The Proximal Bootstrap for Finite-Dimensional Regularized Estimators *

Jessie Li

January 29, 2021

We propose a proximal bootstrap that can consistently estimate the limiting distribution of $\sqrt{n}$ consistent estimators with nonstandard asymptotic distributions in a computationally efficient manner by formulating the proximal bootstrap estimator as the solution to a convex optimization problem, which can have a closed form solution for certain designs. This paper considers the application to finite-dimensional regularized estimators, such as the Lasso, $\ell_1$ norm regularized quantile regression, $\ell_1$ norm support vector regression, and trace regression via nuclear norm regularization.

Keywords: bootstrap, convex optimization, proximal mapping

I Introduction

We propose a computationally efficient bootstrap procedure to conduct pointwise asymptotically valid inference for a large class of $\sqrt{n}$-consistent estimators with nonstandard asymptotic distributions for which standard bootstrap procedures are known to be inconsistent. The application we consider in this paper is finite-dimensional regularized estimators, such as the Lasso (Tibshirani (1996)), $\ell_1$ norm regularized quantile regression (Belloni and Chernozhukov (2011)), $\ell_1$ norm support vector regression (Zhu et al. (2004), Bai et al. (2019)), and trace regression via nuclear norm regularization (Koltchinskii et al. (2011), Moon and Weidner (2018)). Another application that will be explored in a subsequent paper is to constrained optimization problems with a possibly non-smooth and non-convex objective function and a finite number of either estimated or fixed inequality and/or equality constraints, and where the true parameter can lie on the boundary of the constraint set (Andrews (1999), Andrews (2000), Andrews (2002a)).

Motivated by the optimization literature and recent contributions in computationally efficient bootstrap procedures (e.g. Forneron and Ng (2019)), our proximal bootstrap estimator can be expressed as the solution to a convex optimization problem and efficiently computed starting from an initial $\sqrt{n}$-consistent estimator using built-in and freely available software. Additionally, when the sample hessian is proportional to the identity matrix, the proximal bootstrap has a closed form solution. In the case of a smooth sample objective function and no regularization, the proximal bootstrap is very similar to the k-step bootstrap (for $k = 1$) proposed by Davidson and MacKinnon (1999) and investigated further by Andrews (2002b).

* I would like to thank Alex Torgovitsky and the participants of the UChicago Econometrics workshop and the 2011 AEA session (Optimization-conscious Econometrics) for helpful comments and suggestions.

Department of Economics, University of California, Santa Cruz. Email: jeqli@ucsc.edu
The consistency of the proximal bootstrap relies on a scaling sequence (labeled \( \alpha_n \) in this paper) that converges to zero at a slower than \( \sqrt{n} \) rate. The purpose of the slower than \( \sqrt{n} \) rate is to offset the estimation error from the initial \( \sqrt{n} \)-consistent estimator. The purpose of \( \alpha_n \) is similar to that of \( \epsilon_n \) in the numerical bootstrap Hong and Li (2020). However, we want to emphasize that the proximal bootstrap is a different procedure than the numerical bootstrap because it solves a different optimization problem. The proximal bootstrap works only for \( \sqrt{n} \)-consistent estimators but is typically more computationally efficient than the numerical bootstrap.

Section II reviews the concept of proximal mappings from the optimization literature. Section III contains all of the theoretical results demonstrating consistency of the proximal bootstrap for finite-dimensional regularized estimators. Section IV concludes. The online appendix contains the proof of consistency, provides the specific form of the proximal bootstrap estimator for the Lasso, \( \ell_1 \) norm support vector regression (of which \( \ell_1 \) norm regularized quantile regression is a special case), and trace regression via nuclear norm regularization, and also contains a Monte Carlo simulation for the Lasso.

II Proximal Mappings

Given an Euclidean space \( D \) and a function \( r : D \mapsto \mathbb{R} \), the proximal mapping of \( r \) is the operator given by

\[
prox_r(z) = \arg \min_{\beta \in D} \left\{ r(\beta) + \frac{1}{2} \| \beta - z \|_2^2 \right\}
\]

for any \( z \in D \).

Given a function \( r : D \mapsto \mathbb{R} \) and a symmetric positive definite matrix \( H \), the scaled proximal mapping of \( r \) is the operator given by, for \( \| \beta - z \|_H^2 = (\beta - z)' H (\beta - z) \),

\[
prox_{H,r}(z) = \arg \min_{\beta \in D} \left\{ r(\beta) + \frac{1}{2} \| \beta - z \|_H^2 \right\}
\]

for any \( z \in D \).

When \( r \) is a proper closed and convex function then \( prox_r(z) \) is a singleton for any \( z \in D \) (Theorem 6.3 Beck (2017)). The same can be said for \( prox_{H,r}(z) \) (Lee et al. (2014)).

The proximal map often has a closed form solution. For instance, the proximal mapping of the \( \ell_1 \) norm is given by

\[
prox_{\lambda\|\|\|_1}(z) = \arg \min_{\beta} \left\{ \lambda \| \beta \|_1 + \frac{1}{2} \| \beta - z \|_2^2 \right\}
\]

\[
= \text{sign}(z) \max \{|z| - \lambda, 0\}
\]

\[
= (z - \lambda)^+ - (z + \lambda)^-
\]

where \( x^+ \equiv \max(x, 0) \) and \( x^- \equiv -\min(x, 0) \).

