Rate-Adaptive Bootstrap for Possibly Misspecified GMM *

Han Hong†  Jessie Li‡

June 10, 2021

We consider inference for possibly misspecified GMM models based on possibly nonsmooth moment conditions. While misspecified GMM estimators with smooth moments remain $\sqrt{n}$ consistent and asymptotically normal, globally misspecified nonsmooth 2-step GMM estimators are $n^{1/3}$ consistent when either the weighting matrix is fixed or when the weighting matrix is estimated at the $n^{1/3}$ rate or faster. If the weighting matrix is estimated at a rate slower than $n^{1/3}$, the 2-step GMM estimator will have the same rate of convergence as the weighting matrix. Because the estimator’s nonstandard asymptotic distribution cannot be consistently estimated using the standard bootstrap, we propose an alternative rate-adaptive bootstrap procedure that consistently estimates the asymptotic distribution regardless of whether the GMM estimator is smooth or nonsmooth, correctly or incorrectly specified. Monte Carlo simulations for the smooth and nonsmooth cases confirm that our rate-adaptive bootstrap confidence intervals exhibit empirical coverage close to the nominal level.

Keywords: rate-adaptive bootstrap, misspecified GMM, cube-root asymptotics.

1 Introduction

Many GMM models are based on nonsmooth moment conditions that involve indicator functions. Examples include quantile instrumental variables (e.g. Chernozhukov and Hansen (2005) and Honoré and Hu (2004b)) and simulated method of moments that are based on frequency simulators (McFadden (1989) and Pakes and Pollard (1989)). While the asymptotic behavior of nonsmooth GMM estimators has been well established when the model is assumed to be correctly specified, in practice it can happen that the model is misspecified in the sense that the population moment conditions evaluated at the parameter value which minimizes the population GMM objective do not equal zero. Many empirical studies use GMM to obtain parameter estimates even though the J-test rejects the null of correct

---

*We thank Ivan Fernandez-Val, Jean-Jacques Forneron, Hiroaki Kaido, Zhongjun Qu, Yinchu Zhu and participants in conferences and seminars for helpful comments.

†Department of Economics, Stanford University. doubleh@stanford.edu

‡Department of Economics, University of California Santa Cruz. jeqli@ucsc.edu
specification (see Lee (2014) for examples of such studies). The reason is that the J-test is known to have poor size control in small samples. Therefore, it is difficult to know for sure if the model is misspecified, which motivates the need for inference procedures robust to misspecification and nonsmoothness. The study of misspecification is not only important for estimation and inference of model parameters and for model testing and selection, but also important for studying the properties of computational methods (Gao and Hong (2014)).

Misspecified GMM models are studied in, for example, Hall and Inoue (2003), Berkowitz et al. (2012), Guggenberger (2012), Lee (2014), Hansen and Lee (2018), Bonhomme and Weidner (2018), Armstrong and Kolesár (2021), and Cheng et al. (2019). All assume that the sample moment conditions are smooth in the parameters, which allows the GMM estimator to typically remain $\sqrt{n}$ consistent and asymptotically normal. In contrast to the above papers, we show that globally misspecified GMM estimators with nonsmooth, specifically non-directionally differentiable, moments, converge at the cubic-root rate to a nonstandard asymptotic distribution, similar to ones in Kim and Pollard (1990) and Jun et al. (2015).

An insightful paper by Cattaneo et al. (2020) proposes a rate-adaptive bootstrap for M-estimators which does not require knowing the estimator’s rate of convergence to consistently estimate the estimator’s limiting distribution and to construct asymptotically valid confidence intervals. Taking inspiration from their paper, we propose a rate-adaptive bootstrap that consistently estimates the limiting distribution of the GMM estimator regardless of whether the model is correctly or incorrectly specified, smooth or nonsmooth. In the case where the model is correctly specified, our rate-adaptive bootstrap confidence intervals cover the true parameter with the specified nominal coverage probability asymptotically. In the case where the model is incorrectly specified, the rate-adaptive bootstrap confidence intervals achieves the nominal coverage asymptotically for the pseudo-true parameter, which is defined as the parameter which minimizes the population GMM objective function.

Section 2 contains the problem setup and gives the formula for our rate-adaptive bootstrap estimate in the case where the GMM estimator is computed using a fixed weighting matrix $W$. Section 3 shows $n^{1/3}$-consistency and gives the non-normal asymptotic distribution of the misspecified GMM estimator with a fixed weighting matrix and nonsmooth moments. Section 4 demonstrates consistency of the rate-adaptive bootstrap under both correctly specified and misspecified GMM and under both smooth and nonsmooth moments for a fixed weighting matrix. Section 5 studies the 2-step GMM estimator computed using an estimated weighting matrix $W_n$. The rate of convergence of the 2-step GMM estimator is determined by the rate of convergence of $W_n$ to a population analog $W$. If $W_n$ converges in probability to $W$ at a rate equal to or faster than $n^{1/3}$, then the 2-step GMM estimator is $n^{1/3}$-consistent. If $W_n$ converges in probability to $W$ at a rate slower than $n^{1/3}$, then the estimator will have the same rate of convergence as $W_n$. Section 6 discusses how our bootstrap can be applied to the GMM formulation of instrumental variables quantile regression (Chernozhukov and Hansen (2005)), simulated method of moments (McFadden (1989) and Pakes and Pollard (1989)), and dynamic censored regression (Honore and Hu (2004a)). Section 7 contains Monte Carlo simulation results demonstrating that the empirical coverage frequencies of the rate-adaptive bootstrap confidence intervals are close to the nominal level, while the empirical coverage frequencies of the standard bootstrap confidence intervals are far from the nominal level for a simple location model with misspecified nonsmooth moments. In another example with misspecified smooth moments, the rate-adaptive bootstrap
performs just as well as the standard bootstrap in terms of empirical coverage and average interval width. Section 8 concludes.

2 Setup for fixed weighting matrix

Consider a random sample $X_1, X_2, \ldots, X_n$ of independent draws from a probability measure $P$ on a sample space $\mathcal{X}$. Define the empirical measure $P_n \equiv \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$, where $\delta_x$ is the measure that assigns mass 1 at $x$ and zero everywhere else. Denote the bootstrap empirical measure by $P_n^*$, which can refer to the multinomial, wild, or other exchangeable bootstraps. Weak convergence is defined in the sense of Kosorok (2007): $X_n \leadsto X_n$ in the metric space $(\mathbb{D}, d)$ if and only if $\sup_{f \in BL_1} |E^* f(X_n) - Ef(X)| \to 0$ where $BL_1$ is the space of functions $f : \mathbb{D} \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1. Conditional weak convergence in probability is also defined in the sense of Kosorok (2007): $X_n \xrightarrow{w} X$ in the metric space $(\mathbb{D}, d)$ if and only if $\sup_{f \in BL_1} |E^* f(X_n) - Ef(X)| \overset{p}{\to} 0$ and $E^* f(X_n)^* - E^* f(X)^* \overset{p}{\to} 0$ 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, $E^*$ denotes expectation with respect to the bootstrap weights $W$ conditional on the data, and $f(X_n)^*$ and $f(X)^*$ denote measurable majorants and minorants with respect to the joint data (including the weights $W$).

The GMM estimator using a fixed positive definite weighting matrix $W$ and sample moments $\hat{\pi}(\theta) \equiv P_n \pi (\cdot, \theta)$ is given by

$$\hat{\theta}_n \equiv \arg \min_{\theta \in \Theta} Q_n (\theta) = \frac{1}{2} \hat{\pi}(\theta)' W \hat{\pi}(\theta)$$

For $\theta^* \equiv \arg \min_{\theta \in \Theta} Q (\theta) = \frac{1}{2} \pi (\theta)' W \pi (\theta)$, where $\pi (\theta) \equiv P \pi (\cdot, \theta)$, it is well known from standard results in Newey and McFadden (1994) that for correctly specified models where $\pi (\theta^*) = 0$, $\sqrt{n} \left( \hat{\theta}_n - \theta^* \right) \leadsto -(G'WG)^{-1} G'W N(0, P \pi (\cdot, \theta^*) \pi (\cdot, \theta^*))$, where $G = \partial \pi (\theta^*) / \partial \theta$.

Under model misspecification, the asymptotic distribution differs depending on whether the model is smooth or nonsmooth. For smooth models that are globally misspecified in the sense that $\pi (\theta^*) = c$ for a vector of fixed constants $c \neq 0$, Hall and Inoue (2003) showed that $\sqrt{n} \left( \hat{\theta}_n - \theta^* \right) \leadsto N (0, H^{-1} \Omega H^{-1})$ where

$$\Sigma_{11} = P (\pi (\cdot, \theta^*) - \pi (\theta^*)) (\pi (\cdot, \theta^*) - \pi (\theta^*))'$$
$$\Sigma_{12} = P (\pi (\cdot, \theta^*) - \pi (\theta^*)) \pi (\theta^*)' W \left( \partial \pi (\cdot, \theta^*) / \partial \theta - G \right)$$
$$\Sigma_{21} = P \left( \partial \pi (\cdot, \theta^*) / \partial \theta - G \right)' W \pi (\theta^*) (\pi (\cdot, \theta^*) - \pi (\theta^*))'$$
$$\Sigma_{22} = P \left( \partial \pi (\cdot, \theta^*) / \partial \theta - G \right)' W \pi (\theta^*) \pi (\theta^*)' W \left( \partial \pi (\cdot, \theta^*) / \partial \theta - G \right)$$

$$\Omega = G' W \Sigma_{11} W G + \Sigma_{22} + G' W \Sigma_{12} + \Sigma_{21} W G$$
\[
\bar{H} = G'WG + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{jk} \pi_k(\theta^*) H_j
\]

where for each \(j = 1, \ldots, m\), define \(H_j = \frac{\partial^2}{\partial \theta \partial \theta'} \pi_j(\theta^*)\).

For globally misspecified nonsmooth GMM estimators, we will show that \(n^{1/3} \left( \hat{\theta}_n - \theta^* \right) \rightsquigarrow J\), where

\[
J = \arg \min_{h \in \mathbb{R}^d} \left\{ \pi(\theta^*)' W Z_0(h) + \frac{1}{2} h' \bar{H} h \right\}
\]

For \(g(\cdot, \theta) = \pi(\cdot, \theta) - \pi(\cdot, \theta^*)\), \(Z_0(h)\) is a mean zero Gaussian process with covariance kernel

\[
\Sigma_{1/2}(s, t) = \lim_{\alpha \to \infty} \alpha P g \left( \frac{s}{\alpha} \right) g \left( \frac{t}{\alpha} \right)'
\]

Our goal is to develop a rate-adaptive bootstrap procedure to consistently estimate the limiting distribution of the GMM estimator regardless of whether the model is correctly or incorrectly specified, smooth or nonsmooth. In other words, we do not need to know the rate of convergence of the GMM estimator when using the rate-adaptive bootstrap to construct asymptotically valid confidence intervals for \(\theta^*\). The rate-adaptive bootstrap estimate in the case of a fixed weighting matrix is

\[
\hat{\theta}_n^* = \arg \min_{\theta \in \Theta} \left\{ \hat{\pi}(\hat{\theta}_n)' W (P_n^* - P_n) \left( \pi(\cdot, \theta) - \pi(\cdot, \hat{\theta}_n) \right) + \frac{1}{2} (\theta - \hat{\theta}_n)' \left( \hat{G} W G + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{jk} \hat{\pi}_k(\hat{\theta}) \hat{H}_j \right) (\theta - \hat{\theta}_n) + (\theta - \hat{\theta}_n)' \hat{G} W (P_n^* - P_n) \hat{\pi}(\cdot, \hat{\theta}_n) \right\}
\]

For \(\gamma \in \{1/3, 1/2\}\), we will show that the limiting distribution of \(n^{\gamma} \left( \hat{\theta}_n^* - \hat{\theta}_n \right)\) coincides with the limiting distribution of \(n^{\gamma} \left( \hat{\theta}_n - \theta^* \right)\). We do not need to know the value of \(\gamma\) in order to form asymptotically valid confidence intervals for \(\theta^*\) using the empirical distribution of \(\hat{\theta}_n^* - \hat{\theta}_n\). The following steps illustrate how to use the rate-adaptive bootstrap to form asymptotically valid intervals for \(\theta^*\) if we use the multinomial bootstrap empirical measure \(P_n^* \equiv \frac{1}{n} \sum_{i=1}^{n} \mathbb{W}_i \delta_{X_i}\) for the multinomial vector \(\mathbb{W}_n = (\mathbb{W}_n, \ldots, \mathbb{W}_m)\) with probabilities \((1/n, \ldots, 1/n)\) and number of trials \(n\).

1. Compute \(\hat{\theta}_n, \hat{\pi}(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} \pi(X_i, \hat{\theta}_n)\), \(\hat{G}, \hat{H}\).

2. Repeat for \(B\) bootstrap iterations: draw a bootstrap sample \(X_1^*, \ldots, X_n^*\) and compute

\[
\hat{\theta}_n^* = \arg \min_{\theta \in \Theta} \left\{ \hat{\pi}(\hat{\theta}_n)' W \left( \frac{1}{n} \sum_{i=1}^{n} \left( \pi(X_i^*, \theta) - \pi(X_i, \hat{\theta}_n) \right) \right) - \frac{1}{n} \sum_{i=1}^{n} \left( \pi(X_i, \theta) - \pi(X_i, \hat{\theta}_n) \right) \right\}
\]
3. Compute the $1 - \alpha/2$ and $\alpha/2$ percentiles of the empirical distribution of $\hat{\theta}_n - \hat{\theta}_n$. Call them $c_{1-\alpha/2}$ and $c_{\alpha/2}$.

A $1 - \alpha$ two-sided equal-tailed confidence interval for $\theta^*$ can be formed by

$$[\hat{\theta}_n - c_{1-\alpha/2}, \hat{\theta}_n - c_{\alpha/2}]$$

3 Nonsmooth misspecified GMM using a fixed weighting matrix

Throughout the paper, we will impose the following assumptions for different values of $\gamma$ and $\rho$ depending on the rate of convergence of $\hat{\theta}_n$.

**Assumption 1.** For $\hat{Q}_n(\theta) \equiv \frac{1}{2} P_n \pi (\cdot, \theta)^T WP_n \pi (\cdot, \theta)$ and $Q(\theta) \equiv \frac{1}{2} P \pi (\cdot, \theta)^T WP \pi (\cdot, \theta)$, suppose the following conditions are satisfied for some $\gamma \in (\frac{1}{2}, 1)$:

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

(ii) For every open set $C$ that contains $\theta^*$, $\inf_{\theta \in C} Q(\theta) > Q(\theta^*)$.

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

**Assumption 2.** Let $g(\cdot, \theta) \equiv \pi(\cdot, \theta) - \pi(\cdot, \theta^*)$ satisfy the following conditions for some $\rho \in (0, \frac{3}{2})$ and $\gamma = \frac{1}{2(2-\rho)}$:

(i) There exists a $R_0 > 0$ such that the class of functions $G_R = \{g(\cdot, \theta) : \| \theta - \theta^* \| \leq R \}$ is uniformly manageable with envelope function $G_R(\cdot) \equiv \sup_{g \in G_R} |g(\cdot, \theta)|$ for all $R \leq R_0$.

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

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

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

(v) There exists a $0 < R_0 < 1$ such that $PG^2_R = O(R^{2\rho})$ for all $R \leq R_0$. 


(vi) For each \( \eta > 0 \), there exists a \( K \) such that \( PG_R^21\{G_R > K\} < \eta R^{2\rho} \) for \( R \to 0 \).

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

We first show that globally misspecified nonsmooth GMM estimators exhibit cubic-root rate of convergence. To simplify exposition we first consider a fixed weighting matrix \( W \). Later in section 5, we will consider estimated weighting matrices. The cubic-root rate of convergence is obtained when assumptions 1 and 2 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 we have non-directionally differentiable moments. Other types of nonsmooth moments that are directionally differentiable do not have this linear rate of decay and therefore retain the \( \sqrt{n} \) root rate of convergence. See section 6 for examples that distinguish between different types of nonsmooth moments.

**Theorem 1.** Suppose \( \pi (\theta^*) = c \) for a vector of fixed constants \( c \neq 0 \) and that assumptions 1-2 are satisfied for \( \gamma = 1/3 \) and \( \rho = 1/2 \). Then, \( \hat{\theta}_n - \theta^* = o_P(1) \) and

\[
n^{1/3} (\hat{\theta}_n - \theta^*) \to \arg \min_{h \in \mathbb{R}^d} \left\{ \pi (\theta^*)' W Z_{0,1/2} (h) + \frac{1}{2} h' H h \right\}
\]

\[
H = G' W G + \sum_{j=1}^m \sum_{k=1}^m W_{jk} \pi_k (\theta^*) H_j
\]

where \( Z_{0,1/2} (h) \) is a mean zero Gaussian process with covariance kernel

\[
\Sigma_{1/2} (s, t) = \lim_{\alpha \to \infty} \alpha P g \left( \cdot, \theta^* + \frac{s}{\alpha} \right) g \left( \cdot, \theta^* + \frac{t}{\alpha} \right)'
\]

The next theorem shows that the globally misspecified GMM estimator is \( \sqrt{n} \)-consistent when the moments are smooth and when the moments are nonsmooth but remain directionally differentiable.

**Theorem 2.** Suppose \( \pi (\theta^*) = c \) for a vector of fixed constants \( c \neq 0 \) and that assumptions 1-2 are satisfied for \( \gamma = 1/2 \) and \( \rho = 1 \), and \( g (\cdot, \theta) \) is Lipschitz continuous in \( \theta \) with a stochastically bounded Lipschitz constant. Suppose that for each \( \epsilon > 0 \) and \( t \in \mathbb{R}^d \),

\[
\lim_{n \to \infty} P \left\{ \left\| \frac{g \left( \cdot, \theta^* + \frac{1}{\sqrt{n}} \right)}{\pi \left( \cdot, \theta^* \right)} \right\|^2 1 \left\{ \left\| \frac{g \left( \cdot, \theta^* + \frac{1}{\sqrt{n}} \right)}{\pi \left( \cdot, \theta^* \right)} \right\|^2 > \epsilon \sqrt{n} \right\} = 0
\]

Then \( \hat{\theta}_n - \theta^* = o_P(1) \) and

\[
n^{1/2} (\hat{\theta}_n - \theta^*) \to \arg \min_{h \in \mathbb{R}^d} \left\{ \pi (\theta^*)' W Z_{0,1} (h) + h' G' W U_0 + \frac{1}{2} h' H h \right\}
\]

where \( U_0 \sim N \left( 0, P (\pi (\cdot, \theta^*) - \pi (\theta^*)) (\pi (\cdot, \theta^*) - \pi (\theta^*))' \right) \) and \( Z_{0,1} (h) \) is a mean zero Gaussian process with covariance kernel \( \Sigma_1 (s, t) = \lim_{\alpha \to \infty} \alpha^2 P g \left( \cdot, \theta^* + \frac{s}{\alpha} \right) g \left( \cdot, \theta^* + \frac{t}{\alpha} \right)' \).
4 Rate-Adaptive Bootstrap for fixed weighting matrix

We impose the following envelope integrability assumption in order to show that \( n^{\gamma} \left( \hat{\theta}_n - \theta^* \right) \) and \( n^{\gamma} \left( \hat{\theta}_n^* - \hat{\theta}_n \right) \) have the same limiting distribution. The assumption is needed to show bootstrap equicontinuity results so that both the localized empirical process and its bootstrap analog converge weakly to the same limiting process.

**Assumption 3.** For every \( \rho \in \left( 0, \frac{3}{2} \right) \) and \( \gamma = \frac{1}{2(2-\rho)} \), define \( m(\cdot, \theta, h) \equiv n^{\gamma \rho} \left( \pi(\cdot; \theta + \frac{h}{n^{\gamma}}) - \pi(\cdot; \theta) \right) \). For any \( \epsilon \rightarrow 0 \),

\[
\lim_{\lambda \rightarrow \infty} \lim_{n \rightarrow \infty} \sup_{t \geq \lambda} P \left\{ \sup_{h \in \mathbb{R}, \|\theta - \theta^*\| \leq \epsilon_n} \left| m(\cdot, \theta, h) - m(\cdot, \theta^*, h) \right| \right\} > t = 0
\]

Furthermore,

\[
\lim_{\lambda \rightarrow \infty} \lim_{n \rightarrow \infty} \sup_{t \geq \lambda} P \left\{ \sup_{\|\theta - \theta^*\| \leq \epsilon_n} \left| g(\cdot, \theta) - g(\cdot, \theta^*) \right| \right\} > t = 0
\]

**Theorem 3.** Suppose assumptions 1-3 are satisfied, \( \hat{G} \xrightarrow{p} G \), and \( \hat{H}_j \xrightarrow{p} H_j \) for \( j = 1 \ldots m \).

