Statistica Sinica Preprint No: SS-2023-0118	
Title	Linear Hypothesis Testing for High Dimensional Tobit Models
Manuscript ID	SS-2023-0118
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0118
Complete List of Authors	Tate Jacobson and Hui Zou
Corresponding Authors	Tate Jacobson
E-mails	jacobtat@oregonstate.edu

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

Linear Hypothesis Testing for High Dimensional Tobit Models

Tate Jacobson[1] and Hui Zou[2]

[1]*Department of Statistics, Oregon State University*

[2]*School of Statistics, University of Minnesota*

Abstract: Few methods have been developed for conducting statistical inference in high-dimensional left-censored regression. Among the methods that do exist, none are flexible enough to test general linear hypotheses—that is, all hypotheses of the form $\mathrm{H}_0 : \mathbf{C}\boldsymbol{\beta}^*_{\mathcal{M}} = \boldsymbol{t}$. To fill this gap, we introduce partial penalized Tobit tests for testing general linear hypotheses in high-dimensional left-censored data. In particular, we develop partial penalized Wald, score, and likelihood ratio tests for high-dimensional Tobit models. We derive approximate distributions for the partial penalized Tobit test statistics under the null hypothesis and local alternatives in an ultra high-dimensional setting, finding that the tests achieve their nominal size asymptotically and that they are approximately equivalent for large n. In addition, we derive the tests' approximate power in this setting. We propose an alternating direction method of multipliers algorithm to compute the partial penalized test statistics. Through an extensive empirical study, we show that the partial penalized Tobit tests achieve their nominal size and that they are consistent in a finite sample setting. As an application, we analyze data from

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

the AIDS Clinical Trials Group, using our partial penalized Tobit tests to test whether certain HIV mutations are significant predictors of HIV viral load.

Key words and phrases: censoring, high-dimensional statistical inference, hypothesis testing.

1. Introduction

As it has become easier to collect large amounts of data, high dimensional modeling problems have become increasingly common in many domains. For researchers analyzing data with a left-censored response—common in some economic and medical applications—the availability of high dimensional data leads to the challenge of dealing with two modeling complications at once. As a motivating example, we consider the problem of modeling the relationship between human immunodeficiency virus (HIV) viral load and mutations in the HIV genome. The assays used to measure HIV viral load cannot detect the virus if its concentration is below a certain (known) threshold. Rather than discarding these observations, researchers simply record that the viral load is less than or equal to the detection threshold. As a result, the observed viral load is left-censored at the threshold value. At the same time, the number of participants in any given study is typically smaller than the number of unique mutations

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

present in the genomes of the participants' HIV infections. In particular, genome data are often assumed to be *ultra high-dimensional*, meaning that the number of predictors p grows almost exponentially with the number of observations n, as adding new participants to the study introduces many new mutations into the sample.

Researchers analyzing high-dimensional left-censored data need statistical models which meaningfully account for both high-dimensions and censoring. Only a few estimators have been developed for performing regression in this setting: Johnson (2009), Li, Dicker and Zhao (2014), and Soret et al. (2018) extended the Buckley-James estimator for high-dimensional data (Buckley and James, 1979); Müller and Van de Geer (2016) and Zhou and Liu (2016) extended the least absolute deviation (LAD) estimator (Powell, 1984); and Jacobson and Zou (2023) extended the Tobit model (Tobin, 1958). Of these estimators, only those introduced by Müller and Van de Geer (2016) and Jacobson and Zou (2023) possess any theoretical guarantees in the ultra high-dimensional setting—Müller and Van de Geer (2016) established that their lasso-penalized censored LAD estimator is consistent while Jacobson and Zou (2023) proved that their folded-concave penalized Tobit estimator possesses the strong oracle property.

In this study, our interest lies in developing flexible testing procedures

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

for high dimensional left-censored regression. The literature on inference for high-dimensional regression models has grown rapidly in recent years. Several authors have used the desparsifying or debiasing technique (Van de Geer et al., 2014; Zhang and Zhang, 2014; Javanmard and Montanari, 2014; Cai and Guo, 2017; Cai, Guo and Ma, 2021) to develop inferential procedures for lasso type estimators. Zhang and Cheng (2017) introduced a bootstrap-assisted test based on the desparsified lasso estimator for testing $\mathrm{H}_0 : \boldsymbol{\beta}^*_{\mathcal{M}} = \boldsymbol{t}$, where $\boldsymbol{\beta}^*_{\mathcal{M}}$ is a subvector of the true regression coefficient vector $\boldsymbol{\beta}^*$, in high-dimensional generalized linear models (GLMs). Likewise, Ma, Cai and Li (2021) used a debiased lasso estimator to develop a test of $\mathrm{H}_0 : \boldsymbol{\beta}^*_{\mathcal{M}} = \mathbf{0}$ for high-dimensional logistic regression. Ning and Liu (2017) debiased Rao's score test statistic to develop a decorrelated score function for hypothesis testing and constructing confidence regions for generic penalized M-estimators. Similarly, Fang, Ning and Liu (2017) developed decorrelated score, Wald, and partial likelihood ratio tests for high-dimensional Cox regression. See Cai, Guo and Xia (2023) for a thorough review of debiasing methods for high-dimensional inference.

A few recent high-dimensional inference methods do not rely on the debiasing technique. Chang et al. (2021) introduced a method for constructing confidence regions for $\boldsymbol{\theta}_{\mathcal{M}}$, a low-dimensional component of the full param-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

eter vector $\boldsymbol{\theta}$, using general estimating equations. Rather than applying a bias-correction, the authors transformed the estimating functions in the empirical likelihood to reduce the impact of the high-dimensional nuisance parameters $\boldsymbol{\theta}_{\mathcal{M}^c}$ in the estimation. In addition, they discussed the possibility of extending their method to construct confidence regions for transformations of $\boldsymbol{\theta}_{\mathcal{M}}$, including linear transformations $\mathbf{C}\boldsymbol{\theta}_{\mathcal{M}}$. Cui, Guo and Zhong (2018) introduced a procedure which uses a refitted cross-validation estimate of the variance to test the significance of the entire coefficient vector in linear models. Chen, Li and Chen (2023) developed a score test of $\mathrm{H}_0 : \boldsymbol{\beta}^*_{\mathcal{M}} = \boldsymbol{t}$ for non-sparse subvectors $\boldsymbol{\beta}^*_{\mathcal{M}}$ of $\boldsymbol{\beta}^*$ in high-dimensional GLMs. Wang and Cui (2014) introduced a partial penalized likelihood ratio test for testing $\mathrm{H}_0 : \boldsymbol{\beta}^*_{\mathcal{M}} = \mathbf{0}$ in the setting where $p = o\left(n^{1/5}\right)$. Expanding their approach, Shi et al. (2019) proposed partial penalized Wald, score, and likelihood ratio tests for testing general linear hypotheses—that is, all hypotheses of the form $\mathrm{H}_0 : \mathbf{C}\boldsymbol{\beta}^*_{\mathcal{M}} = \boldsymbol{t}$—in GLMs in the ultra high-dimensional setting. Among existing testing procedures for high-dimensional GLMs, only Shi et al.'s (2019) partial penalized tests cover this broad class of testing problems.

None of the studies outlined above examine left-censored data. However, Ning and Liu's (2017) test of $\mathrm{H}_0 : \boldsymbol{\beta}^*_{\mathcal{M}} = \mathbf{0}$ for generic penalized

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

M-estimators and Chang et al.s' (2021) approach to constructing confidence regions for $\boldsymbol{\theta}_{\mathcal{M}}$ using general estimating equations could be adapted for several left-censored regression models. To our best knowledge, only Bradic and Guo (2019) have proposed inferential procedures specifically for high-dimensional left-censored regression: they developed one-step estimators based on the lasso-penalized censored LAD estimator (Müller and Van de Geer, 2016) to construct robust confidence intervals for contrasts of the regression coefficients.

The main goal of this study is to develop flexible testing procedures for high-dimensional left-censored regression based on the penalized Tobit model (Jacobson and Zou, 2023). We extend the partial penalized testing framework of Shi et al. (2019) for its unique flexibility among testing procedures for high-dimensional GLMs. Because the Tobit model is not a GLM, significant effort is required to develop theory for our new procedures. We design our testing procedures to be flexible enough to test general linear hypotheses and to generalize to the ultra high-dimensional setting. In our theoretical study of these tests, we also allow the number of constraints under the null hypothesis and the set of coefficients being tested to grow with n, ensuring that the tests are suitable for a wide range of applications.

We use partial penalized Tobit estimators to develop partial penalized

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

Wald, score, and likelihood ratio tests for high-dimensional Tobit regression. We derive rates of convergence and limiting expressions for partial penalized Tobit estimators with folded concave penalties and, using these estimators, derive approximate distributions for the partial penalized test statistics for large n under the null hypothesis and local alternatives in the ultra high-dimensional setting. From these results, we establish that the partial penalized Tobit tests are approximately equivalent for large n and achieve their nominal size asymptotically. In addition, we derive their approximate power under local alternatives. For our implementation of these testing procedures, we develop an alternating direction method of multipliers (ADMM) algorithm (Boyd et al., 2011) to compute the partial penalized Tobit estimators by minimizing the partial penalized negative Tobit log-likelihood.

