Inference for Constrained Extremum Estimators

Jessie Li*

June 23, 2024

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.

*Department of Economics, University of California Santa Cruz. jeqli@ucsc.edu

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 (Andrews and Guggenberger (2010)). 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 inspiration 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, nonconvex sample objective function as well as constrained GMM estimators with either correctly specified or globally misspecified nonsmooth sample moments. By globally misspecified, 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 inference 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 (1988), Shapiro (1989), Shapiro (1990), Geyer (1994), Knight (2001), and Knight (2006). While several of these papers derive the non-standard asymptotic distributions of various 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 matter for identification. 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 extremum 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 population 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 conduct uniformly valid inference for constrained M-estimators, while Section 3 demonstrates how to conduct uniformly valid inference for constrained GMM estimators. Section 4 contains 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 $\theta \geqslant 0$ is

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

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

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 $\hat{\theta}_n = \arg\max_{\theta \geqslant 0} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(\theta - 1 \leqslant y_i \leqslant \theta + 1)$ as the non-negative center of an interval of length 2 that contains the largest possible fraction of observations. It is well-known that $\hat{\theta}_n \xrightarrow{p} \theta_0 = \max(\beta_0, 0)$.

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

$$\hat{\theta}_n = \arg\min_{\theta \geqslant 0} \frac{1}{2} \hat{\pi}_n(\theta)' \hat{\pi}_n(\theta)$$

If $\tau = 0.5$, the moments are correctly specified and $\hat{\theta}_n$ converges to $\theta_0 = \arg\min_{\theta \geqslant 0} \frac{1}{2}\pi(\theta)' \pi(\theta)$ at the $\sqrt{n}$-rate. If additionally, $\beta_0 \geqslant 0$, meaning the constraint is correctly specified, $\theta_0 = \beta_0$. However, if $\tau \neq 0.5$, and τ is not drifting towards 0.5, then the moments are globally misspecified and $\hat{\theta}_n$ is cubic-root consistent for θ_0 which is different from β_0 even if $\beta_0 \geqslant 0$. We will study this example in the Monte Carlo simulations.

1.2 Notation

Consider a random sample $\mathcal{X}_n = (X_1, X_2, ..., X_n)$ of independent draws from a probability measure P on a sample space $\mathcal{X}$. Define the empirical measure $P_n \equiv \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, where δ_x is the measure that assigns mass 1 at x and zero everywhere else. Denote the bootstrap empirical measure by $P_n^* = \frac{1}{n}\sum_{i=1}^n W_{ni}\delta_{X_i}$, which can refer to the multinomial, wild, or other exchangeable bootstraps. An exchangeable bootstrap requires that $W_n \equiv (W_{n1}, \ldots, W_{nn})$ 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, $P_n^* = \frac{1}{n}\sum_{i=1}^n \left(\xi_i/\bar{\xi}_n\right)\delta_{X_i}$, where ξ_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): $Z_n \rightsquigarrow Z$ in the metric space $(\mathbb{D}, d)$ if and only if $\sup_{f \in BL_1} |E^* f(Z_n) - Ef(Z)| \to 0$ where BL_1 is the space of functions $f : \mathbb{D} \mapsto \mathbb{R}$ 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, $U \geqslant f(Z_n)$, and EU exists. Conditional weak convergence is also defined in the sense of Kosorok (2007): $Z_n \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} Z$ in the metric space $(\mathbb{D}, d)$ if and only if $\sup_{f \in BL_1} |E_{\mathbb{W}} f(Z_n) - Ef(Z)| \xrightarrow{p} 0$ and $E_{\mathbb{W}} f(Z_n)^* - E_{\mathbb{W}} f(Z_n)_* \xrightarrow{p} 0$ for all $f \in BL_1$, where BL_1 is the space of functions $f : \mathbb{D} \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1, $E_{\mathbb{W}}$ denotes expectation with respect to the bootstrap weights $\mathbb{W}$ 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 $\mathbb{W}$). Let $X_n^* = o_P^*(1)$ if $P(|X_n^*| > \epsilon | \mathcal{X}_n) = o_P(1)$ for all $\epsilon > 0$. Also define $M_n^* = O_P^*(1)$ (hence also $O_P(1)$) if $\lim_{m\to\infty} \limsup_{n\to\infty} P(P(|M_n^*| > m|\mathcal{X}_n) > \epsilon) = 0 \; \forall \epsilon > 0$.

2 Constrained M-estimator

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

$$\theta_0 = \arg\min_{\theta \in C} \{\pi(\theta) \equiv P\pi(\cdot, \theta)\}$$

where the constraint set $C \subseteq \Theta$ is a subset of the parameter space $\Theta \subset \mathbb{R}^d$ for fixed d. We assume C is Chernoff regular at θ_0, which means C is well-approximated by a cone K at θ_0 in the sense 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 θ_0 to lie either in the interior or on the boundary of C, and we will estimate θ_0 using a constrained M-estimator given by

$$\hat{\theta}_n = \arg\min_{\theta \in C} \{\hat{\pi}_n(\theta) \equiv P_n \pi(\cdot, \theta)\}$$

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 θ_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 θ_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 θ_0.

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

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

(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 θ_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 $\mathcal{G}_R \equiv \{g(\cdot,\theta) : \|\theta-\theta_0\| \leqslant R\}$ for R near zero is uniformly manageable for the envelope function $G_R(\cdot) \equiv \sup_{g\in\mathcal{G}_R} |g(\cdot,\theta)|$.

(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 θ with a stochastically bounded Lipschitz constant.

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

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

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

(viii) $\pi(\theta) = P\pi(\cdot; \theta)$ *is twice differentiable at* θ_0 *with Jacobian* $l(\theta_0) \equiv \frac{\partial}{\partial\theta}\pi(\theta_0)$ *and 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 θ_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 γ. 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 $\mathcal{G}_R$ 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 $\mathcal{G}_R$ 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 θ_0 by $\hat{\theta}_n$ in the bootstrapped empirical processes.

Assumption 3. *For some $\rho \in (0,1]$ and $\gamma = \frac{1}{2(2-\rho)}$, define $m_n(\cdot,\theta,h) \equiv n^{\gamma\rho}\left(\pi\left(\cdot;\theta+\frac{h}{n^{\gamma}}\right)-\pi(\cdot;\theta)\right)$. For any $\epsilon_n \to 0$ and any compact set $\mathcal{H} \subset \mathbb{R}^d$,*

$$\lim_{\lambda\to\infty}\lim_{n\to\infty}\sup\sup_{t\geqslant\lambda} t^2\ P\left\{\sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n}\left\|\frac{m_n(\cdot,\theta,h)-m_n(\cdot,\theta_0,h)}{1+n^{\gamma}\|\theta-\theta_0\|}\right\|>t\right\}=0.$$

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 $\mathcal{H} \subset \mathbb{R}^d$, there exists some constant $K > 0$ such that $\sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n}\left\|\frac{m_n(\cdot,\theta,h)-m_n(\cdot,\theta_0,h)}{1+n^{\gamma}\|\theta-\theta_0\|}\right\| \leqslant K$. 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 θ_0, which means that the constrained minimizer is the same as the unconstrained minimizer of the population objective.

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

This assumption requires that the sum of the Lagrange multipliers times the constraint gradients must be zero at θ_0. Suppose $C \equiv \{\theta \in \Theta : f_j(\theta) = 0 \text{ for } j \in \mathcal{E}, f_j(\theta) \leqslant 0 \text{ for } j \in \mathcal{I}\}$. The first order KKT condition says that θ_0 solves the population constrained optimization problem if $l(\theta_0) + \sum_{j\in\mathcal{E}\cup\mathcal{I}} \lambda_{0j} f_j'(\theta_0) = 0$. By imposing $l(\theta_0) = 0$, we are requiring that $\sum_{j\in\mathcal{E}\cup\mathcal{I}} \lambda_{0j} f_j'(\theta_0) = 0$. If we further assume linear independence constraint qualification

(LICQ), then $l(\theta_0) = 0$ implies $\lambda_{0j} = 0$ for all $j \in \mathcal{E} \cup \mathcal{I}^*$. The reason is as follows. LICQ says that the gradients of the active constraints $f_j'(\theta_0) \equiv \left.\frac{\partial f_j(\theta)}{\partial \theta}\right|_{\theta=\theta_0}$ for $j \in \mathcal{E} \cup \mathcal{I}^*$, where $\mathcal{I}^* \equiv \{j \in \mathcal{I} : f_j(\theta_0) = 0\}$, are linearly independent. LICQ implies that $f_j'(\theta_0) \neq 0$ for all $j \in \mathcal{E} \cup \mathcal{I}^*$, which means that $\sum_{j \in \mathcal{E} \cup \mathcal{I}} \lambda_{0j} f_j'(\theta_0) = 0$ will imply $\lambda_{0j} = 0$ for all $j \in \mathcal{E} \cup \mathcal{I}^*$. The constraints which are non-active have zero Lagrange multipliers and the weakly active constraints also have zero Lagrange multipliers. However, under LICQ, the strongly active constraints have positive Lagrange multipliers and must be ruled out in order for the first order KKT condition to reduce down to the condition $l(\theta_0) = 0$. In Examples 1-2, the constraint $\theta \geqslant 0$ will be strongly active at $\theta_0 = \max(\beta_0, 0)$ if $\beta_0 < 0$. Assumption 4 allows for weakly active and inactive constraints at θ_0. In Examples 1-2, the constraint $\theta \geqslant 0$ will be weakly active at $\theta_0 = \max(\beta_0, 0)$ if $\beta_0 = 0$ and inactive if $\beta_0 > 0$.

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

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

where $\hat{H} \xrightarrow{p} H_0$ 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 θ is the unconstrained minimizer of the population objective, then the sample objective should achieve its minimum close to θ even if we perturb θ 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 $n^\gamma h' l(\theta_0)$ 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 $1-\alpha$ conditional quantile of $-\inf_{h\in\mathbb{R}^d}\hat{\mathbb{H}}_n(h)$. We will show that $\mathcal{C}^*_{1-\alpha} = \left\{\theta : n^{2\gamma}\left(\hat{\pi}_n(\theta) - \inf_{h\in\mathcal{B}_{\delta_n}}\hat{\pi}_n\left(\theta+\frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right\}$ 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 $c \geqslant 0$ is some constant and $\tau_n \to \infty$ as $n \to \infty$.

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

Theorem 1. *(Uniformly valid inference when constraints are not necessary for identification) Let $\mathcal{P}$ be a class of distributions for which Assumptions 1-4 are satisfied uniformly in $P \in \mathcal{P}$* [1]*, 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 $\liminf_{n\to\infty}\inf_{P\in\mathcal{P}} P\left(\theta(P) \in \mathcal{C}^*_{1-\alpha}\right) \geqslant 1-\alpha$, where*

$$\mathcal{C}^*_{1-\alpha} = \left\{\theta : n^{2\gamma}\left(\hat{\pi}_n(\theta) - \inf_{h\in\mathcal{B}_{\delta_n}}\hat{\pi}_n\left(\theta+\frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right\},$$

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

[1] We define $X_n = o_P(1)$ uniformly in $P \in \mathcal{P}$ if if $\sup_{P\in\mathcal{P}} P(|X_n| > \epsilon) = 0\ \forall\epsilon > 0$. We define $M_n = O_P(1)$ uniformly in $P \in \mathcal{P}$ if $\lim_{m\to\infty}\limsup_{n\to\infty}\sup_{P\in\mathcal{P}} P(|M_n| > m) = 0$. Similarly, $X^*_n = o^*_P(1)$ uniformly in $P \in \mathcal{P}$ if if $\sup_{P\in\mathcal{P}} P(|X^*_n| > \epsilon|\mathcal{X}_n) = o_P(1)$ for all $\epsilon > 0$. $M^*_n = O^*_P(1)$ uniformly in $P \in \mathcal{P}$ if $\lim_{m\to\infty}\limsup_{n\to\infty}\sup_{P\in\mathcal{P}} P(P(|M^*_n| > m|\mathcal{X}_n) > \epsilon) = 0\ \forall\epsilon > 0$.

Remark 1. Although our current setup has assumed C is a fixed (non-random) constraint set, the result in Theorem 1 allows for the constraints to be estimated using the data as long as the sample constrained estimator $\hat{\theta}_n$ is consistent for the population constrained argmin θ_0. The reason is that Theorem 1 does not use the constraints in any way except through the consistency of $\hat{\theta}_n$ for θ_0. There are many cases where $\hat{\theta}_n$ will remain consistent even when the constraints are estimated. For example, Knight (1999) states on page 13 that if the sample constraint set is constructed using convex, finite-valued functions that converge to their population limits, then the indicator function for the sample constraint set will epi-converge in distribution to the indicator function for the population constraint set. If, additionally the sample objective converges uniformly to the population objective, then Theorem 1 in Knight (1999) will imply the sample constrained argmin converges in probability to the population constrained argmin. Other cases where consistency holds are given in Dupacová and Wets (1988), Shapiro (1990), Robinson (1996), and Bonnans and Shapiro (2013), among other papers.

Remark 2. In theory, any choice of $\kappa \in (0, \infty]$ will achieve uniformly valid coverage, but if $\hat{\pi}_n(\theta)$ is nonsmooth or nonconvex, in practice setting $\kappa < \infty$ can help the solver more easily find the solution to $\inf_{h \in \mathcal{B}_{\delta_n}} \hat{\pi}_n \left(\theta + \frac{h}{\sqrt{n}}\right)$. The choice of κ can also affect the conservativeness of the confidence set's coverage, with larger values of κ typically leading to less conservative coverage asymptotically. The reason is that the test statistic's limiting distribution $-\inf_{\{h \in \mathbb{R}^d : \|h\| \leqslant \kappa\}} \mathbb{H}_0(h)$ is closer to $-\inf_{h \in \mathbb{R}^d} \hat{\mathbb{H}}_n(h)$'s limiting distribution $-\inf_{h \in \mathbb{R}^d} \mathbb{H}_0(h)$ for larger values of κ. In our Monte Carlo simulations we saw that $\kappa = 1$ leads to more conservative coverage than $\kappa = 5$, but there is very little difference in coverage between $\kappa = 5$ and $\kappa = \infty$. Additionally, the average interval length is shorter for some parameters when we use $\kappa = 5$ instead of $\kappa = \infty$. Therefore, we recommend that practitioners use a moderate sized, finite value of κ such as $\kappa = 5$.

Remark 3. 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: $CI_{1-\alpha}^{Proj} = \left[\inf_{\theta \in \mathcal{C}_{1-\alpha}^{*}} a'\theta, \sup_{\theta \in \mathcal{C}_{1-\alpha}^{*}} a'\theta\right]$. The uniform asymptotic validity of these projection intervals follows directly from the uniform asymptotic validity of $\mathcal{C}_{1-\alpha}^{*}$.

Now we relax Assumption 4 to allow for possibility that the unconstrained minimizer differs from the constrained minimizer of the population problem $\left(l\left(\theta_0\right) \equiv \frac{\partial}{\partial\theta}\pi\left(\theta_0\right) \neq 0\right)$. Now we allow for possibility that the unconstrained minimizer differs from the constrained minimizer of the population problem $\left(l\left(\theta_0\right) \equiv \frac{\partial}{\partial\theta}Q\left(\theta_0\right) \neq 0\right)$. Under LICQ, $l\left(\theta_0\right) \neq 0$ implies that some constraint(s) are strongly active at θ_0 and are therefore necessary for identification of θ_0. We will modify our test statistic to $n^{2\gamma}\left(\hat{\pi}_n\left(\theta\right) - \inf_{h \in \mathcal{C}_{\delta_n}^{\theta}} \hat{\pi}_n\left(\theta + \frac{h}{n^{\gamma}}\right)\right)$, where $\mathcal{C}_{\delta_n}^{\theta} = \left\{h \in n^{\gamma}\left(C - \theta\right) : \frac{\|h\|}{n^{\gamma}} \leqslant \delta_n\right\}$, 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 $\hat{\pi}_n\left(\theta\right)$ is non-smooth and/or nonconvex, in practice setting $\kappa < \infty$ helps the solver find the solution to $\inf_{h \in \mathcal{C}_{\delta_n}^{\theta}} \hat{\pi}_n\left(\theta + \frac{h}{n^{\gamma}}\right)$. The intuition behind the test statistic is that if θ is the constrained minimizer of the population objective, then the sample objective should achieve its minimum close to θ even if we perturb θ by small deviations while still satisfying the constraints.

