(n)$, $\sum_{i=1}^n x_{ij}^2 (2 + \mathbf{x}_i' \boldsymbol{\delta}^* + g(-\mathbf{x}_i' \boldsymbol{\delta}^*))^2 = O(n)$, and $\sum_{i=1}^n \frac{1}{2} (\mathbf{x}_i' \boldsymbol{\delta}^*)^2 (2 + \mathbf{x}_i' \boldsymbol{\delta}^* + g(-\mathbf{x}_i' \boldsymbol{\delta}^*))^2 = O(n)$ where $j, k \in \{0, \ldots, p\}$;

(A2) $s = O\left(n^{\alpha_1}\right)$, $\log(p) = O(n^{\alpha_2})$, where $\alpha_1, \alpha_2 \in \left(0, \frac{1}{3}\right)$;

and define $M_1 = \max_j n^{-1} \left\| \mathbf{x}_{(j)} \right\|_2^2$ and $M_2 = 16 + 4n^{-1} \sum_{i=1}^n (\mathbf{x}_i' \boldsymbol{\delta}^*)^2$.

Theorem 2. *Suppose that* $Y_i^* = \mathbf{x}_i'\beta^* + \epsilon_i$ *where* $\epsilon_i \overset{iid}{\sim} N(0, \sigma^{*2})$ *and define* $Y_i = Y_i^* \mathbb{1}_{Y_i^* > 0}$ *for* $i = 1, \ldots, n$. *Let* $\hat{\Theta}^{\text{lasso}}$ *denote the solution to the lasso-penalized Tobit model* (2) *with penalty parameter* $\lambda_{\text{lasso}} = A\sqrt{\frac{\log p}{n}}$ *where* $A > \max\left\{4\sqrt{M_1}, \frac{\sqrt{8M_2}}{\gamma^*}\right\}$. *If* (A0) - (A2) *hold, then for large* n, p

$$\left\|\hat{\Theta}^{\text{lasso}} - \Theta^*\right\|_2 \leq \frac{3\sqrt{s+2}\lambda_{\text{lasso}}}{\kappa}$$

with probability at least $1 - b_1 p^{-1} - b_2 e^{-b_3 n^{1-2\alpha_1}}$ *where* b_1, b_2, b_3 *are constants.*

Remark. *Under condition* (A2), $\sqrt{\frac{(s+2)\log p}{n}} \to 0$ *and, by extension, the* ℓ_2 *estimation loss for* $\hat{\Theta}_{\text{lasso}}$ *converges to* 0 *as* $n, p \to \infty$. *As such,* $\hat{\Theta}_{\text{lasso}}$ *is consistent under the* ℓ_2 *norm.*

Theorem 2 follows immediately from the more general finite-sample probability bound given in Theorem S.1 (in supplementary material). Note that we cannot compute $\lambda_{\text{lasso}} = A\sqrt{\frac{\log p}{n}}$ in practice as $\boldsymbol{\delta}^*$ and γ^* are unknown.

4.3 The folded-concave-penalized Tobit estimator

It is now well-understood that a lasso-penalized estimator often does not achieve consistent model selection unless a stringent "irrepresentable condition" (Zhao and Yu, 2006; Zou, 2006) is assumed. To relax this condition, we can try to use a folded-concave-penalized Tobit estimator. We aim to minimize $R_n(\boldsymbol{\delta}, \gamma) = \ell_n(\boldsymbol{\delta}, \gamma) + P_\lambda(\boldsymbol{\delta})$ where $P_\lambda(\boldsymbol{\delta}) = \sum_{j=1}^p P_\lambda(|\delta_j|)$ is a *folded concave penalty*, meaning that $P_\lambda(|t|)$ satisfies

(i) $P_\lambda(t)$ is increasing and concave in $t \in [0, \infty)$ with $P_\lambda(0) = 0$.

(ii) $P_\lambda(t)$ is differentiable in $t \in (0, \infty)$ with $P'_\lambda(0) := P'_\lambda(0+) \geq a_1\lambda$.

(iii) $P'_\lambda(t) \geq a_1\lambda$ for $t \in (0, a_2\lambda]$

(iv) $P'_\lambda(t) = 0$ for $t \in [a\lambda, \infty)$ where $a > a_2$,

where a is pre-specified and a_1 and a_2 are fixed positive constants which depend on the folded concave penalty we choose. Two well-known folded concave penalties are the SCAD

penalty (Fan and Li, 2001), the derivative of which is given by $P'_\lambda(t) = \lambda \mathbb{1}_{t \leq \lambda} + \frac{(a\lambda - t)_+}{a-1} \mathbb{1}_{t > \lambda}$, where $a > 2$, and the MCP (Zhang, 2010), the derivative of which is given by $P'_\lambda(t) = \left(\lambda - \frac{t}{a}\right)_+$, where $a > 1$. One can show that $a_1 = a_2 = 1$ for the SCAD penalty and $a_1 = 1 - a^{-1}$, $a_2 = 1$ for the MCP.

The strongest rationale for using a folded concave penalty is that it can produce an estimator with the *strong oracle property*—that is, an estimator which is equal to the *oracle estimator* with very high probability (Fan and Lv, 2011; Fan et al., 2014). The oracle estimator knows the true support set $\mathcal{A}$ beforehand and, as a result, delivers optimal estimation efficiency. For the Tobit model, the oracle estimator is given by

$$\hat{\Theta}^{\text{oracle}} = \underset{\Theta : \Theta_{\mathcal{A}'^c} = \mathbf{0}}{\arg\min}\, \ell_n(\Theta) \tag{3}$$

Note that we cannot solve (3) in practice because $\mathcal{A}$ is unknown. Instead, the oracle estimator is a theoretical benchmark to compare our estimators against. Because our loss function is convex, the oracle estimator is unique, meaning that

$$\nabla_j \ell_n(\hat{\Theta}^{\text{oracle}}) = 0 \qquad \forall j \in \mathcal{A}' \tag{4}$$

where ∇_j denotes the derivative with respect to the jth element of Θ.

With a folded concave penalty function, the overall objective $R_n(\boldsymbol{\delta}, \gamma)$ may no longer be convex and, consequently, could have multiple local solutions. As such, theory should be developed for a specific, explicitly-defined local solution. We examine the local solution to the folded concave penalization problem which we obtain using the LLA algorithm (Zou and Li, 2008). This choice is inspired by the general theory developed in Fan et al. (2014) where the authors established the strong oracle property of the LLA solution for a wide class of problems. We expect the same result holds for the Tobit model.

The LLA algorithm turns the Tobit model with a folded concave penalty into a sequence of weighted Tobit lasso models. We can use Algorithm 1 to fit each weighted Tobit lasso

model. Algorithm 2 shows the complete details of the LLA algorithm for the Tobit model with a folded concave penalty.

