Rate-Adaptive Bootstrap for Possibly Misspecified GMM *

Han Hong† Jessie Li‡

June 30, 2022

We consider inference for possibly misspecified GMM models based on possibly nonsmooth moment conditions. While it is well known that misspecified GMM estimators with smooth moments remain $\sqrt{n}$ consistent and asymptotically normal, globally misspecified nonsmooth 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. 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.

JEL Classification: C10, C15

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 (Hall and Horowitz (1996)). 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 (Creel et al. (2015)).

Misspecified GMM models are studied in, for example, Hall and Inoue (2003), Berkowitz et al. (2012), Guggenberger (2012), Lee (2014), Hansen and Lee (2021),Bonhomme and Weidner (2018), Giurcanu and Presnell (2018), Armstrong and Kolesár (2021), and Cheng et al. (2019). All assume that the sample moment conditions are smooth (in the sense of twice continuously differentiable) or directionally differentiable (in the sense of Gateaux) in the parameters, which allows the GMM estimator to remain $\sqrt{n}$ consistent and asymptotically normal. In the case of smooth moments, Hall and Inoue (2003) derived the asymptotic distribution of globally misspecified GMM estimators in the sense that the population moments are equal to a vector of fixed nonzero constants that do not approach zero as $n \to \infty$. They show that the globally misspecified smooth GMM estimator is still $\sqrt{n}$-consistent and asymptotically normal, except with a different variance-covariance matrix than the correctly specified case. In contrast, 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). This nonstandard distribution cannot be estimated consistently by any of the current methods for bootstrapping GMM estimators (for example the standard (nonparametric) bootstrap, centered bootstrap of Hall and Horowitz (1996), or empirical likelihood bootstrap of Brown and Newey (2002)) because convergence to this limiting distribution is not locally uniform in the underlying DGP (Lehmann and Romano (2006)). However, other resampling methods such as subsampling (Politis et al. (1999)) or the numerical bootstrap (Hong and Li (2020)) will work, assuming that we know the rate of convergence. In other words, we need to know whether the model is correctly or incorrectly specified because if the nonsmooth GMM estimator is correctly specified, then the asymptotic distribution remains $\sqrt{n}$-consistent and asymptotically normal.

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. They can overcome the inconsistency of the standard bootstrap because they are bootstrapping consistent estimates of the components of the non-standard limiting distribution rather than applying the bootstrap to the objective function of the M-estimator. 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 globally incorrectly specified, smooth or nonsmooth. Our rate-adaptive bootstrap procedure differs from the one in Cattaneo et al. (2020) because our focus is on GMM, which is not handled by their procedure for M-estimators. 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 globally incorrectly specified, the rate-adaptive bootstrap confidence intervals achieve the nominal coverage asymptotically for the pseudo-true parameter, which is defined as the parameter which minimizes the population GMM objective function.

Both Lee (2014) and Giurcanu and Presnell (2018) have proposed bootstrap procedures that are robust to misspecification, but neither allows for the moment conditions to be nonsmooth. Lee (2014) used Hall and Inoue (2003)'s misspecification-robust estimator of the asymptotic variance of GMM to develop a misspecification-robust (MR) bootstrap procedure. We investigate their procedure in Section 7.4's Monte Carlo study and find that our procedure has similar performance to theirs when the moments are smooth. Giurcanu and Presnell (2018)'s procedure is similar to a guidebook on when to use various existing tests and bootstrap methodologies in the literature. They recommend first testing for misspecification using a J-test and then applying either the standard bootstrap, centered bootstrap of Hall and Horowitz (1996), or empirical likelihood bootstrap of Brown and Newey (2002) depending on the outcome of the test. In contrast to Giurcanu and Presnell (2018), our procedure does not test for misspecification but instead adaptively performs inference for the pseudo-true parameter under misspecification. However, we are similar to Giurcanu and Presnell (2018) in that we also find that the choice of the weighting matrix impacts the GMM estimator's asymptotic distribution.

Several important papers have considered another form of misspecification which arises in the context of two-step semiparametric GMM estimators, where the lack of precision in the first stage nonparametric estimator can make traditional normal confidence intervals suffer from extreme undercoverage. Cattaneo and Jansson (2018) propose novel bootstrap percentile confidence intervals which provide an automatic method of bias correction and are therefore "robust" to first stage misspecification. Their intervals are derived from a new bootstrap distributional approximation based on small bandwidth asymptotics. In a recent paper, Cattaneo and Jansson (2021) consider the problem of estimating the average density of a continuously distributed random vector and show that the nonparametric bootstrap can consistently estimate the distribution of the simple plug-in estimator even though the estimator is known to be biased. This automatic bias correction property is qualitatively related to the ability of the rate-adaptive bootstrap to automatically select in or select out certain components of the asymptotic distribution depending on the level of smoothness and specification of the moments.

Section 2 explains in greater detail the different impacts that global misspecification has on the asymptotic distribution of GMM when the moments are smooth versus nonsmooth. We show that misspecification under the nonsmooth case is of more concern because the rate of convergence becomes cubic-root and the asymptotic distribution becomes nonstandard, thus invalidating the standard bootstrap or inference using asymptotic critical values. We explain how our rate-adaptive bootstrap can still provide consistent inference for this nonsmooth case as well as for smooth case under global misspecification. We also provide three examples illustrating the applicability of our method: 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 4 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 5 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 and an instrumental quantile regession model with misspecified non-smooth moments. Section 6 concludes. The Appendix contains additional theoretical results and proofs of the theorems, in addition to another Monte Carlo example with misspecified smooth moments, where the rate-adaptive bootstrap performs just as well as the standard bootstrap in terms of empirical coverage and average interval width.

2 GMM model with fixed weighting matrix

Consider a random sample $X_1, X_2, ..., X_n$ of independent draws from a probability measure P on a sample space $\mathcal{X}$. Define the empirical measure $P_n \equiv \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$, where δ_x is the measure that assigns mass 1 at x and zero everywhere else. Denote the bootstrap empirical measure by P_n^*, which can refer to the multinomial, wild, or other exchangeable bootstraps. Weak convergence is defined in the sense of Kosorok (2007): $X_n \rightsquigarrow X$ 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): $X_n \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} X$ in the metric space $(\mathbb{D}, d)$ if and only if $\sup_{f\in BL_1}|E_{\mathbb{W}}f(X_n) - Ef(X)| \xrightarrow{p} 0$ and $E_{\mathbb{W}}f(X_n)^* - E_{\mathbb{W}}f(X_n)_* \xrightarrow{p} 0$ for all $f \in BL_1$. $E_{\mathbb{W}}$ denotes expectation with respect to the bootstrap weights $\mathbb{W}$ conditional on the data, and $f(X_n)^*$ and $f(X_n)_*$ denote measurable majorants and minorants with respect to the joint data (including the weights $\mathbb{W}$). Let $X_n^* = o_P^*(1)$ if the law of X_n^* is governed by P_n and if $P_n(|X_n^*| > \epsilon) = o_P(1)$ for all $\epsilon > 0$. Also define $M_n^* = O_P^*(1)$ (hence also $O_P(1)$) if $\lim_{m\to\infty}\limsup_{n\to\infty} P(P_n(M_n^* > m) > \epsilon) \to 0$, $\forall\epsilon > 0$.