We minimize over the constrained neighborhood $\mathcal{C}_{\delta_n}^{\theta}$ when constructing our test statistic because the troublesome term $n^{\gamma}h'l\left(\theta_0\right)$ 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 $1-\alpha$ conditional quantile of $-\inf_{h\in\mathbb{R}^d}\hat{\mathbb{H}}_n(h)$. We will show that $\mathcal{C}^*_{1-\alpha}=\left\{\theta : n^{2\gamma}\left(\hat{\pi}_n(\theta)-\inf_{h\in\mathcal{C}^{\theta}_{\delta_n}}\hat{\pi}_n\left(\theta+\frac{h}{n^\gamma}\right)\right)\leqslant \hat{c}^*_{1-\alpha}\right\}$ 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 $\hat{\mathbb{H}}_n(h)$ because we cannot uniformly consistently estimate the tangent cone of the constraint set at θ_0. Since we do not observe θ_0, we would have to replace θ_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 θ_0, which introduces an additional $n^\gamma\left(\hat{\theta}_n-\theta_0\right)$ term. However, this procedure would only be pointwise valid because the convergence of $n^\gamma\left(\hat{\theta}_n-\theta_0\right)$ to its limiting distribution is not uniform over P.

In the next theorem, $J_n(\cdot,P)$ denotes the CDF of $T_n\equiv-\inf_{\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}}\left\{n^{2\gamma}(P_n-P)\left(\pi\left(\cdot,\theta_0+\frac{h}{n^\gamma}\right)-\pi(\cdot,\theta_0)\right)+\frac{1}{2}h'H_0h\right\}$ 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 θ_0: $T_C(\theta_0)\equiv\limsup_{\tau\downarrow 0}\frac{C-\theta_0}{\tau}$. Additionally, $J^*_{\alpha_n}(\cdot,P)$ denotes the conditional CDF of $-\inf_{h\in\mathbb{R}^d}\hat{\mathbb{H}}_n(h)$ under P, and $J^*(\cdot,P)$ denotes the CDF of its limiting distribution under P.

Theorem 2. *(Uniformly valid inference when constraints may be necessary for identification) 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 $J_n^{-1}(1-\alpha,P)$. Then $\liminf_{n\to\infty}\inf_{P\in\mathcal{P}}P\left(\theta(P)\in\mathcal{C}^*_{1-\alpha}\right)\geqslant 1-\alpha$, where*

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

$\mathcal{C}^{\theta}_{\delta_n}=\left\{h\in n^\gamma(C-\theta):\frac{\|h\|}{n^\gamma}\leqslant\delta_n\right\}$, $\delta_n\to 0$ satisfies $n^\gamma\delta_n\to\kappa$ for $\kappa\in(0,\infty]$, and $\hat{c}^*_{1-\alpha}$ is

the $1-\alpha$ *quantile of* $-\inf_{h\in\mathbb{R}^d}\hat{\mathbb{H}}_n(h)$.

Remark 4. When Assumption 4 holds, both the confidence set in Theorem 2 and the confidence set in Theorem 1 will be valid. In this case, we find that the confidence set in Theorem 1 tends to be less conservative than the confidence set in Theorem 1 because the limiting distribution of the test statistic $n^{2\gamma}\left(\hat{\pi}_n(\theta_0)-\inf_{h\in\mathcal{B}_{\delta_n}}\hat{\pi}_n\left(\theta_0+\frac{h}{n^\gamma}\right)\right)$ is closer to the limiting distribution of $-\inf_{h\in\mathbb{R}^d}\hat{\mathbb{H}}_n(h)$. However, if the researcher is unsure of whether 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 $\pi : \mathcal{X} \times \Theta \to \mathbb{R}^m$, sample moments $\hat{\pi}_n(\theta) \equiv P_n \pi(\cdot, \theta)$, and population moments $\pi(\theta) \equiv P\pi(\cdot, \theta)$. 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 θ_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) $\hat{Q}_n\left(\hat{\theta}_n\right) \leqslant \inf_{\theta \in C} \hat{Q}_n(\theta) + o_p\left(n^{-2\gamma}\right)$.

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

(v) C *is Chernoff regular at* θ_0.

Assumption 6. *There exists some* $\rho \in \left\{\frac{1}{2}, 1\right\}$ *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* $\mathcal{G}_R = \{g_j(\cdot,\theta) : \|\theta - \theta_0\| \leqslant R, j = 1, \ldots, m\}$ *for* R *near zero are uniformly manageable for the envelope functions* $G_R(\cdot) \equiv \sup_{g_j \in \mathcal{G}_R} |g_j(\cdot,\theta)|$.

(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* θ *with a stochastically bounded Lipschitz constant.*

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

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

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

(ix) $Pg(\cdot,\theta)$ *is twice differentiable at* θ_0 *with full rank Jacobian matrix* $G = \frac{\partial}{\partial\theta}\pi(\theta_0)$ *and finite Hessian matrices* $H_j = \frac{\partial^2}{\partial\theta\partial\theta'}\pi_j(\theta_0)$ *for* $j = 1 \ldots m$.

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 θ_0, which means that the constrained minimizer of the population objective is the same as the unconstrained minimizer.

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

Assumption 7 will rule out strongly active constraints at θ_0 when LICQ holds. In Example 3, the constraint $\theta \geqslant 0$ will be strongly active at $\theta_0 = \underset{\theta\geqslant 0}{\arg\min}\ \frac{1}{2}\pi(\theta)'\pi(\theta)$ if $\underset{\theta\geqslant 0}{\arg\min}\ \frac{1}{2}\pi(\theta)'\pi(\theta) \neq \underset{\theta\in\mathbb{R}}{\arg\min}\ \frac{1}{2}\pi(\theta)'\pi(\theta)$. If $\tau = 0.5$, Assumption 7 implies that $\theta_0 = \beta_0$, so that both the moments and the non-negativity constraint are correctly specified.

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

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

$\hat{G}$ 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 θ is the unconstrained minimizer of the population objective, then the sample objective should achieve its minimum close to θ even if we perturb θ 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 $1-\alpha$ conditional quantile of $-\inf_{h \in \mathbb{R}^d} \hat{A}_n(h)$. We will show that $\mathcal{C}^*_{1-\alpha} = \left\{\theta : n^{2\gamma}\left(\hat{Q}_n(\theta) - \inf_{h \in \mathcal{B}_{\delta_n}} \hat{Q}_n\left(\theta + \frac{h}{n^{\gamma}}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right\}$ 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 $\epsilon_n \to 0$, $\lim_{\lambda \to \infty} \lim_{n \to \infty} \sup \sup_{t \geqslant \lambda} t^2 \; P\left\{\sup_{\|\theta - \theta_0\| \leqslant \epsilon_n} \left\| \frac{\pi(\cdot, \theta) - \pi(\cdot, \theta_0)}{1 + \sqrt{n}\|\theta - \theta_0\|} \right\| > t\right\} = 0$. Additionally, for each $\epsilon > 0$ and $t \in \mathbb{R}^d$,*

$$\lim_{n \to \infty} P \left\| \begin{pmatrix} \sqrt{n} g\left(\cdot, \theta_0 + \frac{t}{\sqrt{n}}\right) \\ \pi(\cdot, \theta_0) \end{pmatrix} \right\|^2 1\left\{ \left\| \begin{pmatrix} \sqrt{n} g\left(\cdot, \theta_0 + \frac{t}{\sqrt{n}}\right) \\ \pi(\cdot, \theta_0) \end{pmatrix} \right\| > \epsilon\sqrt{n} \right\} = 0.$$

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

Theorem 3. *(Uniformly valid inference when constraints are not necessary for identification) 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* $\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} P\left(\theta(P) \in \mathcal{C}^*_{1-\alpha}\right) \geqslant 1-\alpha$*, where*

$$\mathcal{C}^*_{1-\alpha} = \left\{\theta : n^{2\gamma}\left(\hat{Q}_n(\theta) - \inf_{h\in\mathcal{B}_{\delta_n}} \hat{Q}_n\left(\theta + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right\},$$

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

Now suppose the constraints may be necessary for identification of θ_0 so that Assumption 7 may not hold. We will modify our test statistic to $n^{2\gamma}\left(\hat{Q}_n(\theta) - \inf_{h\in\mathcal{C}^\theta_{\delta_n}} \hat{Q}_n\left(\theta + \frac{h}{n^\gamma}\right)\right)$, where $\mathcal{C}^\theta_{\delta_n} = \left\{h \in n^\gamma(C-\theta) : \frac{\|h\|}{n^\gamma} \leqslant \delta_n\right\}$, 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 θ is the constrained minimizer of the population objective, then the sample objective should achieve its minimum close to θ even if we perturb θ by small deviations while still satisfying the constraints. In theory, the choice of κ 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 $\inf_{h\in\mathcal{C}^\theta_{\delta_n}} \hat{Q}_n\left(\theta + \frac{h}{n^\gamma}\right)$.

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 $1-\alpha$ conditional 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 θ_0. Since we do not observe θ_0, we would have to replace θ_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 θ_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 $\mathcal{C}^*_{1-\alpha} = \left\{\theta : n^{2\gamma}\left(\hat{Q}_n(\theta) - \inf_{h \in \mathcal{C}^\theta_{\delta_n}} \hat{Q}_n\left(\theta + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right\}$ 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 $S_n = -\inf_{h \in \{h \in T_C(\theta_0): \|h\| \leqslant \kappa\}} \left\{n^{2\gamma}\left(\hat{Q}_2\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_3\left(\theta_0 + \frac{h}{n^\gamma}\right)\right) + \frac{1}{2}h'\bar{H}h\right\}$ 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 $J^*_{\alpha_n}(\cdot, P)$ denote the conditional CDF of $-\inf_{h \in \mathbb{R}^d} \hat{A}_n(h)$ under P, and let $J^*(\cdot, P)$ denote the CDF of its limiting distribution under P.

Theorem 4. *(Uniformly valid inference when constraints may be necessary for identification) 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 $\{J^*(\cdot, P) : P \in \mathcal{P}\}$ are equicontinuous at $J_n^{-1}(1-\alpha, P)$. Then $\liminf_{n\to\infty} \inf_{P \in \mathcal{P}} P\left(\theta(P) \in \mathcal{C}^*_{1-\alpha}\right) \geqslant 1-\alpha$, where*

$$\mathcal{C}^*_{1-\alpha} = \left\{\theta : n^{2\gamma}\left(\hat{Q}_n(\theta) - \inf_{h \in \mathcal{C}^\theta_{\delta_n}} \hat{Q}_n\left(\theta + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right\},$$

$\mathcal{C}^\theta_{\delta_n} = \left\{h \in n^\gamma(C - \theta) : \frac{\|h\|}{n^\gamma} \leqslant \delta_n\right\}$, $\delta_n \to 0$ satisfies $n^\gamma \delta_n \to \kappa$ for $\kappa \in (0, \infty]$, and $\hat{c}^*_{1-\alpha}$ is the $1-\alpha$ conditional 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 $\hat{\theta}_n = \arg\min_{\theta \in C} \left\{ \hat{Q}_n(\theta) = \frac{1}{2}\hat{\pi}(\theta)' W_n \hat{\pi}(\theta) \right\}$. 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 $\hat{\theta}_1^C = \arg\min_{\theta \in C} \frac{1}{2}\hat{\pi}(\theta)' W_1 \hat{\pi}(\theta)$. Note that we need to redefine the presumed to be unique pseudo-true parameter to be $\theta_0 = \arg\min_{\theta \in C} \frac{1}{2}\pi(\theta)' W \pi(\theta)$ 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 $W = W\left(\theta_1^C\right)$, 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 $G_1' W_1 \pi\left(\theta_1^C\right) = 0$, where $G_1 = \frac{\partial}{\partial \theta}\pi\left(\theta_1^C\right)$, then $\theta_1^C = \theta_1$ and both $W = W(\theta_1)$ and $W = W\left(\theta_1^C\right)$ lead to the same pseudo-true parameter θ_0. However, if $G' W_1 \pi\left(\theta_1^C\right) \neq 0$, meaning that the constraints matter for identification, then θ_1^C will differ from θ_1, and depending on whether we set $W = W(\theta_1)$ or $W = W\left(\theta_1^C\right)$, we can obtain different values of θ_0. Because we would like to enforce the constraints to identify the parameters, we would typically use $W = W\left(\theta_1^C\right)$ 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 test statistic $n^{2\gamma}\left(\hat{Q}_n(\theta) - \inf_{h \in \mathcal{C}^\theta_{\delta_n}} \hat{Q}_n\left(\theta + \frac{h}{n^\gamma}\right)\right)$ against the conditional quantiles of $-\inf_{h \in \mathbb{R}^d} \hat{B}_n(h)$, where

$$\hat{B}_n(h) = n^{2\gamma}\hat{\pi}_n\left(\hat{\theta}_n\right)' W_n \left(P_n^* - P_n\right)\left(\pi\left(\cdot, \hat{\theta}_n + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \hat{\theta}_n\right)\right) \tag{3}$$

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

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

$$\hat{\theta}_1^{C*}=\underset{\theta\in\frac{C-\hat{\theta}_1^C}{\eta_n}+\hat{\theta}_1^C}{\arg\min}\Bigg\{\hat{\pi}_n\left(\hat{\theta}_1^C\right)'W_1\left(P_n^*-P_n\right)\left(\pi\left(\cdot,\theta\right)-\pi\left(\cdot,\hat{\theta}_1^C\right)\right)$$
$$+\frac{1}{2}\left(\theta-\hat{\theta}_1^C\right)'\left(\hat{G}'W_1\hat{G}+\sum_{j=1}^{m}\sum_{k=1}^{m}W_{1,jk}\hat{\pi}_{nk}\left(\hat{\theta}_1^C\right)\hat{H}_j\right)\left(\theta-\hat{\theta}_1^C\right)$$
$$+\left(\theta-\hat{\theta}_1^C\right)'\hat{G}'W_1\left(P_n^*-P_n\right)\pi\left(\cdot,\hat{\theta}_1^C\right)\Bigg\},\tag{4}$$

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 θ_1^C.

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

$$\hat{\theta}_1^*=\underset{\theta\in\Theta}{\arg\min}\Bigg\{\hat{\pi}_n\left(\hat{\theta}_1\right)'W_1\left(P_n^*-P_n\right)\left(\pi\left(\cdot,\theta\right)-\pi\left(\cdot,\hat{\theta}_1\right)\right)$$
$$+\frac{1}{2}\left(\theta-\hat{\theta}_1\right)'\left(\hat{G}'W_1\hat{G}+\sum_{j=1}^{m}\sum_{k=1}^{m}W_{1,jk}\hat{\pi}_{nk}\left(\hat{\theta}_1\right)\hat{H}_j\right)\left(\theta-\hat{\theta}_1\right)$$
$$+\left(\theta-\hat{\theta}_1\right)'\hat{G}'W_1\left(P_n^*-P_n\right)\pi\left(\cdot,\hat{\theta}_1\right)\Bigg\}\tag{5}$$

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. *γ, W_n and W_n^* can fall into one of the following cases:*

(i) $\gamma = 1/2$, $W_n - W = o_p\left(n^{-1/2}\right)$ *and* $W_n^* - W_n = o_p^*\left(n^{-1/2}\right)$.

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

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

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

Furthermore, If $\gamma = 1/2$, $W_n - W = O_p\left(n^{-1/2}\right)$ *and* $W_n^* - W_n = O_p^*\left(n^{-1/2}\right)$, *then*

$$\begin{pmatrix} \pi\left(\theta_0\right)' W_n n\left(P_n - P\right) g\left(\cdot, \theta_0 + n^{-1/2}h\right) \\ h'G'W_n\sqrt{n}\left(P_n - P\right)\pi\left(\cdot, \theta_0\right) \\ h'G'\sqrt{n}\left(W_n - W\right)\pi\left(\theta_0\right) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \pi\left(\theta_0\right)' W \mathcal{Z}_{0,1}\left(h\right) \\ h'G'WU_0 \\ h'G'\mathcal{W}_0 \end{pmatrix} \text{ and}$$

$$\begin{pmatrix} \hat{\pi}_n\left(\hat{\theta}_n\right)' W_n n\left(P_n^* - P_n\right)\left(\pi\left(\cdot, \hat{\theta}_n + \frac{h}{\sqrt{n}}\right) - \pi\left(\cdot, \hat{\theta}_n\right)\right) \\ h'\hat{G}'W_n\sqrt{n}\left(P_n^* - P_n\right)\pi\left(\cdot, \hat{\theta}_n\right) \\ h'\hat{G}'\sqrt{n}\left(W_n^* - W_n\right)\hat{\pi}_n\left(\hat{\theta}_n\right) \end{pmatrix} \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \begin{pmatrix} \pi\left(\theta_0\right)' W \mathcal{Z}_{0,1}\left(h\right) \\ h'G'WU_0 \\ h'G'\mathcal{W}_0 \end{pmatrix} \text{ in the}$$

product space of locally bounded functions $\left\{\mathbf{B}_{loc}\left(\mathbb{R}^d\right)\right\}^3$ *for some tight random vector* $\mathcal{W}_0$. *Here,* $U_0 \sim N\left(0, P\left(\pi\left(\cdot, \theta_0\right) - \pi\left(\theta_0\right)\right)\left(\pi\left(\cdot, \theta_0\right) - \pi\left(\theta_0\right)\right)'\right)$ *and* $\mathcal{Z}_{0,1}\left(h\right)$ *is a mean zero Gaussian process with covariance kernel* $\Sigma_1\left(s, t\right) = \lim_{\alpha \to \infty} \alpha^2 Pg\left(\cdot, \theta_0 + \frac{s}{\alpha}\right) g\left(\cdot, \theta_0 + \frac{t}{\alpha}\right)'$.