Algorithm 2: The local linear approximation (LLA) algorithm

Initialize $\hat{\Theta}^{(0)} = \hat{\Theta}^{\text{initial}}$ and compute the adaptive weights

$$\hat{\mathbf{w}}^{(0)} = (\hat{w}_1^{(0)}, \ldots, \hat{w}_p^{(0)})' = (P'_\lambda(|\hat{\delta}_1^{(0)}|), \ldots, P'_\lambda(|\hat{\delta}_p^{(0)}|))'.$$

for $m = 1, 2, \ldots$ **do**

Solve the following optimization problem

$\hat{\Theta}^{(m)} = \arg\min_\Theta \ell_n(\Theta) + \sum_{j=1}^p \hat{w}_j^{(m-1)} \cdot |\delta_j|$;

Update the adaptive weight vector with $\hat{w}_j^{(m)} = P'_\lambda(|\hat{\delta}_j^{(m)}|)$ for $j = 1, \ldots, p$.

end

We aim to show that the LLA algorithm finds the oracle estimator in one step and converges to it in two steps with probability rapidly converging to 1 as $n, p \to \infty$. We define $Q_1 = \max_{j \in \mathcal{A} \cup \{0\}} \lambda_{\max}\left(\frac{1}{n}\mathbf{X}'_{(\mathcal{A}\cup\{0\})} \operatorname{diag}\{|\mathbf{X}_{(j)}|\}\mathbf{X}_{(\mathcal{A}\cup\{0\})}\right)$; $Q_2 = \left\|\left(\mathrm{E}\left[\frac{1}{n}\nabla^2_{\mathcal{A}'} \log L_n(\Theta^*)\right]\right)^{-1}\right\|_\infty$; $Q_3 = Q_2 \cdot \left\|\mathrm{E}\left[\frac{1}{n}[\nabla^2 \log L_n(\Theta^*)]_{\mathcal{A}'^c,\mathcal{A}'}\right]\right\|_\infty$; $H^*_{\mathcal{A}',\mathcal{A}'} = \mathrm{E}\left[\frac{1}{n}\nabla^2_{\mathcal{A}'} \log L_n(\Theta^*)\right] \otimes \mathrm{E}\left[\frac{1}{n}\nabla^2_{\mathcal{A}'} \log L_n(\Theta^*)\right]$, where $\otimes$ denotes the Kronecker product; $K_1 = \left\|\mathrm{E}\left[\frac{1}{n}\nabla^2_{\mathcal{A}'} \log L_n(\Theta^*)\right]\right\|_\infty$; and $K_2 = \left\|(H^*_{\mathcal{A}',\mathcal{A}'})^{-1}\right\|_\infty$. We assume the following:

(A3) $||\boldsymbol{\delta}^*_{\mathcal{A}}||_{\min} > (a+1)\lambda$

(A4) $\mathrm{E}\left[\nabla^2_{\mathcal{A}'} \log L_n(\Theta^*)\right] \succ 0$

(A5) $Q_1 = O(1)$, $Q_2 = O(1)$, $Q_3 = O(1)$, $K_1 = O(1)$, and $K_2 = O(1)$;

(A6) $\exists_{C_1, C_2 > 0}$ such that $Q_2 > C_1$ and $\left\|\mathrm{E}\left[\frac{1}{n}[\nabla^2 \log L_n(\Theta^*)]_{\mathcal{A}'^c,\mathcal{A}'}\right]\right\|_\infty > C_2$ for all n.

Theorem 3. *Suppose that* $Y_i^* = \mathbf{x}_i'\beta^* + \epsilon_i$ *where* $\epsilon_i \overset{iid}{\sim} N(0, \sigma^{*2})$ *and define* $Y_i = Y_i^* \mathbb{1}_{Y_i^* > 0}$ *for* $i = 1, \ldots, n$. *Let* $\lambda = B\sqrt{\frac{\log p}{n}}$ *with* $B > \max\left\{\frac{(9Q_3+2)\sqrt{4M_1}}{a_1}, \frac{(9Q_3+2)\sqrt{2M_2}}{a_1\gamma^*}, 4Q_2\sqrt{M_1}, \frac{Q_2\sqrt{8M_2}}{\gamma^*}\right\}$. *Let* $a_0 = \min\{1, a_2\}$. *Suppose that our initial estimator* $\hat{\Theta}^{\text{initial}}$ *satisfies*

$$||\hat{\boldsymbol{\delta}}^{\text{initial}}_{(-0)} - \boldsymbol{\delta}^*_{(-0)}||_{\max} \leq a_0\lambda \tag{5}$$

If (A1) - (A6) *hold, then for large* n, p *the LLA algorithm initialized by* $\hat{\Theta}^{\text{initial}}$ *finds* $\hat{\Theta}^{\text{oracle}}$ *in one iteration with probability at least* $1 - c_1 p^{-1} - c_2 e^{-c_3 n^{1-2\alpha_1}}$ *and converges to* $\hat{\Theta}^{\text{oracle}}$ *after two iterations with probability at least* $1 - d_1 p^{-1} - d_2 e^{-d_3 n^{1-2\alpha_1}}$ *where* $c_1, c_2, c_3, d_1, d_2, d_3$ *are constants.*

Remark. *Under condition* (A2) *both of the probability bounds in Theorem 3 rapidly converge to* 1 *as* $n, p \to \infty$.

Theorem 3 follows immediately from Theorem S.4 (in supplementary material), which provides more general finite-sample bounds. Note that we cannot obtain $\lambda = B\sqrt{\frac{\log p}{n}}$ in practice as $\boldsymbol{\delta}^*$ and γ^* are unknown.

All that remains is to pick an initial estimator which satisfies (5) with high probability. We choose the Tobit lasso as our initial estimator since we already have an estimation loss bound from Theorem 2. The following corollary combines Theorems 2 and 3 to bound the probability that the LLA algorithm initialized by $\hat{\Theta}^{\text{lasso}}$ converges to the oracle estimator in two steps.

Corollary 1. *Suppose that* $Y_i^* = \mathbf{x}_i'\beta^* + \epsilon_i$ *where* $\epsilon_i \overset{iid}{\sim} N(0, \sigma^{*2})$ *and define* $Y_i = Y_i^* \mathbb{1}_{Y_i^* > 0}$ *for* $i = 1, \ldots, n$. *Define* A *and* B *as in Theorems 2 and 3. If conditions* (A0) - (A6) *hold,* $\lambda_{\text{lasso}} = A\sqrt{\frac{\log p}{n}}$ *, and* $\lambda = \max\left\{B\sqrt{\frac{\log p}{n}}, \frac{3\sqrt{s+2}\lambda_{\text{lasso}}}{a_0 \kappa}\right\}$, *then for large* n, p *the LLA algorithm initialized by* $\hat{\Theta}^{\text{lasso}}$ *converges to* $\hat{\Theta}^{\text{oracle}}$ *after two iterations with probability at least* $1 - k_1 p^{-1} - k_2 e^{-k_3 n^{1-2\alpha_1}}$ *where* k_1, k_2, k_3 *are constants.*

Remark. *Under condition* (A2) *the probability bound in Corollary 1 rapidly converges to* 1 *as* $n, p \to \infty$. *As such, Corollary 1 establishes that the two-step LLA estimator initialized by the Tobit lasso possesses the strong oracle property in a high-dimensional setting where* $p \gg n$.