Although it is rarely the case that the scaled proximal map has a closed form solution, it can still be efficiently computed as the solution to a convex optimization problem if \( r \) is convex. Additionally, Friedlander and Goh (2017) show that for certain \( r \) that have a 'Quadratic Support' representation (which is satisfied for many functions such as the \( \ell_1 \) norm, the \( \ell_2 \) norm, and indicators on polyhedral cones, the scaled proximal map can be written as a quadratic optimization problem over conic constraints.
III Proximal Bootstrap

III.a Notation

Consider a random sample $X_1, X_2, \ldots, X_n$ of independent draws from a probability measure $P$ on a sample space $\mathcal{X}$. Define the empirical measure $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$, where $\delta_x$ is the measure that assigns mass 1 at $x$ and zero everywhere else. Denote the bootstrap empirical measure by $P_n^*$, which can refer to the multinomial, wild, or other exchangeable bootstraps. For a measurable vector-valued function $g : \mathcal{X} \mapsto \mathbb{R}^m$ indexed by a parameter $\beta$, denote $P g(\cdot, \beta) = E \left[ g(X_1, \beta) \right]$, $P_n g(\cdot, \beta) = \frac{1}{n} \sum_{i=1}^{n} g(X_i, \beta)$, and $P_n^* g(\cdot, \beta) = \frac{1}{n} \sum_{i=1}^{n} g(X_i^*, \beta)$ for a bootstrap sample $X_1^*, \ldots, X_n^*$. Weak convergence is defined in the sense of Kosorok (2007): $Z_n \rightsquigarrow Z$ in the metric space $(\mathbb{D}, d)$ if and only if $\sup_{f \in BL_1} |E f(Z_n) - E f(Z)| \xrightarrow{P} 0$ where $BL_1$ is the space of functions $f : \mathbb{D} \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1. Conditional weak convergence is also defined in the sense of Kosorok (2007): $Z_n \rightsquigarrow Z$ in the metric space $(\mathbb{D}, d)$ if and only if $\sup_{f \in BL_1} |E_{W} f(Z_n) - E_{W} f(Z)| \xrightarrow{P} 0$ and $E_{W} f(Z_n)^* - E_{W} f(Z_n)^* \xrightarrow{P} 0$ for all $f \in BL_1$, where $BL_1$ is the space of functions $f : \mathbb{D} \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1, $E_{W}$ denotes expectation with respect to the bootstrap weights $W$ conditional on the data, and $f(Z_n)^*$ and $f(Z_n)^*$ denote measurable majorants and minorants with respect to the joint data (including the weights $W$). Let $X_n^* = \sigma_P^*(1)$ if the law of $X_n^*$ is governed by $P_n$ and if $P_n(|X_n^*| > \epsilon) = o_P(1)$ for all $\epsilon > 0$. Also define $M_n^* = O_P^*(1)$ (hence also $O_p(1)$) if $\lim_{m \to \infty} \limsup_{n \to \infty} P \left( P_n(M_n^* > m) > \epsilon \right) \to 0$, $\forall \epsilon > 0$.

III.b Finite-dimensional Regularized Estimators

We first consider $\sqrt{n}$-consistent estimators $\hat{\beta}_n$ that minimize an objective function which can be written as the sum of two functions: the random, possibly non-convex, non-smooth loss function $\hat{Q}_n(\beta)$, and the penalty function $\frac{\lambda_n}{\sqrt{n}} r(\beta)$, where $r : \mathbb{R}^d \mapsto \mathbb{R}$ is a typically convex but non-smooth deterministic function, and $\frac{\lambda_n}{\sqrt{n}} = o(1)$. We assume $d$ is fixed.

$$\hat{\beta}_n = \arg \min_{\beta \in \mathbb{R}^d} \left\{ \hat{Q}_n(\beta) + \frac{\lambda_n}{\sqrt{n}} r(\beta) \right\}$$

We propose a proximal bootstrap estimator $\hat{\beta}_n^*$ that can be efficiently computed using standard, built-in optimization routines starting from an initial $\sqrt{n}$-consistent estimator $\sqrt{n} \left( \hat{\beta}_n - \beta_0 \right) = O_P(1)$, where $\beta_0 = \arg \min_{\beta \in \mathbb{R}^d} Q(\beta)$. One possible $\hat{\beta}_n$ is $\hat{\beta}_n$, but sometimes there are more computationally efficient estimators. For some $\alpha_n \to 0$ and $\alpha_n \sqrt{n} \to \infty$,

$$\hat{\beta}_n^* = \text{prox}_{H_n, \alpha_n \lambda_n r(\cdot)} \left( \hat{\beta}_n - \alpha_n \sqrt{n} H_n^{-1} \left( \hat{l}_n(\hat{\beta}_n) - \hat{l}_n(\hat{\beta}_n) \right) \right)$$

$\hat{l}_n(\hat{\beta}_n)$ is a consistent estimate of $l(\beta_0) = \frac{\partial Q(\beta_0)}{\partial \beta}$, where $Q(\beta)$ is a lower semicontinuous function that is twice differentiable at $\beta_0$, and $\sup_{\beta \in K} \left| \hat{Q}_n(\beta) - Q(\beta) \right| = o_P(1)$ for every compact subset $K$ of $\mathbb{R}^d$. In the case where $\hat{Q}_n(\beta)$ is differentiable, $\hat{l}_n(\beta)$ can simply be the
which often has a closed form solution. Furthermore, for certain types of 
\(\hat{\beta}_n\) is a bootstrap analog of \(\hat{\beta}_n\) using the multinomial, wild, or other exchangeable bootstraps. \(\hat{H}_n\) is a consistent, symmetric, positive definite estimate of the population Hessian \(H_0 = \frac{\partial^2 Q(\beta_0)}{\partial \beta \partial \beta}\).

If \(\hat{H}_n = \frac{1}{n}I_d\) for some constant \(c\), then \(\hat{\beta}_n\) reduces down to an unscaled proximal map which often has a closed form solution.

\[
\hat{\beta}_n = \underset{\beta}{\text{prox}} \left( \frac{1}{2}I_d, c_n \lambda_n r(\cdot) \right) \left( \beta_n - \alpha_n \sqrt{n} c_n I_d \left( \hat{t}_n(\beta_n) - \hat{t}_n(\bar{\beta}_n) \right) \right)
\]

\[
= \underset{\beta}{\text{arg min}} \left\{ \alpha_n \lambda_n r(\beta) + \frac{1}{2} \left\| \beta - \beta_n + \alpha_n \sqrt{n} \left( \hat{t}_n(\beta_n) - \hat{t}_n(\bar{\beta}_n) \right) \right\|^2 \right\}
\]

\[
= \underset{\beta}{\text{arg min}} \left\{ \alpha_n \lambda_n r(\beta) + \frac{1}{2} \left\| \beta - \beta_n + \alpha_n \sqrt{n} \left( \hat{t}_n(\beta_n) - \hat{t}_n(\bar{\beta}_n) \right) \right\|^2 \right\}
\]

