The Proximal Bootstrap for Constrained Estimators *

Jessie Li

November 1, 2021

We demonstrate how to use the proximal bootstrap to pointwise consistently estimate the limiting distribution of $\sqrt{n}$-consistent estimators defined as the solution to a constrained optimization problem with a possibly nonsmooth and nonconvex sample objective function and a constraint set defined by smooth equalities and/or inequalities that can be either fixed or estimated from the data at the $\sqrt{n}$ rate. The proximal bootstrap estimator is typically much faster to compute than the standard bootstrap because it can be written as the solution to a quadratic programming problem. Monte Carlo simulations illustrate the valid coverage of the proximal bootstrap in a boundary constrained nonsmooth GMM model, a conditional logit model with estimated capacity constraints, and a Mathematical programming with equilibrium constraints (MPEC) formulation of the Rust (1987) Bus Engine Replacement model proposed in Su and Judd (2012).

Keywords: bootstrap, convex optimization, constrained optimization, proximal mapping

1 Introduction

This paper considers using the proximal bootstrap estimator proposed in Li (2021) to conduct pointwise asymptotically valid inference for a large class of $\sqrt{n}$-consistent estimators with possibly nonstandard asymptotic distributions for which standard bootstrap procedures fail. The application which we will focus on in this paper is estimators defined by the solution to a constrained optimization problem with a possibly nonsmooth and nonconvex sample objective function and either estimated or fixed smooth inequality and/or equality constraints. A well-known example of a constrained estimator with a nonstandard distribution is the constrained MLE estimator where the true parameter lies on the boundary of the constraint set (Andrews (1999), Andrews (2000), Andrews (2002)).

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 consistent estimator using built-in and freely available software. 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, similar to the $\epsilon_n$ 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 more computationally efficient than the numerical bootstrap. Another novel part of this paper is that we provide a general asymptotic distribution for estimators defined

---

* I would like to thank the participants of the University of Chicago Econometrics workshop (in particular Stéphane Bonhomme, Azeem Shaikh, and Alex Torgovitsky) and the 2021 ASSA session Optimization-conscious Econometrics for helpful comments and suggestions.
by the solution to a constrained optimization problem with equality and/or inequality constraints which can be estimated from the data, while Hong and Li (2020) looked only at estimators with fixed constraints that do not depend on the data. The asymptotic distribution of constrained estimators with estimated constraints is derived using ideas from the optimization literature and encompasses as special cases the results in Geyer (1994), Andrews (1999), Andrews (2000), and Andrews (2002) for constrained estimators with fixed constraint sets and true parameters possibly lying on the boundaries of the constraint sets. Examples of constrained estimators appear in Moon and Schorfheide (2009), Kaido (2016), Hsieh et al. (2018), Kaido et al. (2019), Kaido et al. (2021), and Horowitz and Lee (2019). While several of these papers are concerned with conducting inference on the optimal value of the constrained optimization problem, we are instead interested in conducting inference on the optimal solution.

Section 2 briefly reviews the concept of proximal mappings from the optimization literature. Section 3 contains all the main theoretical results for finite-dimensional constrained estimators with either fixed constraints (subsection 3.2) or estimated constraints (subsection 3.3). Section 4 contains Monte Carlo simulation evidence demonstrating the validity of confidence intervals constructed using the proximal bootstrap for a boundary constrained nonsmooth GMM model, a conditional logit model with estimated capacity constraints, and the Mathematical programming with equilibrium constraints (MPEC) formulation of the Rust (1987) Bus Engine Replacement model proposed in Su and Judd (2012).

2 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

$$\text{prox}_r (z) = \arg \min_{\beta \in D} \left\{ r(\beta) + \frac{1}{2} \| \beta - z \|_2^2 \right\} \text{ for any } z \in D$$

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 \( \| \beta - z \|_H^2 = (\beta - z)' H (\beta - z) \),

$$\text{prox}_{H,r} (z) = \arg \min_{\beta \in D} \left\{ r(\beta) + \frac{1}{2} \| \beta - z \|_H^2 \right\} \text{ for any } z \in D$$

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

3 Proximal Bootstrap

3.1 Notation

Consider a random sample $X_1, X_2, ..., X_n$ of independent draws from a probability measure $P$ on a sample space $X$. 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. Weak convergence is defined in the sense of Kosorok (2007): $Z_n \Rightarrow Z$ in the metric space $(D, d)$ if and only if $\sup_{f \in BL_1} |E f(Z_n) - Ef(Z)| \rightarrow 0$ where $BL_1$ is the space of functions $f: D \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1. Conditional weak convergence is also defined in the sense of Kosorok (2007): $Z_n \overset{P}{\Rightarrow} Z$ in the metric space $(D, d)$.
if and only if $\sup_{f \in BL_1}|E_Wf(Z_n) - E f(Z)| \xrightarrow{P} 0$ and $E_Wf(Z_n) - E_Wf(Z_n)^* \xrightarrow{P} 0$ for all $f \in BL_1$, where $BL_1$ is the space of functions $f: \mathbb{D} \to \mathbb{R}$ 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 $X_n^* = o_p^*(1)$ 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 $\lim_{m \to \infty} \limsup_{n \to \infty} P_n(M_n^* > m) > \epsilon \to 0$, $\forall \epsilon > 0$.

### 3.2 Constrained Estimators with Fixed Constraints

It is well known (see e.g. Andrews (2000)) that the standard bootstrap is inconsistent when the true parameters $\beta_0$ lie on the boundary of the constraint set $C$. Andrews (1999) derives the asymptotic distribution of constrained extremum estimators where the rescaled constraint set $\sqrt{n}(C - \beta_0)$ can be approximated by a convex cone. Geyer (1994) considers a more general case where the cone does not need to be convex. We consider constrained estimators $\hat{\beta}_n = \arg \min_{\beta \in C} \hat{Q}_n(\beta)$, where $\hat{Q}_n(\beta)$ is possibly non-smooth, nonconvex function that converges uniformly to a function $Q(\beta)$ that is twice differentiable at $\beta_0 = \arg \min_{\beta \in C} Q(\beta)$, and $C$ is a closed constraint set that is Chernoff Regular at $\beta_0$.

We will show that the proximal bootstrap is consistent both when $\beta_0$ lies in the interior and on the boundary of $C$. For some $\alpha_n \to 0$ and $\alpha_n \sqrt{n} \to \infty$, define the proximal bootstrap estimator as

$$\hat{\beta}_n^* = \text{prox}_{R_n, \infty 1(\cdot \neq C)} \left( \tilde{\beta}_n - \alpha_n \sqrt{n} \tilde{H}_n \left( \tilde{\beta}_n - \tilde{l}_n(\tilde{\beta}_n) \right) \right)$$

$$= \arg \min_{\beta \in \mathbb{R}^q} \left\{ \infty 1(\beta \notin C) + \alpha_n \sqrt{n} \left( \tilde{H}_n(\tilde{\beta}_n) - \tilde{l}_n(\tilde{\beta}_n) \right)' \left( \beta - \tilde{\beta}_n \right) + \frac{1}{2} \left\| \beta - \tilde{\beta}_n \right\|_{H_n}^2 \right\}$$

$$= \arg \min_{\beta \in C} \left\{ \alpha_n \sqrt{n} \left( \tilde{H}_n(\tilde{\beta}_n) - \tilde{l}_n(\tilde{\beta}_n) \right)' \left( \beta - \tilde{\beta}_n \right) + \frac{1}{2} \left\| \beta - \tilde{\beta}_n \right\|_{H_n}^2 \right\}$$

In the case where $\hat{Q}_n(\beta)$ is differentiable, $\tilde{l}_n(\beta)$ can simply be the Jacobian of $\hat{Q}_n(\beta)$. More generally, to handle non-differentiable $\hat{Q}_n(\beta)$, $\tilde{l}_n(\beta)$ is a subgradient of $\hat{Q}_n(\beta)$. $\tilde{H}_n(\tilde{\beta}_n)$ is the bootstrap analog of $\tilde{l}_n(\tilde{\beta}_n)$. $\tilde{H}_n$ is a consistent, symmetric, positive definite estimate of the population Hessian $H_0 = \frac{\partial^2 Q(\beta_0)}{\partial \beta \partial \beta'}$.

It is well known (see e.g. Geyer (1994)) that if $Q(\beta)$ achieves its minimum over $C$ at some point $\beta_0$ where it has a local quadratic approximation $Q(\beta) = Q(\beta_0) + \frac{1}{2} (\beta - \beta_0)' H_0 (\beta - \beta_0) + o \left( \| \beta - \beta_0 \|^2 \right)$, then $\sqrt{n} \left( \tilde{\beta}_n - \beta_0 \right) \xrightarrow{\text{d}} \mathcal{J} = \arg \min_{h' \in T_C(\beta_0)} \left\{ h' W_0 + \frac{1}{2} h' H_0 h \right\}$, where $T_C(\beta_0) = \limsup_{\tau \to 0} \frac{G_{\beta_0}}{\tau}$ is the tangent cone of $C$ at $\beta_0$. For closed sets $C$ that are Chernoff Regular at $\beta_0$, the limit exists and $T_C(\beta_0) = \lim_{\tau \to 0} \frac{G_{\beta_0}}{\tau}$. In the next theorem, we show that $\frac{\hat{\beta}_n - \beta_0}{\alpha_n} \xrightarrow{P_{W, \infty}} \mathcal{J}$. Before we present the theorem, we list a few assumptions needed for the theorem to hold.

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

**Assumption 1.** (i) $\hat{\beta}_n = \arg \min_{\beta \in C} \hat{Q}_n(\beta)$ is uniformly tight.  $^1$

(ii) $\beta_0 = \arg \min_{\beta \in C} Q(\beta)$ is unique, where $Q(\beta)$ is a lower semicontinuous function that is twice differentiable at $\beta_0$ and $\sup_{\beta \in K} |\hat{Q}_n(\beta) - Q(\beta)| = o_P(1)$ for every compact subset $K$ of $C$.

$^1$For every $\epsilon > 0$, there exists a compact $K \subset C$ with $P \left( \hat{\beta}_n \notin K \right) \geq 1 - \epsilon$ for every $n$.  

3
The next assumption states that $\hat{Q}_n(\beta)$ 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 $\sqrt{n} \left( \hat{I}_n(\beta_0) - l(\beta_0) \right) = O_P(1)$ such that for any $\delta_n \to 0$,

$$\sup_{\|h\| \leq \sqrt{n}\delta_n} \left| \frac{n\hat{Q}_n(\beta_0 + \frac{h}{\sqrt{n}}) - n\hat{Q}_n(\beta_0) - h'\sqrt{n} \left( \hat{I}_n(\beta_0) - l(\beta_0) \right) - \frac{1}{2} h'H_0 h}{1 + \|h\|^2} \right| = o_P(1)$$

Note that assumption 2 effectively implies that $l(\beta_0) = o(1/\sqrt{n})$ because by a first order Taylor expansion of $n\hat{Q}_n(\beta_0 + \frac{h}{\sqrt{n}}) - nQ(\beta_0 + \frac{h}{\sqrt{n}})$ around $n\hat{Q}_n(\beta_0) - nQ(\beta_0)$, followed by a second order Taylor expansion of $nQ(\beta_0 + \frac{h}{\sqrt{n}})$ around $nQ(\beta_0)$,

$$n\hat{Q}_n(\beta_0 + \frac{h}{\sqrt{n}}) - n\hat{Q}_n(\beta_0) = nQ(\beta_0 + \frac{h}{\sqrt{n}}) - nQ(\beta_0) + h'\sqrt{n} \left( \hat{I}_n(\beta_0) - l(\beta_0) \right) + o_P(1)$$

$$= h'\sqrt{n}l(\beta_0) + \frac{1}{2} h'H_0 h + h'\sqrt{n} \left( \hat{I}_n(\beta_0) - l(\beta_0) \right) + o_P(1)$$

In order for assumption 2 to hold, it must be that $l(\beta_0) = o(1/\sqrt{n})$. When $l(\beta_0)$ does not depend on $n$, we will have $l(\beta_0) = 0$. We will relax this assumption in section 3.3 to allow for $l(\beta_0) \neq 0$.

The next assumption is needed to show that $\sqrt{n} \left( \hat{I}_n(\beta_0) - l(\beta_0) \right)$ and $\sqrt{n} \left( \hat{I}^*_n(\beta_0) - \hat{I}_n(\beta_0) \right)$ have the same asymptotic distribution.

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

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

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

**Theorem 1.** Suppose Assumptions 1-4 are satisfied, $C \subset \mathbb{R}^d$ is a non-random closed set that is Chernoff Regular at $\beta_0 = \arg \min_{\beta \in C} Q(\beta)$, and $Q(\beta) = Q(\beta_0) + \frac{1}{2} (\beta - \beta_0)' H_0 (\beta - \beta_0) + o \left( \| \beta - \beta_0 \|^2 \right)$.
For any $\bar{\beta}_n$ such that $\sqrt{n}(\bar{\beta}_n - \beta_0) = O_P(1)$ and $H_n \overset{p}{\to} H_0$, let

$$\hat{\beta}_n^* = \arg\min_{\beta \in C} \alpha_n \sqrt{n} \left( \hat{t}_n(\beta_n) - \hat{t}_n(\bar{\beta}_n) \right)' (\beta - \bar{\beta}_n) + \frac{1}{2} \| \beta - \bar{\beta}_n \|^2_{H_n}$$

and $\hat{\beta}_n - \bar{\beta}_n = o_p(1)$. Then for any sequence $\alpha_n$ such that $\alpha_n \to 0$ and $\sqrt{n} \alpha_n \to \infty$, $\sqrt{n}(\hat{\beta}_n - \beta_0) \overset{w}{\to} J$ and $\hat{\beta}_n - \beta_0$ imply that the conditions of Lemma 4.3 in Geyer (1994) are satisfied, and therefore $\sqrt{n}(\hat{\beta}_n - \beta_0) = O_P(1)$.