5 Simulation Study

In the following simulation study, we compare the Tobit lasso, the two-step Tobit LLA estimator with a SCAD penalty initialized by the Tobit lasso (Tobit LLA), the least-squares lasso, least-squares SCAD, and Soret et al.'s(2018) high-dimensional Buckley-James estimator (SAWCT2018) to determine whether the penalized Tobit models provide an appreciable improvement in prediction, estimation, and selection performance on high-dimensional data with a left-censored response. We compare our methods to SAWCT2018 as it is the best available alternative for high-dimensional left-censored regression. For reference, we set $a = 3.7$ for Tobit LLA's SCAD penalty and the least-squares SCAD penalty throughout these simulations.

For each simulation setting, we generate 100 datasets with 100 training observations and 5000 test observations. We generate an uncensored response from a linear model $y_i^* = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i$, where $\mathbf{x}_i \sim N(0, \Sigma)$ and $\epsilon_i \sim N(0, \sigma^2)$, and left-censor it to create y_i as follows. Let q denote the proportion of the y_i that are left-censored in a simulated dataset. We control q by setting c_q to be the q-quantile of the y_i^* from both the training and test data and censoring the response at c_q—that is, we set $y_i = \max\{y_i^*, c_q\}$.

We have four elements we can vary across our simulation settings: Σ, q, p, and the response generating parameters (β_0, $\boldsymbol{\beta}$, and σ). We run simulations with each of the following covariance structures for the predictors: independent, CS(0.5), CS(0.8), AR1(0.5), and AR1(0.8) (CS(ρ) means that $\Sigma_{ij} = \rho$ for $i \neq j$, $\Sigma_{ii} = 1$ for all i and AR1(ρ) means that $(\Sigma_\rho)_{ij} = \rho^{|i-j|}$ for all i, j). For each covariance structure, we generate datasets with every combination of $q \in \{\frac{1}{8}, \frac{1}{4}, \frac{1}{2}\}$ and $p \in \{50, 500\}$. For all of these simulations, we set $\beta_0 = 3$, $\beta = (5, 1, 0.5, -2, 0.1, 0, \ldots, 0)$, and $\sigma = 1$. All together we examine 30 cases. We group our results into Tables 1, 2, 3, 4, and 5 based on the covariance structure of the predictors and vary p and q within these tables.

To assess the prediction performance of the models, we tune each of the models on the training data using 5-fold CV then compute the MSE of their predictions on the test data. In our simulation results, we report the average test MSE over 100 replications and give its standard error in parentheses. We also include prediction results for the ordinary least squares oracle model (OLS Oracle) and an ordinary least squares model with all of the predictors (OLS) for cases where $p \leq n$.

We use a variety of metrics to compare the parameter estimation and selection performance of the penalized models. To compare the accuracy of the parameter estimates, we report the ℓ_1 loss $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1$ and ℓ_2 loss $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2$. To assess the selection performance of these models, we report the number of false positive (FP) and false negative (FN) variable selections. In our simulation results, we report the average for each metric over 100 replications and give its standard error in parentheses.

5.1 Prediction results

We see a remarkably consistent pattern in our prediction results: in all 30 simulation settings the penalized Tobit models attain the two lowest average test MSEs, with Tobit LLA delivering the best prediction performance in 29 of 30 settings. We see a clear gap in prediction performance separating the three methods which account for censoring (the Tobit lasso, Tobit LLA, and SAWCT2018) from the least squares methods in that the average test MSEs for the OLS methods are, at minimum, nearly double those of the Tobit models and SAWCT2018. As the proportion of censored observations q increases, the test MSEs for the least-squares models climb upwards while the test MSEs for the Tobit models and SAWCT2018 largely remain stable. In particular, we see that the average test MSEs for the least-squares lasso and SCAD are around five times the average test MSE of Tobit LLA when $q = \frac{1}{2}$. The message here is clear: failing to account for censoring in the data can come at a steep price in terms of prediction accuracy, especially when the proportion

Table 1: Simulation Results with Independent Covariates

q	p	Method	MSE	ℓ_2	ℓ_1	FP	FN
$\frac{1}{8}$	50	Lasso	2.37(0.03)	1.5(0.06)	3.14(0.09)	4.6(0.4)	0.8(0.1)
		SCAD	2.22(0.02)	0.98(0.05)	2.19(0.06)	2(0.2)	1(0.1)
		Tobit Lasso	1.08(0.01)	0.24(0.01)	1.61(0.05)	7(0.3)	**0.6(0)**
		Tobit LLA	**1.01(0.01)**	**0.15(0.01)**	**0.81(0.03)**	**1.1(0.1)**	0.9(0)
		SAWCT2018	1.09(0.01)	0.28(0.01)	1.47(0.04)	4.4(0.2)	0.7(0)
		OLS Oracle	2.1(0.01)	-	-	-	-
		OLS	3.95(0.07)	-	-	-	-
	500	Lasso	2.45(0.04)	1.87(0.07)	3.85(0.12)	9.6(0.8)	1.2(0.1)
		SCAD	2.1(0.02)	0.91(0.05)	2.35(0.07)	4.5(0.4)	1.2(0.1)
		Tobit Lasso	1.23(0.01)	0.49(0.02)	2.61(0.05)	14.5(0.5)	**0.9(0)**
		Tobit LLA	**1.04(0.01)**	**0.22(0.01)**	**1.03(0.03)**	**2.3(0.2)**	1(0)
		SAWCT2018	1.28(0.02)	0.6(0.02)	2.51(0.07)	10.6(0.6)	1(0)
		OLS Oracle	1.93(0.01)	-	-	-	-
$\frac{1}{4}$	50	Lasso	3.24(0.04)	4.45(0.13)	4.91(0.09)	4(0.3)	1.2(0.1)
		SCAD	3.05(0.03)	3.42(0.11)	4.07(0.08)	2.1(0.2)	1.4(0.1)
		Tobit Lasso	0.9(0.01)	0.3(0.01)	1.69(0.05)	5.9(0.3)	**0.6(0)**
		Tobit LLA	**0.84(0.01)**	**0.18(0.01)**	**0.88(0.03)**	**1(0.1)**	0.9(0)
		SAWCT2018	0.93(0.01)	0.42(0.02)	1.77(0.04)	4.3(0.2)	0.7(0.1)
		OLS Oracle	2.79(0.01)	-	-	-	-
		OLS	5.49(0.1)	-	-	-	-
	500	Lasso	3.58(0.05)	5.01(0.13)	5.8(0.15)	9.5(0.9)	1.6(0.1)
		SCAD	3.18(0.03)	3.23(0.11)	4.36(0.08)	6(0.5)	1.5(0.1)
		Tobit Lasso	1.12(0.02)	0.61(0.03)	2.78(0.06)	13.4(0.4)	**1(0)**
		Tobit LLA	**0.94(0.01)**	**0.28(0.02)**	**1.13(0.03)**	**2(0.2)**	1.1(0)
		SAWCT2018	1.21(0.02)	0.91(0.04)	2.95(0.08)	10.5(0.5)	**1(0)**
		OLS Oracle	2.88(0.02)	-	-	-	-
$\frac{1}{2}$	50	Lasso	3.84(0.04)	15.6(0.21)	8.26(0.07)	3.3(0.3)	1.7(0.1)
		SCAD	3.66(0.03)	13.73(0.21)	7.57(0.07)	2.2(0.2)	1.9(0.1)
		Tobit Lasso	0.69(0.01)	0.55(0.03)	2.24(0.06)	6.1(0.3)	**0.6(0.1)**
		Tobit LLA	**0.6(0.01)**	**0.31(0.02)**	**1.14(0.04)**	**0.8(0.1)**	1.1(0)
		SAWCT2018	0.74(0.01)	1.43(0.06)	2.87(0.07)	3.6(0.2)	0.8(0.1)
		OLS Oracle	3.45(0.02)	-	-	-	-
		OLS	6.49(0.1)	-	-	-	-
	500	Lasso	3.99(0.04)	18.11(0.18)	9.66(0.16)	8.4(0.9)	2.2(0.1)
		SCAD	3.72(0.03)	15.21(0.19)	8.48(0.06)	5.5(0.5)	2.2(0.1)
		Tobit Lasso	0.93(0.02)	1.42(0.07)	3.57(0.08)	10.2(0.4)	**1.2(0)**
		Tobit LLA	**0.69(0.01)**	**0.49(0.03)**	**1.49(0.04)**	**1.7(0.2)**	1.4(0)
		SAWCT2018	1.11(0.03)	3.57(0.14)	5.09(0.1)	10.1(0.4)	1.3(0)
		OLS Oracle	3.28(0.01)	-	-	-	-