This paper is organized as follows. In Section 2 we introduce general linear hypotheses for high-dimensional Tobit regression. In Section 3 we develop the partial penalized Tobit estimators and examine their asymptotic properties. In Section 4 we introduce the partial penalized Tobit hypothesis tests and derive approximate distributions for their test statistics. In Section 5 we outline our ADMM algorithm for computing the partial penalized Tobit estimators and discuss details of our implementation. In Section 6 we

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

assess the finite sample behavior of the partial penalized Tobit hypothesis tests through an extensive simulation study. In Section 7 we return to our motivating example, applying the partial penalized Tobit hypothesis tests to conduct significance testing for potential HIV drug resistance mutations using data from the AIDS Clinical Trials Group. Theoretical proofs, supporting lemmas, and additional simulation results are given in Sections S.1, S.2, and S.3 of the supplementary material, respectively.

2. Testing Setup

We begin by introducing high-dimensional left-censored data and general linear hypotheses. Suppose we have predictors $\mathbf{x} = (1, x_1, \ldots, x_p)' \in \mathbb{R}^{p+1}$ and a left-censored response $y \geq L$, where L is a known left-censoring point. Following Tobin (1958), we assume that there exists an unobserved latent response variable y^* such that $y = \max\{y^*, L\}$ and that y^* is generated from a linear model, $y^* = \mathbf{x}'\boldsymbol{\beta}^* + \varepsilon$, where $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1^*, \ldots, \beta_p^*) \in \mathbb{R}^{p+1}$ and $\varepsilon \sim N(0, \sigma^{*2})$. We assume that $L = 0$ without loss of generality. Let $\{(y_i, \mathbf{x}_i')\}_{i=1}^n$ be the observed data and assume that the predictors are fixed. In addition to working with a left-censored response, we assume that the data are ultra-high-dimensional, with $\log p = O(n^\eta)$ for some $\eta \in (0, 1)$.

We aim to test general linear hypotheses about $\boldsymbol{\beta}^*$, the true regression

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

coefficient vector. Let $\mathcal{M} \subseteq \{0, 1, \ldots, p\}$ denote the set of indices for the coefficients being tested and let $m = |\mathcal{M}|$. General linear hypotheses are of the form:

$$\mathrm{H}_0 : \mathbf{C}\boldsymbol{\beta}^*_{\mathcal{M}} = \boldsymbol{t}, \tag{2.1}$$

where the constraint matrix $\mathbf{C} \in \mathbb{R}^{r\times m}$ is of full row rank (so none of the constraints placed on $\boldsymbol{\beta}^*$ under H_0 are redundant) and $\boldsymbol{t} \in \mathbb{R}^r$. Note that (2.1) is flexible enough to cover a wide range of null hypotheses, including special testing problems such as $\mathrm{H}_0 : \boldsymbol{\beta}^*_{\mathcal{M}} = \mathbf{0}$ and $\mathrm{H}_0 : \boldsymbol{a}'\boldsymbol{\beta}^*_{\mathcal{M}} = b$ where $a \in \mathbb{R}^m$, $b \in \mathbb{R}$.

3. Partial Penalized Tobit Regression

3.1 Tobit likelihood

Based on Tobin's (1958) latent-variable formulation for a left-censored response, one can derive the following likelihood:

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

where $d_i = \mathbb{1}_{y_i>0}$ and $\Phi(\cdot)$ denotes the standard normal CDF. Note that $\log L_n(\boldsymbol{\beta}, \sigma^2)$ is not concave in $(\boldsymbol{\beta}, \sigma^2)$. As such, we will instead use Olsen's (1978) reparameterization of the Tobit model, $\boldsymbol{\delta} := \boldsymbol{\beta}/\sigma$ and $\gamma := 1/\sigma$, as it gives us a concave log-likelihood $\log L_n(\boldsymbol{\delta}, \gamma)$. The Tobit log-likelihood in

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

$(\boldsymbol{\delta}, \gamma)$ is given, up to an affine transformation, by

$$\log L_n(\boldsymbol{\delta}, \gamma) = \sum_{i=1}^{n} d_i \left\{ \log(\gamma) - \frac{1}{2}(\gamma y_i - \mathbf{x}_i'\boldsymbol{\delta})^2 \right\} + (1 - d_i) \log \left\{ \Phi\left(-\mathbf{x}_i'\boldsymbol{\delta}\right) \right\}. \tag{3.1}$$

Note that we cannot separate the dispersion parameter γ from the regression coefficients $\boldsymbol{\delta}$ in $\log L_n(\boldsymbol{\delta}, \gamma)$. Since we will be working with the entire parameter vector, we define $\boldsymbol{\theta} := (\boldsymbol{\delta}', \gamma)'$ for ease of notation. We use $\boldsymbol{\beta}^*, \sigma^*, \boldsymbol{\delta}^*, \gamma^*$, and $\boldsymbol{\theta}^*$ to denote the true parameter values.

We use the $(\boldsymbol{\delta}, \gamma)$ parameterization solely to facilitate estimation. In taking this approach, we do not lose the ability to estimate and make inferences about $\boldsymbol{\beta}$ and σ. We see that $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ have the same sparsity pattern since $\delta_j = 0$ if and only if $\beta_j = 0$. Moreover, we can express (2.1) in terms of $(\boldsymbol{\delta}, \gamma)$ or in terms of $\boldsymbol{\theta}$, as $\mathrm{H}_0 : \mathbf{C}\boldsymbol{\delta}^*_{\mathcal{M}} = \gamma^* \boldsymbol{t}$ or $\mathrm{H}_0 : \mathbf{C}^*\boldsymbol{\theta}^*_{\mathcal{M}'} = \mathbf{0}$, where $\mathbf{C}^* := \begin{bmatrix} \mathbf{C} & -\boldsymbol{t} \end{bmatrix}$ and $\mathcal{M}' := \mathcal{M} \cup \{p+1\}$ so that $\boldsymbol{\theta}^*_{\mathcal{M}'} = (\boldsymbol{\delta}^*_{\mathcal{M}}{}', \gamma^*)'$. It is the first non-trivial result that the convex reparameterization of Tobit likelihood does not alter the linear nature of hypotheses. We will switch between these equivalent expressions for H_0 as needed.

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

3.2 Partial penalized negative Tobit log-likelihood

Let $\ell_n(\boldsymbol{\theta}) = -\frac{1}{n}\log L_n(\boldsymbol{\theta})$ and $p_\lambda(\cdot)$ be a penalty function. We use the *partial penalized negative Tobit log-likelihood*

$$Q_n(\boldsymbol{\theta}) := \ell_n(\boldsymbol{\theta}) + \sum_{j \in \mathcal{M}^c \setminus \{0\}} p_{\lambda_n}(|\delta_j|), \tag{3.2}$$

to compute the estimators used in our tests. Note that the coefficients being tested in H_0 are not penalized in (3.2). Whereas debiasing techniques for high-dimensional inference remove bias from an estimator after the fact, the partial penalized likelihood avoids introducing bias in the first place. Because we leave $\boldsymbol{\delta}_{\mathcal{M}}$ unpenalized, we can avoid placing a minimum signal strength assumption on $\boldsymbol{\delta}^*_{\mathcal{M}}$, giving our tests power at local alternatives.

3.3 Partial penalized Tobit estimators

We define the following *partial penalized Tobit estimators* based on the partial penalized negative Tobit log-likelihood:

$$\hat{\boldsymbol{\theta}}_0 := \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^{p+2}} Q_n(\boldsymbol{\theta}) \quad \text{subject to } \mathbf{C}^*\boldsymbol{\theta}_{\mathcal{M}'} = \mathbf{0} \tag{3.3}$$

$$\hat{\boldsymbol{\theta}}_a := \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^{p+2}} Q_n(\boldsymbol{\theta}). \tag{3.4}$$

We refer to $\hat{\boldsymbol{\theta}}_0$ as the *reduced model* estimator and $\hat{\boldsymbol{\theta}}_a$ as the *full model* estimator. One can think of the partial penalized estimators as analogues

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

to the constrained and unconstrained maximum likelihood estimators, with $\hat{\boldsymbol{\theta}}_0$ minimizing $Q_n(\boldsymbol{\theta})$ over the set of $\boldsymbol{\theta}$ values which satisfy the constraints of the null hypothesis and $\hat{\boldsymbol{\theta}}_a$ minimizing $Q_n(\boldsymbol{\theta})$ without any constraints. As we will see in Section 4, the partial penalized estimators play similar roles to the constrained and unconstrained maximum likelihood estimators in our partial penalized extensions of the Wald, score, and likelihood ratio test statistics. Before we introduce the tests, however, we will examine the statistical properties of the partial penalized Tobit estimators.