Proof: Assumption 1 implies that $\hat{\beta}_n \overset{p}{\to} \beta_0 = \arg\min_{\beta \in C} Q(\beta)$ (see e.g. Corollary 3.2.3 in van der Vaart and Wellner (1996)). Assumption 2, $Q(\beta) = Q(\beta_0) + \frac{1}{2} (\beta - \beta_0)' H_0 (\beta - \beta_0) + o \left( \| \beta - \beta_0 \|^2 \right)$, and $\hat{\beta}_n \overset{p}{\to} \beta_0$ imply that the conditions of Lemma 4.3 in Geyer (1994) are satisfied, and therefore $\sqrt{n}(\hat{\beta}_n - \beta_0) = O_P(1)$.

To derive its asymptotic distribution, use the centered and scaled parameter $h = \sqrt{n}(\beta - \beta_0)$:

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \arg\min_{h \in \sqrt{n}(C - \beta_0)} \left\{ nQ_n \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - nQ_n (\beta_0) \right\}$$

and

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \arg\min_{h \in \sqrt{n}(C - \beta_0)} \left\{ h' \sqrt{n} \left( \hat{t}_n(\beta_0) - l(\beta_0) \right) + \frac{1}{2} h'H_0 h + o_P(1) \right\}$$

The second line is due to the uniform in $h$ local quadratic expansion of $nQ_n \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - nQ_n (\beta_0)$, which follows from assumption 2.

Then assumption 3 implies the Lindeberg Condition is satisfied and $\sqrt{n}(P_n - P) g(\cdot, \beta_0) \overset{w}{\to} W_0$. Therefore,

$$nQ_n \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - nQ_n (\beta_0) \overset{w}{\to} h'W_0 + \frac{1}{2} h'H_0 h$$

as a process indexed by $h$ in the space of bounded functions $\ell^\infty(K)$ for any compact $K \subset \mathbb{R}^d$. Since $h'W_0 + \frac{1}{2} h'H_0 h$ has a continuous sample path, according to page 5 of Knight (1999),

$$nQ_n \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - nQ_n (\beta_0) \to_{u-d} h'W_0 + \frac{1}{2} h'H_0 h$$

where $\to_{u-d}$ denotes convergence in distribution with respect to the topology of uniform convergence on compact sets. Chernoff Regularity implies that

$$\infty_1 \left( h \notin \sqrt{n}(C - \beta_0) \right) \overset{\delta}{\to} \infty_1 \left( h \notin T_C(\beta_0) \right)$$

where $\overset{\delta}{\to}$ denotes epigraphical convergence as defined in Geyer (1994), page 1997. Therefore, by Theorem 4 of Knight (1999),

$$nQ_n \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - nQ_n (\beta_0) + \infty_1 \left( h \notin \sqrt{n}(C - \beta_0) \right) \to_{e-d} h'W_0 + \frac{1}{2} h'H_0 h + \infty_1 \left( h \notin T_C(\beta_0) \right)$$

where $\to_{e-d}$ denotes epi-convergence in distribution as defined on page 5 of Knight (1999). Then by Theorem 1 of Knight (1999), whose conditions are satisfied because $h'W_0 + \frac{1}{2} h'H_0 h$ almost surely
has a unique minimizer over $T_C(\beta_0)$ due to $C$ being a closed set (see Proposition 4.2 and Theorem 4.4 of Geyer (1994)),

$$ \sqrt{n} \left( \hat{\beta}_n - \beta_0 \right) = \arg \min_{h \in \mathbb{R}^d} \left\{ n \hat{Q}_n \left( \beta_0 + \frac{h}{\sqrt{n}} \right), - n \hat{Q}_n (\beta_0) + \infty \right\} (h \neq \sqrt{n} (C - \beta_0)) $$

$$ \leadsto \arg \min_{h \in \mathbb{R}^d} \left\{ h'W_0 + \frac{1}{2} h' \bar{H}_0 h + \infty \right\} (h \neq T_C(\beta_0)) $$

For the proximal bootstrap, since $\sqrt{n} \alpha_n \to \infty$ and $\sqrt{n} \left( \hat{\beta}_n - \beta_0 \right) = O_P(1)$, $\frac{\beta^*_n - \beta_0}{\alpha_n} = o_P(1)$, where

$$ \frac{\beta^*_n - \beta_0}{\alpha_n} = \arg \min_{h \in \mathbb{R}^d} \left\{ \infty \left( h \neq \frac{C - \beta_0}{\alpha_n} \right) + \alpha_n \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right)' \left( \beta_0 - \beta_n + \alpha_n h \right) + \frac{1}{2} \left\| \beta_0 - \beta_n + \alpha_n h \right\|^2_{\bar{H}_n} \right\} $$

$$ = \arg \min_{h \in \mathbb{R}^d} \left\{ \infty \left( h \neq \frac{C - \beta_0}{\alpha_n} \right) + \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right)' \left( \beta_0 - \beta_n + \alpha_n + h \right) + \frac{1}{2} \left\| \beta_0 - \beta_n + h \right\|^2_{\bar{H}_n} \right\} $$

$$ = \arg \min_{h \in \mathbb{R}^d} \left\{ \infty \left( h \neq \frac{C - \beta_0}{\alpha_n} \right) + h' \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right) + \frac{1}{2} h' \bar{H}_n h + o_P(1) \right\} $$

Assumption 4 implies $\sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right)$ and $\sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right)$ have the same asymptotic distribution. Therefore,

$$ h' \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right) + \frac{1}{2} h' \bar{H}_n h \overset{\mathbb{P}}{\rightarrow} h'W_0 + \frac{1}{2} h' \bar{H}_0 h $$

A bootstrap in probability version of Theorem 4 of Knight (1999) can then be stated to show that

$$ \frac{\beta^*_n - \beta_0}{\alpha_n} \overset{\mathbb{P}}{\rightarrow} \left( \hat{\beta}_n - \hat{\beta}_n \right) + \frac{1}{2} h' \bar{H}_n h + \infty \left( h \neq \frac{C - \beta_0}{\alpha_n} \right) = O_P(1) $$

where $\overset{\mathbb{P}}{\rightarrow}$ denotes epi-convergence of the conditional law of $\hat{\beta}^*_n \overset{\mathbb{P}}{\rightarrow} \beta_0$, which can be equivalently stated as $d \left( \hat{\beta}^*_n, \beta \right) = o_P(1)$, where $d \left( \hat{\beta}^*_n, \beta \right) = \int_0^\infty \max \left\{ \left\| d_{\text{epi}} \hat{\beta}^*_n (v) - d_{\text{epi}} \beta_n (v) \right\| \right\} \exp (-\rho) d\rho$ is the epigraph of $G : \mathbb{R}^d \to \mathbb{R}$.

A modification of Theorem 1 of Knight (1999) to epi-convergence of conditional laws suggests that

$$ \frac{\beta^*_n - \beta_0}{\alpha_n} \overset{\mathbb{P}}{\rightarrow} \left( \hat{\beta}_n - \hat{\beta}_n \right) + o_P(1) \overset{\mathbb{P}}{\rightarrow} \arg \min_{h \in \mathbb{R}^d} \hat{G}_0 \left( \hat{\beta}_n \right) = \mathcal{J} $$

**Remark 1.** We can remove the assumption that $C$ is a closed set by assuming instead that $\mathcal{J} = \arg \min \left\{ h'W_0 + \frac{1}{2} h' \bar{H}_0 h \right\}$ is almost surely unique. This can happen for example if we strengthen the condition on $C$ to Clarke Regularity at $\beta_0$ (see Geyer (1994) page 1997 or Rockafellar et al. (1998) Definition 6.4 page 199 for a definition), which implies that $T_C(\beta_0)$ is a convex cone. Every
We will define the population analog of $C$ to be $C_0 = \{ \beta \in \mathbb{B} : f_{n_jj} (\beta) = 0 \text{ for } j \in \mathcal{E}, f_{n_jj} (\beta) \leq 0 \text{ for } j \in \mathcal{I} \}$, where $\sup_{\beta \in \mathcal{B}} | f_{n_jj} (\beta) - f_{ojo} (\beta) | = o_P(1)$ for all $j \in \mathcal{E} \cup \mathcal{I}$. The pseudo-true parameters are defined as $\hat{\beta}_0 = \arg \min_{\beta \in \mathcal{B}} Q(\beta)$, where $Q(\beta)$ is a lower semicontinuous function that is twice differentiable at $\beta_0$ and $\sup_{\beta \in \mathcal{B}} | \hat{Q}_n (\beta) - Q(\beta) | = o_P(1)$. The pseudo-true parameters differ from the true parameters when the constraints are violated at the unconstrained optimum. If the constraints are satisfied at the unconstrained optimum, then they are the same.

For simplicity, we will impose that the sample and population constraints satisfy Linear Independence Constraint Qualification (LICQ). LICQ is sufficient and necessary to ensure that the set of optimal Lagrange multipliers that satisfy the first order KKT conditions is a singleton (Wachsmuth (2013)). It is also possible to relax LICQ to Mangasarian-Fromovitz constraint qualification (MFCQ) as long as we impose the additional condition that there are unique optimal Lagrange multipliers.
MFCQ is weaker than LICQ because it does not require that the gradients of the equality constraints are linearly independent.

**Assumption 5.** Suppose Linear Independence Constraint Qualification (LICQ) holds at \( \beta_0 \): the gradients of the population equality constraints \( F_{0j} \) for \( j \in \mathcal{I} \) and the binding population inequality constraints \( F_{0j} \) for \( j \in \mathcal{I}^\ast \equiv \{ j \in \mathcal{I} : f_{0j}(\beta_0) = 0 \} \) are linearly independent. Furthermore, the gradients of the estimated equality constraints \( F_{nj} \) for \( j \in \mathcal{I}^\ast \equiv \{ j \in \mathcal{I} : f_{nj}(\beta_0) = 0 \} \) are almost surely linearly independent.

Instead of assumption 2, we now require that the Lagrangian has a uniform local quadratic expansion in \( \sqrt{n} \) neighborhoods of \( \beta_0 \). The importance of using the Lagrangian instead of the objective function is that it allows for the pseudo-true parameters to not be a solution of the unconstrained population optimization problem: \( I(\beta_0) \neq 0 \).

**Assumption 6.** Suppose \( f_{nj} : \mathbb{B} \to \mathbb{R} \) and \( f_{0j} : \mathbb{B} \to \mathbb{R} \) are twice continuously differentiable functions. Define \( \lambda_n \) to be the unique optimal Lagrange multipliers for \( \beta_n \). Define \( \bar{L}_n(\beta) = \bar{Q}_n(\beta) + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_n f_{nj}(\beta), F_{nj}(\beta_0) = \frac{\partial f_{nj}(\beta)}{\partial \beta} \bigg|_{\beta=\beta_0}, F_{0j} = \frac{\partial f_{0j}(\beta)}{\partial \beta} \bigg|_{\beta=\beta_0} \), and \( G_{0j} = \frac{\partial^2 f_{0j}(\beta)}{\partial \beta^2} \bigg|_{\beta=\beta_0} \).

For any \( \delta_n \to 0 \),

\[
\sup_{\|h\| \leq \delta_n} \frac{n\bar{L}_n(\beta_0 + \frac{h}{\sqrt{n}}) - n\bar{L}_n(\beta_0) - h' \sqrt{n}\hat{\lambda}_n(\beta_0) - \frac{1}{2} h' H_0 h - \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} \left( \sqrt{n} F_{nj}(\beta_0)' h + \frac{1}{2} h' G_{0j} h \right)}{1 + \|h\|^2} = o_P(1)
\]

where \( \lambda_0 \) are the unique Lagrange multipliers that satisfy \( \lambda_{0j} f_{0j}(\beta_0) = 0 \) for all \( j \in \mathcal{E} \cup \mathcal{I} \), \( 0 \leq \lambda_{0j} < \infty \) for all \( j \in \mathcal{E} \cup \mathcal{I} \), and \( \nabla L(\beta_0, \lambda_0) = I(\beta_0) + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} F_{0j} = 0 \).

Next, we define the proximal bootstrap estimator. Let \( f_{nj}^*(\beta) \) be the bootstrap analog of \( f_{nj}(\beta) \) and let \( F_{nj}^*(\beta) = \frac{\partial f_{nj}^*(\beta)}{\partial \beta} \). For any \( \beta_n \) such that \( \sqrt{n}(\beta_n - \beta_0) = O_P(1) \), let \( \tilde{F}_{nj} = F_{nj}(\beta_n), F_{nj}^* = F_{nj}^*(\beta_n), G_{nj} = \frac{\partial^2 f_{0j}(\beta)}{\partial \beta^2} \bigg|_{\beta=\bar{\beta}_n} \) for all \( j \), and let \( \bar{\lambda}_n \) be the unique optimal Lagrange multipliers for \( \bar{\beta}_n \). Define \( \hat{\beta}_n^* = \arg \min_{\beta \in C^*} \hat{A}_n^*(\beta) \), for

\[
\hat{A}_n^*(\beta) = \alpha_n \sqrt{n} \left\langle \hat{\theta}_n(\beta_n) - \hat{\theta}_n(\bar{\beta}_n) \right\rangle + \frac{1}{2} \| \beta - \bar{\beta}_n \|_{H_n}^2
+ \sum_{j \in \mathcal{E} \cup \mathcal{I}} \hat{\lambda}_{nj} \left( \alpha_n \sqrt{n} \left( F_{nj}^* - F_{nj} \right)' (\beta - \bar{\beta}_n) + \frac{1}{2} \| \beta - \bar{\beta}_n \|_{G_{nj}}^2 \right)
\]

\( C^* = \{ \beta \in \mathbb{B} : f_{nj}(\beta_n) + F_{nj}'(\beta - \bar{\beta}_n) + \alpha_n \sqrt{n} \left( f_{nj}^*(\bar{\beta}_n) - f_{nj}(\bar{\beta}_n) \right) = 0 \text{ for } j \in \mathcal{E}, f_{nj}(\beta_n) + F_{nj}'(\beta - \bar{\beta}_n) + \alpha_n \sqrt{n} \left( f_{nj}^*(\bar{\beta}_n) - f_{nj}(\bar{\beta}_n) \right) \leq 0 \text{ for } j \in \mathcal{I} \} \)

We now demonstrate consistency of the proximal bootstrap.

**Theorem 2.** Suppose Assumptions 3-4 and 5 - 6 are satisfied in addition to the following:

(i) Suppose \( \hat{\beta}_n \to \beta_0 \), where \( \beta_0 = \arg \min_{\beta \in C^*} Q(\beta) \) is unique.