Table 2: Simulation Results with CS(0.5) Covariates

q	p	Method	MSE	ℓ_2	ℓ_1	FP	FN
$\frac{1}{8}$	50	Lasso	2.02(0.02)	1.6(0.07)	3.33(0.08)	6.5(0.3)	0.9(0.1)
		SCAD	1.93(0.02)	0.92(0.05)	2.05(0.07)	**1.6(0.2)**	1.2(0.1)
		Tobit Lasso	1.07(0.01)	0.41(0.02)	1.94(0.05)	7.3(0.3)	**0.6(0)**
		Tobit LLA	**1(0.01)**	**0.25(0.01)**	**1.04(0.03)**	1.8(0.2)	0.9(0)
		SAWCT2018	1.08(0.01)	0.45(0.02)	1.72(0.04)	5.4(0.2)	0.7(0.1)
		OLS Oracle	1.8(0.01)	-	-	-	-
		OLS	3.48(0.08)	-	-	-	-
	500	Lasso	2.37(0.03)	2.61(0.09)	4.97(0.15)	15.8(0.8)	1.5(0.1)
		SCAD	2.03(0.02)	1.03(0.05)	2.34(0.06)	**4.3(0.3)**	1.6(0.1)
		Tobit Lasso	1.19(0.01)	0.84(0.03)	3.11(0.06)	15.8(0.4)	**1.2(0)**
		Tobit LLA	**1(0.01)**	**0.36(0.02)**	**1.41(0.04)**	5.3(0.3)	**1.2(0)**
		SAWCT2018	1.22(0.01)	0.95(0.03)	3.2(0.07)	15.6(0.5)	**1.2(0)**
		OLS Oracle	1.87(0.01)	-	-	-	-
$\frac{1}{4}$	50	Lasso	2.91(0.03)	4.83(0.15)	5.31(0.11)	5.9(0.3)	1.1(0.1)
		SCAD	2.74(0.03)	3.31(0.13)	3.9(0.09)	1.8(0.2)	1.6(0.1)
		Tobit Lasso	0.91(0.01)	0.47(0.02)	2.03(0.05)	7.2(0.3)	**0.7(0.1)**
		Tobit LLA	**0.84(0.01)**	**0.28(0.02)**	**1.1(0.03)**	**1.4(0.1)**	1(0)
		SAWCT2018	0.95(0.01)	0.67(0.03)	2.01(0.05)	5(0.2)	0.8(0)
		OLS Oracle	2.57(0.01)	-	-	-	-
		OLS	4.81(0.08)	-	-	-	-
	500	Lasso	2.99(0.03)	5.94(0.17)	6.46(0.13)	13.3(0.7)	1.8(0.1)
		SCAD	2.69(0.03)	3.34(0.14)	4.26(0.09)	5.7(0.4)	1.9(0.1)
		Tobit Lasso	1.06(0.01)	1.07(0.05)	3.35(0.08)	14.9(0.5)	**1.2(0)**
		Tobit LLA	**0.89(0.01)**	**0.47(0.03)**	**1.58(0.05)**	**5.4(0.4)**	**1.2(0.1)**
		SAWCT2018	1.11(0.02)	1.33(0.05)	3.58(0.08)	14(0.5)	**1.2(0.1)**
		OLS Oracle	2.41(0.01)	-	-	-	-
$\frac{1}{2}$	50	Lasso	3.15(0.03)	15.52(0.21)	8.63(0.08)	5.4(0.3)	1.6(0.1)
		SCAD	3.12(0.03)	13.32(0.24)	7.58(0.1)	2(0.2)	2.2(0.1)
		Tobit Lasso	0.68(0.01)	0.86(0.05)	2.65(0.08)	6.7(0.3)	**0.9(0.1)**
		Tobit LLA	**0.63(0.01)**	**0.55(0.05)**	**1.52(0.06)**	**1.3(0.1)**	1.3(0.1)
		SAWCT2018	0.77(0.01)	2.13(0.1)	3.4(0.08)	4.2(0.2)	1(0.1)
		OLS Oracle	2.87(0.02)	-	-	-	-
		OLS	5.48(0.08)	-	-	-	-
	500	Lasso	3.48(0.03)	18.3(0.26)	9.76(0.13)	10.1(0.7)	2.4(0.1)
		SCAD	3.25(0.02)	14.92(0.26)	8.14(0.07)	**2.8(0.4)**	2.8(0)
		Tobit Lasso	0.94(0.02)	2.21(0.1)	4.27(0.09)	12.7(0.4)	**1.5(0.1)**
		Tobit LLA	**0.7(0.01)**	**0.82(0.05)**	**1.99(0.06)**	4.4(0.3)	1.6(0.1)
		SAWCT2018	1.02(0.02)	3.81(0.15)	5.36(0.1)	12.9(0.4)	1.6(0.1)
		OLS Oracle	2.85(0.01)	-	-	-	-