If $\gamma = 1/3$, $W_n - W = O_p\left(n^{-1/3}\right)$ *and* $W_n^* - W_n = O_p^*\left(n^{-1/3}\right)$, *then*

$$\begin{pmatrix} \pi\left(\theta_0\right)' W n^{2/3}\left(P_n - P\right) g\left(\cdot, \theta_0 + n^{-1/3}h\right) \\ h'G'n^{1/3}\left(W_n - W\right)\pi\left(\theta_0\right) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \pi\left(\theta_0\right)' W \mathcal{Z}_{0,1/2}\left(h\right) \\ h'G'\mathcal{W}_0 \end{pmatrix} \text{ and}$$

$$\begin{pmatrix} \pi(\theta_0)' W n^{2/3} (P_n^* - P_n) g(\cdot, \theta_0 + n^{-1/3} h) \\ h' G' n^{1/3} (W_n^* - W_n) \pi(\theta_0) \end{pmatrix} \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \begin{pmatrix} \pi(\theta_0)' W \mathcal{Z}_{0,1/2}(h) \\ h' G' \mathcal{W}_0 \end{pmatrix}$$ *in the product space of locally bounded functions* $\left\{\mathbf{B}_{loc}\left(\mathbb{R}^d\right)\right\}^2$ *for some tight random vector* $\mathcal{W}_0$*. Here,* $\mathcal{Z}_{0,1/2}(h)$ *is a mean zero Gaussian process with covariance kernel* $\Sigma_{1/2}(s,t) = \lim_{\alpha\to\infty} \alpha P g\left(\cdot, \theta_0 + \frac{s}{\alpha}\right) g\left(\cdot, \theta_0 + \frac{t}{\alpha}\right)'$.

Let $\hat{c}^*_{1-\alpha}$ be the $1-\alpha$ conditional quantile of $-\inf_{h\in\mathbb{R}^d} \hat{B}_n(h)$. We will show that $\mathcal{C}^*_{1-\alpha} = \left\{\theta : n^{2\gamma}\left(\hat{Q}_n(\theta) - \inf_{h\in\mathcal{C}^\theta_{\delta_n}} \hat{Q}_n\left(\theta + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right\}$ 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 $R_n = -\inf_{h\in\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}} \left\{n^{2\gamma}\bar{Q}_n\left(\theta_0 + \frac{h}{n^\gamma}\right) + \frac{1}{2}h'\bar{H}h\right\}$ under P, where $\bar{Q}_n(\cdot)$ is defined in equation 9, 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. *(Uniformly valid inference when constraints may be necessary for identification) 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* $\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} P\left(\theta(P) \in \mathcal{C}^*_{1-\alpha}\right) \geqslant 1-\alpha$*, where*

$$\mathcal{C}^*_{1-\alpha} = \left\{\theta : n^{2\gamma}\left(\hat{Q}_n(\theta) - \inf_{h\in\mathcal{C}^\theta_{\delta_n}} \hat{Q}_n\left(\theta + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right\},$$

$\mathcal{C}^\theta_{\delta_n} = \left\{h \in n^\gamma(C-\theta) : \frac{\|h\|}{n^\gamma} \leqslant \delta_n\right\}$*,* $\delta_n \to 0$ *satisfies* $n^\gamma\delta_n \to \kappa$ *for* $\kappa \in (0,\infty]$*, and* $\hat{c}^*_{1-\alpha}$ *is the* $1-\alpha$ *conditional 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 γ. In other words, if we are willing to assume that θ_0 is away from the boundary of C, then we can be agnostic about whether the model is correctly specified or globally misspecified (we have to exclude the possibility of local misspecification). Our pointwise valid procedure for constrained GMM with a fixed weighting matrix is as follows:

1. Compute $\hat{\theta}_n = \arg\min_{\theta \in C} \frac{1}{2}\hat{\pi}_n(\theta)' W \hat{\pi}_n(\theta)$, $\hat{\pi}_n\left(\hat{\theta}_n\right) = \frac{1}{n}\sum_{i=1}^n \pi\left(X_i, \hat{\theta}_n\right)$, $\hat{G}$, $\hat{H}_j$ 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

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

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 θ_{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 $\rho(\theta_0)$, where $\rho : \Theta \mapsto \mathbb{R}$, 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 γ. Since we know that $\gamma \geqslant 1/3$ for the class of GMM estimators we consider, the sequence η_n will satisfy $\eta_n/n^\gamma \to 0$ and remove the additional noise caused by centering the constraint set around $\hat{\theta}_n$ instead of the unknown θ_0.

Theorem 6. *Suppose Assumptions 3, 5-6 and 8 hold,* $\hat{G} \xrightarrow{p} G$, *and* $\hat{H}_j \xrightarrow{p} H_j$ *for* $j = 1 \ldots m$. *Also suppose* $\bar{H} = G'WG + \sum_{j=1}^m \sum_{k=1}^m W_{jk}\pi_k(\theta_0) H_j$ *is positive definite. Then,*

$$n^\gamma \left(\hat{\theta}_n - \theta_0\right) \rightsquigarrow \underset{h \in T_C(\theta_0)}{\arg\min}\ A_0(h), \quad n^\gamma \left(\hat{\theta}_n^* - \hat{\theta}_n\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \underset{h \in T_C(\theta_0)}{\arg\min}\ A_0(h)$$

where $\bar{H} = G'WG + \sum_{j=1}^m \sum_{k=1}^m W_{jk}\pi_k(\theta_0) H_j$, $T_C(\theta_0) \equiv \limsup_{\tau \downarrow 0} \frac{C-\theta_0}{\tau}$, $A_0(h) = \pi(\theta_0)' W \mathcal{Z}_{0,1}(h) + h'G'WU_0 + \frac{1}{2}h'\bar{H}h$ *for* $\gamma = 1/2$, *and* $A_0(h) = \pi(\theta_0)' W \mathcal{Z}_{0,1/2}(h) + \frac{1}{2}h'\bar{H}h$ *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* $\Sigma_1(s,t) = \lim_{\alpha\to\infty} \alpha^2 P g\left(\cdot, \theta_0 + \frac{s}{\alpha}\right) g\left(\cdot, \theta_0 + \frac{t}{\alpha}\right)'$, *and* $\mathcal{Z}_{0,1/2}(h)$ *is a mean zero Gaussian process with covariance kernel* $\Sigma_{1/2}(s,t) = \lim_{\alpha\to\infty} \alpha P g\left(\cdot, \theta_0 + \frac{s}{\alpha}\right) g\left(\cdot, \theta_0 + \frac{t}{\alpha}\right)'$.

4 Monte Carlo

4.1 Nonsmooth Location Model

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

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

For $\pi(\cdot,\theta) = [1(y_i \leqslant \theta) - \tau; y_i - \theta]'$, the population moments are $\pi(\theta) = [P(y_i \leqslant \theta) - \tau; Ey_i - \theta]'$. 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, meaning that $\tau = c \neq 0.5$ for some fixed constant c. As demonstrated in Hong and Li (2023), the standard bootstrap will undercover the parameter of interest whenever $\tau \neq 0.5$, with the undercoverage becoming more severe as τ moves further away from 0.5. Additionally, if $\tau = 0.5$, Li (2023) has shown that the standard bootstrap and subsampling percentile intervals will under-cover drifting values of β_0 when we impose a non-negativity constraint. 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 all 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:

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

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 $C = \{\theta : \theta \geqslant 0\}$. We call θ_0 the pseudo true value because it will not be equal to β_0 if $\tau \neq 0.5$ or $\beta_0 < 0$.

We will examine the empirical coverage and average width of the confidence set $\mathcal{C}^*_{1-\alpha} = \left\{\theta : n^{2\gamma}\left(\hat{Q}_n(\theta) - \inf_{h \in \mathcal{C}^{\theta}_{\delta_n}} \hat{Q}_n\left(\theta + \frac{h}{n^{\gamma}}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right\}$, where $\mathcal{C}^{\theta}_{\delta_n} = \left\{h \in n^{\gamma}(C - \theta) : \frac{\|h\|}{n^{\gamma}} \leqslant \delta_n\right\}$, $\delta_n \to 0$ satisfies $n^{\gamma}\delta_n \to \kappa$ for $\kappa \in (0,\infty]$, and $\hat{c}^*_{1-\alpha}$ is the $1-\alpha$ empirical 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 our estimators. Additionally, we estimate G and H using kernel estimators:

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

$K_h(x) = K(x/h)$, $K(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, $K'_h(x) = K'(x/h)$ and $K'(x) = -(2\pi)^{-1/2} x\exp(-x^2/2)$, where h 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 a nominal 95% confidence set when $\beta_0 \geqslant 0$. We consider a range of different values of $\beta_0 \in \left\{0, n^{-1}, n^{-1/2}, n^{-1/3}, n^{-1/4}, n^{-1/6}, 1\right\}$, 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 τ 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 β_0 close to zero (when the constraint becomes weakly active), but approaches the nominal level as β_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 β_0.

Table 2 shows the empirical coverage frequencies and average interval lengths (in parentheses) of a nominal 95% confidence set when $\beta_0 < 0$. We consider a range of different values of $\beta_0 \in \left\{-n^{-1}, -n^{-1/2}, -n^{-1/3}, -n^{-1/4}, -n^{-1/6}, -1\right\}$. Because negative values of β_0 violate the non-negativity constraint, the non-negativity constraint is misspecified and therefore

Table 1: Coverage Frequencies and Average Interval Lengths, $\kappa = 5$

β_0	0	n^{-1}	$n^{-1/2}$	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/6}$	1
				$\tau = 0.1$			
$n = 100$	1.000	1.000	0.990	0.979	0.964	0.956	0.945
	(0.559)	(0.543)	(0.472)	(0.430)	(0.424)	(0.421)	(0.420)
$n = 500$	1.000	1.000	1.000	0.991	0.964	0.963	0.960
	(0.369)	(0.360)	(0.303)	(0.232)	(0.211)	(0.211)	(0.211)
$n = 1000$	1.000	1.000	1.000	0.999	0.980	0.957	0.958
	(0.328)	(0.327)	(0.276)	(0.195)	(0.161)	(0.159)	(0.159)
$n = 5000$	1.000	1.000	1.000	1.000	0.999	0.969	0.970
	(0.291)	(0.289)	(0.264)	(0.186)	(0.103)	(0.085)	(0.086)
				$\tau = 0.3$			
$n = 100$	0.996	0.994	0.971	0.960	0.952	0.947	0.945
	(0.464)	(0.452)	(0.410)	(0.393)	(0.393)	(0.392)	(0.393)
$n = 500$	1.000	1.000	0.995	0.978	0.959	0.956	0.954
	(0.253)	(0.246)	(0.210)	(0.186)	(0.184)	(0.184)	(0.185)
$n = 1000$	1.000	1.000	0.998	0.977	0.957	0.942	0.953
	(0.205)	(0.204)	(0.168)	(0.137)	(0.135)	(0.135)	(0.135)
$n = 5000$	1.000	1.000	1.000	0.999	0.957	0.955	0.969
	(0.157)	(0.155)	(0.132)	(0.078)	(0.066)	(0.066)	(0.067)
				$\tau = 0.5$			
$n = 100$	0.976	0.977	0.962	0.950	0.945	0.944	0.942
	(0.408)	(0.401)	(0.386)	(0.381)	(0.382)	(0.383)	(0.383)
$n = 500$	0.974	0.972	0.965	0.964	0.955	0.959	0.953
	(0.185)	(0.183)	(0.175)	(0.174)	(0.174)	(0.174)	(0.174)
$n = 1000$	0.974	0.973	0.970	0.952	0.955	0.945	0.954
	(0.131)	(0.131)	(0.124)	(0.123)	(0.123)	(0.124)	(0.124)
$n = 5000$	0.976	0.973	0.974	0.944	0.948	0.956	0.956
	(0.059)	(0.058)	(0.056)	(0.056)	(0.056)	(0.056)	(0.056)

strongly active (binding) at $\theta_0 = \underset{\theta \geqslant 0}{\arg\min} Q(\theta)$, which leads to the coverage for θ_0 being quite conservative. However, the average interval lengths are not particularly wide because we are covering the pseudo-true parameter θ_0 rather than the true parameter β_0.

We also examined the coverage of the confidence set constructed using the unconstrained minimum of the objective function: $\mathcal{D}^*_{1-\alpha} = \left\{\theta : n^{2\gamma}\left(\hat{Q}_n(\theta) - \inf_{h \in \mathcal{B}_{\delta_n}} \hat{Q}_n\left(\theta + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right\}$, where $\mathcal{B}_{\delta_n} = \left\{h \in \mathbb{R}^d : \frac{\|h\|}{n^\gamma} \leqslant \delta_n\right\}$ is a shrinking neighborhood and $\delta_n \to 0$ satisfies $n^\gamma \delta_n \to \kappa$ for $\kappa = 5$. This confidence set is only valid when Assumption 7 holds, which rules out several drifting parameters when $\tau \neq 0.5$. If we have correctly specified moments and correctly

Table 2: Coverage Frequencies and Average Interval Lengths, $\kappa = 5$

β_0	$-n^{-1}$	$-n^{-1/2}$	$-n^{-1/3}$	$-n^{-1/4}$	$-n^{-1/6}$	-1
			$\tau = 0.1$			
$n = 100$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.566)	(0.674)	(0.839)	(0.990)	(1.095)	(1.091)
$n = 500$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.368)	(0.435)	(0.570)	(0.642)	(0.644)	(0.635)
$n = 1000$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.330)	(0.380)	(0.486)	(0.513)	(0.510)	(0.504)
$n = 5000$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.283)	(0.295)	(0.299)	(0.297)	(0.296)	(0.294)
			$\tau = 0.3$			
$n = 100$	0.996	1.000	1.000	1.000	1.000	1.000
	(0.467)	(0.554)	(0.708)	(0.876)	(1.048)	(1.087)
$n = 500$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.252)	(0.307)	(0.441)	(0.582)	(0.641)	(0.633)
$n = 1000$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.206)	(0.249)	(0.368)	(0.490)	(0.508)	(0.503)
$n = 5000$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.156)	(0.181)	(0.263)	(0.296)	(0.295)	(0.293)
			$\tau = 0.5$			
$n = 100$	0.980	0.998	1.000	1.000	1.000	1.000
	(0.407)	(0.452)	(0.516)	(0.544)	(0.537)	(0.515)
$n = 500$	0.969	0.998	1.000	1.000	1.000	1.000
	(0.184)	(0.204)	(0.242)	(0.240)	(0.233)	(0.226)
$n = 1000$	0.974	0.998	1.000	1.000	1.000	1.000
	(0.131)	(0.144)	(0.173)	(0.169)	(0.164)	(0.160)
$n = 5000$	0.979	0.999	1.000	1.000	1.000	1.000
	(0.059)	(0.065)	(0.077)	(0.074)	(0.072)	(0.071)

specified constraints, which means $\tau = 0.5$ and $\beta_0 \geqslant 0$, then Assumption 7 will hold. As shown in Table 3, if $\tau = 0.5$, the coverage of $\mathcal{D}^*_{1-\alpha}$ is less conservative than the coverage of $\mathcal{C}^*_{1-\alpha}$, and the average interval lengths are also shorter. The reason is that the limiting distribution of $n^{2\gamma}\left(\hat{Q}_n(\theta) - \inf_{h\in\mathcal{B}_{\delta_n}} \hat{Q}_n\left(\theta + \frac{h}{n^\gamma}\right)\right)$ is closer to the limiting distribution of $-\inf_{h\in\mathbb{R}^d} \hat{A}_n(h)$. However, when $\tau \neq 0.5$, the coverage of the confidence set is far below the nominal level for $\beta_0 = 0$ and also for several of the drifting values of β_0. The faster the rate at which β_0 drifts towards zero, the more severe the undercoverage, and furthermore, the undercoverage worsens with larger values of n.