(ii) Suppose \( \hat{\beta}_n^* - \hat{\beta}_0 = o_P(1) \).
(iii) Suppose \( \nabla^2 \mathcal{L}(\beta_0, \lambda_0) = H_0 + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} G_{0j} \) is positive definite.

(iv) Suppose \( \sqrt{n} \left( f_n(\beta_0) - f_0(\beta_0) \right) \overset{p}{\longrightarrow} U_0, \) a tight random vector, and \( \sqrt{n} \left( \hat{t}_n(\beta_0) - l(\beta_0) \right) + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} \sqrt{n} \left( F_{nj}(\beta_0) - F_{0j}(\beta_0) \right) \overset{p}{\longrightarrow} W_0 + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} V_{0j} \), a tight random vector.

(v) Suppose \( \sqrt{n} \left( f_n^*(\beta_0) - f_n(\beta_0) \right) \overset{P}{\longrightarrow} U_0, \) \( \max_{j \in \mathcal{E} \cup \mathcal{I}} |\lambda_{nj} - \lambda_{0j}| \overset{P}{\longrightarrow} 0, \)

\[
\sqrt{n} \left( \hat{t}_n(\beta_0) - \hat{t}_n(\beta_0) \right) + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} \sqrt{n} \left( F_n^*(\beta_0) - F_n(\beta_0) \right) \overset{P}{\longrightarrow} W_0 + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} V_{0j},
\]

\[
\sup_{\|\beta - \beta_0\| \leq o(1)} \sqrt{n} \left( f_n^*(\beta) - f_n(\beta) - f_n^*(\beta_0) + f_n(\beta_0) \right) = o_p(1), \quad \text{and}
\]

\[
\sup_{\|\beta - \beta_0\| \leq o(1)} \sqrt{n} \left( F_n^*(\beta) - F_n(\beta) - F_n^*(\beta_0) + F_n(\beta_0) \right) = o_p(1).
\]

(vi) \( \bar{H}_n \overset{P}{\longrightarrow} H_0, \) \( \max_{j \in \mathcal{E} \cup \mathcal{I}} \left| G_{nj} - G_{0j} \right| \overset{P}{\longrightarrow} 0, \) and \( \bar{H}_n + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{nj} G_{nj} \) is symmetric and positive definite.

Then, for any sequence \( \alpha_n \) such that \( \alpha_n \rightarrow 0 \) and \( \sqrt{n} \alpha_n \rightarrow \infty, \) \( \sqrt{n} \left( \hat{\beta}_n - \beta_0 \right) \overset{P}{\longrightarrow} \mathcal{J} \) and \( \frac{\hat{\beta}_n - \beta_0}{\alpha_n} \overset{P}{\longrightarrow} \mathcal{J} \), where for \( \mathcal{I}^*_n(\lambda_0) = \{ j \in \mathcal{I}^*: \lambda_{0j} > 0 \} \) and \( \mathcal{I}^*_0(\lambda_0) = \{ j \in \mathcal{I}^*: \lambda_{0j} = 0 \}, \)

\[
\mathcal{J} = \arg \min_{h \in \Sigma} \left\{ h'W_0 + \frac{1}{2} h'H_0 h + \sum_{j \in \mathcal{E} \cup \mathcal{I}^*_n(\lambda_0)} \lambda_{0j} \left( h'V_{0j} + \frac{1}{2} h'G_{0j} h \right) \right\}
\]

\[
\Sigma = \{ h : U_{0j} + F_{0j}' h = 0 \text{ for } j \in \mathcal{E} \cup \mathcal{I}^*_n(\lambda_0), U_{0j} + F_{0j}' h > 0 \text{ for } j \in \mathcal{I}^*_0(\lambda_0) \}
\]

Proof: We can show that consistency implies \( \sqrt{n} \)-consistency using a modified version of the first part of the proof of Theorem 5 on page 141 of Pollard (1984) for allowed for estimated constraints. We need to constrain \( \hat{\beta}_n \) to lie in \( C \) and replace the population objective \( F(\cdot) \) with the population Lagrangian \( \mathcal{L}(\beta_0, \lambda_0) = Q(\beta_0) + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} f_{0j}(\beta_0) \) and the empirical process \( E_n \Delta \) with \( \sqrt{n} \left( \hat{t}_n(\beta_0) - l(\beta_0) \right) + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} \sqrt{n} \left( F_{nj}(\beta_0) - F_{0j}(\beta_0) \right) \). The first order KKT condition \( \nabla \mathcal{L}(\beta_0, \lambda_0) = l(\beta_0) + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} F_{0j}(\beta_0) = 0 \) and positive-definiteness of \( \nabla^2 \mathcal{L}(\beta_0, \lambda_0) = H_0 + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} G_{0j} \) imply the local quadratic expansion \( \mathcal{L}(\beta, \lambda_0) = \mathcal{L}(\beta_0, \lambda_0) + \frac{1}{2} (\beta - \beta_0, H_0, \lambda_0) + o \left( \|\beta - \beta_0\|^2 \right) \) for \( \beta \) in a small neighborhood of \( \beta_0 \). The rest of the arguments are the same as Pollard (1984).

Recall \( \hat{\mathcal{L}}_n(\beta) = \hat{Q}_n(\beta) + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{nj} f_{nj}(\beta) \) is the sample Lagrangian evaluated at the optimal Lagrange multipliers \( \lambda_n \) for \( \hat{\beta}_n \). It is well known that \( \hat{\beta}_n = \arg \min_{\beta \in C} \hat{\mathcal{L}}_n(\beta) \) can be equivalently expressed as \( \hat{\beta}_n = \arg \min_{\beta \in C} \hat{\mathcal{L}}_n(\beta) \) when the first order KKT conditions are satisfied. Shapiro (1990) shows that it is important to use this Lagrangian formulation when deriving the asymptotic distribution of \( \hat{\beta}_n \) because it captures the sampling variation in the objective as well as the estimated constraints.

Additionally, LICQ (and also the weaker Mangasarian-Fromovitz and Abadie Constraint Qualifications) implies that the linearization of the constraint set is sufficient to capture the geometry of the constraints near \( \beta_0 \) (Nocedal and Wright (2006) chapter 12); in particular, the tangent cone of \( C \) at \( \beta_0 \) is equal to the linearized feasible direction set at \( \beta_0 \). We can then use this linearized constraint set to derive the asymptotic distribution of \( \sqrt{n} \left( \hat{\beta}_n - \beta_0 \right) \). Denote the feasible direction
implies uniform convergence over compact sets $\lambda$

Recall $?$

Lagrangian over $\n$

as a process indexed by $\n$

Denote the linearized feasible direction set by $\n$

Therefore, uniformly in $\n$

$h$

Therefore, uniformly in $\n$

$n$

Therefore, uniformly in $\n$

$h$

where the last line follows from the following arguments. First note that assumption 6 in combination with $\n$

with $\n$

implies that for any $\n$

Therefore, uniformly in $\n$

Recall $\n$

and

$h$

as a process indexed by $\n$ in the space of bounded functions $\ell^\infty(K)$ for any compact $K \subset \mathbb{R}^d$. 
Note that since \( \sqrt{n} f_{nj}(\beta_0) + F_{nj}(\beta_0)' h \xrightarrow{p} -\infty \) for \( j \in \mathcal{I} \setminus \mathcal{I}^* \), the nonbinding inequality constraints do not affect the asymptotic distribution under our pointwise asymptotics. Since \( \sqrt{n} f_{nj}(\beta_0) \xrightarrow{p} U_{0j} \), jointly, for all \( j \in \mathcal{E} \cup \mathcal{I}^* \), \( F_n(\beta_0) = F_0 + o_P(1) \), and finite dimensional convergence in distribution implies epi-convergence in distribution for convex functions,

\[
\varnothing 1(h \notin \Sigma_n) \rightarrow_{e-d} \varnothing 1 \left\{ h : U_{0j} + F_{0j}' h = 0 \text{ for } j \in \mathcal{E}, U_{0j} + F_{0j}' h \leq 0 \text{ for } j \in \mathcal{I}^* \right\}
\]

Because we have assumed LICQ at \( \beta_0 \), Theorem 2.1 of Shapiro (1988) implies that minimizing over \( \left\{ h : U_{0j} + F_{0j}' h = 0 \text{ for } j \in \mathcal{E}, U_{0j} + F_{0j}' h \leq 0 \text{ for } j \in \mathcal{I}^* \right\} \) will produce the same set of solutions as minimizing over \( \mathcal{I} = \left\{ h : U_{0j} + F_{0j}' h = 0 \text{ for } j \in \mathcal{E} \cup \mathcal{I}^* (\lambda_0), U_{0j} + F_{0j}' h \leq 0 \text{ for } j \in \mathcal{I}_0^* (\lambda_0) \right\} \). Condition (iii) is a second order sufficient condition and guarantees that the argmin in \( \mathcal{J} \) is unique. Then by the argmin continuous mapping theorem (Theorem 1 of Knight (1999)), \( \arg \min_h \hat{G}_n(h) \rightarrow_{e-d} \arg \min_h G_0(h) \), where

\[
\hat{G}_n(h) = n \hat{Q}_n \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - n \hat{Q}_n(\beta_0) + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{nj}(f_{nj}(\beta_0 + \frac{h}{\sqrt{n}}) - f_{nj}(\beta_0)) + \varnothing 1(h \notin \Sigma_n)
\]

\[
G_0(h) = h' W_0 + \frac{1}{2} h' H_0 h + \sum_{j \in \mathcal{E} \cup \mathcal{I}_0^* (\lambda_0)} \lambda_{0j} \left( h' W_{0j} + \frac{1}{2} h' G_{0j} h \right) + \varnothing 1(h \notin \Sigma)
\]

Next we show consistency of the proximal bootstrap. Note that since \( C^* \) is already a linearized constraint set, the linearized feasible direction set is simply

\[
\Sigma_n^* = \left\{ h : f_{nj}(\beta_n) + \hat{F}_{nj}'(\beta_n - \beta_n + \alpha_n h) + \alpha_n \sqrt{n} \left( f_{nj}(\beta_n) - f_{nj}(\beta_n) \right) = 0 \text{ for } j \in \mathcal{E} \right\}
\]

\[
\left\{ h : f_{nj}(\beta_n) + \hat{F}_{nj}'(\beta_n - \beta_n + \alpha_n h) + \alpha_n \sqrt{n} \left( f_{nj}(\beta_n) - f_{nj}(\beta_n) \right) \leq 0 \text{ for } j \in \mathcal{I} \right\}
\]

\[
= \left\{ h : \frac{f_{nj}(\beta_n)}{\alpha_n} + \hat{F}_{nj}' h + \sqrt{n} \left( f_{nj}(\beta_n) - f_{nj}(\beta_n) \right) + \hat{F}_{nj}' \left( \frac{\beta_n - \beta_n}{\alpha_n} \right) = 0 \text{ for } j \in \mathcal{E}, \right\}
\]

\[
\left\{ h : \frac{f_{nj}(\beta_n)}{\alpha_n} + \hat{F}_{nj}' h + \sqrt{n} \left( f_{nj}(\beta_n) - f_{nj}(\beta_n) \right) + \hat{F}_{nj}' \left( \frac{\beta_n - \beta_n}{\alpha_n} \right) \leq 0 \text{ for } j \in \mathcal{I} \right\}
\]

Note that \( \frac{f_{nj}(\beta_n)}{\alpha_n} \xrightarrow{p} -\infty \) for \( j \in \mathcal{I} \setminus \mathcal{I}^* \) while \( \frac{f_{nj}(\beta_n)}{\alpha_n} \sim \frac{\sqrt{n} (f_{nj}(\beta_n) - f_{nj}(\beta_0))}{\sqrt{n} \alpha_n} \) and \( \sqrt{n} (f_{nj}(\beta_0) - f_{nj}(\beta_0)) \) is \( o_P(1) \) for \( j \in \mathcal{E} \cup \mathcal{I}^* \). Additionally, \( \hat{F}_{nj}' \left( \frac{\beta_n - \beta_n}{\alpha_n} \right) = o_P(1), \hat{F}_n = F_0 + o_P(1), \right\}

\[
\sqrt{n} \left( f_{nj}(\beta_0) - f_{nj}(\beta_0) \right) \xrightarrow{p} U_{0j}, \text{ jointly, for all } j \in \mathcal{E} \cup \mathcal{I}^*, \text{ and } \sqrt{n} \left( f_{nj}(\beta_n) - f_{nj}(\beta_n) \right) \xrightarrow{p} U_{0j}, \text{ jointly, for all } j \in \mathcal{E} \cup \mathcal{I}^* \text{ because } \sup_{|\beta - \beta_n| \leq o(1)} \sqrt{n} \left( f_{nj}(\beta) - f_{nj}(\beta_0) \right) = o_P(1).
\]

Therefore,

\[
\varnothing 1(h \notin \Sigma_n^*) \rightarrow_{e-d} \varnothing 1 \left\{ h : U_{0j} + F_{0j}' h = 0 \text{ for } j \in \mathcal{E}, U_{0j} + F_{0j}' h \leq 0 \text{ for } j \in \mathcal{I}^* \right\}
\]