Table 3: Simulation Results with CS(0.8) Covariates

q	p	Method	MSE	ℓ_2	ℓ_1	FP	FN
$\frac{1}{8}$	50	Lasso	2(0.02)	3.02(0.15)	4.52(0.14)	6(0.3)	1.2(0.1)
		SCAD	1.94(0.02)	1.83(0.13)	2.76(0.11)	**0.9(0.1)**	2.1(0.1)
		Tobit Lasso	1.06(0.01)	1.02(0.05)	2.74(0.08)	6(0.2)	**1(0.1)**
		Tobit LLA	**1.02(0.01)**	**0.74(0.05)**	**1.77(0.06)**	1.2(0.1)	1.5(0.1)
		SAWCT2018	1.16(0.01)	1.61(0.07)	2.71(0.07)	3.4(0.2)	1.2(0.1)
		OLS Oracle	1.79(0.01)	-	-	-	-
		OLS	3.38(0.05)	-	-	-	-
	500	Lasso	2.11(0.03)	4.56(0.17)	6.34(0.22)	13.6(0.7)	1.9(0.1)
		SCAD	1.94(0.02)	2.45(0.17)	3.15(0.11)	**1.3(0.2)**	2.5(0.1)
		Tobit Lasso	1.21(0.01)	1.98(0.08)	4(0.08)	11.3(0.4)	**1.6(0.1)**
		Tobit LLA	**1.07(0.01)**	**1.09(0.07)**	**2.26(0.07)**	4.1(0.3)	1.8(0.1)
		SAWCT2018	1.22(0.01)	2.12(0.08)	3.91(0.07)	10.5(0.3)	**1.6(0.1)**
		OLS Oracle	1.65(0.01)	-	-	-	-
$\frac{1}{4}$	50	Lasso	2.52(0.03)	5.66(0.22)	5.81(0.12)	5.4(0.3)	1.6(0.1)
		SCAD	2.54(0.03)	4.07(0.26)	4.32(0.18)	**0.8(0.1)**	2.3(0.1)
		Tobit Lasso	0.92(0.01)	1.01(0.05)	2.74(0.08)	5.8(0.2)	**1.1(0.1)**
		Tobit LLA	**0.89(0.01)**	**0.86(0.05)**	**1.87(0.06)**	1(0.1)	1.7(0)
		SAWCT2018	1.02(0.01)	1.8(0.08)	2.94(0.07)	3.3(0.2)	1.4(0.1)
		OLS Oracle	2.29(0.01)	-	-	-	-
		OLS	4.32(0.07)	-	-	-	-
	500	Lasso	2.71(0.03)	8.37(0.26)	7.56(0.23)	10.8(0.7)	2.4(0.1)
		SCAD	2.57(0.02)	5.55(0.27)	4.92(0.13)	**1.1(0.2)**	2.7(0)
		Tobit Lasso	1.05(0.02)	2.29(0.11)	4.17(0.09)	11.3(0.4)	**1.6(0.1)**
		Tobit LLA	**0.93(0.01)**	**1.37(0.09)**	**2.48(0.08)**	3.8(0.3)	2(0.1)
		SAWCT2018	1.09(0.02)	2.66(0.11)	4.14(0.09)	9.8(0.3)	1.7(0.1)
		OLS Oracle	2.2(0.01)	-	-	-	-
$\frac{1}{2}$	50	Lasso	2.74(0.01)	17.09(0.27)	9.26(0.12)	4.8(0.3)	2.1(0.1)
		SCAD	2.74(0.02)	14.64(0.28)	7.96(0.1)	**0.5(0.1)**	2.9(0)
		Tobit Lasso	**0.63(0.01)**	**1.56(0.09)**	3.34(0.08)	6(0.3)	**1.2(0.1)**
		Tobit LLA	0.65(0.01)	1.67(0.12)	**2.54(0.09)**	1(0.1)	1.9(0.1)
		SAWCT2018	0.78(0.01)	4.06(0.15)	4.34(0.08)	2.8(0.2)	1.7(0.1)
		OLS Oracle	2.56(0.01)	-	-	-	-
		OLS	4.93(0.07)	-	-	-	-
	500	Lasso	3(0.02)	19.32(0.31)	10.18(0.19)	8.4(0.6)	2.8(0)
		SCAD	2.85(0.02)	14.99(0.23)	8.03(0.06)	**0.4(0.1)**	3(0)
		Tobit Lasso	0.82(0.02)	3.69(0.19)	5.11(0.11)	9.7(0.3)	**2.1(0.1)**
		Tobit LLA	**0.73(0.02)**	**2.46(0.17)**	**3.14(0.09)**	2.8(0.3)	2.4(0.1)
		SAWCT2018	0.89(0.02)	5.25(0.2)	5.51(0.11)	7.2(0.3)	2.3(0.1)
		OLS Oracle	2.65(0.01)	-	-	-	-

