This paper studies constrained extremum estimators with possibly non-standard asymptotic distributions, which includes constrained M-estimators and constrained GMM estimators with possibly nonsmooth, nonconvex objectives. We demonstrate how we can use test-inversion to construct a uniformly valid simultaneous confidence set which covers parameters either in the interior or on the boundary of the constraint set as well as parameters drifting towards the boundary at arbitrary rates. Our method works for constrained estimators that are not $\sqrt{n}$-consistent, but we require that we know the rate of convergence of the estimator.

*Keywords: Constrained M-estimator, Constrained GMM, non-standard asymptotics.*
1 Introduction

This paper studies estimators defined by the solution to a constrained optimization problem with non-random inequality and/or equality constraints and a possibly nonsmooth and nonconvex sample objective function. We are interested in conducting inference on the parameter defined as the solution to the population analog of the sample optimization problem. In particular, we are interested in conducting inference that is uniformly valid across different types of parameters, those that are on the boundary of the constraint set, those that are drifting towards the boundary, and those that are in the interior of the constraint set. It is well known (see e.g. Geyer (1994), Andrews (1999), Andrews (2000), and Andrews (2002)) that when the parameters are on or drifting towards the boundary of the constraint set, the asymptotic distribution of the constrained estimator is non-standard, and traditional inference procedures such as the standard bootstrap will not be pointwise or uniformly valid. Alternative inference procedures such as subsampling (Politis et al. (1999)), the numerical bootstrap (Hong and Li (2020)), or the m-out-of-n bootstrap (Bickel and Sakov (2008)) will be pointwise, but not uniformly valid because they will not produce correct coverage when the parameters are drifting towards the boundary of the constraint set. They will only be valid when the parameter is either in the interior or exactly on the boundary of the constraint set.

We propose a uniformly valid inference method using a simultaneous confidence set constructed by inverting a test statistic based on a local expansion of the constrained minimal value of the objective function around a shrinking neighborhood of the parameter of interest. We benchmark the test statistic against critical values obtained from bootstrapping consistent estimates of the components of the limiting distribution of the objective’s local expansion. This method of bootstrapping components of the limiting distribution takes in-
spiration from Cattaneo et al. (2020), but we differ from them in that we use test inversion
to conduct uniformly valid inference for the constrained argmin of the population problem,
while they are interested in pointwise valid inference for the unconstrained argmin.

Our procedure can handle constrained M-estimators with a possibly nonsmooth, non-
convex sample objective function as well as constrained GMM estimators with either cor-
rectly specified or globally misspecified nonsmooth sample moments. By globally misspec-
ified, we mean that the population moments are equal to fixed nonzero constants that do
not approach zero as $n \to \infty$. Under global misspecification, Hong and Li (2023) show that
GMM estimators with nonsmooth moments exhibit the cubic-root rate of convergence to a
nonstandard limiting distribution. We show how to extend the pointwise valid procedure
for unconstrained GMM estimators in Hong and Li (2023) to conduct uniformly valid in-
ference under constraints. We allow for both fixed and estimated weighting matrices which
can converge at various rates to their probability limits.

The statistics literature contains many papers on constrained estimation such as Shapiro
While several of these papers derive the non-standard asymptotic distributions of vari-
ous constrained estimators, they did not propose a practical inference procedure as we
do. Within econometrics, examples of relevant papers include Andrews (2001), Moon and
Schorfheide (2009), Kaido and Santos (2014), Kaido (2016), Gafarov (2016), Chen et al.
(2018), Ketz (2018), Kaido et al. (2019), Kaido et al. (2021), Horowitz and Lee (2019),
Fang and Seo (2021), Hsieh et al. (2022), Fan and Shi (2023), Ketz and McCloskey (2023),
and Chernozhukov et al. (2023). Some of these papers (e.g. Andrews (2001), Fang and
Seo (2021), Fan and Shi (2023)) are concerned with testing the validity of the constraints.
Instead, we are instead interested in conducting inference on the solution to the population
constrained optimization problem, allowing for the possibility that the constraints are mis-
specified. Additionally, some of these papers (e.g. Gafarov (2016), Hsieh et al. (2022),
Horowitz and Lee (2019), Fan and Shi (2023) ) are concerned with linear constraints or quadratic objective functions, but our method covers a large class of constrained ex-
tremum estimators with possibly nonsmooth, nonconvex objective functions and nonlinear constraints. Additionally, papers such as Geyer (1994), Moon and Schorfheide (2009), and Ketz (2018) require the parameter of interest to be a solution of the unconstrained popu-
lation optimization problem. In contrast, we allow for the possibility that the constraints matter for identification of the parameter, which means the solution of the unconstrained population optimization problem will differ from the solution of the constrained problem. However, we do not allow for partial identification because we require that there be a unique solution to the population constrained optimization problem. Additionally, we require that the objective function be defined at every value in \( \mathbb{R}^d \), which is in contrast to Ketz (2018) who point out that the objective function for random coefficient models cannot be defined for negative values of the variances.

The outline of our paper is as follows. Subsection 1.1 contains examples of constrained estimators and Subsection 1.2 contains the notation. Section 2 demonstrates how to con-
duct uniformly valid inference for constrained M-estimators, while Section 3 demonstrates how to conduct uniformly valid inference for constrained GMM estimators. Section 4 con-
tains Monte Carlo simulation evidence demonstrating the uniform validity of our proposed confidence set for a boundary constrained nonsmooth GMM model. Section 5 concludes, and the Appendix contains proofs of the theorems.
1.1 Examples of Constrained Extremum Estimators

**Example 1.** An example of a constrained estimator with a non-random constraint set is the boundary constrained maximum likelihood estimator in Andrews (2000). Suppose we have a simple location model with i.i.d data:

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

The maximum likelihood estimator subject to the constraint that \( \theta \geq 0 \) is

\[ \hat{\theta}_n = \arg \min_{\theta \geq 0} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \theta)^2 \]

It is well-known that \( \hat{\theta}_n \overset{p}{\rightarrow} \theta_0 = \max (\beta_0, 0) \).

**Example 2.** Another example is a constrained modal estimator similar to Example 3.2.13 in van der Vaart and Wellner (1996). Suppose we have the same simple location model as in the previous example. Define \( \hat{\theta}_n = \arg \max_{\theta \geq 0} \frac{1}{n} \sum_{i=1}^{n} 1 (\theta - 1 \leq y_i \leq \theta + 1) \) as the non-negative center of an interval of length 2 that contains the largest possible fraction of observations. It is well-known that \( \hat{\theta}_n \overset{p}{\rightarrow} \theta_0 = \max (\beta_0, 0) \).

**Example 3.** Another example is a nonsmooth GMM estimator with a non-negativity constraint. Our model is

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

For \( \pi (\theta) = [P (y_i \leq \theta) - \tau, E y_i - \theta]' \) and \( \hat{\pi}_n (\theta) = \left[ \frac{1}{n} \sum_{i=1}^{n} 1 (y_i \leq \theta) - \tau, \frac{1}{n} \sum_{i=1}^{n} y_i - \theta \right]' \),

\[ \hat{\theta}_n = \arg \min_{\theta \geq 0} \frac{1}{2} \hat{\pi}_n (\theta)' \hat{\pi}_n (\theta) \]
If $\tau = 0.5$, the model is correctly specified and $\hat{\theta}_n$ converges to $\theta_0$ at the $\sqrt{n}$-rate. However, if $\tau \neq 0.5$, and $\tau$ is not drifting towards 0.5, then the model is globally misspecified and $\hat{\theta}_n$ is cubic-root consistent for $\theta_0 = \arg \min_{\theta \geq 0} \frac{1}{2} \pi (\theta)' \pi (\theta)$. We will study this example in the Monte Carlo simulations.

1.2 Notation

Consider a random sample $X_n = (X_1, X_2, ..., X_n)$ of independent draws from a probability measure $P$ on a sample space $\mathcal{X}$. Define the empirical measure $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$, where $\delta_x$ is the measure that assigns mass 1 at $x$ and zero everywhere else. Denote the bootstrap empirical measure by $P_n^* = \frac{1}{n} \sum_{i=1}^{n} W_n \delta_{X_i}$, which can refer to the multinomial, wild, or other exchangeable bootstraps. An exchangeable bootstrap requires that $W_n = (W_{n1}, \ldots, W_{nn})$ is an exchangeable vector of nonnegative weights which sum to 1. For the multinomial bootstrap, $W_n$ is a multinomial random vector (independent of the data) with probabilities $(1/n, \ldots, 1/n)$. For the wild bootstrap, $P_n^* = \frac{1}{n} \sum_{i=1}^{n} (\xi_i/\bar{\xi}_n) \delta_{X_i}$, where $\xi_i$ are nonnegative i.i.d. random variables (independent of the data) with finite third moments and $\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^{n} \xi_i$. Weak convergence is defined in the sense of Kosorok (2007): $Z_n \overset{w}{\longrightarrow} Z$ in the metric space $(\mathcal{D}, d)$ if and only if $\sup_{f \in BL_1} |E^* f(Z_n) - E f(Z)| \to 0$ where $BL_1$ is the space of functions $f : \mathcal{D} \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1. $E^* f(Z_n)$ is the outer expectation of $f(Z_n)$, which is the infimum over all $EU$ where $U$ is measurable, $U \succeq f(Z_n)$, and $EU$ exists. Conditional weak convergence is also defined in the sense of Kosorok (2007): $Z_n \overset{p}{\rightharpoonup}_{W} Z$ in the metric space $(\mathcal{D}, d)$ if and only if $\sup_{f \in BL_1} |E_W f(Z_n) - E f(Z)| \xrightarrow{p} 0$ and $E_W f(Z_n)^* - E_W f(Z_n)^* \xrightarrow{p} 0$ for all $f \in BL_1$, where $BL_1$ is the space of functions $f : \mathcal{D} \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1, $E_W$ denotes expectation with respect to the bootstrap weights $W$ conditional on the data, and $f(Z_n)^*$ and $f(Z_n)^*$ denote measurable majorants.
and minorants with respect to the joint data (including the weights \( W \)). Let \( X^*_n = o_p(1) \) if \( P (|X^*_n| > \epsilon |X_n) = o_p(1) \) for all \( \epsilon > 0 \). Also define \( M^*_n = O_p(1) \) (hence also \( O_P(1) \)) if \( \lim_{m \to \infty} \limsup_{n \to \infty} P (|M^*_n| > m |X_n) > \epsilon = 0 \) for all \( \epsilon > 0 \).

### 2 Constrained M-estimator

Suppose we are interested in conducting inference on the parameter given by the solution to the population constrained optimization problem:

\[
\theta_0 = \arg \min_{\theta \in C} \left\{ \pi (\theta) \equiv P \pi (\cdot, \theta) \right\}
\]

where the constraint set \( C \subseteq \mathbb{R}^d \) is Chernoff regular at \( \theta_0 \). A constraint set \( C \) is Chernoff Regular at \( \theta_0 \) if \( C \) is well-approximated by a cone \( K \) at \( \theta_0 \), meaning that \( \inf_{w \in K} \| (\theta - \theta_0) - w \| = o (\| \theta - \theta_0 \|) \) for all \( \theta \in C \), and \( \inf_{\theta \in C} \| (\theta - \theta_0) - w \| = o (\| w \|) \) for all \( w \in K \) (see Theorem 2.1 of Geyer (1994) for more details). We allow for \( \theta_0 \) to lie either in the interior or on the boundary of \( C \), and we will estimate \( \theta_0 \) using a constrained M-estimator given by

\[
\hat{\theta}_n = \arg \min_{\theta \in C} \left\{ \hat{\pi}_n (\theta) \equiv P_n \pi (\cdot, \theta) \right\}
\]

where \( \hat{\pi}_n (\theta) \) may be nonsmooth and/or nonconvex as in Example 2.

It is well known (see e.g. Andrews (2000)) that applying a standard bootstrap procedure to estimate the distribution of the constrained M-estimator is inconsistent when \( \theta_0 \) lies on the boundary of the constraint set \( C \) or is drifting towards the boundary at some rate. Alternative inference procedures such as subsampling or the m-out-of-n bootstrap will be consistent when \( \theta_0 \) is on the boundary, but will not be consistent for parameters that are drifting towards the boundary because they will not be able to consistently estimate the
nonstandard limiting distribution of $\hat{\theta}_n$ (Andrews and Guggenberger (2010)). Instead of estimating the distribution of $\hat{\theta}_n$, we will instead try to estimate the distribution of a scaled difference between two terms involving the objective function $\hat{\pi}_n (\cdot)$. We will show that our procedure is uniformly asymptotically valid over all rates of drift for $\theta_0$.

**Assumption 1.** There exists $\rho \in (0,1]$ and $\gamma = \frac{1}{2(2-\rho)}$ such that the following conditions hold:

(i) $P_{n,\pi} (\cdot, \hat{\theta}_n) \leq \inf_{\theta \in C} P_{n,\pi} (\cdot, \theta) + o_p (n^{-2\gamma})$.

(ii) $\inf_{\theta \in C, \|\theta - \theta_0\| > \epsilon} P_{\pi} (\cdot, \theta) > P_{\pi} (\cdot, \theta_0)$ for all $\epsilon > 0$.

(iii) $\sup_{\theta \in \Theta} \|P_{n,\pi}(\cdot, \theta) - P_{\pi}(\cdot, \theta)\| = o_P(1)$.

(iv) $C$ is Chernoff regular at $\theta_0$.

**Assumption 2.** There exists $\rho \in (0,1]$ and $\gamma = \frac{1}{2(2-\rho)}$ such that $g (\cdot, \theta) = \pi (\cdot, \theta) - \pi (\cdot, \theta_0)$ satisfies the following conditions:

(i) The class $G_R = \{g (\cdot, \theta) : \|\theta - \theta_0\| \leq R\}$ for $R$ near zero is uniformly manageable for the envelope function $G_R (\cdot) = \sup_{g \in G_R} |g (\cdot, \theta)|$.

(ii) $PG^2_R = O(R^{2\rho})$ for $R \to 0$.

(iii) For each $\eta > 0$, there exists a $K$ such that $PG^2_R 1\{G_R > K\} < \eta R^{2\rho}$ for $R$ near 0.

(iv) If $\rho = 1$, $\pi (\cdot, \theta)$ is Lipschitz continuous in $\theta$ with a stochastically bounded Lipschitz constant.

(v) $\Sigma_{\rho}(s, t) = \lim_{\alpha \to \infty} \alpha^{2\rho} P g (\cdot, \theta_0 + \frac{t}{\alpha}) g (\cdot, \theta_0 + \frac{s}{\alpha})'$ exists for each $s, t$ in $\mathbb{R}^d$.

(vi) $\lim_{\alpha \to \infty} \alpha^{2\rho} P |g (\cdot, \theta_0 + \frac{t}{\alpha})|^2 1\{|g (\cdot, \theta_0 + \frac{t}{\alpha})| > \epsilon \alpha^{2(1-\rho)}\} = 0$ for each $\epsilon > 0$ and $t \in \mathbb{R}^d$.  

8
(vii) $P |g(\cdot, \theta_1) - g(\cdot, \theta_2)| = O (\|\theta_1 - \theta_2\|^{2\rho})$ for $\|\theta_1 - \theta_2\| \to 0$.

(viii) $\pi (\theta_0) = P \pi (\cdot; \theta_0)$ is twice differentiable with Jacobian $l (\theta_0) \equiv \frac{\partial}{\partial \theta} \pi (\theta_0)$ and positive definite Hessian $H_0 \equiv \frac{\partial^2}{\partial \theta \partial \theta} \pi (\theta_0)$.

Assumption 1 is needed to show consistency of $\hat{\theta}_n$ for $\theta_0$ while Assumption 2 is needed to derive the asymptotic distribution of $\hat{\theta}_n$ and other statistics which are required for our inference procedure. We assume that the researcher knows the rate of convergence coefficient $\gamma$. The square-root rate of convergence is obtained when Assumption 2 is satisfied for $\gamma = 1/2$ and $\rho = 1$, which occurs for Example 1, while the cubic-root rate of convergence is obtained when Assumption 2 is satisfied for $\gamma = 1/3$ and $\rho = 1/2$, which occurs for Example 2. Manageable classes are defined in Definition 4.1 of Pollard (1989), and an example is all Euclidean classes. A manageable class for a constant envelope is a universal Donsker class in the sense of Dudley (1987). Uniform manageable classes are manageable classes for which a uniform upper bound exists in the maximal inequalities for the corresponding empirical processes. As discussed after Corollary 3.2 of Kim and Pollard (1990), we need to assume $G_R$ are uniformly manageable in order to demonstrate stochastic equicontinuity of certain processes that appear in the expansion of the objective function. We demonstrate stochastic equicontinuity by applying the maximal inequalities in Lemma 3.1 of Kim and Pollard (1990) over the classes $G_R$ for all values of $R$ near zero, rather than a particular value of $R$.

We will impose an additional an envelope integrability condition needed to demonstrate the validity of bootstrapping certain statistics which appear in the benchmarking distribution of our inference procedure. Specifically, the condition is needed to show bootstrap equicontinuity results so that we can replace $\theta_0$ by $\hat{\theta}_n$ in the bootstrapped empirical processes.
Assumption 3. For some $\rho \in (0, 1]$ and $\gamma = \frac{1}{2(1-\rho)}$, define $m_n(\cdot, \theta, h) = n^{\gamma\rho} \left( \pi \left( \cdot; \theta + \frac{h}{n^\gamma} \right) - \pi \left( \cdot; \theta \right) \right)$. For any $\epsilon_n \to 0$ and any compact set $H \subset \mathbb{R}^d$,

$$\lim \limsup_{\lambda \to \infty} \limsup_{n \to \infty} \sup_{t \geq \lambda} P \left\{ \sup_{h \in H, \|\theta-\theta_0\| \leq \epsilon_n} \frac{\left\| m_n(\cdot, \theta, h) - m_n(\cdot, \theta_0, h) \right\|}{1 + n^{\gamma} \|\theta - \theta_0\|} > t \right\} = 0.$$  

We need Assumption 3 to make use of Lemma 4.2 in Wellner and Zhan (1996), which states that stochastic equicontinuity implies bootstrap equicontinuity under a relatively mild envelope (square) integrability assumption (their Assumption A.5). A strong sufficient condition for Assumption 3 is that the envelope is uniformly bounded. For all sufficiently large $n$ such that $\epsilon_n \to 0$ and any compact $H \subset \mathbb{R}^d$, there exists some constant $K > 0$ such that $\sup_{h \in H, \|\theta-\theta_0\| \leq \epsilon_n} \left\| m_n(\cdot, \theta, h) - m_n(\cdot, \theta_0, h) \right\| \leq K$. In the Appendix, we verify that Assumption 3 is satisfied for Examples 1 through 3.

We first consider the case where the constraints are not necessary for identification of $\theta_0$, which means that the constrained minimum is the same as the unconstrained minimum of the population objective. Assumption 4 is effectively ruling out strongly active (binding) constraints at $\theta_0$. In Examples 1-2, the constraint $\theta \geq 0$ will be binding at $\theta_0 = \max(\beta_0, 0)$ if $\beta_0 < 0$.

Assumption 4. $l(\theta_0) = \frac{\partial}{\partial \theta} \pi(\theta_0) = 0$.