Next, we can center and scale the bootstrap estimator to get

\[
\frac{\hat{\beta}_n^* - \beta_0}{\alpha_n} = \arg \min_{h \in \Sigma_n^*} \left\{ \alpha_n \sqrt{n} \left( \hat{t}_n (\beta_n) - \hat{t}_n (\beta_n) \right)' (\beta_0 - \beta_n + \alpha_n h) + \frac{1}{2} \| \beta_0 - \beta_n + \alpha_n h \|^2_{H_n} \right\}
\]

\[
+ \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{nj} \left( \alpha_n \sqrt{n} \left( \hat{F}_{nj}' - \hat{F}_{nj}' \right)' (\beta_0 - \beta_n + \alpha_n h) + \frac{1}{2} \| \beta_0 - \beta_n + \alpha_n h \|^2_{G_{nj}} \right) \}.
\]
\[
\begin{align*}
= \arg \min_{h \in \Sigma_n} & \left\{ \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right)' \left( \frac{\alpha_n}{\beta_0 - \beta_n} + h \right) + \frac{1}{2} \left\| \frac{\beta_0 - \beta_n}{\alpha_n} + h \right\|_{H_n}^2 \right\} \\
& + \sum_{j \in \mathcal{I}} \lambda_{n_j} \left( \sqrt{n} \left( \hat{F}_{nj}^* - \bar{F}_{nj} \right)' \left( \frac{\alpha_n}{\beta_0 - \beta_n} + h \right) + \frac{1}{2} \left\| \frac{\beta_0 - \beta_n}{\alpha_n} + h \right\|_{G_{nj}}^2 \right) \\
= \arg \min_{h \in \Sigma_n} & \left\{ h' \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right) + \frac{1}{2} h' H_n h \right. \\
& + \sum_{j \in \mathcal{I}} \lambda_{n_j} \left( h' \sqrt{n} \left( \hat{F}_{nj}^* - \bar{F}_{nj} \right) + \frac{1}{2} h' G_{nj} h \right) + o_P(1) \left( \lambda_0 \right) \right\} = \mathcal{J}
\end{align*}
\]

where the last line follows from the following arguments. First, note that since \( \tilde{H}_n \xrightarrow{p} H_0, \tilde{G}_{nj} \xrightarrow{p} G_{0j} \) for all \( j \), and the proximal bootstrap Lagrangian is convex in \( h \), we have that uniformly in \( h \),

\[
\begin{align*}
h' \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right) + \frac{1}{2} h' H_n h + \sum_{j \in \mathcal{I}} \lambda_{n_j} \left( h' \sqrt{n} \left( \hat{F}_{nj}^* - \bar{F}_{nj} \right) + \frac{1}{2} h' G_{nj} h \right)
& = h' \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right) + \frac{1}{2} h' H_0 h + \sum_{j \in \mathcal{I}} \lambda_{n_j} \left( h' \sqrt{n} \left( \hat{F}_{nj}^* - \bar{F}_{nj} \right) + \frac{1}{2} h' G_{0j} h \right) + o_P(1)
\end{align*}
\]

Next, note that assumption 4. \( \max_{j \in \mathcal{I}} |\lambda_{nj} - \lambda_0| \xrightarrow{p} 0 \), and \( \sup_{|\beta_0 - \beta_n| \leq o(1)} \sqrt{n} \left( F_{nj}^* (\beta) - F_n (\beta) - F_{nj}^* (\beta_0) + F_n (\beta_0) \right) = o_P(1) \) imply \( \sum_{j \in \mathcal{I}} \lambda_{nj} \sqrt{n} \left( \hat{F}_{nj}^* - \bar{F}_{nj} \right) \xrightarrow{p} W_0 + \sum_{j \in \mathcal{I}} \lambda_0 V_0j \) because

\[
\begin{align*}
& \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right) + \sum_{j \in \mathcal{I}} \lambda_{nj} \sqrt{n} \left( \hat{F}_{nj}^* - \bar{F}_{nj} \right) \\
& = \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right) + \sum_{j \in \mathcal{I}} \lambda_{nj} \sqrt{n} \left( F_{nj}^* (\beta) - F_n (\beta) \right) + \sum_{j \in \mathcal{I}} \left( \lambda_{nj} - \lambda_0 \right) \sqrt{n} \left( \hat{F}_{nj}^* - \bar{F}_{nj} \right)
\end{align*}
\]

and we assumed \( \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right) + \sum_{j \in \mathcal{I}} \lambda_0 \sqrt{n} \left( F_{nj}^* (\beta_0) - F_n (\beta_0) \right) \xrightarrow{p} W_0 + \sum_{j \in \mathcal{I}} \lambda_0 V_0j \).

Additionally, \( \max_{j \in \mathcal{I}} |G_{nj} - G_{0j}| \xrightarrow{p} 0 \) and \( \max_{j \in \mathcal{I}} |\lambda_{nj} - \lambda_0| \xrightarrow{p} 0 \) imply that \( \sum_{j \in \mathcal{I}} \lambda_{nj} G_{nj} \xrightarrow{p} \sum_{j \in \mathcal{I}} \lambda_{0j} G_{0j} \). By convexity of the bootstrap Lagrangian in \( h \), pointwise convergence implies uniform convergence over compact sets \( K \subset \mathbb{R}^d \); therefore,

\[
\begin{align*}
h' \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right) + \frac{1}{2} h' H_0 h + \sum_{j \in \mathcal{I}} \lambda_{n_j} \left( h' \sqrt{n} \left( \hat{F}_{nj}^* - \bar{F}_{nj} \right) + \frac{1}{2} h' G_{0j} h \right)
\end{align*}
\]
\[
\begin{align*}
\frac{\partial}{\partial \varepsilon} h'W_0 + \frac{1}{2} h'H_0 h + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} \left( h'V_{0j} + \frac{1}{2} h'G_{0j} h \right) \\
= h'W_0 + \frac{1}{2} h'H_0 h + \sum_{j \in \mathcal{E} \cup \mathcal{I}^* (\lambda_0)} \lambda_{0j} \left( h'V_{0j} + \frac{1}{2} h'G_{0j} h \right)
\end{align*}
\]

as a process indexed by \( h \) in the space of bounded functions \( \ell^\infty (K) \) for any compact \( K \subset \mathbb{R}^d \).

Finally, note that \( \hat{\beta}_n^* \) is unique because \( \check{H}_n + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{nj} \tilde{G}_{nj} \) is symmetric and positive definite. Then, by a modification of the bootstrap argmin continuous mapping lemma 14.2 in Hong and Li (2020) that replaces weak convergence with epi-convergence, \( \arg \min_{h} \tilde{G}_n^* (h) \overset{\mathcal{D}}{\longrightarrow} \arg \min_{h} \tilde{G}_0^* (h) \) for

\[
\tilde{G}_n^* (h) = h'\sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right) + \frac{1}{2} h'H_n h + \sum_{j \in \mathcal{E} \cup \mathcal{I}^* (\lambda_0)} \lambda_{0j} \left( h'V_{0j} + \frac{1}{2} h'G_{0j} h \right) + \epsilon 1 (h \notin \Sigma^*)
\]

Remark 4. If \( l (\beta_0) = 0 \), which is implied by \( Q (\beta) = Q (\beta_0) + \frac{1}{2} (\beta - \beta_0)' H_0 (\beta - \beta_0) + o \left( \| \beta - \beta_0 \|^2 \right) \), then \( \mathcal{J} \) reduces down to

\[
\mathcal{J} = \arg \min_{h \in \Sigma} \left\{ h'W_0 + \frac{1}{2} h'H_0 h \right\}
\]

\[
\Sigma = \{ h : U_{0j} + F'_{0j} h = 0 \text{ for } j \in \mathcal{E}, U_{0j} + F'_{0j} h \leq 0 \text{ for } j \in \mathcal{I}^* (\lambda_0) \}
\]

This is because by the KKT conditions, \( \lambda_0 \) satisfies \( l (\beta_0) + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} F_{0j} = 0 \), so if \( l (\beta_0) = 0 \), then \( \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} F_{0j} = 0 \). By LICQ, the binding constraint gradients \( F_{0j} \) for \( j \in \mathcal{E} \cup \mathcal{I}^* \) are all nonzero, and furthermore, the optimal Lagrange multipliers for the nonbinding inequality constraints \( j \in \mathcal{I} \setminus \mathcal{I}^* \) are zero by the complementary slackness conditions \( \lambda_{0j} f_{0j} (\beta_0) = 0 \) for all \( j \in \mathcal{E} \cup \mathcal{I} \). Therefore, \( \lambda_{0j} = 0 \) for all \( j \in \mathcal{E} \cup \mathcal{I} \) is a solution to \( \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} F_{0j} = 0 \). Since the set of Lagrange multipliers that satisfy the KKT conditions is a singleton under LICQ, \( \lambda_{0j} = 0 \) for all \( j \in \mathcal{E} \cup \mathcal{I} \) are the unique optimal Lagrange multipliers, which implies \( \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} (h'V_{0j} + \frac{1}{2} h'G_{0j} h) = 0, \mathcal{I}^* (\lambda_0) = \varnothing \), and \( \mathcal{I}^* = \mathcal{I}^*_0 (\lambda_0) \).

In this case, it is easy to extend our theory to the case where the number of constraints is growing with \( n \), assuming that the dimension of \( \beta \) is fixed. We redefine the proximal bootstrap estimator as \( \hat{\beta}_n^* = \arg \min_{\beta \in C^*} \hat{A}_n^* (\beta) \), where

\[
\hat{A}_n^* (\beta) = \alpha_n \sqrt{n} \left( \hat{\beta}_n - \hat{\beta}_n \right) + \frac{1}{2} \left\| \beta - \hat{\beta}_n \right\|^2_{H_n}
\]

\[
C^* = \{ \beta \in \mathbb{R} : f_{nj} (\hat{\beta}_n) + F'_{nj} (\beta - \hat{\beta}_n) + \alpha_n \sqrt{n} \left( f_{nj}^* (\hat{\beta}_n) - f_{nj} (\hat{\beta}_n) \right) = 0 \text{ for } j \in \mathcal{E}_n,
\]

\[
f_{nj} (\beta_0) + F'_{nj} (\beta - \beta_0) + \alpha_n \sqrt{n} \left( f_{nj}^* (\beta_0) - f_{nj} (\hat{\beta}_n) \right) \leq 0 \text{ for } j \in \mathcal{I}^*_n \}
\]

Here, \( \mathcal{I}^*_n = \{ j \in \mathcal{I} : f_{0j} (\beta_0) = 0 \} \), and \( \Sigma_n \) and \( \Sigma^*_n \) are the same as in the proof of Theorem 2 except allowing for \( \mathcal{E}_n \) and \( \mathcal{I}_n \) to depend on \( n \). The limiting distribution of \( \sqrt{n} (\hat{\beta}_n - \beta_0) \) can be difficult
to characterize due to the presence of an infinite number of constraints in the limit as \( n \to \infty \).

To avoid explicitly characterizing the limiting distribution, we will work with the following finite constraint set \( \Sigma \):

\[
\Sigma = \{ h : U_{0j} + F'_{0j}h = 0 \text{ for } j \in \mathcal{E}_n, U_{0j} + F'_{0j}h \leq 0 \text{ for } j \in \mathcal{I}_n^* \}
\]

To demonstrate consistency of the proximal bootstrap, we will show that both \( \infty 1 (h \notin \Sigma_n) \) and \( \infty 1 (h \notin \Sigma_n^*) \) have the same limit (in the sense of epi-convergence in distribution) without explicitly characterizing this limit. Because \( \infty 1 (h \notin \Sigma_n) \) and \( \infty 1 (h \notin \Sigma_n^*) \) are convex functions, to show epi-convergence in distribution, it suffices to show finite dimensional convergence. In particular, we will show that \( \infty 1 (h \notin \Sigma_n^*) - \infty 1 (h \notin \Sigma) \) and \( \infty 1 (h \notin \Sigma_n^*) - \infty 1 (h \notin \Sigma) \) both converge weakly to zero.

To do so, we will need to assume

\[
\sup_{te\mathbb{R}} \left| P \left( \max_{j \in \mathcal{E}_n \cup \mathcal{I}_n^*} \sqrt{n} f_{nj} (\beta_0) \leq t \right) - P \left( \max_{j \in \mathcal{E}_n \cup \mathcal{I}_n^*} U_{0j} \leq t \right) \right| \to 0, \quad \text{and}
\sup_{te\mathbb{R}} \left| P \left( \max_{j \in \mathcal{E}_n \cup \mathcal{I}_n^*} \sqrt{n} f_{nj} (\beta_n) \leq t \right) - P \left( \max_{j \in \mathcal{E}_n \cup \mathcal{I}_n^*} U_{0j} \leq t \right) \right| \overset{p}{\to} 0.
\]

These assumptions can be derived using the results in Chernozhukov et al. (2013) and Chernozhukov et al. (2019) for Gaussian approximation of maxima of sums for high dimensional random vectors. We will also need to assume \( \max_{j \in \mathcal{E}_n \cup \mathcal{I}_n^*} |F_{nj} (\beta) - F_{0j}| = o_p(1) \) and \( \max_{j \in \mathcal{E}_n \cup \mathcal{I}_n^*} |F_{nj} - F_{0j}| = o_p(1) \).