For correctly specified models,

\[
\sqrt{n} \left( \hat{\theta}_n^* - \hat{\theta}_n \right) \xrightarrow{w} (G'WG)^{-1} G'WN \left( 0, P\pi(\cdot, \theta^*) \pi(\cdot, \theta^*)' \right)
\]

For misspecified smooth models or models where the moments are directionally differentiable,

\[
\sqrt{n} \left( \hat{\theta}_n^* - \hat{\theta}_n \right) \xrightarrow{w} N \left( 0, \bar{H}^{-1} \Omega \bar{H}^{-1'} \right)
\]

\[
\Omega = G'WG \Sigma_{11}WG + \Sigma_{22} + G'WG \Sigma_{12} + \Sigma_{21}WG
\]

\[
H = G'WG + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{jk} \pi_k (\theta^*) H_j, \quad H_j = \frac{\partial^2}{\partial \theta \partial \theta'} \pi_j (\theta^*') \quad \text{for } j = 1, \ldots, m
\]

\[
\Sigma_{11} = P \left( \pi(\cdot, \theta^*) - \pi(\theta^*) \right) \left( \pi(\cdot, \theta^*) - \pi(\theta^*) \right)'
\]

\[
\Sigma_{12} = P \left( \pi(\cdot, \theta^*) - \pi(\theta^*) \right) \pi(\theta^*)' W \left( \frac{\partial}{\partial \theta} \pi(\cdot, \theta^*) - G \right)
\]

\[
\Sigma_{21} = P \left( \frac{\partial}{\partial \theta} \pi(\cdot, \theta^*) - G \right)' W \pi(\theta^*) \left( \pi(\cdot, \theta^*) - \pi(\theta^*) \right)'
\]

\[
\Sigma_{22} = P \left( \frac{\partial}{\partial \theta} \pi(\cdot, \theta^*) - G \right)' W \pi(\theta^*) \pi(\theta^*)' W \left( \frac{\partial}{\partial \theta} \pi(\cdot, \theta^*) - G \right)
\]

For globally misspecified nonsmooth models where assumptions 1-2 are satisfied for \( \gamma = 1/3, \rho = 1/2 \),

\[
n^{1/3} \left( \hat{\theta}_n^* - \hat{\theta}_n \right) \xrightarrow{w} \arg \min_{h} \left\{ \pi(\theta^*)' W \mathcal{Z}_{0,1/2}(h) + \frac{1}{2} h' \bar{H} h \right\}
\]
For globally misspecified models where the assumptions in Theorem 2 hold,

\[ n^{1/2} \left( \hat{\theta}_n - \hat{\theta}_n \right) \xrightarrow{P} \arg\min_{h \in \mathbb{R}^d} \left\{ \pi (\theta^*)' W Z_{0,1} (h) + h' G' W U_0 + \frac{1}{2} h' \bar{H} h \right\} \]

5 The case of an estimated weighting matrix

We now consider the case of an estimated weighting matrix. First we show that nonsmooth misspecified GMM has a different asymptotic distribution depending on the rate at which the estimated weighting matrix converges to its probability limit. Next we show that the rate-adaptive bootstrap needs to be modified to include an additional term to capture the variation between the estimated weighting matrix and its probability limit.

Note that we need to redefine the pseudo-true parameter to be \( \hat{\theta}^* = \arg\min_{\theta \in \Theta} \pi (\theta)' W (\theta^*_1) \pi (\theta) \) where \( W (\theta^*_1) \) depends on the 1-step GMM pseudo-true parameter using some fixed weighting matrix \( W_1: \theta^*_1 = \arg\min_{\theta \in \Theta} \pi (\theta)' W_1 \pi (\theta) \). The estimated weighting matrix \( W_n \left( \hat{\theta}_1 \right) \) will depend on the 1-step GMM estimator \( \hat{\theta}_1 = \arg\min_{\theta \in \Theta} \pi (\theta)' W_1 \pi (\theta) \). The next theorem demonstrates that the globally misspecified 2-step GMM estimator \( \hat{\theta}_n = \arg\min_{\theta \in \Theta} \pi (\theta)' W_n \left( \hat{\theta}_1 \right) \pi (\theta) \) with non-directionally differentiable \( \pi (\cdot, \theta) \) will have a different asymptotic distribution depending on the rate at which \( W_n \left( \hat{\theta}_1 \right) \) converges to \( W (\theta^*_1) \). To simplify notation, we will use \( W_n \) to refer to \( W_n \left( \hat{\theta}_1 \right) \) and \( W \) to refer to \( W \left( \theta^*_1 \right) \).

**Theorem 4.** Suppose \( \pi (\theta^*) = c \) for a vector of fixed constants \( c \neq 0 \) and that assumptions 1-2 are satisfied for \( \gamma = 1/3 \) and \( \rho = 1/2 \).

If \( W_n - W = O_p (n^{-1/3}) \), then \( \hat{\theta}_n - \theta^* = o_P (1) \) and for \( \bar{Z}_0 (h) \equiv \pi (\theta^*)' W Z_{0,1/2} (h) \),

\[ n^{1/3} \left( \hat{\theta}_n - \theta^* \right) \rightsquigarrow \arg\min_{h \in \mathbb{R}^d} \left\{ \bar{Z}_0 (h) + \frac{1}{2} h' \bar{H} h \right\} \]

If \( W_n - W = O_p (n^{-1/3}) \) and \( \left( \pi (\theta^*)' W_n^{2/3} (P_n - P) g (\cdot, \theta^* + n^{-1/3} h) \right) \rightsquigarrow \left( \bar{Z}_0 (h) \right) \),

then \( \hat{\theta}_n - \theta^* = o_P (1) \) and

\[ n^{1/3} \left( \hat{\theta}_n - \theta^* \right) \rightsquigarrow \arg\min_{h \in \mathbb{R}^d} \left\{ \bar{Z}_0 (h) + h' G' W_0 + \frac{1}{2} h' \bar{H} h \right\} \]

Now we consider the case of 2-step GMM with either smooth moments or nonsmooth moments that remain directionally differentiable. The rate of convergence of the 2-step GMM estimator will depend on the rate of convergence of the weighting matrix. If the weighting matrix is \( \sqrt{n} \)-consistent, then the 2-step GMM estimator will be \( \sqrt{n} \)-consistent as well. However, if the weighting matrix converges at a slower rate, then the 2-step GMM estimator will also converge at a slower rate.

The following assumption deals with \( \sqrt{n} \)-consistently estimated weighting matrices.
Assumption 4. The weighting matrix $W_n$ satisfies $\sqrt{n}(W_n - W) = \sqrt{n}(P_n - P) \phi(\cdot, \theta_1^\ast) + o_p(1)$ where $\theta_1^\ast$ is the probability limit of the 1-step GMM estimate using a fixed weighting matrix, $\phi(\theta_1^\ast) \equiv P\phi(\cdot, \theta_1^\ast) < \infty$ and $P\phi(\cdot, \theta_1^\ast) \phi(\cdot, \theta_1^\ast)' < \infty$, and the bootstrapped weighting matrix $W_n^\ast$ has the same representation $\sqrt{n}(W_n^\ast - W_n) = \sqrt{n}(P_n^\ast - P_n) \phi(\cdot, \theta_1^\ast) + o_p(1)$. Additionally, for each $\epsilon > 0$ and $t \in \mathbb{R}^d$,

$$
\lim_{n \to \infty} \left\| \left( \begin{array}{l}
g(\cdot, \theta^\ast + \frac{\epsilon}{\sqrt{n}}) \\
\phi(\cdot, \theta^\ast)
\end{array} \right) \right\|^2 \left\{ \left\| \left( \begin{array}{l}
g(\cdot, \theta^\ast + \frac{\epsilon}{\sqrt{n}}) \\
\phi(\cdot, \theta^\ast)
\end{array} \right) \right\| > \epsilon \sqrt{n} \right\} = 0
$$

Theorem 5. Suppose $\pi(\theta^\ast) = c$ for a vector of fixed constants $c \neq 0$ and that assumptions 1-2 are satisfied for $\gamma = 1/2$ and $\rho = 1$, and $g(\cdot, \theta)$ is Lipschitz continuous in $\theta$ with a stochastically bounded Lipschitz constant. If $W_n - W = O_p(n^{-1/2})$ and assumption 4 is satisfied, then $\hat{\theta}_n - \theta^\ast = o_p(1)$ and

$$
\sqrt{n}(\hat{\theta}_n - \theta^\ast) \rightsquigarrow \arg \min_{h \in \mathbb{R}^d} \left\{ \pi(\theta^\ast)' W \mathcal{Z}_{0,1}(h) + h' G' W' U_0 + h' G' \Phi_0 \pi(\theta^\ast) + \frac{1}{2} h' \tilde{H} h \right\}
$$

where $\Phi_0 \pi(\theta^\ast) \sim N(0, P(\phi(\cdot, \theta_1^\ast) - \phi(\theta_1^\ast)) \pi(\theta^\ast) \pi(\theta^\ast)' (\phi(\cdot, \theta_1^\ast) - \phi(\theta_1^\ast))')$. If $W_n - W = o_p(n^{-1/2})$, then

$$
n^{1/2}(\hat{\theta}_n - \theta^\ast) \rightsquigarrow \arg \min_{h \in \mathbb{R}^d} \left\{ \pi(\theta^\ast)' W \mathcal{Z}_{0,1}(h) + h' G' W' U_0 + \frac{1}{2} h' \tilde{H} h \right\}
$$

When we use an estimated weighting matrix, we have to modify the rate-adaptive bootstrap estimate to include an additional term that accounts for the additional variation induced by estimating the weighting matrix:

$$
\hat{\theta}_n^* = \arg \min_{\theta \in \Theta} \left\{ \hat{\pi}(\hat{\theta}_n)^' W_n (P_n^* - P_n) \left( \pi(\cdot, \theta) - \pi(\cdot, \hat{\theta}_n) \right) + \frac{1}{2} \left( \theta - \hat{\theta}_n \right)^' \hat{G}' W_n \hat{G} + \sum_{j=1}^m \sum_{k=1}^m W_{n,jk} \hat{\pi}_k \left( \hat{\theta}_n \right) \hat{H}_j \right\} \left( \theta - \hat{\theta}_n \right) \text{ (5.1)}
$$

where $W_n^* = W_n^* \left( \hat{\theta}_1^\ast \right)$ could potentially depend on the rate-adaptive bootstrap estimator $\hat{\theta}_1^\ast$ using a fixed weighting matrix $W_1$:

$$
\hat{\theta}_1^\ast = \arg \min_{\theta \in \Theta} \left\{ \hat{\pi}(\hat{\theta}_n)^' W_1 (P_n^* - P_n) \left( \pi(\cdot, \theta) - \pi(\cdot, \hat{\theta}_n) \right) + \frac{1}{2} \left( \theta - \hat{\theta}_n \right)^' \hat{G}' W_1 \hat{G} + \sum_{j=1}^m \sum_{k=1}^m W_{1,jk} \hat{\pi}_k \left( \hat{\theta}_n \right) \hat{H}_j \right\} \left( \theta - \hat{\theta}_n \right) \text{ (5.2)}
$$
\( \left( \theta - \hat{\theta}_n \right)' G' W_1 (P_n^* - P_n) \pi \left( \cdot, \hat{\theta}_n \right) \)

The following theorem shows that the rate-adaptive bootstrap is consistent for the limiting distribution of the 2-step GMM estimator under correct specification and different scenarios of global misspecification.

**Theorem 6.** Suppose assumptions 1-3 are satisfied, \( \hat{G} \overset{p}{\rightarrow} G \), and \( \hat{H}_j \overset{p}{\rightarrow} H_j \) for \( j = 1\ldots m \). For correctly specified models,

\[
\sqrt{n} \left( \hat{\theta}_n - \hat{\theta}_n \right) \overset{p}{\rightarrow} \left( G' W G \right)^{-1} G' W N \left( 0, P \pi (\cdot, \theta^*) \pi (\cdot, \theta^*)' \right)
\]

For globally misspecified smooth models or models with directionally differentiable moments where the weighting matrix \( W_n \) satisfies \( \sqrt{n} (W_n - W) = \sqrt{n} (P_n - P) \phi (\cdot, \theta^*_1) + o_p(1) \) and the bootstrapped weighting matrix \( W_n^* \) satisfies \( \sqrt{n} (W_n^* - W_n) = \sqrt{n} (P_n^* - P_n) \phi (\cdot, \theta^*_1) + o_p(1) \),

\[
\sqrt{n} \left( \hat{\theta}_n^* - \hat{\theta}_n \right) \overset{p}{\rightarrow} N \left( 0, \bar{H}^{-1} \Omega W \bar{H}^{-1} \right)
\]

\( \Omega_W = G' W \Sigma_{11} W G + \Sigma_{22} + G' W \Sigma_{12} + \Sigma_{21} W G + G' \Sigma_{33} G + G' \Sigma_{31} W G + G' \Sigma_{32} G + G' \Sigma_{32} \)

\( \bar{H} = G' W G + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{jk} \pi_k (\theta^*) H_j, \quad H_j = \frac{\partial^2}{\partial \theta \partial \theta'} \pi_j (\theta^*) \) for \( j = 1, \ldots, m \)

where \( \Sigma_{11}, \Sigma_{12}, \Sigma_{21}, \) and \( \Sigma_{22} \) are the same as before and

\[
\Sigma_{13} = P (\pi \left( \cdot, \theta^* \right) - \pi (\theta^*))' \pi (\theta^*)' (\phi (\cdot, \theta^*_1) - \phi (\theta^*_1))'
\]

\[
\Sigma_{31} = P (\phi (\cdot, \theta^*_1) - \phi (\theta^*_1))' \pi (\theta^*) (\pi (\cdot, \theta^*) - \pi (\theta^*))'
\]

\[
\Sigma_{23} = P \left( \frac{\partial}{\partial \theta} \pi \left( \cdot, \theta^* \right) - G \right)' W \pi (\theta^*) \pi (\theta^*)' (\phi (\cdot, \theta^*_1) - \phi (\theta^*_1))'
\]

\[
\Sigma_{32} = P (\phi (\cdot, \theta^*_1) - \phi (\theta^*_1))' \pi (\theta^*) \pi (\theta^*)' W \left( \frac{\partial}{\partial \theta} \pi \left( \cdot, \theta^* \right) - G \right)
\]

\[
\Sigma_{33} = P (\phi (\cdot, \theta^*_1) - \phi (\theta^*_1))' \pi (\theta^*) \pi (\theta^*)' (\phi (\cdot, \theta^*_1) - \phi (\theta^*_1))'
\]

For globally misspecified nonsmooth models where assumptions 1 and 2 are satisfied for \( \gamma = 1/3 \) and \( \rho = 1/2 \), if \( W_n - W = o_p (n^{-1/3}) \) and \( W_n^* - W_n = o_p^* (n^{-1/3}) \), then

\[
n^{1/3} \left( \hat{\theta}_n^* - \hat{\theta}_n \right) \overset{p}{\rightarrow} \arg \min_h \left\{ \pi (\theta^*)' W Z_{0,1/2} (h) + \frac{1}{2} h' \bar{H} h \right\}
\]

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

\[
\left( \pi (\theta^*)' W n^{2/3} (P_n^* - P_n) g (\cdot, \theta^* + n^{-1/3} h) \right) \overset{p}{\rightarrow} \left( \pi (\theta^*)' W Z_{0,1/2} (h) \right),
\]

then

\[
n^{1/3} \left( \hat{\theta}_n^* - \hat{\theta}_n \right) \overset{p}{\rightarrow} \arg \min_h \left\{ \pi (\theta^*)' W Z_{0,1/2} (h) + h' G' W_0 + \frac{1}{2} h' \bar{H} h \right\}
\]
For globally misspecified nonsmooth models where assumptions 1 and 2 are satisfied for \( \gamma = 1/2 \) and \( \rho = 1 \), if \( W_n - W = o_p(n^{-1/2}) \) and \( W^*_n - W_n = o^*_p(n^{-1/2}) \), then

\[
\sqrt{n} \left( \hat{\theta}_n - \theta_n \right) \xrightarrow{P} \arg \min_{h \in \mathbb{R}^d} \left\{ h'G'WU_0 + \pi (\theta^*)' W \mathbb{Z}_{0,1} (h) + \frac{1}{2} h' \Phi h \right\}
\]

If instead assumption 4 is satisfied,

\[
\sqrt{n} \left( \hat{\theta}_n - \theta_n \right) \xrightarrow{P} \arg \min_{h \in \mathbb{R}^d} \left\{ h'G'WU_0 + \pi (\theta^*)' W \mathbb{Z}_{0,1} (h) + h' \Phi \pi (\theta^*) + \frac{1}{2} h' \Phi h \right\}
\]

where \( \Phi \pi (\theta^*) \sim N (0, P (\phi (\cdot, \theta^*_1) - \phi (\theta^*_1)) \pi (\theta^*) \pi (\theta^*)' (\phi (\cdot, \theta^*_1) - \phi (\theta^*_1))' \), and

\[
U_0 \sim N (0, P (\pi (\cdot, \theta^*) - \pi (\theta^*)) (\pi (\cdot, \theta^*) - \pi (\theta^*))').
\]

6 Applications

We consider applications of our rate-adaptive bootstrap procedure to the GMM formulation of instrumental variables quantile regression (Chernozhukov and Hansen (2005)), simulated method of moments (McFadden (1989) and Pakes and Pollard (1989)), and Dynamic Censored Regression (Honore and Hu (2004a)).

6.1 GMM Formulation of Instrumental Variable Quantile Regression

The moment conditions are \( \pi (\cdot, \theta) = (\tau - 1 (y_i \leq q (D_i, w_i, \theta))) z_i \), where \( y_i \) is the dependent variable, \( D_i \) is a vector of endogenous regressors, \( w_i \) is a vector of exogenous regressors, \( z_i \) is a vector of instruments, and \( q (\cdot) \) is the quantile response function, assumed to be a monotonic, twice differentiable function. Suppose the quantile response function has a single index variable, \( \hat{\mathbf{D}} \) is a vector of exogenous regressors, \( \hat{\mathbf{W}} \) is a vector of exogenous regressors, \( \hat{\mathbf{z}} \) is a vector of instruments, and \( \hat{\pi} (\cdot) \) is the quantile response function, assumed to be a monotonic, twice differentiable function. Suppose the quantile response function has a single index structure: \( q (D_i, w_i, \theta) = q (x_i \theta) \) for \( x_i \theta = [D_i, w_i] \). For \( \pi (\theta) = E (\tau - F_{y|x,z} (q (x' \theta))) \) \( z \), the Jacobian is \( G = \frac{\partial}{\partial \theta} \pi (\theta^*) = -E f_{y|x,z} (q (x' \theta^*)) zq' (x' \theta^*) x' \) and the \( j \)th element of the Hessian is \( H_j = \frac{\partial^2}{\partial \theta \partial \theta^*} \pi_j (\theta^*) = -E f_{y|x,z} (q (x' \theta^*)) z_j (q' (x' \theta^*)) x x' + E f_{y|x,z} (q (x' \theta^*)) z_j q'' (x' \theta^*) x x' \).

A crucial condition that generates cubic-root convergence in globally misspecified models with non-directionally differentiable moments is when the value of \( \rho \) that satisfies assumption 2(v) is \( \rho = 1/2 \). The class \( G_R \equiv \{ \pi (\cdot, \theta) = \pi (\cdot, \theta^*) : \| \theta - \theta^* \| \leq R \} \) has envelope function

\[
G_R (\cdot) = \sup_{\| \theta - \theta^* \| \leq R} \| \pi (\cdot, \theta) - \pi (\cdot, \theta^*) \|
\]

\[
= \sup_{\| \theta - \theta^* \| \leq R} \| z_i (1 (y_i \leq q (x_i \theta^*)) - 1 (y_i \leq q (x_i \theta))) \|
\]

Therefore,

\[
PG_R^2 \leq E \| z_i \|^2 \{ P (q (x_i \theta^* - R)) \leq y_i \leq q (x_i \theta^*)) + P (q (x_i \theta^* + R)) \leq y_i \leq q (x_i \theta^*)) \}
\]

\[
\leq 2 E \| z_i \|^2 \sup_{\theta \in \Theta} \left\{ f_{y|x,z} (q (x' \theta)) q' (x' \theta) x' \right\} R = O (R)
\]
In the case of a fixed weighting matrix, the asymptotic distribution of the IV quantile regression estimator is given in Theorem 1. We now consider the case of an estimated weighting matrix. The 2-step GMM estimator \( \hat{\theta}_n = \arg \min_{\theta} \hat{\pi}_n (\theta)' W_n \left( \hat{\theta}_1 \right) \hat{\pi}_n (\theta) \) depends on the 1-step GMM estimator \( \hat{\theta}_1 = \arg \min_{\theta} \hat{\pi}_n (\theta)' W_1 \hat{\pi}_n (\theta) \) whose probability limit is \( \theta_1^* = \arg \min_{\theta} (\theta)' W_1 \pi (\theta) \). The pseudo-true parameters are given by \( \theta^* = \arg \min_{\theta} (\theta)' W (\theta_1^*) \pi (\theta) \), where \( W (\theta_1^*) \) is the inverse of the variance-covariance matrix of the population moments

\[
W (\theta_1^*) = \left( E \left[ \pi (\cdot, \theta_1^*) \pi (\cdot, \theta_1^*)' \right] - \pi (\theta_1^*) \pi (\theta_1^*)' \right)^{-1}
\]

The last line follows from the fact that conditional on \( x_i, z_i \), \( \tau - 1 (y_i \leq q (x_i' \theta_1^*)) \) is a Bernoulli random variable that equals \( \tau - 1 \) with probability \( F_{y|x,z} (q (x_i' \theta_1^*)) \) and equals \( \tau \) with probability \( 1 - F_{y|x,z} (q (x_i' \theta_1^*)) \). Therefore,

\[
E \left[ (\tau - 1 (y_i \leq q (x_i' \theta_1^*)))^2 \bigg| z_i \right] = (\tau - 1)^2 F_{y|x,z} (q (x_i' \theta_1^*)) + \tau^2 (1 - F_{y|x,z} (q (x_i' \theta_1^*)))
\]

The estimated weighting matrix is

\[
W_n (\hat{\theta}_1) = \left( \frac{1}{n} \sum_{i=1}^{n} \left( \tau^2 + (1 - 2\tau) \hat{F}_{y|x,z} (q (x_i' \hat{\theta}_1)) \right) z_i z_i' - \hat{\pi}_n (\hat{\theta}_1) \hat{\pi}_n (\hat{\theta}_1)' \right)^{-1}
\]