Define the moment function $\pi : \mathcal{X} \times \Theta \to \mathbb{R}^m$. To simplify exposition we first consider a fixed weighting matrix W. Later in Section 4, we will consider estimated weighting matrices. The GMM estimator using a fixed positive definite weighting matrix W and sample moments $\hat{\pi}_n(\theta) \equiv P_n\pi(\cdot, \theta)$ is given by

$$\hat{\theta}_n \equiv \underset{\theta\in\Theta\subset\mathbb{R}^d}{\arg\min}\ \hat{Q}_n(\theta), \qquad \hat{Q}_n(\theta) \equiv \frac{1}{2}\hat{\pi}_n(\theta)' W \hat{\pi}_n(\theta)$$

We assume the population GMM objective has a unique minimizer $\theta^{\#} \equiv \underset{\theta\in\Theta}{\arg\min}\ Q(\theta)$ where $Q(\theta) \equiv \frac{1}{2}\pi(\theta)' W\pi(\theta)$ and $\pi(\theta) \equiv P\pi(\cdot, \theta)$. It is well known from standard results in Newey and McFadden (1994) that for correctly specified models where $\pi(\theta^{\#}) = 0$, $\sqrt{n}\left(\hat{\theta}_n - \theta^{\#}\right) \rightsquigarrow -(G'WG)^{-1} G'WN\left(0, P\pi(\cdot, \theta^{\#})\pi(\cdot, \theta^{\#})'\right)$, where $G = \frac{\partial}{\partial\theta}\pi(\theta^{\#})$.

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 $\pi(\cdot, \theta)$ is twice continuously differentiable with respect to θ and $\pi(\theta^{\#}) = c$

for a vector of fixed constants $c \neq 0$, Hall and Inoue (2003) showed that $\sqrt{n}\left(\hat{\theta}_n - \theta^{\#}\right) \rightsquigarrow N\left(0, \bar{H}^{-1}\Omega\bar{H}^{-1\prime}\right)$ where

$$\begin{aligned}
\Sigma_{11} &= P\left(\pi\left(\cdot,\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)\left(\pi\left(\cdot,\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)' \\
\Sigma_{12} &= P\left(\pi\left(\cdot,\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)\pi\left(\theta^{\#}\right)' W\left(\frac{\partial}{\partial\theta}\pi\left(\cdot,\theta^{\#}\right) - G\right) \\
\Sigma_{21} &= \Sigma_{12}' \\
\Sigma_{22} &= P\left(\frac{\partial}{\partial\theta}\pi\left(\cdot,\theta^{\#}\right) - G\right)' W\pi\left(\theta^{\#}\right)\pi\left(\theta^{\#}\right)' W\left(\frac{\partial}{\partial\theta}\pi\left(\cdot,\theta^{\#}\right) - G\right) \\
\Omega &= G'W\Sigma_{11}WG + \Sigma_{22} + G'W\Sigma_{12} + \Sigma_{21}WG \\
\bar{H} &= G'WG + \sum_{j=1}^{m}\sum_{k=1}^{m} W_{jk}\pi_k\left(\theta^{\#}\right)H_j
\end{aligned} \tag{2.1}$$

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

Although misspecification changes the asymptotic distribution of smooth estimators, the estimator remains $\sqrt{n}$-consistent, and the nonparametric bootstrap can be used for inference. However, misspecification is a much more serious issue in the nonsmooth case because the rate of convergence becomes cubic-root and the asymptotic distribution becomes non-standard, which invalidates the standard bootstrap. For GMM estimators that are globally misspecified and nonsmooth, specifically non-directionally differentiable (in the sense of Gateaux), we will show that

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

For $g(\cdot,\theta) = \pi(\cdot,\theta) - \pi\left(\cdot,\theta^{\#}\right)$, $\mathcal{Z}_0(h)$ is a mean zero Gaussian process in the space of locally bounded functions $\mathbf{B}_{\text{loc}}\left(\mathbb{R}^d\right)$ equipped with the topology of uniform convergence on compacta. The covariance kernel of $\mathcal{Z}_0(h)$ is

$$\Sigma_{1/2}(s,t) = \lim_{\alpha\to\infty}\alpha Pg\left(\cdot,\theta^{\#} + \frac{s}{\alpha}\right)g\left(\cdot,\theta^{\#} + \frac{t}{\alpha}\right)'$$

We next 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

$$\begin{aligned}
\hat{\theta}_n^* = \arg\min_{\theta\in\Theta}\Big\{&\hat{\pi}_n\left(\hat{\theta}_n\right)' W\left(P_n^* - P_n\right)\left(\pi(\cdot,\theta) - \pi\left(\cdot,\hat{\theta}_n\right)\right) \\
&+ \frac{1}{2}\left(\theta - \hat{\theta}_n\right)'\left(\hat{G}'W\hat{G} + \sum_{j=1}^{m}\sum_{k=1}^{m} W_{jk}\hat{\pi}_k\left(\hat{\theta}_n\right)\hat{H}_j\right)\left(\theta - \hat{\theta}_n\right) \\
&+ \left(\theta - \hat{\theta}_n\right)'\hat{G}'W\left(P_n^* - P_n\right)\pi\left(\cdot,\hat{\theta}_n\right)\Big\}
\end{aligned} \tag{2.2}$$

where $\hat{G}$ is a consistent estimate of G and $\hat{H}_j$ is a consistent estimate of H_j for $j = 1 \dots m$.

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 γ in order to form asymptotically valid confidence intervals for $\theta^{\#}$ using the empirical distribution of $\hat{\theta}_n^* - \hat{\theta}_n$. The intuition for why our rate-adaptive bootstrap procedure is consistent is similar to the arguments given in Cattaneo et al. (2020). Instead of bootstrapping the GMM objective function, we are bootstrapping consistent estimates of the different components that can appear in the asymptotic distribution, depending on whether the model is correctly or incorrectly specified, smooth or nonsmooth. For the case of nonsmooth moments, the first term in (2.2) is used to approximate the Gaussian process $\pi\left(\theta^{\#}\right)' W \mathcal{Z}_0(h)$, while the second term is used to approximate the quadratic mean $\frac{1}{2}h'\bar{H}h$. The third term will disappear asymptotically for nonsmooth models but remain for sufficiently smooth models. We can use the same estimator for both smooth and nonsmooth models because their different rates of convergence will automatically cause the appropriate terms to disappear from the asymptotic distribution.

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 $P_n^* \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbb{W}_{ni}\delta_{X_i}$ for the multinomial vector $\mathbb{W}_n = (\mathbb{W}_{n1}, \dots, \mathbb{W}_{nn})$ with probabilities $(1/n, \dots, 1/n)$ and number of trials n.

1. Compute $\hat{\theta}_n$, $\hat{\pi}_n\left(\hat{\theta}_n\right) = \frac{1}{n}\sum_{i=1}^{n}\pi\left(X_i, \hat{\theta}_n\right)$, $\hat{G}$, $\hat{H}_j$ for $j = 1 \dots m$.

2. Repeat for B bootstrap iterations: draw a bootstrap sample $X_1^*, \dots, X_n^*$ and compute

$$\hat{\theta}_n^* = \underset{\theta\in\Theta}{\arg\min}\left\{\hat{\pi}_n\left(\hat{\theta}_n\right)' W\left(\frac{1}{n}\sum_{i=1}^{n}\left(\pi\left(X_i^*, \theta\right) - \pi\left(X_i^*, \hat{\theta}_n\right)\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\pi\left(X_i, \theta\right) - \pi\left(X_i, \hat{\theta}_n\right)\right)\right)\right.$$
$$+\frac{1}{2}\left(\theta - \hat{\theta}_n\right)'\left(\hat{G}'W\hat{G} + \sum_{j=1}^{m}\sum_{k=1}^{m} W_{jk}\hat{\pi}_k\left(\hat{\theta}\right)\hat{H}_j\right)\left(\theta - \hat{\theta}_n\right)$$
$$\left. + \left(\theta - \hat{\theta}_n\right)'\hat{G}'W\left(\frac{1}{n}\sum_{i=1}^{n}\left(\pi\left(X_i^*, \hat{\theta}_n\right) - \pi\left(X_i, \hat{\theta}_n\right)\right)\right)\right\}$$

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

A $1-\alpha$ two-sided equal-tailed confidence interval for $\theta_k^{\#}$ can be formed by

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

We will use the following notation to denote the stacked confidence intervals for the vector of the parameters:

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

2.1 Asymptotic Distribution for Nonsmooth misspecified GMM using a fixed weighting matrix

Throughout the paper, we will impose the following assumptions. The different values of γ and ρ depend on the rate of convergence of $\hat{\theta}_n$.