3.3.1 Notation

We adopt the following notation throughout our study. Given a matrix $\mathbf{A} \in [a_{ij}]_{n\times m}$ and sets of indices $\mathcal{U} \subseteq \{1,\ldots,m\}$ and $\mathcal{T} \subseteq \{1,\ldots,n\}$, we let $\mathbf{A}_{(\mathcal{U})}$ denote the submatrix consisting of the *columns* of $\mathbf{A}$ with indices in $\mathcal{U}$ and $\mathbf{A}_{\mathcal{T}}$ to denote the submatrix consisting of the *rows* of $\mathbf{A}$ with indices in $\mathcal{T}$. We let $\lambda_{\min}\{\mathbf{A}\}$ and $\lambda_{\max}\{\mathbf{A}\}$ denote the smallest and largest eigenvalues, respectively, of $\mathbf{A}$. Let $\mathbf{A}'$ denote the transpose of $\mathbf{A}$. We will use the following 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 entrywise maximum $\|\mathbf{A}\|_{\max} = \max_{(i,j)} |a_{i,j}|$, and the entrywise minimum $\|\mathbf{A}\|_{\min} = \min_{(i,j)} |a_{i,j}|$. Given a square matrix $\mathbf{B}$, we let $\mathrm{tr}\{\mathbf{B}\}$ denote its

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

trace. Given a vector $\boldsymbol{v} \in \mathbb{R}^q$, we let diag$\{\boldsymbol{v}\}$ denote the diagonal matrix with the entries of $\boldsymbol{v}$ along its diagonal. We define $\|\boldsymbol{v}\|_0 = \#\{j : v_j \neq 0\}$. Given a function $f : \mathbb{R}^m \to \mathbb{R}$, we let $\nabla_{\mathcal{U}} f(\boldsymbol{t})$ and $\nabla^2_{\mathcal{U}} f(\boldsymbol{t})$ denote the gradient and Hessian, respectively, of $f(\boldsymbol{t})$ with respect to $\boldsymbol{t}_{\mathcal{U}}$.

For any symmetric matrix $\mathbf{S} \in \mathbb{R}^{q\times q}$, the spectral theorem guarantees that there exists an orthogonal matrix $\mathbf{U}$ and a diagonal matrix $\mathbf{D} = \text{diag}\{d_1, \dots, d_q\}$ such that $\mathbf{S} = \mathbf{UDU}'$. If $\mathbf{S}$ is positive semidefinite, then the entries of $\mathbf{D}$ are non-negative and we define $\mathbf{D}^{1/2} = \text{diag}\{d_1^{1/2}, \dots, d_q^{1/2}\}$ and $\mathbf{D}^{-1/2} = \text{diag}\{d_1^{-1/2}, \dots, d_q^{-1/2}\}$. In addition, we define $\mathbf{S}^{1/2} = \mathbf{UD}^{1/2}\mathbf{U}'$ and $\mathbf{S}^{-1/2} = \mathbf{UD}^{-1/2}\mathbf{U}'$.

3.3.2 Assumptions

We assume, without loss of generality, that $0 \in \mathcal{M}$ and that the predictors are reordered so that $\mathcal{M} = \{0, \dots, m-1\}$. Let $\mathcal{S} = \{j \in \mathcal{M}^c : \delta_j^* \neq 0\}$ and $s = |\mathcal{S}|$. We assume the following about the true model:

(A1) $\boldsymbol{\beta}^*$ is sparse and satisfies $\mathbf{C}\boldsymbol{\beta}^*_{\mathcal{M}} = \boldsymbol{t} + \boldsymbol{h}_n$, where $\boldsymbol{h}_n \to \mathbf{0}$;
$\lambda_{\max}\{(\mathbf{CC}')^{-1}\} = O(1)$; and
$\|\boldsymbol{h}_n\|_2 = O(\sqrt{\min\{s+m-r+1, r\}/n})$.

Condition (A1) implies that either the null hypothesis (2.1) is true (that is, if $\boldsymbol{h}_n = 0$) or the true model satisfies a sequence of local alternatives.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

We set $p_\lambda(\cdot)$ to be a folded-concave penalty function in order to reduce bias in estimating $\boldsymbol{\delta}_{\mathcal{M}^c}$. As a consequence, (3.2) may be nonconvex and may have multiple local minima. Define $\rho(t;\lambda) = \lambda^{-1}p_\lambda(t)$ for $\lambda > 0$. We assume that $\rho(t;\lambda)$ is increasing and concave on $[0,\infty)$ and is continuously differentiable on $(0,\infty)$. We further assume that $\rho'(0^+;\lambda) > 0$, that $\rho'(0^+;\lambda)$ is independent of λ, and that $\rho'(t;\lambda)$ is increasing in $\lambda \in (0,\infty)$. For any vector $\boldsymbol{v} \in \mathbb{R}^q$, we define $\bar{\boldsymbol{\rho}}(\boldsymbol{v};\lambda) := (\mathrm{sgn}(v_1)\rho'(|v_1|;\lambda),\ldots,\mathrm{sgn}(v_q)\rho'(|v_q|;\lambda))'$, where $\mathrm{sgn}(t)$ denotes the sign function. We define the local concavity of ρ at $\boldsymbol{v} \in \mathbb{R}^q$ with $\|\boldsymbol{v}\|_0 = q$ by

$$\kappa(\rho,\boldsymbol{v},\lambda) := \lim_{\epsilon\to 0^+}\max_{1\le j\le q}\sup_{t_1<t_2\in(|v_j|-\epsilon,|v_j|+\epsilon)} -\frac{\rho'(t_2;\lambda)-\rho'(t_1;\lambda)}{t_2-t_1}.$$

Note that if $\rho(t;\lambda)$ is twice continuously differentiable, then $\kappa(\rho,\boldsymbol{v},\lambda) = \max_{1\le j\le q} -\rho''(|v_j|;\lambda)$. Two popular folded-concave penalties are the SCAD penalty (Fan and Li, 2001), which has derivative $p'_\lambda(t) = \frac{(a\lambda-t)_+}{a-1}\mathbb{1}_{(t>\lambda)} + \lambda\mathbb{1}_{(t\le\lambda)}$ where $a > 2$, and the MCP (Zhang, 2010), which has derivative $p'_\lambda(t) = \left(\lambda - \frac{t}{a}\right)_+$ where $a > 1$.

Define $\mathcal{N}_0 = \{\boldsymbol{\theta} \in \mathbb{R}^{p+2} : \|\boldsymbol{\theta}_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\|_2 \le \sqrt{(s+m+1)\log n/n}$, $\boldsymbol{\theta}_{(\mathcal{M}'\cup\mathcal{S})^c} = \mathbf{0}\}$ and $\kappa_0 = \max_{\boldsymbol{\theta}\in\mathcal{N}_0}\kappa(\rho,\boldsymbol{\theta},\lambda_n)$. Let $d_n = \min_{j\in\mathcal{S}}\frac{|\delta_j^*|}{2}$. We assume the following about the penalty function $p_{\lambda_n}(\cdot)$:

(A2) $\lambda_n\kappa_0 = o(1)$; $\max\{\sqrt{s+m},\sqrt{\log p}\}/\sqrt{n} = o(\lambda_n)$, $\lambda_n = o(d_n)$; and

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

$$p'_{\lambda_n}(d_n) = o((s+m)^{-1/2}n^{-1/2}).$$

The minimum signal strength assumption in (A2) prevents the elements of $\hat{\boldsymbol{\delta}}_{0,\mathcal{S}}$ and $\hat{\boldsymbol{\delta}}_{a,\mathcal{S}}$ from being pushed to 0 by the penalty, enabling us to derive limiting expressions for $\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}}$ and $\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}}$ later in our study. Conditions like (A2) are often assumed to establish asymptotic guarantees for folded-concave penalized estimators (Fan and Lv, 2011; Shi et al., 2019).

As we can see in (3.1), the Tobit likelihood treats censored and uncensored observations differently. We introduce the following notation to express $\ell_n(\boldsymbol{\theta})$ and its derivatives in simpler forms. Let $n_1 = \sum_{i=1}^n d_i$ denote the number of uncensored observations and $n_0 = n - n_1$ denote the number of censored observations among $y_1, \ldots, y_n$. Let $\mathbf{y}_1 \in \mathbb{R}^{n_1}$ be the vector of uncensored response values, $y_i > 0$, and $\mathbf{y}_0 \in \mathbb{R}^{n_0}$ be the vector of censored response values, $y_i = 0$. Define $\mathbf{X}_1 \in \mathbb{R}^{n_1\times(p+1)}$ to be the matrix of predictors corresponding to $\mathbf{y}_1$ and $\mathbf{X}_0 \in \mathbb{R}^{n_0\times(p+1)}$ to be the matrix of predictors corresponding to $\mathbf{y}_0$. We reorder our observations so that $\mathbf{X} = \begin{bmatrix}\mathbf{X}'_0 & \mathbf{X}'_1\end{bmatrix}'$ and $\mathbf{y} = \begin{bmatrix}\mathbf{y}'_0 & \mathbf{y}'_1\end{bmatrix}'$.