For any \( h_1, \ldots, h_k \) where \( k \) is fixed,

\[
P \left( h_1 \in \Sigma_n, \ldots, h_k \in \Sigma_n \right)
= P \left( \bigcap_{i=1}^{k} \{ \sqrt{n} f_{nj} (\beta_0) + F_{nj} (\beta_0) \leq 0 \text{ for } j \in \mathcal{E}_n \} \bigcap \{ \sqrt{n} f_{nj} (\beta_0) \geq 0 \text{ for } j \in \mathcal{I}_n^* \} \right)
\]

\[
= P \left( \bigcap_{i=1}^{k} \{ \sqrt{n} f_{nj} (\beta_0) + F_{nj} (\beta_0) \leq 0 \text{ for } j \in \mathcal{E}_n \} \right)
= P \left( \bigcap_{i=1}^{k} \{ \sqrt{n} f_{nj} (\beta_0) + F_{nj} (\beta_0) \leq 0 \text{ for } j \in \mathcal{E}_n \} \right)
\]

where we have used \( \sup_{1 \leq i \leq k} \sqrt{n} f_{nj} (\beta_0) \overset{p}{\to} -\infty \) for \( j \in \mathcal{I}_n \backslash \mathcal{I}_n^* \), \( \max_{1 \leq i \leq k} |F_{nj} (\beta) - F_{0j}| = o_p(1) \), and \( \max_{j \in \mathcal{E}_n \cup \mathcal{I}_n^*} \left| \frac{\sqrt{n} f_{nj} (\beta_0)}{U_{0j}} \right| = o(1) \).
are the same as in Theorem 2. It follows that for \( c_{1-\alpha} \) the \( 1-\alpha \) quantile of \( \mathcal{J} = \arg \min_{h \in \Sigma} \{ h'W_0 + \frac{1}{2} h'H_0 h \} \),
\[
P \left( \sqrt{n} \left( \hat{\beta}_n - \beta_0 \right) > c_{1-\alpha} \right) \to \alpha.
\]
Similarly, to show finite dimensional convergence in probability of \( \infty (h \notin \Sigma^*_n) \) to \( \infty (h \notin \Sigma) \),
\[
P \left( h_1 \in \Sigma^*_n, \ldots, h_k \in \Sigma^*_n | \mathcal{X}_n \right) = P \left( h_1 \in \Sigma, \ldots, h_k \in \Sigma \right) = P \left( h_1 \in \Sigma^*_n, \ldots, h_k \in \Sigma^*_n | \mathcal{X}_n \right) - P \left( h_1 \in \Sigma, \ldots, h_k \in \Sigma \right)
\]
where we have used \( \bar{F}_{n,j} (\beta_0 - \hat{\beta}_n) = o_P(1), \) \( f_{nj}(\beta_0 - \hat{\beta}_n) = o_P(1), \) \( \frac{\max f_{nj}(\beta_0 - \hat{\beta}_n)_{1 \leq i \leq k}}{\alpha_n} = o_P(1) \) for all \( j \in \mathcal{E}_n \cup \mathcal{I}_n^*, \) \( \sup_{h \in \mathcal{E}_n} \left| P \left( \max_{j_{E_n} \cup \mathcal{I}_n^*} \sqrt{n} (f_{nj}(\beta_0 - \hat{\beta}_n)_{1 \leq i \leq k}) \leq t \right) \right| \to 0, \) and \( \max_{j_{E_n} \cup \mathcal{I}_n^*} |\bar{F}_{n,j} - F_{0,j}| = o_P(1). \) The rest of the arguments are the same as in Theorem 2. It follows that for \( c_{1-\alpha} \) the \( 1-\alpha \) quantile of \( \mathcal{J} = \arg \min_{h \in \Sigma} \{ h'W_0 + \frac{1}{2} h'H_0 h \} \), \( P \left( \alpha^{-1}_n (\hat{\beta}_n - \beta_0) > c_{1-\alpha} \middle| \mathcal{X}_n \right) \to \alpha. \) Since we showed \( P \left( \sqrt{n} (\hat{\beta}_n - \beta_0) > c_{1-\alpha} \right) \to \alpha, \) it follows that \( P \left( \sqrt{n} (\hat{\beta}_n - \beta_0) > c^b_{1-\alpha} \right) \to \alpha, \) where \( c^b_{1-\alpha} \) is the \( 1-\alpha \) empirical quantile of \( \alpha^{-1}_n (\hat{\beta}_n - \beta_0) \).

**Remark 5.** If there are only equality constraints \( f_{nj}(\beta) = 0, \) then the asymptotic distribution reduces down to \( \mathcal{J} = \arg \min_{h \in \Sigma} \{ h'W_0 + \frac{1}{2} h'H_0 h : \Sigma = \{ h : U_{0j} + F'_{0j} h = 0 \} \} \) for \( \Sigma = \{ h : U_{0j} + F'_{0j} h = 0 \} \) for \( j \in \mathcal{E} \). The reason is that \( h \in \Sigma_n = \{ h : \sqrt{n} f_{nj}(\beta_0) + F_{nj}(\beta_0)' h = 0 \} \) for \( j \in \mathcal{E} \) implies \( F_{nj}(\beta_0)' h = -\sqrt{n} f_{nj}(\beta_0), \) so
\[
\arg \min_{h \in \Sigma_n} \left\{ \frac{h}{{\hat{\lambda}}_n (\beta_0) + \frac{h}{{\sqrt{n}}} - {\hat{\lambda}}_n (\beta_0)} \right\}
\]

\[
= \arg \min_{h \in \Sigma_n} \left\{ \frac{h}{{\hat{\lambda}}_n (\beta_0) + \frac{h}{{\sqrt{n}}} - {\hat{\lambda}}_n (\beta_0)} - \sum_{j \in \mathcal{E}} \lambda_{0j} \left( \sqrt{n} F_{nj}(\beta_0)' h + \frac{1}{2} h' G_{0j} h \right) + o_P(1) \right\}
\]

15
where the last line follows from standard arguments in Amemiya (1985) section 1.4.1 or Newey and McFadden (1994) section 9.1 (for clarity Lemma 5.1 repeats these arguments). If \( W_0 \) and \( U_0 \) are multivariate normal, then the asymptotic distribution will be multivariate normal.

If \( l(\beta_0) = 0 \) or if the constraints are linear, then \( \sum_{j \in E} \lambda_{0j} G_{0j} = 0 \) and \( B_0 = H_0, \) so \( \mathcal{J} = -H_0^{-1} \left( I - F_0 \left( F_0'H_0^{-1} F_0 \right)^{-1} \right) W_0 - F_0'B_0^{-1} \left( F_0'B_0^{-1} F_0 \right)^{-1} U_0. \)

**Remark 6.** If strict complementarity holds, meaning \( \lambda_{0j} > 0 \) whenever \( f_{0j}(\beta_0) = 0, \) then \( \mathcal{T}^* = \mathcal{T}_{+}^* (\lambda_0) \) and the asymptotic distribution reduces down to

\[
\mathcal{J} = \arg \min_{h \in \Sigma} \left\{ h'W_0 + \frac{1}{2} h'H_0 h + \sum_{j \in \mathcal{E} \cup \mathcal{T}_{+}^* (\lambda_0)} \lambda_{0j} \left( h' V_{0j} + \frac{1}{2} h' G_{0j} h \right) \right\}
\]

for \( \Sigma = \{ h : U_{0j} + F_{0j}' h = 0 \text{ for } j \in \mathcal{E} \cup \mathcal{T}_{+}^* (\lambda_0) \}. \) Just like in the previous remark, we can express

\[
\mathcal{J} = -B_0^{-1} \left( I - F_0 \left( F_0'B_0^{-1} F_0 \right)^{-1} \right) \left( W_0 + \sum_{j \in \mathcal{E} \cup \mathcal{T}_{+}^* (\lambda_0)} \lambda_{0j} V_{0j} \right) - B_0^{-1} F_0 \left( F_0'B_0^{-1} F_0 \right)^{-1} U_0,
\]

where \( B_0 = H_0 + \sum_{j \in \mathcal{E} \cup \mathcal{T}_{+}^* (\lambda_0)} \lambda_{0j} G_{0j}. \) If \( W_0, V_0, \) and \( U_0 \) are multivariate normal, then \( \mathcal{J} \) will also be multivariate normal.

If \( l(\beta_0) = 0, \) then \( \sum_{j \in \mathcal{E} \cup \mathcal{T}} \lambda_{0j} \left( h' V_{0j} + \frac{1}{2} h' G_{0j} h \right) = 0 \) and \( \mathcal{T}_{+}^* (\lambda_0) = \emptyset, \) so \( \mathcal{J} \) reduces down to

\[
\mathcal{J} = \arg \min_{h \in \Sigma} \{ h'W_0 + \frac{1}{2} h'H_0 h \}, \text{ for } \Sigma = \{ h : U_{0j} + F_{0j}' h = 0 \text{ for } j \in \mathcal{E} \}.
\]

**Remark 7.** In the case where there are only inequality constraints, we can also obtain a closed form solution for the asymptotic distribution \( \mathcal{J}. \) When there are no equality constraints, \( \Sigma = \{ h : U_{0j} + F_{0j}' h \leq 0 \text{ for } j \in \mathcal{T}^* \}. \) It follows from Lemma 5.2 that

\[
\mathcal{J} = \max \left\{ -B_0^{-1} \left( I - F_0^+ \left( F_0'B_0^{-1} F_0^+ \right)^{-1} F_0'B_0^{-1} \right) \left( W_0 + \sum_{j \in \mathcal{T}_{+}^* (\lambda_0)} \lambda_{0j} V_{0j} \right) \right. \\
- B_0^{-1} F_0^+ \left( F_0'B_0^{-1} F_0^+ \right)^{-1} U_{0+}, -B_0^{-1} \left( W_0 + \sum_{j \in \mathcal{T}_{+}^* (\lambda_0)} \lambda_{0j} V_{0j} \right) \right\}
\]

where \( F_{0+} \) is the matrix of \( F_{0j} \) for \( j \in \mathcal{T}_{+}^* (\lambda_0), \) \( U_{0+} \) is the vector of \( U_{0j} \) for \( j \in \mathcal{T}_{+}^* (\lambda_0), \) and \( B_0 = H_0 + \sum_{j \in \mathcal{T}_{+}^* (\lambda_0)} \lambda_{0j} G_{0j}. \) A special case of this is the constrained maximum likelihood example
in Andrews (2000). He imposes a nonnegativity constraint \( \mu \geq 0 \) for a normal mean model (with variance 1) and shows that the asymptotic distribution of the maximum likelihood estimator is \( \mathcal{J} = \max \{ Z, 0 \} \) (where \( Z \sim N(0, 1) \)) if the true mean equals 0. We can obtain this asymptotic distribution by setting \( F_0 = 1, G_0 = 0, V_0 = 0, U_0 = 0, B_0 = H_0 = 1, \) and \( W_0 = Z \).

If there are no binding inequality constraints, then \( \mathcal{I}^*_+ (\lambda_0) = \emptyset, \sum_{j \in \mathcal{I}^*_+ (\lambda_0)} \lambda_{0j} G_{0j} = 0, \) and \( H_0 = B_0, \) so the asymptotic distribution reduces down to \( \mathcal{J} = -H_0^{-1} W_0, \) which will be multivariate normal if \( W_0 \) is multivariate normal.

**Remark 8.** In the case of fixed constraints \( f_n (\beta) = f_0 (\beta) \) that do not depend on the data, if \( l (\beta_0) \) may not be zero, the proximal bootstrap estimator is

\[
\hat{\beta}_n^* = \arg \min_{\beta \in \mathcal{C}^*} \alpha_n \sqrt{n} \left( \hat{\ell}_n (\beta_n) - \hat{l}_n (\beta_n) \right)' \left( \beta - \beta_n \right) + \frac{1}{2} \left( \beta - \beta_n \right)' B_n \left( \beta - \beta_n \right) + \frac{1}{2} \sum_{j \in \mathcal{I}} \lambda_{nj} \left\| \beta - \beta_n \right\|_{G_{nj}}^2
\]

\( C^* = \{ \beta \in \mathcal{B} : f_{nj} (\beta_n) + \tilde{F}_{nj}' (\beta - \beta_n) = 0 \text{ for } j \in \mathcal{E}, f_{nj} (\beta_n) + \tilde{F}_{nj}' (\beta - \beta_n) \leq 0 \text{ for } j \in \mathcal{I} \} \)

If \( l (\beta_0) = 0, \) which is implied by \( Q (\beta) = Q (\beta_0) + \frac{1}{2} (\beta - \beta_0)' H_0 (\beta - \beta_0) + o \left( \| \beta - \beta_0 \|^2 \right) \), then the proximal bootstrap estimator can be defined as

\[
\hat{\beta}_n^* = \arg \min_{\beta \in \mathcal{C}^*} \alpha_n \sqrt{n} \left( \hat{\ell}_n (\beta_n) - \hat{l}_n (\beta_n) \right)' \left( \beta - \beta_n \right) + \frac{1}{2} \left( \beta - \beta_n \right)' B_n \left( \beta - \beta_n \right)
\]

\( C^* = \{ \beta \in \mathcal{B} : f_{nj} (\beta_n) + \tilde{F}_{nj}' (\beta - \beta_n) = 0 \text{ for } j \in \mathcal{E}, f_{nj} (\beta_n) + \tilde{F}_{nj}' (\beta - \beta_n) \leq 0 \text{ for } j \in \mathcal{I} \} \)

The asymptotic distribution when \( l (\beta_0) \) may not be zero can be derived as follows:

\[
n \hat{Q}_n \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - n \hat{Q}_n (\beta_0) + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{nj} n \left( f_{0j} \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - f_{0j} (\beta_0) \right)
\]

\[
= h' \sqrt{n} \left( \hat{\ell}_n (\beta_0) - l (\beta_0) \right) + \frac{1}{2} h' H_0 h + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} \left( \sqrt{n} (F_{0j} - F_{0j})' h + \frac{1}{2} h' G_{0j} h \right) + o_P (1)
\]

\[
= h' \sqrt{n} \left( \hat{\ell}_n (\beta_0) - l (\beta_0) \right) + \frac{1}{2} h' H_0 h + \frac{1}{2} \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} h' G_{0j} h + o_P (1)
\]

\[
\sim h' W_0 + \frac{1}{2} h' H_0 h + \frac{1}{2} \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} h' G_{0j} h
\]

Furthermore since \( \sqrt{n} f_{nj} (\beta_0) = \sqrt{n} f_{0j} (\beta_0) = 0 \) for all \( j \in \mathcal{E} \cup \mathcal{I}^* \),

\[
\Sigma = \{ h : F_{0j}' h = 0 \text{ for } j \in \mathcal{E}, F_{0j}' h \leq 0 \text{ for } j \in \mathcal{I}^* \}
\]

Therefore, \( \mathcal{J} = \arg \min_{h \in \Sigma} \left\{ h' W_0 + \frac{1}{2} h' H_0 h + \frac{1}{2} \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} h' G_{0j} h \right\} \).

When \( l (\beta_0) = 0, \) since \( \lambda_{0j} = 0 \) for all \( j \in \mathcal{E} \cup \mathcal{I}, \) \( \mathcal{J} \) reduces down to the asymptotic distribution in Theorem 1.

\[
n \hat{Q}_n \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - n \hat{Q}_n (\beta_0) + \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{nj} n \left( f_{0j} \left( \beta_0 + \frac{h}{\sqrt{n}} \right) - f_{0j} (\beta_0) \right)
\]

\[
= h' \sqrt{n} \left( \hat{\ell}_n (\beta_0) - l (\beta_0) \right) + \frac{1}{2} h' H_0 h + o_P (1)
\]

\[
\sim h' W_0 + \frac{1}{2} h' H_0 h
\]
Since LICQ is satisfied (which implies the Tangent cone is equal to the linearized feasible set),

\[ \Sigma = \{ h : F_{0j}'h = 0 \text{ for } j \in \mathcal{E}, F_{0j}'h \leq 0 \text{ for } j \in \mathcal{I}^* \} = T_C(\beta_0) \]

Therefore, \( \mathcal{J} = \arg \min_{h \in T_C(\beta_0)} \{ h'W_0 + \frac{1}{2}h'H_0h \} \).