Assumption 1. *For $\hat{Q}_n(\theta) \equiv \frac{1}{2} P_n \pi(\cdot,\theta)' W P_n \pi(\cdot,\theta)$ and $Q(\theta) \equiv \frac{1}{2} P\pi(\cdot,\theta)' W P\pi(\cdot,\theta)$, suppose the following conditions are satisfied for $\rho \in \left\{\frac{1}{2}, 1\right\}$ and $\gamma = \frac{1}{2(2-\rho)}$:*

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

(ii) $\inf_{\theta\in\Theta:\|\theta-\theta^{\#}\|>\epsilon} Q(\theta) > Q\left(\theta^{\#}\right)$ *for all* $\epsilon > 0$.

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

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

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

(i) $\theta^{\#}$ *is an interior point of* Θ.

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

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

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

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

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

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

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

Assumption 1 is needed to show consistency of $\hat{\theta}_n$ for $\theta^{\#}$ while Assumption 2 is needed to derive its asymptotic distribution. "Uniformly manageable" classes of functions are discussed after Corollary 3.2 of Kim and Pollard (1990) and are similar to universal Donsker classes, as discussed after Definition 4.1 of Pollard (1989).

Similar to Kim and Pollard (1990), 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 $\pi(\cdot,\theta)$

is not Hadamard directionally differentiable, such as the ones that appear in the GMM formulation of IV quantile regression or simulated method of moments. Other types of nonsmooth moments that are directionally differentiable do not have this linear rate of decay and therefore retain the $\sqrt{n}$ rate of convergence. We now provide some examples that distinguish between different types of nonsmooth moments.

Example 1. GMM Formulation of Instrumental Variable Quantile Regression (IVQR) This example studies how to do inference in the case of possible misspecification of moments in Chernozhukov and Hansen (2005))'s IVQR GMM estimator. The IVQR estimator can be used to estimate quantile treatment effects under non-compliance, and under correct specification, the estimator is known to be $\sqrt{n}$-consistent and asymptotically normal. However, if the moments are (globally) misspecified, which can happen for example if the instruments are invalid, then the estimator is cubic-root consistent and has a non-standard asymptotic distribution.

The moment conditions for IVQR are non-smooth, in particular non-directionally differentiable, because $\pi(\cdot,\theta) = (\tau - 1(y_i \le q(d_i, w_i, \theta)))\, z_i$, 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, which has a single index structure $q(d_i, w_i, \theta) = q(x_i'\theta)$ for $x_i' = [d_i, w_i]$. Additionally, $q(\cdot)$ is assumed to be a monotonic, twice differentiable function. For $\pi(\theta) = E\left(\tau - F_{y|x,z}\left(q\left(x'\theta^{\#}\right)\right)\right) z$, the Jacobian is $G = \frac{\partial}{\partial\theta}\pi\left(\theta^{\#}\right) = -Ef_{y|x,z}\left(q\left(x'\theta^{\#}\right)\right) zq'\left(x'\theta^{\#}\right)x'$ and the jth element of the Hessian is $H_j = \frac{\partial^2}{\partial\theta\partial\theta'}\pi_j\left(\theta^{\#}\right) = -Ef'_{y|x,z}\left(q\left(x'\theta^{\#}\right)\right) z_j\left(q'\left(x'\theta^{\#}\right)\right)^2 xx' + Ef_{y|x,z}\left(q\left(x'\theta^{\#}\right)\right) z_j q''\left(x'\theta^{\#}\right) xx'$. We will assume that $F_{y|x,z}$ is absolutely continuous and the assumptions in Chernozhukov and Hansen (2005) are satisfied so that G and H_j are well-defined.

A crucial condition that generates cubic-root convergence in globally misspecified models with non-directionally differentiable moments is when the value of ρ that satisfies Assumption 2(iii)-(viii) is $\rho = 1/2$. In the Appendix, we show this is true for this example.

Example 2. Simulated Method of Moments Simulated method of moments has a wide range of applications especially in discrete choice models where an agent's choice probabilities are too complicated to calculate analytically (McFadden (1989) and Pakes and Pollard (1989)). Instead, we take simulation draws from some assumed distribution for the errors and using the empirical frequency simulator to estimate the choice probabilities. In this example, we consider a binary discrete choice model but the results are easily generalizable to multivariate discrete choice models.

The moment conditions are $\pi(\cdot,\theta) = \left(y_i - \frac{1}{S}\sum_{s=1}^{S} 1\left(h\left(x_i'\theta\right) + \eta_{is} > 0\right)\right) z_i$, 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, $h(\cdot)$ is a monotonic, twice differentiable function, and $\{\eta_{is}\}_{s=1}^{S}$ are individual i's simulation draws from an absolutely continuous distribution $F_{\eta|x,z}$ with density function $f_{\eta|x,z}$ symmetric around zero. For $\pi(\theta) = E\left(y - F_{\eta|x,z}\left(h\left(x'\theta\right)\right)\right) z$, the Jacobian is $G = \frac{\partial}{\partial\theta}\pi\left(\theta^{\#}\right) = -Ef_{\eta|x,z}\left(h\left(x'\theta^{\#}\right)\right) zh'\left(x'\theta^{\#}\right)x'$ and the jth element of the Hessian is $H_j = \frac{\partial^2}{\partial\theta\partial\theta'}\pi_j\left(\theta^{\#}\right) = -Ef'_{\eta|x,z}\left(h\left(x'\theta^{\#}\right)\right) z_j\left(h'\left(x'\theta^{\#}\right)\right)^2 xx' + Ef_{\eta|x,z}\left(h\left(x'\theta^{\#}\right)\right) z_j h''\left(x'\theta^{\#}\right) xx'$. Assume that McFadden (1989)'s assumptions are satisfied so that G and H_j are well defined.

We verify in the Appendix that the value of ρ that satisfies Assumption 2(iii)-(viii) is

$\rho = 1/2$.

Example 3. Dynamic Censored Regression Honore and Hu (2004a) consider estimation of a panel data censored regression model with lagged dependent variables: $y_{it} = y_{it-1}\theta + \alpha_i + \epsilon_{it}$ where $\{\epsilon_{it}\}_{t=1}^{T}$ is a sequence of i.i.d. random variables conditional on (y_{i0}, α_i). They show that GMM using $\pi(\cdot,\theta) = \max\{0, y_{it} - y_{it-1}\theta\} - y_{it-1}$ as the moment conditions will be $\sqrt{n}$ consistent and asymptotically normal and that the true parameter uniquely satisfies the population moments under correct specification. Define the stacked moments as $\pi(\theta) = [\pi_2(\theta), \ldots, \pi_T(\theta)]'$ where $\pi_t(\theta) = E[\max\{0, y_{it} - y_{it-1}\theta\} - y_{it-1}]$ $= E[1(y_{it} > y_{it-1}\theta)(y_{it} - y_{it-1}\theta) - y_{it-1}]$. The Jacobian is $G = [G_2, \ldots, G_T]'$ for $G_t = -E\left[y_{it-1}1\left(y_{it} > y_{it-1}\theta^{\#}\right)\right]$, and the Hessians are $H_t = E\left[y_{it-1}f_{y_{it}|y_{it-1}}\left(y_{it-1}\theta^{\#}\right)\right]$ for $t = 2, \ldots, T$.