Let the influence function representation of the 1-step GMM estimator be given by

\[
\sqrt{n} \left( \hat{\theta}_1 - \theta_1^* \right) = \sqrt{n} (P_n - P) \kappa (\cdot, \theta_1^*) + o_p (1)
\]

where \( \kappa (\cdot, \theta_1^*) = - (G'W_1 G)^{-1} G'W_1 \pi (\cdot, \theta_1^*) \). The influence function representation of the estimated weighting matrix is

\[
\sqrt{n} \left( W_n (\hat{\theta}_1) - W (\theta_1^*) \right) = \sqrt{n} (P_n - P) \psi (\cdot, \theta_1^*) W (\theta_1^*)^{-1} + \frac{\partial W (\theta_1^*)}{\partial \theta^*} \sqrt{n} (P_n - P) \kappa (\cdot, \theta_1^*) + o_p (1)
\]

\[
= -W (\theta_1^*)^{-1} \sqrt{n} (P_n - P) \psi (\cdot, \theta_1^*) W (\theta_1^*)^{-1} + \frac{\partial W (\theta_1^*)}{\partial \theta^*} \sqrt{n} (P_n - P) \kappa (\cdot, \theta_1^*) + o_p (1)
\]

\[
= \sqrt{n} (P_n - P) \phi (\cdot, \theta_1^*) + o_p (1)
\]
We can obtain the expression for $\sqrt{n} \left( P_n - P \right) \psi (\cdot, \theta_1^*)$ by rewriting

$$
\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} \left( \pi (\cdot, \theta_1^*) \pi (\cdot, \theta_1^*)' - E \left[ \pi (\cdot, \theta_1^*) \pi (\cdot, \theta_1^*)' \right] \right) - \left( \hat{\pi}_n (\theta_1^*) \hat{\pi}_n (\theta_1^*)' - \pi (\theta_1^*) \pi (\theta_1^*)' \right) \right)
$$

$$
= \sqrt{n} \left( \tau^2 \frac{1}{n} \sum_{i=1}^{n} (z_i z_i' - E [z_i z_i']) + (1 - 2\tau) \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( 1 (y_j \leq q (x_i^* \theta_1^*)) z_i z_i' - E \left[ F_{y|x,z} (q (x_i^* \theta_1^*)) z_i z_i' \right] \right) \right)
$$

$$
- \left( \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \frac{1}{2} - 1 (y_i \leq q (x_i^* \theta_1^*)) \left( \frac{1}{2} - 1 (y_j \leq q (x_j^* \theta_1^*)) \right) z_i z_i' - \pi (\theta_1^*) \pi (\theta_1^*)' \right) \right)
$$

$$
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tau^2 (z_i z_i' - E [z_i z_i']) + (1 - 2\tau) \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E [g (w_i, w_j) | w_i] + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E [g (w_i, w_j) | w_j] \right)
$$

$$
- \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E [h (w_i, w_j) | w_i] + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E [h (w_i, w_j) | w_j] \right) + o_p(1)
$$

$$
= \sqrt{n} (P_n - P) \psi (\cdot, \theta_1^*) + o_p(1)
$$

where the second to last equality follows from the fact that $\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} g (w_i, w_j)$ and $\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} h (w_i, w_j)$ are non-degenerate V-statistics which have the same asymptotic distribution as the non-degenerate U-statistics $\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g (w_i, w_j)$ and $\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h (w_i, w_j)$ which have the following the decompositions:

$$
\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g (w_i, w_j) = \frac{1}{n} \sum_{i=1}^{n} E [g (w_i, w_j) | w_i] + \frac{1}{n} \sum_{j=1}^{n} E [g (w_i, w_j) | w_j] + o_p \left( \frac{1}{\sqrt{n}} \right)
$$

$$
\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h (w_i, w_j) = \frac{1}{n} \sum_{i=1}^{n} E [h (w_i, w_j) | w_i] + \frac{1}{n} \sum_{j=1}^{n} E [h (w_i, w_j) | w_j] + o_p \left( \frac{1}{\sqrt{n}} \right)
$$

The bootstrapped weighting matrix is computed using the multinomial bootstrap and an initial rate-adaptive bootstrap estimator $\hat{\theta}_1^*$ computed using a fixed weighting matrix.

$$
W_n^* \left( \hat{\theta}_1^* \right) = \left( \tau^2 \frac{1}{n} \sum_{i=1}^{n} z_i^* z_i'' + (1 - 2\tau) \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( 1 (y_j \leq q (x_i^* \hat{\theta}_1^*)) \right) z_i^* z_i'' - \pi_n^* \left( \hat{\theta}_1^* \right) \pi_n^* \left( \hat{\theta}_1^* \right)' \right)^{-1}
$$

We can show that the bootstrapped weighting matrix has the same influence function representation:

$$
\sqrt{n} \left( W_n^* \left( \hat{\theta}_1^* \right) - W_n (\theta_1^*) \right)
$$
\[ \sqrt{n} \left( W_n^* \left( \hat{\theta}_i^* \right) - W_n \left( \theta_i^* \right) \right) = \sqrt{n} \left( W_n^* \left( \theta_i^* \right) - W_n \left( \theta_i^* \right) \right) + \frac{\partial W_n^* \left( \theta_i^* \right)}{\partial \theta} \sqrt{n} \left( \hat{\theta}_i^* - \theta_i^* \right) - \frac{\partial W_n \left( \theta_i^* \right)}{\partial \theta} \sqrt{n} \left( \hat{\theta}_i - \theta_i \right) + o_p(1) \]

\[ = \sqrt{n} \left( W_n^* \left( \theta_i^* \right) - W_n \left( \theta_i^* \right) \right) + \frac{\partial W \left( \theta_i^* \right)}{\partial \theta} \sqrt{n} \left( \hat{\theta}_i^* - \theta_i^* \right) + o_p(1) \]

\[ = -W \left( \theta_i^* \right)^{-1} \sqrt{n} \left( P_n^* - P_n \right) \psi \left( \cdot, \theta_i^* \right) W \left( \theta_i^* \right)^{-1} + \frac{\partial W \left( \theta_i^* \right)}{\partial \theta} \sqrt{n} \left( P_n^* - P_n \right) \kappa \left( \cdot, \theta_i^* \right) + o_p(1) \]

\[ = \sqrt{n} \left( P_n^* - P_n \right) \phi \left( \cdot, \theta_i^* \right) + o_p(1) \]

where \( \sqrt{n} \left( W_n^* \left( \theta_i^* \right) - W_n \left( \theta_i^* \right) \right) = -W \left( \theta_i^* \right)^{-1} \sqrt{n} \left( P_n^* - P_n \right) \psi \left( \cdot, \theta_i^* \right) W \left( \theta_i^* \right)^{-1} + o_p(1) \) follows from the consistency of the multinomial bootstrap for V-statistics of order 2 (see Theorem 3.1 in Bickel and Freedman (1981)) as long as \( \int g^2 \left( w_1, w_2 \right) dF \left( w_1 \right) dF \left( w_2 \right) < \infty \), \( \int g^2 \left( w, w \right) dF \left( w \right) < \infty \), \( \int h^2 \left( w_1, w_2 \right) dF \left( w_1 \right) dF \left( w_2 \right) < \infty \), and \( \int h^2 \left( w, w \right) dF \left( w \right) < \infty \).

### 6.2 Simulated Method of Moments

The moment conditions are \( \pi \left( \cdot, \theta \right) = \left( y_i - \frac{1}{S} \sum_{s=1}^{S} 1 \left( h \left( x_i^s \theta + \eta_i s > 0 \right) \right) z_i, \right. \) where \( y_i \in \{0, 1\} \) is the choice of individual \( i \), \( z_i \) is a vector of instruments, \( x_i \) is a vector of covariates, \( h \left( \cdot \right) \) is a monotonic, twice differentiable function, and \( \{ \eta_i s \}_{s=1}^{S} \) are individual \( i \)'s simulation draws from a mean zero, symmetric distribution. For \( \pi \left( \theta \right) = E \left( y - F_{\eta | x, z} \left( h \left( x \theta \right) \right) \right) z \), the Jacobian is \( G = \frac{\partial}{\partial \theta} \pi \left( \theta^* \right) = -E f_{\eta | x, z} \left( h \left( x \theta^* \right) \right) z h' \left( x \theta^* \right) x \) and the \( j \)th element of the Hessian is \( H_j = \frac{\partial^2}{\partial \theta \partial \theta} \pi_j \left( \theta^* \right) = -E f_{\eta | x, z} \left( h \left( x \theta^* \right) \right) z_j h' \left( x \theta^* \right)^2 xx' + E f_{\eta | x, z} \left( h \left( x \theta^* \right) \right) z_j h'' \left( x \theta^* \right) xx' \).

We now verify that the value of \( \rho \) that satisfies assumption 2(v) is \( \rho = 1/2 \). The class \( G_R \equiv \{ \pi \left( \cdot, \theta \right) - \pi \left( \cdot, \theta^* \right) : \| \theta - \theta^* \| \leq R \} \) has envelope function

\[ G_R \left( \cdot \right) = \sup_{\| \theta - \theta^* \| \leq R} \left\| \pi \left( \cdot, \theta \right) - \pi \left( \cdot, \theta^* \right) \right\| \]

\[ = \sup_{\| \theta - \theta^* \| \leq R} \left\| \frac{1}{S} \sum_{s=1}^{S} \left( 1 \left( h \left( x_i^s \theta + \eta_i s > 0 \right) - 1 \left( h \left( x_i^s \theta + \eta_i s > 0 \right) \right) \right) \right\| \]

Therefore,

\[ PC_R^2 \leq E \left| z_i \right|^2 \left( P \left( h \left( x_i^* \theta^* - R \right) \right) \right) \leq -\eta_i s \leq -h \left( x_i^s \theta^* \right) \right) + P \left( h \left( x_i^s \left( \theta^* + R \right) \right) \right) \leq \eta_i s \leq h \left( x_i^s \theta^* \right) \right) \]

\[ \leq 2E \left| z_i \right|^2 \sup_{\theta \in \Theta} \left\{ f_{\eta | x, z} \left( h \left( x \theta \right) \right) h' \left( x \theta \right) x \right\} R = O \left( R \right) \]

Just as in the previous example, the pseudo-true parameters are given by \( \theta^* = \arg \min_{\theta} \pi \left( \theta \right) W \left( \theta_i^* \right) \pi \left( \theta \right) \), where \( W \left( \theta_i^* \right) \) is the inverse of the variance-covariance matrix of the population moments:

\[ W \left( \theta_i^* \right) = \left( E \left[ \pi \left( \cdot, \theta_i^* \right) \pi \left( \cdot, \theta_i^* \right)' \right] - \pi \left( \theta_i^* \right) \pi \left( \theta_i^* \right)' \right)^{-1} \]

\[ = \left( E \left[ y_i - \frac{1}{S} \sum_{s=1}^{S} \left( 1 \left( h \left( x_i^s \theta_i^* + \eta_i s \right) > 0 \right) \right) \right] \right)^2 z_i z_i' - \pi \left( \theta_i^* \right) \pi \left( \theta_i^* \right)' \right)^{-1} \]
\[
= \left( E \left[ E \left[ \left( y_i - \frac{1}{S} \sum_{s=1}^{S} 1 \left( h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) \right) \left| x_i, z_i \right] \right] z_i z_i' \right] - \pi (\theta_1) \pi (\theta_1)' \right)^{-1}
\]

The estimated weighting matrix is

\[
W_n \left( \hat{\theta}_1 \right) = \left( \frac{1}{n} \sum_{i=1}^{n} \left( y_i - \frac{1}{S} \sum_{s=1}^{S} 1 \left( h \left( x_i^s \hat{\theta}_1 \right) + \eta_i > 0 \right) \right)^2 z_i z_i' - \hat{\pi}_n \left( \hat{\theta}_1 \right) \hat{\pi}_n \left( \hat{\theta}_1 \right)' \right)^{-1}
\]

Under the condition that \( n/S \to 0 \), we can derive an influence function representation of the estimated weighting matrix similar to the previous example:

\[
\sqrt{n} \left( W_n \left( \hat{\theta}_1 \right) - W \left( \theta_1^* \right) \right) = \sqrt{n} \left( W_n \left( \theta_1^* \right) - W \left( \theta_1^* \right) \right) + \frac{\partial W_n \left( \theta_1^* \right)}{\partial \theta} \sqrt{n} \left( \hat{\theta}_1 - \theta_1^* \right) + o_p(1)
\]

\[
= -W \left( \theta_1^* \right)^{-1} \sqrt{n} \left( P_n - P \right) \psi \left( \cdot, \theta_1^* \right) W \left( \theta_1^* \right)^{-1} + \frac{\partial W \left( \theta_1^* \right)}{\partial \theta} \sqrt{n} \left( P_n - P \right) \kappa \left( \cdot, \theta_1^* \right) + o_p(1)
\]

We can obtain the expression for \( \sqrt{n} \left( P_n - P \right) \psi \left( \cdot, \theta_1^* \right) \) by rewriting

\[
\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} \left( \pi \left( \cdot, \theta_1^* \right) \pi \left( \cdot, \theta_1^* \right)' - E \left[ \pi \left( \cdot, \theta_1^* \right) \pi \left( \cdot, \theta_1^* \right)' \right] \right) \right) = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} \left( y_i - \frac{1}{S} \sum_{s=1}^{S} 1 \left( h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) \right)^2 z_i z_i' - E \left[ \left( y_i - \frac{1}{S} \sum_{s=1}^{S} 1 \left( h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) \right)^2 z_i z_i' \right] \right)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( y_i^2 z_i z_i' - E \left[ y_i^2 z_i z_i' \right] \right) - \frac{1}{n} \sum_{i=1}^{n} \sum_{s=1}^{S} \sum_{t=1}^{S} \left( y_i h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) z_i z_i' - E \left[ y_i h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right] z_i z_i' \right)
\]

\[
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{s=1}^{S} \sum_{t=1}^{S} \sum_{s=1}^{S} \sum_{t=1}^{S} \left( \left( y_i y_j z_i z_j' - E \left[ y_i y_j z_i z_j' \right] \right) \right) + \left( h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) \left( h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) z_i z_i' - E \left[ \left( h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) \right] z_i z_i' \right)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( y_i^2 z_i z_i' - E \left[ y_i^2 z_i z_i' \right] \right) - \frac{1}{n} \sum_{i=1}^{n} \sum_{s=1}^{S} \sum_{t=1}^{S} \left( y_i h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) z_i z_i' - E \left[ y_i h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right] z_i z_i' \right)
\]

\[
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{s=1}^{S} \sum_{t=1}^{S} \sum_{s=1}^{S} \sum_{t=1}^{S} \left( \left( y_i y_j z_i z_j' - E \left[ y_i y_j z_i z_j' \right] \right) \right) + \left( h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) \left( h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) z_i z_i' - E \left[ \left( h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) \right] z_i z_i' \right)
\]

\[
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( y_i^2 z_i z_i' - E \left[ y_i^2 z_i z_i' \right] \right) - \frac{1}{n} \sum_{i=1}^{n} \sum_{s=1}^{S} \sum_{t=1}^{S} \left( y_i h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) z_i z_i' - E \left[ y_i h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right] z_i z_i' \right)
\]

\[
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( y_i^2 z_i z_i' - E \left[ y_i^2 z_i z_i' \right] \right) - \frac{1}{n} \sum_{i=1}^{n} \sum_{s=1}^{S} \sum_{t=1}^{S} \left( y_i h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right) z_i z_i' - E \left[ y_i h \left( x_i^s \theta_1 \right) + \eta_i > 0 \right] z_i z_i' \right)
\]
\[
+ 2 \frac{\sqrt{n}}{\sqrt{S}} \sum_{s=1}^{S} E \left[ q(w_i, w_j, \eta_s) | \eta_s \right] - 2 \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E \left[ r \left( w_i, w_j, \eta_s, \eta_i \right) | w_i \right] + 2 \frac{\sqrt{n}}{\sqrt{S}} \sum_{s=1}^{S} E \left[ r \left( w_i, w_j, \eta_s, \eta_i \right) | \eta_s \right] + a_p(1)
= \sqrt{n} (P_n - P) \psi (\cdot, \theta^*_1) + a_p(1)
\]

where the second to last equality follows from the Hoeffding decompositions for two-sample and one-sample U-statistics and the last equality follows from the condition that \( n/S \to 0 \).

The bootstrapped weighting matrix is computed using the multinomial bootstrap and an initial rate-adaptive bootstrap estimator \( \hat{\theta}_1^* \) computed using a fixed weighting matrix.

\[
W_n \left( \hat{\theta}_1^* \right) = \left( \frac{1}{n} \sum_{i=1}^{n} \left( y_i - \frac{1}{S} \sum_{s=1}^{S} 1 \left( h \left( x_i^\top \hat{\theta}_1^* \right) + \eta_{is} > 0 \right) \right)^2 \right)^{-1}
\]

We can show that the bootstrapped weighting matrix has the same influence function representation using bootstrap consistency results for U-statistics (see e.g. Arcones and Gine (1992)).

### 6.3 Dynamic Censored Regression

\( \pi (\cdot, \theta) = \max \left\{ 0, y_{it} - y_{it-1} \theta \right\} - y_{it-1} \) is the moment condition in the simple censored regression model \( y_{it} = y_{it-1} \theta + \alpha_i + \epsilon_{it} \) where \( \epsilon_{it}\) is a sequence of i.i.d. random variables conditional on \((y_{i0}, \alpha_i)\). For \( \pi (\theta) = E \left[ \max \left\{ 0, y_{it} - y_{it-1} \theta \right\} - y_{it-1} \right] = E \left[ 1 \left( y_{it} > y_{it-1} \theta \right) \left( y_{it} - y_{it-1} \theta \right) - y_{it-1} \right], \)

\( G = -E \left[ y_{it-1} \left( y_{it} > y_{it-1} \theta^* \right) \right] \) and \( H = E \left[ y_{it-1} f_{y_{it-1}} \left( y_{it-1} \theta^* \right) \right] \).

Even though \( \pi (\cdot, \theta) \) is nonsmooth, the \( \sqrt{n} \) rate of convergence arises because \( \pi (\cdot, \theta) \) remains directionally differentiable. We check that the value of \( \rho \) that satisfies assumption 2(v) is \( \rho = 1 \) instead of \( \rho = 1/2 \) as in the previous two examples.

The class \( \mathcal{G}_R \equiv \{ \pi (\cdot, \theta) - \pi (\cdot, \theta^*) : \| \theta - \theta^* \| \leq R \} \) has envelope function

\[
G_R (\cdot) = \sup_{ \| \theta - \theta^* \| \leq R } \| \pi (\cdot, \theta) - \pi (\cdot, \theta^*) \|
= \sup_{ \| \theta - \theta^* \| \leq R } \| \max \left\{ 0, y_{it} - y_{it-1} \theta \right\} - \max \left\{ 0, y_{it} - y_{it-1} \theta^* \right\} \|
\leq \sup_{ \| \theta - \theta^* \| \leq R } \| \max \left\{ 0, y_{it} - y_{it-1} \theta, -y_{it} + y_{it-1} \theta^*, y_{it-1} \left( \theta^* - \theta \right) \right\} \|
\]

Therefore,

\[
P G_R^2 \leq E \max \left\{ 0, \sup_{ \| \theta - \theta^* \| \leq R } \| y_{it} - y_{it-1} \theta \|^2, \| -y_{it} + y_{it-1} \theta^* \|^2, \| y_{it-1} \|^2 R^2 \right\}
= O \left( R^2 \right)
\]

Just as in the previous example, \( W \left( \hat{\theta}_1^* \right) \) is the inverse of the variance-covariance matrix of the population moments

\[
W \left( \hat{\theta}_1^* \right) = \left( E \left[ \pi (\cdot, \theta_1^*)^2 \right] - \pi (\theta_1^*)^2 \right)^{-1}
= \left( E \left[ \left( \max \left\{ 0, y_{it} - y_{it-1} \theta_1^* \right\} - y_{it-1} \right)^2 \right] - \pi (\theta_1^*)^2 \right)^{-1}
\]
We can obtain the expression for initial rate-adaptive bootstrap estimator

\[ \sqrt{n} \left( W_n \left( \hat{\theta}_1 \right) - W (\theta^*_1) \right) \]

\[ = \sqrt{n} (W_n (\theta^*_1) - W (\theta^*_1)) + \frac{\partial W_n (\theta^*_1)}{\partial \theta^*} \sqrt{n} (\hat{\theta}_1 - \theta^*_1) + o_p(1) \]

\[ = -W (\theta^*_1)^{-1} \sqrt{n} (P_n - P) \psi (\cdot, \theta^*_1) W (\theta^*_1)^{-1} + \frac{\partial W (\theta^*_1)}{\partial \theta^*} \sqrt{n} (P_n - P) \kappa (\cdot, \theta^*_1) + o_p(1) \]

\[ = \sqrt{n} (P_n - P) \phi (\cdot, \theta^*_1) + o_p(1) \]

We can obtain the expression for \( \sqrt{n} (P_n - P) \psi (\cdot, \theta^*_1) \) using similar arguments as in the previous example:

\[ \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} \left( \pi (\cdot, \theta^*_1)^2 - E \left[ \pi (\cdot, \theta^*_1)^2 \right] \right) - \left( \hat{\pi}_n (\theta^*_1)^2 - \pi (\theta^*_1)^2 \right) \right) \]

\[ = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \max \left\{ 0, y_{it} - y_{it-1} \theta^*_1 \right\} - y_{it-1} \right)^2 - E \left[ \max \left\{ 0, y_{it} - y_{it-1} \theta^*_1 \right\} - y_{it-1} \right]^2 \]

\[ - \left( \frac{\sqrt{n}}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \max \left\{ 0, y_{it} - y_{it-1} \theta^*_1 \right\} - y_{it-1} \right) \left( \max \left\{ 0, y_{jt} - y_{jt-1} \theta^*_1 \right\} - y_{jt-1} \right) - E \left[ \max \left\{ 0, y_{it} - y_{it-1} \theta^*_1 \right\} - y_{it-1} \right]^2 \right) \]

\[ = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \max \left\{ 0, y_{it} - y_{it-1} \theta^*_1 \right\} - y_{it-1} \right)^2 - E \left[ \max \left\{ 0, y_{it} - y_{it-1} \theta^*_1 \right\} - y_{it-1} \right]^2 \]

\[ - \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E [g (y_{i}, y_{j}) | y_{i}] + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E [g (y_{i}, y_{j}) | y_{j}] \right) + o_p(1) \]

\[ = \sqrt{n} (P_n - P) \psi (\cdot, \theta^*_1) + o_p(1) \]

The bootstrapped weighting matrix is computed using the multinomial bootstrap and an initial rate-adaptive bootstrap estimator \( \hat{\theta}_1 \) computed using a fixed weighting matrix.

\[ W_n \left( \hat{\theta}_1^* \right) = \left( \frac{1}{n} \sum_{i=1}^{n} \left( \max \left\{ 0, y_{it}^* - y_{it-1}^* \theta^*_1 \right\} - y_{it-1}^* \right)^2 - \hat{\pi}_n (\theta^*_1)^2 \right)^{-1} \]

\[ = \left( \frac{1}{n} \sum_{i=1}^{n} \left( \max \left\{ 0, y_{it}^* - y_{it-1}^* \theta^*_1 \right\} - y_{it-1}^* \right)^2 - 2y_{it-1}^* \max \left\{ 0, y_{it}^* - y_{it-1}^* \theta^*_1 \right\} + (y_{it-1}^*)^2 \right) - \hat{\pi}_n (\theta^*_1)^2 \right)^{-1} \]

The bootstrapped weighting matrix has the same influence function representation using bootstrap consistency results for U-statistics (see e.g. Arcones and Gine (1992)).
7 Monte Carlo

7.1 Nonsmooth Case

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

\[ y_i = \theta_0 + \epsilon_i, i = 1, \ldots, n \]

where \( \epsilon_i \sim N(0, 1) \) and \( \theta_0 = 0 \).