Using this new notation, we can express $\ell_n(\boldsymbol{\theta})$ as

$$\ell_n(\boldsymbol{\theta}) = -\frac{n_1}{n}\log(\gamma) + \frac{1}{2n}(\gamma\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\delta})'(\gamma\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\delta}) - \frac{1}{n}\sum_{i=1}^{n_0}\log\{\Phi(-\mathbf{x}'_i\boldsymbol{\delta})\}$$

Let $g(s) = \phi(s)/\Phi(s) = \nabla\log(\Phi(s))$, where $\phi(\cdot)$ denotes the standard nor-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

mal density, and let $\boldsymbol{g}(\boldsymbol{\delta}) = (g(-\mathbf{x}_1'\boldsymbol{\delta}), \ldots, g(-\mathbf{x}_{n_0}'\boldsymbol{\delta}))'$. One can show that

$$\nabla \ell_n(\boldsymbol{\theta}) = \frac{1}{n}\begin{bmatrix} -\mathbf{X}_1'(\gamma\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\delta}) + \mathbf{X}_0'\boldsymbol{g}(\boldsymbol{\delta}) \\ -n_1\gamma^{-1} + \mathbf{y}_1'(\gamma\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\delta}) \end{bmatrix}.$$

Define $h(s) = g(s)(s + g(s))$. It is straightforward to show that $g'(s) = -g(s)(s + g(s)) = -h(s)$. As such, we can express the Hessian of $\ell_n(\boldsymbol{\theta})$ as

$$\begin{aligned}\nabla^2 \ell_n(\boldsymbol{\theta}) &= \frac{1}{n}\begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_0'\mathbf{D}(\boldsymbol{\delta})\mathbf{X}_0 & -\mathbf{X}_1'\mathbf{y}_1 \\ -\mathbf{y}_1'\mathbf{X}_1 & \mathbf{y}_1'\mathbf{y}_1 + n_1\gamma^{-2} \end{bmatrix} \\ &= \frac{1}{n}\begin{bmatrix} \mathbf{X}' \\ -\mathbf{y}' \end{bmatrix}\begin{bmatrix} \mathbf{D}(\boldsymbol{\delta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_1} \end{bmatrix}\begin{bmatrix} \mathbf{X} & -\mathbf{y} \end{bmatrix} + \frac{1}{n}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & n_1\gamma^{-2} \end{bmatrix}\end{aligned}$$

where $\mathbf{D}(\boldsymbol{\delta})$ is a $n_0 \times n_0$ diagonal matrix with $[\mathbf{D}(\boldsymbol{\delta})]_{ii} = h(-\mathbf{x}_i'\boldsymbol{\delta})$ for $i = 1, \ldots, n_0$ and $\mathbf{I}_q$ denotes a $q \times q$ identity matrix. We define

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{X}_{(\mathcal{M}\cup\mathcal{S})}' \\ -\mathbf{y}' \end{bmatrix}\begin{bmatrix} \mathbf{D}(\boldsymbol{\delta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_1} \end{bmatrix}\begin{bmatrix} \mathbf{X}_{(\mathcal{M}\cup\mathcal{S})} & -\mathbf{y} \end{bmatrix}$$

and assume the following additional conditions are satisfied:

(A3) $\exists c_H > 0$ such that $\inf_{\boldsymbol{\theta}\in\mathcal{N}_0} \lambda_{\min}\{\mathrm{E}[\mathbf{H}(\boldsymbol{\theta})]\} \geq nc_H$ for all n,
$\lambda_{\max}\{\mathrm{E}[\mathbf{H}(\boldsymbol{\theta}^*)]\} = O(n)$,
$\mathrm{E}\left[\lambda_{\max}^2\left\{\begin{bmatrix} \mathbf{X}_{(\mathcal{M}\cup\mathcal{S})}' \\ -\mathbf{y}' \end{bmatrix}\begin{bmatrix} \mathbf{X}_{(\mathcal{M}\cup\mathcal{S})} & -\mathbf{y} \end{bmatrix}\right\}\right] = O(n^2)$,
$\max_j \lambda_{\max}\left\{\mathbf{X}_{(\mathcal{M}\cup\mathcal{S})}' \operatorname{diag}\{|\mathbf{X}_{(j)}|\}\mathbf{X}_{(\mathcal{M}\cup\mathcal{S})}\right\} = O(n)$, and
$\left\|\mathrm{E}\left[\mathbf{X}_{((\mathcal{M}\cup\mathcal{S})^c)}'\begin{bmatrix} \mathbf{D}(\boldsymbol{\delta}^*) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_1} \end{bmatrix}\begin{bmatrix} \mathbf{X}_{(\mathcal{M}\cup\mathcal{S})} & -\mathbf{y} \end{bmatrix}\right]\right\|_\infty = O(n)$;

(A4) $\max_j \left\|\mathbf{X}_{(j)}\right\|_2 = O(\sqrt{n})$, $\max_{j,k}\sum_{i=1}^n x_{ij}^2 x_{ik}^2 = O(n)$, $\sum_{i=1}^n (\mathbf{x}_i'\boldsymbol{\delta}^*)^2 =$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

$O(n)$, $\max_j \sum_{i=1}^n x_{ij}^2\{2+\mathbf{x}_i'\boldsymbol{\delta}^*+g(-\mathbf{x}_i'\boldsymbol{\delta}^*)\}^2=O(n)$,

$\sum_{i=1}^n \frac{1}{2}(\mathbf{x}_i'\boldsymbol{\delta}^*)^2\{2+\mathbf{x}_i'\boldsymbol{\delta}^*+g(-\mathbf{x}_i'\boldsymbol{\delta}^*)\}^2=O(n)$, $\max_{i,j}|x_{ij}|=O(1)$,

and $\max_i |\mathbf{x}_i\boldsymbol{\delta}^*|=O(1)$, where $j,k\in\{0,\ldots,p\}$; and

(A5) $\log(p)=O(n^\eta)$ for some $\eta\in(0,1)$,

3.3.3 Asymptotic results for partial penalized Tobit estimators

Let $\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}=\mathrm{E}\,[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)]$, $\widetilde{\mathbf{C}}=\begin{bmatrix}\mathbf{C} & \mathbf{0} & -\boldsymbol{t}\end{bmatrix}$, and $\boldsymbol{\Psi}=\widetilde{\mathbf{C}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'$.

Theorem 1. *Suppose that* (A1) - (A5) *hold and that* $(s+m)^3\log(s+m)=o(n)$. *Then there exist local solutions* $\hat{\boldsymbol{\theta}}_0$ *and* $\hat{\boldsymbol{\theta}}_a$ *to* (3.3) *and* (3.4) *satisfying* $\hat{\boldsymbol{\theta}}_{0,(\mathcal{M}'\cup\mathcal{S})^c}=\hat{\boldsymbol{\theta}}_{a,(\mathcal{M}'\cup\mathcal{S})^c}=\mathbf{0}$ *with probability converging to* 1 *as* $n\to\infty$,
$\left\|\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}}-\boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2=O_p(\sqrt{(s+m-r+1)/n})$, *and*
$\left\|\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}}-\boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2=O_p(\sqrt{(s+m+1)/n})$.
Additionally,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}}-\boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}})=\frac{1}{\sqrt{n}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*)+o_p(1), \tag{3.5}$$

and

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}}-\boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}})=&\frac{1}{\sqrt{n}}\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}(\mathbf{I}_{s+m+1}-\mathbf{P}_n)\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*)\\&-\sqrt{n}\gamma^*\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\mathbf{P}_n\boldsymbol{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}}\begin{bmatrix}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\boldsymbol{h}_n\\ \mathbf{0}\end{bmatrix}+o_p(1)\end{aligned} \tag{3.6}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

where $\mathbf{P}_n = \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2}\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\widetilde{\mathbf{C}}\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2}$, *a projection matrix.*

Note that $\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}$ is positive definite by (A3), making the expression for $\mathbf{P}_n$ valid. Theorem 1 establishes that there exist $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_a$ that are estimation consistent and selection consistent. In addition, (3.5) and (3.6) provide limiting expressions for the partial penalized Tobit estimators with $\sqrt{n}$-scaling. We will leverage these properties in deriving approximate distributions for our test statistics in the next section.

4. Partial Penalized Tobit Tests

4.1 Partial penalized Tobit test statistics

We use the partial penalized Tobit estimators, $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_a$, to develop testing procedures for high-dimensional Tobit regression based on the Wald test, the score test, and the likelihood ratio test. To reflect our use of the partial penalized negative Tobit log-likelihood as our objective for $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_a$ in (3.3) and (3.4), we refer to our testing procedures as the *partial penalized Wald test*, the *partial penalized score test*, and the *partial penalized likelihood ratio test*.