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

Theorem 1. *Suppose $\pi\left(\theta^{\#}\right) = c$ 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*

$$n^{1/3}\left(\hat{\theta}_n - \theta^{\#}\right) \rightsquigarrow \arg\min_{h\in\mathbb{R}^d}\left\{\pi\left(\theta^{\#}\right)' W \mathcal{Z}_{0,1/2}(h) + \frac{1}{2}h'\bar{H}h\right\}$$
$$\bar{H} = G'WG + \sum_{j=1}^{m}\sum_{k=1}^{m} W_{jk}\pi_k\left(\theta^{\#}\right)H_j$$

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

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

In the Appendix, we show that the globally misspecified GMM estimator is $\sqrt{n}$-consistent when the moments are nonsmooth but remain directionally differentiable. In the correctly specified case, a reduction to the standard result of Newey and McFadden (1994) previously mentioned is achieved.

3 Rate-Adaptive Bootstrap for fixed weighting matrix

We impose the following envelope integrability assumptions 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. There are some differences between our assumption and the ones in Cattaneo et al. (2020) because Cattaneo et al. (2020) show bootstrap equicontinuity using the maximal inequalities in Pollard (1989) whereas we make use of Lemma 4.2 in Wellner and Zhan (1996), which states that stochastic equicontinuity implies bootstrap equicontinuity under an envelope integrability assumption.

Assumption 3. *For $\rho \in \left\{\frac{1}{2}, 1\right\}$ and $\gamma = \frac{1}{2(2-\rho)}$, define $m_n(\cdot, \theta, h) \equiv n^{\gamma\rho}\left(\pi\left(\cdot; \theta + \frac{h}{n^\gamma}\right) - \pi(\cdot; \theta)\right)$.*
(i) For any $\epsilon_n \to 0$,

$$\lim_{\lambda\to\infty} \lim_{n\to\infty} \sup \sup_{t\geq\lambda} t^2 \; P\left\{\sup_{h\in\mathbb{R}^d, \|\theta-\theta^\#\|\leq\epsilon_n} \left\| \frac{m_n(\cdot,\theta,h) - m_n(\cdot,\theta^\#,h)}{1+n^\gamma\|\theta-\theta^\#\|}\right\| > t\right\} = 0$$

(ii) Furthermore, if Assumptions 1- 2 are satisfied for $\gamma = 1/2$ and $\rho = 1$, then for any $\epsilon_n \to 0$,

$$\lim_{\lambda\to\infty} \lim_{n\to\infty} \sup \sup_{t\geq\lambda} t^2 \; P\left\{\sup_{\|\theta-\theta^\#\|\leq\epsilon_n} \left\| \frac{\pi(\cdot,\theta) - \pi(\cdot,\theta^\#)}{1+\sqrt{n}\|\theta-\theta^\#\|}\right\| > t\right\} = 0$$

Sufficient conditions for (i) and (ii) are that the envelopes are uniformly bounded. For all sufficiently large n such that $\epsilon_n \to 0$, there exists some constants $C_1 > 0$ and $C_2 > 0$ such that $\sup_{h\in\mathbb{R}^d, \|\theta-\theta^\#\|\leq\epsilon_n} \left\| \frac{m_n(\cdot,\theta,h)-m_n(\cdot,\theta^\#,h)}{1+n^\gamma\|\theta-\theta^\#\|}\right\| \leq C_1$, and $\sup_{\|\theta-\theta^\#\|\leq\epsilon_n} \left\|\frac{\pi(\cdot,\theta)-\pi(\cdot,\theta^\#)}{1+\sqrt{n}\|\theta-\theta^\#\|}\right\| \leq C_2$.

The next theorem illustrates consistency of the rate-adaptive bootstrap for correctly specified and globally misspecified models which can be either smooth or nonsmooth. Under correct specification, the asymptotic distribution is normal for smooth and nonsmooth moments. Under global misspecification, the asymptotic distribution is normal in the smooth case but in the nonsmooth case, it is nonstandard.

Theorem 2. *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,

$$\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} -\left(G'WG\right)^{-1} G'WN\left(0, P\pi\left(\cdot, \theta^\#\right)\pi\left(\cdot, \theta^\#\right)'\right)$$

For globally misspecified models with twice continuously differentiable $\pi(\cdot, \theta)$, if Assumptions 1- 2 are satisfied for $\gamma = 1/2, \rho = 1$,

$$\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} N\left(0, \bar{H}^{-1}\Omega\bar{H}^{-1'}\right)$$

where Ω and $\bar{H}$ are defined in equation (2.1). If instead Assumptions 1- 2 are satisfied for $\gamma = 1/3, \rho = 1/2$,

$$n^{1/3}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \arg\min_h \left\{\pi\left(\theta^\#\right)' W \mathcal{Z}_{0,1/2}(h) + \frac{1}{2}h'\bar{H}h\right\}$$

4 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 presumed to be unique pseudo-true parameter to be $\theta^{\#} = \arg\min_{\theta\in\Theta} \pi(\theta)' W\left(\theta_1^{\#}\right)\pi(\theta)$ where $W\left(\theta_1^{\#}\right)$ depends on the presumed to be unique 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 $\hat{\theta}_1 = \arg\min_{\theta\in\Theta} \hat{\pi}(\theta)' W_1 \hat{\pi}(\theta)$. The next theorem demonstrates that the globally misspecified 2-step GMM estimator $\hat{\theta}_n = \arg\min_{\theta\in\Theta} \hat{\pi}(\theta)' W_n\left(\hat{\theta}_1\right)\hat{\pi}(\theta)$ 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 $W\left(\theta_1^{\#}\right)$. 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 3. *Suppose $\pi\left(\theta^{\#}\right) = c$ 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 $W_n - W = o_p\left(n^{-1/3}\right)$, then $\hat{\theta}_n - \theta^{\#} = o_P(1)$ and for $\bar{\mathcal{Z}}_0(h) \equiv \pi\left(\theta^{\#}\right)' W \mathcal{Z}_{0,1/2}(h)$,

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

If $W_n - W = O_p\left(n^{-1/3}\right)$ and $\begin{pmatrix} \pi\left(\theta^{\#}\right)' W n^{2/3}(P_n - P)\, g\left(\cdot,\theta^{\#} + n^{-1/3}h\right) \\ h'G' n^{1/3}(W_n - W)\,\pi\left(\theta^{\#}\right)\end{pmatrix} \rightsquigarrow \begin{pmatrix} \bar{\mathcal{Z}}_0(h) \\ h'G'\mathcal{W}_0\end{pmatrix}$ *in the product space of locally bounded functions $\left\{\mathbf{B}_{loc}\left(\mathbb{R}^d\right)\right\}^2$ for some tight random vector $\mathcal{W}_0$, then $\hat{\theta}_n - \theta^{\#} = o_P(1)$ and*

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

In the case where the moments are smooth, our estimated weighting matrix W_n typically satisfies the following assumption which states that W_n is $\sqrt{n}$-consistent with an influence function representation, and that the bootstrapped weighting matrix W_n^* shares the same influence function representation.