Even if there is no closed form for \(\hat{\beta}_n\), it is still the solution to a convex optimization problem assuming \(r(\beta)\) is convex:

\[
\hat{\beta}_n = \underset{\beta}{\text{arg min}} \left\{ \alpha_n \lambda_n r(\beta) + \frac{1}{2} \left\| \beta - \beta_n + \alpha_n \sqrt{n} \hat{H}_n^{-1} \left( \hat{t}_n(\beta_n) - \hat{t}_n(\bar{\beta}_n) \right) \right\|^2 \right\}
\]

Furthermore, for certain types of \(r(\beta)\), we can use proposition 4.1 in Friedlander and Goh (2017) to efficiently compute the proximal bootstrap by solving a quadratic optimization problem over conic constraints. For example, if \(r(\beta) = \|\beta\|_1\),

\[
\hat{\beta}_n = \hat{H}_n^{-1} \left( \hat{H}_n \left( \beta_n - \alpha_n \sqrt{n} \hat{H}_n^{-1} \left( \hat{t}_n(\beta_n) - \hat{t}_n(\bar{\beta}_n) \right) \right) - \alpha_n \lambda_n \gamma \right)
\]

\[
\gamma = \underset{\gamma \in \{ \gamma : 1 \leq \gamma \leq 1 \}}{\text{arg min}} \left\{ \alpha_n \lambda_n \gamma \right\} \hat{H}_n^{-1} \left( \beta_n - \alpha_n \sqrt{n} \hat{H}_n^{-1} \left( \hat{t}_n(\beta_n) - \hat{t}_n(\bar{\beta}_n) \right) \right)^T \gamma
\]

Remark 1. In the case of \(r(\beta) = 0\), smooth \(\hat{Q}_n(\beta)\), and \(\beta_n\) that satisfies \(\hat{t}_n(\bar{\beta}_n) = 0\), the proximal bootstrap is similar to the k-step bootstrap (for \(k = 1\)) proposed by Davidson and MacKinnon (1999) and investigated further by Andrews (2002b), except with an additional scaling factor of \(\alpha_n \sqrt{n}\):

\[
\hat{\beta}_n = \beta_n - \alpha_n \sqrt{n} \hat{H}_n^{-1} \left( \hat{t}_n(\beta_n) - \hat{t}_n(\bar{\beta}_n) \right) = \beta_n - \alpha_n \sqrt{n} \hat{H}_n^{-1} \hat{t}_n(\bar{\beta}_n)
\]

If \(\alpha_n = 1/\sqrt{n}\) in this case, then the proximal bootstrap coincides with the 1-step bootstrap.
III.c Assumptions

The first assumption is needed to show consistency of \( \hat{\beta}_n \) for \( \beta_0 \).

**Assumption 1.** (i) \( \hat{\beta}_n = \arg \min_{\beta \in \mathbb{R}^d} \left\{ \hat{Q}_n (\beta) + \frac{n}{\sqrt{n}} r (\beta) \right\} \) is uniformly tight. (ii) \( \beta_0 = \arg \min_{\beta \in \mathbb{R}^d} Q (\beta) \) is unique, where \( Q (\beta) \) is a lower semicontinuous function that is twice differentiable at \( \beta_0 \) and \( \sup_{\beta \in K} \left| \hat{Q}_n (\beta) - Q (\beta) \right| = o_p(1) \) for every compact subset \( K \) of \( \mathbb{R}^d \).

The next assumption states that the objective function admits a uniform local quadratic approximation around \( \sqrt{n} \) neighborhoods of \( \beta_0 \). It is needed to derive the asymptotic distribution of \( \sqrt{n} (\hat{\beta}_n - \beta_0) \).

**Assumption 2.** There exists a symmetric, positive definite \( H_0 \) and \( \sqrt{n} (\hat{\beta}_n - \beta_0) = O_p(1) \) such that for any \( \delta_n \to 0 \),

\[
\sup_{\|h\| \leq \sqrt{n} \delta_n} \left| \frac{n \hat{Q}_n (\beta_0 + h/\sqrt{n}) - n \hat{Q}_n (\beta_0) - h' \sqrt{n} \left( \hat{\beta}_n (\beta_0) - l (\beta_0) \right) - \frac{1}{2} h' H_0 h}{1 + \|h\|^2} \right| = o_p(1)
\]

The next assumption is needed to show that \( \sqrt{n} (\hat{\beta}_n - \beta_0) \) and \( \sqrt{n} (\hat{\beta}_n - \hat{\beta}_n) \) have the same asymptotic distribution.

**Assumption 3.** There exists a function \( g : \mathcal{X} \mapsto \mathbb{R} \) indexed by a parameter \( \beta \in \mathbb{R}^d \) such that for any \( \beta \in \mathbb{R}^d \), \( \sqrt{n} (\hat{\beta}_n (\beta) - l(\beta)) = \sqrt{n} (P_n - P) g (\cdot, \beta) + o_p(1) \) and \( \sqrt{n} (\hat{\beta}_n (\beta) - \hat{\beta}_n (\beta)) = \sqrt{n} (P_n^* - P_n) g (\cdot, \beta) + o_p(1) \), where \( \lim_{n \to \infty} \|g (\cdot, \beta_0)\|^2 1 (\|g (\cdot, \beta_0)\| > \epsilon \sqrt{n}) = 0 \) for each \( \epsilon > 0 \).

The next assumption is needed to show stochastic equicontinuity of \( \sqrt{n} (\hat{\beta}_n (\beta) - l(\beta)) \) and bootstrap equicontinuity results which will be used to show \( \sqrt{n} (\hat{\beta}_n (\beta) - \hat{\beta}_n (\beta)) \) and \( \sqrt{n} (\hat{\beta}_n (\beta_0) - \hat{\beta}_n (\beta_0)) \) have the same asymptotic distribution.

**Assumption 4.** (i) \( \mathcal{G}_R = \{ g (\cdot, \beta) - g (\cdot, \beta_0) : \| \beta - \beta_0 \| \leq R \} \) is a Donsker class for some \( R > 0 \), and \( P (g (\cdot, \beta) - g (\cdot, \beta_0))^2 \to 0 \) for \( \beta \to \beta_0 \). (ii) \( \lim_{n \to \infty} \sup_{\beta \in \mathbb{R}^d} \sup_{t \geq \lambda} \frac{P \left\{ \left| \frac{g (\cdot, \beta) - g (\cdot, \beta_0)}{1 + \sqrt{n} \| \beta - \beta_0 \|} \right| > t \right\}}{\left| \frac{g (\cdot, \beta) - g (\cdot, \beta_0)}{1 + \sqrt{n} \| \beta - \beta_0 \|} \right|} \to 0 \) for any \( \delta_n \to 0 \).