**Remark 9.** Alternatively, we can define the proximal bootstrap estimator as \( \hat{\beta}_n^* = \arg \min_{\beta \in C^*} \hat{A}_n^*(\beta) \), where

\[
\hat{A}_n^*(\beta) = \alpha_n \sqrt{n} \left( \tilde{t}_n(\beta) - \hat{t}_n(\beta) \right)'(\beta - \hat{\beta}_n) + \frac{1}{2} \| \beta - \hat{\beta}_n \|_{H_n}^2 \\
+ \sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{nj} \left( \alpha_n \sqrt{n} \left( \tilde{F}_{nj}(\beta) - \hat{F}_{nj}(\beta) \right)'(\beta - \hat{\beta}_n) + \frac{1}{2} \| \beta - \hat{\beta}_n \|_{G_{nj}}^2 \right)
\]

\( C^* = \{ \beta \in \mathcal{B} : f_{nj}(\beta) + \alpha_n \sqrt{n} \left( f_{nj}^*(\beta) - f_{nj}(\beta) \right) = 0 \text{ for } j \in \mathcal{E}, \]

\( f_{nj}(\beta) + \alpha_n \sqrt{n} \left( f_{nj}^*(\beta) - f_{nj}(\beta) \right) \leq 0 \text{ for } j \in \mathcal{I} \}

The feasible direction set is

\( \mathcal{F}_n^* = \{ h : f_{nj}(\beta_0 + \alpha_n h) + \alpha_n \sqrt{n} \left( f_{nj}^*(\beta) - f_{nj}(\beta) \right) = 0 \text{ for } j \in \mathcal{E}, \]

\( f_{nj}(\beta_0 + \alpha_n h) + \alpha_n \sqrt{n} \left( f_{nj}^*(\beta) - f_{nj}(\beta) \right) \leq 0 \text{ for } j \in \mathcal{I} \}

and the linearized feasible direction set is

\[
\Sigma_n^* = \left\{ h : \frac{f_{nj}(\beta_0)}{\alpha_n} + F_{nj}(\beta_0)'h + \sqrt{n} \left( f_{nj}^*(\beta) - f_{nj}(\beta) \right) = 0 \text{ for } j \in \mathcal{E}, \right. \\
\left. \frac{f_{nj}(\beta_0)}{\alpha_n} + F_{nj}(\beta_0)'h + \sqrt{n} \left( f_{nj}^*(\beta) - f_{nj}(\beta) \right) \leq 0 \text{ for } j \in \mathcal{I} \right\}
\]

Note that since \( \frac{f_{nj}(\beta_0)}{\alpha_n} + F_{nj}(\beta_0)'h + \sqrt{n} \left( f_{nj}^*(\beta) - f_{nj}(\beta) \right) \overset{p}{\to} -\infty \text{ for } j \in \mathcal{I} \cup \mathcal{I}^* \), the nonbinding inequality constraints do not affect the asymptotic distribution under our pointwise asymptotics. Since \( \frac{f_{nj}(\beta_0)}{\alpha_n} = \sqrt{n} \left( f_{nj}^*(\beta_0) - f_{nj}(\beta_0) \right) \overset{p}{\to} U_{0j} \), jointly, for all \( j \in \mathcal{E} \cup \mathcal{I}^* \), and \( \sqrt{n} \left( f_{nj}^*(\beta) - f_{nj}(\beta) \right) \overset{p}{\to} U_{0j} \), jointly, for all \( j \in \mathcal{E} \cup \mathcal{I}^* \), it follows that \( \propto 1 \left( h \notin \Sigma_n^* \right) \overset{P}{\to} \propto 1 \left( h \notin \Sigma \right) \).

Therefore, this nonlinearized bootstrap estimator has the same asymptotic distribution as the linearized version in Theorem 2.

### 4 Monte Carlo Simulations

#### 4.1 Boundary Constrained Nonsmooth GMM

We consider a simple location model with i.i.d data:

\[ y_i = \beta_0 + \epsilon_i, \quad \epsilon_i \sim N(0,1), \quad \beta_0 = 0 \]

For \( \pi(\cdot, \beta) = [1(y_i \leq \beta) - \tau; y_i - \beta]' \), let the population and sample moments be

\[ \pi(\beta) = [P(y_i \leq \beta) - 0.5; Ey_i - \beta]', \quad \hat{\pi}_n(\beta) = \left[ \frac{1}{n} \sum_{i=1}^{n} 1(y_i \leq \beta) - 0.5; \frac{1}{n} \sum_{i=1}^{n} y_i - \beta \right]' \]
We solve the following constrained GMM problem:

$$\hat{\beta}_n = \arg\min_{\beta \geq 0} Q_n (\beta) = \hat{\pi}_n (\beta)^T \hat{\pi}_n (\beta)$$

We use Matlab’s built-in fmincon solver to compute $\hat{\beta}_n = \hat{\beta}_n$ and also

$$\hat{\beta}_n^* = \arg\min_{\beta \in C} \left\{ \alpha_n \sqrt{n} \left( \hat{I}_n (\beta) - \hat{I}_n (\beta) \right) \right\} ,$$

where $H_n = \hat{G}'_n \hat{G}_n + \hat{L}_n \hat{\pi}_n (\beta_n)$, $\hat{I}_n (\beta) = \hat{G}'_n \hat{\pi}_n (\beta_n)$, and

$$\hat{G}_n = \frac{1}{\bar{h}} \sum_{i=1}^{n} K_h \left( \hat{\theta}_n \right) , \hat{G}_n^* = \frac{1}{\bar{h}} \sum_{i=1}^{n} K_h \left( \hat{y}_i - \hat{\theta}_n \right) , \hat{L}_n = \frac{1}{\bar{h} \bar{h}^2} \sum_{i=1}^{n} K_h^{' \prime} \left( \hat{y}_i - \hat{\theta}_n \right) \right)$$

and $K_h (x) = K(x/h), K(x) = (2\pi)^{-1/2} \exp^{-x^2/2}, K_h^{' \prime} (x) = K^{' \prime} (x/h)$ and $K^{' \prime} (x) = -(2\pi)^{-1/2} x \exp^{-x^2/2}$.

We use the Silverman’s rule of thumb bandwidth $\bar{h} = 1.06n^{-1/5}$.

We consider five different sample sizes $n \in \{100, 500, 1000, 5000, 10000\}$ and three different $\alpha_n$’s for each $n$: $\alpha_n \in \{n^{-1/3}, n^{-1/4}, n^{-1/6}, n^{-1/8}, n^{-1/10}\}$. We use 5000 bootstrap iterations and 2000 Monte Carlo simulations. Empirical coverage frequencies for equal-tailed nominal 95% confidence intervals and average interval lengths are reported in table 1.

<table>
<thead>
<tr>
<th>$n$</th>
<th>100</th>
<th>500</th>
<th>1000</th>
<th>5000</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_n = n^{-1/3}$</td>
<td>0.969</td>
<td>0.975</td>
<td>0.971</td>
<td>0.972</td>
<td>0.968</td>
</tr>
<tr>
<td>(0.216)</td>
<td>(0.095)</td>
<td>(0.067)</td>
<td>(0.029)</td>
<td>(0.021)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_n = n^{-1/4}$</td>
<td>0.969</td>
<td>0.975</td>
<td>0.971</td>
<td>0.972</td>
<td>0.968</td>
</tr>
<tr>
<td>(0.210)</td>
<td>(0.092)</td>
<td>(0.065)</td>
<td>(0.029)</td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_n = n^{-1/6}$</td>
<td>0.969</td>
<td>0.975</td>
<td>0.971</td>
<td>0.972</td>
<td>0.968</td>
</tr>
<tr>
<td>(0.206)</td>
<td>(0.091)</td>
<td>(0.064)</td>
<td>(0.028)</td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_n = n^{-1/8}$</td>
<td>0.969</td>
<td>0.975</td>
<td>0.971</td>
<td>0.972</td>
<td>0.968</td>
</tr>
<tr>
<td>(0.204)</td>
<td>(0.090)</td>
<td>(0.063)</td>
<td>(0.028)</td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_n = n^{-1/10}$</td>
<td>0.969</td>
<td>0.975</td>
<td>0.971</td>
<td>0.972</td>
<td>0.968</td>
</tr>
<tr>
<td>(0.203)</td>
<td>(0.090)</td>
<td>(0.063)</td>
<td>(0.028)</td>
<td>(0.020)</td>
<td></td>
</tr>
</tbody>
</table>

We now compare the proximal bootstrap with the centered standard bootstrap estimator $\hat{\beta}_n^{**} = \arg\min_{\beta \in C} \left( \hat{\pi}_n^* (\beta) - \hat{\pi}_n (\hat{\beta}_n) \right)^T \left( \hat{\pi}_n^* (\beta) - \hat{\pi}_n (\hat{\beta}_n) \right)$. Empirical coverage frequencies for equal-tailed nominal 95% confidence intervals and average interval lengths are reported in table 2. Interestingly, the coverage frequencies are similar, although the intervals are wider.

<table>
<thead>
<tr>
<th>$n$</th>
<th>100</th>
<th>500</th>
<th>1000</th>
<th>5000</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.968</td>
<td>0.976</td>
<td>0.974</td>
<td>0.972</td>
<td>0.968</td>
</tr>
<tr>
<td>(0.236)</td>
<td>(0.107)</td>
<td>(0.076)</td>
<td>(0.034)</td>
<td>(0.024)</td>
<td></td>
</tr>
</tbody>
</table>
4.2 Conditional Logit Model with Estimated Inequality Constraints

We generate data according to \( y_{ij} = 1 \left( y_{ij}^* > y_{ik}^* \forall k \neq j \right) \), where the utility of individual \( i = 1 \ldots n \) from picking choice \( j = 1 \ldots J \) is given by \( y_{ij}^* = \beta_0 x_{ij} + \epsilon_{ij} \), for \( x_i \sim N \left( \begin{pmatrix} 1 \\ 2 \\ \vdots \\ J \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & \ldots & 0.5 \\ 0.5 & 1 & \ldots & 0.5 \\ \vdots & \vdots & \ddots & \vdots \\ 0.5 & 0.5 & \ldots & 1 \end{pmatrix} \right) \)

and \( \epsilon_{ij} \overset{i.i.d.}{\sim} \) Type 1 Extreme Value. We set \( \beta_0 = 0.1 \). The constrained MLE estimator maximizes the log-likelihood subject to the constraints that the share of individuals who pick each choice cannot exceed the supply of that choice. These inequality constraints can be viewed as capacity constraints similar to the ones in de Palma et al. (2007) which state that the equilibrium demand for each housing unit should not exceed the supply of that housing unit. For \( P_{ij} = \frac{\exp(\beta_{ij})}{\sum_l \exp(\beta_{il})} \),

\[
\hat{\beta}_n = \arg \max_\beta \ln L(\beta) = \frac{1}{nj} \sum_{i=1}^n \sum_{j=1}^J y_{ij} \ln P_{ij}
\]

s.t. \( \frac{1}{n} \sum_{i=1}^n P_{ij} \leq \bar{b}_j \) for all \( j = 1 \ldots J \)

where \( \bar{b}_j = \frac{1}{10^6} \sum_{i=1}^{10^6} \frac{\exp(\beta_{0,\tilde{x}_{ij}})}{\sum_{l} \exp(\beta_{0,l})} \) for \( \tilde{x}_{ij} \) drawn independently from the same distribution as \( x_{ij} \).

We use Matlab’s built-in fmincon solver to compute \( \hat{\beta}_n = \hat{\beta}_n \) and also \( \hat{\beta}_n^* = \arg \min_\beta \hat{\lambda}_n(\beta), \) where \( \hat{\lambda}_n(\beta) = \sum_{j=1}^J \hat{\lambda}_{nj} \left( \alpha_n \sqrt{n} \left( \hat{F}_{nj} - \hat{F}_{nj} \right)' \right) \left( \beta - \hat{\beta}_n \right) + \frac{1}{2} \| \beta - \hat{\beta}_n \|^2_{H_n} \)

\( C^* = \{ f_{nj}(\beta_n) + \tilde{F}_{nj}'(\beta - \beta_n) + \alpha_n \sqrt{n} \left( f_{nj}^*(\beta_n) - f_{nj}(\beta_n) \right) \leq 0 \text{ for } j \in T \} \)

\[
\hat{\lambda}_n(\beta) = \frac{\partial \ln L(\beta)}{\partial \beta} = -\frac{1}{nj} \sum_{i=1}^n \sum_{j=1}^J (y_{ij} - P_{ij}) x_{ij}
\]

\[
H_n(\beta) = -\frac{\partial^2 \ln L(\beta)}{\partial \beta \partial \beta'} = \frac{1}{nj} \sum_{i=1}^n \sum_{j=1}^J P_{ij} \left( x_{ij} - \sum_l P_{il} x_{il} \right) \left( x_{ij} - \sum_l P_{il} x_{il} \right)'
\]

\[
F_{nj}(\beta) = \frac{1}{n} \sum_{i=1}^n \frac{\partial P_{ij}}{\partial \beta} = \frac{1}{n} \sum_{i=1}^n P_{ij} \frac{\partial \ln P_{ij}}{\partial \beta} = \frac{1}{n} \sum_{i=1}^n P_{ij} \left( x_{ij} - \sum_l P_{il} x_{il} \right)
\]

\[
G_{nj}(\beta) = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 P_{ij}}{\partial \beta \partial \beta'} = \frac{1}{n} \sum_{i=1}^n P_{ij} \frac{\partial^2 \ln P_{ij}}{\partial \beta \partial \beta'} = \frac{1}{n} \sum_{i=1}^n P_{ij} \left( x_{ij} - \sum_l P_{il} x_{il} \right)' - \frac{1}{n} \sum_{i=1}^n P_{ij} \sum_l \frac{\partial P_{il}}{\partial \beta} x_{il}'
\]

\[
= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^J P_{ij} \left( x_{ij} - \sum_l P_{il} x_{il} \right)' \left( x_{ij} - \sum_l P_{il} x_{il} \right)'
\]
\[- \frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{n} P_{il}P_{il} \left( x_{il} - \sum_{m} P_{im}x_{im} \right) x_{il}' \]

We consider \( n \in \{100, 500, 1000\} \), \( J = 20 \), and \( \alpha_n \in \{n^{-1/3}, n^{-1/4}, n^{-1/6}, n^{-1/8}, n^{-1/10}\} \). Empirical coverage frequencies for equal-tailed nominal 95% confidence intervals and average interval lengths are reported in table 3. We use \( B = 2000 \) bootstrap iterations and \( R = 1000 \) Monte Carlo simulations.