The main idea of our inference procedure is that we will benchmark the distribution of a test statistic $n^{2\gamma} \left( \hat{\pi}_n(\theta) - \inf_{h \in B_{\delta_n}} \hat{\pi}_n(\theta + \frac{h}{n^\gamma}) \right)$ against the empirical distribution of $-\inf_{h \in \mathbb{R}^d} \hat{H}_n(h)$, where $B_{\delta_n} = \left\{ h \in \mathbb{R}^d : \frac{|h|}{n^\gamma} \leq \delta_n \right\}$ is a shrinking neighborhood, $\delta_n \to 0$ satisfies $n^{\gamma}\delta_n \to \kappa$ for $\kappa \in (0, \infty]$, and

$$\hat{H}_n(h) = n^{2\gamma} \left( P_n^* - P_n \right) \left( \pi \left( \cdot; \hat{\theta}_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot; \hat{\theta}_n \right) \right) + \frac{1}{2} h' \hat{H} h, \quad (1)$$
where \( \hat{H} \overset{P}{\rightarrow} H_0 \) and the bootstrap empirical measure \( P_n^* \) can refer to either the multinomial, wild, or other exchangeable bootstraps. The intuition behind the test statistic is that if \( \theta \) is the unconstrained minimum of the population objective, then the sample objective should achieve its minimum close to \( \theta \) even if we perturb \( \theta \) by small deviations that shrink to zero as \( n \rightarrow \infty \). We are able to ignore the constraints when constructing our test statistic because the troublesome term \( n^2 h' l(\theta_0) \) in the asymptotic expansion of the test statistic disappears when \( l(\theta_0) = 0 \). Since the constraints are not present in the test statistic, our benchmarking distribution also does not need to use the constraints.

Let \( \hat{c}_{1-\alpha}^* \) be the \( 1 - \alpha \) empirical quantile of \( -\inf_{h \in \mathbb{R}^d} \hat{H}_n(h) \). We will show that \( \mathcal{C}^*_{1-\alpha} = \left\{ \theta : n^{2\gamma} \left( \hat{\pi}_n(\theta) - \inf_{h \in B_n} \hat{\pi}_n(\theta + \frac{h}{n^{\gamma}}) \right) \leq \hat{c}_{1-\alpha}^* \} \) is a uniformly asymptotically valid nominal \( 1 - \alpha \) confidence set for \( \theta(P) = \theta_0 \). By uniformly valid inference, we mean inference that is uniformly valid across parameters that are either in the interior or on the boundary of the constraint set or are drifting towards the boundary of the constraint set at arbitrary rates. For Examples 1-2, this means that we can handle parameters of the form \( \theta_0 = c/\tau_n \) where \( c \geq 0 \) is some constant and \( \tau_n \rightarrow \infty \) as \( n \rightarrow \infty \).

In the next theorem, \( J_n(\cdot, P) \) denotes the CDF of \( n^{2\gamma} \left( \hat{\pi}_n(\theta_0) - \inf_{h \in B_n} \hat{\pi}_n(\theta_0 + \frac{h}{n^{\gamma}}) \right) \) under \( P \), and \( J(\cdot, P) \) denotes the CDF of its limiting distribution under \( P \). Similarly, \( J^*_n(\cdot, P) \) denotes the conditional CDF of \( -\inf_{h \in \mathbb{R}^d} \hat{H}_n(h) \) under \( P \), and \( J^*(\cdot, P) \) denotes the CDF of its limiting distribution under \( P \).

**Theorem 1.** Let \( \mathcal{P} \) be a class of distributions for which Assumptions 1-4 are satisfied uniformly in \( P \in \mathcal{P} \), and \( \{ J(\cdot, P) : P \in \mathcal{P} \} \) and \( \{ J^*(\cdot, P) : P \in \mathcal{P} \} \) are equicontinuous at

---

\(^1\)We define \( X_n = o_P(1) \) uniformly in \( P \in \mathcal{P} \) if \( \sup_{P \in \mathcal{P}} \mathbb{P}(|X_n| > \epsilon) = 0 \) \( \forall \epsilon > 0 \). We define \( M_n = O_P(1) \) uniformly in \( P \in \mathcal{P} \) if \( \lim_{m \rightarrow \infty} \sup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}} \mathbb{P}(|M_n| > m) = 0 \). Similarly, \( X_n^* = o_P^*(1) \) uniformly in \( P \in \mathcal{P} \) if \( \sup_{P \in \mathcal{P}} \mathbb{P}(|X_n^*| > \epsilon|X_n|) = o_P(1) \) for all \( \epsilon > 0 \). \( M_n^* = O_P^*(1) \) uniformly in \( P \in \mathcal{P} \) if \( \lim_{m \rightarrow \infty} \sup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}} \mathbb{P}(|M_n^*| > m|X_n| > \epsilon) = 0 \) \( \forall \epsilon > 0 \).
\[ J_n^{-1}(1 - \alpha, P). \text{Then} \lim \inf_{n \to \infty} \inf_{\theta \in P} P(\theta (P) \in C^*_{1-\alpha}) \geq 1 - \alpha, \text{where} \]

\[ C^*_{1-\alpha} = \left\{ \theta : n^{2\gamma} \left( \hat{\pi}_n (\theta) - \inf_{h \in B_{\delta_n}} \hat{\pi}_n \left( \theta + \frac{h}{n^{\gamma}} \right) \right) \leq \hat{c}^*_{1-\alpha} \right\}, \]

\[ B_{\delta_n} = \left\{ h \in \mathbb{R}^d : \frac{|h|}{n^{\gamma}} \leq \delta_n \right\}, \delta_n \to 0 \text{ satisfies } n^{\gamma} \delta_n \to \kappa \text{ for } \kappa \in (0, \infty], \text{ and } \hat{c}^*_{1-\alpha} \text{ is the } 1 - \alpha \text{ quantile of } - \inf_{h \in \mathbb{R}^d} \hat{H}_n (h). \]

**Remark 1.** If we would like to construct a nominal \( 1 - \alpha \) confidence set for a subvector \( \gamma_0 = a' \theta_0 \), where \( a \) is a known vector, we could use projection: \( \text{CI}^{\text{Proj}}_{1-\alpha} = \left[ \inf_{\theta \in C^*_{1-\alpha}} a' \theta, \sup_{\theta \in C^*_{1-\alpha}} a' \theta \right] \). The uniform asymptotic validity of these projection intervals follows directly from the uniform asymptotic validity of \( C^*_{1-\alpha} \).

Now we relax Assumption 4 to allow for possibility that the unconstrained minimum differs from the constrained minimum of the population problem \((\frac{\partial}{\partial \theta} \pi (\theta_0) \neq 0)\). This means that the constraints could matter for identification of \( \theta_0 \). We will modify our test statistic to \( n^{2\gamma} \left( \hat{\pi}_n (\theta) - \inf_{h \in C^*_{\delta_n}} \hat{\pi}_n \left( \theta + \frac{h}{n^{\gamma}} \right) \right) \), where \( C^*_{\delta_n} = \left\{ h \in n^{\gamma} (C - \theta) : \frac{|h|}{n^{\gamma}} \leq \delta_n \right\} \), and \( \delta_n \to 0 \) satisfies \( n^{\gamma} \delta_n \to \kappa \) for \( \kappa \in (0, \infty] \). In theory, any choice of \( \kappa \in (0, \infty] \) will achieve uniformly valid coverage, but if \( \hat{\pi}_n (\theta) \) is nonsmooth and/or nonconvex, in practice setting \( \kappa < \infty \) helps the solver find the solution to \( \inf_{h \in C^*_{\delta_n}} \hat{\pi}_n \left( \theta + \frac{h}{n^{\gamma}} \right) \). The intuition behind the test statistic is that if \( \theta \) is the constrained minimum of the population objective, then the sample objective should achieve its minimum close to \( \theta \) even if we perturb \( \theta \) by small deviations while still satisfying the constraints.

We cannot ignore the constraints when constructing our test statistic because the troublesome term \( n^{\gamma} h' l (\theta_0) \) in the asymptotic expansion of the test statistic can only be signed when we minimize over the constraint set instead of the entire parameter space. We will need to use the sign to find another statistic that stochastically dominates the test statistic.
and has a well-defined limiting distribution. We can then compare this other statistic to the benchmarking statistic used to form critical values and demonstrate uniform validity of our inference procedure.

Let $\hat{c}^{*}_{1-\alpha}$ be the $1-\alpha$ empirical quantile of $-\inf_{h \in \mathbb{R}^d} \hat{H}_n (h)$. We will show that $\mathcal{C}^{*}_{1-\alpha} = \left\{ \theta : n^{2\gamma} \left( \hat{\pi}_n (\theta) - \inf_{h \in \mathcal{C}^0_n} \hat{\pi}_n \left( \theta + \frac{h}{n^{\gamma}} \right) \right) \leq \hat{c}^{*}_{1-\alpha} \right\}$ is a uniformly asymptotically valid nominal $1-\alpha$ confidence set for $\theta (P) = \theta_0$. We are still benchmarking the test statistic against the unconstrained minimum of $\hat{H}_n (h)$ because we cannot uniformly consistently estimate the tangent cone of the constraint set at $\theta_0$. Since we do not observe $\theta_0$, we would have to replace $\theta_0$ by $\hat{\theta}_n$ and use a sequence $\eta_n \to \infty$ satisfying $\eta_n / n^{\gamma} \to 0$ in order to remove the additional noise caused by centering the constraint set at $\hat{\theta}_n$ instead of $\theta_0$, which introduces an additional $n^{\gamma} (\hat{\theta}_n - \theta_0)$ term. However, this procedure would only be pointwise valid because the convergence of $n^{\gamma} (\hat{\theta}_n - \theta_0)$ to its limiting distribution is not uniform over $P$.

In the next theorem, $J_n (\cdot, P)$ denotes the CDF of

$$T_n \equiv - \inf_{\{ h \in T_C (\theta_0) : \| h \| \leq \kappa \}} \left\{ n^{2\gamma} (P_n - P) \left( \pi (\cdot, \theta_0 + \frac{h}{n^{\gamma}}) - \pi (\cdot, \theta_0) \right) + \frac{1}{2} h' H_0 h \right\} \text{ under } P,$$

and $J (\cdot, P)$ denotes the CDF of $T_n$'s limiting distribution under $P$. Here, $T_C (\theta_0)$ is the tangent cone of $C$ at $\theta_0$: $T_C (\theta_0) \equiv \lim \sup_{\tau \downarrow 0} \frac{C - \theta_0}{\tau}$. Additionally, $J^*_n (\cdot, P)$ denotes the conditional CDF of $- \inf_{h \in \mathbb{R}^d} \hat{H}_n (h)$ under $P$, and $J^* (\cdot, P)$ denotes the CDF of its limiting distribution under $P$.

**Theorem 2.** Let $\mathcal{P}$ be a class of distributions for which Assumptions 1-3 are satisfied uniformly in $P \in \mathcal{P}$, and $\{ J (\cdot, P) : P \in \mathcal{P} \}$ and $\{ J^* (\cdot, P) : P \in \mathcal{P} \}$ are equicontinuous at $J^{-1}_n (1-\alpha, P)$. Then $\lim \inf \inf_{P \in \mathcal{P}} P \left( \theta (P) \in \mathcal{C}^*_{1-\alpha} \right) \geq 1 - \alpha$, where

$$\mathcal{C}^*_{1-\alpha} = \left\{ \theta : n^{2\gamma} \left( \hat{\pi}_n (\theta) - \inf_{h \in \mathcal{C}^0_n} \hat{\pi}_n \left( \theta + \frac{h}{n^{\gamma}} \right) \right) \leq \hat{c}^{*}_{1-\alpha} \right\}.$$
Remark 2. When Assumption 4 holds, both the confidence set in Theorem 2 and the confidence set in Theorem 1 will be valid. However, because it is difficult for the researcher to know if Assumption 4 holds, we recommend using the confidence set in Theorem 2.

3 Constrained GMM

We can extend our uniformly valid inference method to constrained GMM estimators with either smooth or nonsmooth moments and which are either correctly specified or globally misspecified. As shown in Hong and Li (2023), GMM estimators with nonsmooth (non-directionally differentiable) moments have the cubic-root rate of convergence when they are globally misspecified and the square-root rate of convergence when they are correctly specified. Throughout the paper, we presume that the user knows the rate of convergence, which can typically be accomplished by first testing for misspecification using a J-test. The reason we require knowing the rate of convergence is because we are interested in conducting uniformly valid inference, which is inference that is uniformly valid across parameters drifting towards the boundary of the constraint set at arbitrary rates. If we were not interested in uniformly valid inference, we could use an alternative pointwise valid procedure that does not require knowing the rate of convergence. Details are in Section 3.3.

3.1 Fixed Weighting Matrix

We first consider the case of a fixed weighting matrix $W$ which is assumed to be symmetric positive definite. Define the moment function $\pi : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^m$, sample moments $\hat{\pi}_n (\theta) \equiv \sum_{i=1}^{n} \pi \left( x_i, \theta \right)$, and the empirical moment function $\hat{\pi}_n (\theta) = \frac{1}{n} \sum_{i=1}^{n} \pi \left( x_i, \theta \right)$.
\( P_n \pi (\cdot, \theta), \) and population moments \( \pi (\theta) = P \pi (\cdot, \theta). \) The constrained GMM estimator and the pseudo-true parameter are given by

\[
\hat{\theta}_n = \arg \min_{\theta \in C} \left\{ \hat{Q}_n (\theta) = \frac{1}{2} \hat{\pi}_n (\theta)' W \hat{\pi}_n (\theta) \right\}
\]

\[
\theta_0 = \arg \min_{\theta \in C} \left\{ Q (\theta) = \frac{1}{2} \pi (\theta)' W \pi (\theta) \right\}.
\]

We will propose a uniformly valid inference procedure that will be able to handle the case of correctly specified moments \( \pi (\theta_0) = 0, \) where \( \theta_0 \) can be either on the boundary of the constraint set or drifting towards the boundary of the constraint set at arbitrary rates. We also allow for the possibility that \( \pi (\theta_0) = c \) for some fixed constants \( c \neq 0, \) meaning that the moments are globally misspecified. However, we do not allow for locally misspecified models where \( \pi (\theta_0) = c/\sqrt{n} \) is drifting towards zero at the \( \sqrt{n} \) rate. It remains an open question whether we can conduct uniformly valid inference for locally misspecified constrained GMM models.

We will slightly modify Assumptions 1 and 2 to handle constrained GMM estimators:

**Assumption 5.** There exists some \( \rho \in \left\{ \frac{1}{2}, 1 \right\} \) and \( \gamma = \frac{1}{2(2-\rho)} \) such that the following conditions are satisfied:

(i) \( \hat{Q}_n \left( \hat{\theta}_n \right) \leq \inf_{\theta \in C} \hat{Q}_n (\theta) + o_p (n^{-2\gamma}). \)

(ii) \( \inf_{\theta \in C : \| \theta - \theta_0 \| > \epsilon} Q (\theta) > Q (\theta_0) \) for all \( \epsilon > 0. \)

(iii) \( \sup_{\theta \in \Theta} \| P_n \pi (\cdot, \theta) - P \pi (\cdot, \theta) \| = o_P (1). \)

(iv) \( \sup_{\theta \in \Theta} P | \pi (\cdot, \theta) | < \infty. \)

(v) \( C \) is Chernoff regular at \( \theta_0. \)
Assumption 6. There exists some $\rho \in \{\frac{1}{2}, 1\}$ and $\gamma = \frac{1}{2(2-\rho)}$ such that $g(\cdot, \theta) \equiv \pi(\cdot, \theta) - \pi(\cdot, \theta_0)$ satisfies the following conditions:

(i) The classes of functions $G_R = \{g_j(\cdot, \theta) : \|\theta - \theta_0\| \leq R, j = 1, \ldots, m\}$ for $R$ near zero are uniformly manageable for the envelope functions $G_R(\cdot) \equiv \sup_{g_j \in G_R} |g_j(\cdot, \theta)|$.

(ii) $PG_R^2 = O(R^{2\rho})$ for $R \to 0$.

(iii) For each $\eta > 0$, there exists a $K$ such that $PG_R^2 1\{G_R > K\} < \eta R^{2\rho}$ for $R$ near 0.

(iv) If $\rho = 1$, $\pi(\cdot, \theta)$ is Lipschitz continuous in $\theta$ with a stochastically bounded Lipschitz constant.

(v) $\sum_\rho(s, t) = \lim_{\alpha \to \infty} \alpha^{2\rho} PG(\cdot, \theta_0 + \frac{t}{\alpha}) g(\cdot, \theta_0 + \frac{t}{\alpha})'$ exists for each $s, t$ in $\mathbb{R}^d$.

(vi) $\lim_{\alpha \to \infty} \alpha^{2\rho} P g(\cdot, \theta_0 + \frac{t}{\alpha}) \|g(\cdot, \theta_0 + \frac{t}{\alpha})\|^2 1\{\|g(\cdot, \theta_0 + \frac{t}{\alpha})\| > \epsilon \alpha^{2(1-\rho)}\} = 0$ for each $\epsilon > 0$ and $t \in \mathbb{R}^d$.

(vii) $P\|g(\cdot, \theta_1) - g(\cdot, \theta_2)\| = O(\|\theta_1 - \theta_2\|^{2\rho})$ for $\|\theta_1 - \theta_2\| \to 0$.

(ix) $Pg(\cdot, \theta)$ is twice differentiable at $\theta_0$ with full rank Jacobian matrix $G = \frac{\partial}{\partial \theta_0} \pi(\theta_0)$ and positive definite Hessian matrices $H_j = \frac{\partial^2}{\partial \theta_0^2} \pi_j(\theta_0)$ for $j = 1 \ldots m$.

(x) $\bar{H} = G'WG + \sum_{j=1}^m \sum_{k=1}^m W_{jk} \pi_k(\theta_0) H_j$ is positive definite.

Similar to Kim and Pollard (1990), the cubic-root rate of convergence is obtained when Assumptions 5 and 6 are satisfied for $\gamma = 1/3$ and $\rho = 1/2$. In particular, this amounts to a linear rate of decay of $PG_R^2$. Usually the linear rate of decay arises when $\pi(\cdot, \theta)$ is not directionally differentiable, such as the ones that appear in the GMM formulation of IV quantile regression (Chernozhukov and Hansen (2005)) or simulated method of moments (McFadden (1989) and Pakes and Pollard (1989)). Other types of nonsmooth moments that are directionally differentiable, such as in dynamic censored regression (Honore and
Hu (2004)), do not have this linear rate of decay and therefore retain the \( \sqrt{n} \) rate of convergence. More details about IV quantile regression, simulated method of moments, and dynamic censored regression can be found in Hong and Li (2023).

We will now describe our inference procedure. We first consider the case when the constraints are not necessary for identification of \( \theta_0 \), which means that the constrained minimum of the population objective is the same as the unconstrained minimum.

Assumption 7. \( G'W\pi (\theta_0) = 0 \), where \( G = \frac{\partial}{\partial \theta} \pi (\theta_0) \).