We define the partial penalized Wald test statistic based on $\sqrt{n}\mathbf{C}^*\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'}$. As an immediate consequence of (3.5) in Theorem 1, we see that $\sqrt{n}\mathbf{C}^*\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'}$ has asymptotic variance $\mathbf{C}^*\left[\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\right]_{\mathcal{M}',\mathcal{M}'}\mathbf{C}^{*\prime}$. Let $\hat{\mathcal{S}}_a = \{j \in \mathcal{M}^c : \hat{\boldsymbol{\delta}}_{a,j} \neq$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

$0\}$ denote the sample version of $\mathcal{S}$ for the full model. Theorem 1 establishes that $\hat{\mathcal{S}}_a = \mathcal{S}$ with probability converging to 1 as $n \to \infty$. As such, we define the *partial penalized Wald test statistic* by

$$T_W := (\mathbf{C}^* \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'})' \left(\mathbf{C}^* \left[\{ -\nabla^2_{\mathcal{M}' \cup \hat{\mathcal{S}}_a} \log L_n(\hat{\boldsymbol{\theta}}_a) \}^{-1} \right]_{\mathcal{M}',\mathcal{M}'} \mathbf{C}^{*\prime} \right)^{-1} \mathbf{C}^* \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'}.$$

Let $\hat{\mathcal{S}}_0 = \{ j \in \mathcal{M}^c : \hat{\boldsymbol{\delta}}_{0,j} \neq 0 \}$ denote the sample version of $\mathcal{S}$ for the reduced model. We define the *partial penalized score test statistic* by

$$T_S := \left\{ \nabla_{\mathcal{M}' \cup \hat{\mathcal{S}}_0} \log L_n(\hat{\boldsymbol{\theta}}_0) \right\}' \left\{ -\nabla^2_{\mathcal{M}' \cup \hat{\mathcal{S}}_0} \log L_n(\hat{\boldsymbol{\theta}}_0) \right\}^{-1} \nabla_{\mathcal{M}' \cup \hat{\mathcal{S}}_0} \log L_n(\hat{\boldsymbol{\theta}}_0).$$

Lastly, we define the *partial penalized likelihood ratio test statistic* by

$$T_L := 2\{ \log L_n(\hat{\boldsymbol{\theta}}_a) - \log L_n(\hat{\boldsymbol{\theta}}_0) \}.$$

The partial penalized Tobit test statistics are analogous to their low-dimensional counterparts. The key difference is that the estimators used to compute them minimize the partial penalized negative log-likelihood rather than the unpenalized negative log-likelihood.

4.2 Testing procedure

Let $T \in \{T_W, T_S, T_L\}$ be any one of the partial penalized Tobit test statistics. For a given significance level $\alpha \in (0,1)$, we reject $\mathrm{H}_0 : \mathbf{C}^* \boldsymbol{\theta}^*_{\mathcal{M}'} = \mathbf{0}$ when $T > \chi^2_\alpha(r)$, where $\chi^2_\alpha(r)$ denotes the upper-α quantile of a χ^2 distribution

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

with r degrees of freedom.

4.3 Approximate distributions of the test statistics

The following theorem supports our choice of $T > \chi^2_\alpha(r)$ as the rejection rule for the partial penalized Tobit hypothesis tests.

Theorem 2. *Suppose that* (A1) - (A5) *hold and that* $(s+m)^3\log(s+m) = o(n)$*. Then* T_W*,* T_S*, and* T_L *evaluated at the partial penalized estimators* $\hat{\boldsymbol{\theta}}_0$ *and* $\hat{\boldsymbol{\theta}}_a$ *in Theorem 1 satisfy*

$$\sup_x \left|P(T \le x) - P(\chi^2(r,\nu_n) \le x)\right| \to 0 \tag{4.1}$$

as $n \to \infty$ *for* $T = T_W$*,* T_S*, or* T_L*, where* $\chi^2(r,\nu_n)$ *is a noncentral chi-square random variable with* r *degrees of freedom and noncentrality parameter* $\nu_n = n\gamma^{*2}\boldsymbol{h}_n'\boldsymbol{\Psi}^{-1}\boldsymbol{h}_n$*.*

It is important to note Theorem 2 does not state that our partial penalized test statistics converge in distribution to $\chi^2(r,\nu_n)$ random variables. Because r can diverge with n the notion of convergence in distribution is not well-defined in this setting. Instead, Theorem 2 provides that for $T \in \{T_W, T_S, T_L\}$ and any $x \in \mathbb{R}$ the difference between $P(T \le x)$ and $P(\chi^2(r,\nu_n) \le x)$ converges to 0 and, as such, the distribution of T is well-approximated by a $\chi^2(r,\nu_n)$ distribution for large n.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

Theorem 2 leads to a few immediate corollaries. Under H_0, $\boldsymbol{h}_n = \mathbf{0}$ and, by extension, $\nu_n = 0$. As such, Theorem 2 implies that under the null hypothesis $\lim_n P(T > \chi^2_\alpha(r)) = \alpha$ for $T = T_W$, T_S, or T_L—that is, the partial penalized tests asymptotically achieve their nominal size. Theorem 2 further establishes that the partial penalized tests have approximate power $P(\chi^2(r, \nu_n) > \chi^2_\alpha(r))$ for large n under the alternatives specified in (A1). Under these alternatives, $\boldsymbol{h}_n \neq \mathbf{0}$ and, by extension, $\nu_n \geq 0$. Since $\chi^2(r, \nu_n)$ is stochastically larger than $\chi^2(r)$ this implies that the partial penalized tests have approximate power at least α for large n. In addition, Theorem 2 establishes that for $T_1, T_2 \in \{T_W, T_S, T_L\}$,

$$\sup_x |P(T_1 \leq x) - P(T_2 \leq x)| \to 0$$

as $n \to \infty$. That is, T_W, T_S, and T_L are approximately equivalent. We sum up these findings in the following corollary.

Corollary 1. *Suppose that* (A1) - (A5) *hold and that* $(s+m)^3 \log(s+m) = o(n)$. *Then* T_W, T_S, *and* T_L *evaluated at the partial penalized estimators* $\hat{\boldsymbol{\theta}}_0$ *and* $\hat{\boldsymbol{\theta}}_a$ *in Theorem 1 satisfy the following:*

- *Under* H_0, *for any significance level* $0 < \alpha < 1$,

$$\lim_{n \to \infty} P(T > \chi^2_\alpha(r)) = \alpha$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

for $T = T_W$, T_S, *or* T_L, *where* $\chi^2_\alpha(r)$ *denotes the upper-*α *quantile of a* χ^2 *distribution with* r *degrees of freedom.*

- *Under the alternative* $\mathbf{C}\boldsymbol{\beta}^*_{\mathcal{M}} = \boldsymbol{t} + \boldsymbol{h}_n$, *where* $\|\boldsymbol{h}_n\|_2 = O(\sqrt{\min\{s+m-r+1, r\}/n})$, *for any significance level* $0 < \alpha < 1$,

$$\lim_{n\to\infty} \left| P(T > \chi^2_\alpha(r)) - P(\chi^2(r, \nu_n) > \chi^2_\alpha(r)) \right| = 0$$

for $T = T_W$, T_S, *or* T_L, *where* $\nu_n = n\gamma^{*2}\boldsymbol{h}_n'\boldsymbol{\Psi}^{-1}\boldsymbol{h}_n$.

- *For* $T_1, T_2 \in \{T_W, T_S, T_L\}$,

$$\sup_x |P(T_1 \le x) - P(T_2 \le x)| \to 0 \quad as \quad n \to \infty.$$

5. Implementation Details

5.1 Computing the partial penalized Tobit estimators

We develop algorithms based on the alternating direction method of multipliers (ADMM) to compute the partial penalized Tobit estimators. Here we focus on computing the reduced model estimator $\hat{\boldsymbol{\theta}}_0$, though the same approach can be used to compute $\hat{\boldsymbol{\theta}}_a$. For fixed $\lambda > 0$, we can express the constrained optimization problem (3.3) as

$$\hat{\boldsymbol{\theta}}_0^\lambda = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^{p+2}} \ \ell_n(\boldsymbol{\theta}) + \sum_{j\in\mathcal{M}^c\setminus\{0\}} p_\lambda(|\delta_j|) \quad \text{subject to } \mathbf{C}\boldsymbol{\delta}_{\mathcal{M}} = \gamma\boldsymbol{t}.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

By introducing dummy variables $\boldsymbol{\eta} \in \mathbb{R}^{p-m}$, we can rewrite this problem as

$$(\hat{\boldsymbol{\theta}}_0^\lambda, \hat{\boldsymbol{\eta}}_0^\lambda) = \underset{\substack{\boldsymbol{\theta}\in\mathbb{R}^{p+2} \\ \boldsymbol{\eta}\in\mathbb{R}^{p-m}}}{\arg\min}\ \ell_n(\boldsymbol{\theta}) + \sum_{j=1}^{p-m} p_\lambda(|\eta_j|) \quad \text{subject to } \mathbf{C}\boldsymbol{\delta}_{\mathcal{M}} = \gamma \boldsymbol{t},\ \boldsymbol{\delta}_{\mathcal{M}^c} = \boldsymbol{\eta}, \tag{5.1}$$

separating the objective into a component depending only on $\boldsymbol{\theta}$ and a component depending only on $\boldsymbol{\eta}$. The augmented Lagrangian is

$$\begin{aligned} L_\rho(\boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\nu}) = \ell_n(\boldsymbol{\theta}) &+ \sum_{j=1}^{p-m} p_\lambda(|\eta_j|) \\ &+ \frac{\rho}{2}\left\|\mathbf{C}\boldsymbol{\delta}_{\mathcal{M}} - \gamma\boldsymbol{t} + \frac{\boldsymbol{\nu}_1}{\rho}\right\|_2^2 + \frac{\rho}{2}\left\|\boldsymbol{\delta}_{\mathcal{M}^c} - \boldsymbol{\eta} + \frac{\boldsymbol{\nu}_2}{\rho}\right\|_2^2 \end{aligned} \tag{5.2}$$

with Lagrangian penalty parameter $\rho > 0$ and dual variables $\boldsymbol{\nu}_1 \in \mathbb{R}^r$, $\boldsymbol{\nu}_2 \in \mathbb{R}^{p-m}$, and $\boldsymbol{\nu} = (\boldsymbol{\nu}_1', \boldsymbol{\nu}_2')'$.