Table 3: Proximal Bootstrap Empirical Coverage Frequencies and Average Interval Lengths

<table>
<thead>
<tr>
<th></th>
<th>( n = 100 )</th>
<th>( n = 500 )</th>
<th>( n = 1000 )</th>
<th>( n = 2000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( J = 20 )</td>
<td>( J = 20 )</td>
<td>( J = 20 )</td>
<td>( J = 20 )</td>
</tr>
<tr>
<td>( \alpha_n = n^{-1/3} )</td>
<td>0.917</td>
<td>0.923</td>
<td>0.946</td>
<td>0.929</td>
</tr>
<tr>
<td></td>
<td>(0.0016)</td>
<td>(0.0007)</td>
<td>(0.0005)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>( \alpha_n = n^{-1/4} )</td>
<td>0.924</td>
<td>0.935</td>
<td>0.952</td>
<td>0.946</td>
</tr>
<tr>
<td></td>
<td>(0.0016)</td>
<td>(0.0007)</td>
<td>(0.0005)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>( \alpha_n = n^{-1/6} )</td>
<td>0.922</td>
<td>0.939</td>
<td>0.952</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>(0.0016)</td>
<td>(0.0007)</td>
<td>(0.0005)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>( \alpha_n = n^{-1/8} )</td>
<td>0.922</td>
<td>0.927</td>
<td>0.945</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.0007)</td>
<td>(0.0005)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>( \alpha_n = n^{-1/10} )</td>
<td>0.918</td>
<td>0.920</td>
<td>0.944</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.0007)</td>
<td>(0.0005)</td>
<td>(0.0004)</td>
</tr>
</tbody>
</table>

We also consider the case where \( J \) is growing with \( n \). Proximal bootstrap empirical coverage frequencies for equal-tailed nominal 95% confidence intervals and average interval lengths are reported in table 4. The results are computed using \( B = 2000, R = 1000 \).

Table 4: Proximal Bootstrap Empirical Coverage Frequencies and Average Interval Lengths

<table>
<thead>
<tr>
<th></th>
<th>( n = 100 )</th>
<th>( n = 500 )</th>
<th>( n = 100 )</th>
<th>( n = 500 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( J = 50 )</td>
<td>( J = 50 )</td>
<td>( J = 100 )</td>
<td>( J = 100 )</td>
</tr>
<tr>
<td>( \alpha_n = n^{-1/3} )</td>
<td>0.928</td>
<td>0.936</td>
<td>0.932</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.0003)</td>
<td>(0.0006)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>( \alpha_n = n^{-1/4} )</td>
<td>0.936</td>
<td>0.940</td>
<td>0.939</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.0003)</td>
<td>(0.0006)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>( \alpha_n = n^{-1/6} )</td>
<td>0.941</td>
<td>0.946</td>
<td>0.944</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.0003)</td>
<td>(0.0006)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>( \alpha_n = n^{-1/8} )</td>
<td>0.939</td>
<td>0.946</td>
<td>0.945</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.0003)</td>
<td>(0.0006)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>( \alpha_n = n^{-1/10} )</td>
<td>0.938</td>
<td>0.947</td>
<td>0.947</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.0003)</td>
<td>(0.0006)</td>
<td>(0.0003)</td>
</tr>
</tbody>
</table>

We now compare the proximal bootstrap with the standard multinominal bootstrap estimator. Empirical coverage frequencies for equal-tailed nominal 95% confidence intervals and average interval lengths are reported in table 5. The standard bootstrap undercovers for all values of \( n \) and the intervals are sometimes wider.

21
Table 5: Standard Bootstrap Empirical Coverage Frequencies and Average Interval Lengths

<table>
<thead>
<tr>
<th></th>
<th>n = 100</th>
<th>n = 500</th>
<th>n = 1000</th>
<th>n = 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>B = 2000</td>
<td>J = 20</td>
<td>0.922</td>
<td>0.911</td>
<td>0.919</td>
</tr>
<tr>
<td>R = 1000</td>
<td>(0.0018)</td>
<td>(0.0008)</td>
<td>(0.0006)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>B = 2000</td>
<td>J = 20</td>
<td>0.926</td>
<td>0.907</td>
<td>0.910</td>
</tr>
<tr>
<td>R = 2000</td>
<td>(0.0018)</td>
<td>(0.0008)</td>
<td>(0.0006)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>n = 100</td>
<td>J = 50</td>
<td>n = 500</td>
<td>n = 100</td>
<td>n = 500</td>
</tr>
<tr>
<td>B = 2000</td>
<td>J = 50</td>
<td>0.922</td>
<td>0.928</td>
<td>0.911</td>
</tr>
<tr>
<td>R = 1000</td>
<td>(0.0008)</td>
<td>(0.0004)</td>
<td>(0.0007)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>B = 2000</td>
<td>(0.0008)</td>
<td>(0.0004)</td>
<td>(0.0007)</td>
<td>(0.0003)</td>
</tr>
</tbody>
</table>

4.3 Rust (1987) Bus Engine Replacement Model

We apply our method to conduct inference for the Mathematical Programming with Equilibrium Constraints (MPEC) formulation of the Rust (1987) Bus Engine Replacement model. Su and Judd (2012) indicate that the MPEC estimator can be bootstrapped, although they do not provide an analysis of the empirical coverage frequencies of bootstrap confidence intervals. We find that our proximal bootstrap method performs equally good in terms of coverage and is more than twice as fast as the standard bootstrap.

Using the code accompanying Su and Judd (2012), we generate data using the following parameters used in their paper: discount factor $\beta = 0.975$, replacement cost $RC = 11.7257$, operating cost parameter $\theta_1 = 2.4569$, and transition probabilities $\theta_3 = \left[ 0.0937, 0.4475, 0.4459, 0.0127, 0.0002 \right]^T$.

The MPEC objective function is the log likelihood which is a function of both the structural parameters and the choice-specific value functions $EV(x, d)$ given the data $\left( (x_i^t, d_i^t)_{t=1}^T \right)_{i=1}^M$, where $x_i^t$ is the mileage of bus $i$ in period $t$ and $d_i^t$ is the replacement decision for bus $i$ in period $t$.

$$L(\theta_1, \theta_3, RC, EV) = \frac{1}{M} \sum_{i=1}^M \sum_{t=2}^T \log \left( \frac{\exp \left[ \nu (x_i^t, d_i^t; \theta_1, RC) + \beta EV (x_i^t, d_i^t) \right]}{\sum_{d' \in \{0,1\}} \exp \left[ \nu (x_i^t, d'; \theta_1, RC) + \beta EV (x_i^t, d') \right]} \right)$$

$$+ \frac{1}{M} \sum_{i=1}^M \sum_{t=2}^T \log (p_3 (x_{i,t} | x_{i,t-1}, d_{i,t-1}, \theta_3))$$

The constraints are the fixed point equations defining the discretized choice-specific value functions $EV(x, d)$ for mileage constrained to lie on a grid $\hat{x} = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_K\}$:

$$EV(\hat{x}_k, d) = \sum_{x'} \log \left( \sum_{d' \in \{0,1\}} \exp \left[ \nu (x', d'; \theta_1, RC) + \beta EV (x', d') \right] \right) p_3 (x' | \hat{x}_k, d, \theta_3)$$

Given the current state $\hat{x}_k$, the next period mileage $x' \in \{\hat{x}_k, \hat{x}_{k+1}, \hat{x}_{k+2}, \hat{x}_{k+3}, \hat{x}_{k+4}\}$ can move up at most 4 grid points if the engine is not replaced. If the engine is replaced, the mileage resets to $\hat{x}_1$. Su and Judd (2012)’s code chooses the mileage grid to be $\hat{x} = \{1, 2, 3, \ldots, 175\}$. The utility
function in their code is defined as

\[
\nu(x, d; \theta_1, RC) = \begin{cases} 
-0.001x\theta_1, & d = 0 \\
-RC - 0.001\theta_1, & d = 1
\end{cases}
\]

The transition probabilities \( p_3(x' | \hat{x}_k, d, \theta_3) \) are equal to \( \theta_3 \) if replacement does not occur. The only values of the choice-specific value functions we need to estimate are the ones corresponding to no replacement \( EV = [EV(\hat{x}_1, 0), EV(\hat{x}_2, 0), \ldots, EV(\hat{x}_K, 0)] \) because \( EV(\hat{x}_k, 1) = EV(\hat{x}_1, 0) \) for all \( k \), as pointed out in footnote 9 of Su and Judd (2012). Notice that because the mileage grid is fixed, the constraints do not depend on the data. Define the parameters \( \theta = (\theta_1, \theta_3', RC, EV)' \), and the constraint set \( C = \{f_j(\theta) = 0 \text{ for } j \in \mathcal{E}, f_j(\theta) \leq 0 \text{ for } j \in \mathcal{T}\} \), where \( f_j(\theta) \) includes the \( EV \) fixed point equations as well as the constraints on the transition probabilities satisfying \( 0 \leq \theta_3 \leq 1 \) and \( \sum_j \theta_{3j} = 1 \). Because our asymptotics are large \( M \), fixed \( T \), the rate of convergence of our estimator is \( \sqrt{M} \). For some \( \alpha_M \rightarrow 0 \) and \( \sqrt{M} \alpha_M \rightarrow \infty \), and a \( \sqrt{M} \) consistent estimator \( \bar{\theta}_M \), the proximal bootstrap estimator is given by

\[
\hat{\theta}_M^* = \arg \min_{\theta \in C^*} \alpha_M \sqrt{M} \left( \bar{\theta}_M - \bar{\theta}_M \right)' \left( \theta - \bar{\theta}_M \right) + \frac{1}{2} \left\| \theta - \bar{\theta}_M \right\|^2_{\hat{B}_M} \\
C^* = \{f_j(\bar{\theta}_M) + F_j(\theta - \bar{\theta}_M) = 0 \text{ for } j \in \mathcal{E}, f_j(\bar{\theta}_M) + F_j(\theta - \bar{\theta}_M) \leq 0 \text{ for } j \in \mathcal{T}\}
\]

We follow Su and Judd (2012) and use Knitro to compute \( \bar{\theta}_M = \hat{\theta}_M \) as well as \( \hat{\theta}_M^* \), although in principle the built-in Matlab nonlinear optimization solvers should also find the solution given enough time to search the parameter space. Because \( l(\theta_0) = 0 \) in this model, we do not need to include the Lagrange multiplier term in the objective function.

Tables 6-8 show the empirical coverage frequencies and average interval lengths for two-sided equal tailed nominal 95\% proximal bootstrap confidence intervals computed using \( B = 1000 \) bootstrap iterations and \( R = 2000 \) Monte Carlo simulations. We consider 6 different values of \( M \in \{500, 1000, 2000, 4000, 5000, 6000\} \) and three different values of \( \alpha_M \in \{M^{-1/3}, M^{-1/4}, M^{-1/6}\} \). Due to time constraints on the server, we were unable to obtain results for the standard bootstrap using the same values of \( M, B, \) and \( R \), but the results should be similar given that the standard bootstrap is consistent in this example.
Table 6: Proximal Bootstrap Coverage Frequencies and Average Interval Lengths for $\alpha_M = M^{-1/3}$

<table>
<thead>
<tr>
<th>$M$</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
<th>5000</th>
<th>6000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>0.925</td>
<td>0.946</td>
<td>0.949</td>
<td>0.942</td>
<td>0.948</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td>(0.520)</td>
<td>(0.373)</td>
<td>(0.264)</td>
<td>(0.187)</td>
<td>(0.167)</td>
<td>(0.152)</td>
</tr>
<tr>
<td>$\theta_{30}$</td>
<td>0.951</td>
<td>0.947</td>
<td>0.945</td>
<td>0.933</td>
<td>0.932</td>
<td>0.935</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>0.955</td>
<td>0.944</td>
<td>0.951</td>
<td>0.948</td>
<td>0.94</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\theta_{32}$</td>
<td>0.949</td>
<td>0.952</td>
<td>0.944</td>
<td>0.942</td>
<td>0.942</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\theta_{33}$</td>
<td>0.957</td>
<td>0.95</td>
<td>0.949</td>
<td>0.951</td>
<td>0.96</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>RC</td>
<td>0.927</td>
<td>0.95</td>
<td>0.949</td>
<td>0.946</td>
<td>0.946</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>(1.683)</td>
<td>(1.204)</td>
<td>(0.853)</td>
<td>(0.604)</td>
<td>(0.540)</td>
<td>(0.492)</td>
</tr>
</tbody>
</table>

Table 7: Proximal Bootstrap Coverage Frequencies and Average Interval Lengths for $\alpha_M = M^{-1/4}$

<table>
<thead>
<tr>
<th>$M$</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
<th>5000</th>
<th>6000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>0.923</td>
<td>0.949</td>
<td>0.95</td>
<td>0.941</td>
<td>0.949</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>(0.520)</td>
<td>(0.372)</td>
<td>(0.264)</td>
<td>(0.187)</td>
<td>(0.167)</td>
<td>(0.153)</td>
</tr>
<tr>
<td>$\theta_{30}$</td>
<td>0.952</td>
<td>0.948</td>
<td>0.94</td>
<td>0.935</td>
<td>0.934</td>
<td>0.937</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>0.954</td>
<td>0.942</td>
<td>0.95</td>
<td>0.946</td>
<td>0.944</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\theta_{32}$</td>
<td>0.952</td>
<td>0.95</td>
<td>0.941</td>
<td>0.943</td>
<td>0.939</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\theta_{33}$</td>
<td>0.959</td>
<td>0.95</td>
<td>0.95</td>
<td>0.949</td>
<td>0.958</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>RC</td>
<td>0.927</td>
<td>0.952</td>
<td>0.95</td>
<td>0.945</td>
<td>0.949</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>(1.683)</td>
<td>(1.204)</td>
<td>(0.853)</td>
<td>(0.604)</td>
<td>(0.540)</td>
<td>(0.493)</td>
</tr>
</tbody>
</table>