Assumption 4. *The weighting matrix W_n satisfies $\sqrt{n}(W_n - W) = \sqrt{n}(P_n - P)\,\phi\left(\cdot,\theta_1^{\#}\right) + o_p(1)$ where $\theta_1^{\#}$ is the probability limit of the 1-step GMM estimate using a fixed weighting matrix, $P\left\|vech\left(\phi\left(\cdot,\theta_1^{\#}\right)\right)\right\|^2 < \infty$, and the bootstrapped weighting matrix W_n^* has the same representation $\sqrt{n}(W_n^* - W_n) = \sqrt{n}(P_n^* - P_n)\,\phi\left(\cdot,\theta_1^{\#}\right) + o_p^*(1)$. Additionally, for each $\epsilon > 0$ and $t \in \mathbb{R}^d$,*

$$\lim_{n\to\infty} P\left\|\begin{pmatrix}\sqrt{n}g\left(\cdot,\theta^{\#} + \frac{t}{\sqrt{n}}\right) \\ \pi\left(\cdot,\theta^{\#}\right) \\ vech\left(\phi\left(\cdot,\theta_1^{\#}\right)\right)\end{pmatrix}\right\|^2 1\left\{\left\|\begin{pmatrix}\sqrt{n}g\left(\cdot,\theta^{\#} + \frac{t}{\sqrt{n}}\right) \\ \pi\left(\cdot,\theta^{\#}\right) \\ vech\left(\phi\left(\cdot,\theta_1^{\#}\right)\right)\end{pmatrix}\right\| > \epsilon\sqrt{n}\right\} = 0$$

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:

$$\begin{aligned}\hat{\theta}_n^* = \arg\min_{\theta\in\Theta} \Big\{ & \hat{\pi}_n\left(\hat{\theta}_n\right)' W_n\left(P_n^* - P_n\right)\left(\pi\left(\cdot,\theta\right) - \pi\left(\cdot,\hat{\theta}_n\right)\right) \\ & + \frac{1}{2}\left(\theta - \hat{\theta}_n\right)'\left(\hat{G}'W_n\hat{G} + \sum_{j=1}^{m}\sum_{k=1}^{m} W_{n,jk}\hat{\pi}_k\left(\hat{\theta}_n\right)\hat{H}_j\right)\left(\theta - \hat{\theta}_n\right) \\ & + \left(\theta - \hat{\theta}_n\right)'\hat{G}'W_n\left(P_n^* - P_n\right)\pi\left(\cdot,\hat{\theta}_n\right) \\ & + \left(\theta - \hat{\theta}_n\right)'\hat{G}'\left(W_n^* - W_n\right)\hat{\pi}_n\left(\hat{\theta}_n\right)\Big\} \end{aligned} \tag{4.1}$$

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

$$\begin{aligned}\hat{\theta}_1^* = \arg\min_{\theta\in\Theta} \Big\{ & \hat{\pi}_n\left(\hat{\theta}_n\right)' W_1\left(P_n^* - P_n\right)\left(\pi\left(\cdot,\theta\right) - \pi\left(\cdot,\hat{\theta}_n\right)\right) \\ & + \frac{1}{2}\left(\theta - \hat{\theta}_n\right)'\left(\hat{G}'W_1\hat{G} + \sum_{j=1}^{m}\sum_{k=1}^{m} W_{1,jk}\hat{\pi}_k\left(\hat{\theta}_n\right)\hat{H}_j\right)\left(\theta - \hat{\theta}_n\right) \\ & + \left(\theta - \hat{\theta}_n\right)'\hat{G}'W_1\left(P_n^* - P_n\right)\pi\left(\cdot,\hat{\theta}_n\right)\Big\} \end{aligned} \tag{4.2}$$

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 4. *Suppose Assumptions 1-3 are satisfied, $\hat{G} \xrightarrow{p} G$, and $\hat{H}_j \xrightarrow{p} H_j$ for $j = 1\ldots m$.*

(i) For correctly specified models, when $W_n - W = o_p(1)$ and $W_n^* - W_n = o_p^*(1)$,

$$\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} -\left(G'WG\right)^{-1} G'WN\left(0, P\pi\left(\cdot,\theta^{\#}\right)\pi\left(\cdot,\theta^{\#}\right)'\right)$$

(ii) For globally misspecified models with twice continuously differentiable $\pi(\cdot,\theta)$ where Assumptions 1 and 2 are satisfied for $\gamma = 1/2$ and $\rho = 1$, and the weighting matrix W_n satisfies Assumption 4,

$$\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} N\left(0, \bar{H}^{-1}\Omega_W\bar{H}^{-1\prime}\right)$$

$$\begin{aligned}\Omega_W = & G'W\Sigma_{11}WG + \Sigma_{22} + G'W\Sigma_{12} + \Sigma_{21}WG \\ & + G'\Sigma_{33}G + G'W\Sigma_{13}G + G'\Sigma_{31}WG + \Sigma_{23}G + G'\Sigma_{32}\end{aligned}$$

where $\bar{H}$, Σ_{11}, Σ_{12}, Σ_{21}, and Σ_{22} are the same as in equation (2.1) and

$$\begin{aligned}
\Sigma_{13} &= P\left(\pi\left(\cdot,\theta^{\#}\right)-\pi\left(\theta^{\#}\right)\right)\pi\left(\theta^{\#}\right)'\left(\phi\left(\cdot,\theta_1^{\#}\right)-\phi\left(\theta_1^{\#}\right)\right)' \\
\Sigma_{31} &= \Sigma_{13}' \\
\Sigma_{23} &= P\left(\frac{\partial}{\partial\theta}\pi\left(\cdot,\theta^{\#}\right)-G\right)' W\pi\left(\theta^{\#}\right)\pi\left(\theta^{\#}\right)'\left(\phi\left(\cdot,\theta_1^{\#}\right)-\phi\left(\theta_1^{\#}\right)\right)' \\
\Sigma_{32} &= \Sigma_{23}' \\
\Sigma_{33} &= P\left(\phi\left(\cdot,\theta_1^{\#}\right)-\phi\left(\theta_1^{\#}\right)\right)\pi\left(\theta^{\#}\right)\pi\left(\theta^{\#}\right)'\left(\phi\left(\cdot,\theta_1^{\#}\right)-\phi\left(\theta_1^{\#}\right)\right)'
\end{aligned}$$

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

$$n^{1/3}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \arg\min_h \left\{\pi\left(\theta^{\#}\right)' W \mathcal{Z}_{0,1/2}(h) + \frac{1}{2}h'\bar{H}h\right\}$$

If $W_n - W = O_p\left(n^{-1/3}\right)$, $W_n^* - W_n = O_p^*\left(n^{-1/3}\right)$, and
$\begin{pmatrix} \pi\left(\theta^{\#}\right)' W n^{2/3}\left(P_n^* - P_n\right) g\left(\cdot,\theta^{\#} + n^{-1/3}h\right) \\ h'G'n^{1/3}\left(W_n^* - W_n\right)\pi\left(\theta^{\#}\right) \end{pmatrix} \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \begin{pmatrix} \pi\left(\theta^{\#}\right)' W \mathcal{Z}_{0,1/2}(h) \\ h'G'\mathcal{W}_0 \end{pmatrix}$ *in* $\left\{\mathbf{B}_{loc}\left(\mathbb{R}^d\right)\right\}^2$,
then

$$n^{1/3}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \arg\min_h \left\{\pi\left(\theta^{\#}\right)' W \mathcal{Z}_{0,1/2}(h) + h'G'\mathcal{W}_0 + \frac{1}{2}h'\bar{H}h\right\}$$