We will benchmark the distribution of the test statistic \( n^{2\gamma} \left( \hat{Q}_n (\theta) - \inf_{h \in B_n} \hat{Q}_n (\theta + \frac{h}{n}) \right) \) against the empirical distribution of \(- \inf_{h \in \mathcal{B}_{\delta_n}} \hat{A}_n (h)\), where \( \mathcal{B}_{\delta_n} = \{ h \in \mathbb{R}^d : \frac{|h|}{n^{\gamma}} \leq \delta_n \} \) is a shrinking neighborhood, \( \delta_n \to 0 \) satisfies \( n^{\gamma} \delta_n \to \kappa \) for \( \kappa \in (0, \infty) \), and

\[
\hat{A}_n (h) = n^{2\gamma} \hat{\pi}_n \left( \hat{\theta}_n \right)' W \left( P^*_n - P_n \right) \left( \pi \left( \cdot, \hat{\theta}_n + \frac{h}{n^{\gamma}} \right) - \pi \left( \cdot, \hat{\theta}_n \right) \right) + \frac{1}{2} h' \left( \hat{G}'W \hat{G} + \sum_{j=1}^m \sum_{k=1}^m W_{jk} \hat{\pi}_{nk} \left( \hat{\theta}_n \right) \hat{H}_j \right) h + n^{\gamma} h' \hat{G}'W \left( P^*_n - P_n \right) \pi \left( \cdot, \hat{\theta}_n \right) \tag{2}
\]

\( \hat{G} \) is a consistent estimate of \( G \), and \( \hat{H}_j \) is a consistent estimate of \( H_j = \frac{\partial^2}{\partial \theta \partial \theta} \pi_j (\theta_0) \) for \( j = 1 \ldots m \). The intuition behind the test statistic is that if \( \theta \) is the unconstrained minimum of the population objective, then the sample objective should achieve its minimum close to \( \theta \) even if we perturb \( \theta \) by small deviations that shrink to zero as \( n \to \infty \). We are able to ignore the constraints when constructing our test statistic because the troublesome term \( n^{\gamma} h' G'W \pi (\theta_0) \) in the asymptotic expansion of the test statistic disappears when \( G'W \pi (\theta_0) = 0 \). Since the constraints are not present in the test statistic, our benchmarking distribution also does not need to use the constraints. Let \( \hat{c}_{1-\alpha}^* \) be the \( 1 - \alpha \) empirical quan-
tile of $- \inf_{h \in \mathbb{R}^d} \hat{A}_n(h)$. We will show that $C_{1-\alpha} = \left\{ \theta : n^{2\gamma} \left( \hat{Q}_n(\theta) - \inf_{h \in B_{\delta_n}} \hat{Q}_n(\theta + \frac{h}{n^{\gamma}}) \right) \leq \hat{c}_{1-\alpha} \right\}$ is a uniformly asymptotically valid nominal $1 - \alpha$ confidence set for $\theta(P) = \theta_0$.

We will impose an additional assumption to show bootstrap equicontinuity results which are necessary to demonstrate the validity of our inference procedure when the GMM estimator is $\sqrt{n}$-consistent. This assumption is also used in Hong and Li (2023).

**Assumption 8.** If Assumptions 5 and 6 are satisfied for $\gamma = 1/2$, $\rho = 1$, then for any $\epsilon_n \to 0$, $\lim \limsup_{\lambda \to \infty} \sup_{\lambda \geq \lambda} P \left\{ \sup_{\|\theta - \theta_0\| \leq \epsilon_n} \frac{\|\pi(\cdot, \theta) - \pi(\cdot, \theta_0)\|}{1 + \sqrt{n}} > t \right\} = 0$. Additionally, for each $\epsilon > 0$ and $t \in \mathbb{R}^d$, $\lim_{n \to \infty} P \left\{ \left( \sqrt{n^\gamma} \left( \pi(\cdot, \theta_0 + \frac{t}{\sqrt{n}}) \right) - \pi(\cdot, \theta_0) \right) : \pi(\cdot, \theta_0) \right\} > \epsilon \sqrt{n} \right\} = 0.

In the next theorem, $J_n(\cdot, P)$ denotes the CDF of $n^{2\gamma} \left( \hat{Q}_n(\theta_0) - \inf_{h \in B_{\delta_n}} \hat{Q}_n(\theta_0 + \frac{h}{n^{\gamma}}) \right)$ under $P$, and $J(\cdot, P)$ denotes the CDF of its limiting distribution under $P$. Similarly, $J^*_n(\cdot, P)$ denotes the conditional CDF of $- \inf_{h \in B_{\delta_n}} \hat{A}_n(h)$ under $P$, and $J^*(\cdot, P)$ denotes the CDF of its limiting distribution under $P$.

**Theorem 3.** Let $\mathcal{P}$ be a class of distributions for which Assumptions 3, 5-7 and 8 are satisfied uniformly in $P \in \mathcal{P}$, and $\{J(\cdot, P) : P \in \mathcal{P}\}$ and $\{J^*(\cdot, P) : P \in \mathcal{P}\}$ are equicontinuous at $J_n^{-1}(1 - \alpha, P)$. Then $\liminf_{n \to \infty} \inf_{P \in \mathcal{P}} P \left( \theta(P) \in C_{1-\alpha} \right) \geq 1 - \alpha$, where

$$C_{1-\alpha} = \left\{ \theta : n^{2\gamma} \left( \hat{Q}_n(\theta) - \inf_{h \in B_{\delta_n}} \hat{Q}_n \left( \theta + \frac{h}{n^{\gamma}} \right) \right) \leq \hat{c}_{1-\alpha} \right\},$$

$$B_{\delta_n} = \left\{ h \in \mathbb{R}^d : \frac{\|h\|}{n^{\gamma}} \leq \delta_n \right\}, \delta_n \to 0 \text{ satisfies } n^{\gamma} \delta_n \to \kappa \text{ for } \kappa \in (0, \infty], \text{ and } \hat{c}_{1-\alpha} \text{ is the } 1 - \alpha \text{ quantile of } - \inf_{h \in B_{\delta_n}} \hat{A}_n(h).$$

Now suppose the constraints may be necessary for identification of $\theta_0$ so that Assump-
tion 7 may not hold. We will modify our test statistic to 

\[ n^{2\gamma} \left( \hat{Q}_n (\theta) - \inf_{h \in C^\delta_{\hat{\theta}_n}} \hat{Q}_n (\theta + \frac{h}{n^{\gamma}}) \right), \]

where \( C^\delta_{\hat{\theta}_n} = \{ h \in n^{\gamma} (C - \theta) : \frac{|h|}{n^{\gamma}} \leq \delta_n \} \), and \( \delta_n \to 0 \) satisfies \( n^{\gamma} \delta_n \to \kappa \) for \( \kappa \in (0, \infty) \). The intuition behind the test statistic is that if \( \theta \) is the constrained minimum of the population objective, then the sample objective should achieve its minimum close to \( \theta \) even if we perturb \( \theta \) by small deviations while still satisfying the constraints. In theory, the choice of \( \kappa \) does not matter for achieving uniformly valid coverage, but if \( \hat{Q}_n (\theta) \) is nonsmooth and/or nonconvex, in practice setting \( \kappa < \infty \) helps the solver find the solution to \( \inf_{h \in C^\delta_{\hat{\theta}_n}} \hat{Q}_n (\theta + \frac{h}{n^{\gamma}}) \).

We cannot ignore the constraints when constructing our test statistic because the troublesome term \( n^{\gamma} h'G'W\pi (\theta_0) \) in the asymptotic expansion of the test statistic can only be signed when we minimize over the constraint set instead of the entire parameter space. We will need to use the sign to find another statistic that stochastically dominates the test statistic and has a well-defined limiting distribution. We can then compare this other statistic to the benchmarking statistic used to form critical values and demonstrate uniform validity of our inference procedure.

Let \( \hat{c}_{1-\alpha}^* \) be the \( 1 - \alpha \) empirical quantile of \( -\inf_{h \in \mathbb{R}^d} \hat{A}_n (h) \). We are still benchmarking against the unconstrained minimum of \( \hat{A}_n (h) \) because we cannot uniformly consistently estimate the tangent cone of the constraint set at \( \theta_0 \). Since we do not observe \( \theta_0 \), we would have to replace \( \theta_0 \) by \( \hat{\theta}_n \) and use a sequence \( \eta_n \to \infty \) satisfying \( \eta_n / n^{\gamma} \to 0 \) in order to remove the additional noise caused by centering the constraint set at \( \hat{\theta}_n \) instead of \( \theta_0 \), which introduces an additional \( n^{\gamma} \left( \hat{\theta}_n - \theta_0 \right) \) term. However, this procedure would be only pointwise valid because the convergence of \( n^{\gamma} \left( \hat{\theta}_n - \theta_0 \right) \) to its limiting distribution is not uniform over \( P \).

We will show that \( C_{1-\alpha}^* = \left\{ \theta : n^{2\gamma} \left( \hat{Q}_n (\theta) - \inf_{h \in C^\delta_{\hat{\theta}_n}} \hat{Q}_n (\theta + \frac{h}{n^{\gamma}}) \right) \leq \hat{c}_{1-\alpha}^* \right\} \) is a uniformly asymptotically valid nominal \( 1 - \alpha \) confidence set for \( \theta (P) = \theta_0 \). Our inference procedure is
uniformly valid across parameters that are either in the interior or on the boundary of the constraint set or are drifting towards the boundary at arbitrary unknown rates. Let \( J_n (\cdot, P) \) be the CDF of \( S_n = - \inf_{h \in \mathcal{H}_C (\theta_0): \|h\| < \kappa} \left\{ n^{2 \gamma} \left( \hat{Q}_2 (\theta_0 + \frac{h}{n^{\gamma}}) + \hat{Q}_3 (\theta_0 + \frac{h}{n^{\gamma}}) \right) + \frac{1}{2} h' \hat{H} h \right\} \) under \( P \), where \( \hat{Q}_2 (\cdot) \) and \( \hat{Q}_3 (\cdot) \) are defined in 6, and let \( J (\cdot, P) \) denote the CDF of \( S_n \)'s limiting distribution under \( P \). Additionally, let \( J^\ast (\cdot, P) \) denote the CDF of \( - \inf_{h \in \mathbb{R}^d} \hat{A}_n (h) \) under \( P \), and let \( J^\ast (\cdot, P) \) denote the CDF of its limiting distribution under \( P \).

**Theorem 4.** Let \( \mathcal{P} \) be a class of distributions for which Assumptions 3, 5-6 and 8 are satisfied uniformly in \( P \in \mathcal{P} \), and \( \{ J (\cdot, P) : P \in \mathcal{P} \} \) and \( \{ J^\ast (\cdot, P) : P \in \mathcal{P} \} \) are equicontinuous at \( J_n^{-1} (1 - \alpha, P) \). Then \( \lim \inf_{n \to \infty} \inf_{P \in \mathcal{P}} P (\theta (P) \in \mathcal{C}^\ast_{1 - \alpha}) \geq 1 - \alpha \), where

\[
\mathcal{C}^\ast_{1 - \alpha} = \left\{ \theta : n^{2 \gamma} \left( \hat{Q}_n (\theta) - \inf_{h \in \mathcal{C}^{\delta_n}_{1 - \alpha}} \hat{Q}_n \left( \theta + \frac{h}{n^{\gamma}} \right) \right) \leq \hat{c}^\ast_{1 - \alpha} \right\},
\]

\( \mathcal{C}^{\delta_n}_{1 - \alpha} = \left\{ h \in n^{\gamma} (C - \theta) : \frac{|h|}{n^{\gamma}} \leq \delta_n \right\} \), \( \delta_n \to 0 \) satisfies \( n^{\gamma} \delta_n \to \kappa \) for \( \kappa \in (0, \infty) \), and \( \hat{c}^\ast_{1 - \alpha} \) is the \( 1 - \alpha \) quantile of \( - \inf_{h \in \mathbb{R}^d} \hat{A}_n (h) \).

3.2 Estimated Weighting Matrix

We now consider the case of an estimated weighting matrix where the 2-step GMM estimator is \( \hat{\theta}_n = \arg \min_{\theta \in \mathcal{C}} \left\{ Q_n (\theta) = \frac{1}{2} \hat{\pi} (\theta)' W_n \hat{\pi} (\theta) \right\} \). It can be that the estimated weighting matrix \( W_n = W_n \left( \hat{\theta}_1 \right) \) depends on the unconstrained 1-step GMM estimator \( \hat{\theta}_1 = \arg \min_{\theta \in \Theta} \frac{1}{2} \hat{\pi} (\theta)' W_1 \hat{\pi} (\theta) \), or it can be that \( W_n = W_n \left( \hat{\theta}_1^C \right) \) depends on the constrained 1-step GMM estimator \( \hat{\theta}_1^C = \arg \min_{\theta \in \mathcal{C}} \frac{1}{2} \hat{\pi} (\theta)' W_1 \hat{\pi} (\theta) \). Note that we need to redefine the presumed to be unique pseudo-true parameter to be \( \theta_0 = \arg \min_{\theta \in \mathcal{C}} \frac{1}{2} \pi (\theta)' W \pi (\theta) \) where \( W \) depends on the presumed to be unique 1-step GMM pseudo-true parameter using some fixed weighting matrix \( W_1 \). If we are using the unconstrained 1-step GMM estimator, then \( W = W (\theta_1) \),
where \( \theta_1 = \arg\min_{\theta \in \Theta} \frac{1}{2} \pi (\theta)' W_1 \pi (\theta) \). If we are using the constrained 1-step GMM estimator, then \( W = W (\theta_1^C) \), where \( \theta_1^C = \arg\min_{\theta \in C} \frac{1}{2} \pi (\theta)' W_1 \pi (\theta) \). The choice of which \( W \) to use can be determined by whether the constraints matter for identification of the parameters. If \( G_1' W_1 \pi (\theta_1^C) = 0 \), where \( G_1 = \frac{1}{\sqrt{n}} \pi (\theta_1^C) \), then \( \theta_1^C = \theta_1 \) and both \( W = W (\theta_1) \) and \( W = W (\theta_1^C) \) lead to the same pseudo-true parameter \( \theta_0 \). However, if \( G' W_1 \pi (\theta_1^C) \neq 0 \), meaning that the constraints matter for identification, then \( \theta_1^C \) will differ from \( \theta_1 \), and depending on whether we set \( W = W (\theta_1) \) or \( W = W (\theta_1^C) \), we can obtain different values of \( \theta_0 \). Because we would like to enforce the constraints to identify the parameters, we would typically use \( W = W (\theta_1^C) \) in this case.

The presence of the estimated weighting matrix adds an additional source of variation which needs to be accounted for when constructing our confidence set. We will benchmark the distribution of the test statistic \( n^{2\gamma} \left( \hat{Q}_n (\theta) - \inf_{h \in C_{\theta_n}^\alpha_h} Q_n (\theta + \frac{h}{n^{\gamma}}) \right) \) against the empirical distribution of \(- \inf_{h \in \mathbb{R}^d} \hat{B}_n (h) \), where

\[
\hat{B}_n (h) = n^{2\gamma} \tilde{\pi}_n \left( \tilde{\theta}_n \right)' W_n (P_n^* - P_n) \pi \left( \cdot, \tilde{\theta}_n + \frac{h}{n^{\gamma}} \right) - \pi \left( \cdot, \tilde{\theta}_n \right) \\
+ \frac{1}{2} h' \left( \hat{G}' W_n \hat{G} + \sum_{j=1}^m \sum_{k=1}^m W_{n,jk} \tilde{\pi}_{nk} (\tilde{\theta}_n) \hat{H}_j \right) h \\
+ n^{\gamma} h' \hat{G}' W_n (P_n^* - P_n) \pi \left( \cdot, \tilde{\theta}_n \right) \\
+ n^{\gamma} h' \hat{G}' (W_{n}^* - W_n) \tilde{\pi}_n \left( \tilde{\theta}_n \right).
\]

\( W_n^* \) is the bootstrap analog of the weighting matrix and depends on a consistent bootstrap analog of the 1-step estimator using a fixed weighting matrix \( W_1 \). If we are using the
constrained 1-step estimator, then \( W_n^* = W_n^* \left( \hat{\theta}_1^{C*} \right) \), where

\[
\hat{\theta}_1^{C*} = \arg \min_{\theta \in \Theta} \left\{ \tilde{\pi}_n \left( \hat{\theta}_1^{C*} \right)' W_1 (P_n^* - P_n) \left( \pi (\cdot, \theta) - \pi (\cdot, \hat{\theta}_1^{C}) \right) \right. \\
+ \frac{1}{2} \left( \theta - \hat{\theta}_1^{C*} \right)' \left( \hat{G}' W_1 \hat{G} + \sum_{j=1}^{m} \sum_{k=1}^{m} W_1, jk \tilde{\pi}_{nk} (\hat{\theta}_1^{C*}) \hat{H}_j \right) \left( \theta - \hat{\theta}_1^{C*} \right) \]
\begin{align}
&+ \left( \theta - \hat{\theta}_1^{C*} \right)' \hat{G}' W_1 (P_n^* - P_n) \pi \left( \cdot, \hat{\theta}_1^{C*} \right) \right\},
\end{align}

and \( \eta_n \to \infty \) is a sequence that satisfies \( \eta_n/n^\gamma \to 0 \). The purpose of this sequence is to remove the additional noise caused by centering the constraint set around \( \hat{\theta}_1^{C} \) instead of the unknown \( \theta_1^{C} \).

If we are using the unconstrained 1-step estimator, then \( W_n^* = W_n^* \left( \hat{\theta}_1^{*} \right) \), where

\[
\hat{\theta}_1^{*} = \arg \min_{\theta \in \Theta} \left\{ \tilde{\pi}_n \left( \hat{\theta}_1^{*} \right)' W_1 (P_n^* - P_n) \left( \pi (\cdot, \theta) - \pi (\cdot, \hat{\theta}_1) \right) \right. \\
+ \frac{1}{2} \left( \theta - \hat{\theta}_1^{*} \right)' \left( \hat{G}' W_1 \hat{G} + \sum_{j=1}^{m} \sum_{k=1}^{m} W_1, jk \tilde{\pi}_{nk} (\hat{\theta}_1) \hat{H}_j \right) \left( \theta - \hat{\theta}_1^{*} \right) \]
\begin{align}
&+ \left( \theta - \hat{\theta}_1^{*} \right)' \hat{G}' W_1 (P_n^* - P_n) \pi \left( \cdot, \hat{\theta}_1^{*} \right) \right\}.
\end{align}

We will impose an additional assumption regarding the allowable rates of convergence of the estimated weighting matrix. For a given rate of convergence of the estimator, we allow for the weighting matrix to converge at either the same rate or a faster rate. We also require joint weak convergence assumptions for terms involving the estimated weighting matrix and the empirical processes that appear from expanding the moments. The joint weak convergence assumptions are also used in Hong and Li (2023).

**Assumption 9.** \( \gamma, W_n \) and \( W_n^* \) can fall into one of the following cases:
(i) \( \gamma = 1/2, W_n - W = o_p(n^{-1/2}) \) and \( W_n^* - W_n = o_p^*(n^{-1/2}) \).

(ii) \( \gamma = 1/2, W_n - W = O_p(n^{-1/2}) \) and \( W_n^* - W_n = O_p^*(n^{-1/2}) \).

(iii) \( \gamma = 1/3, W_n - W = o_p(n^{-1/3}) \) and \( W_n^* - W_n = o_p^*(n^{-1/3}) \).

(iv) \( \gamma = 1/3, W_n - W = O_p(n^{-1/3}) \) and \( W_n^* - W_n = O_p^*(n^{-1/3}) \).