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

\[ \pi(\theta) = [P(y_i \leq \theta) - \tau; Ey_i - \theta]' \]

The model can not be correctly specified as long as \( \tau \neq 0 \).

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

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

\[ Q(\theta) = \pi(\theta)' \pi(\theta) = (P(y_i \leq \theta) - \tau)^2 + (Ey_i - \theta)^2 \]

The pseudo true value \( \theta^* = \arg \min_{\theta \in \Theta} Q(\theta) \) is given by the root of the following equation:

\[ f_y(\theta^*) (F_y(\theta^*) - \tau) + \theta^* = 0. \]

We examine the empirical coverage frequencies of nominal 95\% equal-tailed rate-adaptive bootstrap confidence intervals:

\[ \left[ \hat{\theta}_n - c_{0.975}, \hat{\theta}_n - c_{0.025} \right] \]

where \( c_{0.975} \) and \( c_{0.025} \) are the 97.5th and 2.5th percentiles of \( \hat{\theta}_n^* - \hat{\theta}_n \). Recall that

\[ \hat{\theta}_n^* = \arg \min_{\theta \in \Theta} \left\{ \hat{\pi}(\hat{\theta}_n)' W \left( \left( \hat{\pi}^*(\theta) - \hat{\pi}^*(\hat{\theta}_n) \right) - \left( \hat{\pi}(\theta) - \hat{\pi}(\hat{\theta}_n) \right) \right) \right\} \\
+ \frac{1}{2} (\theta - \hat{\theta}_n)' \left( \hat{G}' W \hat{G} + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{jk} \hat{\pi}_k(\hat{\theta}_n) \hat{H}_j \right) (\theta - \hat{\theta}_n) \\
+ \left( \theta - \hat{\theta}_n \right)' \hat{G}' W \left( \hat{\pi}^* (\hat{\theta}_n) - \hat{\pi}(\hat{\theta}_n) \right) \}

where \( \hat{\pi}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \pi(y_i, \theta) \) and \( \hat{\pi}^*(\theta) = \frac{1}{n} \sum_{i=1}^{n} \pi(y_i^*, \theta) \).

\[ \hat{G} = \left[ \begin{array}{c} \frac{1}{nh} \sum_{i=1}^{n} K_h(y_i - \hat{\theta}_n) \\ -1 \end{array} \right] \quad \hat{H} = \left[ \begin{array}{c} \frac{1}{nh^2} \sum_{i=1}^{n} K_h^2(y_i - \hat{\theta}_n) \\ 0 \end{array} \right] \]
for \( K_h(x) = K(x/h), K_h'(x) = K'(x/h), K(x) = (2\pi)^{-1/2} e^{-x^2/2}, \) and \( K'(x) = -(2\pi)^{-1/2} x e^{-x^2/2}. \)

We use the Silverman’s Rule of Thumb bandwidth \( h = 1.06\text{std}(y)n^{-1/5}, \) but the results are robust to other choices of the bandwidth such as on the order of \( n^{-1/3}, n^{-1/4}, n^{-1/6}, \) or \( n^{-1/10}. \)

Table 1 shows the rate-adaptive bootstrap empirical coverage frequencies for \( \theta^* \) (along with the average widths of the confidence intervals in parentheses) for \( \tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}, \) \( n \in \{200, 800, 1600, 3200, 6400\}, B = 1000 \) bootstrap iterations, and \( R = 1000 \) Monte Carlo simulations.

We can see that the empirical coverage frequency is quite close to the nominal level of 95% even at smaller sample sizes.

Table 1: Rate-adaptive Bootstrap Empirical Coverage Frequencies

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 200 )</td>
<td>0.945</td>
<td>0.950</td>
<td>0.951</td>
<td>0.954</td>
<td>0.940</td>
</tr>
<tr>
<td>(0.331)</td>
<td>(0.299)</td>
<td>(0.279)</td>
<td>(0.298)</td>
<td>(0.330)</td>
<td></td>
</tr>
<tr>
<td>( n = 800 )</td>
<td>0.944</td>
<td>0.947</td>
<td>0.948</td>
<td>0.950</td>
<td>0.951</td>
</tr>
<tr>
<td>(0.181)</td>
<td>(0.156)</td>
<td>(0.140)</td>
<td>(0.156)</td>
<td>(0.181)</td>
<td></td>
</tr>
<tr>
<td>( n = 1600 )</td>
<td>0.954</td>
<td>0.959</td>
<td>0.950</td>
<td>0.949</td>
<td>0.958</td>
</tr>
<tr>
<td>(0.135)</td>
<td>(0.114)</td>
<td>(0.099)</td>
<td>(0.113)</td>
<td>(0.134)</td>
<td></td>
</tr>
<tr>
<td>( n = 3200 )</td>
<td>0.956</td>
<td>0.944</td>
<td>0.955</td>
<td>0.951</td>
<td>0.959</td>
</tr>
<tr>
<td>(0.101)</td>
<td>(0.083)</td>
<td>(0.070)</td>
<td>(0.083)</td>
<td>(0.101)</td>
<td></td>
</tr>
<tr>
<td>( n = 6400 )</td>
<td>0.949</td>
<td>0.947</td>
<td>0.944</td>
<td>0.945</td>
<td>0.954</td>
</tr>
<tr>
<td>(0.076)</td>
<td>(0.061)</td>
<td>(0.049)</td>
<td>(0.061)</td>
<td>(0.076)</td>
<td></td>
</tr>
</tbody>
</table>

We also examine the empirical coverage frequencies of nominal 95% equal-tailed standard bootstrap confidence intervals \([\hat{\theta}_n - d_{0.975}, \hat{\theta}_n - d_{0.025}].\) where \( d_{0.975} \) and \( d_{0.025} \) are the 97.5th and 2.5th percentiles of \( \tilde{\theta}_1 - \hat{\theta}_n \) for \( \tilde{\theta}_1 = \arg\min_{\theta \in \Theta} \left( \hat{\pi}^*(\theta) - \hat{\pi} \left( \hat{\theta}_n \right) \right)' \left( \hat{\pi}^*(\theta) - \hat{\pi} \left( \hat{\theta}_n \right) \right) \).

From table 2, we can see that the standard bootstrap performs fine under correct specification, but the performance deteriorates as \( \tau \) moves further away from 0.5.

Table 2: Centered Standard Bootstrap Empirical Coverage Frequencies

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 200 )</td>
<td>0.913</td>
<td>0.922</td>
<td>0.942</td>
<td>0.913</td>
<td>0.901</td>
</tr>
<tr>
<td>(0.277)</td>
<td>(0.277)</td>
<td>(0.277)</td>
<td>(0.277)</td>
<td>(0.277)</td>
<td></td>
</tr>
<tr>
<td>( n = 800 )</td>
<td>0.881</td>
<td>0.917</td>
<td>0.953</td>
<td>0.929</td>
<td>0.883</td>
</tr>
<tr>
<td>(0.139)</td>
<td>(0.139)</td>
<td>(0.139)</td>
<td>(0.139)</td>
<td>(0.139)</td>
<td></td>
</tr>
<tr>
<td>( n = 1600 )</td>
<td>0.864</td>
<td>0.892</td>
<td>0.939</td>
<td>0.913</td>
<td>0.843</td>
</tr>
<tr>
<td>(0.098)</td>
<td>(0.098)</td>
<td>(0.098)</td>
<td>(0.098)</td>
<td>(0.098)</td>
<td></td>
</tr>
<tr>
<td>( n = 3200 )</td>
<td>0.832</td>
<td>0.894</td>
<td>0.940</td>
<td>0.884</td>
<td>0.812</td>
</tr>
<tr>
<td>(0.070)</td>
<td>(0.070)</td>
<td>(0.070)</td>
<td>(0.070)</td>
<td>(0.070)</td>
<td></td>
</tr>
<tr>
<td>( n = 6400 )</td>
<td>0.823</td>
<td>0.894</td>
<td>0.954</td>
<td>0.894</td>
<td>0.785</td>
</tr>
<tr>
<td>(0.049)</td>
<td>(0.049)</td>
<td>(0.049)</td>
<td>(0.049)</td>
<td>(0.049)</td>
<td></td>
</tr>
</tbody>
</table>
Now we consider the case of an estimated weighting matrix. The variance-covariance matrix of the moments is

\[ E(\pi(\cdot, \theta) - \pi(\theta)) (\pi(\cdot, \theta) - \pi(\theta))' = E[\pi(\cdot, \theta) \pi(\cdot, \theta)'] - \pi(\theta) \pi(\theta)'. \]

\[ = E\left[ \begin{array}{c}
1 (y_i \leq \theta) - 2\tau 1 (y_i \leq \theta) + \tau^2 \\
1 (y_i \leq \theta) y_i - \theta (y_i \leq \theta) - \tau y_i + \tau \theta
\end{array} \right] - \pi(\theta) \pi(\theta)'
\]

\[ = \left[ \begin{array}{cc}
F_y(\theta) - 2\tau F_y(\theta) + \tau^2 \\
- f_y(\theta) - \theta F_y(\theta) + \tau \theta
\end{array} \right] - \pi(\theta) \pi(\theta)'
\]

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

\[ W_n(\hat{\theta}_1) = \left[ \begin{array}{ccc}
\hat{F}_y(\hat{\theta}_1) - \hat{F}_y(\hat{\theta}_1)^2 & - \hat{f}_y(\hat{\theta}_1) & 1 \\
- \hat{f}_y(\hat{\theta}_1) & 1
\end{array} \right]^{-1}
\]

where \( \hat{\theta}_1 = \arg \min_{\theta \in \Theta} \hat{\pi}(\theta)' \hat{\pi}(\theta), \hat{f}_y(\hat{\theta}_1) = \frac{1}{nh} \sum_{i=1}^n K_h(y_i - \hat{\theta}_1), \hat{F}_y(\hat{\theta}_1) = \frac{1}{n} \sum_{i=1}^n 1 \left( y_i \leq \hat{\theta}_1 \right). \)

For \( \hat{\theta}_n = \arg \min_{\theta \in \Theta} \hat{\pi}(\theta)' W_n(\hat{\theta}_1) \hat{\pi}(\theta), \) the rate-adaptive bootstrap estimate is

\[ \hat{\theta}_n^* = \arg \min_{\theta \in \Theta} \left\{ \hat{\pi}(\hat{\theta}_n)' W_n(\hat{\theta}_1) \left( \left( \hat{\pi}^*(\theta) - \hat{\pi}^*(\hat{\theta}_n) \right) - \left( \hat{\pi}(\theta) - \hat{\pi}(\hat{\theta}_n) \right) \right) \right\}
\]

\[ + \frac{1}{2} (\theta - \hat{\theta}_n)' \left( \hat{G}^* W_n(\hat{\theta}_1) \hat{G} + \sum_{j=1}^m \sum_{k=1}^m W_{n,jk} (\hat{\theta}_1) \hat{\pi}_k (\hat{\theta}_n) \hat{H}_j \right) (\theta - \hat{\theta}_n)
\]

\[ + (\theta - \hat{\theta}_n)' \hat{G}^* (W_n^*(\hat{\theta}_1^*) - W_n(\hat{\theta}_1)) \hat{\pi}(\hat{\theta}_n) \}
\]

The bootstrapped weighting matrix is

\[ W_n^*(\hat{\theta}_1^*) = \left[ \begin{array}{ccc}
\hat{F}_y^*(\hat{\theta}_1^*) - \hat{F}_y^*(\hat{\theta}_1^*)^2 & - \hat{f}_y^*(\hat{\theta}_1^*) & 1 \\
- \hat{f}_y^*(\hat{\theta}_1^*) & 1
\end{array} \right]^{-1}
\]

where \( \hat{\theta}_1^* \) is the rate-adaptive bootstrap estimate using \( W = I, \hat{f}_y^*(\hat{\theta}_1^*) = \frac{1}{nh} \sum_{i=1}^n K_h^* (y_i - \hat{\theta}_1^*), \) and \( \hat{F}_y^*(\hat{\theta}_1^*) = \frac{1}{n} \sum_{i=1}^n 1 \left( y_i^* \leq \hat{\theta}_1^* \right). \)

We are interested in the rate-adaptive bootstrap empirical coverage frequencies for \( \theta^* = \arg \min_{\theta \in \Theta} \theta^* W(\theta^*) \pi(\theta) \) where \( W(\theta^*_1) = \left[ \begin{array}{ccc}
F_y(\theta^*_1) - F_y(\theta^*_1)^2 & - f_y(\theta^*_1) & 1 \\
- f_y(\theta^*_1) & 1
\end{array} \right]^{-1} \) and \( \theta^*_1 = \arg \min_{\theta \in \Theta} \theta^* \pi(\theta). \)

20
Table 3 shows the empirical coverage frequencies of nominal 95% equal-tailed rate-adaptive bootstrap confidence intervals:

$$\left[ \hat{\theta}_n - c_{0.975}, \hat{\theta}_n - c_{0.025} \right]$$

where $c_{0.975}$ and $c_{0.025}$ are the 97.5th and 2.5th percentiles of $\hat{\theta}_n^* - \hat{\theta}_n$. We used $B = 1000$ bootstrap iterations and $R = 1000$ Monte Carlo simulations. There is some slight under-coverage for the case of $\tau = 0.5$ and over-coverage for the other values of $\tau$, but the performance is much better than the standard bootstrap intervals shown in Table 4.

Table 3: Rate-adaptive Bootstrap Empirical Coverage Frequencies

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 200$</td>
<td>0.970</td>
<td>0.963</td>
<td>0.941</td>
<td>0.965</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td>(1.224)</td>
<td>(0.682)</td>
<td>(0.286)</td>
<td>(0.693)</td>
<td>(1.228)</td>
</tr>
<tr>
<td>$n = 800$</td>
<td>0.987</td>
<td>0.957</td>
<td>0.948</td>
<td>0.967</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>(0.688)</td>
<td>(0.370)</td>
<td>(0.143)</td>
<td>(0.370)</td>
<td>(0.702)</td>
</tr>
<tr>
<td>$n = 1600$</td>
<td>0.984</td>
<td>0.966</td>
<td>0.954</td>
<td>0.967</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>(0.520)</td>
<td>(0.293)</td>
<td>(0.101)</td>
<td>(0.293)</td>
<td>(0.523)</td>
</tr>
<tr>
<td>$n = 3200$</td>
<td>0.990</td>
<td>0.960</td>
<td>0.934</td>
<td>0.952</td>
<td>0.990</td>
</tr>
<tr>
<td></td>
<td>(0.394)</td>
<td>(0.235)</td>
<td>(0.071)</td>
<td>(0.234)</td>
<td>(0.394)</td>
</tr>
<tr>
<td>$n = 6400$</td>
<td>0.978</td>
<td>0.947</td>
<td>0.944</td>
<td>0.957</td>
<td>0.982</td>
</tr>
<tr>
<td></td>
<td>(0.311)</td>
<td>(0.189)</td>
<td>(0.050)</td>
<td>(0.189)</td>
<td>(0.313)</td>
</tr>
</tbody>
</table>

Table 4 shows the empirical coverage frequencies of nominal 95% equal-tailed standard bootstrap confidence intervals

$$\left[ \hat{\theta}_n - d_{0.975}, \hat{\theta}_n - d_{0.025} \right]$$

where $d_{0.975}$ and $d_{0.025}$ are the 97.5th and 2.5th percentiles of $\tilde{\theta}_2 - \hat{\theta}_n$, for $\tilde{\theta}_2 = \arg \min_{\theta \in \Theta} \left( \hat{\pi}^*(\theta) - \hat{\pi} \left( \hat{\theta}_n \right) \right)' W_n^* \left( \tilde{\theta}_1 \right) \left( \hat{\pi}^*(\theta) - \hat{\pi} \left( \hat{\theta}_n \right) \right)$ and $\tilde{\theta}_1 = \arg \min_{\theta \in \Theta} \left( \hat{\pi}^*(\theta) - \hat{\pi} \left( \hat{\theta}_n \right) \right)'. \left( \hat{\pi}^*(\theta) - \hat{\pi} \left( \hat{\theta}_n \right) \right)$. There is under-coverage across all values of $\tau$, even for the correctly specified case of $\tau = 0.5$.

Table 4: Standard Bootstrap Equal-tailed Coverage Frequencies and Interval Lengths

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 200$</td>
<td>0.697</td>
<td>0.842</td>
<td>0.951</td>
<td>0.835</td>
<td>0.688</td>
</tr>
<tr>
<td></td>
<td>(0.344)</td>
<td>(0.329)</td>
<td>(0.310)</td>
<td>(0.331)</td>
<td>(0.338)</td>
</tr>
<tr>
<td>$n = 800$</td>
<td>0.647</td>
<td>0.766</td>
<td>0.957</td>
<td>0.791</td>
<td>0.648</td>
</tr>
<tr>
<td></td>
<td>(0.174)</td>
<td>(0.172)</td>
<td>(0.160)</td>
<td>(0.171)</td>
<td>(0.174)</td>
</tr>
<tr>
<td>$n = 1600$</td>
<td>0.606</td>
<td>0.757</td>
<td>0.966</td>
<td>0.752</td>
<td>0.613</td>
</tr>
<tr>
<td></td>
<td>(0.125)</td>
<td>(0.122)</td>
<td>(0.114)</td>
<td>(0.123)</td>
<td>(0.125)</td>
</tr>
<tr>
<td>$n = 3200$</td>
<td>0.555</td>
<td>0.711</td>
<td>0.976</td>
<td>0.685</td>
<td>0.533</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.088)</td>
<td>(0.081)</td>
<td>(0.088)</td>
<td>(0.090)</td>
</tr>
<tr>
<td>$n = 6400$</td>
<td>0.533</td>
<td>0.682</td>
<td>0.963</td>
<td>0.666</td>
<td>0.516</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.063)</td>
<td>(0.057)</td>
<td>(0.063)</td>
<td>(0.064)</td>
</tr>
</tbody>
</table>
7.2 IV Quantile Regression

The data generating process is

\[ y_i = \alpha_0 + \beta_0 d_i + u_i, \quad \left( \begin{array}{c} u_i \\ d_i \end{array} \right) \overset{i.i.d.}{\sim} N(\mu, \Omega), \quad \mu = \left( \begin{array}{c} \delta \\ 1 \end{array} \right), \quad \Omega = \left( \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \]

It follows then that

\[ u_i | d_i \sim N(\delta, 1) \]

\[ P(u_i \leq 0|d_i) = \Phi(-\delta) \]

\[ y_i | d_i \sim N(\alpha_0 + \beta_0 d_i + \delta, 1) \]

The population moments are for \( z_i = (1 \quad d_i)' \),

\[ \pi(\theta) = E \left[ \left( \frac{1}{2} - 1 \left( y_i \leq \alpha + \beta d_i \right) \right) z_i \right] \]

\[ = E \left[ \left( \frac{1}{2} - F_{y|d}(\alpha + \beta d_i) \right) z_i \right] \]

\[ = E \left[ \left( \frac{1}{2} - \Phi(\alpha - \alpha_0 + (\beta - \beta_0) d_i - \delta) \right) z_i \right] \]

The sample moments are

\[ \hat{\pi}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{2} - 1 \left( y_i \leq \alpha + \beta d_i \right) \right) z_i \]

Note that if \( \delta = 0 \), then we have a correctly specified model for median regression. For values of \( \delta \neq 0 \), the model is misspecified. Because the researcher is not able to observe \( \delta \), it is desirable to have a procedure that will perform valid inference for the true parameters \( \theta_0 = (\alpha_0, \beta_0)' \) when \( \delta = 0 \), and also will perform valid inference for the pseudo-true parameters \( \theta^* = (\alpha^*, \beta^*)' = \arg \min_{\theta} \pi(\theta)' W \pi(\theta) \) when \( \delta \neq 0 \). We first consider the case of a fixed weighting matrix \( W = I \).