Table 4: Simulation Results with AR1(0.5) Covariates

q	p	Method	MSE	ℓ_2	ℓ_1	FP	FN
$\frac{1}{8}$	50	Lasso	2.25(0.02)	1.58(0.07)	3.03(0.07)	3.8(0.3)	1.3(0.1)
		SCAD	2.13(0.02)	1.09(0.05)	2.34(0.06)	1.8(0.2)	1.4(0.1)
		Tobit Lasso	1.02(0.01)	0.32(0.01)	1.7(0.04)	6(0.3)	**0.8(0)**
		Tobit LLA	**0.97(0.01)**	**0.26(0.02)**	**1.05(0.03)**	**1.1(0.1)**	1(0.1)
		SAWCT2018	1.03(0.01)	0.38(0.02)	1.55(0.03)	3.9(0.2)	1(0)
		OLS Oracle	2.02(0.01)	-	-	-	-
		OLS	3.88(0.07)	-	-	-	-
	500	Lasso	2.78(0.04)	2.37(0.09)	4.18(0.13)	8.9(0.7)	1.8(0)
		SCAD	2.48(0.03)	1.33(0.07)	2.83(0.09)	4.5(0.4)	1.7(0.1)
		Tobit Lasso	1.25(0.01)	0.69(0.03)	2.73(0.05)	13(0.5)	1.3(0)
		Tobit LLA	**1.07(0.01)**	**0.36(0.02)**	**1.35(0.04)**	**2.4(0.2)**	**1.2(0)**
		SAWCT2018	1.3(0.02)	0.82(0.03)	2.64(0.06)	9.7(0.6)	1.4(0.1)
		OLS Oracle	2.23(0.01)	-	-	-	-
$\frac{1}{4}$	50	Lasso	3.34(0.04)	4.66(0.14)	5.08(0.11)	4.3(0.3)	1.5(0.1)
		SCAD	3.23(0.04)	3.64(0.13)	4.19(0.1)	2.3(0.2)	1.7(0.1)
		Tobit Lasso	0.93(0.01)	0.4(0.02)	1.94(0.05)	6.5(0.3)	**0.9(0)**
		Tobit LLA	**0.88(0.01)**	**0.33(0.02)**	**1.19(0.04)**	**1(0.1)**	1.2(0)
		SAWCT2018	0.94(0.01)	0.53(0.02)	1.87(0.03)	4.2(0.2)	1(0)
		OLS Oracle	2.98(0.02)	-	-	-	-
		OLS	5.64(0.1)	-	-	-	-
	500	Lasso	3.72(0.05)	5.39(0.15)	6(0.13)	9.1(0.8)	1.9(0)
		SCAD	3.51(0.04)	3.74(0.14)	4.85(0.09)	7.1(0.5)	2(0.1)
		Tobit Lasso	1.15(0.02)	0.86(0.03)	2.95(0.05)	12.6(0.4)	1.4(0)
		Tobit LLA	**0.97(0.01)**	**0.46(0.02)**	**1.5(0.04)**	**2.5(0.2)**	**1.3(0)**
		SAWCT2018	1.23(0.02)	1.13(0.04)	3.18(0.07)	11.1(0.6)	1.5(0.1)
		OLS Oracle	3.03(0.01)	-	-	-	-
$\frac{1}{2}$	50	Lasso	3.88(0.03)	15.66(0.19)	8.44(0.07)	3.4(0.3)	1.8(0.1)
		SCAD	3.87(0.04)	14.01(0.2)	7.84(0.09)	2(0.2)	2.3(0.1)
		Tobit Lasso	0.68(0.01)	0.67(0.03)	2.3(0.05)	5.4(0.2)	**1.1(0)**
		Tobit LLA	**0.63(0.01)**	**0.53(0.03)**	**1.52(0.04)**	**0.9(0.1)**	1.5(0.1)
		SAWCT2018	0.73(0.01)	1.44(0.06)	2.87(0.06)	3(0.2)	1.3(0)
		OLS Oracle	3.54(0.01)	-	-	-	-
		OLS	6.78(0.1)	-	-	-	-
	500	Lasso	4.08(0.04)	16.79(0.21)	9.1(0.11)	6(0.6)	2.3(0)
		SCAD	4.02(0.03)	14.17(0.21)	8.25(0.08)	3.8(0.5)	2.7(0.1)
		Tobit Lasso	0.82(0.02)	1.45(0.06)	3.56(0.08)	10(0.4)	1.6(0)
		Tobit LLA	**0.64(0.01)**	**0.7(0.04)**	**1.83(0.05)**	**2.3(0.2)**	**1.5(0.1)**
		SAWCT2018	0.94(0.02)	2.93(0.1)	4.57(0.09)	8.5(0.5)	1.7(0)
		OLS Oracle	3.52(0.02)	-	-	-	-

Table 5: Simulation Results with AR1(0.8) Covariates

q	p	Method	MSE	ℓ_2	ℓ_1	FP	FN
$\frac{1}{8}$	50	Lasso	2.01(0.02)	2.01(0.1)	3.57(0.1)	4.5(0.3)	1.6(0.1)
		SCAD	1.93(0.02)	1.65(0.09)	2.72(0.07)	**1.1(0.1)**	2.1(0)
		Tobit Lasso	1.04(0.01)	**0.69(0.03)**	2.17(0.06)	4.8(0.2)	**1.3(0.1)**
		Tobit LLA	**0.99(0.01)**	**0.69(0.05)**	**1.72(0.05)**	1.5(0.1)	1.8(0)
		SAWCT2018	1.06(0.01)	0.9(0.04)	2.06(0.04)	2.6(0.2)	1.6(0)
		OLS Oracle	1.8(0.01)	-	-	-	-
		OLS	3.42(0.07)	-	-	-	-
	500	Lasso	2.53(0.04)	3.42(0.13)	4.69(0.09)	9.1(0.6)	2.1(0)
		SCAD	2.23(0.02)	2.18(0.1)	3.3(0.06)	**3.9(0.4)**	2.5(0.1)
		Tobit Lasso	1.26(0.02)	**1.51(0.06)**	3.31(0.05)	11.9(0.4)	**2(0)**
		Tobit LLA	**1.19(0.02)**	1.69(0.1)	**2.84(0.07)**	4.1(0.4)	2.2(0.1)
		SAWCT2018	1.3(0.02)	1.57(0.06)	3.46(0.07)	12(0.5)	**2(0)**
		OLS Oracle	1.9(0.01)	-	-	-	-
$\frac{1}{4}$	50	Lasso	2.91(0.03)	5.08(0.18)	5.45(0.1)	4(0.3)	1.8(0)
		SCAD	2.84(0.03)	4.23(0.19)	4.53(0.13)	1.6(0.2)	2.5(0.1)
		Tobit Lasso	0.91(0.01)	0.83(0.04)	2.36(0.06)	4.8(0.2)	**1.4(0.1)**
		Tobit LLA	**0.85(0.01)**	**0.71(0.05)**	**1.78(0.05)**	**1.5(0.2)**	1.8(0)
		SAWCT2018	0.95(0.01)	1.14(0.05)	2.34(0.05)	2.5(0.2)	1.7(0)
		OLS Oracle	2.55(0.01)	-	-	-	-
		OLS	5.04(0.09)	-	-	-	-
	500	Lasso	3.62(0.04)	7.15(0.2)	6.6(0.09)	8.1(0.6)	2.4(0.1)
		SCAD	3.28(0.02)	4.81(0.2)	5.07(0.1)	4.2(0.4)	2.8(0)
		Tobit Lasso	1.2(0.02)	**1.99(0.09)**	3.72(0.07)	10.5(0.4)	2(0)
		Tobit LLA	**1.08(0.02)**	2.03(0.1)	**3.04(0.07)**	**3.3(0.3)**	2.4(0.1)
		SAWCT2018	1.25(0.03)	2.14(0.09)	4.11(0.09)	12.1(0.5)	**1.9(0)**
		OLS Oracle	2.82(0.02)	-	-	-	-
$\frac{1}{2}$	50	Lasso	3.33(0.03)	15.47(0.23)	8.59(0.08)	3.3(0.3)	2.3(0.1)
		SCAD	3.27(0.02)	13.93(0.24)	7.81(0.07)	**1(0.2)**	2.9(0)
		Tobit Lasso	0.67(0.01)	**1.14(0.06)**	2.74(0.07)	4.1(0.2)	**1.6(0.1)**
		Tobit LLA	**0.65(0.01)**	1.33(0.1)	**2.31(0.09)**	**1(0.1)**	2.1(0.1)
		SAWCT2018	0.75(0.01)	2.2(0.08)	3.34(0.06)	1.8(0.1)	1.9(0)
		OLS Oracle	3.05(0.02)	-	-	-	-
		OLS	5.93(0.1)	-	-	-	-
	500	Lasso	3.71(0.04)	18.32(0.2)	9.58(0.16)	6.1(0.7)	2.6(0)
		SCAD	3.36(0.02)	15.27(0.22)	8.38(0.07)	2.6(0.3)	3(0)
		Tobit Lasso	1.04(0.03)	3.33(0.15)	4.51(0.08)	8(0.4)	**2.1(0)**
		Tobit LLA	**0.86(0.02)**	**2.98(0.11)**	**3.53(0.05)**	**1.7(0.2)**	2.9(0)
		SAWCT2018	1.15(0.03)	4.81(0.19)	5.75(0.1)	9.7(0.5)	**2.1(0)**
		OLS Oracle	3.07(0.02)	-	-	-	-

of censored observations is high.