Furthermore, if \( \gamma = 1/2, W_n - W = O_p(n^{-1/2}) \) and \( W_n^* - W_n = O_p^*(n^{-1/2}) \), then
\[
\begin{pmatrix}
\pi(\theta_0)' W_n n (P_n - P) g(\cdot, \theta_0 + n^{-1/2}h) \\
h' G' W_n \sqrt{n} (P_n - P) \pi(\cdot, \theta_0)
\end{pmatrix} \overset{v}{\longrightarrow} \begin{pmatrix}
\pi(\theta_0)' W Z_{0,1}(h) \\
h' G' W U_0
\end{pmatrix}
\]
and
\[
\begin{pmatrix}
\hat{\pi}_n(\hat{\theta}_n)' W_n n (P_n^* - P_n) \pi(\cdot, \hat{\theta}_n) \\
h' \hat{G}' W_n \sqrt{n} (P_n^* - P_n) \pi(\cdot, \hat{\theta}_n)
\end{pmatrix} \overset{p}{\longrightarrow} \begin{pmatrix}
\pi(\theta_0)' W Z_{0,1}(h) \\
h' G' W U_0
\end{pmatrix}
\]
are asymptotically equivalent in the product space of locally bounded functions \( \{B_{loc}(\mathbb{R}^d)\}^3 \) for some tight random vector \( W_0 \).

Here, \( U_0 \sim N(0, P(\pi(\cdot, \theta_0) - \pi(\theta_0))(\pi(\cdot, \theta_0) - \pi(\theta_0))') \) and \( Z_{0,1}(h) \) is a mean zero Gaussian process with covariance kernel \( \Sigma_1(s,t) = \lim_{\alpha \to \infty} \alpha^2 P g(\cdot, \theta_0 + \frac{s}{\alpha}) g(\cdot, \theta_0 + \frac{t}{\alpha})' \).

If \( \gamma = 1/3, W_n - W = O_p(n^{-1/3}) \) and \( W_n^* - W_n = O_p^*(n^{-1/3}) \), then
\[
\begin{pmatrix}
\pi(\theta_0)' W_n^{2/3} (P_n - P) g(\cdot, \theta_0 + n^{-1/3}h) \\
h' G' n^{1/3} (W_n - W) \pi(\theta_0)
\end{pmatrix} \overset{v}{\longrightarrow} \begin{pmatrix}
\pi(\theta_0)' W Z_{0,1/2}(h) \\
h' G' W_0
\end{pmatrix}
\]
and
\[
\begin{pmatrix}
\pi(\theta_0)' W_n^{2/3} (P_n^* - P_n) g(\cdot, \theta_0 + n^{-1/3}h) \\
h' \hat{G}' n^{1/3} (W_n^* - W_n) \pi(\theta_0)
\end{pmatrix} \overset{p}{\longrightarrow} \begin{pmatrix}
\pi(\theta_0)' W Z_{0,1/2}(h) \\
h' \hat{G}' W_0
\end{pmatrix}
\]
in the product space of locally bounded functions \( \{B_{loc}(\mathbb{R}^d)\}^2 \) for some tight random vector \( W_0 \). Here, \( Z_{0,1/2}(h) \) is a mean zero Gaussian process with covariance kernel
\[
\Sigma_{1/2}(s,t) = \lim_{\alpha \to \infty} \alpha P g(\cdot, \theta_0 + \frac{s}{\alpha}) g(\cdot, \theta_0 + \frac{t}{\alpha})'.
\]

Let \( \hat{c}_{1-\alpha}^* \) be the \( 1 - \alpha \) empirical quantile of \( \inf_{h \in \mathbb{R}^d} \hat{B}_n(h) \). We will show that
\[ C_{1-\alpha}^* = \left\{ \theta : n^{2\gamma} \left( \hat{Q}_n(\theta) - \inf_{h \in C_{\delta_n}^0} \hat{Q}_n \left( \theta + \frac{h}{n^\gamma} \right) \right) \leq \hat{c}_{1-\alpha}^* \right\} \] is a uniformly asymptotically valid nominal \( 1 - \alpha \) confidence set for \( \theta(P) = \theta_0 \). In the next theorem, we let \( J_n(\cdot, P) \) denote the CDF of \( R_n = -\inf_{h \in \mathcal{T}_C(\theta_0) : \|h\| \leq \kappa} \left\{ n^{2\gamma} \hat{Q}_n(\theta_0 + \frac{h}{n^\gamma}) + \frac{1}{2} h' \hat{H} h \right\} \) under \( P \), where \( \hat{Q}_n(\cdot) \) is defined in equation 7, and let \( J(\cdot, P) \) denote the CDF of \( R_n \)'s limiting distribution under \( P \). Additionally, let \( J_{\alpha_n}^*(\cdot, P) \) denote the conditional CDF of \( -\inf_{h \in \mathbb{R}^d} \hat{B}_n(h) \) under \( P \), and let \( J^*(\cdot, P) \) denote the CDF of its limiting distribution under \( P \).

**Theorem 5.** Let \( \mathcal{P} \) be a class of distributions for which Assumptions 3, 5-6 and 8-9 are satisfied uniformly in \( P \in \mathcal{P} \), and \( \{ J(\cdot, P) : P \in \mathcal{P} \} \) and \( \{ J^*(\cdot, P) : P \in \mathcal{P} \} \) are equicontinuous at \( J_n^{-1}(1 - \alpha, P) \). Then \( \lim_{n \to \infty} \inf_{P \in \mathcal{P}} P \left( \theta(P) \in C_{1-\alpha}^* \right) \geq 1 - \alpha \), where

\[ C_{1-\alpha}^* = \left\{ \theta : n^{2\gamma} \left( \hat{Q}_n(\theta) - \inf_{h \in C_{\delta_n}^0} \hat{Q}_n \left( \theta + \frac{h}{n^\gamma} \right) \right) \leq \hat{c}_{1-\alpha}^* \right\}, \]

\[ C_\delta^0 = \left\{ h \in n^\gamma (C - \theta) : \frac{|h|}{n^\gamma} \leq \delta_n \right\}, \delta_n \to 0 \text{ satisfies } n^\gamma \delta_n \to \kappa \text{ for } \kappa \in (0, \infty], \text{ and } \hat{c}_{1-\alpha}^* \text{ is the } 1 - \alpha \text{ quantile of } -\inf_{h \in \mathbb{R}^d} \hat{B}_n(h). \]

### 3.3 Pointwise Valid Rate-Adaptive Inference for Constrained GMM

If we were only interested in pointwise valid, rather than uniformly valid inference for
\[ \theta_0 = \arg\min_{\theta \in C} \frac{1}{2} \pi(\theta)' W \pi(\theta), \]
we can conduct rate-adaptive inference, which does not require knowing the rate of convergence \( \gamma \). The procedure for constrained GMM with a fixed weighting matrix is as follows:

1. Compute \( \hat{\theta}_n = \arg\min_{\theta \in C} \frac{1}{2} \hat{\pi}_n(\theta)' W \hat{\pi}_n(\theta), \hat{\pi}_n(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^n \pi(X_i, \hat{\theta}_n), \hat{G}, \hat{H}_j \) for \( j = 1 \ldots m \). Pick \( \eta_n \to \infty \) which is a sequence that satisfies \( \eta_n/n^{1/3} \to 0 \).
2. Repeat for \( B \) bootstrap iterations: draw a bootstrap sample \( X_1^*, \ldots, X_n^* \) and compute

\[
\hat{\theta}_n^* = \arg\min_{\theta \in C_{\eta_n}^{\hat{\theta}_n}} \left\{ \hat{\pi}_n \left( \hat{\theta}_n \right)' W \left( \frac{1}{n} \sum_{i=1}^{n} \left( \pi \left( X_i^*, \theta \right) - \pi \left( X_i^*, \hat{\theta}_n \right) \right) \right)^2 + \frac{1}{2} \left( \theta - \hat{\theta}_n \right)' \left( \hat{G}' W \hat{G} + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{jk} \hat{\pi}_{nk} \left( \hat{\theta}_n \right) \hat{H}_j \right) \left( \theta - \hat{\theta}_n \right) \right\}.
\]

3. For \( k = 1, \ldots, d \), compute the \( 1 - \alpha/2 \) and \( \alpha/2 \) percentiles of the empirical distribution of \( \hat{\theta}_{nk}^* - \hat{\theta}_{nk} \). Call them \( c_{k,1-\alpha/2} \) and \( c_{k,\alpha/2} \).

A nominal \( 1 - \alpha \) two-sided equal-tailed confidence interval for \( \theta_{0k} \) can be formed by

\[
\left[ \hat{\theta}_{nk} - c_{k,1-\alpha/2}, \hat{\theta}_{nk} - c_{k,\alpha/2} \right].
\]

A nominal \( 1 - \alpha \) confidence interval for \( \rho \left( \theta_0 \right) \), where \( \rho : \Theta \rightarrow \mathbb{R} \), can be formed using the \( 1 - \alpha/2 \) and \( \alpha/2 \) percentiles of the empirical distribution of \( \rho \left( \hat{\theta}_n^* \right) - \rho \left( \hat{\theta}_n \right) \), denoted \( c_{\rho,1-\alpha/2} \) and \( c_{\rho,\alpha/2} \): \( \left[ \rho \left( \hat{\theta}_n \right) - c_{\rho,1-\alpha/2}, \rho \left( \hat{\theta}_n \right) - c_{\rho,\alpha/2} \right] \).

We will show that the proposed bootstrap procedure will pointwise consistently estimate the distribution of \( \hat{\theta}_n \) without having to know the rate of convergence coefficient \( \gamma \). Since we know that \( \gamma \geq 1/3 \) for the class of GMM estimators we consider, the sequence \( \eta_n \) will satisfy \( \eta_n / n^{\gamma} \rightarrow 0 \) and remove the additional noise caused by centering the constraint set around \( \hat{\theta}_n \) instead of the unknown \( \theta_0 \).

**Theorem 6.** Suppose Assumptions 3, 5-6 and 8 hold, \( \hat{G} \overset{p}{\rightarrow} G \), and \( \hat{H}_j \overset{p}{\rightarrow} H_j \) for \( j = 1 \ldots m \).

\[
n^{\gamma} \left( \hat{\theta}_n - \theta_0 \right) \overset{\sim}{\rightarrow} \arg\min_{h \in T_{C_{\theta_0}}} \mathbb{H}_0 \left( h \right), \quad n^{\gamma} \left( \hat{\theta}_n^* - \hat{\theta}_n \right) \overset{\mathbb{W}}{\rightarrow} \arg\min_{h \in T_{C_{\theta_0}}} \mathbb{H}_0 \left( h \right)
\]

where \( \bar{H} = G'WG + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{jk} \pi_k \left( \theta_0 \right) H_j, T_C \left( \theta_0 \right) = \lim sup_{\tau \downarrow 0} C_{\theta_0}^{\eta_{\tau}}, \mathbb{H}_0 \left( h \right) = \pi \left( \theta_0 \right)' W Z_{0,1} \left( h \right) + \ldots
\]
\[ h'G'W_0 + \frac{1}{2} h'\hat{H}h \text{ for } \gamma = 1/2, \text{ and } \mathbb{H}_0(h) = \pi(\theta_0)' W \mathcal{Z}_{0,1/2}(h) + \frac{1}{2} h'\hat{H}h \text{ when } \gamma = 1/3. \]

Here, \( U_0 \sim N\left(0, P\left(\pi(\cdot, \theta_0) - \pi(\theta_0)\right) \left(\pi(\cdot, \theta_0) - \pi(\theta_0)\right)'\right) \), \( \mathcal{Z}_{0,1}(h) \) is a mean zero Gaussian process with covariance kernel \( \Sigma_1(s,t) = \lim_{\alpha \to \infty} \alpha^2 P g(\cdot, \theta_0 + \frac{z}{\alpha}) g(\cdot, \theta_0 + \frac{L}{\alpha})' \), and \( \mathcal{Z}_{0,1/2}(h) \) is a mean zero Gaussian process with covariance kernel \( \Sigma_{1/2}(s,t) = \lim_{\alpha \to \infty} \alpha P g(\cdot, \theta_0 + \frac{z}{\alpha}) g(\cdot, \theta_0 + \frac{L}{\alpha})' \).

4 Monte Carlo

4.1 Nonsmooth Location Model

Consider a simple location model with i.i.d data,

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

For \( \pi(\cdot, \theta) = [1 \left( y_i \leq \theta \right) - \tau; y_i - \theta]' \), the population moments are \( \pi(\theta) = [P \left( y_i \leq \theta \right) - \tau; Ey_i - \theta]' \).

The model cannot be correctly specified as long as \( \tau \neq 0.5 \), and the nonsmoothness of the moments leads to a cubic-root rate of convergence under global misspecification. The standard bootstrap will undercover the parameter of interest whenever \( \tau \neq 0.5 \), with the undercoverage becoming more severe as \( \tau \) moves further away from 0.5. The rate-adaptive bootstrap in Hong and Li (2023) can conduct pointwise valid inference for misspecified GMM models without constraints, but here we would like to be able to conduct inference that is uniformly valid across drifting parameters while imposing a non-negativity constraint.

We first use a fixed weighting matrix \( W = I \), and consider the following GMM criterion function and its probability limit:

\[ \hat{Q}_n(\theta) = \frac{1}{2} \hat{\pi}_n(\theta)' \hat{\pi}_n(\theta) = \frac{1}{2} \left( \frac{1}{n} \sum_{i=1}^{n} \left[ \left( y_i \leq \theta \right) - \tau \right] \right)^2 + \frac{1}{2} \left( \frac{1}{n} \sum_{i=1}^{n} y_i - \theta \right)^2 \]
\[ Q(\theta) = \frac{1}{2} \pi'(\theta)'\pi(\theta) = \frac{1}{2} (P(y_\ell \leq \theta) - \tau)^2 + \frac{1}{2} (Ey_i - \theta)^2. \]

We are interested in conducting uniformly valid inference on the pseudo true value given by \( \theta_0 = \arg\min_{\theta \in C} Q(\theta) \), where the constraint set is \( C = \{ \theta : \theta \geq 0 \} \). We call \( \theta_0 \) the pseudo true value because it will not be equal to \( \beta_0 \) if \( \tau \neq 0.5 \) or \( \beta_0 < 0 \). If we wish to conduct inference on \( \beta_0 \), we must have correctly specified moments and correctly specified constraints, which means \( \tau = 0.5 \) and \( \beta_0 \geq 0 \).

We use the confidence set \( C_{1-\alpha}^* = \left\{ \theta : n^{2\gamma} \left( \hat{Q}_n(\theta) - \inf_{h \in C_{\theta_0}^\delta} \hat{Q}_n(\theta + \frac{h}{n^{\gamma}}) \right) \leq \hat{c}_{1-\alpha}^* \right\} \), where \( C_{\theta_0}^\delta = \left\{ h \in n^{\gamma} (C - \theta) : \frac{|h|}{n^{\gamma}} \leq \delta_n \right\} \), \( \delta_n \to 0 \) satisfies \( n^{\gamma} \delta_n \to \kappa \) for \( \kappa \in (0, \infty) \), and \( \hat{c}_{1-\alpha}^* \) is the \( 1 - \alpha \) quantile of \( \inf_{h \in \mathbb{R}^d} \hat{A}_n(h) \) for \( \hat{A}_n(h) \) given in equation 2. In all of our simulations, we used \( \kappa = 5 \) and Matlab’s patternsearch routine to compute the constrained GMM estimator \( \hat{\theta}_n = \arg\min_{\theta \in C} \hat{Q}_n(\theta) \). Additionally, we estimate \( G \) and \( H \) using kernel estimators:

\[
\hat{G}_n = \begin{bmatrix}
\frac{1}{nh} \sum_{i=1}^n K_h \left( y_i - \hat{\theta}_n \right) \\
-1
\end{bmatrix}, \quad \hat{H}_n = \begin{bmatrix}
\frac{1}{nh^2} \sum_{i=1}^n K'_h \left( y_i - \hat{\theta}_n \right) \\
0
\end{bmatrix},
\]

\[ K_h(x) = K(x/h), \quad K(x) = (2\pi)^{-1/2} \exp(-x^2/2), \quad K'_h(x) = K'(x/h) \)

\[ K'(x) = -(2\pi)^{-1/2} x \exp(-x^2/2), \text{ where } h \text{ is Silverman’s rule-of-thumb bandwidth } h = 1.06n^{-1/5}. \]

We also tried different bandwidth values and the results were not affected.

Table 1 shows the empirical coverage frequencies and average interval lengths (in parentheses) of nominal 95% confidence intervals. We consider a range of different values of \( \beta_0 \in \{0, n^{-1}, n^{-1/2}, n^{-1/3}, n^{-1/4}, n^{-1/6}, 1\} \), where \( n \in \{100, 500, 1000, 5000\} \). We consider three different values of \( \tau \in \{0.1, 0.3, 0.5\} \), where the first two values of \( \tau \) correspond to misspecified models with the cubic-root rate of convergence and the last value corresponds to a correctly specified model with the square-root rate of convergence. We use \( B = 1000 \)
bootstrap iterations, and \( R = 2000 \) Monte Carlo simulations. The coverage is above the nominal level for values of \( \beta_0 \) close to zero (when the constraint becomes weakly active), but approaches the nominal level as \( \beta_0 \) becomes more positive (when the constraint becomes inactive). For any given sample size, the average interval length is fairly small and does not change much across the different values of \( \beta_0 \).

Table 1: Coverage Frequencies and Average Interval Lengths

<table>
<thead>
<tr>
<th>( \beta_0 )</th>
<th>0</th>
<th>( n^{-1} )</th>
<th>( n^{-1/2} )</th>
<th>( n^{-1/3} )</th>
<th>( n^{-1/4} )</th>
<th>( n^{-1/6} )</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 100 )</td>
<td>1.000</td>
<td>1.000</td>
<td>0.990</td>
<td>0.979</td>
<td>0.964</td>
<td>0.956</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td>(0.559)</td>
<td>(0.543)</td>
<td>(0.472)</td>
<td>(0.430)</td>
<td>(0.424)</td>
<td>(0.421)</td>
<td>(0.420)</td>
</tr>
<tr>
<td>( n = 500 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.991</td>
<td>0.964</td>
<td>0.963</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td>(0.369)</td>
<td>(0.360)</td>
<td>(0.303)</td>
<td>(0.232)</td>
<td>(0.211)</td>
<td>(0.211)</td>
<td>(0.211)</td>
</tr>
<tr>
<td>( n = 1000 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.980</td>
<td>0.957</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td>(0.328)</td>
<td>(0.327)</td>
<td>(0.276)</td>
<td>(0.195)</td>
<td>(0.161)</td>
<td>(0.159)</td>
<td>(0.159)</td>
</tr>
<tr>
<td>( n = 5000 )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.969</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>(0.291)</td>
<td>(0.289)</td>
<td>(0.264)</td>
<td>(0.186)</td>
<td>(0.103)</td>
<td>(0.085)</td>
<td>(0.086)</td>
</tr>
</tbody>
</table>

Table 2 shows the empirical coverage frequencies and average interval lengths (in parentheses) of nominal 95% confidence intervals. We consider a range of different values of \( \beta_0 \in \{0, -n^{-1}, -n^{-1/2}, -n^{-1/3}, -n^{-1/4}, -n^{-1/6}, -1\} \). Because negative values of \( \beta_0 \) violate
the non-negativity constraint, the non-negativity constraint is misspecified and therefore strongly active (binding) at $\theta_0 = \arg\min_{\theta \geq 0} Q(\theta)$, which leads to the coverage for $\theta_0$ being quite conservative. However, the average interval lengths are not particularly wide because we are covering the pseudo-true parameter $\theta_0$ rather than the true parameter $\beta_0$.