Using the scaled form of the augmented Lagrangian (5.2), we develop the ADMM algorithm given in Algorithm 1 to solve (5.1). We update $\boldsymbol{\theta}$ using a Newton-Raphson algorithm. We use the SCAD penalty in our implementation, both because it is a folded-concave penalty and because using it gives us a closed-form solution for the $\boldsymbol{\eta}$ updates.

5.2 Selecting the penalty parameter

We set $\lambda_{\max}$ to be the smallest λ such that $\delta_j = 0$ for all $j \in \mathcal{M}^c$ based on the Karush-Kuhn-Tucker conditions. We then set $\lambda_{\min} = c \cdot \lambda_{\max}$ where c is

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

Algorithm 1: ADMM algorithm for $\hat{\boldsymbol{\theta}}_0$

Initialize $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\eta}^{(0)}, \boldsymbol{\nu}^{(0)})$;

repeat

Update $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}^{(k+1)} = \arg\min_{\boldsymbol{\theta}} \left\{ \ell_n(\boldsymbol{\theta}) + \frac{\rho}{2} \left\| \mathbf{C}\boldsymbol{\delta}_{\mathcal{M}} - \gamma \boldsymbol{t} + \frac{\boldsymbol{\nu}_1^{(k)}}{\rho} \right\|_2^2 + \frac{\rho}{2} \left\| \boldsymbol{\delta}_{\mathcal{M}^c} - \boldsymbol{\eta}^{(k)} + \frac{\boldsymbol{\nu}_2^{(k)}}{\rho} \right\|_2^2 \right\};$$

Update $\boldsymbol{\eta}$:

$$\boldsymbol{\eta}^{(k+1)} = \arg\min_{\boldsymbol{\eta}} \left\{ \sum_{j=1}^{p-m} p_\lambda(|\eta_j|) + \frac{\rho}{2} \left\| \boldsymbol{\delta}_{\mathcal{M}^c}^{(k+1)} - \boldsymbol{\eta} + \frac{\boldsymbol{\nu}_2^{(k)}}{\rho} \right\|_2^2 \right\};$$

Dual update: $\boldsymbol{\nu}^{(k+1)} = \boldsymbol{\nu}^{(k)} + \rho \begin{pmatrix} \mathbf{C}\boldsymbol{\delta}_{\mathcal{M}}^{(k+1)} - \gamma^{(k+1)}\boldsymbol{t} \\ \boldsymbol{\delta}_{\mathcal{M}^c}^{(k+1)} - \boldsymbol{\eta}^{(k+1)} \end{pmatrix}$;

$k = k + 1$;

until *primal and dual residuals are sufficiently small*;

a small constant. We compute $\hat{\boldsymbol{\theta}}_0^\lambda$ along a path of λ values which are evenly spaced on the log-scale between $\lambda_{\min}$ and $\lambda_{\max}$. To speed up computation, we warm start the ADMM algorithm for computing $\hat{\boldsymbol{\theta}}^{\lambda_k}$ with the computed solution for the previous penalty parameter value λ_{k-1}, setting $\boldsymbol{\theta}^{(0)} = \hat{\boldsymbol{\theta}}^{\lambda_{k-1}}$, $\boldsymbol{\eta}^{(0)} = \hat{\boldsymbol{\eta}}^{\lambda_{k-1}}$, and $\boldsymbol{\nu}^{(0)} = \hat{\boldsymbol{\nu}}^{\lambda_{k-1}}$ in Algorithm 1.

We select $\hat{\lambda}$ based on the following information criterion:

$$\hat{\lambda} = \arg\min_{\lambda} \left\{ n\, \ell_n(\hat{\boldsymbol{\theta}}_0^\lambda) + c_n \left\| \hat{\boldsymbol{\theta}}_0^\lambda \right\|_0 \right\}, \tag{5.3}$$

where $c_n = \max\{\log n, \log(\log n) \log p\}$, and use $\hat{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_0^{\hat{\lambda}}$ as our final reduced model estimator. Our choice of c_n is motivated by selection consistency guarantees for BIC (Schwarz, 1978) and GIC (Fan and Tang, 2013) in the fixed-p and ultra-high dimensional settings, respectively. We take

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

the maximum of the BIC and GIC penalties to cover both settings. While Schwarz (1978) and Fan and Tang (2013) only examine GLMs, similar arguments could be used to extend their guarantees to the Tobit model.

6. Simulation Study

In the following simulation study, we examine the finite sample performance of the partial penalized Tobit Wald, score, and likelihood ratio tests in a variety of settings. In this section, we run the partial penalized tests with significance level $\alpha = 0.05$ and report their estimated rejection probabilities. Throughout our empirical study, we set $a = 3.7$ for the SCAD penalty and set $\rho = 1$ in the ADMM algorithm.

We examine 24 different simulation settings. For each simulation setting, we generate 600 datasets with 200 observations each. We generate the censored response y_i by first generating an uncensored response from a linear model $y_i^* = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$, where $\mathbf{x}_i \sim N(0, \Sigma)$ and $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, then setting $y_i = \max\{y_i^*, 0\}$ for $i = 1, \ldots, 200$. We generate these data with $\beta_0 = 1$ and $\boldsymbol{\beta} = (2, -2 - h_1, \mathbf{0}_{p-2})$, varying h_1 to create different test cases.

We vary Σ, p, and h_1 to create our different simulation settings. We examine settings where (1) $\Sigma = \mathbf{I}_p$ and (2) $\Sigma_{ij} = 0.5^{|i-j|}$ for all i, j (we

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

refer to these as these independent and AR1(0.5) settings, respectively). For each covariance structure, we run simulations with every combination of $p \in \{50, 250, 400\}$ and $h_1 \in \{0, 0.1, 0.2, 0.4\}$. We set $\sigma = 1$ across all simulation settings.

We test the following hypotheses at significance level $\alpha = 0.05$ in each simulation setting:

- $\mathrm{H}_0^{(1)} : \beta_1 + \beta_2 = 0$
- $\mathrm{H}_0^{(2)} : \beta_2 = -2$
- $\mathrm{H}_0^{(3)} : \beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$
- $\mathrm{H}_0^{(4)} : \beta_1 + \beta_2 = 0,\ \beta_2 = -2$, and $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$.

When $h_1 = 0$, each of these null hypotheses is true. Based on our theoretical results, we would expect the rejection probabilities for the tests to be close to their nominal $\alpha = 0.05$ significance level in these cases. As h_1 increases, the alternative gets farther from the null and we would expect that rejection probabilities for the tests to increase. We include $\mathrm{H}_0^{(4)}$, a test of multiple hypotheses, to illustrate the flexibility of our testing procedures.

Table 1 shows the estimated rejection probabilities for the partial penalized Wald, score, and likelihood ratio tests in simulations where $\Sigma = \mathbf{I}_p$. These estimates are based on 600 replications, with standard errors given

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

in parentheses. Focusing first on simulations where the null is true (i.e. $h_1 = 0$), we see that the partial penalized Tobit tests all achieve estimated rejection probabilities near the nominal significance level of $\alpha = 0.05$ for all four null hypotheses. These results are consistent with the large-sample guarantees given in Corollary 1, which provide that the tests' rejection rates will converge to α. As h_1 increases and the data generating model moves further from satisfying the null hypothesis, we see that the rejection probabilities for all three partial penalized Tobit tests rapidly increase. Notably, the estimated rejection probabilities are similar across simulations with $p = 50$, $p = 250$, and $p = 400$ for each null $\mathrm{H}_0^{(i)}$ and each value of h_1. This suggests that the partial penalized Tobit tests are not adversely affected by p growing as n remains fixed. Moreover, the estimated rejection probabilities for all three tests are close to each other in every simulation setting. Comparing results across the null hypotheses being tested, we see that the estimated rejection probabilities for the tests of $\mathrm{H}_0^{(3)}$ are lower than for the other hypotheses. This suggests that increasing the number of coefficients being tested decreases the power of the partial penalized tests.

Table 2 shows the estimated rejection probabilities for the partial penalized Tobit tests in simulations where $\Sigma_{ij} = 0.5^{|i-j|}$ for all i, j. The results in these cases are consistent with the results from the simulations with $\Sigma = \mathbf{I}_p$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

As in those simulations, we see that (i) when H_0 is true the estimated rejection probabilities for the tests are all near the nominal level of $\alpha = 0.05$, (ii) as h_1 increases so does the estimated power of each test, and (iii) the estimated rejection probabilities for the tests do not appreciably change as p increases. The clearest difference in this second set of simulations is that, in cases where $h_1 > 0$, the estimated rejection probabilities for the tests of $H_0^{(1)}$ and $H_0^{(3)}$ are markedly higher than in the simulations with independent predictors.