(i) will imply stochastic equicontinuity, which in combination with the envelope function integrability condition in (ii) will imply bootstrap equicontinuity. A sufficient condition for (ii) is \( \sup_{g (\cdot, \beta) \in \mathcal{G}_R} \left\| \frac{g (\cdot, \beta) - g (\cdot, \beta_0)}{1 + \sqrt{n} \| \beta - \beta_0 \|} \right\| \leq C \) for some constant \( C \).
The next assumption states that \( r(\beta) \) is closed, convex and Hadamard directionally differentiable at \( \beta_0 \). It is needed to derive the asymptotic distribution of \( \sqrt{n}(\hat{\beta}_n - \beta_0) \).

**Assumption 5.** \( r : \mathbb{R}^d \rightarrow \mathbb{R} \) is a proper closed, convex function, and there is a continuous map \( r'_\beta : \mathbb{R}^d \rightarrow \mathbb{R} \) such that for all \( h_n \rightarrow h \in \mathbb{R}^d \), \( \lim_{\alpha_n \downarrow 0} \left| r(\beta_0 + \alpha_n h_n) - r(\beta_0) - r'_\beta(h) \right| = 0 \).

The next theorem demonstrates consistency of the proximal bootstrap by showing that the limiting distribution of \( \hat{\beta}_n \) coincides with the limiting distribution of \( \sqrt{n}(\hat{\beta}_n - \beta_0) \).

**Theorem 1.** Suppose Assumptions 1-5 are satisfied and \( \lambda_n \rightarrow \lambda_0 \in [0, \infty) \). Then for any \( \hat{\beta}_n \) such that \( \sqrt{n}(\hat{\beta}_n - \beta_0) = O_p(1) \), for any \( \hat{H}_n \) that is a consistent, symmetric, positive definite estimate of \( H_0 \), and for any sequence \( \alpha_n \) such that \( \alpha_n \rightarrow 0 \) and \( \sqrt{n}\alpha_n \rightarrow \infty \),

\[
\sqrt{n}(\hat{\beta}_n - \beta_0) \stackrel{\mathcal{D}}{\rightarrow} J \text{ and } \frac{\hat{\beta}_n - \beta_0}{\alpha_n} \stackrel{\mathcal{D}}{\rightarrow} J, \text{ where } J = \arg \min_{h \in \mathbb{R}^d} \{ \lambda_0 r'_\beta(h) + h'W_0 + \frac{1}{2} h'H_0 h \}
\]

and \( W_0 \sim N(0, P(g(\cdot, \beta_0) - Pg(\cdot, \beta_0)) (g(\cdot, \beta_0) - Pg(\cdot, \beta_0))') \).

### IV Conclusion

We have proposed a computationally efficient proximal bootstrap estimator that consistently estimates the limiting distribution of \( \sqrt{n} \)-consistent estimators for which the standard bootstrap is known to be inconsistent. This paper has considered the application to finite-dimensional regularized estimators; an application to constrained estimators will be explored in a subsequent paper.

### References


Andrews, Donald W.K., “Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space,” *Econometrica*, 2000, 68 (2), 399–405. 1


Bai, Yuehao, Hung Ho, Guillaume A Pouliot, and Joshua KC Shea, “Inference for Support Vector Regression under l1 Regularization,” 2019. 1


Belloni, Alexandre and Victor Chernozhukov, “\( \ell_1 \)-penalized quantile regression in high-dimensional sparse models,” *The Annals of Statistics*, 2011, 39 (1), 82–130. 1, 8


Online Appendix

V Examples of Regularized Estimators

V.a LASSO

\[
\hat{\beta}_n = \arg\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \frac{\lambda_n}{\sqrt{n}} \|\beta\|_1 \right\}
\]

\[
r'_{\beta_0}(h) = \sum_{j=1}^{p} (h_j \text{sign}(\beta_{0j}) \mathbb{1}(\beta_{0j} \neq 0) + |h_j| \mathbb{1}(\beta_{0j} = 0))
\]

\[
\hat{\beta}_n^* = \arg\min_{\beta} \alpha_n \lambda_n \|\beta\|_1 + \alpha_n \sqrt{n} \left( \hat{\ell}_n(\beta_n) - \hat{\ell}_n(\bar{\beta}_n) \right)' (\beta - \bar{\beta}_n) + \frac{1}{2} \|\beta - \bar{\beta}_n\|_{H_n}^2
\]

\[
\ell(\beta_0) = -E[x_i(y_i - x_i^T \beta)], \quad \hat{\ell}_n(\bar{\beta}_n) = -\frac{1}{n} \sum_{i=1}^{n} x_i (y_i - x_i^T \bar{\beta}_n)
\]

\[
H_0 = E[x_i x_i'], \quad H_n = \frac{1}{n} \sum_{i=1}^{n} x_i x_i'
\]

Examples of \(\hat{\ell}_n^*(\bar{\beta}_n)\) include the multinomial and wild bootstrap analogs of \(\hat{\ell}_n(\bar{\beta}_n)\):

\[
\hat{\ell}_n^*(\bar{\beta}_n) = -\frac{1}{n} \sum_{i=1}^{n} x_i^T (y_i^* - x_i^T \bar{\beta}_n), \quad \hat{\ell}_n^*(\bar{\beta}_n) = -\frac{1}{n} \sum_{i=1}^{n} (\xi_i - \tilde{\xi}) x_i (y_i - x_i^T \bar{\beta}_n)
\]

where \(\xi_i\) are i.i.d. variables with variance 1 and finite 3rd moment and \(\tilde{\xi} = \frac{1}{n} \sum_{i=1}^{n} \xi_i\).