Table 2: Coverage Frequencies and Average Interval Lengths

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>0</th>
<th>$-n^{-1}$</th>
<th>$-n^{-1/2}$</th>
<th>$-n^{-1/3}$</th>
<th>$-n^{-1/4}$</th>
<th>$-n^{-1/6}$</th>
<th>$-1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau = 0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.556)</td>
</tr>
<tr>
<td>(0.367)</td>
</tr>
<tr>
<td>(0.327)</td>
</tr>
<tr>
<td>(0.283)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau = 0.3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.460)</td>
</tr>
<tr>
<td>(0.252)</td>
</tr>
<tr>
<td>(0.205)</td>
</tr>
<tr>
<td>(0.156)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.404)</td>
</tr>
<tr>
<td>(0.184)</td>
</tr>
<tr>
<td>(0.130)</td>
</tr>
<tr>
<td>(0.059)</td>
</tr>
</tbody>
</table>

Now we consider the case of an estimated weighting matrix. The variance-covariance
matrix of the moments is

\[ E(\pi(\cdot, \theta) - \pi(\theta)) \pi(\cdot, \theta) - \pi(\theta))' = \begin{bmatrix} F_y(\theta) - F_y(\theta)^2 & -f_y(\theta) \\ -f_y(\theta) & 1 \end{bmatrix}. \]

We consider using an estimate of the inverse of the variance-covariance matrix of the moments as our weighting matrix:

\[ W_n(\hat{\theta}_1^C) = \begin{bmatrix} \hat{F}_y(\hat{\theta}_1^C) - \hat{F}_y(\hat{\theta}_1^C)^2 & -\hat{f}_y(\hat{\theta}_1^C) \\ -\hat{f}_y(\hat{\theta}_1^C) & 1 \end{bmatrix}^{-1}, \]

where \( \hat{\theta}_1^C = \arg\min_{\theta \geq 0} \frac{1}{2} \hat{\pi}_n(\theta)' \hat{\pi}_n(\theta) \) is the constrained 1-step estimator using the identity weighting matrix, \( \hat{f}_y(\hat{\theta}_1^C) = \frac{1}{nh} \sum_{i=1}^n K_h(y_i - \hat{\theta}_1^C) \), and \( \hat{F}_y(\hat{\theta}_1^C) = \frac{1}{n} \sum_{i=1}^n 1(y_i \leq \hat{\theta}_1^C) \).

The bootstrapped weighting matrix is computed using the multinomial bootstrap:

\[ W_n^*(\hat{\theta}_1^{C*}) = \begin{bmatrix} \hat{F}_y^*(\hat{\theta}_1^{C*}) - \hat{F}_y(\hat{\theta}_1^{C*})^2 & -\hat{f}_y^*(\hat{\theta}_1^{C*}) \\ -\hat{f}_y^*(\hat{\theta}_1^{C*}) & 1 \end{bmatrix}^{-1}, \]

where \( \hat{\theta}_1^{C*} \) is the constrained bootstrap estimate in equation 4 using \( W_1 = I \) and \( \eta_n = n^{1/4} \), \( \hat{f}_y^*(\hat{\theta}_1^{C*}) = \frac{1}{nh} \sum_{i=1}^n K_h(y_i^* - \hat{\theta}_1^{C*}) \), and \( \hat{F}_y^*(\hat{\theta}_1^{C*}) = \frac{1}{n} \sum_{i=1}^n 1(y_i^* \leq \hat{\theta}_1^{C*}) \). We use the same Silverman’s Rule of Thumb bandwidth as before \( h = 1.06\text{std}(y)n^{-1/5} \).

We want to conduct uniformly valid inference on the pseudo true value given by

\[ \theta_0 = \arg\min_{\theta \geq 0} \frac{1}{2} \pi(\theta)' W(\theta) \pi(\theta) \] where \( W(\theta) = \begin{bmatrix} F_y(\theta) - F_y(\theta)^2 & -f_y(\theta) \\ -f_y(\theta) & 1 \end{bmatrix}^{-1} \)

and \( \theta_1^C = \arg\min_{\theta \geq 0} \frac{1}{2} \pi(\theta)' \pi(\theta) \). Our uniformly asymptotically valid nominal \( 1-\alpha \) confidence set is

\[ C_{1-\alpha}^* = \left\{ \theta : n^{2\gamma} \left( \hat{Q}_n(\theta) - \inf_{h \in C_{\eta_n}^h} \hat{Q}_n(\theta + \frac{h}{n}) \right) \leq \hat{c}_{1-\alpha}^* \right\}, \] where \( \hat{Q}_n(\theta) = \frac{1}{2} \hat{\pi}_n(\theta)' W_n(\hat{\theta}_1^C) \hat{\pi}_n(\theta) \),

\[ C_\delta_n^\eta = \left\{ h \in n^{\gamma} (C - \theta) : n^{\eta_\delta_n} \leq \delta_n \right\}, \delta_n \to 0 \] satisfies \( n^{\gamma_\delta_n} \delta_n \to \kappa \), and \( \hat{c}_{1-\alpha}^* \) is the \( 1-\alpha \) quantile.
of $- \inf_{h \in \mathbb{R}^d} \hat{B}_n (h)$ for $\hat{B}_n (h)$ given in equation 3.

Table 3: Coverage Frequencies and Average Interval Lengths

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>0</th>
<th>$n^{-1}$</th>
<th>$n^{-1/2}$</th>
<th>$n^{-1/3}$</th>
<th>$n^{-1/4}$</th>
<th>$n^{-1/6}$</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100$</td>
<td>0.987</td>
<td>0.983</td>
<td>0.985</td>
<td>0.986</td>
<td>0.975</td>
<td>0.961</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>1.462</td>
<td>1.490</td>
<td>1.435</td>
<td>1.460</td>
<td>1.533</td>
<td>1.498</td>
<td>1.513</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.986</td>
<td>0.982</td>
<td>0.986</td>
<td>0.972</td>
<td>0.970</td>
<td>0.963</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td>0.901</td>
<td>0.888</td>
<td>0.864</td>
<td>0.871</td>
<td>0.861</td>
<td>0.903</td>
<td>0.881</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.977</td>
<td>0.980</td>
<td>0.983</td>
<td>0.981</td>
<td>0.969</td>
<td>0.960</td>
<td>0.959</td>
</tr>
<tr>
<td></td>
<td>0.705</td>
<td>0.712</td>
<td>0.681</td>
<td>0.685</td>
<td>0.687</td>
<td>0.698</td>
<td>0.704</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>0.983</td>
<td>0.979</td>
<td>0.986</td>
<td>0.975</td>
<td>0.964</td>
<td>0.961</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td>0.401</td>
<td>0.400</td>
<td>0.396</td>
<td>0.379</td>
<td>0.384</td>
<td>0.383</td>
<td>0.397</td>
</tr>
</tbody>
</table>

| $\tau = 0.1$ |

| $n = 100$ | 0.980 | 0.981 | 0.971 | 0.956 | 0.930 | 0.930 | 0.933 |
|           | 0.855 | 0.885 | 0.814 | 0.781 | 0.808 | 0.786 | 0.774 |
| $n = 500$ | 0.981 | 0.980 | 0.981 | 0.958 | 0.935 | 0.929 | 0.938 |
|           | 0.501 | 0.499 | 0.471 | 0.452 | 0.437 | 0.440 | 0.436 |
| $n = 1000$ | 0.975 | 0.979 | 0.980 | 0.963 | 0.938 | 0.928 | 0.926 |
|           | 0.401 | 0.400 | 0.382 | 0.368 | 0.362 | 0.354 | 0.358 |
| $n = 5000$ | 0.982 | 0.978 | 0.982 | 0.959 | 0.942 | 0.952 | 0.946 |
|           | 0.251 | 0.250 | 0.242 | 0.230 | 0.225 | 0.222 | 0.224 |

| $\tau = 0.3$ |

| $n = 100$ | 0.969 | 0.968 | 0.954 | 0.935 | 0.924 | 0.924 | 0.916 |
|           | 0.410 | 0.414 | 0.395 | 0.400 | 0.401 | 0.404 | 0.406 |
| $n = 500$ | 0.966 | 0.963 | 0.960 | 0.950 | 0.945 | 0.940 | 0.948 |
|           | 0.191 | 0.189 | 0.182 | 0.185 | 0.186 | 0.185 | 0.185 |
| $n = 1000$ | 0.970 | 0.963 | 0.961 | 0.936 | 0.950 | 0.939 | 0.947 |
|           | 0.136 | 0.136 | 0.131 | 0.133 | 0.133 | 0.133 | 0.133 |
| $n = 5000$ | 0.973 | 0.976 | 0.966 | 0.946 | 0.947 | 0.955 | 0.951 |
|           | 0.062 | 0.062 | 0.060 | 0.060 | 0.061 | 0.061 | 0.061 |

| $\tau = 0.5$ |

Tables 3 and 4 show the empirical coverage frequencies and average interval lengths (in parentheses) of nominal 95% confidence intervals. We consider a range of different values of $\beta_0 \in \{0, n^{-1}, n^{-1/2}, n^{-1/3}, n^{-1/4}, n^{-1/6}, 1\}$, where $n \in \{100, 500, 1000, 5000\}$. We consider three different values of $\tau \in \{0.1, 0.3, 0.5\}$ and use $B = 1000$ bootstrap iterations and $R = 2000$ Monte Carlo simulations. For the positive values of $\beta_0$, the coverage is close to the nominal level for sufficiently large values of $n$ and is less conservative when $\beta_0$ is further away from the boundary of the constraint set. For the negative values of $\beta_0$, the
parameter average interval lengths are not particularly wide because we are covering the pseudo-true $n$

\[ \text{Table 4: Coverage Frequencies and Average Interval Lengths} \]

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>0</th>
<th>$-n^{-1}$</th>
<th>$-n^{-1/2}$</th>
<th>$-n^{-1/3}$</th>
<th>$-n^{-1/4}$</th>
<th>$-n^{-1/6}$</th>
<th>$-1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100$</td>
<td>0.987</td>
<td>0.984</td>
<td>0.989</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\tau = 0.1$</td>
<td>1.243</td>
<td>1.246</td>
<td>1.300</td>
<td>1.393</td>
<td>1.447</td>
<td>1.470</td>
<td>1.524</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.986</td>
<td>0.984</td>
<td>0.992</td>
<td>0.996</td>
<td>0.996</td>
<td>0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>$\tau = 0.3$</td>
<td>0.830</td>
<td>0.818</td>
<td>0.841</td>
<td>0.887</td>
<td>0.895</td>
<td>0.846</td>
<td>0.760</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.977</td>
<td>0.984</td>
<td>0.986</td>
<td>0.996</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\tau = 0.5$</td>
<td>0.679</td>
<td>0.682</td>
<td>0.682</td>
<td>0.720</td>
<td>0.720</td>
<td>0.660</td>
<td>0.572</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>0.983</td>
<td>0.980</td>
<td>0.991</td>
<td>0.993</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

\[ \begin{align*}
\tau &= 0.1 \\
(1.243) & (1.246) (1.300) (1.393) (1.447) (1.470) (1.524) \\
\tau &= 0.3 \\
(0.830) & (0.818) (0.841) (0.887) (0.895) (0.846) (0.760) \\
\tau &= 0.5 \\
(0.679) & (0.682) (0.682) (0.720) (0.720) (0.660) (0.572) \\
(0.983) & (0.980) (0.991) (0.993) (0.999) (1.000) (1.000) \\
\end{align*} \]

\[ \begin{align*}
\tau &= 0.1 \\
(1.243) & (1.246) (1.300) (1.393) (1.447) (1.470) (1.524) \\
\tau &= 0.3 \\
(0.830) & (0.818) (0.841) (0.887) (0.895) (0.846) (0.760) \\
\tau &= 0.5 \\
(0.679) & (0.682) (0.682) (0.720) (0.720) (0.660) (0.572) \\
(0.983) & (0.980) (0.991) (0.993) (0.999) (1.000) (1.000) \\
\end{align*} \]

The non-negativity constraint is misspecified and therefore strongly active at $\theta_0 = \arg \min_{\theta \geq 0} Q(\theta)$. This misspecification causes the coverage for $\theta_0$ to be quite conservative. However, the average interval lengths are not particularly wide because we are covering the pseudo-true parameter $\theta_0$ rather than the true parameter $\beta_0$. 

32
5 Conclusion

We have proposed an inference procedure for parameters defined by the solution to constrained optimization problems with non-random constraints. We allow the sample objective to be nonsmooth, nonconvex, and the rate of convergence of the constrained estimator to be different from the $\sqrt{n}$ rate, thus allowing for constrained M-estimators with nonstandard limiting distributions as well as globally misspecified nonsmooth constrained GMM estimators. We have demonstrated that our confidence set has uniformly valid coverage across a range of different parameters which can be either in the interior or on the boundary of the constraint set or are drifting towards the boundary at arbitrary rates.

6 Appendix

6.1 Proofs of Theorems

6.1.1 Proof of Theorem 1

Consider any sequence \( \{ P^{(n)} \in \mathcal{P} : n \geq 1 \} \) that determines \( \theta_n = \theta \left( P^{(n)} \right) \) and the laws of all random variables. Denote the empirical measure as \( P_n \) and the bootstrap empirical measure as \( P^*_n \). Consistency of \( \hat{\theta}_n \) for \( \theta_n \) follows from Assumption 1 and constraining \( \theta \) to lie in \( C \) when applying Corollary 3.2.3 in van der Vaart and Wellner (1996). We already showed in the proof of Theorem 4.1 of Hong and Li (2020) that under Assumption 2, \( n^{2\gamma} \left( P_n - P^{(n)} \right) \left( \pi \left( \cdot, \theta_n + \frac{h_n}{n^\delta} \right) - \pi \left( \cdot, \theta_n \right) \right) \) converges in finite dimensional distribution to a mean zero Gaussian process \( Z_{0,\rho}(h) \) with covariance kernel

\[
\Sigma_{\rho}(s, t) = \lim_{\alpha \to \infty} \alpha^{2\rho} P^{(n)} g \left( \cdot, \theta_n + \frac{s}{\alpha} \right) g \left( \cdot, \theta_n + \frac{t}{\alpha} \right)'.
\]
Additionally, \( n^{2\gamma} P^{(n)} \left( \pi \left( \cdot, \theta_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \theta_n \right) \right) = h' n^{\gamma} l (\theta_n) + \frac{1}{2} h' H_0 h + o(1) = \frac{1}{2} h' H_0 h + o(1) \) since we assumed in Assumption 4 that \( l(\theta_n) = 0 \). Therefore,

\[
\begin{align*}
& n^{2\gamma} \hat{\pi}_n \left( \theta_n + \frac{h}{n^\gamma} \right) - n^{2\gamma} \hat{\pi}_n (\theta_n) \\
& = n^{2\gamma} (P_n - P^{(n)}) \left( \pi \left( \cdot, \theta_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \theta_n \right) \right) + \frac{1}{2} h' H_0 h + o_P(1) \\
& \rightsquigarrow Z_{0,\rho} (h) + \frac{1}{2} h' H_0 h,
\end{align*}
\]

as a process indexed by \( h \) in the space of locally bounded functions \( B_{\text{loc}} (\mathbb{R}^d) \) equipped with the topology of uniform convergence on compacta.

Theorem 3.6.13 in van der Vaart and Wellner (1996) or Theorem 2.6 in Kosorok (2007) then implies that the bootstrapped process \( n^{2\gamma} (P_n^* - P_n) \left( \pi \left( \cdot, \theta_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \theta_n \right) \right) \) is consistent for the same limiting process as \( n^{2\gamma} (P_n - P^{(n)}) \left( \pi \left( \cdot, \theta_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \theta_n \right) \right) \):

\[
\begin{align*}
& n^{2\gamma} (P_n^* - P_n) \left( \pi \left( \cdot, \theta_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \theta_n \right) \right) \xrightarrow{\mathbb{P}} Z_{0,\rho} (h).
\end{align*}
\]

Next we show that for every sequence of positive numbers \( \{\epsilon_n\} \) converging to zero,

\[
\begin{align*}
n^{2\gamma} \sup_{d \in \mathcal{D}(n)} \left\| (P_n - P^{(n)}) d \right\| = o_P(1),
\end{align*}
\]

where \( \mathcal{D}(n) = \{d(\cdot, \theta_n, h_1, h_2) = g(\cdot; \theta_n + n^{-\gamma} h_1) - g(\cdot; \theta_n + n^{-\gamma} h_2) \text{ such that} \max(\|h_1\|,\|h_2\|) \leq M \text{ and } \|h_1 - h_2\| \leq \epsilon_n\} \). Note that \( \mathcal{D}(n) \) has envelope function \( D_n = 2G_{R(n)} \) where \( R(n) = Mn^{-\gamma} \).
The maximal inequality states that

\[ n^{2\gamma} P^{(n)} \sup_{d \in \mathcal{D}(n)} \left| (P_n - P^{(n)}) d \right| \leq P^{(n)} \sqrt{n^{2\gamma} P_n D_n^2 J} \left( \frac{n^{2\gamma} \sup_{d \in \mathcal{D}(n)} P_n d^2}{n^{2\gamma} P_n D_n^2} \right). \]

Note that \( P^{(n)} n^{2\gamma} P_n D_n^2 = O \left( n^{2\gamma} \left( n^{-\gamma} \right)^2 \right) = O(1) \). Next, for each \( K > 0 \), \( P^{(n)} \sup_{d \in \mathcal{D}(n)} d^2 1 \{ D_n > K \} \leq P^{(n)} P_n \sup_{d \in \mathcal{D}(n)} P_n |d| \leq P^{(n)} P_n D_n^2 1 \{ D_n > K \} + K \sup_{d \in \mathcal{D}(n)} P_n |d| \)

\[ + K P^{(n)} \sup_{d \in \mathcal{D}(n)} |P_n| - P^{(n)}|d| \]. For the first term, for large enough \( K \), there exists some \( \eta > 0 \) such that \( P^{(n)} P_n D_n^2 1 \{ D_n > K \} < \eta n^{-2\gamma} \). For the second term, \( K \sup_{d \in \mathcal{D}(n)} P_n |d| = O \left( n^{-2\gamma} \epsilon_n \right) = o \left( n^{-2\gamma} \right) \). For the third term, if \( \rho < 1 \),

\[ KP^{(n)} \sup_{d \in \mathcal{D}(n)} |P_n| - P^{(n)}|d| \leq Kn^{-1/2} J(1) \sqrt{P^{(n)} D_n^2} \]

\[ = O \left( n^{-\gamma(\rho+1/2)} \right) = O \left( n^{-2\gamma} \right) = o \left( n^{-2\gamma} \right) \]

In the case where \( \rho = 1 \) and \( \gamma = 1/2 \), because we assumed in Assumption 2 that \( \pi(\cdot, \theta) \) is Lipschitz in \( \theta \) with a stochastically bounded Lipschitz constant, we have that \( D_n = O_{P^{(n)}} \left( n^{-1/2} \epsilon_n \right) \). We can then use the maximal inequality in Section 3.1 of Kim and Pollard (1990) to show \( K P^{(n)} \sup_{d \in \mathcal{D}(n)} |P_n| - P^{(n)}|d| \leq Kn^{-1/2} J(1) \sqrt{P^{(n)} D_n^2} = O(n^{-1} \epsilon_n) = o(n^{-1}) \).

Therefore, \( P^{(n)} \sup_{d \in \mathcal{D}(n)} P_n d_n^2 = o(1) \).