The bootstrapped sample moments using the multinomial bootstrap are

\[ \hat{\pi}^*_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{2} - 1 \left( y_i^* \leq \alpha + \beta d_i^* \right) \right) z_i^* \]

The population Jacobian and Hessians are

\[ G(\theta) = -E \left[ f_{y|d}(\alpha + \beta d_i) z_i z_i' \right] \]

\[ H_1(\theta) = -E \left[ f_{y|d}(\alpha + \beta d_i) z_i z_i' \right] \]

\[ H_2(\theta) = -E \left[ f_{y|d}(\alpha + \beta d_i) d_i z_i z_i' \right] \]

22
Their estimates are
\[ \hat{G} = -\frac{1}{nh} \sum_{i=1}^{n} K_h \left( y_i - \hat{\alpha}_n - \hat{\beta}_n d_i \right) z_i z'_i \]
\[ \hat{H}_1 = -\frac{1}{nh^2} \sum_{i=1}^{n} K'_h \left( y_i - \hat{\alpha}_n - \hat{\beta}_n d_i \right) z_i z'_i \]
\[ \hat{H}_2 = -\frac{1}{nh^2} \sum_{i=1}^{n} K'_h \left( y_i - \hat{\alpha}_n - \hat{\beta}_n d_i \right) d_i z_i z'_i \]

The rate-adaptive bootstrap estimator in the case of a fixed weighting matrix \( W \) is
\[ \hat{\theta}^*_n = \arg \min_{\theta \in \Theta} \left\{ \hat{\pi}_n \left( \hat{\theta}_n \right)' W (P_n^* - P_n) \left( \pi (\cdot, \theta) - \pi (\cdot, \hat{\theta}_n) \right) \right\} \]
\[ + \frac{1}{2} \left( \theta - \hat{\theta}_n \right)' \left( \hat{G}' W \hat{G} + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{j,k} \hat{\pi}_n(k, \hat{\theta}_n) \hat{H}_j \right) \left( \theta - \hat{\theta}_n \right) \]
\[ + \left( \theta - \hat{\theta}_n \right)' \hat{G}' W (P_n^* - P_n) \pi (\cdot, \hat{\theta}_n) \]

Tables 5 and 6 compare the empirical coverage frequencies and average interval lengths of nominal 95% equal-tailed confidence intervals \( [\hat{\theta}_n - c_{0.975}, \hat{\theta}_n - c_{0.025}] \) constructed using the rate-adaptive bootstrap estimator and the centered standard bootstrap estimator \( \hat{\theta}^*_n = \arg \min_{\theta} \left( \hat{\pi}_n (\theta) - \hat{\pi}_n (\hat{\theta}_n) \right)' \hat{\pi}_n (\hat{\theta}_n) \). We use \( B = 2000 \) bootstrap iterations, \( R = 5000 \) Monte Carlo simulations, and two values of \( \delta \in \{0.1, 0.9\} \). The standard bootstrap undercovers for all values of \( n \) while the rate-adaptive bootstrap achieves coverage close to the nominal level.

| Table 5: Rate-Adaptive vs. Standard Bootstrap, \( \delta = 0.1 \), fixed weighting matrix \( W = I \) |
|-----------------|----|----|----|----|----|
| \( n \)        | 200 | 800 | 1600 | 3200 | 6400 |
| Rate-adaptive   | 0.979 | 0.958 | 0.953 | 0.949 | 0.955 |
| (0.385)         | (0.240) | (0.173) | (0.124) | (0.088) |
|                 | 0.974 | 0.956 | 0.951 | 0.958 | 0.955 |
| (0.306)         | (0.176) | (0.125) | (0.088) | (0.062) |
| Standard        | 0.835 | 0.915 | 0.923 | 0.929 | 0.938 |
| (0.271)         | (0.231) | (0.171) | (0.123) | (0.087) |
|                 | 0.820 | 0.912 | 0.925 | 0.938 | 0.938 |
| (0.243)         | (0.170) | (0.123) | (0.087) | (0.062) |

Now consider the case of an estimated weighting matrix. Let \( W (\theta^*_1) = \text{plim} \ W_n (\hat{\theta}_1) \) be the probability limit of an estimated weighting matrix computed using an initial GMM estimator \( \hat{\theta}_1 = \arg \min_{\theta} \hat{\pi}_n (\theta)' \hat{\pi}_n (\theta) \) whose probability limit is \( \theta^*_1 = \arg \min_{\theta} \pi (\theta)' \pi (\theta) \). The pseudo-true parameters are given by \( \theta^* = \arg \min_{\theta} \pi (\theta)' W (\theta^*_1) \pi (\theta) \), which can be computed
by solving

\[
0 = \frac{\partial \pi (\theta^*)'}{\partial \theta} W (\theta_1^*) \pi (\theta^*)
\]

\[
= - E \left[ f_{y | d} (\alpha^* + \beta^* d_i) z_i z_i' \right] W (\theta_1^*) E \left[ \left( \frac{1}{2} - F_{y | d} (\alpha^* + \beta^* d_i) \right) z_i \right]
\]

\[
= - E \left[ \left( \frac{1}{2} - \phi (\alpha^* - \alpha_0 + (\beta^* - \beta_0) d_i - \delta) \right) z_i \right] W (\theta_1^*) E \left[ \left( \frac{1}{2} - \Phi (\alpha^* - \alpha_0 + (\beta^* - \beta_0) d_i - \delta) \right) z_i \right]
\]

\[
W (\theta_1^*) = \left( E \left[ \pi (\cdot, \theta_1^*) \pi (\cdot, \theta_1^*)' \right] - \pi (\theta_1^*) \pi (\theta_1^*)' \right)^{-1}
\]

\[
= \left( E \left[ \left( \frac{1}{2} - 1 (y_i \leq \alpha^* + \beta^* d_i) \right)^2 z_i z_i' \right] - \pi (\theta_1^*) \pi (\theta_1^*)' \right)^{-1}
\]

\[
= \left( E \left[ \left( \frac{1}{2} - 1 (y_i \leq \alpha^* + \beta^* d_i) \right)^2 \mid z_i \right] z_i z_i' \right) - \pi (\theta_1^*) \pi (\theta_1^*)' \right)^{-1}
\]

\[
= \left( \frac{1}{4} E [z_i z_i'] - \pi (\theta_1^*) \pi (\theta_1^*)' \right)^{-1}
\]

The last line follows from the fact that conditional on \( z_i \), \( \frac{1}{2} - 1 (y_i \leq \alpha^* + \beta^* d_i) \) is a Bernoulli random variable that equals \( \frac{1}{2} \) with probability \( F_{y | d} (\alpha^* + \beta^* d_i) \) and equals \( \frac{1}{2} \) with probability \( 1 - F_{y | d} (\alpha^* + \beta^* d_i) \). Therefore, \( E \left[ \left( \frac{1}{2} - 1 (y_i \leq \alpha^* + \beta^* d_i) \right)^2 \mid z_i \right] = \frac{1}{4} \).

The estimated weighting matrix is

\[
W_n \equiv W_n (\hat{\theta}_1) = \left( \frac{1}{4} n \sum_{i=1}^{n} z_i z_i' - \hat{\pi}_n (\hat{\theta}_1) \hat{\pi}_n (\hat{\theta}_1)' \right)^{-1}
\]

The bootstrapped weighting matrix is computed using the multinomial bootstrap and an initial rate-adaptive bootstrap estimator \( \hat{\theta}_1^* \) computed using a fixed weighting matrix \( W = I \).

\[
W_n^* \equiv W_n^* (\hat{\theta}_1^*) = \left( \frac{1}{4} n \sum_{i=1}^{n} z_i^* z_i'^* - \hat{\pi}_n (\hat{\theta}_1^*) \hat{\pi}_n (\hat{\theta}_1^*)' \right)^{-1}
\]

Table 6: Rate-Adaptive vs. Standard Bootstrap, \( \delta = 0.9 \), fixed weighting matrix \( W = I \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>200</th>
<th>800</th>
<th>1600</th>
<th>3200</th>
<th>6400</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate-adaptive</td>
<td>0.959</td>
<td>0.954</td>
<td>0.951</td>
<td>0.951</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td>(0.477)</td>
<td>(0.251)</td>
<td>(0.176)</td>
<td>(0.125)</td>
<td>(0.088)</td>
</tr>
<tr>
<td></td>
<td>0.971</td>
<td>0.959</td>
<td>0.951</td>
<td>0.956</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td>(0.294)</td>
<td>(0.174)</td>
<td>(0.124)</td>
<td>(0.088)</td>
<td>(0.062)</td>
</tr>
<tr>
<td>Standard</td>
<td>0.864</td>
<td>0.915</td>
<td>0.926</td>
<td>0.932</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>(0.421)</td>
<td>(0.247)</td>
<td>(0.175)</td>
<td>(0.123)</td>
<td>(0.087)</td>
</tr>
<tr>
<td></td>
<td>0.844</td>
<td>0.921</td>
<td>0.924</td>
<td>0.935</td>
<td>0.939</td>
</tr>
<tr>
<td></td>
<td>(0.230)</td>
<td>(0.171)</td>
<td>(0.123)</td>
<td>(0.087)</td>
<td>(0.062)</td>
</tr>
</tbody>
</table>
The rate-adaptive bootstrap estimator in the case of an estimated weighting matrix is

\[
\hat{\theta}^*_n = \arg \min_{\theta \in \Theta} \left\{ \pi_n(\hat{\theta}_n) W_n(P^*_n - P_n) \left( \pi(\cdot, \theta) - \pi(\cdot, \hat{\theta}_n) \right) \right.
\]

\[
+ \frac{1}{2} (\theta - \hat{\theta}_n)' \left( \hat{G}' W_n \hat{G} + \sum_{j=1}^m \sum_{k=1}^m W_{n,jk} \pi_n,k(\hat{\theta}) \hat{H}_j \right) (\theta - \hat{\theta}_n)
\]

\[
+ (\theta - \hat{\theta}_n)' \hat{G}' W_n (P^*_n - P_n) \pi(\cdot, \hat{\theta}_n)
\]

\[
+ (\theta - \hat{\theta}_n)' \hat{G}' W_n^* W_n (\hat{\theta}) - \hat{\theta}_n \pi(\cdot, \hat{\theta}_n)
\}\}

Tables 7 and 8 compare the empirical coverage frequencies and average interval lengths of nominal 95% equal-tailed confidence intervals \([\hat{\theta}_n - c_{0.975}, \hat{\theta}_n - c_{0.025}]\) constructed using the rate-adaptive bootstrap estimator and the centered standard bootstrap estimator \(\tilde{\theta}^*_n = \arg \min_{\theta} \left\{ \pi^*_n(\theta) - \pi_n(\hat{\theta}_n) \right\} W_n^* W_n (\hat{\theta}) - \hat{\theta}_n \pi(\cdot, \hat{\theta}_n) \). We use \(B = 2000\) bootstrap iterations, \(R = 5000\) Monte Carlo simulations, and two values of \(\delta \in \{0.1, 0.9\}\). The standard bootstrap undercovers for all values of \(n\) while the rate-adaptive bootstrap achieves coverage close to the nominal level.

Table 7: Rate-Adaptive vs. Standard Bootstrap, \(\delta = 0.1\), estimated weighting matrix

<table>
<thead>
<tr>
<th>(n)</th>
<th>200</th>
<th>800</th>
<th>1600</th>
<th>3200</th>
<th>6400</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate-adaptive</td>
<td>0.958</td>
<td>0.952</td>
<td>0.951</td>
<td>0.950</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td>(0.503)</td>
<td>(0.250)</td>
<td>(0.177)</td>
<td>(0.125)</td>
<td>(0.089)</td>
</tr>
<tr>
<td></td>
<td>0.957</td>
<td>0.953</td>
<td>0.950</td>
<td>0.956</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td>(0.357)</td>
<td>(0.177)</td>
<td>(0.125)</td>
<td>(0.088)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>Standard</td>
<td>0.855</td>
<td>0.918</td>
<td>0.921</td>
<td>0.932</td>
<td>0.939</td>
</tr>
<tr>
<td></td>
<td>(0.357)</td>
<td>(0.245)</td>
<td>(0.174)</td>
<td>(0.123)</td>
<td>(0.087)</td>
</tr>
<tr>
<td></td>
<td>0.864</td>
<td>0.921</td>
<td>0.921</td>
<td>0.937</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>(0.281)</td>
<td>(0.175)</td>
<td>(0.123)</td>
<td>(0.087)</td>
<td>(0.062)</td>
</tr>
</tbody>
</table>

Table 8: Rate-Adaptive vs. Standard Bootstrap, \(\delta = 0.9\), estimated weighting matrix

<table>
<thead>
<tr>
<th>(n)</th>
<th>200</th>
<th>800</th>
<th>1600</th>
<th>3200</th>
<th>6400</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate-adaptive</td>
<td>0.949</td>
<td>0.951</td>
<td>0.951</td>
<td>0.949</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td>(0.501)</td>
<td>(0.250)</td>
<td>(0.177)</td>
<td>(0.125)</td>
<td>(0.088)</td>
</tr>
<tr>
<td></td>
<td>0.957</td>
<td>0.953</td>
<td>0.951</td>
<td>0.957</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td>(0.356)</td>
<td>(0.177)</td>
<td>(0.125)</td>
<td>(0.088)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>Standard</td>
<td>0.880</td>
<td>0.916</td>
<td>0.921</td>
<td>0.932</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>(0.452)</td>
<td>(0.247)</td>
<td>(0.174)</td>
<td>(0.123)</td>
<td>(0.087)</td>
</tr>
<tr>
<td></td>
<td>0.878</td>
<td>0.921</td>
<td>0.919</td>
<td>0.939</td>
<td>0.939</td>
</tr>
<tr>
<td></td>
<td>(0.287)</td>
<td>(0.175)</td>
<td>(0.123)</td>
<td>(0.087)</td>
<td>(0.062)</td>
</tr>
</tbody>
</table>
7.3 Smooth Case

Now suppose we consider the data combination example in section 7.1 of Lee (2014). Suppose we observe \((y_i, z_i) \in \mathbb{R}^2\), and our goal is to estimate \(\theta = Ez_i\). Suppose we think that the mean of \(y_i\) is 0, and we would like to exploit this information to get more accurate estimates of \(\theta\). Our moments are

\[
\pi_1 (\cdot, \theta) = y_i, \quad \pi_2 (\cdot, \theta) = z_i - \theta,
\]

However, suppose the actual mean of \(y_i\) is \(\delta \neq 0\), so the model is misspecified. We generate data as

\[
\begin{pmatrix} y_i \\ z_i \end{pmatrix} \overset{i.i.d}{\sim} N \left( \begin{pmatrix} \delta \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right)
\]

As shown in the supplemental appendix of Lee (2014), the 1-step GMM estimator (using the identity weighting matrix) is \(\hat{\theta}_1 = \bar{z}\) and the 2-step GMM estimator using the optimal weighting matrix \(W_n = \begin{pmatrix} S^2_y & S_{yz} \\ S_{yz} & S^2_z \end{pmatrix}^{-1} = \frac{S^2_y}{S^2_y - S^2_{yz}} \begin{pmatrix} S^2_z & -S_{yz} \\ -S_{yz} & S^2_y \end{pmatrix}\) is \(\hat{\theta}_2 = \bar{z} - \frac{S_{yz}}{S^2_y} \bar{y}\).

We would like to compare the performance of our rate-adaptive bootstrap to the standard bootstrap estimators \(\hat{\theta}_1^* = \bar{z}^*\) and \(\hat{\theta}_2^* = \bar{z}^* - \frac{S_{yz}^*}{S^2_y} \bar{y}^*\).

It turns out that the rate-adaptive bootstrap 1-step GMM estimator is numerically identical to the standard bootstrap 1-step GMM estimator. We can see this by noting that

\[
\hat{\theta}_1^* = \arg\min_{\theta \in \Theta} \left\{ \frac{1}{2} (\theta - \hat{\theta}_1)^2 + (\theta - \hat{\theta}_1)^2 G^*(P^*_n - P_n) \pi (\cdot, \hat{\theta}_1) \right\} = \arg\min_{\theta \in \Theta} \left\{ \frac{1}{2} (\theta - \bar{z})^2 - (\theta - \bar{z}) (\bar{z}^* - \bar{z}) \right\} = \bar{z}^*
\]

The rate-adaptive 2-step GMM estimator differs from the standard bootstrap 2-step GMM estimator:

\[
\hat{\theta}_2^* = \arg\min_{\theta \in \Theta} \left\{ \frac{1}{2} (\theta - \hat{\theta}_2)^2 \hat{G} W_n \hat{G} + (\theta - \hat{\theta}_2)^2 \hat{G} W_n (P^*_n - P_n) \pi (\cdot, \hat{\theta}_2) \right\} = \arg\min_{\theta \in \Theta} \left\{ \frac{1}{2} (\theta - \bar{z})^2 \frac{S^2_y}{S^2_y S^2_z - S^2_{yz}} + (\theta - \bar{z}) \frac{(\bar{y}^* - \bar{y}) - S^2_y (\bar{z}^* - \bar{z})}{S^2_y S^2_z - S^2_{yz}} \right\}
\]

\[
\implies \hat{\theta}_2^* = \bar{z}^* - \frac{S_{yz}^* \bar{y}^* - S^2_y S^2_z - S^2_{yz}}{S^2_y S^2_z - S^2_{yz}} \frac{(S^2_y S^2_z - S^2_{yz})}{S^2_y S^2_z - S^2_{yz}} \bar{y}
\]

We examine the empirical coverage frequencies of nominal 95% equal-tailed rate-adaptive bootstrap confidence intervals \([\hat{\theta}_2 - c_{0.975}, \hat{\theta}_2 - c_{0.025}]\), where \(c_{0.975}\) and \(c_{0.025}\) are the 97.5th and 2.5th quantiles, respectively.
and 2.5th percentiles of $\hat{\theta}_2^* - \hat{\theta}_2$. We also examine the empirical coverage frequencies of nominal 95\% equal-tailed standard bootstrap confidence intervals: $[\hat{\theta}_2 - d_{0.975}, \hat{\theta}_2 - d_{0.025}]$, where $d_{0.975}$ and $d_{0.025}$ are the 97.5th and 2.5th percentiles of $\hat{\theta}_2^* - \hat{\theta}_2$. We also examine the empirical coverage frequencies of Lee (2014)'s nominal 95\% MR bootstrap confidence intervals. We use $B = 5000$ bootstrap iterations and $R = 5000$ Monte Carlo simulations.

From tables 9 and 10 which correspond to $\delta = 1$ and $\delta = 0.1$ respectively, we can see that the rate-adaptive bootstrap performs similarly to the standard and MR bootstraps in terms of both coverage and confidence interval width. Results for symmetric confidence intervals are very similar and available upon request.

| Table 9: Empirical Coverage Frequencies for $\delta = 1$ |
|---|---|---|---|---|---|---|
| $n$ | 200 | 800 | 1600 | 3200 | 6400 | 9600 |
| Rate-adaptive | 0.947 | 0.952 | 0.946 | 0.955 | 0.951 | 0.951 |
| (0.343) | (0.170) | (0.120) | (0.085) | (0.060) | (0.049) |
| Standard | 0.944 | 0.951 | 0.946 | 0.955 | 0.950 | 0.951 |
| (0.339) | (0.170) | (0.120) | (0.085) | (0.060) | (0.049) |
| MR | 0.944 | 0.950 | 0.946 | 0.955 | 0.949 | 0.951 |
| (0.339) | (0.170) | (0.120) | (0.085) | (0.060) | (0.049) |

| Table 10: Empirical Coverage Frequencies for $\delta = 0.1$ |
|---|---|---|---|---|---|---|
| $n$ | 200 | 800 | 1600 | 3200 | 6400 | 9600 |
| Rate-adaptive | 0.947 | 0.948 | 0.948 | 0.950 | 0.949 | 0.952 |
| (0.240) | (0.121) | (0.085) | (0.060) | (0.043) | (0.035) |
| Standard | 0.947 | 0.948 | 0.948 | 0.949 | 0.949 | 0.952 |
| (0.241) | (0.121) | (0.085) | (0.060) | (0.043) | (0.035) |
| MR | 0.948 | 0.949 | 0.949 | 0.950 | 0.949 | 0.953 |
| (0.241) | (0.121) | (0.085) | (0.060) | (0.043) | (0.035) |

8 Conclusion

We have demonstrated that globally misspecified GMM estimators with nonsmooth (non-directionally differentiable) moments have a cubic-root rate or slower rate of convergence to a nonstandard asymptotic distribution, hence invalidating the standard bootstrap for inference. We have proposed an alternative inference procedure that does not require knowing the rate of convergence to consistently estimate the limiting distribution and is thus robust to global misspecification and nonsmoothness. Our rate-adaptive bootstrap provides asymptotically valid inference for the true parameter when the model is correctly specified and for the pseudo-true parameter when the model is globally misspecified.
9 Appendix

Proof for Theorem 1  The consistency proof is a direct application of Corollary 3.2.3 in van der Vaart and Wellner (1996) after we show that $\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q(\theta)| = o_p(1)$. For a fixed weighting matrix $W$,

$$
sup_{\theta \in \Theta} \left| P_n \pi \left( \cdot, \theta \right)' WP_n \pi \left( \cdot, \theta \right) - P \pi \left( \cdot, \theta \right)' WP \pi \left( \cdot, \theta \right) \right|
\leq sup_{\theta \in \Theta} \left( P_n \pi \left( \cdot, \theta \right) - P \pi \left( \cdot, \theta \right) \right)' W \left( P_n \pi \left( \cdot, \theta \right) - P \pi \left( \cdot, \theta \right) \right)
\leq sup_{\theta \in \Theta} \|P_n \pi \left( \cdot, \theta \right) - P \pi \left( \cdot, \theta \right)\| W sup_{\theta \in \Theta} \|P_n \pi \left( \cdot, \theta \right) - P \pi \left( \cdot, \theta \right)\|
= o_p(1)
$$

Next we show that $n^{1/3} \left( \hat{\theta}_n - \theta^* \right) = O_p(1)$. Define $\hat{Q}_n(\theta) = \sqrt{n} \left( P_n - P \right) g \left( \cdot, \theta \right)$, $\hat{g} \left( \theta \right) = P_n g \left( \cdot, \theta \right)$, and $g(\theta) = P g \left( \cdot, \theta \right)$. Then $\hat{\pi}(\theta) = g(\theta) + \hat{\pi}(\theta^*) + \hat{\eta}(\theta)$, where $\hat{\eta}(\theta) = \frac{1}{\sqrt{n}} \hat{Q}_n(\theta)$.