Table 3: Coverage Frequencies and Average Interval Lengths using Unconstrained Objective

β_0	0	n^{-1}	$n^{-1/2}$	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/6}$	2
				$\tau = 0.1$			
$n = 100$	0.783	0.805	0.943	0.953	0.941	0.947	0.938
	(0.379)	(0.377)	(0.371)	(0.365)	(0.368)	(0.365)	(0.364)
$n = 500$	0.256	0.282	0.600	0.953	0.953	0.943	0.959
	(0.201)	(0.200)	(0.198)	(0.194)	(0.190)	(0.190)	(0.190)
$n = 1000$	0.058	0.057	0.232	0.854	0.953	0.935	0.945
	(0.153)	(0.153)	(0.151)	(0.148)	(0.144)	(0.144)	(0.144)
$n = 5000$	0.000	0.000	0.000	0.025	0.849	0.950	0.950
	(0.084)	(0.084)	(0.083)	(0.080)	(0.078)	(0.077)	(0.077)
				$\tau = 0.3$			
$n = 100$	0.898	0.920	0.947	0.948	0.945	0.946	0.942
	(0.354)	(0.352)	(0.347)	(0.346)	(0.347)	(0.346)	(0.345)
$n = 500$	0.699	0.724	0.928	0.957	0.951	0.943	0.961
	(0.177)	(0.176)	(0.174)	(0.172)	(0.171)	(0.171)	(0.173)
$n = 1000$	0.493	0.509	0.824	0.953	0.950	0.937	0.949
	(0.130)	(0.130)	(0.128)	(0.126)	(0.126)	(0.126)	(0.126)
$n = 5000$	0.013	0.011	0.071	0.908	0.947	0.944	0.957
	(0.065)	(0.065)	(0.064)	(0.062)	(0.062)	(0.061)	(0.062)
				$\tau = 0.5$			
$n = 100$	0.954	0.949	0.947	0.952	0.942	0.947	0.945
	(0.342)	(0.341)	(0.342)	(0.341)	(0.342)	(0.340)	(0.342)
$n = 500$	0.946	0.942	0.944	0.963	0.958	0.948	0.959
	(0.166)	(0.166)	(0.166)	(0.166)	(0.166)	(0.165)	(0.166)
$n = 1000$	0.952	0.942	0.948	0.945	0.952	0.945	0.959
	(0.120)	(0.120)	(0.119)	(0.119)	(0.119)	(0.120)	(0.119)
$n = 5000$	0.956	0.952	0.948	0.945	0.954	0.957	0.954
	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)

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

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

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

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

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

The bootstrapped weighting matrix is computed using the multinomial bootstrap:

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

where $\hat{\theta}_1^{C*}$ is the constrained bootstrap estimate in equation 4 using $W_1 = I$ and $\eta_n = n^{1/4}$, $\hat{f}_y^*\left(\hat{\theta}_1^{C*}\right) = \frac{1}{nh}\sum_{i=1}^n K_h\left(y_i^* - \hat{\theta}_1^{C*}\right)$, and $\hat{F}_y^*\left(\hat{\theta}_1^{C*}\right) = \frac{1}{n}\sum_{i=1}^n 1\left(y_i^* \leqslant \hat{\theta}_1^{C*}\right)$. 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 $\theta_0 = \underset{\theta \geqslant 0}{\arg\min} \frac{1}{2}\pi(\theta)' W\left(\theta_1^C\right)\pi(\theta)$ where $W\left(\theta_1^C\right) = \begin{bmatrix} F_y\left(\theta_1^C\right) - F_y\left(\theta_1^C\right)^2 & -f_y\left(\theta_1^C\right) \\ -f_y\left(\theta_1^C\right) & 1 \end{bmatrix}^{-1}$ and $\theta_1^C = \underset{\theta \geqslant 0}{\arg\min} \frac{1}{2}\pi(\theta)'\pi(\theta)$. Our uniformly asymptotically valid nominal $1-\alpha$ confidence set is $\mathcal{C}_{1-\alpha}^* = \left\{\theta : n^{2\gamma}\left(\hat{Q}_n(\theta) - \inf_{h \in \mathcal{C}_{\delta_n}^\theta} \hat{Q}_n\left(\theta + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}_{1-\alpha}^*\right\}$, where $\hat{Q}_n(\theta) = \frac{1}{2}\hat{\pi}_n(\theta)' W_n\left(\hat{\theta}_1^C\right)\hat{\pi}_n(\theta)$, $\mathcal{C}_{\delta_n}^\theta = \left\{h \in n^\gamma(C - \theta) : \frac{\|h\|}{n^\gamma} \leqslant \delta_n\right\}$, $\delta_n \to 0$ satisfies $n^\gamma \delta_n \to \kappa$, and $\hat{c}_{1-\alpha}^*$ is the $1-\alpha$ conditional quantile of $-\inf_{h \in \mathbb{R}^d} \hat{B}_n(h)$ for $\hat{B}_n(h)$ given in equation 3.

Tables 4 and 5 show the empirical coverage frequencies and average interval lengths (in parentheses) of nominal 95% confidence intervals when $\kappa = 5$. We consider a range of different values of $\beta_0 \in \pm\left\{0, n^{-1}, n^{-1/2}, n^{-1/3}, n^{-1/4}, n^{-1/6}, 1\right\}$, 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 itera-

Table 4: Coverage Frequencies and Average Interval Lengths, $\kappa = 5$

β_0	0	n^{-1}	$n^{-1/2}$	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/6}$	1
				$\tau = 0.1$			
$n = 100$	0.987	0.983	0.985	0.986	0.975	0.961	0.977
	(1.462)	(1.490)	(1.435)	(1.460)	(1.533)	(1.498)	(1.513)
$n = 500$	0.986	0.982	0.986	0.972	0.970	0.963	0.957
	(0.901)	(0.888)	(0.864)	(0.871)	(0.861)	(0.903)	(0.881)
$n = 1000$	0.977	0.980	0.983	0.981	0.969	0.960	0.959
	(0.705)	(0.712)	(0.681)	(0.685)	(0.687)	(0.698)	(0.704)
$n = 5000$	0.983	0.979	0.986	0.975	0.964	0.961	0.956
	(0.401)	(0.400)	(0.396)	(0.379)	(0.384)	(0.383)	(0.397)
				$\tau = 0.3$			
$n = 100$	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)
				$\tau = 0.5$			
$n = 100$	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)

tions and $R = 2000$ Monte Carlo simulations. For the positive values of β_0, the coverage is close to the nominal level for sufficiently large values of n and is less conservative when β_0 is further away from the boundary of the constraint set. For the negative values of β_0, the non-negativity constraint is misspecified and therefore strongly active at $\theta_0 = \arg\min_{\theta \geqslant 0} Q(\theta)$. This misspecification causes the coverage for θ_0 to be quite conservative. However, the average interval lengths are not particularly wide because we are covering the pseudo-true parameter θ_0 rather than the true parameter β_0. Additional Monte Carlo simulations for the cases of $\kappa = \infty$ and $\kappa = 1$ are in the Appendix Section 6.3.

Table 5: Coverage Frequencies and Average Interval Lengths, $\kappa = 5$

β_0	$-n^{-1}$	$-n^{-1/2}$	$-n^{-1/3}$	$-n^{-1/4}$	$-n^{-1/6}$	-1
			$\tau = 0.1$			
$n = 100$	0.984	0.989	0.999	0.999	1.000	1.000
	(1.246)	(1.300)	(1.393)	(1.447)	(1.470)	(1.524)
$n = 500$	0.984	0.992	0.996	0.996	0.999	1.000
	(0.818)	(0.841)	(0.887)	(0.895)	(0.846)	(0.760)
$n = 1000$	0.984	0.986	0.996	0.999	1.000	1.000
	(0.682)	(0.682)	(0.720)	(0.720)	(0.660)	(0.572)
$n = 5000$	0.980	0.991	0.993	0.999	1.000	1.000
	(0.401)	(0.406)	(0.415)	(0.407)	(0.366)	(0.315)
			$\tau = 0.3$			
$n = 100$	0.979	0.993	0.999	0.999	1.000	1.000
	(0.823)	(0.906)	(1.014)	(1.140)	(1.246)	(1.322)
$n = 500$	0.981	0.991	0.999	1.000	1.000	1.000
	(0.498)	(0.526)	(0.601)	(0.689)	(0.709)	(0.677)
$n = 1000$	0.980	0.989	0.998	1.000	1.000	1.000
	(0.402)	(0.420)	(0.486)	(0.562)	(0.564)	(0.535)
$n = 5000$	0.982	0.992	0.997	1.000	1.000	1.000
	(0.252)	(0.259)	(0.298)	(0.341)	(0.325)	(0.304)
			$\tau = 0.5$			
$n = 100$	0.970	0.996	0.999	1.000	1.000	1.000
	(0.419)	(0.467)	(0.557)	(0.655)	(0.783)	(1.437)
$n = 500$	0.964	0.998	1.000	1.000	1.000	1.000
	(0.191)	(0.211)	(0.280)	(0.355)	(0.495)	(1.174)
$n = 1000$	0.967	0.997	1.000	1.000	1.000	1.000
	(0.136)	(0.150)	(0.208)	(0.279)	(0.413)	(1.132)
$n = 5000$	0.977	0.999	1.000	1.000	1.000	1.000
	(0.062)	(0.068)	(0.106)	(0.163)	(0.284)	(1.071)

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 $\left\{P^{(n)} \in \mathcal{P} : n \geqslant 1\right\}$ 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 θ_n follows from Assumption 1 and constraining θ 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, $n^{2\gamma}\left(P_n - P^{(n)}\right)\left(\pi\left(\cdot, \theta_n + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta_n\right)\right)$ converges in finite dimensional distribution to a mean zero Gaussian process $\mathcal{Z}_{0,\rho}(h)$ 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, $n^{2\gamma} P^{(n)}\left(\pi\left(\cdot, \theta_n + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta_n\right)\right) = h' n^\gamma l\left(\theta_n\right) + \frac{1}{2} h' H_0 h + o(1) = \frac{1}{2} h' H_0 h + o(1)$ since we assumed in Assumption 4 that $l\left(\theta_n\right) = 0$. Therefore,

$$\begin{aligned}
&n^{2\gamma}\hat{\pi}_n\left(\theta_n + \frac{h}{n^\gamma}\right) - n^{2\gamma}\hat{\pi}_n\left(\theta_n\right) \\
&= n^{2\gamma}\left(P_n - P^{(n)}\right)\left(\pi\left(\cdot, \theta_n + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta_n\right)\right) + \frac{1}{2} h' H_0 h + o_{P^{(n)}}(1) \\
&\rightsquigarrow \mathcal{Z}_{0,\rho}(h) + \frac{1}{2} h' H_0 h,
\end{aligned}$$

as a process indexed by h in the space of locally bounded functions $\mathbf{B}_{\text{loc}}\left(\mathbb{R}^d\right)$ 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 $n^{2\gamma}\left(P_n^* - P_n\right)\left(\pi\left(\cdot, \theta_n + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta_n\right)\right)$ is consistent for the same limiting process as $n^{2\gamma}\left(P_n - P^{(n)}\right)\left(\pi\left(\cdot, \theta_n + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta_n\right)\right)$:

$$n^{2\gamma}\left(P_n^* - P_n\right)\left(\pi\left(\cdot, \theta_n + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta_n\right)\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{Z}_{0,\rho}\left(h\right).$$

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

$$n^{2\gamma} \sup_{d \in \mathscr{D}(n)} \left|\left(P_n - P^{(n)}\right) d\right| = o_{P^{(n)}}(1),$$

where $\mathscr{D}(n) = \{d\left(\cdot, \theta_n, h_1, h_2\right) = g\left(\cdot; \theta_n + n^{-\gamma} h_1\right) - g\left(\cdot; \theta_n + n^{-\gamma} h_2\right)$ such that $\max(\|h_1\|, \|h_2\|) \leqslant M$ and $\|h_1 - h_2\| \leqslant \epsilon_n\}$. Note that $\mathscr{D}(n)$ has envelope function $D_n = 2G_{R(n)}$ where $R(n) = Mn^{-\gamma}$.

The maximal inequality states that

$$n^{2\gamma} P^{(n)} \sup_{d \in \mathscr{D}(n)} \left|\left(P_n - P^{(n)}\right) d\right| \leqslant P^{(n)} \sqrt{n^{2\gamma\rho} P_n D_n^2}\, J\left(\frac{n^{2\gamma\rho} \sup_{d \in \mathscr{D}(n)} P_n d^2}{n^{2\gamma\rho} P_n D_n^2}\right).$$

Note that $P^{(n)} n^{2\gamma\rho} P_n D_n^2 = O\left(n^{2\gamma\rho}\left(n^{-\gamma}\right)^{2\rho}\right) = O(1)$. Next, for each $K > 0$, $P^{(n)} \sup_{d \in \mathscr{D}(n)} P_n d^2 \leqslant P^{(n)} P_n \sup_{d \in \mathscr{D}(n)} d^2 1\left\{D_n > K\right\} + K P^{(n)} \sup_{d \in \mathscr{D}(n)} P_n |d| \leqslant P^{(n)} P_n D_n^2 1\left\{D_n > K\right\} + K \sup_{d \in \mathscr{D}(n)} P^{(n)} |d|$ $+ K P^{(n)} \sup_{d \in \mathscr{D}(n)} \left|P_n |d| - P^{(n)} |d|\right|$. For the first term, for large enough K, there exists some $\eta > 0$ such that $P^{(n)} P_n D_n^2 1\left\{D_n > K\right\} < \eta n^{-2\gamma\rho}$. For the second term, $K \sup_{d \in \mathscr{D}(n)} P^{(n)} |d| =$

$O\left(n^{-2\gamma\rho}\epsilon_n\right) = o\left(n^{-2\gamma\rho}\right)$. For the third term, if $\rho < 1$,

$$
\begin{aligned}
KP^{(n)} \sup_{d\in\mathscr{D}(n)} \left|P_n|d| - P^{(n)}|d|\right| &\leqslant Kn^{-1/2}J(1)\sqrt{P^{(n)}D_n^2} \\
&= O\left(n^{-(\gamma\rho+1/2)}\right) = O\left(n^{-2\gamma}\right) = o\left(n^{-2\gamma\rho}\right)
\end{aligned}
$$

In the case where $\rho = 1$ and $\gamma = 1/2$, because we assumed in Assumption 2 that $\pi(\cdot,\theta)$ is Lipschitz in θ 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 $KP^{(n)}\sup_{\mathscr{D}(n)}\left|P_n|d_j| - P^{(n)}|d_j|\right| < Kn^{-\frac{1}{2}}J(1)\sqrt{P^{(n)}D_n^2} = O(n^{-1}\epsilon_n) = o(n^{-1})$. Therefore, $P^{(n)}n\sup_{\mathscr{D}(n)}P_n d_j^2 = o(1)$.

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

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

Since $\lim_{\lambda\to\infty}\lim_{n\to\infty}\sup\sup_{t\geqslant\lambda}t^2\left\{\sup_{h\in\mathcal{H},\|\theta-\theta_n\|\leqslant\epsilon_n}\left|\frac{m_n(\cdot,\theta,h)-m_n(\cdot,\theta_n,h)}{1+n^\gamma\|\theta-\theta_n\|}\right| > t\right\} = 0$ 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

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

Therefore, since $\hat{H} \xrightarrow{p} H$,

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

$$\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{Z}_{0,\rho}(h) + \frac{1}{2}h'H_0h \equiv \mathbb{H}_0(h)$$

Then the continuous mapping results in Lemma 10.11 of Kosorok (2007) imply $-\inf_{h\in\mathbb{R}^d}\hat{\mathbb{H}}_n(h) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} -\inf_{h\in\mathbb{R}^d}\mathbb{H}_0(h)$. Also by the continuous mapping theorem,

$$n^{2\gamma}\left(\hat{\pi}_n(\theta_n) - \inf_{h\in\mathcal{B}_{\delta_n}}\hat{\pi}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right) \rightsquigarrow -\inf_{h\in\{h\in\mathbb{R}^d:\|h\|\leqslant\kappa\}}\mathbb{H}_0(h)$$

Since $J\left(\cdot, P^{(n)}\right)$ 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 $\epsilon > 0$ small enough. Since $J^*\left(\cdot, P^{(n)}\right)$ is also equicontinuous at $J_n^{-1}\left(1-\alpha, P^{(n)}\right)$, we have for any $P^{(n)}$ and ϵ small enough, $J^*_{\alpha_n}\left(x_n, P^{(n)}\right) - J^*\left(x_n, P^{(n)}\right) = o_{P^{(n)}}(1)$.