We have shown \( n^{2\gamma} \sup_{d \in \mathcal{D}(n)} \left| (P_n - P^{(n)}) d \right| = o_{P^{(n)}}(1) \), which implies that

\[ n^{2\gamma} \sup_{h \in \mathcal{H}} \left| (P_n - P^{(n)}) \left( \pi \left( \hat{\theta}_n + \frac{h}{n^{\gamma}} \right) - \pi \left( \hat{\theta}_n \right) \right) - \left( \pi \left( \hat{\theta}_n + \frac{h}{n^{\gamma}} \right) - \pi \left( \hat{\theta}_n \right) \right) \right| \]

\[ = o_{P^{(n)}} \left( 1 + n^{\gamma} \left\| \hat{\theta}_n - \theta_n \right\| \right) = o_{P^{(n)}}(1) \]

Since \( \lim_{\lambda \to \infty} \lim_{n \to \infty} \sup_{t \geq \lambda} \left\{ \sup_{h \in \mathcal{H}, |\theta - \hat{\theta}_n| < \epsilon_n} \frac{m_n(\cdot, \theta, h) - m_n(\cdot, \hat{\theta}_n, h)}{1 + n^{\gamma} |\theta - \hat{\theta}_n|} > t \right\} = 0 \) for any \( \epsilon_n \to 0 \) and any
compact set $\mathcal{H} \subset \mathbb{R}^d$ by Assumption 3, Lemma 4.2 in Wellner and Zhan (1996) implies that

$$
\begin{align*}
\sup_{h \in \mathcal{H}} n^{2\gamma} & \left| (P_n^* - P_n) \left( \pi \left( \cdot, \hat{\theta}_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \hat{\theta}_n \right) - \left( \pi \left( \cdot, \theta_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \theta_n \right) \right) \right| \\
& = \mathcal{O}_{P(n)} \left( 1 + n^\gamma \left\| \hat{\theta}_n - \theta_n \right\| \right) = \mathcal{O}_{P(n)} (1)
\end{align*}
$$

Therefore, since $\hat{H} \overset{P}{\to} H$,

$$
\hat{H}_n (h) = n^{2\gamma} (P_n^* - P_n) \left( \pi \left( \cdot, \hat{\theta}_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \hat{\theta}_n \right) \right) + \frac{1}{2} h' \hat{H} h
$$

$$
\xrightarrow{\mathbb{P}} \mathcal{Z}_{0, \rho} (h) + \frac{1}{2} h' H_0 h \equiv \mathbb{H}_0 (h)
$$

Then the continuous mapping results in Lemma 10.11 of Kosorok (2007) imply $- \inf_{h \in \mathbb{R}^d} \hat{H}_n (h) \xrightarrow{\mathbb{P}} - \inf \mathbb{H}_0 (h)$. Additionally, since $\frac{\sqrt{n^{\gamma\rho}}}{n^{2\gamma}} = 1$,

$$
n^{2\gamma} \left( \hat{\pi}_n (\theta_n) - \inf_{h \in \mathcal{B}_n} \hat{\pi}_n \left( \theta_n + \frac{h}{n^\gamma} \right) \right) \xrightarrow{\mathbb{P}} - \inf_{h \in \{ h \in \mathbb{R}^d \mid \| h \| \leq \kappa \}} \mathbb{H}_0 (h)
$$

Since $J \left( \cdot, P^{(n)} \right)$ is equicontinuous at $J^{-1}_n \left( 1 - \alpha, P^{(n)} \right)$, we have for $x_n = J^{-1}_n \left( 1 - \alpha - \epsilon, P^{(n)} \right)$, $J_n \left( x_n, P^{(n)} \right) - J \left( x_n, P^{(n)} \right) = o(1)$ for any $P^{(n)}$ and $\epsilon$ small enough. Since $J^* \left( \cdot, P^{(n)} \right)$ is also equicontinuous at $J^{-1}_n \left( 1 - \alpha, P^{(n)} \right)$, we have for any $P^{(n)}$ and $\epsilon$ small enough,

$$
J^{-1}_n \left( x_n, P^{(n)} \right) - J^* \left( x_n, P^{(n)} \right) = o_P (1).
$$

Note that $- \inf_{h \in \{ h \in \mathbb{R}^d \mid \| h \| \leq \kappa \}} \mathbb{H}_0 (h) \leq - \inf_{h \in \mathbb{R}^d} \mathbb{H}_0 (h)$ for any realizations of the random variables, which means $J^* \left( x_n, P^{(n)} \right) < J \left( x_n, P^{(n)} \right)$ for all $n$ large enough. Then, for all $\epsilon > 0$ and $n$ large enough, there exists $\delta > 0$ such that $P^{(n)} \left( J^{-1}_n \left( x_n, P^{(n)} \right) - J_n \left( x_n, P^{(n)} \right) > \epsilon \right) \leq \delta$. If $J^{-1}_n \left( x_n, P^{(n)} \right) - J_n \left( x_n, P^{(n)} \right) \leq \epsilon$, then $J^{-1}_n \left( 1 - \alpha - \epsilon, P^{(n)} \right) \leq J^{-1}_n \left( 1 - \alpha, P^{(n)} \right)$.

Then, using arguments similar to those in Lemma A.1 (vi) of Romano and Shaikh (2012),
for all $\epsilon > 0$ and $n$ large enough,

$$P^{(n)} \left( n^{2\gamma} \left( \hat{\pi}_n(\theta_n) - \inf_{h \in B_{h_n}} \hat{\pi}_n(\theta_n + \frac{h}{n^\gamma}) \right) \leq J_{\alpha_n}^\ast \left( 1 - \alpha, P^{(n)} \right) \right)$$

$$\geq P^{(n)} \left( n^{2\gamma} \left( \hat{\pi}_n(\theta_n) - \inf_{h \in B_{h_n}} \hat{\pi}_n(\theta_n + \frac{h}{n^\gamma}) \right) \leq J_{\alpha_n}^\ast \left( 1 - \alpha, P^{(n)} \right) \right)$$

$$\geq P^{(n)} \left( n^{2\gamma} \left( \hat{\pi}_n(\theta_n) - \inf_{h \in B_{h_n}} \hat{\pi}_n(\theta_n + \frac{h}{n^\gamma}) \right) \leq J_{\alpha_n}^\ast \left( 1 - \alpha, P^{(n)} \right) \right)$$

$$- P^{(n)} \left( J_{\alpha_n}^\ast \left( x_n, P^{(n)} \right) - J_n \left( x_n, P^{(n)} \right) > \epsilon \right)$$

$$\geq 1 - \alpha - \epsilon - \delta$$

Since $\epsilon$ and $\delta$ can be arbitrarily small, $\liminf_{n \to \infty} P^{(n)} \left( n^{2\gamma} \left( \hat{\pi}_n(\theta_n) - \inf_{h \in B_{h_n}} \hat{\pi}_n(\theta_n + \frac{h}{n^\gamma}) \right) \leq \hat{c}_{1-\alpha}^\ast \right) \geq 1 - \alpha$. For $\rho = \liminf_{n \to \infty} \inf_{P \in P} \left( n^{2\gamma} \left( \hat{\pi}_n(\theta_n) - \inf_{h \in B_{h_n}} \hat{\pi}_n(\theta_n + \frac{h}{n^\gamma}) \right) \leq \hat{c}_{1-\alpha}^\ast \right)$, we can find a sequence $\{P^{(n)} \in P\}$ such that $\rho = \liminf_{n \to \infty} P^{(n)} \left( n^{2\gamma} \left( \hat{\pi}_n(\theta_n) - \inf_{h \in B_{h_n}} \hat{\pi}_n(\theta_n + \frac{h}{n^\gamma}) \right) \leq \hat{c}_{1-\alpha}^\ast \right)$.

Find a subsequence $n_k$ of $n$ for which $\theta_n$ converges, with its limit denoted $\theta$. The same arguments as above applied to such a subsequence imply

$$\liminf_{n_k \to \infty} P^{(n_k)} \left( n_k^{2\gamma} \left( \hat{\pi}_{n_k}(\theta_{n_k}) - \inf_{h \in B_{h_{n_k}}} \hat{\pi}_{n_k}(\theta_{n_k} + \frac{h}{n_k^\gamma}) \right) \leq \hat{c}_{1-\alpha}^\ast \right) \geq 1 - \alpha. \text{ Since } \{P^{(n_k)}, \theta_{n_k}\}$$

is a subsequence of $\{P^{(n)}, \theta_n\}$, $\rho = \liminf_{n_k \to \infty} P^{(n_k)} \left( n_k^{2\gamma} \left( \hat{\pi}_{n_k}(\theta_{n_k}) - \inf_{h \in B_{h_{n_k}}} \hat{\pi}_{n_k}(\theta_{n_k} + \frac{h}{n_k^\gamma}) \right) \leq \hat{c}_{1-\alpha}^\ast \right) \geq 1 - \alpha.$

\[\blacksquare\]

6.1.2 Proof of Theorem 2

Recall that $n^{2\gamma} P \left( \pi \left( \cdot, \theta_0 + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \theta_0 \right) \right) = h' n^\gamma l(\theta_0) + \frac{1}{2} h' H_0 h + o(1)$. Additionally, Chernoff regularity implies that $+\infty \left( h \notin n^\gamma \left( C - \theta_0 \right) \right) \overset{\epsilon}{\rightarrow} +\infty \left( h \notin T_C(\theta_0) \right)$, where $T_C(\theta_0) \equiv \limsup_{\tau \to 0} \frac{\theta_0 - \theta_0}{\tau}$. Note that when $\theta_0 \in \partial C$, $h' l(\theta_0) \geq 0$ for all $h \in T_C(\theta_0)$. Otherwise, there would exist some descent direction $h \in T_C(\theta_0)$ that reduces the value of the objective function and $\theta_0$ would not be the constrained minimum of the population objective. Then, for
any \( c \in \mathbb{R} \),

\[
\limsup_{n \to \infty} \sup_{P \in \mathcal{P}} \left( n^{2\gamma} \left( \hat{\pi}_n (\theta_0) - \inf_{\hat{P} \in \mathcal{P}} \left( \hat{\pi}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) \right) > c \right) = \limsup_{n \to \infty} \sup_{P \in \mathcal{P}} \left( -n^{2\gamma} \inf_{\hat{P} \in \mathcal{P}} \left\{ (P_n - P) \left( \pi (\cdot, \theta_0 + \frac{h}{n^{\gamma}}) - \pi (\cdot, \theta_0) \right) + h' (\theta_0) + \frac{1}{2} h' H_0 h \right\} > c \right)
\]

Because the test statistic \( n^{2\gamma} \left( \hat{\pi}_n (\theta_0) - \inf_{\hat{P} \in \mathcal{P}} \left( \hat{\pi}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) \right) \) does not have a well-defined limiting distribution when \( h' n^{\gamma} (\theta_0) \neq 0 \), we will instead define \( J_n (\cdot, P) \) as the CDF of \( -\inf_{\{h \in \mathbb{R}^d : |h| \leq \kappa\}} \left\{ n^{2\gamma} (P_n - P) \left( \pi (\cdot, \theta_0 + \frac{h}{n^{\gamma}}) - \pi (\cdot, \theta_0) \right) + h' (\theta_0) + \frac{1}{2} h' H_0 h \right\} \) under \( P \), and \( J (\cdot, P) \) as the CDF of its limiting distribution \( -\inf_{\{h \in \mathbb{R}^d : |h| \leq \kappa\}} \mathbb{H}_0 (h) \) under \( P \). The definition of \( J^*_{\alpha_n} (\cdot, P) \) remains the same as in Theorem 1, denoting the conditional CDF of \( -\inf_{\{h \in \mathbb{R}^d : |h| \leq \kappa\}} \mathbb{H}_0 (h) \) under \( P \), and \( J^* (\cdot, P) \) still denotes the CDF of its limiting distribution \( -\inf_{\{h \in \mathbb{R}^d : |h| \leq \kappa\}} \mathbb{H}_0 (h) \) under \( P \). Note that \( -\inf_{\{h \in \mathbb{R}^d : |h| \leq \kappa\}} \mathbb{H}_0 (h) \) is stochastically dominated by \( -\inf_{\{h \in \mathbb{R}^d : |h| \leq \kappa\}} \mathbb{H}_0 (h) \). The result follows from modifying the proof of Theorem 1 to incorporate the new definitions of \( J_n (\cdot, P) \) and \( J (\cdot, P) \), and replacing \( n^{2\gamma} \left( \hat{\pi}_n (\theta_0) - \inf_{\hat{P} \in \mathcal{P}} \left( \hat{\pi}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) \right) \) by \( n^{2\gamma} \left( \hat{\pi}_n (\theta_0) - \inf_{\hat{P} \in \mathcal{P}} \left( \hat{\pi}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) \right) \).

6.1.3 Proof of Theorem 3

Consider any sequence \( \{P^{(n)} \in \mathcal{P} : n \geq 1\} \) that determines \( \theta_n = \theta (P^{(n)}) \) and the laws of all random variables.

Define \( \hat{G}_n (\theta) = \sqrt{n} \left( P_n - P^{(n)} \right) g (\cdot, \theta), \hat{g} (\theta) = P_n g (\cdot, \theta), \) and \( g (\theta) = P^{(n)} g (\cdot, \theta) \). Then \( \hat{\pi}_n (\theta) = g (\theta) + \hat{\pi}_n (\theta_n) + \hat{\eta}_n (\theta) \), where \( \hat{\eta}_n (\theta) = \frac{1}{\sqrt{n}} \hat{G}_n (\theta) \). Recall that \( \hat{Q}_n (\theta) = \)

38
where

\[ Q_1 (\theta) = \frac{1}{2} g (\theta)^r W g (\theta) + g (\theta)^r W \pi (\theta_n), \quad \hat{Q}_3 (\theta) = \pi (\theta_n)^r W \hat{\eta}_n (\theta) \]

\[ \hat{Q}_2 (\theta) = \frac{1}{2} \hat{\eta}_n (\theta)^r W \hat{\eta}_n (\theta) + g (\theta)^r W (\hat{\pi}_n (\theta_n) - \pi (\theta_n)) + g (\theta)^r W \hat{\eta}_n (\theta) + (\hat{\pi}_n (\theta_n) - \pi (\theta_n))^r W \hat{\eta}_n (\theta). \]

(6)

We showed in the proofs of Theorems 1 and 5 of Hong and Li (2023) that

\[ n^{2 \gamma} \left( \hat{Q}_n (\theta_n + n^{-\gamma} h) - \hat{Q}_n (\theta_n) \right) = n^{2 \gamma} Q_1 (\theta_n + n^{-\gamma} h) + n^{2 \gamma} \left( \hat{Q}_2 (\theta_n + n^{-\gamma} h) + \hat{Q}_3 (\theta_n + n^{-\gamma} h) \right) \]

where \( n^{2 \gamma} \left( \hat{Q}_2 (\theta_n + n^{-\gamma} h) + \hat{Q}_3 (\theta_n + n^{-\gamma} h) \right) \xrightarrow{\gamma} \pi (\theta_n)^r W Z_{0,1/2} (h) \) if \( \gamma = 1/3 \) and

\[ n^{2 \gamma} \left( \hat{Q}_2 (\theta_n + n^{-\gamma} h) + \hat{Q}_3 (\theta_n + n^{-\gamma} h) \right) \xrightarrow{\gamma} \pi (\theta_n)^r W Z_{0,1} (h) + h' G' W U_0 \] if \( \gamma = 1/2 \). Additionally,

\[ n^{2 \gamma} Q_1 (\theta_n + n^{-\gamma} h) = n^\gamma h' \frac{\partial Q_1 (\theta_n)}{\partial \theta} + \frac{1}{2} h' \frac{\partial^2 Q_1 (\theta_n)}{\partial \theta^2} \hat{h} + o (1) \]

\[ = n^\gamma h' G' W \pi (\theta_n) + \frac{1}{2} h' \hat{H} h + o (1) \]

\[ = \frac{1}{2} h' \hat{H} h + o (1) \]

since \( Q_1 (\theta) \) achieves the minimal value of 0 at \( \theta_n \) when the constraints are not necessary for identification of \( \theta_n \). Therefore,

\[ n^{2 \gamma} \left( \hat{Q}_n (\theta_n + n^{-\gamma} h) - \hat{Q}_n (\theta_n) \right) \]

\[ = n^{2 \gamma} \left( \hat{Q}_2 (\theta_n + n^{-\gamma} h) + \hat{Q}_3 (\theta_n + n^{-\gamma} h) \right) + \frac{1}{2} h' \hat{H} h + o_p (1) \]

\[ \xrightarrow{\gamma} A_0 (h) \equiv \begin{cases} \pi (\theta_n)^r W Z_{0,1/2} (h) + \frac{1}{2} h' \hat{H} h & \text{if } \gamma = 1/3 \\ \pi (\theta_n)^r W Z_{0,1} (h) + h' G' W U_0 + \frac{1}{2} h' \hat{H} h & \text{if } \gamma = 1/2 \end{cases} \]

as a process indexed by \( h \) in the space of locally bounded functions \( B_{\text{loc}} (\mathbb{R}^d) \) equipped...
with the topology of uniform convergence on compacta.

We already showed in Theorems 2 and 6 of Hong and Li (2023) that under Assumptions 3, 5 - 8, \( \hat{A}_n(h) \xrightarrow{\mathbb{P}} A_0(h) \), where

\[
\hat{A}_n(h) = n^{2\gamma} \hat{\pi}_n \left( \hat{\theta}_n \right) W \left( P_n^* - P_n \right) \left( \pi \left( \cdot, \hat{\theta}_n + \frac{h}{n^{\gamma}} \right) - \pi \left( \cdot, \theta_n \right) \right) + \frac{1}{2} h' \left( \hat{G}' W \hat{G} + \sum_{j=1}^m \sum_{k=1}^m W_{jk} \hat{\pi}_{nk} \left( \hat{\theta}_n \right) \hat{H}_j \right) h + n^\gamma h' \hat{G}' W \left( P_n^* - P_n \right) \pi \left( \cdot, \hat{\theta}_n \right).
\]

Then the continuous mapping results in Lemma 10.11 of Kosorok (2007) imply \(- \inf_{h \in \mathbb{R}^d} \hat{A}_n(h) \xrightarrow{\mathbb{P}} - \inf_{h \in \mathbb{R}^d} A_0(h) \). Additionally,

\[
n^{2\gamma} \left( \hat{Q}_n \left( \theta_n \right) - \inf_{h \in \mathcal{B}_{\epsilon_n}} \hat{Q}_n \left( \theta_n + \frac{h}{n^{\gamma}} \right) \right) \xrightarrow{\mathbb{P}} - \inf_{h \in \left\{ h \in \mathbb{R}^d : |h| \leq \epsilon \right\}} A_0(h).
\]

Since \( J(\cdot, P^{(n)}) \) is equicontinuous at \( J_n^{-1} \left( 1 - \alpha, P^{(n)} \right) \), we have for \( x_n = J_n^{-1} \left( 1 - \alpha - \epsilon, P^{(n)} \right) \),

\[
J_n \left( x_n, P^{(n)} \right) - J \left( x_n, P^{(n)} \right) = o(1)
\]

for any \( P^{(n)} \) and \( \epsilon \) small enough. Since \( J^* (\cdot, P^{(n)}) \) is also equicontinuous at \( J_n^{-1} \left( 1 - \alpha, P^{(n)} \right) \), we have for any \( P^{(n)} \) and \( \epsilon \) small enough,

\[
J^*_{\alpha_n} \left( x_n, P^{(n)} \right) - J^* \left( x_n, P^{(n)} \right) = o_{P^{(n)}}(1).
\]

Note that \(- \inf_{h \in \left\{ h \in \mathbb{R}^d : |h| \leq \epsilon \right\}} A_0(h) \leq - \inf_{h \in \mathbb{R}^d} A_0(h) \) for any realizations of the random variables, which means \( J^* \left( x_n, P^{(n)} \right) < J \left( x_n, P^{(n)} \right) \) for all \( n \) large enough. Then, for all \( \epsilon > 0 \) and \( n \) large enough, there exists \( \delta > 0 \) such that \( P^{(n)} \left( J^*_{\alpha_n} \left( x_n, P^{(n)} \right) - J_n \left( x_n, P^{(n)} \right) > \epsilon \right) \leq \delta \). If \( J^*_{\alpha_n} \left( x_n, P^{(n)} \right) - J_n \left( x_n, P^{(n)} \right) \leq \epsilon \), then \( J_n^{-1} \left( 1 - \alpha - \epsilon, P^{(n)} \right) \leq J_{\alpha_n}^{-1} \left( 1 - \alpha, P^{(n)} \right) \).