Supplementary simulations

We present additional simulation results in Sections S.3 and S.4 of the supplementary material. In Section S.3, we examine the empirical distributions of the p-values for the partial penalized Tobit tests in simulations where the null hypothesis is true. These additional results provide evidence that the test statistics approximately follow a $\chi^2(r)$ distribution under the null, as suggested by Corollary 1. In Section S.4, we conduct additional studies examining the effects of the sample size n and the correlation coefficient ρ in the AR1(ρ) correlation structure for the predictors on the power of the partial penalized Tobit tests.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

Table 1: Estimated rejection probabilities when $\Sigma = \mathbf{I}_p$

	p = 50			p = 250			p = 400		
	LRT	Wald	Score	LRT	Wald	Score	LRT	Wald	Score
h_1					$H_0^{(1)}$				
0	5.33 (0.92)	5.17 (0.9)	5.17 (0.9)	4.67 (0.86)	4.67 (0.86)	4.67 (0.86)	4.67 (0.86)	4.67 (0.86)	4.67 (0.86)
0.1	15.17 (1.46)	14.83 (1.45)	14.83 (1.45)	15 (1.46)	15 (1.46)	15 (1.46)	13.67 (1.4)	13.67 (1.4)	13.67 (1.4)
0.2	39.5 (2)	39.83 (2)	39.83 (2)	42.83 (2.02)	42.83 (2.02)	42.83 (2.02)	38.5 (1.99)	38.5 (1.99)	38.33 (1.98)
0.4	89.5 (1.25)	89.5 (1.25)	89.5 (1.25)	92.17 (1.1)	92 (1.11)	92 (1.11)	91.67 (1.13)	91.5 (1.14)	91.5 (1.14)
h_1					$H_0^{(2)}$				
0	5.67 (0.94)	5.33 (0.92)	5.17 (0.9)	5.5 (0.93)	5.67 (0.94)	5.83 (0.96)	5.5 (0.93)	5.5 (0.93)	5.5 (0.93)
0.1	17 (1.53)	17 (1.53)	16.83 (1.53)	17.5 (1.55)	17.17 (1.54)	17 (1.53)	18 (1.57)	17.67 (1.56)	17.83 (1.56)
0.2	49 (2.04)	49 (2.04)	49.17 (2.04)	54.83 (2.03)	54.5 (2.03)	54.5 (2.03)	53 (2.04)	52.5 (2.04)	52.67 (2.04)
0.4	97.83 (0.59)	97.67 (0.62)	97.67 (0.62)	97.83 (0.59)	97.83 (0.59)	97.83 (0.59)	97.83 (0.59)	97.83 (0.59)	97.83 (0.59)
h_1					$H_0^{(3)}$				
0	3.5 (0.75)	3.5 (0.75)	3.5 (0.75)	5.17 (0.9)	5.17 (0.9)	5.17 (0.9)	7 (1.04)	7 (1.04)	7 (1.04)
0.1	10 (1.22)	9.83 (1.22)	9.67 (1.21)	10 (1.22)	10 (1.22)	10 (1.22)	8.33 (1.13)	8.33 (1.13)	8.33 (1.13)
0.2	22.17 (1.7)	22.33 (1.7)	22.33 (1.7)	21.5 (1.68)	21.5 (1.68)	21.5 (1.68)	24.33 (1.75)	24.33 (1.75)	24.33 (1.75)
0.4	65.17 (1.95)	65.33 (1.94)	65.33 (1.94)	67 (1.92)	67 (1.92)	67 (1.92)	66.17 (1.93)	65.83 (1.94)	65.67 (1.94)
h_1					$H_0^{(4)}$				
0	4.17 (0.82)	4.67 (0.86)	4.5 (0.85)	5.83 (0.96)	5.67 (0.94)	5.67 (0.94)	5.5 (0.93)	5.17 (0.9)	5.17 (0.9)
0.1	15.67 (1.48)	15 (1.46)	14.83 (1.45)	12.33 (1.34)	12.17 (1.33)	12 (1.33)	12.67 (1.36)	12 (1.33)	12 (1.33)
0.2	43.83 (2.03)	43 (2.02)	43 (2.02)	41.33 (2.01)	40.33 (2)	40.17 (2)	37.5 (1.98)	36.33 (1.96)	36.33 (1.96)
0.4	95.67 (0.83)	95.67 (0.83)	95.5 (0.85)	94.67 (0.92)	94.33 (0.94)	94.33 (0.94)	97 (0.7)	96.67 (0.73)	96.67 (0.73)

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

Table 2: Estimated rejection probabilities when $\Sigma_{ij} = 0.5^{|i-j|}$ for all i, j

	p = 50			p = 250			p = 400		
	LRT	Wald	Score	LRT	Wald	Score	LRT	Wald	Score
h_1					$H_0^{(1)}$				
0	5.83 (0.96)	5.83 (0.96)	5.83 (0.96)	6 (0.97)	6 (0.97)	6.67 (1.02)	3.67 (0.77)	3.83 (0.78)	4 (0.8)
0.1	21.5 (1.68)	21.5 (1.68)	21.5 (1.68)	19 (1.6)	19.17 (1.61)	19 (1.6)	20.5 (1.65)	20.5 (1.65)	20.5 (1.65)
0.2	58 (2.01)	58 (2.01)	58 (2.01)	58 (2.01)	58 (2.01)	58 (2.01)	58.5 (2.01)	58.5 (2.01)	58.5 (2.01)
0.4	98 (0.57)	98.17 (0.55)	98.17 (0.55)	98.5 (0.5)	98.5 (0.5)	98.5 (0.5)	98.83 (0.44)	98.83 (0.44)	98.83 (0.44)
h_1					$H_0^{(2)}$				
0	6 (0.97)	6.17 (0.98)	6.33 (0.99)	6 (0.97)	5.83 (0.96)	5.83 (0.96)	6 (0.97)	5.83 (0.96)	5.83 (0.96)
0.1	15.5 (1.48)	15.83 (1.49)	15.67 (1.48)	13.83 (1.41)	14 (1.42)	13.83 (1.41)	14.67 (1.44)	14.33 (1.43)	14.5 (1.44)
0.2	44.67 (2.03)	44.67 (2.03)	44.33 (2.03)	46.67 (2.04)	46.17 (2.04)	45.5 (2.03)	42.17 (2.02)	42 (2.01)	42 (2.01)
0.4	95.5 (0.85)	95.83 (0.82)	95.83 (0.82)	96.17 (0.78)	96 (0.8)	96 (0.8)	95 (0.89)	95 (0.89)	95 (0.89)
h_1					$H_0^{(3)}$				
0	5.33 (0.92)	5.33 (0.92)	5.17 (0.9)	5.67 (0.94)	5.67 (0.94)	5.83 (0.96)	5.83 (0.96)	5.83 (0.96)	6.17 (0.98)
0.1	14.5 (1.44)	14.33 (1.43)	14.33 (1.43)	13.67 (1.4)	13.67 (1.4)	13.67 (1.4)	13.83 (1.41)	13.83 (1.41)	13.83 (1.41)
0.2	39.33 (1.99)	39.5 (2)	39.33 (1.99)	43.5 (2.02)	43.5 (2.02)	43.5 (2.02)	44 (2.03)	44 (2.03)	43.67 (2.02)
0.4	94.17 (0.96)	94.33 (0.94)	94.33 (0.94)	91.83 (1.12)	91.83 (1.12)	91.83 (1.12)	94.17 (0.96)	94.17 (0.96)	94.17 (0.96)
h_1					$H_0^{(4)}$				
0	4.33 (0.83)	4.17 (0.82)	4.33 (0.83)	6.83 (1.03)	6.33 (0.99)	6.33 (0.99)	5.17 (0.9)	5.17 (0.9)	5.17 (0.9)
0.1	13.5 (1.4)	13.5 (1.4)	13.5 (1.4)	14.83 (1.45)	14 (1.42)	14 (1.42)	15.33 (1.47)	14.83 (1.45)	14.83 (1.45)
0.2	46 (2.03)	45.5 (2.03)	45.5 (2.03)	48.33 (2.04)	47.83 (2.04)	47.83 (2.04)	49 (2.04)	47.83 (2.04)	47.83 (2.04)
0.4	97.17 (0.68)	97.33 (0.66)	97.33 (0.66)	99 (0.41)	99 (0.41)	99 (0.41)	97.33 (0.66)	97.17 (0.68)	97.17 (0.68)

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

7. HIV Drug Resistance Mutation Testing

Human immunodeficiency virus (HIV) can mutate rapidly in HIV-infected patients taking antiretroviral drugs. To manage the virus, physicians monitor patients' viral loads and conduct genotypic testing to check for known HIV drug resistance mutations (DRMs) (Shafer, 2002). Research identifying DRMs and quantifying their effects on HIV viral load is critical to supporting this ongoing therapy (Shafer, 2006). As noted in the introduction, two challenges arise in modeling the relationship between HIV viral load and mutations in the virus' genome: (i) the observed viral load is left-censored at a known detection threshold and (ii) the number of mutations in a study of HIV-infected patients typically far exceeds the number of participants in the study. Given these challenges, our partial penalized Tobit hypothesis tests are well-suited for conducting significance testing for potential DRMs using HIV viral load data.