Note that $-\inf_{h\in\{h\in\mathbb{R}^d:\|h\|\leqslant\kappa\}}\mathbb{H}_0(h) \leqslant -\inf_{h\in\mathbb{R}^d}\mathbb{H}_0(h)$ 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 $P^{(n)}\left(J^*_{\alpha_n}\left(x_n, P^{(n)}\right) - J_n\left(x_n, P^{(n)}\right) > \epsilon\right) \leqslant \delta$. If $J^*_{\alpha_n}\left(x_n, P^{(n)}\right) - J_n\left(x_n, P^{(n)}\right) \leqslant \epsilon$, then $J_n^{-1}\left(1-\alpha-\epsilon, P^{(n)}\right) \leqslant J^{*-1}_{\alpha_n}\left(1-\alpha, P^{(n)}\right)$. Then, using arguments similar to those in Lemma A.1 (vi) of Romano and Shaikh (2012), for all $\epsilon > 0$ and n large enough,

$$\begin{aligned}
&P^{(n)}\left(n^{2\gamma}\left(\hat{\pi}_n(\theta_n) - \inf_{h\in\mathcal{B}_{\delta_n}}\hat{\pi}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right) \leqslant J^{*-1}_{\alpha_n}\left(1-\alpha, P^{(n)}\right)\right) \\
&\geqslant P^{(n)}\left(n^{2\gamma}\left(\hat{\pi}_n(\theta_n) - \inf_{h\in\mathcal{B}_{\delta_n}}\hat{\pi}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right) \leqslant J^{*-1}_{\alpha_n}\left(1-\alpha, P^{(n)}\right) \cap J^*_{\alpha_n}\left(x_n, P^{(n)}\right) - J_n\left(x_n, P^{(n)}\right) \leqslant \epsilon\right) \\
&\geqslant P^{(n)}\left(n^{2\gamma}\left(\hat{\pi}_n(\theta_n) - \inf_{h\in\mathcal{B}_{\delta_n}}\hat{\pi}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right) \leqslant J^{-1}_{n}\left(1-\alpha-\epsilon, P^{(n)}\right) \cap J^*_{\alpha_n}\left(x_n, P^{(n)}\right) - J_n\left(x_n, P^{(n)}\right) \leqslant \epsilon\right) \\
&\geqslant P^{(n)}\left(n^{2\gamma}\left(\hat{\pi}_n(\theta_n) - \inf_{h\in\mathcal{B}_{\delta_n}}\hat{\pi}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right) \leqslant J^{-1}_{n}\left(1-\alpha-\epsilon, P^{(n)}\right)\right) \\
&- P^{(n)}\left(J^*_{\alpha_n}\left(x_n, P^{(n)}\right) - J_n\left(x_n, P^{(n)}\right) > \epsilon\right) \\
&\geqslant 1-\alpha-\epsilon-\delta
\end{aligned}$$

Since ϵ and δ can be arbitrarily small, $\liminf_{n\to\infty} P^{(n)}\left(n^{2\gamma}\left(\hat{\pi}_n(\theta_n) - \inf_{h\in\mathcal{B}_{\delta_n}}\hat{\pi}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right) \geqslant$

$1-\alpha$. For $\rho = \liminf_{n\to\infty} \inf_{P\in\mathcal{P}} P\left(n^{2\gamma}\left(\hat{\pi}_n\left(\theta_n\right) - \inf_{h\in\mathcal{B}_{\delta_n}} \hat{\pi}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right)$, we can find a sequence $\left\{P^{(n)} \in \mathcal{P}\right\}$ such that $\rho = \liminf_{n\to\infty} P^{(n)}\left(n^{2\gamma}\left(\hat{\pi}_n\left(\theta_n\right) - \inf_{h\in\mathcal{B}_{\delta_n}} \hat{\pi}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right)$. Find a subsequence n_k of n for which θ_n converges, with its limit denoted θ. The same arguments as above applied to such a subsequence imply

$\liminf_{n_k\to\infty} P^{(n_k)}\left(n_k^{2\gamma}\left(\hat{\pi}_{n_k}\left(\theta_{n_k}\right) - \inf_{h\in\mathcal{B}_{\delta_{n_k}}} \hat{\pi}_{n_k}\left(\theta_{n_k} + \frac{h}{n_k^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right) \geqslant 1-\alpha$. Since $\left\{P^{(n_k)}, \theta_{n_k}\right\}$ is a subsequence of $\left\{P^{(n)}, \theta_n\right\}$, $\rho = \liminf_{n_k\to\infty} P^{(n_k)}\left(n_k^{2\gamma}\left(\hat{\pi}_{n_k}\left(\theta_{n_k}\right) - \inf_{h\in\mathcal{B}_{\delta_{n_k}}} \hat{\pi}_{n_k}\left(\theta_{n_k} + \frac{h}{n_k^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right) \geqslant 1-\alpha$.

■

6.1.2 Proof of Theorem 2

Recall that $n^{2\gamma} P\left(\pi\left(\cdot, \theta_0 + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta_0\right)\right) = h' n^\gamma l\left(\theta_0\right) + \frac{1}{2} h' H_0 h + o(1)$. Additionally, Chernoff regularity implies that $+\infty 1\left(h \notin n^\gamma\left(C - \theta_0\right)\right) \xrightarrow{e} +\infty 1\left(h \notin T_C\left(\theta_0\right)\right)$, where $T_C\left(\theta_0\right) \equiv \limsup_{\tau\downarrow 0} \frac{C-\theta_0}{\tau}$. Note that when θ_0 is the constrained minimizer, $h' l\left(\theta_0\right) \geqslant 0$ for all $h \in T_C\left(\theta_0\right)$. Otherwise, there would exist some descent direction $h \in T_C\left(\theta_0\right)$ that reduces the value of the objective function and θ_0 would not be the constrained minimizer anymore. A proof of this result is on pages 325-326 of Nocedal and Wright (2006). Then, for any $c \in \mathbb{R}$,

$$
\begin{aligned}
&\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(n^{2\gamma}\left(\hat{\pi}_n\left(\theta_0\right) - \inf_{h\in\mathcal{C}^{\theta_0}_{\delta_n}} \hat{\pi}_n\left(\theta_0 + \frac{h}{n^\gamma}\right)\right) > c\right) \\
=&\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(-n^{2\gamma} \inf_{h\in\mathcal{C}^{\theta_0}_{\delta_n}} \left\{\left(P_n - P\right)\left(\pi\left(\cdot, \theta_0 + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta_0\right)\right) + P\left(\pi\left(\cdot, \theta_0 + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta_0\right)\right)\right\} > c\right) \\
\leqslant&\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(-\inf_{h\in\mathcal{C}^{\theta_0}_{\delta_n}} \left\{n^{2\gamma}\left(P_n - P\right)\left(\pi\left(\cdot, \theta_0 + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta_0\right)\right) + h' n^\gamma l\left(\theta_0\right) + \frac{1}{2} h' H_0 h\right\} > c\right) \\
\leqslant&\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(-\inf_{\left\{h\in T_C\left(\theta_0\right): \|h\|\leqslant\kappa\right\}} \left\{n^{2\gamma}\left(P_n - P\right)\left(\pi\left(\cdot, \theta_0 + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta_0\right)\right) + \frac{1}{2} h' H_0 h\right\} > c\right)
\end{aligned}
$$

Therefore, $n^{2\gamma}\left(\hat{\pi}_n\left(\theta_0\right) - \inf_{h\in\mathcal{C}^{\theta_0}_{\delta_n}} \hat{\pi}_n\left(\theta_0 + \frac{h}{n^\gamma}\right)\right)$ is asymptotically first order stochastically dominated by $-\inf_{\left\{h\in T_C\left(\theta_0\right): \|h\|\leqslant\kappa\right\}} \left\{n^{2\gamma}\left(P_n - P\right)\left(\pi\left(\cdot, \theta_0 + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta_0\right)\right) + \frac{1}{2} h' H_0 h\right\}$. Because the test statistic $n^{2\gamma}\left(\hat{\pi}_n\left(\theta_0\right)-\inf_{h\in\mathcal{C}_{\delta_n}^{\theta_0}}\hat{\pi}_n\left(\theta_0+\frac{h}{n^\gamma}\right)\right)$ may not have a well-defined limiting distribution when $h'n^\gamma l\left(\theta_0\right)\neq 0$, we will instead define $J_n\left(\cdot,P\right)$ as the CDF of $-\inf_{\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}}\left\{n^{2\gamma}\left(P_n-P\right)\left(\pi\left(\cdot,\theta_0+\frac{h}{n^\gamma}\right)-\pi\left(\cdot,\theta_0\right)\right)+\frac{1}{2}h'H_0h\right\}$ under P, and $J\left(\cdot,P\right)$ as the CDF of its limiting distribution $-\inf_{\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}}\mathbb{H}_0\left(h\right)$ under P. The definition of $J_{\alpha_n}^*\left(\cdot,P\right)$ remains the same as in Theorem 1, denoting the conditional CDF of $-\inf_{h\in\mathbb{R}^d}\hat{\mathbb{H}}_n\left(h\right)$ under P, and $J^*\left(\cdot,P\right)$ still denotes the CDF of its limiting distribution $-\inf_{h\in\mathbb{R}^d}\mathbb{H}_0\left(h\right)$ under P. Note that $-\inf_{\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}}\mathbb{H}_0\left(h\right)$ is stochastically dominated by $-\inf_{h\in\mathbb{R}^d}\mathbb{H}_0\left(h\right)$. The result follows from modifying the proof of Theorem 1 to incorporate the new definitions of $J_n\left(\cdot,P\right)$ and $J\left(\cdot,P\right)$, and replacing $n^{2\gamma}\left(\hat{\pi}_n\left(\theta_n\right)-\inf_{h\in\mathcal{B}_{\delta_n}}\hat{\pi}_n\left(\theta_n+\frac{h}{n^\gamma}\right)\right)$ by $n^{2\gamma}\left(\hat{\pi}_n\left(\theta_n\right)-\inf_{h\in\mathcal{C}_{\delta_n}^{\theta_n}}\hat{\pi}_n\left(\theta_n+\frac{h}{n^\gamma}\right)\right)$. ∎

6.1.3 Proof of Theorem 3

Consider any sequence $\left\{P^{(n)}\in\mathcal{P}:n\geqslant 1\right\}$ that determines $\theta_n=\theta\left(P^{(n)}\right)$ and the laws of all random variables. Consistency of $\hat{\theta}_n$ for θ_n follows from constraining θ 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 $\sup_{\theta\in C}\left|\hat{Q}_n(\theta)-Q(\theta)\right|=o_P(1)$.

Define $\hat{\mathcal{G}}_n\left(\theta\right)=\sqrt{n}\left(P_n-P^{(n)}\right)g\left(\cdot,\theta\right)$, $\hat{g}\left(\theta\right)=P_ng\left(\cdot,\theta\right)$, and $g\left(\theta\right)=P^{(n)}g\left(\cdot,\theta\right)$. Then $\hat{\pi}_n\left(\theta\right)=g\left(\theta\right)+\hat{\pi}_n\left(\theta_n\right)+\hat{\eta}_n\left(\theta\right)$, where $\hat{\eta}_n\left(\theta\right)=\frac{1}{\sqrt{n}}\hat{\mathcal{G}}_n\left(\theta\right)$. Recall that $\hat{Q}_n\left(\theta\right)=\frac{1}{2}\hat{\pi}_n\left(\theta\right)'W\hat{\pi}_n\left(\theta\right)$. Write $\hat{Q}_n\left(\theta\right)-\hat{Q}_n\left(\theta_n\right)=Q_1\left(\theta\right)+\hat{Q}_2\left(\theta\right)+\hat{Q}_3\left(\theta\right)$, where

$$
\begin{aligned}
Q_1\left(\theta\right)=&\frac{1}{2}g\left(\theta\right)'Wg\left(\theta\right)+g\left(\theta\right)'W\pi\left(\theta_n\right),\quad \hat{Q}_3\left(\theta\right)=\pi\left(\theta_n\right)'W\hat{\eta}_n\left(\theta\right)\\
\hat{Q}_2\left(\theta\right)=&\frac{1}{2}\hat{\eta}_n\left(\theta\right)'W\hat{\eta}_n\left(\theta\right)+g\left(\theta\right)'W\left(\hat{\pi}_n\left(\theta_n\right)-\pi\left(\theta_n\right)\right)+g\left(\theta\right)'W\hat{\eta}_n\left(\theta\right)+\left(\hat{\pi}_n\left(\theta_n\right)-\pi\left(\theta_n\right)\right)'W\hat{\eta}_n\left(\theta\right).
\end{aligned}
\tag{6}
$$

We showed in Theorems 1 and 5 of Hong and Li (2023) that under Assumptions 5 - 6, $n^{2\gamma}\left(\hat{Q}_n(\theta_n+n^{-\gamma}h)-\hat{Q}_n(\theta_n)\right)=n^{2\gamma}Q_1(\theta_n+n^{-\gamma}h)+n^{2\gamma}\left(\hat{Q}_2(\theta_n+n^{-\gamma}h)+\hat{Q}_3(\theta_n+n^{-\gamma}h)\right)$ where $n^{2\gamma}\left(\hat{Q}_2(\theta_n+n^{-\gamma}h)+\hat{Q}_3(\theta_n+n^{-\gamma}h)\right)\rightsquigarrow\pi(\theta_n)' W\mathcal{Z}_{0,1/2}(h)$ if $\gamma=1/3$ and $n^{2\gamma}\left(\hat{Q}_2(\theta_n+n^{-\gamma}h)+\hat{Q}_3(\theta_n+n^{-\gamma}h)\right)\rightsquigarrow\pi(\theta_n)' W\mathcal{Z}_{0,1}(h)+h'G'WU_0$ if $\gamma=1/2$. Additionally,

$$
\begin{aligned}
n^{2\gamma}Q_1\left(\theta_n+n^{-\gamma}h\right)&=n^{\gamma}h'\frac{\partial Q_1(\theta_n)}{\partial\theta}+\frac{1}{2}h'\frac{\partial^2 Q_1(\theta_n)}{\partial\theta\partial\theta'}h+o(1)\\
&=n^{\gamma}h'G'W\pi(\theta_n)+\frac{1}{2}h'\bar{H}h+o(1)\\
&=\frac{1}{2}h'\bar{H}h+o(1)
\end{aligned}
$$

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

$$
\begin{aligned}
&n^{2\gamma}\left(\hat{Q}_n\left(\theta_n+n^{-\gamma}h\right)-\hat{Q}_n(\theta_n)\right)\\
&=n^{2\gamma}\left(\hat{Q}_2\left(\theta_n+n^{-\gamma}h\right)+\hat{Q}_3\left(\theta_n+n^{-\gamma}h\right)\right)+\frac{1}{2}h'\bar{H}h+o_{P(n)}(1)\\
&\rightsquigarrow A_0(h)\equiv\begin{cases}\pi(\theta_n)' W\mathcal{Z}_{0,1/2}(h)+\frac{1}{2}h'\bar{H}h & \text{if } \gamma=1/3\\ \pi(\theta_n)' W\mathcal{Z}_{0,1}(h)+h'G'WU_0+\frac{1}{2}h'\bar{H}h & \text{if } \gamma=1/2\end{cases}
\end{aligned}
$$

as a process indexed by h in the space of locally bounded functions $\mathbf{B}_{\text{loc}}\left(\mathbb{R}^d\right)$ 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, $\hat{A}_n(h)\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}}A_0(h)$, where

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

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

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

$$n^{2\gamma}\left(\hat{Q}_n(\theta_n)-\inf_{h\in\mathcal{B}_{\delta_n}}\hat{Q}_n\left(\theta_n+\frac{h}{n^{\gamma}}\right)\right)\rightsquigarrow-\inf_{h\in\{h\in\mathbb{R}^d:\|h\|\leqslant\kappa\}}A_0(h)$$