Then, using arguments similar to those in Lemma A.1 (vi) of Romano and Shaikh (2012), for all \( \epsilon > 0 \) and \( n \) large enough,

\[
P^{(n)} \left( n^{2\gamma} \left( \hat{Q}_n \left( \theta_n \right) - \inf_{h \in \mathcal{B}_{\epsilon_n}} \hat{Q}_n \left( \theta_n + \frac{h}{n^{\gamma}} \right) \right) \right) \leq J^*_{\alpha_n^{-1}} \left( 1 - \alpha, P^{(n)} \right) \]

\[
\geq P^{(n)} \left( n^{2\gamma} \left( \hat{Q}_n \left( \theta_n \right) - \inf_{h \in \mathcal{B}_{\epsilon_n}} \hat{Q}_n \left( \theta_n + \frac{h}{n^{\gamma}} \right) \right) \right) \leq J^*_{\alpha_n^{-1}} \left( 1 - \alpha, P^{(n)} \right) \cap J^*_{\alpha_n} \left( x_n, P^{(n)} \right) - J_n \left( x_n, P^{(n)} \right) \leq \epsilon \)
\]
\[
\begin{align*}
&\geq p^{(n)} \left( n^{2\gamma} \left( \hat{Q}_n(\theta_n) - \inf_{h \in B_{\delta n}} \hat{Q}_n \left( \theta_n + \frac{h}{n^{\gamma}} \right) \right) \right) \leq J^{-1} \left( 1 - \alpha - \epsilon, p^{(n)} \right) \cap J_{\alpha, n}^{*} \left( \tau_n, p^{(n)} \right) - J_n \left( \tau_n, p^{(n)} \right) \leq \epsilon \\
&\geq p^{(n)} \left( n^{2\gamma} \left( \hat{Q}_n(\theta_n) - \inf_{h \in B_{\delta n}} \hat{Q}_n \left( \theta_n + \frac{h}{n^{\gamma}} \right) \right) \right) \leq J^{-1} \left( 1 - \alpha - \epsilon, p^{(n)} \right) \sum_{k} \left( J_{\alpha, n}^{*} \left( \tau_n, p^{(n)} \right) - J_n \left( \tau_n, p^{(n)} \right) \right) > \epsilon \\
&\geq 1 - \alpha - \epsilon - \delta
\end{align*}
\]

Since \( \epsilon \) and \( \delta \) can be arbitrarily small, \( \liminf_{n \to \infty} p^{(n)} \left( n^{2\gamma} \left( \hat{Q}_n(\theta_n) - \inf_{h \in B_{\delta n}} \hat{Q}_n \left( \theta_n + \frac{h}{n^{\gamma}} \right) \right) \right) \leq \hat{c}^{\ast}_{1 - \alpha} \geq 1 - \alpha. \) For \( \rho = \liminf_{n \to \infty} \inf_{P \in \mathcal{P}} \left( n^{2\gamma} \left( \hat{Q}_n(\theta_n) - \inf_{h \in B_{\delta n}} \hat{Q}_n \left( \theta_n + \frac{h}{n^{\gamma}} \right) \right) \right) \leq \hat{c}^{\ast}_{1 - \alpha} \), we can find a sequence \( \{P^{(n)} \in \mathcal{P}\} \) such that \( \rho = \liminf_{n \to \infty} p^{(n)} \left( n^{2\gamma} \left( \hat{Q}_n(\theta_n) - \inf_{h \in B_{\delta n}} \hat{Q}_n \left( \theta_n + \frac{h}{n^{\gamma}} \right) \right) \right) \leq \hat{c}^{\ast}_{1 - \alpha} \). Find a subsequence \( n_k \) of \( n \) for which \( \theta_n \) converges, with its limit denoted \( \theta \). The same arguments as above applied to such a subsequence imply

\[
\liminf_{n_k \to \infty} p^{(n_k)} \left( n_k^{2\gamma} \left( \hat{Q}_n(\theta_{n_k}) - \inf_{h \in B_{\delta n_k}} \hat{Q}_n \left( \theta_{n_k} + \frac{h}{n_k^{\gamma}} \right) \right) \right) \leq \hat{c}^{\ast}_{1 - \alpha} \geq 1 - \alpha. \]

Since \( \{P^{(n_k)} \cap \theta_{n_k}\} \) is a subsequence of \( \{P^{(n)} \cap \theta_n\} \), \( \rho = \liminf_{n_k \to \infty} p^{(n_k)} \left( n_k^{2\gamma} \left( \hat{Q}_n(\theta_{n_k}) - \inf_{h \in B_{\delta n_k}} \hat{Q}_n \left( \theta_{n_k} + \frac{h}{n_k^{\gamma}} \right) \right) \right) \leq \hat{c}^{\ast}_{1 - \alpha} \geq 1 - \alpha.


\[6.1.4 \text{ Proof of Theorem 4}\]

Recall \( n^{2\gamma} Q_1(\theta_0 + n^{-\gamma} h) = n^{\gamma} h'G'W \pi(\theta_0) + \frac{1}{2} h'\tilde{H} h + o(1) \). Additionally, Chernoff regularity implies that \( + \infty 1(h \notin n^{\gamma} (C - \theta_0)) \to 1 + \infty 1(h \notin C(\theta_0)) \), where \( T_C(\theta_0) \equiv \limsup_{\tau \downarrow 0}^{C - \theta_0} \). When \( \theta_0 \in \partial C \), \( h'G'W \pi(\theta_0) \geq 0 \) for all \( h \in T_C(\theta_0) \). Otherwise, there would exist some descent direction \( h \in T_C(\theta_0) \) that reduces the value of the objective function and \( \theta_0 \) would not be the constrained minimum of the population objective. Then, for any \( c \in \mathbb{R} \),

\[
\limsup_{n \to \infty} \mathbb{P} \left( n^{2\gamma} \left( \hat{Q}_n(\theta_0) - \inf_{h \in C_{\theta_0}} \hat{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) > c \right) = \limsup_{n \to \infty} \mathbb{P} \left( -n^{2\gamma} \inf_{h \in C_{\theta_0}} \left\{ \hat{Q}_2 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \hat{Q}_3 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + Q_1 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right\} > c \right)
\]

\[
\leq \limsup_{n \to \infty} \mathbb{P} \left( -\inf_{h \in C_{\theta_0}} \left\{ n^{2\gamma} \left( \hat{Q}_2 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \hat{Q}_3 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) + n^{\gamma} h'G'W \pi(\theta_0) + \frac{1}{2} h'\tilde{H} h \right\} > c \right)
\]
\[
\leq \limsup_{n \to \infty} \sup_{P \in \mathcal{P}} \left( - \inf_{h \in \{h \in T_C(\theta_0); |h| \leq \kappa\}} \left\{ n^{2\gamma} \left( \hat{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) - \inf_{h \in C_n^{\alpha}} \hat{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) + \frac{1}{2} h' \hat{H} h \right\} > c \right) 
\]

Because the test statistic \( n^{2\gamma} \left( \hat{Q}_n \left( \theta_0 \right) - \inf_{h \in C_n^{\alpha}} \hat{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) \) does not have a well-defined limiting distribution when \( h' n^\gamma G W \pi \left( \theta_0 \right) \neq 0 \), we will instead define \( J_n \left( \cdot, P \right) \) as the CDF of its limiting distribution 

\[
J_n \left( \cdot, P \right) = \inf_{h \in \{h \in T_C(\theta_0); |h| \leq \kappa\}} \left\{ n^{2\gamma} \left( \hat{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \hat{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) + \frac{1}{2} h' \hat{H} h \right\} \quad \text{under } P, 
\]

and \( J \left( \cdot, P \right) \) as the CDF of its limiting distribution 

\[
J \left( \cdot, P \right) = \inf_{h \in \{h \in T_C(\theta_0); |h| \leq \kappa\}} A_0 \left( h \right) \quad \text{under } P. 
\]

The definition of \( J_n \left( \cdot, P \right) \) remains the same as in Theorem 3, denoting the conditional CDF of \( - \inf_{h \in \mathbb{R}^d} \hat{A}_n \left( h \right) \) under \( P \), and \( J \left( \cdot, P \right) \) still denotes the CDF of its limiting distribution 

\[
- \inf_{h \in \mathbb{R}^d} A_0 \left( h \right) \quad \text{under } P. 
\]

Note that \( - \inf_{h \in T_C(\theta_0); |h| \leq \kappa} A_0 \left( h \right) \) is stochastically dominated by \( - \inf_{h \in \mathbb{R}^d} A_0 \left( h \right) \). The result follows from modifying the proof of Theorem 3 to incorporate the new definitions of \( J_n \left( \cdot, P \right) \) and \( J \left( \cdot, P \right) \), and replacing \( n^{2\gamma} \left( \hat{Q}_n \left( \theta_n \right) - \inf_{h \in B_{\mathbb{R}^d}} \hat{Q}_n \left( \theta_n + \frac{h}{n^{\gamma}} \right) \right) \) by \( n^{2\gamma} \left( \hat{Q}_n \left( \theta_n \right) - \inf_{h \in C_n^{\alpha}} \hat{Q}_n \left( \theta_n + \frac{h}{n^{\gamma}} \right) \right) \).

### 6.1.5 Proof of Theorem 5

Consistency of \( \hat{\theta}_1 \) for \( \theta_1 \) follows from Corollary 3.2.3 in van der Vaart and Wellner (1996) or Theorem 5.7 in Van der Vaart (2000) since the equation array in the proof of Theorem 2.6 in Newey and McFadden (1994) in combination with Assumptions 5(iii) and (iv) imply 

\[
\sup_{\theta \in \Theta} \left| \hat{\pi} \left( \theta \right)' W_1 \hat{\pi} \left( \theta \right) - \pi \left( \theta \right)' W_1 \pi \left( \theta \right) \right| = o_P(1). \]

Consistency of \( \hat{\theta}_1^C \) for \( \theta_1^C \) follows from constraining \( \theta \) to lie in \( C \) when applying Corollary 3.2.3 in van der Vaart and Wellner (1996) or Theorem 5.7 in Van der Vaart (2000). The arguments in the proofs of Theorems 2 and 6 of Hong and Li (2023) imply \( \hat{\theta}_1^* \) is consistent for \( \theta_1 \). We can then construct weighting matrices that satisfy \( W_n \left( \hat{\theta}_1 \right) - W \left( \theta_1 \right) = o_P(1) \) and \( W_n^* \left( \hat{\theta}_1^* \right) = W_n \left( \hat{\theta}_1 \right) = o_P^*(1) \). Additionally, the arguments in the proof of Theorem 6 imply that \( \hat{\theta}_1^{C^*} \) is consistent for \( \theta_1^C \), so that we can construct weighting matrices that satisfy 

\[
W_n \left( \hat{\theta}_1^C \right) - W \left( \theta_1^C \right) = o_P(1) \text{ and } W_n^* \left( \hat{\theta}_1^{C^*} \right) - W_n \left( \hat{\theta}_1^C \right) = o_P^*(1). 
\]
We showed in Theorems 4 and 8 of Hong and Li (2023) that \( \hat{B}_n(h) \overset{p}{\longrightarrow} B_0(h) \), where

\[
\hat{B}_n(h) = n^{2\gamma} \hat{\pi}_n \left( \hat{\theta}_n \right) W_n \left( P_n - P_n \right) \left( \pi \left( \cdot, \hat{\theta}_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \hat{\theta}_n \right) \right) \\
+ \frac{1}{2} \hat{h}' \left( G' W_n \hat{G} + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{n,jk} \hat{\pi}_{nk} \left( \hat{\theta}_n \right) \hat{H}_j \right) h \\
+ n^{\gamma} h' \hat{G}' W_n \left( P_n - P_n \right) \pi \left( \cdot, \hat{\theta}_n \right) \\
+ n^{\gamma} h' \hat{G}' \left( W_n^* - W_n \right) \hat{\pi}_n \left( \hat{\theta}_n \right).
\]

If \( W_n - W = O_P(n^{-\gamma}) \) and \( W_n^* - W_n = O_P(n^{-\gamma}) \), then \( B_0(h) = \pi (\theta_0)' W Z_{0,1}(h) + h' G' W_0 U_0 + \frac{1}{2} h' \tilde{H} h + h' G' W_0 \) for \( \rho = 1 \) and \( B_0(h) = \pi (\theta_0)' W Z_{0,1/2}(h) + \frac{1}{2} h' \tilde{H} h + h' G' W_0 \) when \( \rho = 1/2 \). If \( W_n - W = o_P(n^{-\gamma}) \) and \( W_n^* - W_n = o_P(n^{-\gamma}) \), then \( B_0(h) = \pi (\theta_0)' W Z_{0,1}(h) + h' G' W_0 U_0 + \frac{1}{2} h' \tilde{H} h \) for \( \rho = 1 \) and \( B_0(h) = \pi (\theta_0)' W Z_{0,1/2}(h) + \frac{1}{2} h' \tilde{H} h \) when \( \rho = 1/2 \).

We showed in Theorem 3 of Hong and Li (2023) that

\[
n^{2\gamma} \left( \hat{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) - \hat{Q}_n (\theta_0) \right) \\
= n^{2\gamma} \left( Q_1 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \hat{Q}_2 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \hat{Q}_3 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \\
+ \hat{Q}_4 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \hat{Q}_5 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \hat{Q}_6 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) \\
= n^{2\gamma} \left( Q_1 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \hat{Q}_2 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \hat{Q}_3 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \hat{Q}_4 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) + o_P(1) \\
\rightsquigarrow B_0(h),
\]

\[
Q_1(\theta) = \frac{1}{2} g'(\theta)' W g(\theta) + g(\theta)' W \pi(\theta_0), \quad \hat{Q}_3 (\theta) = \pi (\theta_0)' W \hat{\eta}_n (\theta) \\
\hat{Q}_2 (\theta) = \frac{1}{2} \hat{\eta}_n (\theta)' W \hat{\eta}_n (\theta) + g(\theta)' W \left( \hat{\pi}_n (\theta_0) - \pi (\theta_0) \right) + g(\theta)' W \hat{\eta}_n (\theta) + \left( \hat{\pi}_n (\theta_0) - \pi (\theta_0) \right)' W \hat{\eta}_n (\theta)
\]

43
\[ \hat{Q}_1(\theta) = \frac{1}{2} g(\theta)' (W_n - W) g(\theta) + g(\theta)' (W_n - W) \pi (\theta_0) \]
\[ \hat{Q}_5(\theta) = g(\theta)' (W_n - W) (\hat{\pi}_n (\theta_0) - \pi (\theta_0)) \]
\[ + g(\theta)' (W_n - W) \hat{\pi}_n (\theta) + (\hat{\pi}_n (\theta_0) - \pi (\theta_0))' (W_n - W) \hat{\pi}_n (\theta) \]
\[ \hat{Q}_6(\theta) = \pi (\theta_0)' (W_n - W) \hat{\pi}_n (\theta) + \frac{1}{2} \pi (\theta_0)' (W_n - W) \hat{\pi}_n (\theta). \]

Recall \( n^{2\gamma} q_1 (\theta_0 + n^{-\gamma} h) = n^{\gamma} h' G' W \pi (\theta_0) + \frac{1}{2} h' \check{H} h + o(1) \). Additionally, Chernoff regularity implies that \( + \infty \) \( (h \notin n^{\gamma} (C - \theta_0)) \leadsto + \infty \) \( (h \notin T_C (\theta_0)) \), where \( T_C (\theta_0) \equiv \limsup_{\gamma \to 0} \frac{C - \theta_0}{\gamma} \). When \( \theta_0 \in \partial C \), \( h' G' W \pi (\theta_0) \geq 0 \) for all \( h \in T_C (\theta_0) \). Otherwise, there would exist some descent direction \( h \in T_C (\theta_0) \) that reduces the value of the objective function and \( \theta_0 \) would not be the constrained minimum of the population objective. Also, when \( \theta_0 \in \text{int} C \), \( G' W \pi (\theta_0) = 0 \). Define

\[ \bar{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) = \hat{Q}_2 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \hat{Q}_3 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \hat{Q}_4 \left( \theta_0 + \frac{h}{n^{\gamma}} \right). \] (7)

Then, for any \( c \in \mathbb{R} \),

\[ \limsup_{n \to \infty} \sup_{P \in \mathcal{P}} \left( n^{2\gamma} \left( \hat{Q}_n (\theta_0) - \inf_{h \in \partial C_n} \hat{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) > c \right) \]
\[ = \limsup_{n \to \infty} \sup_{P \in \mathcal{P}} \left( - n^{2\gamma} \inf_{h \in \partial C_n} \left\{ \bar{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + Q_1 \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right\} > c \right) \]
\[ \leq \limsup_{n \to \infty} \sup_{P \in \mathcal{P}} \left( - \inf_{h \in \partial C_n} \left\{ n^{2\gamma} \bar{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + n^{\gamma} h' G' W \pi (\theta_0) + \frac{1}{2} h' \check{H} h \right\} > c \right) \]
\[ \leq \limsup_{n \to \infty} \sup_{P \in \mathcal{P}} \left( - \inf_{h \in \partial T_C (\theta_0)} n^{2\gamma} \bar{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) + \frac{1}{2} h' \check{H} h \right) > c \]

Because the test statistic \( n^{2\gamma} \left( \hat{Q}_n (\theta_0) - \inf_{h \in \partial C_n} \hat{Q}_n \left( \theta_0 + \frac{h}{n^{\gamma}} \right) \right) \) does not have a well-defined limiting distribution when \( h' n^{\gamma} G' W \pi (\theta_0) \neq 0 \), we will instead define \( J_n (\cdot; P) \) as the
CDF of \( - \inf_{h \in \{ h \in T_C(\theta_0): |h| \leq \kappa \}} \{ n^{2\gamma} \hat{Q}_n (\theta_0 + \frac{h}{n}) + \frac{1}{2} h' \tilde{H} h \} \) under \( P \), and \( J (\cdot, P) \) as the CDF of its limiting distribution \( - \inf_{\{ h \in T_C(\theta_0): |h| \leq \kappa \}} B_0 (h) \) under \( P \). Let \( J^*_{\alpha_n} (\cdot, P) \) denote the conditional CDF of \( - \inf_{h \in \mathbb{R}^d} \hat{B}_n (h) \) under \( P \) and let \( J^* (\cdot, P) \) denote the CDF of its limiting distribution \( - \inf_{h \in \mathbb{R}^d} B_0 (h) \) under \( P \). Note that \( - \inf_{\{ h \in T_C(\theta_0): |h| \leq \kappa \}} B_0 (h) \) is stochastically dominated by \( - \inf_{h \in \mathbb{R}^d} B_0 (h) \). The result follows from modifying the proof of Theorem 3 to incorporate the new definitions of \( J_n (\cdot, P) \), \( J (\cdot, P) \), \( J^*_{\alpha_n} (\cdot, P) \), \( J^* (\cdot, P) \), and replacing \( n^{2\gamma} \left( \hat{Q}_n (\theta_n) - \inf_{h \in B_{\delta_n}} \hat{Q}_n (\theta_n + \frac{h}{n}) \right) \) by \( n^{2\gamma} \left( \hat{Q}_n (\theta_n) - \inf_{h \in C_{\delta_n}^{\alpha_n}} \hat{Q}_n (\theta_n + \frac{h}{n}) \right) \).