5 Monte Carlo

5.1 Nonsmooth Location Model

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 $\pi(\cdot,\theta) = \left[1\left(y_i \le \theta\right) - \tau; y_i - \theta\right]'$, let the population moments be

$$\pi(\theta) = \left[P\left(y_i \le \theta\right) - \tau; Ey_i - \theta\right]'$$

The model can not be correctly specified as long as $\tau \neq 0.5$. First consider using a fixed weighting matrix $W = I$, and consider the following GMM criterion function and its probability limit:

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

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

$$f_y\left(\theta^{\#}\right)\left(F_y\left(\theta^{\#}\right) - \tau\right) + \theta^{\#} = 0.$$

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

$$\begin{aligned}
\hat{\theta}_n^* = \underset{\theta \in \Theta}{\arg\min} \Bigg\{ & \hat{\pi}_n\left(\hat{\theta}_n\right)' W \left(\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n^*\left(\hat{\theta}_n\right)\right) - \left(\hat{\pi}(\theta) - \hat{\pi}\left(\hat{\theta}_n\right)\right)\right) \\
& + \frac{1}{2}\left(\theta - \hat{\theta}_n\right)' \left(\hat{G}' W \hat{G} + \sum_{j=1}^{m} \sum_{k=1}^{m} W_{jk} \hat{\pi}_k\left(\hat{\theta}_n\right) \hat{H}_j\right)\left(\theta - \hat{\theta}_n\right) \\
& + \left(\theta - \hat{\theta}_n\right)' \hat{G}' W \left(\hat{\pi}_n^*\left(\hat{\theta}_n\right) - \hat{\pi}\left(\hat{\theta}_n\right)\right) \Bigg\}
\end{aligned}$$

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

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

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 $n^{-1/10}$.

The first 3 columns of Table 1 show the rate-adaptive bootstrap empirical coverage frequencies for $\theta^{\#}$ (along with the average widths of the confidence intervals in parentheses) for $\tau \in \{0.1, 0.3, 0.5\}$, $n \in \{200, 800, 1600, 3200, 6400\}$, $B = 1000$ bootstrap iterations, and $R = 1000$ Monte Carlo simulations. Due to the symmetry of the problem, similar results will hold for $\tau \in \{0.7, 0.9\}$ and are available upon request. The remaining columns show 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 $\tilde{\theta}_1 - \hat{\theta}_n$ for $\tilde{\theta}_1 = \arg\min_{\theta \in \Theta}\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)'\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)$. We can see that the standard bootstrap performs fine under correct specification (which corresponds to $\tau = 0.5$), but the performance deteriorates as τ moves away from 0.5, with more severe undercoverage for the smaller values of τ. In contrast, the empirical coverage frequency of the rate-adaptive bootstrap is quite close to the nominal level of 95% for all values of τ, even at the smaller sample sizes.

Table 1: Rate-adaptive Bootstrap Empirical Coverage Frequencies

	Rate-adaptive Bootstrap			Standard Bootstrap		
τ	0.1	0.3	0.5	0.1	0.3	0.5
$n = 200$	0.945	0.950	0.951	0.913	0.922	0.942
	(0.331)	(0.299)	(0.279)	(0.277)	(0.277)	(0.277)
$n = 800$	0.944	0.947	0.948	0.881	0.917	0.953
	(0.181)	(0.156)	(0.140)	(0.139)	(0.139)	(0.139)
$n = 1600$	0.954	0.959	0.950	0.864	0.892	0.939
	(0.135)	(0.114)	(0.099)	(0.098)	(0.098)	(0.098)
$n = 3200$	0.956	0.944	0.955	0.832	0.894	0.940
	(0.101)	(0.083)	(0.070)	(0.070)	(0.070)	(0.070)
$n = 6400$	0.949	0.947	0.944	0.823	0.894	0.954
	(0.076)	(0.061)	(0.049)	(0.049)	(0.049)	(0.049)

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

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

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

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

where $\hat{\theta}_1 = \arg\min_{\theta\in\Theta} \hat{\pi}_n\left(\theta\right)' \hat{\pi}_n\left(\theta\right)$, $\hat{f}_y\left(\hat{\theta}_1\right) = \frac{1}{nh}\sum_{i=1}^n K_h\left(y_i-\hat{\theta}_1\right)$, $\hat{F}_y\left(\hat{\theta}_1\right)=\frac{1}{n}\sum_{i=1}^n 1\left(y_i\le\hat{\theta}_1\right)$. For $\hat{\theta}_n = \arg\min_{\theta\in\Theta}\hat{\pi}_n\left(\theta\right)' W_n\left(\hat{\theta}_1\right)\hat{\pi}_n\left(\theta\right)$, the rate-adaptive bootstrap estimate is

$$\begin{aligned}\hat{\theta}_n^* = \underset{\theta\in\Theta}{\arg\min} \Big\{ & \hat{\pi}_n\left(\hat{\theta}_n\right)' W_n\left(\hat{\theta}_1\right)\left(\left(\hat{\pi}_n^*\left(\theta\right)-\hat{\pi}_n^*\left(\hat{\theta}_n\right)\right)-\left(\hat{\pi}\left(\theta\right)-\hat{\pi}\left(\hat{\theta}_n\right)\right)\right) \\ & +\frac{1}{2}\left(\theta-\hat{\theta}_n\right)'\left(\hat{G}'W_n\left(\hat{\theta}_1\right)\hat{G}+\sum_{j=1}^m\sum_{k=1}^m W_{n,jk}\left(\hat{\theta}_1\right)\hat{\pi}_k\left(\hat{\theta}_n\right)\hat{H}_j\right)\left(\theta-\hat{\theta}_n\right) \\ & +\left(\theta-\hat{\theta}_n\right)'\hat{G}'W_n\left(\hat{\theta}_1\right)\left(\hat{\pi}_n^*\left(\hat{\theta}_n\right)-\hat{\pi}\left(\hat{\theta}_n\right)\right) \\ & +\left(\theta-\hat{\theta}_n\right)'\hat{G}'\left(W_n^*\left(\hat{\theta}_1^*\right)-W_n\left(\hat{\theta}_1\right)\right)\hat{\pi}_n\left(\hat{\theta}_n\right)\Big\}\end{aligned}$$

The bootstrapped weighting matrix is

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

where $\hat{\theta}_1^*$ is the rate-adaptive bootstrap estimate using $W = I$, $\hat{f}_y^*\left(\hat{\theta}_1^*\right) = \frac{1}{nh}\sum_{i=1}^n K_h\left(y_i^* - \hat{\theta}_1^*\right)$, and $\hat{F}_y^*\left(\hat{\theta}_1^*\right) = \frac{1}{n}\sum_{i=1}^n 1\left(y_i^* \leq \hat{\theta}_1^*\right)$. We use the same Silverman's Rule of Thumb bandwidth as before $h = 1.06\text{std}(y)n^{-1/5}$.