If \(\bar{H}_n = \frac{1}{c} I_d\), \(\hat{\beta}_n^*\) has a closed form solution:

\[
\hat{\beta}_n^* = \text{prox}_{\alpha_n \lambda_n \|\cdot\|_1} \left( \bar{\beta}_n - c \alpha_n \sqrt{n} \left( \hat{\ell}_n^*(\bar{\beta}_n) - \hat{\ell}_n(\bar{\beta}_n) \right) \right) = \left( \bar{\beta}_n - c \alpha_n \sqrt{n} \left( \hat{\ell}_n^*(\bar{\beta}_n) - \hat{\ell}_n(\bar{\beta}_n) \right) - c \alpha_n \lambda_n \right)^+ - \left( \bar{\beta}_n - c \alpha_n \sqrt{n} \left( \hat{\ell}_n^*(\bar{\beta}_n) - \hat{\ell}_n(\bar{\beta}_n) \right) + c \alpha_n \lambda_n \right)^-
\]

where \(x^+ \equiv \max(x, 0)\) and \(x^- \equiv -\min(x, 0)\).

V.b \(\ell_1\)-norm support vector regression

The \(\ell_1\)-norm support vector regression (SVR) estimator of Zhu et al. (2004) is similar to the \(\ell_1\) penalized quantile regression estimator of Belloni and Chernozhukov (2011):

\[
\hat{\beta}_n = \arg\min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} (\rho_\tau(y_i - x_i^T \beta) - \kappa)^+ + \frac{\lambda_n}{\sqrt{n}} \|\beta\|_1 \right\}
\]
The objective uses a relaxed version of the check function:

\[
(\rho_\tau (y_i - x'_i\beta) - \kappa)^+ \\
= \{(1 - \tau) 1 (y_i - x'_i\beta \leq 0) + \tau 1 (y_i - x'_i\beta > 0)\} |y_i - x'_i\beta| - \kappa)^+ \\
= \begin{cases} 
(1 - \tau) (y_i - x'_i\beta - \kappa) 1 (y_i - x'_i\beta - \kappa > 0), & y_i - x'_i\beta \leq 0 \\
(\tau (y_i - x'_i\beta - \kappa)) 1 (\tau (y_i - x'_i\beta - \kappa > 0), & y_i - x'_i\beta > 0 \\
\end{cases} \\
= \begin{cases} 
(1 - \tau) (y_i - x'_i\beta - \kappa) 1 (y_i < x'_i\beta - \frac{\kappa}{1 - \tau}), & y_i - x'_i\beta \leq 0 \\
(\tau (y_i - x'_i\beta - \kappa)) 1 (\tau > x'_i\beta + \frac{\kappa}{\tau}), & y_i - x'_i\beta > 0 \\
\end{cases} \\
= (\tau (y_i - x'_i\beta - \kappa) 1 (y_i > x'_i\beta + \frac{\kappa}{\tau}) - (1 - \tau) (y_i - x'_i\beta + \kappa) 1 (y_i < x'_i\beta - \frac{\kappa}{1 - \tau})
\]

The proximal bootstrap estimator is

\[
\hat{\beta}^*_n = \arg\min_{\beta} \alpha_n \lambda_n \|\beta\|_1 + \alpha_n \sqrt{n} \left( \hat{l}^*_n (\hat{\beta}_n) - \hat{l}_n (\hat{\beta}_n) \right)' (\beta - \hat{\beta}_n) + \frac{1}{2} \|\beta - \hat{\beta}_n\|^2_{\bar{H}_n} \\
\hat{l}_n (\hat{\beta}_n) \text{ is a consistent estimate of } l(\beta_0) \text{ using } \hat{\beta}_n:
\]

\[
l(\beta_0) = -E \left( x_i \left( \tau 1 (y_i > x'_i\beta_0 - \frac{\kappa}{\tau}) - (1 - \tau) 1 (y_i < x'_i\beta_0 - \frac{\kappa}{1 - \tau}) \right) \right) \\
\hat{l}_n (\hat{\beta}_n) = -\frac{1}{n} \sum_{i=1}^n x_i \left( \tau 1 (y_i > x'_i\hat{\beta}_n + \frac{\kappa}{\tau}) - (1 - \tau) 1 (y_i < x'_i\hat{\beta}_n - \frac{\kappa}{1 - \tau}) \right)
\]

The population Hessian and its consistent estimate using \( \hat{\beta}_n \) are given by

\[
H_0 = E \left[ x_i x'_i \left( \tau f_{y|x} (x'_i\beta_0 + \frac{\kappa}{\tau}) + (1 - \tau) f_{y|x} (x'_i - \frac{\kappa}{1 - \tau}) \right) \right] \\
\bar{H}_n = \frac{1}{n} \sum_{i=1}^n x_i x'_i \left( \tau \hat{f}_{y|x} (x'_i\hat{\beta}_n + \frac{\kappa}{\tau}) + (1 - \tau) \hat{f}_{y|x} (x'_i\hat{\beta}_n - \frac{\kappa}{1 - \tau}) \right)
\]

An example of \( \hat{f}_{y|x} (y) \) is \( \frac{1}{n} \sum_{i=1}^n K_h (y) \), where \( K_h (y) = \frac{1}{h} K (y/h) \) and \( K (u) \) is a kernel function that is symmetric around 0 and integrates to 1.

**V.c Trace Regression via Nuclear Norm Regularization**

\[
\hat{\Theta}_n = \arg\min_{\Theta \in \mathbb{R}^{d_1 \times d_2}} \left\{ \frac{1}{n} \sum_{i=1}^n (y_i - tr (\Theta X_i)) + \lambda_n \|\Theta\|_* \right\}
\]

9
where \( \| \Theta \|_* = \sum_{j=1}^{d_1 \wedge d_2} \sigma_j(\Theta) \) is the nuclear norm of \( \Theta \), and \( \sigma_j(\Theta) \) is the \( j \)th largest singular value of \( \Theta \).

\[
\hat{\Theta}_n^* = \arg\min_{\Theta} \alpha_n \lambda_n \| \Theta \|_* + \alpha_n \sqrt{n} \left( \hat{i}_n(\Theta_n) - \hat{i}_n(\Theta) \right)'(\Theta - \Theta) + \frac{1}{2} \| \Theta - \Theta_n \|_{H_n}^2
\]

\[
\hat{i}_n(\Theta_n) = -\frac{1}{n} \sum_{i=1}^{n} X_i \left( y_i - tr(\hat{\Theta}_n X_i) \right), \quad \hat{H}_n = \frac{1}{n} \sum_{i=1}^{n} X_i X_i'
\]

\[
r'_\beta_0(h) = \sum_{j=1}^{d_1 \wedge d_2} (h_j | 1 (\sigma_j(\Theta_0) \neq 0) + |h_j| 1 (\sigma_j(\Theta_0) = 0))
\]