We analyze data from the AIDS Clinical Trials Group's OPTIONS trial (Gandhi et al., 2020), which were obtained from the Stanford HIV Drug Resistance Database (Shafer, 2006). The participants recruited for the OPTIONS trial were HIV-infected individuals who had been taking protease inhibitor (PI)-based treatment and were experiencing virological failure. Researchers gave each participant an optimized drug regimen based

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

on their treatment history and randomly assigned participants with moderate drug resistance to either add nucleoside reverse transcriptase inhibitors (NRTIs) to their drug regimens or omit NRTIs from their drug regimens. Participants with highly drug-resistant HIV infections were all given drug regimens which included NRTIs.

We model HIV viral load at a 12 week follow-up appointment from the time of treatment assignment. Our predictors include indicator variables for protease (PR) and reverse transcriptase (RT) gene mutations in HIV, indicator variables for antiretroviral drugs in patients' treatment regimens, baseline viral load, observation week, and HIV subtype. Our sample includes $n = 407$ participants and $p = 1295$ total predictors. At 12 weeks, 35.6% of participants in this sample had viral loads which fell below the assays' detection threshold of 50 copies/mL and were, therefore, left-censored at 50 copies/mL. As in previous studies of HIV viral load, we log-transform the response so that it is approximately normally distributed (Soret et al., 2018; Jacobson and Zou, 2023).

The Stanford HIV Drug Resistance Database catalogs HIV mutations which have been identified as potential DRMs (Shafer, 2006). The following NRTI resistance mutations from their list are present in our sample: $\mathcal{M}_{\text{NRTI}} = \{$M41L, K65R, D67N, T69TN, T69TA, K70R, K70E, L74V, L74I,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

Y115F, Q151M, M184V, L210W, T215Y, T215F, K219E, K219Q}. We test the hypothesis $\boldsymbol{\beta}_{\mathcal{M}_{\mathrm{NRTI}}} = \mathbf{0}$ at significance level 0.05 using our partial penalized Tobit hypothesis tests. The p-values for the partial penalized likelihood ratio, Wald, and score tests of this hypothesis are $p_L = 1.562 \times 10^{-2}$, $p_W = 1.698 \times 10^{-2}$, and $p_S = 5.144 \times 10^{-5}$, respectively, all well below the significance level of 0.05. As such, we reject the null hypothesis that $\boldsymbol{\beta}_{\mathcal{M}_{\mathrm{NRTI}}} = \mathbf{0}$ and conclude that at least one of the RT mutations in $\mathcal{M}_{\mathrm{NRTI}}$ is a significant predictor of HIV viral load.

8. Concluding Remarks

In this paper we have developed hypothesis tests for high-dimensional left-censored regression based on the partial penalized negative Tobit log likelihood. Our partial penalized Tobit tests are designed for testing general linear hypotheses, making them more broadly applicable than other inferential procedures for high-dimensional left-censored data. We have derived approximate distributions for our partial penalized Tobit test statistics in an ultra high-dimensional setting in which the number of predictors, the number of constraints under the null hypothesis, and the number of coefficients being tested can all grow with the number of observations. In doing so, we have shown that our partial penalized Tobit tests are approximately

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

equivalent for large n and achieve their nominal size asymptotically and have derived their approximate power under local alternatives. In addition, we have shown how the proposed tests can be implemented via an ADMM algorithm. In our empirical study, we have presented strong evidence that our partial penalized Tobit tests achieve their nominal size under the null and are consistent in a finite-sample setting. Lastly, we used our tests to conduct significance testing in HIV viral load data. The code for the proposed tests will be made publicly available on the first author's GitHub site.

We see several possible avenues for future research. While our procedures can be used to test a broad range of hypotheses, the number of coefficients being tested has to be relatively small. Future studies could extend the score test of Chen, Li and Chen (2023) for high-dimensional left-censored regression to test $\mathrm{H}_0 : \boldsymbol{\beta}^*_{\mathcal{M}} = \boldsymbol{t}$ when $|\mathcal{M}|$ can grow at a faster rate. As another extension, one could adapt the general estimating equations approach of Chang et al. (2021) to construct confidence intervals for general transformations $S(\boldsymbol{\beta}_{\mathcal{M}})$ of the coefficients in the Tobit model.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

Supplementary Material

The online supplementary file contains technical proofs, intermediate theoretical results, and additional simulation results.

Acknowledgments

We thank the referees and the associate editor for their insightful comments, which greatly improved this article. This work is supported in part by NSF grants DMS 1915842 and DMS 2015120.

References

Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2011). Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends® in Machine Learning* **3**, 1–122.

Bradic, J. and Guo, J. (2019). Generalized M-estimators for High-dimensional Tobit I Models. *Electronic Journal of Statistics* **13**, 582–645.

Buckley, J. and James, I. (1979). Linear Regression with Censored Data. *Biometrika* **66**, 429–436.

Cai, T. T. and Guo, Z. (2017). Confidence Intervals for High-dimensional Linear Regression: Minimax Rates and Adaptivity. *Annals of Statistics* **45**, 615–646.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

Cai, T. T., Guo, Z. and Ma, R. (2021). Statistical Inference for High-dimensional Generalized Linear Models with Binary Outcomes. *Journal of the American Statistical Association* **116**, 1–14.

Cai, T. T., Guo, Z. and Xia, Y. (2023). Statistical Inference and Large-scale Multiple Testing for High-dimensional Regression Models. *arXiv preprint, https://arxiv.org/abs/2301.10392/* .

Chang, J., Chen, S. X., Tang, C. Y. and Wu, T. T. (2021). High-dimensional empirical likelihood inference. *Biometrika* **108**, 127–147.

Chen, J., Li, Q. and Chen, H. Y. (2023). Testing Generalized Linear Models with High-dimensional Nuisance Parameters. *Biometrika* **110**, 83–99.

Cui, H., Guo, W. and Zhong, W. (2018). Test for High-Dimensional Regression Coefficients Using Refitted Cross-Validation Variance Estimation. *The Annals of Statistics* **46**, 958–988.

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.

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

Fan, Y. and Tang, C. Y. (2013). Tuning Parameter Selection in High Dimensional Penalized

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

Likelihood. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **75**, 531–552.

Fang, E. X., Ning, Y. and Liu, H. (2017). Testing and Confidence Intervals for High Dimensional Proportional Hazards Models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **79**, 1415–1437.

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. (2023). High-dimensional Censored Regression via the Penalized Tobit Likelihood. *Journal of Business & Economic Statistics* .

Javanmard, A. and Montanari, A. (2014). Confidence Intervals and Hypothesis Testing for High-Dimensional Regression. *Journal of Machine Learning Research* **15**, 2869–2909.

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**, 251–268.

Ma, R., Cai, T. T. and Li, H. (2021). Global and Simultaneous Hypothesis Testing for High-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

Dimensional Logistic Regression Models. *Journal of the American Statistical Association* **116**, 984–998.

Müller, P. and Van de Geer, S. (2016). Censored Linear Model in High Dimensions. *TEST* **25**, 75–92.

Ning, Y. and Liu, H. (2017). A General Theory of Hypothesis Tests and Confidence Regions for Sparse High Dimensional Models. *Annals of Statistics* **45**, 158–195.

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

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

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics* **6**, 461–464.

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

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

Shi, C., Song, R., Chen, Z. and Li, R. (2019). Linear Hypothesis Testing for High Dimensional Generalized Linear Models. *Annals of Statistics* **47**, 2671–2703.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

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**, 1–13.

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

Van de Geer, S., Bühlmann, P., Ritov, Y. and Dezeure, R. (2014). On Asymptotically Optimal Confidence Regions and Tests for High-Dimensional Models. *Annals of Statistics* **42**, 1166–1202.

Wang, S. and Cui, H. (2014). Partial Penalized Likelihood Ratio Test under Sparse Case. *arXiv preprint, https://arxiv.org/abs/1312.3723/* .

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

Zhang, C.-H. and Zhang, S. (2014). Confidence Intervals for Low Dimensional Parameters in High Dimensional Linear Models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **76**, 217–242.

Zhang, X. and Cheng, G. (2017). Simultaneous Inference for High-Dimensional Linear Models. *Journal of the American Statistical Association* **112**, 757–768.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0118

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.

Tate Jacobson

Department of Statistics, Oregon State University, Corvallis, OR, 97331, U.S.A.

E-mail: tate.jacobson@oregonstate.edu

Hui Zou

School of Statistics, University of Minnesota, Minneapolis, MN, 55455, U.S.A.

E-mail: zouxx019@umn.edu

Statistica Sinica