We are interested in the rate-adaptive bootstrap empirical coverage frequencies for $\theta^\# = \underset{\theta\in\Theta}{\arg\min}\pi(\theta)' W\left(\theta_1^\#\right)\pi(\theta)$ where $W\left(\theta_1^\#\right) = \begin{bmatrix} F_y\left(\theta_1^\#\right) - F_y\left(\theta_1^\#\right)^2 & -f_y\left(\theta_1^\#\right) \\ -f_y\left(\theta_1^\#\right) & 1 \end{bmatrix}^{-1}$ and $\theta_1^\# = \underset{\theta\in\Theta}{\arg\min}\pi(\theta)'\pi(\theta)$. The first 5 columns of Table 2 show 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 undercoverage for the case of $\tau = 0.5$ and over-coverage for the other values of τ, but the performance is much better than the standard bootstrap intervals shown in the remaining columns: $\left[\hat{\theta}_n - d_{0.975}, \hat{\theta}_n - d_{0.025}\right]$, where $d_{0.975}$ and $d_{0.025}$ are the 97.5th and 2.5th percentiles of $\tilde{\theta}_2 - \hat{\theta}_n$, for $\tilde{\theta}_2 = \arg\min_{\theta\in\Theta}\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)' W_n^*\left(\tilde{\theta}_1\right)\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)$ and $\tilde{\theta}_1 = \arg\min_{\theta\in\Theta}\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)'\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)$. The standard bootstrap has undercoverage across all values of τ, except for the correctly specified case of $\tau = 0.5$. This is to be expected because the standard bootstrap is inconsistent under misspecification for nonsmooth models. However, the rate-adaptive bootstrap will be consistent.

Table 2: Rate-adaptive Bootstrap Empirical Coverage Frequencies

	Rate-adaptive Bootstrap					Standard Bootstrap				
τ	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
$n = 200$	0.970	0.963	0.941	0.965	0.965	0.697	0.842	0.951	0.835	0.688
	(1.224)	(0.682)	(0.286)	(0.693)	(1.228)	(0.344)	(0.329)	(0.310)	(0.331)	(0.338)
$n = 800$	0.987	0.957	0.948	0.967	0.988	0.647	0.766	0.957	0.791	0.648
	(0.688)	(0.370)	(0.143)	(0.370)	(0.702)	(0.174)	(0.172)	(0.160)	(0.171)	(0.174)
$n = 1600$	0.984	0.966	0.954	0.967	0.977	0.606	0.757	0.966	0.752	0.613
	(0.520)	(0.293)	(0.101)	(0.293)	(0.523)	(0.125)	(0.122)	(0.114)	(0.123)	(0.125)
$n = 3200$	0.990	0.960	0.934	0.952	0.990	0.555	0.711	0.976	0.685	0.533
	(0.394)	(0.235)	(0.071)	(0.234)	(0.394)	(0.089)	(0.088)	(0.081)	(0.088)	(0.090)
$n = 6400$	0.978	0.947	0.944	0.957	0.982	0.533	0.682	0.963	0.666	0.516
	(0.311)	(0.189)	(0.050)	(0.189)	(0.313)	(0.064)	(0.063)	(0.057)	(0.063)	(0.064)

5.2 Quantile Regression

Motivated by Chernozhukov and Hong (2003) and Chernozhukov and Hansen (2005), we consider the following data generating process for $\alpha_0 = \beta_0 = 1$:

$$y_i = \alpha_0 + \beta_0 d_i + u_i, \quad \begin{pmatrix} u_i \\ d_i \end{pmatrix} \overset{i.i.d.}{\sim} N(\mu, \Omega), \quad \mu = \begin{pmatrix} \delta \\ 1 \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

It follows then that

$$\begin{aligned} u_i| \, d_i &\sim N(\delta, 1) \\ P(u_i \leq 0| \, d_i) &= \Phi(-\delta) \\ y_i| \, d_i &\sim N(\alpha_0 + \beta_0 d_i + \delta, 1) \end{aligned}$$

The population moments are for $z_i = \begin{pmatrix} 1 & d_i \end{pmatrix}'$,

$$\begin{aligned} \pi(\theta) &= E\left[\left(\frac{1}{2} - 1(y_i \leq \alpha + \beta d_i)\right) z_i\right] \\ &= E\left[\left(\frac{1}{2} - F_{y|d}(\alpha + \beta d_i)\right) z_i\right] \\ &= E\left[\left(\frac{1}{2} - \Phi(\alpha - \alpha_0 + (\beta - \beta_0) d_i - \delta)\right) z_i\right] \end{aligned}$$

The sample moments are

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

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 δ, 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

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

The population Jacobian and Hessians are

$$\begin{aligned} G(\theta) &= -E\left[f_{y|d}(\alpha + \beta d_i) z_i z_i'\right] \\ H_1(\theta) &= -E\left[f_{y|d}'(\alpha + \beta d_i) z_i z_i'\right] \\ H_2(\theta) &= -E\left[f_{y|d}'(\alpha + \beta d_i) d_i z_i z_i'\right] \end{aligned}$$

Their estimates are

$$\hat{G} = -\frac{1}{nh}\sum_{i=1}^{n} K_h\left(y_i - \hat{\alpha}_n - \hat{\beta}_n d_i\right) z_i z_i'$$

$$\hat{H}_1 = -\frac{1}{nh^2}\sum_{i=1}^{n} K_h'\left(y_i - \hat{\alpha}_n - \hat{\beta}_n d_i\right) z_i z_i'$$

$$\hat{H}_2 = -\frac{1}{nh^2}\sum_{i=1}^{n} K_h'\left(y_i - \hat{\alpha}_n - \hat{\beta}_n d_i\right) d_i z_i z_i'$$

where $K_h(x) = K(x/h)$, $K_h'(x) = K'(x/h)$, $K(x) = (2\pi)^{-1/2} e^{-x^2/2}$, $K'(x) = -(2\pi)^{-1/2} x e^{-x^2/2}$, and $h = 0.2$. The rate-adaptive bootstrap estimator in the case of a fixed weighting matrix W is

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

Tables 3 and 4 compare the empirical coverage frequencies and average interval lengths of nominal 95% equal-tailed confidence intervals $\left[\hat{\theta}_n - c_{0.975}, \hat{\theta}_n - c_{0.025}\right]$ constructed using the rate-adaptive bootstrap estimator and the centered standard bootstrap estimator $\tilde{\theta}_1 = \arg\min_{\theta}\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)'\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)$. 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 for both parameters. The standard bootstrap has the worst undercoverage for smaller values of n, and the coverage improves for larger n. The improvement could be due to the nature of the data generating process giving moments that become smoother at n increases.

Table 3: Rate-Adaptive vs. Standard Bootstrap, $\delta = 0.1$, fixed weighting matrix $W = I$

	n	200	800	1600	3200	6400
Rate-adaptive	α_0	0.979	0.958	0.953	0.949	0.955
		(0.385)	(0.240)	(0.173)	(0.124)	(0.088)
	β_0	0.974	0.956	0.951	0.958	0.955
		(0.306)	(0.176)	(0.125)	(0.088)	(0.062)
Standard	α_0	0.835	0.915	0.923	0.929	0.938
		(0.271)	(0.231)	(0.171)	(0.123)	(0.087)
	β_0	0.820	0.912	0.925	0.938	0.938
		(0.243)	(0.170)	(0.123)	(0.087)	(0.062)

Table 4: Rate-Adaptive vs. Standard Bootstrap, $\delta = 0.9$, fixed weighting matrix $W = I$

	n	200	800	1600	3200	6400
	α_0	0.959	0.954	0.951	0.951	0.956
Rate-adaptive		(0.477)	(0.251)	(0.176)	(0.125)	(0.088)
	β_0	0.971	0.959	0.951	0.956	0.956
		(0.294)	(0.174)	(0.124)	(0.088)	(0.062)
	α_0	0.864	0.915	0.926	0.932	0.940
Standard		(0.421)	(0.247)	(0.175)	(0.123)	(0.087)
	β_0	0.844	0.921	0.924	0.935	0.939
		(0.230)	(0.171)	(0.123)	(0.087)	(0.062)