Recall that $\hat{Q}_n(\theta) = \frac{1}{2} \hat{\pi}(\theta)' W \hat{\pi}(\theta)$. Write $\hat{Q}_n(\theta) - \hat{Q}_n(\theta^*) = Q_1(\theta) + Q_2(\theta) + \hat{Q}_3(\theta)$, where

$$
Q_1(\theta) = \frac{1}{2} g(\theta)' W g(\theta) + g(\theta)' W \pi(\theta^*) \quad \hat{Q}_3(\theta) = \pi(\theta^*)' W \hat{\eta}(\theta)
Q_2(\theta) = \frac{1}{2} \hat{\eta}(\theta)' W \hat{\eta}(\theta) + g(\theta)' W (\hat{\pi}(\theta^*) - \pi(\theta^*)) + g(\theta)' W \hat{\eta}(\theta) + (\hat{\pi}(\theta^*) - \pi(\theta^*))' W \hat{\eta}(\theta).
$$

Note that the Taylor expansion of $Q_1(\theta)$ around $\theta^*$ is $Q_1(\theta) = Q_1(\theta^*) + (\theta - \theta^*)' \frac{\partial Q_1(\theta^*)}{\partial \theta} + \frac{1}{2} (\theta - \theta^*)' \frac{\partial^2 Q_1(\theta^*)}{\partial \theta^2} (\theta - \theta^*) + o(\|\theta - \theta^*\|^2) = \frac{1}{2} (\theta - \theta^*)' (\hat{H} + o(1)) (\theta - \theta^*)$ since $\frac{\partial^2 Q_1(\theta^*)}{\partial \theta^2} = G' W g(\theta^*) + G' W \pi(\theta^*) = 0$ and $\frac{\partial^2 Q_1(\theta^*)}{\partial \theta^2} = H$. Next apply Kim and Pollard (1990) Lemma 4.1 to $\hat{\eta}(\theta)$, and in turn $\hat{Q}_3(\theta)$: $\forall \epsilon > 0, \exists M_{n,3} = O_P(1)$ such that

$$
|\hat{Q}_3(\theta)| \leq \epsilon \|\theta - \theta^*\|^2 + n^{-2/3} M_{n,3}^2.
$$

The 1st, 3rd, and 4th terms in $\hat{Q}_2(\theta)$ are all of the form $o_P(1) \hat{\eta}(\theta)$, hence are also bounded by $\epsilon \|\theta - \theta^*\|^2 + n^{-2/3} M_{n,2}^2$. 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

$$
|g(\theta)' W (\hat{\pi}(\theta^*) - \pi(\theta^*))| = O_p \left( \frac{\|\theta - \theta^*\|}{\sqrt{n}} \right) \leq \epsilon \|\theta - \theta^*\|^2 + n^{-2/3} M_{n,22}^2.
$$

Therefore, $\forall \epsilon > 0, \exists M_n = O_P(1)$ such that $|Q_1(\theta) + Q_2(\theta) + \hat{Q}_3(\theta)| \leq \epsilon \|\theta - \theta^*\|^2 + n^{-2/3} M_n^2$. Note that $|Q \left( \hat{\theta}_n \right) - Q \left( \theta^* \right)| \leq |P g \left( \cdot, \hat{\theta}_n \right)' WP g \left( \cdot, \hat{\theta}_n \right)| \leq \left\| P g \left( \cdot, \hat{\theta}_n \right) \right\|^2$. Pick an $\epsilon$ such that $\|P g \left( \cdot, \theta^* \right)\|^2 \leq -2\epsilon \|\theta - \theta^*\|^2$ for $\theta$ in a neighborhood of $\theta^*$. When $\hat{\theta}_n$ lies in this neighborhood,

$$
-o_p \left( n^{-2/3} \right) \leq \inf_{\theta \in \Theta} \hat{Q}_n(\theta) - \hat{Q}_n \left( \hat{\theta}_n \right) \leq |\hat{Q}_n \left( \hat{\theta}_n \right) - \hat{Q}_n \left( \theta^* \right)|
$$

28
\[ \leq \left| \hat{Q}_n \left( \hat{\theta}_n \right) - \hat{Q}_n \left( \theta^* \right) \right| + \left| Q \left( \hat{\theta}_n \right) - Q \left( \theta^* \right) \right| \]
\[ \leq \varepsilon \left\| \hat{\theta}_n - \theta^* \right\|^2 + n^{-2/3} M_n^2 + \left\| P g \left( \cdot, \hat{\theta}_n \right) \right\|^2 \]
\[ \leq \varepsilon \left\| \hat{\theta}_n - \theta^* \right\|^2 + n^{-2/3} M_n^2 - 2 \varepsilon \left\| \hat{\theta}_n - \theta^* \right\|^2 \]
\[ \Rightarrow \left\| \hat{\theta}_n - \theta^* \right\| \leq \varepsilon^{-1/2} n^{-1/3} M_n^2 + o_p \left( n^{-1/3} \right) = O_p \left( n^{-1/3} \right) \]

Therefore, \( \hat{\theta}_n - \theta^* = O_p \left( n^{-1/3} \right) \).

Next note \( \hat{h} = n^{1/3} \left( \hat{\theta}_n - \theta^* \right) \) where \( \hat{h} = \arg \min_h n^{2/3} \hat{Q}_n \left( \theta^* + n^{-1/3} h \right) \). It will follow from the argmax continuous mapping theorem that \( \hat{h} \rightsquigarrow \arg \min_h \pi \left( \theta^* \right) \left( W \mathcal{Z}_{0,1/2} \left( h \right) + \frac{1}{2} h' \hat{H} h \right) \) if we can show that

\[ n^{2/3} \left( \hat{Q}_n \left( \theta^* + n^{-1/3} \right) - \hat{Q}_n \left( \theta^* \right) \right) \rightsquigarrow \pi \left( \theta^* \right) \left( W \mathcal{Z}_{0,1/2} \left( h \right) + \frac{1}{2} h' \hat{H} h \right) \]

as a process indexed by \( h \) in the space of bounded functions \( \ell^\infty \left( K \right) \equiv \{ f : K \mapsto \mathbb{R} \text{ such that } \| f \|_\infty < \infty \} \) for any compact \( K \subset \mathbb{R}^d \). Since \( Q_1 \left( \theta^* + n^{-1/3} h \right) = Q_1 \left( \theta^* \right) + n^{-1/3} h' \frac{\partial Q_1 \left( \theta^* \right)}{\partial \theta} + \frac{1}{2} n^{-2/3} h' \frac{\partial^2 Q_1 \left( \theta^* \right)}{\partial \theta \partial \theta} h + o \left( n^{-2/3} \right) \), \( n^{2/3} Q_1 \left( \theta^* + n^{-1/3} h \right) = \frac{1}{2} h' \hat{H} h + o \left( 1 \right) \).

It remains to show that \( n^{2/3} \left( \hat{Q}_2 \circ \hat{Q}_3 \right) \left( \theta^* + n^{-1/3} h \right) \rightsquigarrow \mathcal{Z}_0 \left( h \right) \). First note that assumption 2(iv) implies that the Lindeberg condition is satisfied. Then the Lindeberg-Feller CLT implies that \( S_n \left( h \right) \equiv n^{2/3} \hat{Q}_n \left( \theta^* + n^{-1/3} h \right) = n^{1/6} \hat{g}_n \left( \theta^* + n^{-1/3} h \right) \) converges in finite dimensional distribution to a mean zero Gaussian process \( \mathcal{Z}_0 \left( h \right) \) with covariance kernel \( \Sigma_{1/2} \left( s, t \right) = \lim_{\alpha \to \infty} \alpha P g \left( \cdot, \theta^* + \frac{s}{\alpha} \right) g \left( \cdot, \theta^* + \frac{t}{\alpha} \right) \).

To show that \( S_n \left( h \right) \) is stochastically equicontinuous, it suffices to show that for every sequence of positive numbers \( \{ \delta_n \} \) converging to zero,

\[ n^{2/3} E_{\mathcal{D}(n)} \sup_{\mathcal{D}(n)} \left| P_n d - P d \right| = o \left( 1 \right) \quad (9.1) \]

where \( \mathcal{D}(n) = \{ d(\cdot, \theta^*, h_1, h_2) = g(\cdot; \theta^* + n^{-1/3} h_1) - g(\cdot; \theta^* + n^{-1/3} h_2) \text{ such that } \max \left( \| h_1 \|, \| h_2 \| \right) \leq M \text{ and } \| h_1 - h_2 \| \leq \delta_n \} \). Note that \( \mathcal{D}(n) \) has envelope function \( D_n = 2G R(n) \) where \( R(n) = M n^{-1/3} \).

Using the Maximal Inequality in section 3.1 of Kim and Pollard (1990), for sufficiently large \( n \), splitting up the expectation according to whether \( n^{1/3} P_n D_n^2 \leq \eta \) for each \( \eta > 0 \), and applying the Cauchy-Schwarz inequality,

\[ n^{2/3} E_{\mathcal{D}(n)} \sup_{\mathcal{D}(n)} \left| P_n d - P d \right| \leq E \sqrt{n^{1/3} P_n D_n^2 J} \left( \frac{n^{1/3} \sup_{\mathcal{D}(n)} P_n d^2}{n^{1/3} P_n D_n^2} \right) \]
\[ \leq \sqrt{n J} \left( 1 + \sqrt{E n^{1/3} P_n D_n^2} \right) \left( \min \left( 1, \frac{1}{\eta n^{1/3} \sup_{\mathcal{D}(n)} P_n d^2} \right) \right). \]
To show that this is $o(1)$ for each fixed $\eta > 0$, first note that by assumption 2(v), $E_n^{1/3} P_n D_n^2 = 4n^{1/3} EC_R^2 = o(n^{1/3} R(n)) = O(1)$ since $R(n) = M n^{-1/3}$. The proof will then be complete if $n^{1/3} \sup_{\mathbb{P}(n)} P_n d^2 = o_p(1)$. Next, for each $K > 0$ write $E_n \sup_{\mathbb{P}(n)} P_n d^2 \leq E P_n \sup_{\mathbb{P}(n)} \{D_n > K\} + K E \sup_{\mathbb{P}(n)} P_n d | |d| | + K E \sup_{\mathbb{P}(n)} P_n |d| - |P| |d|$. By assumption 2(vi), $E P_n D_n^2 \{D_n > K\} < \eta n^{-1/3}$ for large enough $K$. By assumption 2(vii) and the definition of $\mathbb{P}(n)$, $K \sup_{\mathbb{P}(n)} P_n |d| = O(n^{-1/3} \delta_n) = o(n^{-1/3})$. By assumption 2(v) and the maximal inequality in section 3.1 of Kim and Pollard (1990), $K \sup_{\mathbb{P}(n)} P_n |d| - |P| |d| < Kn^{1/2} J(1) \sqrt{P D_n^2} = O(n^{-2/3}) = o(n^{-1/3})$. Therefore, $E_n^{1/3} \sup_{\mathbb{P}(n)} P_n d^2 = o(1)$.

We have shown that $W_n(h) \sim \hat{Z}_0(h)$, which implies that $n^{1/3} \hat{Q}_3(\theta^* + n^{-1/3} h) \sim \hat{Z}_0(h)$. Since the 1st, 3rd and 4th terms in $n^{1/3} \hat{Q}_2(\theta^* + n^{-1/3} h)$ are all of the form $o_p(1) n^{1/3} \check{\eta}_n(\theta^* + n^{-1/3} h)$, they all converge in probability to 0. For the 2nd term there,

$$n^{2/3} |g(\theta^* + n^{-1/3} h)' W (\hat{\pi}(\theta^*) - \pi(\theta^*))| = n^{2/3} O_p \left( \frac{||n^{-1/3} h||}{\sqrt{n}} \right) = O_P(h n^{-1/6}) = o_P(1).$$

Therefore $n^{2/3} \hat{Q}_2(\theta^* + n^{-1/3} h) = o_P(1)$. By Slutsky’s Theorem,

$$n^{2/3} \left( \hat{Q}_1 + \hat{Q}_2 + \hat{Q}_3 \right)(\theta^* + n^{-1/3} h) \sim \hat{Z}_0(h) + \frac{1}{2} h' \hat{H} h.$$

Lemma 2.6 in Kim and Pollard (1990) implies that the Gaussian process $-\hat{Z}_0(h)$ has a unique maximum, which implies that $\hat{Z}_0(h)$ has a unique minimum. In combination with the fact that $\frac{1}{2} h' \hat{H} h$ is a convex function of $h$, there is a unique $\hat{h} = n^{1/3} \left( \hat{\theta}_n - \theta^* \right)$ that minimizes $\hat{Z}_0(h) + \frac{1}{2} h' \hat{H} h$. The result follows from the argmin continuous mapping theorem (Theorem 2.7 in Kim and Pollard (1990)).

**Proof for Theorem 2** The consistency proof is the same as in Theorem 1. Next we show that $n^{1/2} \left( \hat{\theta}_n - \theta^* \right) = O_P(1)$. Recall that $\hat{Q}_n(\theta) - \hat{Q}_n(\theta^*) = Q_1(\theta) + \hat{Q}_2(\theta) + \hat{Q}_3(\theta)$, where

$$Q_1(\theta) = \frac{1}{2} g(\theta)' W g(\theta) + g(\theta)' W \pi(\theta^*) \quad \hat{Q}_3(\theta) = \pi(\theta^*)' W \hat{\eta}_n(\theta)$$

$$\hat{Q}_2(\theta) = \frac{1}{2} \hat{\eta}_n(\theta)' W \hat{\eta}_n(\theta) + g(\theta)' W (\hat{\pi}(\theta^*) - \pi(\theta^*)) + g(\theta)' W \hat{\eta}_n(\theta) + (\pi(\theta^*) - \pi(\theta^*))' W \hat{\eta}_n(\theta).$$

where $\hat{\eta}_n(\theta) = (P_n - P) g(\cdot, \theta), \hat{g}(\theta) = \hat{P}_n g(\cdot, \theta), \text{ and } g(\theta) = P g(\cdot, \theta)$. Note that the Taylor expansion of $Q_1(\theta)$ around $\theta^*$ is $Q_1(\theta) = Q_1(\theta^*) + (\theta - \theta^*)' \nabla Q_1(\theta^*) + \frac{1}{2} (\theta - \theta^*)' \nabla^2 Q_1(\theta^*) (\theta - \theta^*) + o(||\theta - \theta^*||^2) = \frac{1}{2} (\theta - \theta^*)' (\hat{H} + o(1)) (\theta - \theta^*)$ since $\nabla Q_1(\theta^*) = G' W g(\theta^*) + G' W \pi(\theta^*) = 0$ and $\nabla^2 Q(\theta^*) = \hat{H}$. Next apply a modified version of Kim and Pollard (1990) Lemma 4.1 with
\( \gamma = \frac{1}{2}, \rho = 1, \) to \( \hat{\eta}_n(\theta) \), and in turn \( \hat{Q}_3(\theta) : \forall \epsilon > 0, \exists M_{n,3} = O_P(1) \) such that
\[
|\hat{Q}_3(\theta)| \leq \epsilon \|\theta - \theta^*\|^2 + n^{-1}M_{n,3}^2.
\]
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 \( \epsilon \|\theta - \theta^*\|^2 + n^{-1}M_{n,2}^2 \). 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
\[
|g(\theta)W(\hat{\pi}(\theta^*) - \pi(\theta^*))| = O_P\left(\frac{\|\theta - \theta^*\|}{\sqrt{n}}\right) \leq \epsilon \|\theta - \theta^*\|^2 + n^{-1}M_{n,22}^2.
\]
Therefore, \( \forall \epsilon > 0, \exists M_n = O_P(1) \) such that \( |Q_1(\theta) + \hat{Q}_2(\theta) + \hat{Q}_3(\theta)| \leq \epsilon \|\theta - \theta^*\|^2 + n^{-1}M_n^2 \).

Next note \( \hat{\theta}_n = \arg \min h \hat{Q}_n(\theta^* + n^{-1/2}h) \). It will follow from the argmax continuous mapping theorem that \( \hat{\theta}_n \rightsquigarrow \arg \min_{h \in \mathbb{R}^d} \{ \pi(\theta^*)'WZ_{0,1}(h) + h'G'WU_0 + \frac{1}{2}h'Hh \} \) if we can show that
\[
\frac{1}{n} \left( \hat{Q}_n(\theta^* + n^{-1/2}h) - \hat{Q}_n(\theta^*) \right) \rightsquigarrow \pi(\theta^*)'WZ_{0,1}(h) + h'G'WU_0 + \frac{1}{2}h'Hh
\]
as a process indexed by \( h \) in the space of bounded functions \( \ell^\infty(K) \equiv \{ f : K \rightarrow \mathbb{R} \text{ such that } \|f\|_{\infty} < \infty \} \) for any compact \( K \subset \mathbb{R}^d \). Since \( Q_1(\theta^* + n^{-1/2}h) = Q_1(\theta^*) + n^{-1/2}h'\frac{\partial Q_1(\theta^*)}{\partial \theta} + \frac{1}{2}n^{-1}h'Q_1''(\theta^*)h + o(n^{-1}), nQ_1(\theta^* + n^{-1/2}h) = \frac{1}{2}h'HHh + o(1) \).

It remains to show that \( n \left( \frac{\hat{Q}_2 + \hat{Q}_3}{\theta^* + n^{-1/2}h} \right) \rightsquigarrow \pi(\theta^*)'WZ_{0,1}(h) + h'G'WU_0 \) since the 1st, 3rd and 4th terms in \( n\hat{Q}_2(\theta^* + n^{-1/2}h) \) are all of the form \( O_P(1) n \hat{\eta}_n(\theta^* + n^{-1/2}h) \),

\( \text{The main revisions to Lemma 4.1 of Kim and Pollard (1990) are redefining } A(n,j) = (j - 1)n^{-\gamma} \leq ||\theta|| \leq jn^{-\gamma} \), bounding the jth summand in \( P(M_n > m) \) by \( n^{\gamma}P_{\sup_{|\theta| \leq jn^{-\gamma}} |P_n g(\cdot; \theta) - Pg(\cdot, \theta)|^2} \left[ n (j - 1)^2 + m^2 \right]^2 \), where the numerator is further bounded by \( n^{3\gamma} (n^{-1}C'j n^{-\gamma(2\rho)}) = C'j \).
they all converge in probability to 0. For the 2nd term, we can Taylor expand \( g \left( \theta^* + n^{-1/2}h \right) \) around \( \theta^* \):

\[
\sqrt{n} g \left( \theta^* + n^{-1/2}h \right)' W \sqrt{n} (\hat{\pi}(\theta^*) - \pi(\theta^*)) = h' (G + o(1))' W \sqrt{n} (\hat{\pi}(\theta^*) - \pi(\theta^*))
\]

Since we assumed the joint Lindeberg condition: for each \( \epsilon > 0 \) and \( t \in \mathbb{R}^d \),

\[
\lim_{n \to \infty} P \left\{ \left\| \frac{1}{\sqrt{n}} \left( g \left( \cdot, \theta^* + \frac{\cdot}{\sqrt{n}} \right) - g \left( \cdot, \theta^* \right) \right) \right\| > \epsilon \right\} = 0
\]

the Lindeberg-Feller CLT implies that \( S_n(h) \equiv \left( \frac{1}{\sqrt{n}} W n\hat{\eta} \left( \theta^* + n^{-1/2}h \right) \right) \) converges in finite dimensional distribution to \( \left( \frac{\pi(\theta^*)' W Z_{0,1}(h)}{h'G'WU_0} \right) \), where \( Z_{0,1}(h) \) is a mean zero Gaussian process with covariance kernel \( \Sigma_1(s,t) = \lim_{\alpha \to \infty} \alpha^2 P g(\cdot, \theta^* + \frac{\cdot}{\alpha}) g(\cdot, \theta^* + \frac{\cdot}{\alpha})' \), and \( U_0 \sim N(0, P (\pi(\cdot, \theta^*) - \pi(\theta^*)) (\pi(\cdot, \theta^*) - \pi(\theta^*))' \).

Since \( h'G'W \sqrt{n} (\hat{\pi}(\theta^*) - \pi(\theta^*)) \) is a linear (and therefore convex) function of \( h \), pointwise convergence implies uniform convergence over compact sets \( K \subset \mathbb{R}^d \) (Pollard (1991)). Therefore, to show that \( S_n(h) \) is stochastically equicontinuous, it suffices to show that for every sequence of positive numbers \( \{\delta_n\} \) converging to zero,

\[
 nE\sup_{\mathcal{D}(n)} |P_n d - Pd| = o(1) \tag{9.2}
\]

where \( \mathcal{D}(n) = \{d(\cdot, \theta^*, h_1, h_2) = g(\cdot, \theta^* + n^{-1/2}h_1) - g(\cdot, \theta^* + n^{-1/2}h_2) \text{ such that } \max(\|h_1\|, \|h_2\|) \leq M \text{ and } \|h_1 - h_2\| \leq \delta_n \} \). Note that \( \mathcal{D}(n) \) has envelope function \( D_n = 2G_{R(n)} \) where \( R(n) = Mn^{-1/2} \).

Using the Maximal Inequality in section 3.1 of Kim and Pollard (1990), for sufficiently large \( n \), splitting up the expectation according to whether \( nP_nD_n^2 \leq \eta \) for each \( \eta > 0 \), and applying the Cauchy-Schwarz inequality,

\[
 nE\sup_{\mathcal{D}(n)} |P_n d - Pd| \leq E \sqrt{nP_n D_n^2} J \left( \frac{n\sup_{\mathcal{D}(n)} P_n d^2}{nP_n D_n^2} \right) \leq \sqrt{\eta} J(1) + \sqrt{E nP_n D_n^2} \frac{\left[ EJ^2 \left( \min \left( \frac{1}{\eta}, \frac{1}{n\sup_{\mathcal{D}(n)} P_n d^2} \right) \right) \right]}{\eta}.
\]

To show that this is \( o(1) \) for each fixed \( \eta > 0 \), first, note that by assumption 2(v), \( EnP_n D_n^2 = 4nEC_{R(n)}^2 = O(\eta R(n)^2) = O(1) \) since \( R(n) = Mn^{-1/2} \). The proof will then be complete if \( n\sup_{\mathcal{D}(n)} P_n d^2 = o_p(1) \).