Narrowing our focus to the three models which account for censoring, we see that both of the penalized Tobit models achieve lower average test MSEs than SAWCT2018 in all 30 simulation settings and that Tobit LLA achieves the lowest average test MSE in most cases, often by a comfortable margin. In particular, we see that Tobit LLA gains a larger edge over the Tobit lasso and SAWCT2018 in simulations settings with $p = 500$ relative to those with $p = 50$.

5.2 Estimation results

Turning to estimation performance, we see patterns similar to those that emerged in our prediction comparison. The penalized Tobit models' estimates have the two lowest average ℓ_2 losses in all 30 simulation settings (the Tobit LLA estimates have the lowest average ℓ_2 loss overall in 27 of 30 settings). In addition, the Tobit LLA estimates deliver the lowest average ℓ_1 loss in every simulation setting.

As in the prediction comparison there is a clear gap between the least squares methods and the models which account for censoring, with the latter consistently having far lower ℓ_2 and ℓ_1 estimation losses. In many cases, the average ℓ_2 losses for the Tobit and SAWCT2018 estimates differ from those of the least squares estimates by an order of magnitude. Additionally, we once again find that the gap in estimation performance between the models that account for censoring and the least squares models grows as the proportion of censored observations increases to $q = \frac{1}{2}$.

Among the models which account for censoring, the penalized Tobit models' estimates consistently achieve lower average ℓ_2 losses than the SAWCT2018 estimates. Shifting our focus to the ℓ_1 loss, we see that the Tobit LLA estimates achieve markedly lower average ℓ_1 losses than the Tobit lasso and SAWCT2018 estimates in every setting. The competition between the Tobit lasso and SAWCT2018, however, is closer, with the Tobit lasso estimates

achieving a lower average ℓ_1 loss than the SAWCT2018 estimates in just 18 of 30 settings.

5.3 Selection results

Our variable selection results are somewhat mixed. While the penalized Tobit models and SAWCT2018 consistently deliver lower average false negative counts than the least squares models, the differences are relatively small. At the same time, the SCAD and Tobit LLA models consistently make fewer false positive variable selections than the other models. Beyond that, neither SCAD nor Tobit LLA appears to have a clear edge in making fewer false positive selections, though Tobit LLA has a lower average false positive count in 19 of 30 settings.

Overall, the penalized Tobit models deliver comparable (if slightly superior) selection performances to the least squares models and SAWCT2018 in this study. These results further suggest that modelers may prefer to use the Tobit lasso if their goal is to minimize false negative variable selections and Tobit LLA if their goal is to minimize false positive variable selections.

5.4 Takeaways

Tobit LLA clearly outperformed competing methods in this simulation study, providing more accurate predictions and parameter estimates than the alternatives. Because it also has stronger theoretical guarantees than the Tobit lasso, we ultimately recommend Tobit LLA for analyzing high-dimensional left-censored data.

6 HIV Viral Load and Drug Resistance

Due to its short replication cycle and high mutation rate, human immunodeficiency virus (HIV) can rapidly develop drug resistance mutations (DRMs) in HIV-infected patients

receiving antiretroviral therapy. To counter this, guidelines recommended physicians regularly monitor HIV viral load and, if a patient's treatment regimen is failing to suppress the virus, conduct genotypic testing to check for DRMs so they may update the patient's drug regimen appropriately (Shafer, 2002).

There is a substantial literature devoted to identifying DRMs and quantifying the degree of resistance they provide against different antiretroviral treatments (Shafer, 2006). One way to accomplish this is by modeling the relationship between HIV viral load and mutations in the virus's genome. This poses two difficulties: (1) the observed viral load is left-censored because the assays used to measure it cannot detect concentrations below certain thresholds and (2) genome data are inherently high-dimensional. As we established in our simulation study, it is necessary to use a model which accounts for censoring when analyzing these kind of data. As such, we will use Tobit LLA and SAWCT2018 to model HIV viral load and identify potential DRMs.

Our data for this example come from the OPTIONS trial by the AIDS Clinical Trials Group (Gandhi et al., 2020) and were downloaded from the Stanford HIV Drug Resistance Database (Shafer, 2006). The OPTIONS trial study population consisted of 413 HIV-infected individuals receiving protease inhibitor (PI)-based treatment and experiencing virological failure. Each participant was given an optimized antiretroviral regimen based on their viral drug resistance and treatment history. Participants with moderate drug resistance were randomly assigned to either add nucleoside reverse transcriptase inhibitors (NRTIs) to their optimized regimens or omit NRTIs from their optimized regimens. Participants with highly drug-resistant HIV all received optimized regimens which included NRTIs.

We use Tobit LLA and SAWCT2018 to model HIV viral load 12 weeks after drug regimen assignment as a function of HIV genotypic mutations, current drug regimen, baseline viral load, observation week, and HIV subtype using a sample with $p \gg n$ and a moderate

Table 6: Prediction Accuracy on HIV Viral Load Data

Model	Tobit Loss
SAWCT2018	2.04 (0.02)
Tobit LLA	**1.45 (0.01)**

amount of left-censoring. Our data come from the $n = 407$ participants who returned for their 12-week follow-up evaluations and include $p = 1295$ predictors, most of which are indicators for protease (PR) and reverse transcriptase (RT) gene mutations. The assays used to measure HIV viral load in the OPTIONS trial had a detection threshold of 50 copies/mL. At their 12-week evaluations, 35.6% of study participants had viral loads which were at or below this lower limit and, consequently, undetectable. Given this limited information about these censored viral loads, investigators recorded them as falling at the lower limit of 50 copies/mL. We use $\log_{10}$-HIV viral load as our response, as it is often assumed to be normally distributed (Soret et al., 2018).