In the case of \( H_n = \frac{1}{c} I_{d_1} \), the proximal bootstrap has a closed form:

\[
\hat{\Theta}_n^* = \text{prox}_{\alpha_n \lambda_n \| \cdot \|_*} \left( \hat{\Theta}_n - \alpha_n \sqrt{n} \left( \hat{i}_n(\Theta_n) - \hat{i}_n(\Theta) \right) \right) = U \Sigma_{\alpha_n \lambda_n} V'^T
\]

where \( \Sigma_{\alpha_n \lambda_n} = \text{diag} \{ \max(\Sigma_1 - \alpha_n \lambda_n, 0), \max(\Sigma_2 - \alpha_n \lambda_n, 0), \ldots, \max(\Sigma_{d_1 \wedge d_2} - \alpha_n \lambda_n, 0) \} \), and for \( j = 1 \ldots d_1 \wedge d_2 \), \( \Sigma_j \) are the singular values of \( \Theta_n - \alpha_n \sqrt{n} \left( \hat{i}_n(\Theta_n) - \hat{i}_n(\Theta) \right) \).

\section{VI Proof of Theorem 1}

Assumption 1 implies that the conditions of part 2 of Corollary 3.2.3 of \textit{van der Vaart and Wellner (1996)} are satisfied, and therefore \( \hat{\beta}_n \xrightarrow{p} \beta_0 = \arg\min_{\beta \in \mathbb{R}^d} Q(\beta) \). To derive its asymptotic distribution, use the centered and scaled parameter \( h = \sqrt{n}(\beta - \beta_0) \):

\[
\sqrt{n} \left( \hat{\beta}_n - \beta_0 \right) = \arg\min_h \left\{ n\hat{Q}_n \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - n\hat{Q}_n(\beta_0) + \lambda_n \sqrt{n} r \left( \beta_0 + \frac{h}{\sqrt{n}} \right) \right\}
\]

\[
= \arg\min_h \left\{ h' \sqrt{n} \left( \hat{i}_n(\beta_0) - l(\beta_0) \right) + \frac{1}{2} h' H_0 h + \lambda_n \left( r \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - r(\beta_0) \right) \frac{1}{1/\sqrt{n}} \right\} + o_p(1)
\]

\[
\xrightarrow{\sim} \arg\min_h \left\{ \lambda_0 r'_{\beta_0}(h) + h' W_0 + \frac{1}{2} h' H_0 h \right\}
\]

The second line is due to the uniform in \( h \) local quadratic expansion of \( n\hat{Q}_n \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - n\hat{Q}_n(\beta_0) \), which follows from assumption 2. The last line follows from the following arguments. Assumption 3 implies the Lindeberg Condition is satisfied and \( \sqrt{n}(P_n - P) g(\cdot, \beta_0) \xrightarrow{\sim} W_0 \). Assumption 5 implies \( \frac{r(\beta_0 + \frac{h}{\sqrt{n}}) - r(\beta_0)}{1/\sqrt{n}} \rightarrow r'_{\beta_0}(h) \) for each \( h \in \mathbb{R}^d \) and that \( r'_{\beta_0}(h) \) is a convex function of \( h \). Since \( h' \sqrt{n} \left( \hat{i}_n(\beta_0) - l(\beta_0) \right) + \frac{1}{2} h' H_0 h + \lambda_n \left( r \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - r(\beta_0) \right) \frac{1}{1/\sqrt{n}} \) is a convex function of \( h \), pointwise convergence implies uniform convergence over compact sets \( K \subset \mathbb{R}^d \) (Pollard (1991)). Therefore,

\[
n\hat{Q}_n \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - n\hat{Q}_n(\beta_0) + \lambda_n \sqrt{n} \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - \lambda_n \sqrt{n} r(\beta_0) \xrightarrow{\sim} h' W_0 + \frac{1}{2} h' H_0 h + \lambda_0 r'_{\beta_0}(h)
\]
as a process indexed by $h$ in the space of bounded functions $\ell^\infty(K)$ for any compact $K \subset \mathbb{R}^d$. Convexity implies $\lambda_0 r_{\beta_0}'(h) + h'W_0 + \frac{1}{2}h'H_0h$ has a unique minimum, so by the argmin continuous mapping theorem (Theorem 3.2.2 in van der Vaart and Wellner (1996)),

$$ \sqrt{n} \left( \hat{\beta}_n - \beta_0 \right) \xrightarrow{p} \mathcal{J}. $$

Now we show $\hat{\beta}_n - \beta_0$. Since $\alpha_n \to 0$ and $\alpha_n \lambda_n \to 0$ imply $\alpha_n \lambda_n r(h + \beta_0) = o(1)$ and $\alpha_n \sqrt{n} \bar{H}_n \left( \hat{\alpha}^*_n ( \hat{\beta}_n ) - \hat{\alpha}^*_n ( \hat{\beta}_n ) \right) = o_p^*(1)$,

$$ \hat{\beta}_n - \beta_0 = \arg \min_h \left\{ \alpha_n \lambda_n r \left( h + \beta_0 \right) + \frac{1}{2} \left\| h + \beta_0 - \hat{\beta}_n + \alpha_n \sqrt{n} \bar{H}_n^{-1} \left( \hat{\alpha}^*_n ( \hat{\beta}_n ) - \hat{\alpha}^*_n ( \hat{\beta}_n ) \right) \right\}^2 \right\} $$

$$ = \arg \min_h \left\{ \frac{1}{2} h'H_0h + h'H_0 \left( \beta_0 - \beta_0 \right) + \frac{1}{2} \left\| \beta_0 - \beta_0 \right\|_{H_0}^2 \right\} + o_p(1) $$

$$ = \beta_0 - \beta_0 + o_p(1) = o_p(1) $$

The second line follows from convexity of the proximal bootstrap objective function, which implies the difference between $\alpha_n \lambda_n r \left( h + \beta_0 \right) + \frac{1}{2} \left\| h + \beta_0 - \hat{\beta}_n + \alpha_n \sqrt{n} \bar{H}_n^{-1} \left( \hat{\alpha}^*_n ( \hat{\beta}_n ) - \hat{\alpha}^*_n ( \hat{\beta}_n ) \right) \right\}^2$

and $\frac{1}{2} \left\| h + \beta_0 - \beta_0 \right\|_{H_0}^2 = \frac{1}{2} h'H_0h + h'H_0 \left( \beta_0 - \beta_0 \right) + \frac{1}{2} \left\| \beta_0 - \beta_0 \right\|_{H_0}^2$ converges uniformly in probability to zero over any compact subset of $\mathbb{R}^d$.