For each \( K > 0 \) write \( E\sup_{\mathcal{D}(n)} P_n d^2 \leq E\sup_{\mathcal{D}(n)} P_n d^2 \{D_n > K\} + KE\sup_{\mathcal{D}(n)} |P_n d| \leq E\sup_{\mathcal{D}(n)} P_n d^2 \{D_n > K\} + K \sup_{\mathcal{D}(n)} P_n d^2 + KE\sup_{\mathcal{D}(n)} |P_n d| - P|d| \). By assumption 2(vi), \( E\sup_{\mathcal{D}(n)} P_n d^2 \{D_n > K\} < \eta n^{-1} \) for
large enough $K$. By assumption 2(vii) and the definition of $\mathcal{D}(n)$, $K \sup P|d| = O(n^{-1}\delta_n) = o(n^{-1})$. Using the assumption that $g(\cdot, \theta)$ is Lipschitz in $\theta$, so that $D_n = O_P(n^{-1/2}\delta_n)$, and the maximal inequality in section 3.1 of Kim and Pollard (1990), $K \sup P|d| - P|d| < K n^{-\frac{3}{2}} J(1) / \sqrt{PD_n^2} = O(n^{-1}\delta_n) = o(n^{-1})$. Therefore, $\sup P_n |d| = o(1)$. It follows that

$$\left(\pi(\theta^*)' W n\hat{\gamma}_n (\theta^* + n^{-1/2}h) + h' G W \sqrt{n}(\hat{\pi}(\theta^*) - \pi(\theta^*))\right) \Rightarrow \left(\pi(\theta^*)' W Z_{0,1} (h) + h' G W U_0\right)$$

as a process indexed by $h$ in the space of bounded functions $\ell^\infty(K) \equiv \{f : K \mapsto \mathbb{R} \text{ such that } \|f\|_\infty < \infty\}$ for any compact $K \subset \mathbb{R}^d$. By Slutsky’s Theorem,

$$n \left(\hat{Q}_1 + \hat{Q}_2 + \hat{Q}_3\right) (\theta^* + n^{-1/2}h) \Rightarrow \pi(\theta^*)' W Z_{0,1} (h) + h' G W U_0 + \frac{1}{2} h' \tilde{H} h.$$ 

Lemma 2.6 in Kim and Pollard (1990) implies that the Gaussian process $\pi(\theta^*)' W Z_{0,1} (h)$ has a unique minimum. In combination with the fact that $h' G W U_0 + \frac{1}{2} h' \tilde{H} h$ is a convex function of $h$, there is a unique $\hat{h} = n^{1/2} (\hat{\theta}_n - \theta^*)$ that minimizes $\pi(\theta^*)' W Z_{0,1} (h) + h' G W U_0 + \frac{1}{2} h' \tilde{H} h$. The result follows from the argmin continuous mapping theorem (Theorem 2.7 in Kim and Pollard (1990)).

**Proof for Theorem 3** Equation 5.2 implies that for $\hat{h}^* = n^{\gamma} (\hat{\theta}_n - \hat{\theta}_n),$

$$\hat{h}^* = \arg\min_{h \in \mathbb{R}^d} \pi(\hat{\theta}_n)' W n^{\gamma^2} \sqrt{n} (P_n^* - P_n) \left(\pi(\cdot, \hat{\theta}_n + \frac{h}{n^{\gamma}}) - \pi(\cdot, \hat{\theta}_n)\right) + \frac{\sqrt{n} n^{\gamma^2}}{2 n^{2\gamma}} h' G W \hat{G} + \sum_{j=1}^m \sum_{k=1}^m W_{jk} n^{\gamma^2} h(\hat{\theta}_n) \hat{H}_j h + n^{\gamma^2} h' G' W \sqrt{n} (P_n^* - P_n) \pi(\cdot, \hat{\theta}_n)$$

Assumptions 2 (iii) and (iv) imply the Lindeberg condition is satisfied, so by the Lindeberg-Feller CLT, $W_n(h) \equiv n^{\gamma^2} \sqrt{n} (P_n - P) \left(\pi(\cdot, \theta^* + \frac{h}{n^{\gamma}}) - \pi(\cdot, \theta^*)\right)$ converges in finite dimensional distribution to a mean zero Gaussian process $Z_{0,\rho}(h)$ with covariance kernel $\Sigma_{\rho}(s, t) = \lim_{\alpha \to \infty} a^{2\rho} P g(\cdot, \theta^* + \frac{h}{\alpha}) g(\cdot, \theta^* + \frac{h}{\alpha})'$. We already showed in Theorem 1 that $W_n(h)$ is stochastically equicontinuous for $\rho = 1/2, \gamma = 1/3$.

Therefore, $S_n(h) \Rightarrow Z_{0,\rho}(h)$ as a process indexed by $h$ in $\ell^\infty(K)$ for any compact $K \subset \mathbb{R}^d$. Theorem 3.6.13 in van der Vaart and Wellner (1996) or Theorem 2.6 in Kosorok (2007) then implies that the bootstrapped process $n^{\gamma^2} \sqrt{n} (P_n^* - P_n) \left(\pi(\cdot, \theta^* + \frac{h}{n^{\gamma}}) - \pi(\cdot, \theta^*)\right)$ is consistent for the same limiting process as $S_n(h):$

$$n^{\gamma^2} \sqrt{n} (P_n^* - P_n) \left(\pi(\cdot, \theta^* + \frac{h}{n^{\gamma}}) - \pi(\cdot, \theta^*)\right) \Rightarrow Z_{0,\rho}(h)$$

33
Under the envelope integrability assumption 3, Lemma 4.2 in Wellner and Zhan (1996) implies that
\[
\sup_{h \in \mathbb{R}^d} \left\| n^{\gamma \rho} \sqrt{n} (P_n^* - P_n) \left( \pi \left( \cdot, \hat{\theta}_n + \frac{h}{n^{\gamma}} \right) - \pi \left( \cdot, \hat{\theta}_n \right) - \left( \pi \left( \cdot, \theta^* + \frac{h}{n^{\gamma}} \right) - \pi \left( \cdot, \theta^* \right) \right) \right\| = o_p \left( 1 + n^{\gamma} \left\| \hat{\theta}_n - \theta^* \right\| \right) = o_p^* (1)
\]
In combination with the fact that \( \hat{\pi} \left( \hat{\theta}_n \right) \xrightarrow{p} \pi (\theta^*) \),
\[
\hat{\pi} \left( \hat{\theta}_n \right)' W n^{\gamma \rho} \sqrt{n} (P_n^* - P_n) \left( \pi \left( \cdot, \hat{\theta}_n + \frac{h}{n^{\gamma}} \right) - \pi \left( \cdot, \hat{\theta}_n \right) - \left( \pi \left( \cdot, \theta^* + \frac{h}{n^{\gamma}} \right) - \pi \left( \cdot, \theta^* \right) \right) \xrightarrow{\mathcal{W}} \pi (\theta^*)' W \mathcal{Z}_{0, \rho} (h)
\]
For the second term, note that since \( \sqrt{\frac{n^{\gamma \rho}}{n^{2 \gamma}}} = 1 \), \( \hat{G} \xrightarrow{p} G, \hat{H}_j \xrightarrow{p} H_j \) for \( j = 1 \ldots m \), and \( \hat{\pi} \left( \hat{\theta}_n \right) \xrightarrow{p} \pi (\theta^*) \),
\[
\frac{\sqrt{n^{\gamma \rho}}}{2n^{2 \gamma}} h' \left( \hat{G}' W \hat{G} + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{jk} \hat{\pi}_k \left( \hat{\theta}_n \right) \hat{H}_j \right) h \xrightarrow{p} \frac{1}{2} h' \left( G' W G + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{jk} \pi (\theta^*) H_j \right) h \equiv \frac{1}{2} h' \bar{H} h
\]
When \( \gamma = 1/3 \) and \( \rho = 1/2 \), \( \frac{n^{\gamma \rho}}{n^{2 \gamma}} = o(1) \), which implies that the third term is \( o_p^* (1) \):
\[
\frac{n^{\gamma \rho}}{n^{2 \gamma}} h' \hat{G}' W \sqrt{n} (P_n^* - P_n) \pi \left( \cdot, \hat{\theta}_n \right) = o_p^* (1)
\]
It then follows from a bootstrapped version of the argmin continuous mapping theorem (see Lemma 14.2 in Hong and Li (2020) for proof)
\[
\hat{h}^* \xrightarrow{\mathcal{W}} \arg \min_{h \in \mathbb{R}^d} \left\{ \pi (\theta^*)' W \mathcal{Z}_{0, \rho} (h) + \frac{1}{2} h' \bar{H} h \right\}
\]
For misspecified nonsmooth models with \( \gamma = 1/2 \) and \( \rho = 1 \), \( \frac{n^{\gamma \rho}}{n^{2 \gamma}} = 1 \), so the third term also contributes to the asymptotic distribution.

Next, since we showed in Theorem 2 that
\[
\left( \hat{\pi} (\theta^*)' W n^{\gamma \rho} \sqrt{n} (P_n^* - P_n) \left( \pi \left( \cdot, \theta^* + \frac{h}{n^{\gamma}} \right) - \pi \left( \cdot, \theta^* \right) \right) \right) \xrightarrow{\mathcal{W}} \left( \pi (\theta^*)' W \mathcal{Z}_{0, \rho} (h) \right) \xrightarrow{\mathcal{W}} \left( \pi (\theta^*)' W \mathcal{Z}_{0, \rho} (h) \right)
\]
is stochastically equicontinuous, under the envelope integrability assumption 3, Lemma 4.2 in Wellner and Zhan (1996) implies that the process is bootstrap equicontinuous. Therefore,
\[
\left( \hat{\pi} \left( \hat{\theta}_n \right)' W n^{\gamma \rho} \sqrt{n} (P_n^* - P_n) \left( \pi \left( \cdot, \hat{\theta}_n + \frac{h}{n^{\gamma}} \right) - \pi \left( \cdot, \hat{\theta}_n \right) \right) \right) \xrightarrow{\mathcal{W}} \left( \pi (\theta^*)' W \mathcal{Z}_{0, \rho} (h) \right) \xrightarrow{\mathcal{W}} \left( \pi (\theta^*)' W \mathcal{Z}_{0, \rho} (h) \right)
\]
And it follows from a bootstrapped version of the argmin continuous mapping theorem (see Lemma 14.2 in Hong and Li (2020) for proof)
\[
\hat{h}^* \xrightarrow{\mathcal{W}} \arg \min_{h \in \mathbb{R}^d} \left\{ \pi (\theta^*)' W \mathcal{Z}_{0, 1} (h) + h' G' W U_0 + \frac{1}{2} h' \bar{H} h \right\}
\]
Under correct model specification, $\pi(\theta^*) = 0$, so the first term $\pi(\theta^*)'WZ_{0,1}(h)$ disappears and

$$
\hat{h}^* \overset{P}{\sim} \underset{h \in \mathbb{R}^d}{\arg\min} \left\{ \frac{1}{2} h'G'WGh + h'G'WU_0 \right\} = - (G'WG)^{-1} G'WN (0, P\pi(\cdot, \theta^*)\pi(\cdot, \theta^*))'
$$

For smooth models that are misspecified, $\hat{\pi}(\hat{\theta}_n) W_n (P_n - P) \left( \pi(\cdot, \hat{\theta}_n + \frac{h}{n^{1/2}}) - \pi(\cdot, \hat{\theta}_n) \right) \overset{P}{\sim} W_{0,1}(h) W(\pi(\theta^*), \theta^*) = h'Z_0'W(\theta^*)$ which is a mean zero Gaussian random variable with covariance matrix $P \left( \frac{\partial}{\partial \theta} \pi(\cdot, \theta^*) - G \right)' W(\pi(\theta^*)) W(\pi(\theta^*))' W \left( \frac{\partial}{\partial \theta} \pi(\cdot, \theta^*) - G \right)$.

Furthermore, the joint distribution of $Z_0 W(\pi(\theta^*))$ and $U_0$ is given by

$$
\begin{pmatrix}
U_0 \\
Z_0 W(\pi(\theta^*))
\end{pmatrix}
\sim
N
\left( 0, 
\begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix}
\right)
$$

\[
\Sigma_{11} = P(\pi(\cdot, \theta^*) - \pi(\theta^*))' (\pi(\cdot, \theta^*) - \pi(\theta^*))' \\
\Sigma_{12} = P(\pi(\cdot, \theta^*) - \pi(\theta^*))' W \left( \frac{\partial}{\partial \theta} \pi(\cdot, \theta^*) - G \right) \\
\Sigma_{21} = P \left( \frac{\partial}{\partial \theta} \pi(\cdot, \theta^*) - G \right)' W(\pi(\theta^*)) (\pi(\cdot, \theta^*) - \pi(\theta^*))' \\
\Sigma_{22} = P \left( \frac{\partial}{\partial \theta} \pi(\cdot, \theta^*) - G \right)' W(\pi(\theta^*)) (\pi(\cdot, \theta^*) - G)'
\]

Therefore, the asymptotic distribution is given by

$$
\hat{h}^* \overset{P}{\sim} \underset{h \in \mathbb{R}^d}{\arg\min} \left\{ h'Z_0'W(\pi(\theta^*)) + h'G'WU_0 + \frac{1}{2} h'\bar{H}h \right\} = -\bar{H}^{-1} N (0, G'W\Sigma_{11}WG + \Sigma_{22} + G'W\Sigma_{12} + \Sigma_{21}WG)
$$

\[\blacksquare\]

**Proof for Theorem 4** The consistency proof is a direct application of Corollary 3.2.3 in van der Vaart and Wellner (1996) after we show that $\underset{\theta \in \Theta}{\sup} | \tilde{Q}_n(\theta) - Q(\theta) | = o_p(1)$. Since $W_n - W = o_p(1)$, $\underset{\theta \in \Theta}{\sup} \| P_n \pi(\cdot, \theta) - P \pi(\cdot, \theta) \| = o_p(1)$, and $\underset{\theta \in \Theta}{\sup} P_n \pi(\cdot, \theta) = O_p(1)$,

$$
\underset{\theta \in \Theta}{\sup} | P_n \pi(\cdot, \theta)' W_n P_n \pi(\cdot, \theta) - P \pi(\cdot, \theta)' WP \pi(\cdot, \theta) |
\leq \underset{\theta \in \Theta}{\sup} (P_n \pi(\cdot, \theta) - P \pi(\cdot, \theta))' W (P_n \pi(\cdot, \theta) - P \pi(\cdot, \theta)) + \underset{\theta \in \Theta}{\sup} | P_n \pi(\cdot, \theta)' (W_n - W) P_n \pi(\cdot, \theta) |
\leq \underset{\theta \in \Theta}{\sup} \| P_n \pi(\cdot, \theta) - P \pi(\cdot, \theta) \|' W \underset{\theta \in \Theta}{\sup} \| P_n \pi(\cdot, \theta) - P \pi(\cdot, \theta) \| + \underset{\theta \in \Theta}{\sup} | P_n \pi(\cdot, \theta)' (W_n - W) P_n \pi(\cdot, \theta) |
= o_p(1)
$$
Next, write $\hat{Q}_n (\theta) - \hat{Q}_n (\theta^*) = Q_1 (\theta) + \hat{Q}_2 (\theta) + \hat{Q}_3 (\theta) + \hat{Q}_4 (\theta) + \hat{Q}_5 (\theta) + \hat{Q}_6 (\theta)$, where

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

\[
\hat{Q}_2 (\theta) = \frac{1}{2} \hat{n}_n (\theta)' W \hat{n}_n (\theta) + g (\theta)' W (\hat{\pi} (\theta^*) - \pi (\theta^*)) + g (\theta)' W \hat{n}_n (\theta) + (\hat{\pi} (\theta^*) - \pi (\theta^*))' W \hat{n}_n (\theta)
\]

\[
\hat{Q}_4 (\theta) = \frac{1}{2} g (\theta)' (W_n - W) g (\theta) + g (\theta)' (W_n - W) \pi (\theta^*)
\]

\[
\hat{Q}_5 (\theta) = g (\theta)' (W_n - W) (\hat{\pi} (\theta^*) - \pi (\theta^*)) + g (\theta)' (W_n - W) \hat{n}_n (\theta) + (\hat{\pi} (\theta^*) - \pi (\theta^*))' (W_n - W) \hat{n}_n (\theta)
\]

\[
\hat{Q}_6 (\theta) = \pi (\theta^*)' (W_n - W) \hat{n}_n (\theta) + \frac{1}{2} \hat{n}_n (\theta)' (W_n - W) \hat{n}_n (\theta)
\]

We already showed in Theorem 1 that $\forall \epsilon > 0$, there exists $M_n = O_p (1)$ such that $|Q_1 (\theta) + \hat{Q}_2 (\theta) + \hat{Q}_3 (\theta)| \leq \epsilon \|\theta - \theta^*\|^2 + n^{-2/3} M_n^2$.

Next recall that Kim and Pollard (1990) Lemma 4.1 applied to $\hat{n}_n (\theta)$, and in turn $\hat{Q}_6 (\theta) = o_p (1) \hat{n}_n (\theta)$ implies that $\forall \epsilon > 0$, $\exists M_{n, 6} = O_p (1)$ such that

\[
|\hat{Q}_6 (\theta)| \leq \epsilon \|\theta - \theta^*\|^2 + n^{-2/3} M_{n, 6}^2.
\]

The 2nd and 3rd terms in $\hat{Q}_5 (\theta)$ are also of the form $o_p (1) \hat{n}_n (\theta)$, hence are also bounded by $\epsilon \|\theta - \theta^*\|^2 + n^{-2/3} M_{n, 51}^2$, for some $M_{n, 51} = O_p (1)$ and $\forall \epsilon > 0$. The 1st term in $\hat{Q}_5 (\theta)$ can also be bounded by, for some $M_{n, 52} = O_p (1)$ and $\forall \epsilon > 0$,

\[
|g (\theta)' (W_n - W) (\hat{\pi} (\theta^*) - \pi (\theta^*))| = o_p \left( \frac{\|\theta - \theta^*\|}{\sqrt{n}} \right) \leq \epsilon \|\theta - \theta^*\|^2 + n^{-2/3} M_{n, 52}^2.
\]

If $W_n - W = O_p (n^{-\gamma})$ for $1/3 \leq \gamma \leq 1/2$, then $\frac{\partial \hat{Q}_4 (\theta^*)}{\partial \theta} = G' (W_n - W) g (\theta^*) + G' (W_n - W) \pi (\theta^*) = O_p (n^{-\gamma})$. Taylor expanding $\hat{Q}_4 (\theta)$ around $\theta^*$ gives for some $M_{n, 4} = O_p (1)$ and $\forall \epsilon > 0$,

\[
\hat{Q}_4 (\theta) = \hat{Q}_4 (\theta^*) + (\theta - \theta^*)' \frac{\partial \hat{Q}_4 (\theta^*)}{\partial \theta} + \frac{1}{2} (\theta - \theta^*)' \frac{\partial^2 \hat{Q}_4 (\theta^*)}{\partial \theta \partial \theta^*} (\theta - \theta^*) + o_p (\|\theta - \theta^*\|^2)
\]

\[
= \frac{1}{2} (\theta - \theta^*)' \left( \frac{\partial^2 \hat{Q}_4 (\theta^*)}{\partial \theta \partial \theta^*} + o_p (1) \right) (\theta - \theta^*) + o_p \left( \frac{\|\theta - \theta^*\|}{n^\gamma} \right)
\]

\[
\leq \epsilon \|\theta - \theta^*\|^2 + n^{-2\gamma} M_{n, 4}^2.
\]

Then $\forall \epsilon > 0$, there exists $M_n = O_p (1)$ such that $|\hat{Q} (\theta)| = |Q_1 (\theta) + \hat{Q}_2 (\theta) + \hat{Q}_3 (\theta) + \hat{Q}_4 (\theta) + \hat{Q}_5 (\theta) + \hat{Q}_6 (\theta)| \leq \epsilon \|\theta - \theta^*\|^2 + n^{-2/3} M_n^2$. Note that $|\hat{Q} (\hat{n}_n) - Q (\theta^*)| \leq |P g (\cdot, \hat{n}_n)' W g (\cdot, \hat{n}_n)| \leq \|P g (\cdot, \hat{n}_n)\|^2$. Pick an $\epsilon$ such that $\|P g (\cdot, \theta)\|^2 \leq -2 \epsilon \|\theta - \theta^*\|^2$ for $\theta$ in a neighborhood of $\theta^*$. When $\hat{n}_n$ lies in this neighborhood,

\[
-o_p (n^{-2/3}) \leq \inf_{\theta \in \Theta} \hat{Q}_n (\theta) - \hat{Q}_n (\hat{n}_n) \leq \left| \hat{Q}_n (\hat{n}_n) - \hat{Q}_n (\theta^*) \right| \leq \left| \hat{Q}_n (\hat{n}_n) - \hat{Q}_n (\theta^*) \right| + \left| Q (\hat{n}_n) - Q (\theta^*) \right|
\]