6.1.6 Proof of Theorem 6

Recall \( \hat{Q}_n (\theta) - \hat{Q}_n (\theta_0) = Q_1 (\theta) + \hat{Q}_2 (\theta) + \hat{Q}_3 (\theta) \), where

\[
Q_1 (\theta) = \frac{1}{2} g (\theta)' W g (\theta) + g (\theta)' W \pi (\theta_0), \quad \hat{Q}_3 (\theta) = \pi (\theta_0)' W \hat{\eta}_n (\theta)
\]

\[
\hat{Q}_2 (\theta) = \frac{1}{2} \hat{\eta}_n (\theta)' W \hat{\eta}_n (\theta) + g (\theta)' W (\hat{\pi}_n (\theta_0) - \pi (\theta_0)) + g (\theta)' W \hat{\eta}_n (\theta) + (\hat{\pi}_n (\theta_0) - \pi (\theta_0))' W \hat{\eta}_n (\theta).
\]

Consistency of \( \hat{\theta}_n \) for \( \theta_0 \) follows from constraining \( \theta \) to lie in \( C \) when applying Corollary 3.2.3 in van der Vaart and Wellner (1996) or Theorem 5.7 in Van der Vaart (2000) since the equation array in the proof of Theorem 2.6 in Newey and McFadden (1994) in combination with Assumptions 5(iii) and (iv) imply \( \sup_{\theta \in C} |Q_n (\theta) - Q (\theta)| = o_P (1) \).

Apply Kim and Pollard (1990) Lemma 4.1 to \( \hat{\eta}_n (\theta) \), and in turn \( \hat{Q}_3 (\theta) \): \( \forall \epsilon > 0 \), \( \exists M_{n,3} = O_P (1) \) such that

\[
|\hat{Q}_3 (\theta)| \leq \epsilon \| \theta - \theta_0 \|^2 + n^{-2\gamma} M_{n,3}^2.
\]

The 1st, 3rd, and 4th terms in \( \hat{Q}_2 (\theta) \) are all of the form \( o_P (1) \hat{\eta}_n (\theta) \), hence are also
bounded by \( \epsilon \| \theta - \theta_0 \|^2 + n^{-2\gamma} M_{n,2}^2 \). For the 2nd term in \( \hat{Q}_2 (\theta) \), for \( n \) large enough, \( \forall \epsilon > 0, \exists M_{n,22} = O_P (1) \) such that

\[
|g (\theta)' W (\hat{\theta}_n (\theta_0) - \pi (\theta_0))| = O_P \left( \frac{\| \theta - \theta_0 \|}{\sqrt{n}} \right) \leq \epsilon \| \theta - \theta_0 \|^2 + n^{-2\gamma} M_{n,22}^2 .
\]

Therefore, \( \forall \epsilon > 0, \exists M_n = O_P (1) \) such that \( |\hat{Q}_2 (\theta) + \hat{Q}_3 (\theta)| \leq \epsilon \| \theta - \theta_0 \|^2 + n^{-2\gamma} M_n^2 \).

Taylor expanding \( Q_1 (\theta) \) around \( \theta_0 \) while constraining \( \theta \) to lie in \( C \), \( Q_1 (\theta) = Q_1 (\theta_0) + (\theta - \theta_0)' \frac{\partial Q_1(\theta_0)}{\partial \theta} + \frac{1}{2} (\theta - \theta_0)' \frac{\partial^2 Q_1(\theta_0) }{\partial \theta^2} (\theta - \theta_0) + o (\| \theta - \theta_0 \|^2) = (\theta - \theta_0)' G' W \pi (\theta_0) + \frac{1}{2} (\theta - \theta_0)' (\bar{H} + o (1)) (\theta - \theta_0) \) since \( \frac{\partial Q_1(\theta_0)}{\partial \theta} = G' W \pi (\theta_0) \) and \( \frac{\partial^2 Q_1(\theta_0) }{\partial \theta^2} = \bar{H} \). Because \( \bar{H} \) is positive definite and \( (\theta - \theta_0)' G' W \pi (\theta_0) \geq 0 \) for all \( \theta \in C \), there exists \( K > 0 \) and a small enough neighborhood of \( \theta_0 \) such that \( Q_1 (\theta) \geq K \| \theta - \theta_0 \|^2 \). By consistency of \( \hat{\theta}_n \) for \( \theta_0 \), with probability approaching 1, \( Q_1 (\hat{\theta}_n) \geq K \| \hat{\theta}_n - \theta_0 \|^2 \). Then,

\[
Q_1 (\hat{\theta}_n) + \hat{Q}_2 (\hat{\theta}_n) + \hat{Q}_3 (\hat{\theta}_n) = Q_n (\hat{\theta}_n) - \hat{Q}_n (\theta_0) \leq Q_n (\hat{\theta}_n) - \inf_{\theta \in C} Q_n (\theta) \leq o_P (n^{-2\gamma}) .
\]

Choose \( \epsilon \) so that \( K - \epsilon > 0 \). Then,

\[
o_P (n^{-2\gamma}) \geq Q_1 (\hat{\theta}_n) - \epsilon \| \hat{\theta}_n - \theta_0 \|^2 - n^{-2\gamma} M_n^2 \geq (K - \epsilon) \| \hat{\theta}_n - \theta_0 \|^2 - n^{-2\gamma} M_n^2 \implies \| \hat{\theta}_n - \theta_0 \|^2 \leq (K - \epsilon)^{-1} n^{-2\gamma} M_n^2 + o_P (n^{-2\gamma}) = O_P (n^{-2\gamma}) .
\]

It follows that \( n^\gamma (\hat{\theta}_n - \theta_0) = O_P (1) \).

Using the arguments in Theorems 2 and 6 of Hong and Li (2023), \( \hat{H}_n (h) \xrightarrow{P} \hat{H}_0 (h), \)
where

\[
\hat{\mathbb{H}}_n (h) = \hat{\pi}_n (\hat{\theta}_n) ' W n^{\gamma_0} \sqrt{n} (P^*_n - P_n) \left( \pi (\cdot, \hat{\theta}_n + \frac{h}{n^{\gamma}}) - \pi (\cdot, \hat{\theta}_n) \right) \\
+ \frac{\sqrt{n} n^{\gamma_0}}{2n^{2\gamma}} h' \left( \hat{G}' W \hat{G} + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{jk} \hat{\pi}_{nk} (\hat{\theta}_n) \hat{H}_j \right) h \\
+ \frac{n^{\gamma_0}}{n^{\gamma}} h' \hat{G}' W \sqrt{n} (P^*_n - P_n) \pi (\cdot, \hat{\theta}_n). 
\]

For \( \hat{H} = G' W G + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{jk} \pi_k (\theta_0) H_j \), \( \mathbb{H}_0 (h) = \pi (\theta_0)' W Z_{0,1} (h) + h' G' W U_0 + \frac{1}{2} h' \hat{H} h \) for \( \rho = 1 \) and \( \mathbb{H}_0 (h) = \pi (\theta_0)' W Z_{0,1/2} (h) + \frac{1}{2} h' \hat{H} h \) when \( \rho = 1/2 \).

Chernoff regularity implies that for any sequence \( a(n) \to \infty, +\infty \), \( h \notin a(n) (C - \theta_0) \to \infty \), \( +\infty \). Since \( n^{\gamma}/\eta_n \to \infty \) and \( n^{\gamma} (\hat{\theta}_n - \theta_0) = O_P (1) \),

\[
\omega_n^* (h) = +\infty \left( h \notin \frac{n^{\gamma} (C - \theta_0)}{\eta_n} - \frac{n^{\gamma} (\hat{\theta}_n - \theta_0)}{\eta_n} \right) \\
= +\infty \left( h \notin \frac{n^{\gamma} (C - \theta_0)}{\eta_n} - o_p (1) \right) \\
\to \omega (h) \equiv +\infty \left( h \notin T_C (\theta_0) \right).
\]

where \( T_C (\theta_0) = \limsup_{\tau \to 0} \frac{C - \theta_0}{\tau} \). By a bootstrap in probability version of Theorem 4 in Knight (1999),

\[
\hat{\mathbb{H}}_n (h) + \omega_n^* (h) \overset{e-d}{\to} \mathbb{H}_0 (h) + \omega (h).
\]

where \( \overset{e-d}{\to} \) denotes epi-convergence of the conditional law of \( \hat{G}_n^* \) to \( \mathbb{G}_0 \), which can be equivalently stated as \( \sup_{f \in BL_1} |E_{\mathbb{W}} f (\hat{G}_n^*) - E f (\mathbb{G}_0) | \overset{p}{\to} 0 \) and \( E_{\mathbb{W}} f (\hat{G}_n^*)^* - E_{\mathbb{W}} f (\hat{G}_n^*) \overset{p}{\to} 0 \) for all \( f \in BL_1 \), where \( BL_1 \) is the class of Lipschitz norm 1 functions with respect to the metric of epi-convergence defined as \( d (\hat{G}_n^*, \mathbb{G}_0) = \int_0^\infty \max \left\{ |d_{e-pi} \hat{G}_n^* (v) - d_{e-pi} \mathbb{G}_0 (v) | : |v| \leq \rho \right\} \exp (-\rho) d\rho \),

47
where \( d_C(v) = \inf \{|v - u| : u \in C\} \) for a non-empty closed subset of \( \mathbb{R}^{d+1} \), and epi \( G(h) = \{(h, \alpha) : G(h) \leqslant \alpha\} \) is the epigraph of \( G : \mathbb{R}^d \rightarrow \mathbb{R} \). Lemma 2.6 in Kim and Pollard (1990) implies that the Gaussian processes \( Z_{0,1/2}(h) \) and \( Z_{0,1}(h) \) have a unique minimum. In combination with the fact that \( \frac{1}{2} h \tilde{H} h \) is a convex function of \( h \), there is a unique \( h \) that minimizes \( \|H_0(h) + \omega(h)\| \). By a modification of the bootstrap argmin continuous mapping lemma 14.2 in Hong and Li (2020) that replaces weak convergence with epi-convergence,\n\n\[
\alpha_n \left( \hat{\theta}_n - \theta_0 \right) = \arg\min_{h \in \mathbb{R}^d} \left\{ \frac{1}{n} \|H_n(h) + \omega_n(h)\| \right. \\
\left. \tilde{\omega} \arg\min_{h \in \mathbb{R}^d} \{ \|H_0(h) + \omega(h)\| \right\},
\]
which coincides with the limiting distribution of \( \alpha_n \left( \hat{\theta}_n - \theta_0 \right) \).

\[\blacksquare\]

### 6.2 Verification of Assumption 3

For the Andrews (2000) example, \( \pi(\cdot, \theta) = (y_i - \theta)^2 \), \( m_n(\cdot, \theta, h) = \sqrt{n} \left( (y_i - \theta - \frac{h}{\sqrt{n}})^2 - (y_i - \theta)^2 \right) \), and

\[
m_n(\cdot, \theta, h) - m_n(\cdot, \theta_0, h) = \sqrt{n} \left( \pi(\cdot, \theta + \frac{h}{\sqrt{n}}) - \pi(\cdot, \theta) - \left( \pi(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) - \pi(\cdot, \theta_0) \right) \right)
\]

\[
= \sqrt{n} \left( (y_i - \theta - \frac{h}{\sqrt{n}})^2 - (y_i - \theta)^2 - \left( (y_i - \theta_0 - \frac{h}{\sqrt{n}})^2 - (y_i - \theta_0)^2 \right) \right)
\]

\[
= \sqrt{n} \left( -2 (y_i - \theta_0 - \frac{h}{\sqrt{n}}) + 2 (y_i - \theta_0) \right) (\theta - \theta_0)
\]

\[
= 2h (\theta - \theta_0)
\]

\[
\sup_{h \in H, |\theta - \theta_0| \leqslant \epsilon_n} \|m_n(\cdot, \theta, h) - m_n(\cdot, \theta_0, h)\| \leqslant K \text{ will be satisfied if we take } K = 2\sup_{h \in H} |h|.
\]

Another way Assumption 3 can be satisfied is if \( E \left[ \sup_{h \in H, |\theta - \theta_0| \leqslant \epsilon_n} \|m_n(\cdot, \theta, h) - m_n(\cdot, \theta_0, h)\|^{2+\delta} \right] < \)
\( \infty \) for all \( n \) and any \( \delta > 0 \). For Example 2, there exists some constant \( C > 0 \) such that

\[
E \left[ \sup_{h \in \mathcal{H}, \| \theta - \theta_0 \| \leq \epsilon_n} \left| \frac{m_n (\cdot, \theta, h) - m_n (\cdot, \theta_0, h)}{1 + n \gamma \| \theta - \theta_0 \|} \right|^{2+\delta} \right] = E \left[ \sup_{h \in \mathcal{H}, \| \theta - \theta_0 \| \leq \epsilon_n} \left| \frac{n^{1/3} \left| \pi (\cdot, \theta + hn^{-1/3}) - \pi (\cdot, \theta) - (\pi (\cdot, \theta_0 + hn^{-1/3}) - \pi (\cdot, \theta_0)) \right|^{2+\delta}}{1 + n^{1/3} \| \theta - \theta_0 \|^{2+\delta}} \right] \leq E \left[ \sup_{h \in \mathcal{H}, \| \theta - \theta_0 \| \leq \epsilon_n} \left| \frac{n^{1/3} \left| 1 (\theta + \frac{h}{n^{1/3}} - 1 \leq y_i \leq \theta + \frac{h}{n^{1/3}} + 1) - 1 (\theta - 1 \leq y_i \leq \theta + 1) \right|}{1 + n^{1/3} \| \theta - \theta_0 \|^{2+\delta}} \right] \right.
\]

\[
+E \left[ \sup_{h \in \mathcal{H}, \| \theta - \theta_0 \| \leq \epsilon_n} \left| \frac{n^{1/3} \left| 1 (\theta_0 + \frac{h}{n^{1/3}} - 1 \leq y_i \leq \theta_0 + \frac{h}{n^{1/3}} + 1) - 1 (\theta_0 - 1 \leq y_i \leq \theta_0 + 1) \right|}{1 + n^{1/3} \| \theta - \theta_0 \|^{2+\delta}} \right] \right. 
\]

\[
\leq E \left[ \sup_{h \in \mathcal{H}, \| \theta - \theta_0 \| \leq \epsilon_n} \left| \frac{n^{1/3} \left| 1 (\theta + \frac{h}{n^{1/3}} - 1 \leq y_i \leq \theta + \frac{h}{n^{1/3}} + 1) - 1 (\theta - 1 \leq y_i \leq \theta + 1) \right|}{1 + n^{1/3} \| \theta - \theta_0 \|^{2+\delta}} \right] \right. 
\]

\[
\leq \sup_{h \in \mathcal{H}, \| \theta - \theta_0 \| \leq \epsilon_n} \left( \frac{Ch}{1 + n \gamma \| \theta - \theta_0 \|} \right)^{2+\delta} < \infty
\]

For Example 3, \( m_n (\cdot, \theta, h) = n^{\gamma} \left[ 1 (y_i \leq \theta + hn^{-\gamma}) - \tau, y_i = \theta - hn^{-\gamma} \right] - n^{\gamma} \left[ 1 (y_i \leq \theta - \tau, y_i = \theta) \right] = n^{\gamma} \left[ 1 (y_i \leq \theta + hn^{-\gamma}) - 1 (y_i \leq \theta), -hn^{-\gamma} \right] \implies \left[ m_{n1} (\cdot, \theta, h), m_{n2} (\cdot, \theta, h) \right] \), where \( \gamma = 1/2 \) if \( \tau = 0.5 \) and \( \gamma = 1/3 \) if \( \tau \neq 0.5 \). We can verify \( E \left[ \sup_{h \in \mathcal{H}, \| \theta - \theta_0 \| \leq \epsilon_n} \left| \frac{m_n (\cdot, \theta, h) - m_n (\cdot, \theta_0, h)}{1 + n \gamma \| \theta - \theta_0 \|} \right|^{2+\delta} \right] < \infty \) for all \( n \) and any \( \delta > 0 \) because there exists some constant \( C > 0 \) such that

\[
E \left[ \sup_{h \in \mathcal{H}, \| \theta - \theta_0 \| \leq \epsilon_n} \left| \frac{m_{n1} (\cdot, \theta, h) - m_{n1} (\cdot, \theta_0, h)}{1 + n \gamma \| \theta - \theta_0 \|} \right|^{2+\delta} \right] \leq E \left[ \sup_{h \in \mathcal{H}, \| \theta - \theta_0 \| \leq \epsilon_n} \left| n^{\gamma} \left[ 1 (y_i \leq \theta + hn^{-\gamma}) - 1 (y_i \leq \theta) \right] \right| \right. 
\]

\[
+E \left[ \sup_{h \in \mathcal{H}, \| \theta - \theta_0 \| \leq \epsilon_n} \left| n^{\gamma} \left[ 1 (y_i \leq \theta_0 + hn^{-\gamma}) - 1 (y_i \leq \theta_0) \right] \right| \right. 
\]

\[
\leq \sup_{h \in \mathcal{H}, \| \theta - \theta_0 \| \leq \epsilon_n} \left( \frac{Ch}{1 + n \gamma \| \theta - \theta_0 \|} \right)^{2+\delta} < \infty
\]

Additionally, \( E \left[ \sup_{h \in \mathcal{H}, \| \theta - \theta_0 \| \leq \epsilon_n} \left| \frac{m_{n2} (\cdot, \theta, h) - m_{n2} (\cdot, \theta_0, h)}{1 + n \gamma \| \theta - \theta_0 \|} \right|^{2+\delta} \right] = 0 \).
References


Andrews, Donald WK, “Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space,” *Econometrica*, 2000, 68 (2), 399–405. 2, 5, 7, 48


Fan, Yanqin and Xuetao Shi, “Wald, QLR, and score tests when parameters are subject to linear inequality constraints,” *Journal of Econometrics*, 2023, 235 (2), 2005–2026. 3, 4


Horowitz, Joel L and Sokbae Lee, “Non-asymptotic inference in a class of optimization problems,” 2019. 3, 4

Hsieh, Yu-Wei, Xiaoxia Shi, and Matthew Shum, “Inference on estimators defined by mathematical programming,” *Journal of Econometrics*, 2022, 226 (2), 248–268. 3, 4


Ketz, Philipp, “Subvector inference when the true parameter vector may be near or at the boundary,” *Journal of Econometrics*, 2018, 207 (2), 285–306. 3, 4

_ and Adam McCloskey, “Short and simple confidence intervals when the directions of some effects are known,” *Review of Economics and Statistics*, 2023, pp. 1–44. 3


Knight, Keith, “Epi-convergence in distribution and stochastic equi-semicontinuity,” *Unpublished manuscript*, 1999, 37. 47