Table 8: Proximal Bootstrap Coverage Frequencies and Average Interval Lengths for $\alpha_M = M^{-1/6}$

<table>
<thead>
<tr>
<th>$M$</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
<th>5000</th>
<th>6000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>0.924</td>
<td>0.947</td>
<td>0.949</td>
<td>0.943</td>
<td>0.948</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>(0.520)</td>
<td>(0.372)</td>
<td>(0.264)</td>
<td>(0.187)</td>
<td>(0.167)</td>
<td>(0.152)</td>
</tr>
<tr>
<td>$\theta_{30}$</td>
<td>0.952</td>
<td>0.95</td>
<td>0.941</td>
<td>0.933</td>
<td>0.933</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>0.955</td>
<td>0.942</td>
<td>0.949</td>
<td>0.949</td>
<td>0.944</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\theta_{32}$</td>
<td>0.951</td>
<td>0.951</td>
<td>0.942</td>
<td>0.944</td>
<td>0.94</td>
<td>0.946</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\theta_{33}$</td>
<td>0.96</td>
<td>0.951</td>
<td>0.949</td>
<td>0.951</td>
<td>0.959</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>RC</td>
<td>0.925</td>
<td>0.951</td>
<td>0.95</td>
<td>0.947</td>
<td>0.948</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>(1.666)</td>
<td>(1.201)</td>
<td>(0.852)</td>
<td>(0.603)</td>
<td>(0.540)</td>
<td>(0.493)</td>
</tr>
</tbody>
</table>
5 Appendix

5.1 Convexification Theorem

When the fixed set $C = C' \cap B$ can be expressed as the intersection of a set $C'$ which is Clarke Regular at $\beta_0$ and another set $B$ which is a convex subset of $\mathbb{R}^d$ such that for a sequence $\alpha_n$ such that $\alpha_n \to 0$ and $\sqrt{n}\alpha_n \to \infty$, $\frac{B - \beta_0}{\alpha_n}$ are all contained in some bounded region of $\mathbb{R}^d$, we can define an alternative bootstrap estimator that is the solution to a convex optimization problem:

$$\hat{h}^* = \arg \min_{h \in \text{conv} \left( \frac{C' - \beta_0}{\alpha_n} \right) \cap \frac{B - \beta_0}{\alpha_n}} \left\{ \sqrt{n} \left( \hat{l}_n (\beta_n) - \hat{l}_n (\beta_n) \right)' h + \frac{1}{2} h' H_n h \right\}$$

where $\text{conv}(A)$ is the convex hull of a set $A$ (the set of all convex combinations of points in $A$). If $C'$ is a finite collection of points, we can use any off-the-shelf algorithm to compute the $d$-dimensional convex hull of $\left( \frac{C' - \beta_0}{\alpha_n} \right)$.

A popular one is the Quickhull algorithm of Barber et al. (1996) that is implemented as the convhulln function in Matlab.

**Theorem 3.** Suppose Assumptions 3-4 are satisfied, and $\beta_0 = \text{plim} \hat{\beta}_n = \arg \min_{\beta \in C = C' \cap B} Q(\beta)$ is unique for a set $C' \subset \mathbb{R}^d$ that is Clarke Regular at $\beta_0$ and a convex subset $B$ of $\mathbb{R}^d$ such that for a sequence $\alpha_n$ such that $\alpha_n \to 0$ and $\sqrt{n}\alpha_n \to \infty$, $\frac{B - \beta_0}{\alpha_n}$ are all contained in some bounded region of $\mathbb{R}^d$. Suppose $\hat{Q}_n (\hat{\beta}_n) \leq \inf_{\beta \in C} \hat{Q}_n (\beta) + o_P(n^{-1})$ and $\hat{h}^* = \arg \min_{h \in \text{conv} \left( \frac{C' - \beta_0}{\alpha_n} \right) \cap \frac{B - \beta_0}{\alpha_n}} \left\{ \sqrt{n} \left( \hat{l}_n (\beta_n) - \hat{l}_n (\beta_n) \right)' h + \frac{1}{2} h' H_n h \right\}$.

For any $\hat{\beta}_n$ such that $\sqrt{n} (\hat{\beta}_n - \beta_0) = O_P(1)$ and $H_n \overset{p}{\to} H_0$, $\sqrt{n} (\hat{\beta}_n - \beta_0) \overset{p}{\to} \mathcal{J}$ and $\hat{h}^* \overset{p}{\to} \mathcal{J}$, where $\mathcal{J} = \arg \min_{h \in T_C(\beta_0)} \{ h' W_0 + \frac{1}{2} h' H_0 h \}$, $W_0 \sim N \left( 0, P (g (\cdot, \beta_0) - P g (\cdot, \beta_0)) (g (\cdot, \beta_0) - P g (\cdot, \beta_0))' \right)$, and $T_C (\beta_0) = \lim_{\tau \downarrow 0} \frac{C' - \beta_0}{\alpha_n}$.

Proof: The proof of $\sqrt{n} (\hat{\beta}_n - \beta_0) \overset{p}{\to} \mathcal{J}$ is the same as theorem 1. To show $\hat{h}^* \overset{p}{\to} \mathcal{J}$, note that $\hat{h}^* = \tilde{h}^* - \frac{\beta_0 - \beta_0}{\alpha_n}$, where

$$\tilde{h}^* = \arg \min_{h \in \text{conv} \left( \frac{C' - \beta_0}{\alpha_n} \right) \cap \frac{B - \beta_0}{\alpha_n}} \left\{ \sqrt{n} \left( \hat{l}_n (\beta_n) - \hat{l}_n (\beta_n) \right)' h + \frac{1}{2} h' H_n h \right\}$$

Since the sets $\frac{B - \beta_0}{\alpha_n}$ are all convex and contained in some bounded region of $\mathbb{R}^d$, we can use Theorem 4.30 in Rockafellar et al. (1998) to show that

$$\lim_{\alpha_n \to 0} \text{conv} \left( \frac{C' - \beta_0}{\alpha_n} \right) \cap \frac{B - \beta_0}{\alpha_n} = \lim_{\alpha_n \to 0} \text{conv} \left( \frac{C' - \beta_0}{\alpha_n} \right) \cap \frac{B - \beta_0}{\alpha_n} = \text{conv} \left( \lim_{\alpha_n \to 0} \frac{C' - \beta_0}{\alpha_n} \right) \cap \frac{B - \beta_0}{\alpha_n} = \text{conv} (T_C (\beta_0))$$

Since $C'$ is Clarke Regular at $\beta_0$ and $B$ is convex, $C = C' \cap B$ is also Clarke Regular at $\beta_0$ which implies that $T_C (\beta_0)$ is a convex cone and

$$\text{conv} (T_C (\beta_0)) = T_C (\beta_0)$$
It follows that
\[ \infty \left( h \notin \text{conv} \left( \frac{C' - \beta_0}{\alpha_n} \right) \cap \frac{B - \beta_0}{\alpha_n} \right) \Rightarrow \infty \left( h \notin T_C (\beta_0) \right) \]

Similar arguments as in Theorem 1 suggest that
\[ h' \sqrt{n} \left( I_n (\bar{\beta}_n) - \hat{I}_n (\bar{\beta}_n) \right) + \frac{1}{2} h' \tilde{H}_n h + \infty \left( h \notin \text{conv} \left( \frac{C' - \beta_0}{\alpha_n} \right) \cap \frac{B - \beta_0}{\alpha_n} \right) \]
\[ \Rightarrow \frac{p}{e-d} h' W_0 + \frac{1}{2} h' H_0 h + \infty \left( h \notin T_C (\beta_0) \right) \]

which implies \( \hat{h}^* \xrightarrow{\mathcal{P}} J \), and since \( \hat{h}^* = \hat{h}^* - \frac{\beta - \beta_0}{\alpha_n} = \hat{h}^* + o_p(1) \), \( \hat{h}^* \xrightarrow{\mathcal{P}} J \) as well.

### 5.2 Constrained Least Squares

**Lemma 5.1.** Suppose \( H_0 \in \mathbb{R}^d \times \mathbb{R}^d \) is nonsingular, \( R \in \mathbb{R}^d \times \mathbb{R}^m \) has rank \( m \), and \( \Delta_n = O_P(1) \). Then

\[ h^* = \arg \min_{R'h = \delta} h' \Delta_n + \frac{1}{2} h' Bh \]
\[ = -B^{-1} \left( I - R (R' B^{-1} R)^{-1} R' B^{-1} \right) \Delta_n + B^{-1} R (R' B^{-1} R)^{-1} \Delta_n \]

**Proof:** The Lagrangian and KKT conditions are
\[ \mathcal{L} = h' \Delta_n + \frac{1}{2} h' Bh + \lambda \circ (R'h - \delta) \]
\[ \Delta_n + Bh + R\lambda = 0 \]
\[ R'h - \delta = 0 \]

The first KKT condition says \( h^* = -B^{-1} (\Delta_n + R\lambda) \). Substituting into the second KKT condition,
\[ -R'B^{-1} (\Delta_n + R\lambda) = \delta \iff \lambda = -(R' B^{-1} R)^{-1} (\delta + R' B^{-1} \Delta_n) \]

Therefore,
\[ h^* = -B^{-1} \Delta_n + B^{-1} R (R' B^{-1} R)^{-1} (\delta + R' B^{-1} \Delta_n) \]
\[ = -B^{-1} \left( I - R (R' B^{-1} R)^{-1} R' B^{-1} \right) \Delta_n + B^{-1} R (R' B^{-1} R)^{-1} \Delta_n \]

**Lemma 5.2.** Suppose \( H_0 \in \mathbb{R}^d \times \mathbb{R}^d \) is nonsingular, \( R_A \in \mathbb{R}^d \times \mathbb{R}^{m_A} \) has rank \( m_A \), and \( \Delta_n = O_P(1) \), where \( R_A \) denotes the submatrix of \( R \in \mathbb{R}^d \times \mathbb{R}^m \) corresponding to the binding constraints. Then
\[ h^* = \arg \min_{R'h \leq \delta} h' \Delta_n + \frac{1}{2} h' H_0 h \]
\[ = \max \left( -H_0^{-1} \left( I - R_A (R_A H_0^{-1} R_A)^{-1} R_A H_0^{-1} \right) \Delta_n + H_0^{-1} R_A (R_A H_0^{-1} R_A)^{-1} \delta_A, -H_0^{-1} \Delta_n \right) \]

where \( \delta_A \) denotes the subvector of \( \delta \) corresponding to the binding constraints.
Proof: The Lagrangian and KKT Conditions are
\[ \mathcal{L} = h' \Delta_n + \frac{1}{2} h'H_0h + \sum_{i=1}^{m} \mu_i (R_i'h - \delta_i) \]

\[ \Delta_n + H_0h + R\mu = 0 \]
\[ \mu_i \geq 0, \mu_i (R_i'h - \delta_i) = 0 \forall i = 1...m \]

The first KKT condition says \( h^+ = -H_0^{-1} (\Delta_n + R\mu) \). The second says that if \( \mu_i > 0 \), then \( R_i'h^+ - \delta_i = 0 \), meaning the inequality constraint is binding. The assumption that \( R\Lambda \) has rank \( m\Lambda \) implies linear independence constraint qualification is satisfied which means the set of Lagrange multipliers that satisfy the KKT conditions is a singleton (Wachsmuth (2013)). Let the Lagrange multipliers corresponding to binding constraints be denoted \( \mu\Lambda \). The Lagrange multipliers corresponding to nonbinding constraints are zero. Therefore \( R\mu = R\Lambda\mu\Lambda \). Stacking the equations \( R_i'h^+ - \delta_i = 0 \) for the binding constraints, and accounting for the possibility that \( \mu_i = 0 \) for the binding constraints (since strict complementarity may not hold),

\[ R\Lambda'h^+ - \delta\Lambda = -R\Lambda'H_0^{-1} (\Delta_n + R\Lambda\mu\Lambda) - \delta\Lambda = 0 \implies \mu\Lambda = \max \left( - (R\Lambda'H_0^{-1}R\Lambda)^{-1} (R\Lambda'H_0^{-1}\Delta_n + \delta\Lambda), 0 \right) \]

Therefore,

\[ h^+ = -H_0^{-1} (\Delta_n + R\Lambda\mu\Lambda) \]
\[ = \max \left( -H_0^{-1}\Delta_n + H_0^{-1}R\Lambda (R\Lambda'H_0^{-1}R\Lambda)^{-1} (R\Lambda'H_0^{-1}\Delta_n + \delta\Lambda), -H_0^{-1}\Delta_n \right) \]
\[ = \max \left( -H_0^{-1} \left( I - R\Lambda (R\Lambda'H_0^{-1}R\Lambda)^{-1} R\Lambda'H_0^{-1} \right) \Delta_n + H_0^{-1}R\Lambda (R\Lambda'H_0^{-1}R\Lambda)^{-1} \delta\Lambda, -H_0^{-1}\Delta_n \right) \]

References

Amemiya, Takeshi, Advanced Econometrics, Harvard University Press, 1985. 7, 16

Andrews, Donald W.K., “Estimation when a parameter is on a boundary,” Econometrica, 1999, 67 (6), 1341–1383. 1, 2, 3

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, 2, 3, 17


Beck, Amir, First-order methods in optimization, Vol. 25, SIAM, 2017. 2


Horowitz, Joel L and Sokbae Lee, “Non-asymptotic inference in a class of optimization problems,” 2019. 2

Hsieh, Yu-Wei, Xiaoxia Shi, and Matthew Shum, “Inference on estimators defined by mathematical programming,” Available at SSRN 3041040, 2018. 2


_ , Francesca Molinari, and Jörg Stoye, “Confidence intervals for projections of partially identified parameters,” Econometrica, 2019, 87 (4), 1397–1432. 2


Knight, Keith, “Epi-convergence in distribution and stochastic equi--semicontinuity,” Unpublished manuscript, 1999, 37. 5, 6, 11


28