To derive $\hat{\beta}_n^*$’s asymptotic distribution, first note that because $\sqrt{n} \left( \hat{\beta}_n - \beta_0 \right) = O_p(1)$

and $\sqrt{n} \alpha_n \to \infty$, 

$$ \frac{\hat{\beta}_n^* - \beta_0}{\alpha_n} = \frac{\hat{\beta}_n^* - \beta_0}{\alpha_n} - \frac{\sqrt{n} \left( \hat{\beta}_n - \beta_0 \right)}{\sqrt{n} \alpha_n} = \frac{\hat{\beta}_n^* - \beta_0}{\alpha_n} + o_p(1) $$

It therefore suffices to show that $\frac{\hat{\beta}_n^* - \beta_0}{\alpha_n} \xrightarrow{p_{\mathcal{W}}} \mathcal{J}$. To do this, use the centered and scaled parameter $h = (\beta - \beta_0) / \alpha_n$:

$$ \frac{\hat{\beta}_n^* - \beta_0}{\alpha_n} = \arg \min_h \left\{ \alpha_n \lambda_n r \left( \frac{h + \beta_0 - \beta_0}{\alpha_n} + \frac{1}{2} h'H_0h \right) \right\} $$

We have used $\frac{\beta_0 - \beta_0}{\alpha_n} = \frac{\sqrt{n} \left( \beta_0 - \beta_0 \right)}{\sqrt{n} \alpha_n} = o_p(1)$, $H_n \xrightarrow{p_{\mathcal{W}}} H_0$, the assumption of directional differentiability of $r(\beta)$ at $\beta_0$, and the following arguments. Assumption 4(i) says $G_R \equiv \{ g(\cdot, \beta) - g(\cdot, \beta_0) \} \leq R$ is a Donsker class for some $R > 0$, and $P \left( g(\cdot, \beta) - g(\cdot, \beta_0) \right)^2 \to 0$ for $\beta \to \beta_0$. By Lemma 3.3.5 of van der Vaart and Wellner (1996), $\sqrt{n} \left( P_n - P \right) g(\cdot, \beta)$ is stochastically equicontinuous, which implies

$$ \left\| \sqrt{n} \left( P_n - P \right) \left( g(\cdot, \beta_0) - g(\cdot, \beta_0) \right) \right\| = o_p(1 + \sqrt{n} \left\| \beta_n - \beta_0 \right\|) = o_p(1) \]
Stochastic equicontinuity and the envelope integrability condition in assumption 4(ii) imply that the assumptions of Lemma 4.2 in Wellner and Zhan (1996) are satisfied. Therefore, 
\[ \sqrt{n} \left( P_n^* - P_n \right) g \left( \cdot, \beta \right) \] is bootstrap equicontinuous, which implies
\[ \left\| \sqrt{n} \left( P_n^* - P_n \right) (g \left( \cdot, \beta_n \right) - g \left( \cdot, \beta_0 \right) ) \right\| = o_p^* \left( 1 + \sqrt{n} \left\| \beta_n - \beta_0 \right\| \right) = o_p^* (1) \]

Therefore, 
\[ h' \sqrt{n} \left( \hat{\beta}_n (\beta_n) - \hat{\beta}_n (\beta_0) \right) = h' \sqrt{n} \left( P_n^* - P_n \right) g \left( \cdot, \beta_0 \right) + h' \sqrt{n} \left( P_n^* - P_n \right) (g \left( \cdot, \beta_n \right) - g \left( \cdot, \beta_0 \right) ) + o_p^* (1) \xrightarrow{p} h' W_0. \]
By convexity, pointwise convergence implies uniform convergence over compact sets \( K \subset \mathbb{R}^d \), so
\[ \lambda_n \left( \frac{r \left( \beta_0 + \alpha_n h \right) - r (\beta_0)}{\alpha_n} \right) + h' \sqrt{n} \left( \hat{t}_n (\beta_n) - \hat{t}_n (\beta_0) \right) + \frac{1}{2} h' H_n h \xrightarrow{p} \lambda_0 r' (\beta_0) + h' W_0 + \frac{1}{2} h' H_0 h \]
as a process indexed by \( h \) in the space of bounded functions \( \ell^c (K) \) for any compact \( K \subset \mathbb{R}^d \).

\[ \frac{\hat{\beta}_n - \beta_0}{\alpha_n} \xrightarrow{p} J \] follows from the bootstrap version of the argmin continuous mapping theorem (see Lemma 14.2 in Hong and Li (2020)).

\[ \text{Monte Carlo Simulation for Finite-dimensional Lasso} \]

We consider the following data generating process:
\[ y_i = x_i' \beta_0 + \epsilon_i, \quad \beta_0 = \left( 1 \quad 0 \quad 0 \quad 0 \quad 0 \right) ', \quad x_i \sim N \left( 0, I_5 + 0.5 (\tau - I_5) \right), \quad \epsilon_i \sim N \left( 0, 1 \right) \]

We compute the Lasso estimator \( \hat{\beta}_n = \arg \min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - x_i' \beta)^2 + \frac{\lambda_n}{\sqrt{n}} \left\| \beta \right\|_1 \right\} \) using the CVX modeling software in Matlab developed by Grant and Boyd (2009). The proximal bootstrap estimator \( \hat{\beta}_n^* = \arg \min_{\beta} \alpha_n \lambda_n \left\| \beta \right\|_1 + \frac{\lambda_n}{\sqrt{n}} \left( \hat{t}_n (\beta_0) - \hat{t}_n (\beta_n) \right) \) for \( \beta_n = \hat{\beta}_n, \quad \hat{H}_n = \frac{1}{n} \sum_{i=1}^{n} x_i' x_i, \quad \hat{t}_n (\beta_n) = -\frac{1}{n} \sum_{i=1}^{n} x_i (y_i - x_i' \beta_n), \quad \text{and} \quad \hat{t}_n^* (\beta_n) = -\frac{1}{n} \sum_{i=1}^{n} x_i (y_i^* - x_i' \beta_n), \)
is computed using the fminunc Matlab function so that we can run the code in parallel (the current version of CVX does not support parallel code for loops). We also tried using the fmincon Matlab function, and the results were the same.