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

Note that $-\inf_{h\in\{h\in\mathbb{R}^d:\|h\|\leqslant\kappa\}}A_0(h)\leqslant-\inf_{h\in\mathbb{R}^d}A_0(h)$ 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 $P^{(n)}\left(J_{\alpha_n}^{*}\left(x_n,P^{(n)}\right)-J_n\left(x_n,P^{(n)}\right)>\epsilon\right)\leqslant\delta$. If $J_{\alpha_n}^{*}\left(x_n,P^{(n)}\right)-J_n\left(x_n,P^{(n)}\right)\leqslant\epsilon$, then $J_n^{-1}\left(1-\alpha-\epsilon,P^{(n)}\right)\leqslant J_{\alpha_n}^{*-1}\left(1-\alpha,P^{(n)}\right)$. Then, using arguments similar to those in Lemma A.1 (vi) of Romano and Shaikh (2012), for all $\epsilon>0$ and n large enough,

$$\begin{aligned}
&P^{(n)}\left(n^{2\gamma}\left(\hat{Q}_n(\theta_n)-\inf_{h\in\mathcal{B}_{\delta_n}}\hat{Q}_n\left(\theta_n+\frac{h}{n^{\gamma}}\right)\right)\leqslant J_{\alpha_n}^{*-1}\left(1-\alpha,P^{(n)}\right)\right)\\
&\geqslant P^{(n)}\left(n^{2\gamma}\left(\hat{Q}_n(\theta_n)-\inf_{h\in\mathcal{B}_{\delta_n}}\hat{Q}_n\left(\theta_n+\frac{h}{n^{\gamma}}\right)\right)\leqslant J_{\alpha_n}^{*-1}\left(1-\alpha,P^{(n)}\right)\cap J_{\alpha_n}^{*}\left(x_n,P^{(n)}\right)-J_n\left(x_n,P^{(n)}\right)\leqslant\epsilon\right)\\
&\geqslant P^{(n)}\left(n^{2\gamma}\left(\hat{Q}_n(\theta_n)-\inf_{h\in\mathcal{B}_{\delta_n}}\hat{Q}_n\left(\theta_n+\frac{h}{n^{\gamma}}\right)\right)\leqslant J_n^{-1}\left(1-\alpha-\epsilon,P^{(n)}\right)\cap J_{\alpha_n}^{*}\left(x_n,P^{(n)}\right)-J_n\left(x_n,P^{(n)}\right)\leqslant\epsilon\right)\\
&\geqslant P^{(n)}\left(n^{2\gamma}\left(\hat{Q}_n(\theta_n)-\inf_{h\in\mathcal{B}_{\delta_n}}\hat{Q}_n\left(\theta_n+\frac{h}{n^{\gamma}}\right)\right)\leqslant J_n^{-1}\left(1-\alpha-\epsilon,P^{(n)}\right)\right)\\
&-P^{(n)}\left(J_{\alpha_n}^{*}\left(x_n,P^{(n)}\right)-J_n\left(x_n,P^{(n)}\right)>\epsilon\right)\\
&\geqslant 1-\alpha-\epsilon-\delta
\end{aligned}$$

Since ϵ and δ can be arbitrarily small, $\liminf_{n\to\infty} P^{(n)}\left(n^{2\gamma}\left(\hat{Q}_n(\theta_n) - \inf_{h\in\mathcal{B}_{\delta_n}} \hat{Q}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right) \geqslant 1-\alpha$. For $\rho = \liminf_{n\to\infty}\inf_{P\in\mathcal{P}} P\left(n^{2\gamma}\left(\hat{Q}_n(\theta_n) - \inf_{h\in\mathcal{B}_{\delta_n}} \hat{Q}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right)$, we can find a sequence $\left\{P^{(n)} \in \mathcal{P}\right\}$ such that $\rho = \liminf_{n\to\infty} P^{(n)}\left(n^{2\gamma}\left(\hat{Q}_n(\theta_n) - \inf_{h\in\mathcal{B}_{\delta_n}} \hat{Q}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right)$. Find a subsequence n_k of n for which θ_n converges, with its limit denoted θ. The same arguments as above applied to such a subsequence imply $\liminf_{n_k\to\infty} P^{(n_k)}\left(n_k^{2\gamma}\left(\hat{Q}_n(\theta_{n_k}) - \inf_{h\in\mathcal{B}_{\delta_n}} \hat{Q}_n\left(\theta_{n_k} + \frac{h}{n_k^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right) \geqslant 1-\alpha$. Since $\left\{P^{(n_k)}, \theta_{n_k}\right\}$ is a subsequence of $\left\{P^{(n)}, \theta_n\right\}$, $\rho = \liminf_{n_k\to\infty} P^{(n_k)}\left(n_k^{2\gamma}\left(\hat{Q}_n(\theta_{n_k}) - \inf_{h\in\mathcal{B}_{\delta_n}} \hat{Q}_n\left(\theta_{n_k} + \frac{h}{n_k^\gamma}\right)\right) \leqslant \hat{c}^*_{1-\alpha}\right) \geqslant 1-\alpha$.

■

6.1.4 Proof of Theorem 4

Recall $n^{2\gamma} Q_1(\theta_0 + n^{-\gamma} h) = n^\gamma h' G' W \pi(\theta_0) + \frac{1}{2} h' \bar{H} h + o(1)$. Additionally, Chernoff regularity implies that $+\infty 1(h \notin n^\gamma (C - \theta_0)) \xrightarrow{e} +\infty 1(h \notin T_C(\theta_0))$, where $T_C(\theta_0) \equiv \limsup_{\tau\downarrow 0} \frac{C-\theta_0}{\tau}$. When θ_0 is the constrained minimizer, $h'G'W\pi(\theta_0) \geqslant 0$ 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 θ_0 would not be the constrained minimizer anymore. Then, for any $c \in \mathbb{R}$,

$$
\begin{aligned}
&\limsup_{n\to\infty}\sup_{P\in\mathcal{P}} P\left(n^{2\gamma}\left(\hat{Q}_n(\theta_0) - \inf_{h\in\mathcal{C}^{\theta_0}_{\delta_n}} \hat{Q}_n\left(\theta_0 + \frac{h}{n^\gamma}\right)\right) > c\right) \\
&= \limsup_{n\to\infty}\sup_{P\in\mathcal{P}} P\left(-n^{2\gamma} \inf_{h\in\mathcal{C}^{\theta_0}_{\delta_n}} \left\{\hat{Q}_2\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_3\left(\theta_0 + \frac{h}{n^\gamma}\right) + Q_1\left(\theta_0 + \frac{h}{n^\gamma}\right)\right\} > c\right) \\
&\leqslant \limsup_{n\to\infty}\sup_{P\in\mathcal{P}} P\left(-\inf_{h\in\mathcal{C}^{\theta_0}_{\delta_n}} \left\{n^{2\gamma}\left(\hat{Q}_2\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_3\left(\theta_0 + \frac{h}{n^\gamma}\right)\right) + n^\gamma h' G' W \pi(\theta_0) + \frac{1}{2} h' \bar{H} h\right\} > c\right) \\
&\leqslant \limsup_{n\to\infty}\sup_{P\in\mathcal{P}} P\left(-\inf_{h\in\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}} \left\{n^{2\gamma}\left(\hat{Q}_2\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_3\left(\theta_0 + \frac{h}{n^\gamma}\right)\right) + \frac{1}{2} h' \bar{H} h\right\} > c\right)
\end{aligned}
$$

Therefore $n^{2\gamma}\left(\hat{Q}_n(\theta_0) - \inf_{h\in\mathcal{C}^{\theta_0}_{\delta_n}} \hat{Q}_n\left(\theta_0 + \frac{h}{n^\gamma}\right)\right)$ is asymptotically first order stochastically dominated by $-\inf_{h\in\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}} \left\{n^{2\gamma}\left(\hat{Q}_2\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_3\left(\theta_0 + \frac{h}{n^\gamma}\right)\right) + \frac{1}{2} h' \bar{H} h\right\}$. Because the test statistic $n^{2\gamma}\left(\hat{Q}_n(\theta_0)-\inf_{h\in\mathcal{C}_{\delta_n}^{\theta_0}}\hat{Q}_n\left(\theta_0+\frac{h}{n^\gamma}\right)\right)$ may 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 $-\inf_{h\in\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}}\left\{n^{2\gamma}\left(\hat{Q}_2\left(\theta_0+\frac{h}{n^\gamma}\right)+\hat{Q}_3\left(\theta_0+\frac{h}{n^\gamma}\right)\right)+\frac{1}{2}h'\bar{H}h\right\}$ under P, and $J(\cdot,P)$ as the CDF of its limiting distribution $-\inf_{\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}}A_0(h)$ under P. The definition of $J^*_{\alpha_n}(\cdot,P)$ remains the same as in Theorem 3, denoting the conditional CDF of $-\inf_{h\in\mathbb{R}^d}\hat{A}_n(h)$ under P, and $J^*(\cdot,P)$ still denotes the CDF of its limiting distribution $-\inf_{h\in\mathbb{R}^d}A_0(h)$ under P. Note that $-\inf_{\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}}A_0(h)$ is stochastically dominated by $-\inf_{h\in\mathbb{R}^d}A_0(h)$. The result follows from modifying the proof of Theorem 3 to incorporate the new definitions of $J_n(\cdot,P)$ and $J(\cdot,P)$, and replacing $n^{2\gamma}\left(\hat{Q}_n(\theta_n)-\inf_{h\in\mathcal{B}_{\delta_n}}\hat{Q}_n\left(\theta_n+\frac{h}{n^\gamma}\right)\right)$ by $n^{2\gamma}\left(\hat{Q}_n(\theta_n)-\inf_{h\in\mathcal{C}_{\delta_n}^{\theta_n}}\hat{Q}_n\left(\theta_n+\frac{h}{n^\gamma}\right)\right)$. ■

6.1.5 Proof of Theorem 5

Consistency of $\hat{\theta}_n$ for θ_0 follows from constraining θ 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) and $W_n-W=o_P(1)$ imply $\sup_{\theta\in C}\left|\hat{Q}_n(\theta)-Q(\theta)\right|=o_P(1)$.

Under Assumptions 3, 5-6 and 8-9, we showed in Theorems 4 and 8 of Hong and Li (2023) that $\hat{B}_n(h)\underset{\mathbb{W}}{\overset{\mathbb{P}}{\leadsto}}B_0(h)$, where

$$\begin{aligned}\hat{B}_n(h)=&\,n^{2\gamma}\hat{\pi}_n\left(\hat{\theta}_n\right)'W_n\left(P_n^*-P_n\right)\left(\pi\left(\cdot,\hat{\theta}_n+\frac{h}{n^\gamma}\right)-\pi\left(\cdot,\hat{\theta}_n\right)\right)\\&+\frac{1}{2}h'\left(\hat{G}'W_n\hat{G}+\sum_{j=1}^m\sum_{k=1}^m W_{n,jk}\hat{\pi}_{nk}\left(\hat{\theta}_n\right)\hat{H}_j\right)h\\&+n^\gamma h'\hat{G}'W_n\left(P_n^*-P_n\right)\pi\left(\cdot,\hat{\theta}_n\right)\\&+n^\gamma h'\hat{G}'\left(W_n^*-W_n\right)\hat{\pi}_n\left(\hat{\theta}_n\right).\end{aligned}\tag{7}$$

If $W_n - W = O_P(n^{-\gamma})$ and $W_n^* - W_n = O_P^*(n^{-\gamma})$, then $B_0(h) = \pi(\theta_0)' W \mathcal{Z}_{0,1}(h) + h'G'WU_0 + \frac{1}{2}h'\bar{H}h + h'G'\mathcal{W}_0$ for $\rho = 1$ and $B_0(h) = \pi(\theta_0)' W \mathcal{Z}_{0,1/2}(h) + \frac{1}{2}h'\bar{H}h + h'G'\mathcal{W}_0$ when $\rho = 1/2$. If $W_n - W = o_P(n^{-\gamma})$ and $W_n^* - W_n = o_P^*(n^{-\gamma})$, then $B_0(h) = \pi(\theta_0)' W \mathcal{Z}_{0,1}(h) + h'G'WU_0 + \frac{1}{2}h'\bar{H}h$ for $\rho = 1$ and $B_0(h) = \pi(\theta_0)' W \mathcal{Z}_{0,1/2}(h) + \frac{1}{2}h'\bar{H}h$ when $\rho = 1/2$.

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

$$\begin{aligned}
& n^{2\gamma}\left(\hat{Q}_n\left(\theta_0 + \frac{h}{n^\gamma}\right) - \hat{Q}_n(\theta_0)\right) \\
&= n^{2\gamma}\left(Q_1\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_2\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_3\left(\theta_0 + \frac{h}{n^\gamma}\right)\right. \\
&\qquad \left. + \hat{Q}_4\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_5\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_6\left(\theta_0 + \frac{h}{n^\gamma}\right)\right) \\
&= n^{2\gamma}\left(Q_1\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_2\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_3\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_4\left(\theta_0 + \frac{h}{n^\gamma}\right)\right) + o_P(1) \\
& \rightsquigarrow B_0(h).
\end{aligned} \tag{8}$$

$$\begin{aligned}
Q_1(\theta) =& \frac{1}{2} g(\theta)' W g(\theta) + g(\theta)' W \pi(\theta_0), \quad \hat{Q}_3(\theta) = \pi(\theta_0)' W \hat{\eta}_n(\theta) \\
\hat{Q}_2(\theta) =& \frac{1}{2}\hat{\eta}_n(\theta)' W \hat{\eta}_n(\theta) + g(\theta)' W(\hat{\pi}_n(\theta_0) - \pi(\theta_0)) + g(\theta)' W \hat{\eta}_n(\theta) + (\hat{\pi}_n(\theta_0) - \pi(\theta_0))' W \hat{\eta}_n(\theta) \\
\hat{Q}_4(\theta) =& \frac{1}{2} g(\theta)' (W_n - W) g(\theta) + g(\theta)' (W_n - W) \pi(\theta_0) \\
\hat{Q}_5(\theta) =& g(\theta)' (W_n - W)(\hat{\pi}_n(\theta_0) - \pi(\theta_0)) \\
& + g(\theta)' (W_n - W) \hat{\eta}_n(\theta) + (\hat{\pi}_n(\theta_0) - \pi(\theta_0))' (W_n - W) \hat{\eta}_n(\theta) \\
\hat{Q}_6(\theta) =& \pi(\theta_0)' (W_n - W) \hat{\eta}_n(\theta) + \frac{1}{2}\hat{\eta}_n(\theta)' (W_n - W) \hat{\eta}_n(\theta).
\end{aligned}$$

Recall $n^{2\gamma} Q_1(\theta_0 + n^{-\gamma}h) = n^\gamma h'G'W\pi(\theta_0) + \frac{1}{2}h'\bar{H}h + o(1)$. Additionally, Chernoff regularity implies that $+\infty 1(h \notin n^\gamma(C - \theta_0)) \xrightarrow{e} +\infty 1(h \notin T_C(\theta_0))$, where $T_C(\theta_0) \equiv \limsup_{\tau \downarrow 0} \frac{C - \theta_0}{\tau}$. When θ_0 is the constrained minimizer, $h'G'W\pi(\theta_0) \geqslant 0$ 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 θ_0 would not be the constrained minimizer anymore. Define

$$\bar{Q}_n\left(\theta_0 + \frac{h}{n^\gamma}\right) \equiv \hat{Q}_2\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_3\left(\theta_0 + \frac{h}{n^\gamma}\right) + \hat{Q}_4\left(\theta_0 + \frac{h}{n^\gamma}\right). \tag{9}$$

Then, for any $c \in \mathbb{R}$,

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(n^{2\gamma}\left(\hat{Q}_n(\theta_0) - \inf_{h\in\mathcal{C}_{\delta_n}^{\theta_0}} \hat{Q}_n\left(\theta_0 + \frac{h}{n^\gamma}\right)\right) > c\right) \\
&= \lim_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(-n^{2\gamma} \inf_{h\in\mathcal{C}_{\delta_n}^{\theta_0}} \left\{\bar{Q}_n\left(\theta_0 + \frac{h}{n^\gamma}\right) + Q_1\left(\theta_0 + \frac{h}{n^\gamma}\right)\right\} > c\right) \\
&\leqslant \lim_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(-\inf_{h\in\mathcal{C}_{\delta_n}^{\theta_0}} \left\{n^{2\gamma}\bar{Q}_n\left(\theta_0 + \frac{h}{n^\gamma}\right) + n^\gamma h'G'W\pi(\theta_0) + \frac{1}{2}h'\bar{H}h\right\} > c\right) \\
&\leqslant \lim_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(-\inf_{h\in\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}} \left\{n^{2\gamma}\bar{Q}_n\left(\theta_0 + \frac{h}{n^\gamma}\right) + \frac{1}{2}h'\bar{H}h\right\} > c\right)
\end{aligned}$$

Because the test statistic $n^{2\gamma}\left(\hat{Q}_n(\theta_0) - \inf_{h\in\mathcal{C}_{\delta_n}^{\theta_0}} \hat{Q}_n\left(\theta_0 + \frac{h}{n^\gamma}\right)\right)$ may 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 $-\inf_{h\in\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}} \left\{n^{2\gamma}\bar{Q}_n\left(\theta_0 + \frac{h}{n^\gamma}\right) + \frac{1}{2}h'\bar{H}h\right\}$ under P, and $J(\cdot, P)$ as the CDF of its limiting distribution $-\inf_{\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}} B_0(h)$ 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 $-\inf_{\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}} B_0(h)$ 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 $n^{2\gamma}\left(\hat{Q}_n(\theta_n) - \inf_{h\in\mathcal{B}_{\delta_n}} \hat{Q}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right)$ by $n^{2\gamma}\left(\hat{Q}_n(\theta_n) - \inf_{h\in\mathcal{C}_{\delta_n}^{\theta_n}} \hat{Q}_n\left(\theta_n + \frac{h}{n^\gamma}\right)\right)$.