We start by comparing the prediction performance of Tobit LLA and SAWCT2018 in terms of the Tobit loss in order to assess overall model fit. We randomly split the data into a training set of 326 observations and a test set of 81 observations, using stratified sampling to ensure that the training and test sets have similar proportions of left-censored observations. We repeat this process 50 times. Within each of the 50 training sets, we tune Tobit LLA and SAWCT2018 using 5-fold CV. Table 6 reports the average Tobit loss across the 50 test sets, with the standard error in parentheses, for each model.

Our primary interest is in the predictors selected by the models, as they may include potential DRMs. We tune Tobit LLA and SAWCT2018 using 5-fold CV then fit them to the entire dataset. Tobit LLA selects a sparse model with only three predictors: the RT mutation M184V, baseline viral load, and whether the participant is taking raltegravir (RAL), an integrase strand transfer inhibitor (INSTI) included in some of the patients' optimized regimens. SAWCT2018, on the other hand, selects 51 mutations (including

M184V), baseline viral load, and whether the patient is taking RAL or the protease inhibitor saquinavir. While it is possible that the 50 other mutations selected by SAWCT2018 include additional DRMs, the superior prediction performance of the sparse Tobit LLA model suggests that M184V is uniquely important for predicting HIV viral load in this population. It seems far more likely that SAWCT2018 is selecting unimportant mutations, reducing its utility as a method for identifying potential DRMs.

The Tobit LLA model provides some interesting insights into HIV drug resistance. Most importantly, M184V stands out as the sole mutation selected by Tobit LLA. This selection is supported by other research: based on an extensive review of the HIV drug resistance literature, the Stanford HIV Drug Resistance Database lists M184V as a major NRTI resistance mutation (Shafer, 2006). It is also notable that Tobit LLA did not select any NRTIs as important predictors. This is consistent with Gandhi et al.'s (2020) finding that participants who added NRTIs to their regimes did not experience significantly higher rates of virological failure than those who omitted NRTIs from their regimes.

7 Discussion

As high-dimensional data become increasingly common across disciplines, we expect the need for reliable, theoretically-supported techniques for high-dimensional left-censored regression to grow. The penalized Tobit models we introduce in this paper fill several gaps in the literature for high-dimensional left-censored regression. They are among the first models in this area with theoretical guarantees in the setting where $p \gg n$ and the lasso-initialized two-step LLA estimator for folded-concave penalized Tobit regression is the very first to possess the strong oracle property. In addition, our penalized Tobit models provide the first high-dimensional extensions of the enduringly popular Tobit model.

Our penalized Tobit models also perform well empirically. In an extensive simulation

study, our penalized Tobit models delivered superior prediction and estimation performance relative to least squares models and the best available alternative for high-dimensional left-censored regression. When applied to real high-dimensional left-censored HIV viral load data, the Tobit LLA estimator delivered more accurate predictions and selected a more parsimonious model than the best available alternative.

Acknowledgments

We thank the referees and the editor whose thoughtful comments significantly improved this article.

Funding

This work is supported in part by NSF DMS 1915842 and 2015120.

SUPPLEMENTARY MATERIAL

penalized_tobit_proofs: This supplementary file contains intermediate results and technical proofs for the results in this paper.

References

Amemiya, T. (1984), "Tobit Models: A Survey," *Journal of Econometrics*, 24(1-2), 3–61.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009), "Simultaneous Analysis of Lasso and Dantzig Selector," *Annals of Statistics*, 37(4), 1705–1732.

Buckley, J. and James, I. (1979), "Linear Regression with Censored Data," *Biometrika*, 66(3), 429.

Fan, J. and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties," *Journal of the American Statistical Association*, 96(456), 1348–1360.

Fan, J. and Lv, J. (2011), "Nonconcave Penalized Likelihood with NP-dimensionality," *IEEE Transactions on Information Theory*, 57(8), 5467–5484.

Fan, J., Xue, L., and Zou, H. (2014), "Strong Oracle Optimality of Folded Concave Penalized Estimation," *Annals of Statistics*, 42(3), 819–849.

Friedman, J., Hastie, T., and Tibshirani, R. (2010), "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, 33, 1–22.

Gandhi, R. T., Tashima, K. T., Smeaton, L. M., Vu, V., Ritz, J., Andrade, A., Eron, J. J., Hogg, E., and Fichtenbaum, C. J. (2020), "Long-term Outcomes in a Large Randomized Trial of HIV-1 Salvage Therapy: 96-Week Results of AIDS Clinical Trials Group A5241 (OPTIONS)," *Journal of Infectious Diseases*, 221, 1407–1415.

Jacobson, T. and Zou, H. (2022), "High-dimensional Censored Regression via the Penalized Tobit Likelihood," *arXiv preprint arXiv:2203.02601.*

Johnson, B. A. (2009), "On Lasso for Censored Data," *Electronic Journal of Statistics*, 3, 485–506.

Li, Y., Dicker, L., and Zhao, S. D. (2014), "The Dantzig Selector for Censored Linear Regression Models," *Statistica Sinica*, 24(1), 251–268.

Müller, P. and van de Geer, S. (2016), "Censored Linear Model in High Dimensions: Penalised Linear Regression on High-dimensional Data with Left-censored Response Variable," *TEST*, 25(1), 75–92.

Olsen, R. J. (1978), "Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model," *Econometrica*, 46(5), 1211–1215.

Powell, J. L. (1984), "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.

Shafer, R. W. (2002), "Genotypic Testing for Human Immunodeficiency Virus Type 1 Drug Resistance," *Clinical Microbiology Reviews*, 15, 247–277.

—— (2006), "Rationale and Uses of a Public HIV Drug-resistance Database," *Journal of Infectious Diseases*, 194, 51–58.

Soret, P., Avalos, M., Wittkop, L., Commenges, D., and Thiébaut, R. (2018), "Lasso Regularization for Left-censored Gaussian Outcome and High-dimensional Predictors," *BMC Medical Research Methodology*, 18(1), 1–13.

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B*, 58(1), 267–288.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26(1), 24–36.

Yang, Y. and Zou, H. (2013), "An Efficient Algorithm for Computing the HHSVM and its Generalizations," *Journal of Computational and Graphical Statistics*, 22(2), 396–415.

Zhang, C. H. (2010), "Nearly Unbiased Variable Selection Under Minimax Concave Penalty," *Annals of Statistics*, 38(2), 894–942.

Zhao, P. and Yu, B. (2006), "On Model Selection Consistency of Lasso," *Journal of Machine Learning Research*, 7, 2541–2563.

Zhou, X. and Liu, G. (2016), "LAD-Lasso Variable Selection for Doubly Censored Median Regression Models," *Communications in Statistics - Theory and Methods*, 45, 3658–3667.

Zou, H. (2006), "The Adaptive Lasso and its Oracle Properties," *Journal of the American Statistical Association*, 101(476), 1418–1429.

Zou, H. and Li, R. (2008), "One-step Sparse Estimates in Nonconcave Penalized Likelihood Models," *Annals of Statistics*, 36(4), 1509–1533.