We consider five different sample sizes \( n \in \{100, 500, 1000, 5000, 10000\} \), three different \( \alpha_n \)'s for each \( n \): \( \alpha_n \in \{ n^{-1/3}, n^{-1/4}, n^{-1/6} \} \), and two choices of \( \lambda_n \in \{ 0.1, 0.5 \} \). We use 5000 bootstrap iterations and 2000 Monte Carlo simulations. Empirical coverage frequencies for equal-tailed nominal 95% confidence intervals \( \left[ \hat{\beta}_n - \frac{c_{1-\tau} \lambda_n}{\sqrt{n}}, \hat{\beta}_n - \frac{c_{1-\tau} \lambda_n}{\sqrt{n}} \right] \), where \( c_{1-\tau} \) is the \( \tau \)-th percentile of \( \frac{\hat{\beta}_n^* - \beta_0}{\alpha_n} \), and average interval lengths are reported in tables 1-3. Although the proximal bootstrap undercovers for smaller sample sizes, it achieves coverage very close to 95% for sufficiently large \( n \).
Table 1: Proximal Bootstrap Coverage Frequencies and Interval Lengths for $\alpha_n = n^{-1/3}$

<table>
<thead>
<tr>
<th>$\lambda_n = 0.1$</th>
<th>$\lambda_n = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>100</td>
</tr>
<tr>
<td>0.940</td>
<td>0.940</td>
</tr>
<tr>
<td>0.922</td>
<td>0.944</td>
</tr>
<tr>
<td>0.935</td>
<td>0.945</td>
</tr>
<tr>
<td>0.929</td>
<td>0.947</td>
</tr>
<tr>
<td>0.950</td>
<td>0.940</td>
</tr>
<tr>
<td>0.949</td>
<td>0.944</td>
</tr>
<tr>
<td>0.933</td>
<td>0.945</td>
</tr>
<tr>
<td>0.929</td>
<td>0.947</td>
</tr>
<tr>
<td>0.951</td>
<td>0.940</td>
</tr>
<tr>
<td>0.949</td>
<td>0.944</td>
</tr>
<tr>
<td>0.933</td>
<td>0.945</td>
</tr>
<tr>
<td>0.929</td>
<td>0.947</td>
</tr>
</tbody>
</table>

Table 2: Proximal Bootstrap Coverage Frequencies and Interval Lengths for $\alpha_n = n^{-1/4}$

<table>
<thead>
<tr>
<th>$\lambda_n = 0.1$</th>
<th>$\lambda_n = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>100</td>
</tr>
<tr>
<td>0.930</td>
<td>0.940</td>
</tr>
<tr>
<td>0.921</td>
<td>0.944</td>
</tr>
<tr>
<td>0.936</td>
<td>0.945</td>
</tr>
<tr>
<td>0.929</td>
<td>0.947</td>
</tr>
<tr>
<td>0.950</td>
<td>0.940</td>
</tr>
<tr>
<td>0.949</td>
<td>0.944</td>
</tr>
<tr>
<td>0.933</td>
<td>0.945</td>
</tr>
<tr>
<td>0.929</td>
<td>0.947</td>
</tr>
<tr>
<td>0.951</td>
<td>0.940</td>
</tr>
<tr>
<td>0.949</td>
<td>0.944</td>
</tr>
<tr>
<td>0.933</td>
<td>0.945</td>
</tr>
<tr>
<td>0.929</td>
<td>0.947</td>
</tr>
</tbody>
</table>

13
We also compare the proximal bootstrap to the standard multinomial bootstrap estimator \( \hat{\beta}^{**} = \arg\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i^* - x_i^* \beta)^2 + \frac{\lambda_n}{\sqrt{n}} \| \beta \|_1 \right\} \). Empirical coverage frequencies for equal-tailed nominal 95% confidence intervals \( \left[ \hat{\beta}_n - \frac{d_{\tau} \sqrt{\lambda_n}}{\sqrt{n}}, \hat{\beta}_n - \frac{d_{\tau} \sqrt{\lambda_n}}{\sqrt{n}} \right] \), where \( d_{\tau} \) is the \( \tau \)-th percentile of \( \sqrt{n} \left( \hat{\beta}^{**}_n - \hat{\beta}_n \right) \), and average interval lengths are reported in table 4. We use 5000 bootstrap iterations and 2000 Monte Carlo simulations. Interestingly, for the case of \( \lambda_n = 0.1 \), the standard bootstrap coverage frequencies are close to the nominal level. The surprisingly good coverage of the standard bootstrap under certain DGPs is also documented in section 6.2 of Chatterjee and Lahiri (2011). However, when we use \( \lambda_n = 0.5 \), the standard bootstrap undercovers for the nonzero parameter and overcovers for the zero parameters. Additionally, the standard bootstrap confidence intervals are on average wider than the proximal bootstrap confidence intervals.

Table 4: Proximal Bootstrap Coverage Frequencies and Interval Lengths for \( \lambda_n = n^{-1/6} \)

<table>
<thead>
<tr>
<th>Table 4: Standard Bootstrap Coverage Frequencies and Interval Lengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_n = 0.1 )</td>
</tr>
<tr>
<td>( n )</td>
</tr>
<tr>
<td>0.947</td>
</tr>
<tr>
<td>(0.509)</td>
</tr>
<tr>
<td>0.950</td>
</tr>
<tr>
<td>(0.477)</td>
</tr>
<tr>
<td>0.963</td>
</tr>
<tr>
<td>(0.478)</td>
</tr>
<tr>
<td>0.959</td>
</tr>
<tr>
<td>(0.474)</td>
</tr>
<tr>
<td>0.958</td>
</tr>
<tr>
<td>(0.476)</td>
</tr>
</tbody>
</table>