■

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

$$\begin{aligned}
Q_1(\theta) &= \frac{1}{2} g(\theta)' W g(\theta) + g(\theta)' W \pi(\theta_0), \quad \hat{Q}_3(\theta) = \pi(\theta_0)' W \hat{\eta}_n(\theta) \\
\hat{Q}_2(\theta) &= \frac{1}{2} \hat{\eta}_n(\theta)' W \hat{\eta}_n(\theta) + g(\theta)' W (\hat{\pi}_n(\theta_0) - \pi(\theta_0)) + g(\theta)' W \hat{\eta}_n(\theta) + (\hat{\pi}_n(\theta_0) - \pi(\theta_0))' W \hat{\eta}_n(\theta).
\end{aligned}$$

Consistency of $\hat{\theta}_n$ for θ_0 follows from constraining θ 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 $\sup_{\theta \in C} \left| \hat{Q}_n(\theta) - Q(\theta) \right| = o_P(1)$.

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

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

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

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

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

Taylor expanding $Q_1(\theta)$ around θ_0 while constraining θ to lie in C, $Q_1(\theta) = Q_1(\theta_0) + (\theta - \theta_0)' \frac{\partial Q_1(\theta_0)}{\partial \theta} + \frac{1}{2} (\theta - \theta_0)' \frac{\partial^2 Q_1(\theta_0)}{\partial \theta \partial \theta'} (\theta - \theta_0) + o(\|\theta - \theta_0\|^2) = (\theta - \theta_0)' G' W \pi(\theta_0)$ $+\frac{1}{2} (\theta - \theta_0)' \left(\bar{H} + o(1)\right) (\theta - \theta_0)$ since $\frac{\partial Q_1(\theta_0)}{\partial \theta} = G' W \pi(\theta_0)$ and $\frac{\partial^2 Q_1(\theta_0)}{\partial \theta \partial \theta'} = \bar{H}$. Because $\bar{H}$ is positive definite and $(\theta-\theta_0)' G' W \pi(\theta_0) \geqslant 0$ for all $\theta \in C$, there exists $K>0$ and a small enough neighborhood of θ_0 such that $Q_1(\theta) \geqslant K\|\theta-\theta_0\|^2$. By consistency of $\hat{\theta}_n$ for θ_0, with probability approaching 1, $Q_1\left(\hat{\theta}_n\right) \geqslant K\left\|\hat{\theta}_n-\theta_0\right\|^2$. Then,

$$Q_1\left(\hat{\theta}_n\right)+\hat{Q}_2\left(\hat{\theta}_n\right)+\hat{Q}_3\left(\hat{\theta}_n\right)=\hat{Q}_n\left(\hat{\theta}_n\right)-\hat{Q}_n(\theta_0) \leqslant \hat{Q}_n\left(\hat{\theta}_n\right)-\inf_{\theta \in C} \hat{Q}_n(\theta) \leqslant o_P\left(n^{-2\gamma}\right).$$

Choose ϵ so that $K-\epsilon>0$. Then,

$$\begin{aligned}
o_P\left(n^{-2\gamma}\right) &\geqslant Q_1\left(\hat{\theta}_n\right)-\epsilon\left\|\hat{\theta}_n-\theta_0\right\|^2-n^{-2\gamma}M_n^2 \\
&\geqslant (K-\epsilon)\left\|\hat{\theta}_n-\theta_0\right\|^2-n^{-2\gamma}M_n^2 \\
\Longrightarrow \left\|\hat{\theta}_n-\theta_0\right\|^2 &\leqslant (K-\epsilon)^{-1} n^{-2\gamma}M_n^2+o_P\left(n^{-2\gamma}\right)=O_P\left(n^{-2\gamma}\right).
\end{aligned}$$

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

Using the arguments in Theorems 2 and 6 of Hong and Li (2023), $\hat{A}_n(h) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} A_0(h)$, where

$$\begin{aligned}
\hat{A}_n(h)=\;&\hat{\pi}_n\left(\hat{\theta}_n\right)' W n^{\gamma\rho}\sqrt{n}\left(P_n^*-P_n\right)\left(\pi\left(\cdot,\hat{\theta}_n+\frac{h}{n^\gamma}\right)-\pi\left(\cdot,\hat{\theta}_n\right)\right) \\
&+\frac{\sqrt{n}n^{\gamma\rho}}{2n^{2\gamma}}h'\left(\hat{G}'W\hat{G}+\sum_{j=1}^m\sum_{k=1}^m W_{jk}\hat{\pi}_{nk}\left(\hat{\theta}_n\right)\hat{H}_j\right)h \\
&+\frac{n^{\gamma\rho}}{n^\gamma}h'\hat{G}'W\sqrt{n}\left(P_n^*-P_n\right)\pi\left(\cdot,\hat{\theta}_n\right).
\end{aligned}$$

For $\bar{H}=G'WG+\sum_{j=1}^m\sum_{k=1}^m W_{jk}\pi_k(\theta_0)H_j$, $A_0(h)=\pi(\theta_0)'W\mathcal{Z}_{0,1}(h)+h'G'WU_0+\frac{1}{2}h'\bar{H}h$ for $\rho=1$ and $A_0(h)=\pi(\theta_0)'W\mathcal{Z}_{0,1/2}(h)+\frac{1}{2}h'\bar{H}h$ when $\rho=1/2$.

Chernoff regularity implies that for any sequence $a(n)\to\infty$, $+\infty 1\left(h\notin a(n)(C-\theta_0)\right) \xrightarrow{e}$

$+\infty 1\left(h \notin T_C\left(\theta_0\right)\right)$. Since $n^\gamma / \eta_n \rightarrow \infty$ and $n^\gamma\left(\hat{\theta}_n-\theta_0\right)=O_P(1)$,

$$
\begin{aligned}
\omega_n^*(h) & =+\infty 1\left(h \notin \frac{n^\gamma\left(C-\theta_0\right)}{\eta_n}-\frac{n^\gamma\left(\hat{\theta}_n-\theta_0\right)}{\eta_n}\right) \\
& =+\infty 1\left(h \notin \frac{n^\gamma\left(C-\theta_0\right)}{\eta_n}-o_p(1)\right) \\
& \xrightarrow{e} \omega(h) \equiv+\infty 1\left(h \notin T_C\left(\theta_0\right)\right) .
\end{aligned}
$$

where $T_C\left(\theta_0\right) \equiv \limsup\limits_{\tau \downarrow 0} \frac{C-\theta_0}{\tau}$. By a bootstrap in probability version of Theorem 4 in Knight (1999),

$$
\hat{A}_n(h)+\omega_n^*(h) \underset{e-d}{\xrightarrow{p}} A_0(h)+\omega(h) .
$$

where $\underset{e-d}{\xrightarrow{p}}$ denotes epi-convergence of the conditional law of $\hat{\mathbb{G}}_n^*$ to $\mathbb{G}_0$, which can be equivalently stated as $\sup_{f \in BL_1}\left|E_{\mathbb{W}} f\left(\hat{\mathbb{G}}_n^*\right)-E f\left(\mathbb{G}_0\right)\right| \xrightarrow{p} 0$ and $E_{\mathbb{W}} f\left(\hat{\mathbb{G}}_n^*\right)^*-E_{\mathbb{W}} f\left(\hat{\mathbb{G}}_n^*\right)_* \xrightarrow{p} 0$ 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 $d\left(\hat{\mathbb{G}}_n^*, \mathbb{G}_0\right)=\int_0^{\infty} \max \left\{\left|d_{\text{epi } \hat{\mathbb{G}}_n^*}(v)-d_{\text{epi } \mathbb{G}_0}(v)\right|:|v| \leqslant \rho\right\} \exp (-\rho) d \rho$, 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 $G: \mathbb{R}^d \mapsto \mathbb{R}$. Lemma 2.6 in Kim and Pollard (1990) implies that the Gaussian processes $\mathcal{Z}_{0,1/2}(h)$ and $\mathcal{Z}_{0,1}(h)$ have a unique minimum. In combination with the fact that $\frac{1}{2} h^{\prime} \bar{H} h$ is a convex function of h, there is a unique h that minimizes $A_0(h)+\omega(h)$. By a modification of the bootstrap argmin continuous mapping lemma 14.2 in Hong and Li (2020) that replaces weak convergence with epi-convergence, $n^\gamma\left(\hat{\theta}_n^*-\hat{\theta}_n\right)=\underset{h \in \mathbb{R}^d}{\arg \min }\left\{\hat{A}_n(h)+\omega_n^*(h)\right\} \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \underset{h \in \mathbb{R}^d}{\arg \min }\left\{A_0(h)+\omega(h)\right\}$, which coincides with the limiting distribution of $n^\gamma\left(\hat{\theta}_n-\theta_0\right)$.

∎

6.2 Verification of Assumption 3

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

$$
\begin{aligned}
& m_n(\cdot,\theta,h) - m_n(\cdot,\theta_0,h) \\
=& \sqrt{n}\left(\pi\left(\cdot;\theta+\frac{h}{\sqrt{n}}\right) - \pi(\cdot;\theta) - \left(\pi\left(\cdot,\theta_0+\frac{h}{\sqrt{n}}\right) - \pi(\cdot,\theta_0)\right)\right) \\
=& \sqrt{n}\left(\left(y_i-\theta-\frac{h}{\sqrt{n}}\right)^2 - (y_i-\theta)^2 - \left(\left(y_i-\theta_0-\frac{h}{\sqrt{n}}\right)^2 - (y_i-\theta_0)^2\right)\right) \\
=& \sqrt{n}\left(-2\left(y_i-\theta_0-\frac{h}{\sqrt{n}}\right) + 2(y_i-\theta_0)\right)(\theta-\theta_0) \\
=& 2h(\theta-\theta_0)
\end{aligned}
$$

$\sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n} \left\|\frac{m_n(\cdot,\theta,h)-m_n(\cdot,\theta_0,h)}{1+n^\gamma\|\theta-\theta_0\|}\right\| \leqslant K$ will be satisfied if we take $K = 2\sup_{h\in\mathcal{H}}|h|$.

Another way Assumption 3 can be satisfied is if $E\left[\sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n}\left\|\frac{m_n(\cdot,\theta,h)-m_n(\cdot,\theta_0,h)}{1+n^{1/3}\|\theta-\theta_0\|}\right\|^{2+\delta}\right] < \infty$ for all n and any $\delta>0$. For Example 2, there exists some constant $C>0$ such that

$$
\begin{aligned}
& E\left[\sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n}\left|\frac{m_n(\cdot,\theta,h)-m_n(\cdot,\theta_0,h)}{1+n^\gamma\|\theta-\theta_0\|}\right|^{2+\delta}\right] \\
=& E\left[\sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n}\frac{n^{1/3}\left|\pi\left(\cdot;\theta+hn^{-1/3}\right)-\pi(\cdot;\theta)-\left(\pi\left(\cdot,\theta_0+hn^{-1/3}\right)-\pi(\cdot,\theta_0)\right)\right|^{2+\delta}}{\left(1+n^{1/3}\|\theta-\theta_0\|\right)^{2+\delta}}\right] \\
\leqslant& E\left[\sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n}\frac{n^{1/3}\left|\mathbf{1}\left(\theta+\frac{h}{n^{1/3}}-1\leqslant y_i\leqslant\theta+\frac{h}{n^{1/3}}+1\right)-\mathbf{1}(\theta-1\leqslant y_i\leqslant\theta+1)\right|}{\left(1+n^{1/3}\|\theta-\theta_0\|\right)^{2+\delta}}\right] \\
+& E\left[\sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n}\frac{n^{1/3}\left|\mathbf{1}\left(\theta_0+\frac{h}{n^{1/3}}-1\leqslant y_i\leqslant\theta_0+\frac{h}{n^{1/3}}+1\right)-\mathbf{1}(\theta_0-1\leqslant y_i\leqslant\theta_0+1)\right|}{\left(1+n^{1/3}\|\theta-\theta_0\|\right)^{2+\delta}}\right] \\
\leqslant& \sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n}\frac{Ch}{\left(1+n^{1/3}\|\theta-\theta_0\|\right)^{2+\delta}} < \infty
\end{aligned}
$$

For Example 3, $m_n(\cdot,\theta,h) = n^\gamma\left[1\left(y_i\leqslant\theta+hn^{-\gamma}\right)-\tau, y_i-\theta-hn^{-\gamma}\right]' - n^\gamma\left[1(y_i\leqslant\theta)-\tau, y_i-\theta\right]' = n^\gamma\left[1\left(y_i\leqslant\theta+hn^{-\gamma}\right)-1(y_i\leqslant\theta), -hn^{-\gamma}\right]' \equiv \left[m_{n1}(\cdot,\theta,h), m_{n2}(\cdot,\theta,h)\right]'$, where $\gamma=1/2$ if

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

$$
\begin{aligned}
&E\left[\sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n}\left|\frac{m_{n1}(\cdot,\theta,h)-m_{n1}(\cdot,\theta_0,h)}{1+n^{\gamma}\|\theta-\theta_0\|}\right|^{2+\delta}\right]\\
\leqslant &E\left[\sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n}\frac{n^{\gamma}\left|1\left(y_i\leqslant\theta+hn^{-\gamma}\right)-1\left(y_i\leqslant\theta\right)\right|}{\left(1+n^{\gamma}\|\theta-\theta_0\|\right)^{2+\delta}}\right]\\
+&E\left[\sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n}\frac{n^{\gamma}\left|1\left(y_i\leqslant\theta_0+hn^{-\gamma}\right)-1\left(y_i\leqslant\theta_0\right)\right|}{\left(1+n^{\gamma}\|\theta-\theta_0\|\right)^{2+\delta}}\right]\\
\leqslant &\sup_{h\in\mathcal{H},\|\theta-\theta_0\|\leqslant\epsilon_n}\frac{Ch}{\left(1+n^{\gamma}\|\theta-\theta_0\|\right)^{2+\delta}} < \infty
\end{aligned}
$$

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

6.3 Additional Monte Carlo Simulations

We now examine the coverage frequency and average interval length using $\kappa = \infty$ in the fixed weighting matrix setup. In the case of $\beta_0 \geqslant 0$, Table 6 shows that the coverage frequencies and average interval lengths are similar to the case of $\kappa = 5$.

In the case of $\beta_0 < 0$, Table 7 shows that the coverage is not affected, but the average interval length can be longer when we use $\kappa = \infty$, especially for the more negative values of β_0. This suggests that it is better to use a finite $\kappa < \infty$, although if the value of κ is too small, the coverage can be more conservative, as the next table shows.

We next examine the coverage frequency and average interval length using $\kappa = 1$ in the fixed weighting matrix setup. In the case of $\beta_0 \geqslant 0$, Table 8 shows that the coverage can be more conservative than in the case of $\kappa = 5$. For example, when $\tau = 0.5$, all values of β_0 lead to coverage that is above 98% when $\kappa = 1$, but the coverage was under 98% when $\kappa = 5$. For the other values of τ, the coverage is similar between $\kappa = 1$ and $\kappa = 5$. Even

Table 6: Coverage Frequencies and Average Interval Lengths, $\kappa = \infty$

β_0	0	n^{-1}	$n^{-1/2}$	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/6}$	1
				$\tau = 0.1$			
$n = 100$	1.000	1.000	0.992	0.980	0.961	0.962	0.946
	(0.515)	(0.503)	(0.432)	(0.388)	(0.384)	(0.379)	(0.377)
$n = 500$	1.000	1.000	1.000	0.992	0.968	0.955	0.967
	(0.358)	(0.352)	(0.296)	(0.223)	(0.202)	(0.200)	(0.200)
$n = 1000$	1.000	1.000	1.000	0.999	0.978	0.953	0.958
	(0.323)	(0.321)	(0.271)	(0.189)	(0.155)	(0.153)	(0.153)
$n = 5000$	1.000	1.000	1.000	1.000	1.000	0.971	0.970
	(0.289)	(0.288)	(0.262)	(0.183)	(0.102)	(0.083)	(0.084)
				$\tau = 0.3$			
$n = 100$	0.996	0.994	0.971	0.959	0.950	0.952	0.944
	(0.421)	(0.412)	(0.370)	(0.353)	(0.354)	(0.352)	(0.352)
$n = 500$	1.000	1.000	0.994	0.974	0.956	0.949	0.961
	(0.242)	(0.239)	(0.203)	(0.177)	(0.175)	(0.176)	(0.177)
$n = 1000$	1.000	1.000	0.998	0.974	0.954	0.945	0.956
	(0.200)	(0.199)	(0.164)	(0.132)	(0.129)	(0.130)	(0.130)
$n = 5000$	1.000	1.000	1.000	0.996	0.953	0.958	0.967
	(0.155)	(0.154)	(0.131)	(0.076)	(0.065)	(0.065)	(0.065)
				$\tau = 0.5$			
$n = 100$	0.976	0.975	0.963	0.951	0.943	0.948	0.943
	(0.365)	(0.361)	(0.345)	(0.341)	(0.342)	(0.341)	(0.342)
$n = 500$	0.972	0.965	0.965	0.964	0.959	0.946	0.958
	(0.177)	(0.175)	(0.167)	(0.166)	(0.166)	(0.166)	(0.166)
$n = 1000$	0.976	0.971	0.964	0.946	0.951	0.944	0.957
	(0.126)	(0.127)	(0.121)	(0.119)	(0.119)	(0.120)	(0.119)
$n = 5000$	0.978	0.976	0.968	0.944	0.953	0.955	0.955
	(0.058)	(0.058)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)

though the coverage can be more conservative for $\kappa = 1$, the average interval lengths can also be shorter, especially for values of β_0 that are closer to zero. For example, when $\tau = 0.1$ and $\beta_0 = n^{-1}$, the average interval length is 0.066 when $\kappa = 1$ and 0.289 when $\kappa = 5$. In the case of $\beta_0 < 0$ and $\tau = 0.5$, Table 9 shows that the coverage is more conservative when using $\kappa = 1$, but the average interval lengths are shorter for all values of τ.