36
We already showed in Theorem 1 that $n^{2/3}Q_1 \left( \theta^* + n^{-1/3}h \right) = \frac{1}{2} h' \bar{H} h + o(1)$, $n^{2/3}Q_3 \left( \theta^* + n^{-1/3}h \right) \rightsquigarrow \pi(\theta^*)' W Z_0(h)$, and $n^{2/3}Q_2 \left( \theta^* + n^{-1/3}h \right) = o_p(1)$. Furthermore, if $W_n - W = O_p \left( n^{-1/3} \right)$,

$$n^{2/3} \hat{Q}_4 \left( \theta^* + n^{-1/3}h \right) = \frac{1}{2} n^{2/3} g \left( \theta^* + n^{-1/3}h \right)' (W_n - W) g \left( \theta^* + n^{-1/3}h \right) + n^{2/3} g \left( \theta^* + n^{-1/3}h \right)' (W_n - W) \pi \left( \theta^* \right)$$

$$= n^{2/3} O_p \left( \frac{\|n^{-1/3}h\|}{n^{1/3}} \right) + O_p \left( \frac{\|n^{-1/3}h\|}{n^{1/3}} \right) O_p(1) + o_p(1)$$

$$n^{2/3} \hat{Q}_5 \left( \theta^* + n^{-1/3}h \right) = n^{2/3} g \left( \theta^* + n^{-1/3}h \right)' (W_n - W) \left( \hat{\pi}(\theta^*) - \pi(\theta^*) \right) + g \left( \theta^* + n^{-1/3}h \right)' \left( W_n - W \right) n^{2/3} \hat{\eta}_n \left( \theta^* + n^{-1/3}h \right) + \left( \hat{\pi}(\theta^*) - \pi(\theta^*) \right)' \left( W_n - W \right) n^{2/3} \hat{\eta}_n \left( \theta^* + n^{-1/3}h \right)$$

$$= n^{2/3} O_p \left( \frac{\|n^{-1/3}h\|}{n^{5/6}} \right) + O_p \left( \frac{\|n^{-1/3}h\|}{n^{1/3}} \right) O_p(1) + o_p(1)$$

$$n^{2/3} \hat{Q}_6 \left( \theta^* + n^{-1/3}h \right) = \pi(\theta^*)' (W_n - W) n^{2/3} \hat{\eta}_n \left( \theta^* + n^{-1/3}h \right) + \frac{1}{2} \hat{\eta}_n \left( \theta^* + n^{-1/3}h \right)' \left( W_n - W \right) n^{2/3} \hat{\eta}_n \left( \theta^* + n^{-1/3}h \right)$$

$$= O_p \left( n^{-1/3} \right) O_p(1) + o_p(1)$$

By assumption,

$$\left( \begin{array}{c} \pi(\theta^*)' W n^{2/3} \left( P_n - P \right) g \left( \cdot, \theta^* + n^{-1/3}h \right) \\ h' G' n^{1/3} (W_n - W) \pi(\theta^*) \end{array} \right) \rightsquigarrow \left( \begin{array}{c} \pi(\theta^*)' W Z_{0,1/2}(h) \\ h' G' W_0 \end{array} \right)$$

Therefore, by Slutsky’s theorem and the argmin continuous mapping theorem,

$$n^{1/3} \hat{\theta}_n - \theta^* \rightsquigarrow \arg \min_{h \in \mathbb{R}^d} \left\{ \pi(\theta^*)' W \pi(\theta^*)' W Z_{0,1/2}(h) + h' G' W_0 + \frac{1}{2} h' \bar{H} h \right\}$$

If $W_n - W = o_p \left( n^{-1/3} \right)$, $n^{1/3} \left( \hat{\theta}_n - \theta^* \right) \rightsquigarrow \arg \min_{h \in \mathbb{R}^d} \left\{ \pi(\theta^*)' W \pi(\theta^*)' W Z_{0,1/2}(h) + \frac{1}{2} h' \bar{H} h \right\}$ because

$$n^{2/3} \hat{Q}_4 \left( \theta^* + n^{-1/3}h \right) = n^{2/3} O_p \left( \frac{\|n^{-1/3}h\|}{n^{1/3}} \right) + n^{2/3} O_p \left( \frac{\|n^{-1/3}h\|}{n^{1/3}} \right)$$
Proof for Theorem 5  The consistency proof is the same as in Theorem 4. Next, write
\[ \hat{Q}_n (\theta) - \hat{Q}_n (\theta^*) = Q_1 (\theta) + \hat{Q}_2 (\theta) + \hat{Q}_3 (\theta) + \hat{Q}_4 (\theta) + \hat{Q}_5 (\theta) + \hat{Q}_6 (\theta), \]
where
\[ Q_1 (\theta) = \frac{1}{2} g (\theta)^\prime W (\theta) + g (\theta)^\prime W \pi (\theta^*), \quad \hat{Q}_3 (\theta) = \pi (\theta)^\prime W \hat{\eta}_n (\theta) \]
\[ \hat{Q}_2 (\theta) = \frac{1}{2} \hat{\eta}_n (\theta)^\prime W \hat{\eta}_n (\theta) + g (\theta)^\prime W (\hat{\pi} (\theta^*) - \pi (\theta^*)) + g (\theta)^\prime W \hat{\eta}_n (\theta) + (\hat{\pi} (\theta^*) - \pi (\theta^*))^\prime W \hat{\eta}_n (\theta) \]
\[ \hat{Q}_4 (\theta) = \frac{1}{2} g (\theta)^\prime (W_n - W) g (\theta) + g (\theta)^\prime (W_n - W) \pi (\theta^*) \]
\[ \hat{Q}_5 (\theta) = g (\theta)^\prime (W_n - W) (\hat{\pi} (\theta^*) - \pi (\theta^*)) + g (\theta)^\prime (W_n - W) \hat{\eta}_n (\theta) + (\hat{\pi} (\theta^*) - \pi (\theta^*))^\prime (W_n - W) \hat{\eta}_n (\theta) \]
\[ \hat{Q}_6 (\theta) = \pi (\theta^*)^\prime (W_n - W) \hat{\eta}_n (\theta) + \frac{1}{2} \hat{\eta}_n (\theta)^\prime (W_n - W) \hat{\eta}_n (\theta) \]

We already showed in Theorem 2 that \( \forall \epsilon > 0 \), there exists \( M_n = O_p (1) \) such that \( |Q_1 (\theta) + \hat{Q}_2 (\theta) + \hat{Q}_3 (\theta)| \leq \epsilon \|\theta - \theta^*\|^2 + n^{-1} M_n^2 \).

A modified version of Kim and Pollard (1990) Lemma 4.1 applied to \( \hat{\eta}_n (\theta) \), and in turn \( \hat{Q}_6 (\theta) = o_p (1) \), implies that \( \forall \epsilon > 0, \exists M_{n,6} = O_P (1) \) such that
\[ |\hat{Q}_6 (\theta)| \leq \epsilon \|\theta - \theta^*\|^2 + n^{-1} M_{n,6}^2. \]

The 2nd and 3rd terms in \( \hat{Q}_5 (\theta) \) are also of the form \( o_p (1) \hat{\eta}_n (\theta) \), hence are also bounded by \( \epsilon \|\theta - \theta^*\|^2 + n^{-1} M_{n,51}^2 \). The 1st term in \( \hat{Q}_5 (\theta) \) can also be bounded by, for some \( M_{n,52} = O_p (1) \) and \( \forall \epsilon > 0, \)
\[ |g (\theta)^\prime (W_n - W) (\hat{\pi} (\theta^*) - \pi (\theta^*))| = o_p \left( \frac{1}{\sqrt{n}} \right) \leq \epsilon \|\theta - \theta^*\|^2 + n^{-1} M_{n,52}^2. \]

Then \( \forall \epsilon > 0 \), there exists \( M_n = O_p (1) \) such that \( |\hat{Q} (\theta)| = |Q_1 (\theta) + \hat{Q}_2 (\theta) + \hat{Q}_3 (\theta) + \hat{Q}_4 (\theta) + \hat{Q}_5 (\theta) + \hat{Q}_6 (\theta)| \leq \epsilon \|\theta - \theta^*\|^2 + n^{-1} M_n^2 \). Similar arguments as in Theorem 4 suggest that \( \hat{\theta}_n - \theta^* = O_p (n^{-1/2}) \).

We already showed in Theorem 2 that \( nQ_1 (\theta^* + n^{-1/2} h) = \frac{1}{2} h^\prime \hat{H} h + o(1) \), and \( n\hat{Q}_2 (\theta^* + n^{-1/2} h) + n\hat{Q}_3 (\theta^* + n^{-1/2} h) \rightarrow \pi (\theta^*)^\prime W Z_{0,1} (h) + h^\prime G^\prime W U_0. \) Furthermore, if \( W_n - W = O_p (n^{-1/2}) \),
\[ n\hat{Q}_4 (\theta^* + n^{-1/2} h) = \frac{1}{2} n g (\theta^* + n^{-1/2} h)^\prime (W_n - W) g (\theta^* + n^{-1/2} h) \]
\[ + n g (\theta^* + n^{-1/2} h)^\prime (W_n - W) \pi (\theta^*) \]
\[ = n O_p \left( \frac{\|n^{-1/2} h\|^2}{n^{1/2}} \right) + n^{1/2} \left\{ g (\theta^*)^\prime + h^\prime G^\prime n^{-1/2} \right\} + o_p (1) \left( n^{1/2} (W_n - W) \pi (\theta^*) \right) \]
\[ \rightarrow h^\prime G^\prime n^{1/2} (W_n - W) \pi (\theta^*) + o_p (1) \]
\[ \sim h^\prime G^\prime \Phi_0 \pi (\theta^*) \]
\[ n\hat{Q}_5(\theta^* + n^{-1/2}h) = n g(\theta^* + n^{-1/2}h) \left( W_n - W \right) (\hat{\pi}(\theta^*) - \pi(\theta^*)) + g(\theta^* + n^{-1/2}h) \left( W_n - W \right) n\hat{\eta}_n (\theta^* + n^{-1/2}h) + (\hat{\pi}(\theta^*) - \pi(\theta^*)) \left( W_n - W \right) n\hat{\eta}_n (\theta^* + n^{-1/2}h) = nO_p \left( \frac{\|n^{-1/2}h\|}{n} \right) + O_p \left( \frac{\|n^{-1/2}h\|}{n^{1/2}} \right) O_p(1) + O_p(n^{-1}) O_p(1) = O_p(n^{-1/2}) + O_p(n^{-1}) + O_p(n^{-1}) = o_p(1) \]
\[ n\hat{Q}_6(\theta^* + n^{-1/2}h) = \pi(\theta^*) \left( W_n - W \right) n\hat{\eta}_n (\theta^* + n^{-1/2}h) + \frac{1}{2} n\hat{\eta}_n (\theta^* + n^{-1/2}h) \left( W_n - W \right) n\hat{\eta}_n (\theta^* + n^{-1/2}h) = O_p(n^{-1/2}) O_p(1) + O_p(n^{-1}) O_p(n^{-1/2}) O_p(1) = o_p(1) \]

The joint Lindeberg condition is satisfied by assumption 4: for each \( \epsilon > 0 \) and \( t \in \mathbb{R}^d \),
\[
\lim_{n \to \infty} P \left\{ \left\| \begin{pmatrix} g(\cdot, \theta^* + \frac{t}{\sqrt{n}}) \\ \pi(\cdot, \theta^*) \\ \phi(\cdot, \theta_1^*) \end{pmatrix} \right\|_2^2 \leq \epsilon \sqrt{n} \} = 0
\]

Therefore, by the Lindeberg-Feller CLT and stochastic equicontinuity arguments similar to those in Theorem 2,
\[
\begin{pmatrix} \pi(\theta^*) W_n n(P_n - P) g(\cdot, \theta^* + n^{-1/2}h) \\ hG'W_n \sqrt{n}(P_n - P) \pi(\cdot, \theta^*) \\ hG'n^{1/2} \sqrt{n}(P_n - P) \phi(\cdot, \theta_1^*) \pi(\theta^*) \end{pmatrix} \sim \begin{pmatrix} \pi(\theta^*) W \mathcal{Z}_{0,1}(h) \\ hG'W U_0 \\ hG'\Phi_0 \pi(\theta^*) \end{pmatrix}
\]

By assumption 4, \( \sqrt{n} (W_n - W) = \sqrt{n} (P_n - P) \phi(\cdot, \theta_1^*) + o_p(1) \); therefore,
\[
\begin{pmatrix} \pi(\theta^*) W_n n(P_n - P) g(\cdot, \theta^* + n^{-1/2}h) \\ hG'W_n \sqrt{n}(P_n - P) \pi(\cdot, \theta^*) \\ hG'n^{1/2} \sqrt{n}(W_n - W) \pi(\theta^*) \end{pmatrix} \sim \begin{pmatrix} \pi(\theta^*) W \mathcal{Z}_{0,1}(h) \\ hG'W U_0 \\ hG'\Phi_0 \pi(\theta^*) \end{pmatrix}
\]

By Slutsky’s theorem and the argmin continuous mapping theorem,
\[
n^{1/2} \left( \hat{\theta}_n - \theta^* \right) \sim \arg\min_{h \in \mathbb{R}^d} \left\{ \pi(\theta^*) W \mathcal{Z}_{0,1}(h) + h'G'W'U_0 + h'G'\Phi_0 \pi(\theta^*) + \frac{1}{2} h'\bar{H}h \right\}
\]

If \( W_n - W = o_p(n^{-1/2}) \),
\[
n\hat{Q}_4(\theta^* + n^{-1/2}h) = no_p \left( \frac{\|n^{-1/2}h\|^2}{\sqrt{n}} \right) + no_p \left( \frac{\|n^{-1/2}h\|}{n^{1/2}} \right) = o_p \left( \frac{1}{\sqrt{n}} \right) + o_p(1) = o_p(1)
\]

which implies \( n^{1/2} \left( \hat{\theta}_n - \theta^* \right) \sim \arg\min_{h \in \mathbb{R}^d} \left\{ \pi(\theta^*) W \mathcal{Z}_{0,1}(h) + h'G'W'U_0 + \frac{1}{2} h'\bar{H}h \right\} \).
Proof for Theorem 6  Equation 5.1 implies that for $\hat{h}^* = n^\gamma (\hat{\theta}_n - \hat{\theta}_n)$,

$$
\hat{h}^* = \arg\min_{h \in \mathbb{R}^d} \pi (\hat{\theta}_n) W_n n^{3\rho} \sqrt{n} \left( P_n^* - P_n \right) \left( \pi \left( \cdot, \hat{\theta}_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \hat{\theta}_n \right) \right)
$$

$$
+ \sqrt{n} n^{\rho} h' \left( \hat{G}'W_n \hat{G} + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{njk} \hat{\pi}_k \left( \hat{\theta}_n \right) \hat{H}_j \right) h
$$

$$
+ \frac{n^{3\rho}}{n^\gamma} h' \hat{G}' W_n \sqrt{n} \left( P_n^* - P_n \right) \pi \left( \cdot, \hat{\theta}_n \right)
$$

$$
+ \frac{n^{3\rho}}{n^\gamma} h' \hat{G}' \left( W_n^* - W_n \right) \hat{\pi} \left( \hat{\theta}_n \right)
$$

We already showed in Theorem 3 that

$$
\pi \left( \hat{\theta}_n \right) W_n n^{3\rho} \sqrt{n} \left( P_n^* - P_n \right) \left( \pi \left( \cdot, \hat{\theta}_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \hat{\theta}_n \right) \right) \xrightarrow{p} \pi (\theta^*)' W \mathcal{Z}_{0,\rho} (h)
$$

Consistency of $W_n$ for $W$ implies that

$$
\pi \left( \hat{\theta}_n \right) W_n n^{3\rho} \sqrt{n} \left( P_n^* - P_n \right) \left( \pi \left( \cdot, \hat{\theta}_n + \frac{h}{n^\gamma} \right) - \pi \left( \cdot, \hat{\theta}_n \right) \right) \xrightarrow{p} \pi (\theta^*)' W \mathcal{Z}_{0,\rho} (h)
$$

We also showed in Theorem 3 that

$$
\frac{n^{3\rho}}{2n^\gamma} h' \left( \hat{G}'W_n \hat{G} + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{njk} \hat{\pi}_k \left( \hat{\theta}_n \right) \hat{H}_j \right) h \xrightarrow{p} \frac{1}{2} h' \hat{H} h
$$

For misspecified nonsmooth models with $\gamma = 1/3$ and $\rho = 1/2$, the third term is $o_p(1)$:

$$
n^{-\gamma/2} h' \hat{G}' W_n \sqrt{n} \left( P_n^* - P_n \right) \pi \left( \cdot, \hat{\theta}_n \right) = o_p(1)
$$

If $W_n - W = O_p \left( n^{-1/3} \right)$ and $W_n^* - W_n = o_p \left( n^{-1/3} \right)$, the fourth term is also $o_p(1)$:

$$
h' \hat{G}' n^{1/3} \left( W_n^* - W_n \right) \hat{\pi} \left( \hat{\theta}_n \right) = o_p(1)
$$

Therefore, only the first two terms contribute to the asymptotic distribution. It follows from a bootstrapped version of the argmin continuous mapping theorem that

$$
\hat{h}^* \xrightarrow{p} \arg\min_{h \in \mathbb{R}^d} \left\{ \pi (\theta^*)' W \mathcal{Z}_{0,1/2} (h) + \frac{1}{2} h' \hat{H} h \right\}
$$

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

$$
\left( \pi (\theta^*)' W^{2/3} \left( P_n^* - P_n \right) g \left( \cdot, \theta^* + n^{-1/3} h \right) \right) \xrightarrow{p} \left( \pi (\theta^*)' W \mathcal{Z}_{0,1/2} (h) \right) h' \hat{G}' W_0
$$

By using the envelope integrability assumption 3 and invoking Lemma 4.2 in Wellner and Zhan (1996),

$$
\left( \hat{\pi}_n \left( \hat{\theta}_n \right) W_n n^{2/3} \left( P_n^* - P_n \right) g \left( \cdot, \hat{\theta}_n + n^{-1/3} h \right) \right) \xrightarrow{p} \left( \pi (\theta^*)' W \mathcal{Z}_{0,1/2} (h) \right) h' \hat{G}' W_0
$$
It follows from a bootstrapped version of the argmin continuous mapping theorem that

\[ \hat{h}^* \xrightarrow{p} \arg \min_{h \in \mathbb{R}^d} \left\{ \pi (\theta^*)' W Z_{0,1/2} (h) + h' G' \mathcal{W}_0 + \frac{1}{2} h' \tilde{H} h \right\} \]

For misspecified nonsmooth models with \( \rho = 1, \gamma = 1/2 \), we already showed in Theorem 5

\[ \begin{pmatrix} \pi (\theta^*)' W_n (P_n - P) g \left( \cdot, \theta^* + n^{-1/2} h \right) \\ h' G' W_n \sqrt{n} (P_n - P) \pi \left( \cdot, \theta^* \right) \\ h' G' n^{1/2} \sqrt{n} (W_n - W) \pi \left( \cdot, \theta^* \right) \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \pi (\theta^*)' W Z_{0,1} (h) \\ h' G' W W_0 \\ h' G' \Phi_0 \pi (\theta^*) \end{pmatrix} \]

Under assumption 4, the bootstrapped weighting matrix can be written as \( \sqrt{n} (W_n^* - W_n) = \sqrt{n} (P_n^* - P_n) \phi (\cdot, \theta_1^*) + o_p(1) \). Therefore,

\[ \begin{pmatrix} \hat{\pi} (\hat{\theta}_n)' W_n n (P_n^* - P_n) g \left( \cdot, \hat{\theta}_n + n^{-1/2} h \right) \\ h' G' \sqrt{n} (P_n^* - P_n) \pi \left( \cdot, \hat{\theta}_n \right) \\ h' G' n^{1/2} \sqrt{n} (W_n^* - W_n) \pi \left( \cdot, \hat{\theta}_n \right) \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \pi (\theta^*)' W Z_{0,1} (h) \\ h' G' W W_0 \\ h' G' \Phi_0 \pi (\theta^*) \end{pmatrix} \]

It follows from a bootstrapped version of the argmin continuous mapping theorem that

\[ \hat{h}^* \xrightarrow{p} \arg \min_{h \in \mathbb{R}^d} \left\{ h' G' W W_0 + \pi (\theta^*)' W Z_{0,1} (h) + h' G' \Phi_0 \pi (\theta^*) + \frac{1}{2} h' \tilde{H} h \right\} . \]

For misspecified smooth models where \( \rho = 1 \) and \( \gamma = 1/2 \), \( \pi (\theta^*)' W Z_{0,1} (h) = h' Z_0 W \pi (\theta^*) \), and the joint distribution of \( U_0, Z_0 W \pi (\theta^*) \), and \( \Phi_0 \pi (\theta^*) \) is given by

\[ \begin{pmatrix} U_0 \\ Z_0 W \pi (\theta^*) \\ \Phi_0 \pi (\theta^*) \end{pmatrix} \sim N \left( 0, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix} \right) \]

Then the asymptotic distribution is given by

\[ \hat{h}^* \xrightarrow{p} \arg \min_{h \in \mathbb{R}^d} \left\{ h' G' W W_0 + h' Z_0 W \pi (\theta^*) + h' G' \Phi_0 \pi (\theta^*) + \frac{1}{2} h' \tilde{H} h \right\} = N \left( 0, \tilde{H}^{-1} \Omega_W \tilde{H}^{-1} \right) \]

\[ \Omega_W \equiv G' W \Sigma_{11} W G + \Sigma_{22} + G' W \Sigma_{12} + \Sigma_{21} W G + G' \Sigma_{33} G + G' W \Sigma_{13} G + G' \Sigma_{31} W G + \Sigma_{23} G + G' \Sigma_{32} \]

Under correct model specification, \( \pi (\theta^*) = 0 \), so the second and third terms disappear:

\[ \hat{h}^* \xrightarrow{p} \arg \min_{h \in \mathbb{R}^d} \left\{ \frac{1}{2} h' G' W G h + h' G' W U_0 \right\} = (G' W G)^{-1} G' W N \left( 0, P \pi (\cdot, \theta^*) \pi (\cdot, \theta^*)' \right) \]

\[ \blacksquare \]
References


Bonhomme, Stéphane and Martin Weidner, “Minimizing sensitivity to model misspecification,” working paper, 2018. 2


Gao, Jiti and Han Hong, “A Computational implementation of GMM,” working paper, 2014. 2


42