Some intuition for why coverage can be more conservative when $\kappa = 1$ comes from

Table 7: Coverage Frequencies and Average Interval Lengths, $\kappa = \infty$

β_0	$-n^{-1}$	$-n^{-1/2}$	$-n^{-1/3}$	$-n^{-1/4}$	$-n^{-1/6}$	-1
			$\tau = 0.1$			
$n = 100$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.525)	(0.634)	(0.802)	(0.985)	(1.238)	(2.237)
$n = 500$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.358)	(0.427)	(0.570)	(0.728)	(1.006)	(2.247)
$n = 1000$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.325)	(0.375)	(0.501)	(0.652)	(0.917)	(2.249)
$n = 5000$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.289)	(0.315)	(0.401)	(0.520)	(0.760)	(2.245)
			$\tau = 0.3$			
$n = 100$	0.996	1.000	1.000	1.000	1.000	1.000
	(0.426)	(0.515)	(0.669)	(0.846)	(1.095)	(2.114)
$n = 500$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.243)	(0.298)	(0.433)	(0.589)	(0.868)	(2.125)
$n = 1000$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.202)	(0.245)	(0.364)	(0.513)	(0.779)	(2.126)
$n = 5000$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.155)	(0.180)	(0.264)	(0.383)	(0.624)	(2.124)
			$\tau = 0.5$			
$n = 100$	0.980	0.998	1.000	1.000	1.000	1.000
	(0.367)	(0.424)	(0.551)	(0.713)	(0.957)	(1.986)
$n = 500$	0.969	0.998	1.000	1.000	1.000	1.000
	(0.176)	(0.202)	(0.305)	(0.451)	(0.730)	(1.999)
$n = 1000$	0.975	0.998	1.000	1.000	1.000	1.000
	(0.127)	(0.144)	(0.233)	(0.376)	(0.642)	(2.001)
$n = 5000$	0.978	0.999	1.000	1.000	1.000	1.000
	(0.058)	(0.066)	(0.130)	(0.245)	(0.486)	(1.999)

examining the stochastic dominance argument in the proof of Theorem 4. For $\kappa' > \kappa$,

$$\begin{aligned}
&\limsup_{n\to\infty}\sup_{P\in\mathcal{P}} P\left(-\inf_{h\in\{h\in T_C(\theta_0):\|h\|\leqslant\kappa\}}\left\{n^{2\gamma}\left(\hat{Q}_2\left(\theta_0+\frac{h}{n^\gamma}\right)+\hat{Q}_3\left(\theta_0+\frac{h}{n^\gamma}\right)\right)+\frac{1}{2}h'\bar{H}h\right\}>c\right)\\
\leqslant&\limsup_{n\to\infty}\sup_{P\in\mathcal{P}} P\left(-\inf_{h\in\{h\in T_C(\theta_0):\|h\|\leqslant\kappa'\}}\left\{n^{2\gamma}\left(\hat{Q}_2\left(\theta_0+\frac{h}{n^\gamma}\right)+\hat{Q}_3\left(\theta_0+\frac{h}{n^\gamma}\right)\right)+\frac{1}{2}h'\bar{H}h\right\}>c\right)
\end{aligned}$$

So the limiting distribution of $-\inf_{h\in\{h\in T_C(\theta_0):\|h\|\leqslant\kappa'\}}\left\{n^{2\gamma}\left(\hat{Q}_2\left(\theta_0+\frac{h}{n^\gamma}\right)+\hat{Q}_3\left(\theta_0+\frac{h}{n^\gamma}\right)\right)+\frac{1}{2}h'\bar{H}h\right\}$, which is $-\inf_{\{h\in T_C(\theta_0):\|h\|\leqslant\kappa'\}} A_0(h)$, is closer to the benchmarking statistic's asymptotic distribution $-\inf_{h\in\mathbb{R}^d} A_0(h)$ when $\kappa' > \kappa$. However, because this comparison is between limiting

Table 8: Coverage Frequencies and Average Interval Lengths, $\kappa = 1$

β_0	0	n^{-1}	$n^{-1/2}$	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/6}$	1
				$\tau = 0.1$			
$n = 100$	1.000	1.000	0.992	0.980	0.962	0.963	0.949
	(0.318)	(0.323)	(0.357)	(0.386)	(0.398)	(0.396)	(0.393)
$n = 500$	1.000	1.000	1.000	0.993	0.968	0.959	0.969
	(0.163)	(0.164)	(0.176)	(0.200)	(0.205)	(0.205)	(0.205)
$n = 1000$	1.000	1.000	1.000	0.999	0.977	0.954	0.959
	(0.126)	(0.126)	(0.133)	(0.152)	(0.156)	(0.156)	(0.156)
$n = 5000$	1.000	1.000	1.000	1.000	1.000	0.971	0.971
	(0.066)	(0.066)	(0.067)	(0.072)	(0.086)	(0.085)	(0.085)
				$\tau = 0.3$			
$n = 100$	0.996	0.993	0.973	0.963	0.950	0.952	0.947
	(0.322)	(0.324)	(0.345)	(0.354)	(0.357)	(0.356)	(0.355)
$n = 500$	1.000	1.000	0.994	0.972	0.956	0.949	0.962
	(0.168)	(0.168)	(0.177)	(0.177)	(0.176)	(0.176)	(0.177)
$n = 1000$	1.000	1.000	0.998	0.974	0.956	0.943	0.955
	(0.130)	(0.130)	(0.136)	(0.131)	(0.130)	(0.130)	(0.130)
$n = 5000$	1.000	1.000	1.000	0.997	0.954	0.957	0.967
	(0.067)	(0.067)	(0.069)	(0.072)	(0.065)	(0.064)	(0.065)
				$\tau = 0.5$			
$n = 100$	0.992	0.987	0.979	0.982	0.976	0.981	0.977
	(0.315)	(0.324)	(0.387)	(0.429)	(0.439)	(0.439)	(0.439)
$n = 500$	0.989	0.988	0.986	0.988	0.983	0.987	0.990
	(0.143)	(0.146)	(0.179)	(0.207)	(0.209)	(0.208)	(0.209)
$n = 1000$	0.994	0.990	0.981	0.985	0.981	0.981	0.987
	(0.105)	(0.105)	(0.130)	(0.149)	(0.149)	(0.150)	(0.150)
$n = 5000$	0.993	0.990	0.987	0.983	0.983	0.985	0.986
	(0.048)	(0.048)	(0.059)	(0.068)	(0.068)	(0.068)	(0.068)

distributions rather than finite sample distributions, it is difficult to know for sure how conservative the coverage will be for any finite n. For example, in the simulations above, when $\kappa = 1$, $n = 5000$, $\beta_0 = n^{-1/4}$, and $\tau = 0.3$, the coverage is 0.954, which slightly less conservative than the coverage of 0.957 when $\kappa = 5$. However, for $\tau = 0.5$, the coverage is 0.983 when $\kappa = 1$, which is more conservative than the coverage of 0.948 when $\kappa = 5$.

Table 9: Coverage Frequencies and Average Interval Lengths, $\kappa = 1$

β_0	$-n^{-1}$	$-n^{-1/2}$	$-n^{-1/3}$	$-n^{-1/4}$	$-n^{-1/6}$	-1
			$\tau = 0.1$			
$n = 100$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.316)	(0.283)	(0.255)	(0.239)	(0.223)	(0.200)
$n = 500$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.163)	(0.153)	(0.143)	(0.137)	(0.132)	(0.120)
$n = 1000$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.126)	(0.122)	(0.116)	(0.112)	(0.108)	(0.102)
$n = 5000$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.066)	(0.065)	(0.064)	(0.063)	(0.061)	(0.059)
			$\tau = 0.3$			
$n = 100$	0.996	1.000	1.000	1.000	1.000	1.000
	(0.318)	(0.286)	(0.254)	(0.236)	(0.214)	(0.200)
$n = 500$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.167)	(0.156)	(0.141)	(0.134)	(0.129)	(0.120)
$n = 1000$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.130)	(0.123)	(0.114)	(0.110)	(0.106)	(0.100)
$n = 5000$	1.000	1.000	1.000	1.000	1.000	1.000
	(0.067)	(0.066)	(0.063)	(0.061)	(0.060)	(0.059)
			$\tau = 0.5$			
$n = 100$	0.991	0.999	1.000	1.000	1.000	1.000
	(0.308)	(0.234)	(0.169)	(0.138)	(0.111)	(0.100)
$n = 500$	0.990	1.000	1.000	1.000	1.000	1.000
	(0.142)	(0.106)	(0.066)	(0.054)	(0.048)	(0.040)
$n = 1000$	0.992	1.000	1.000	1.000	1.000	1.000
	(0.104)	(0.079)	(0.048)	(0.039)	(0.035)	(0.030)
$n = 5000$	0.990	1.000	1.000	1.000	1.000	1.000
	(0.048)	(0.036)	(0.020)	(0.017)	(0.015)	(0.014)

References

Andrews, Donald W.K., "Estimation when a parameter is on a boundary," *Econometrica*, 1999, *67* (6), 1341–1383. 2

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, 52

_ , "Testing when a parameter is on the boundary of the maintained hypothesis," *Econometrica*, 2001, *69* (3), 683–734. 3

Andrews, Donald W.K., "Generalized method of moments estimation when a parameter is on a boundary," *Journal of Business & Economic Statistics*, 2002, *20* (4), 530–544. 2

Andrews, Donald WK and Patrik Guggenberger, "Asymptotic size and a problem with subsampling and with the m out of n bootstrap," *Econometric Theory*, 2010, *26* (2), 426–468. 2, 8

Bickel, Peter J and Anat Sakov, "On the choice of m in the m out of n bootstrap and confidence bounds for extrema," *Statistica Sinica*, 2008, pp. 967–985. 2

Bonnans, J Frédéric and Alexander Shapiro, *Perturbation analysis of optimization problems*, Springer Science & Business Media, 2013. 13

Cattaneo, Matias D, Michael Jansson, and Kenichi Nagasawa, "Bootstrap-Based Inference for Cube Root Asymptotics," *Econometrica*, 2020, *88* (5), 2203–2219. 3

Chen, Xiaohong, Timothy M Christensen, and Elie Tamer, "Monte Carlo confidence sets for identified sets," *Econometrica*, 2018, *86* (6), 1965–2018. 3

Chernozhukov, Victor and Christian Hansen, "An IV model of quantile treatment effects," *Econometrica*, 2005, *73* (1), 245–261. 18

_ , Whitney K Newey, and Andres Santos, "Constrained Conditional Moment Restriction Models," *Econometrica*, 2023, *91* (2), 709–736. 3

der Vaart, Aad W Van, *Asymptotic statistics*, Vol. 3, Cambridge University Press, 2000. 42, 46, 49

Dudley, RM, "Universal Donsker Classes and Metric Entropy," *Annals of Probability*, 1987, *15* (4), 1306–1326. 9

Dupacová, Jitka and Roger Wets, "Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems," *The annals of statistics*, 1988, *16* (4), 1517–1549. 13

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

Fang, Zheng and Juwon Seo, "A projection framework for testing shape restrictions that form convex cones," *Econometrica*, 2021, *89* (5), 2439–2458. 3

Gafarov, Bulat, "Inference on scalar parameters in set–identified affine models," Technical Report, Mimeo: UC Davis 2016. 3, 4

Geyer, Charles J, "On the asymptotics of constrained M-estimation," *The Annals of Statistics*, 1994, pp. 1993–2010. 2, 3, 4, 7

Hong, Han and Jessie Li, "The numerical bootstrap," *The Annals of Statistics*, 2020, *48* (1), 397–412. 2, 37, 51

_ **and Jessie. Li**, "Rate-Adaptive Bootstrap for Possibly Misspecified GMM," *forthcoming at Econometric Theory*, 2023. 3, 16, 19, 20, 25, 29, 43, 46, 47, 50

Honore, Bo E and Luojia Hu, "Estimation of cross sectional and panel data censored regression models with endogeneity," *Journal of Econometrics*, 2004, *122* (2), 293–316. 19

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

Kaido, Hiroaki, "A dual approach to inference for partially identified econometric models," *Journal of Econometrics*, 2016, *192* (1), 269–290. 3

_ **and Andres Santos**, "Asymptotically efficient estimation of models defined by convex moment inequalities," *Econometrica*, 2014, *82* (1), 387–413. 3

_ **, Francesca Molinari, and Jörg Stoye**, "Confidence intervals for projections of partially identified parameters," *Econometrica*, 2019, *87* (4), 1397–1432. 3

_ **,** _ **, and** _ , "Constraint qualifications in partial identification," *Econometric Theory*, 2021, pp. 1–24. 3

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

Kim, Jeankyung and David Pollard, "Cube root asymptotics," *The Annals of Statistics*, 1990, pp. 191–219. 9, 18, 39, 49, 51

Knight, Keith, "Epi-convergence in distribution and stochastic equi-semicontinuity," *Unpublished manuscript*, 1999, *37*. 13, 51

_ , "Limiting distributions of linear programming estimators," *Extremes*, 2001, *4* (2), 87–103. 3

_ , "Asymptotic theory for M-estimators of boundaries," in "The Art of Semiparametrics," Springer, 2006, pp. 1–21. 3

Kosorok, Michael R, *Introduction to empirical processes and semiparametric inference*, Springer, 2007. 6, 38, 40, 44

Li, Jessie, "The Proximal Bootstrap for Constrained Estimators," *working paper*, 2023. 29

McFadden, Daniel, "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica: Journal of the Econometric Society*, 1989, pp. 995–1026. 19

Moon, Hyungsik Roger and Frank Schorfheide, "Estimation with overidentifying inequality moment conditions," *Journal of Econometrics*, 2009, *153* (2), 136–154. 3, 4

Newey, W. and D. McFadden, "Large Sample Estimation and Hypothesis Testing," in R. Engle and D. McFadden, eds., *Handbook of Econometrics, Vol. 4*, North Holland, 1994, pp. 2113–2241. 42, 46, 49

Nocedal, Jorge and Stephen Wright, *Numerical optimization*, Springer Science & Business Media, 2006. 41

Pakes, A. and D. Pollard, "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 1989, *57* (5), 1027–1057. 19

Politis, D., J. Romano, and M. Wolf, *Subsampling*, Springer Series in Statistics, 1999. 2

Pollard, David, "Asymptotics via empirical processes," *Statistical Science*, 1989, pp. 341–354. 9

Robinson, Stephen M, "Analysis of sample-path optimization," *Mathematics of Operations Research*, 1996, *21* (3), 513–528. 13

Romano, Joseph P and Azeem M Shaikh, "On the uniform asymptotic validity of subsampling and the bootstrap," *The Annals of Statistics*, 2012, *40* (6), 2798–2822. 40, 44

Shapiro, Alexander, "Sensitivity analysis of nonlinear programs and differentiability properties of metric projections," *SIAM Journal on Control and Optimization*, 1988, *26* (3), 628–645. 3

_ , "Asymptotic properties of statistical estimators in stochastic programming," *The Annals of Statistics*, 1989, pp. 841–858. 3

_ , "On differential stability in stochastic programming," *Mathematical Programming*, 1990, *47* (1-3), 107–116. 3, 13

van der Vaart, AW and Jon Wellner, *Weak Convergence and Empirical Processes*, Springer, 1996. 5, 37, 38, 42, 46, 49

Wellner, Jon A and Yihui Zhan, "Bootstrapping Z-estimators," *University of Washington Department of Statistics Technical Report*, 1996, *308.* 10, 39