Now consider the case of an estimated weighting matrix. Let $W\left(\theta_1^{\#}\right) = \text{plim}\, W_n\left(\hat{\theta}_1\right)$ 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 $\theta^{\#} = \arg\min_{\theta} \pi(\theta)' W\left(\theta_1^{\#}\right) \pi(\theta)$, which can be computed by solving

$$
\begin{aligned}
0 =& \frac{\partial \pi\left(\theta^{\#}\right)'}{\partial \theta} W\left(\theta_1^{\#}\right) \pi\left(\theta^{\#}\right) \\
=& -E\left[f_{y|d}\left(\alpha^* + \beta^* d_i\right) z_i z_i'\right] W\left(\theta_1^{\#}\right) E\left[\left(\frac{1}{2} - F_{y|d}\left(\alpha^* + \beta^* d_i\right)\right) z_i\right] \\
=& -E\left[\left(\frac{1}{2} - \phi\left(\alpha^* - \alpha_0 + \left(\beta^* - \beta_0\right) d_i - \delta\right)\right) z_i\right] W\left(\theta_1^{\#}\right) E\left[\left(\frac{1}{2} - \Phi\left(\alpha^* - \alpha_0 + \left(\beta^* - \beta_0\right) d_i - \delta\right)\right) z_i\right]
\end{aligned}
$$

$W\left(\theta_1^{\#}\right)$ is the inverse of the variance-covariance matrix of the population moments

$$
\begin{aligned}
W\left(\theta_1^{\#}\right) &= \left(E\left[\pi\left(\cdot, \theta_1^{\#}\right) \pi\left(\cdot, \theta_1^{\#}\right)'\right] - \pi\left(\theta_1^{\#}\right) \pi\left(\theta_1^{\#}\right)'\right)^{-1} \\
&= \left(E\left[\left(\frac{1}{2} - 1\left(y_i \leq \alpha^* + \beta^* d_i\right)\right)^2 z_i z_i'\right] - \pi\left(\theta_1^{\#}\right) \pi\left(\theta_1^{\#}\right)'\right)^{-1} \\
&= \left(E\left[E\left[\left(\frac{1}{2} - 1\left(y_i \leq \alpha^* + \beta^* d_i\right)\right)^2 \middle| z_i\right] z_i z_i'\right] - \pi\left(\theta_1^{\#}\right) \pi\left(\theta_1^{\#}\right)'\right)^{-1} \\
&= \left(\frac{1}{4} E\left[z_i z_i'\right] - \pi\left(\theta_1^{\#}\right) \pi\left(\theta_1^{\#}\right)'\right)^{-1}
\end{aligned}
$$

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

The estimated weighting matrix is

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

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 $W = I$.

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

The rate-adaptive bootstrap estimator in the case of an estimated weighting matrix is

$$\begin{aligned}
\hat{\theta}_n^* = \quad \arg\min_{\theta\in\Theta} \quad & \left\{\hat{\pi}_n\left(\hat{\theta}_n\right)' W_n\left(P_n^* - P_n\right)\left(\pi\left(\cdot,\theta\right) - \pi\left(\cdot,\hat{\theta}_n\right)\right)\right. \\
& + \frac{1}{2}\left(\theta - \hat{\theta}_n\right)'\left(\hat{G}' W_n \hat{G} + \sum_{j=1}^{m}\sum_{k=1}^{m} W_{n,jk}\hat{\pi}_{n,k}\left(\hat{\theta}\right)\hat{H}_j\right)\left(\theta - \hat{\theta}_n\right) \\
& + \left(\theta - \hat{\theta}_n\right)' \hat{G}' W_n\left(P_n^* - P_n\right)\pi\left(\cdot,\hat{\theta}_n\right) \\
& \left. + \left(\theta - \hat{\theta}_n\right)' \hat{G}'\left(W_n^* - W_n\right)\hat{\pi}_n\left(\hat{\theta}_n\right)\right\}
\end{aligned}$$

Tables 5 and 6 compare the empirical coverage frequencies and average interval lengths of nominal 95% equal-tailed confidence intervals $\left[\hat{\theta}_n - c_{0.975}, \hat{\theta}_n - c_{0.025}\right]$ constructed using the rate-adaptive bootstrap estimator and the centered standard bootstrap estimator $\tilde{\theta}_2 = \arg\min_{\theta\in\Theta}\left(\hat{\pi}_n^*\left(\theta\right) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)' W_n^*\left(\tilde{\theta}_1\right)\left(\hat{\pi}_n^*\left(\theta\right) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)$, where the weighting matrix depends on $\tilde{\theta}_1 = \arg\min_{\theta\in\Theta}\left(\hat{\pi}_n^*\left(\theta\right) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)'\left(\hat{\pi}_n^*\left(\theta\right) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)$. 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. The fact that the standard bootstrap's coverage approaches the nominal level as n increases could be due to the particularities of the data generating process which lead to moments that become smoother as n increases.

6 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.

Table 5: Rate-Adaptive vs. Standard Bootstrap, $\delta = 0.1$, estimated weighting matrix

	n	200	800	1600	3200	6400
	α_0	0.958	0.952	0.951	0.950	0.956
Rate-adaptive		(0.503)	(0.250)	(0.177)	(0.125)	(0.089)
	β_0	0.957	0.953	0.950	0.956	0.956
		(0.357)	(0.177)	(0.125)	(0.088)	(0.063)
	α_0	0.855	0.918	0.921	0.932	0.939
Standard		(0.357)	(0.245)	(0.174)	(0.123)	(0.087)
	β_0	0.864	0.921	0.921	0.937	0.938
		(0.281)	(0.175)	(0.123)	(0.087)	(0.062)

Table 6: Rate-Adaptive vs. Standard Bootstrap, $\delta = 0.9$, estimated weighting matrix

	n	200	800	1600	3200	6400
	α_0	0.949	0.951	0.951	0.949	0.955
Rate-adaptive		(0.501)	(0.250)	(0.177)	(0.125)	(0.088)
	β_0	0.957	0.953	0.951	0.957	0.957
		(0.356)	(0.177)	(0.125)	(0.088)	(0.063)
	α_0	0.880	0.916	0.921	0.932	0.940
Standard		(0.452)	(0.247)	(0.174)	(0.123)	(0.087)
	β_0	0.878	0.921	0.919	0.939	0.939
		(0.287)	(0.175)	(0.123)	(0.087)	(0.062)

7 Appendix

7.1 Additional Results for Misspecified GMM with Directionally Differentiable Moments

Asymptotic distribution under fixed weighting matrix

Theorem 5. *Suppose $\pi\left(\theta^{\#}\right) = c$ for a vector of fixed constants $c \neq 0$ and that Assumptions 1- 2 are satisfied for $\gamma = 1/2$ and $\rho = 1$, and $\pi(\cdot, \theta)$ is Lipschitz continuous in θ with a stochastically bounded Lipschitz constant. Suppose that for each $\epsilon > 0$ and $t \in \mathbb{R}^d$,*

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

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

$$\sqrt{n}\left(\hat{\theta}_n - \theta^{\#}\right) \rightsquigarrow \arg\min_{h\in\mathbb{R}^d} \left\{ \pi\left(\theta^{\#}\right)' W \mathcal{Z}_{0,1}(h) + h'G'WU_0 + \frac{1}{2} h'\bar{H}h \right\}$$

where $U_0 \sim N\left(0, P\left(\pi\left(\cdot, \theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)\left(\pi\left(\cdot, \theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)'\right)$ and $\mathcal{Z}_{0,1}(h)$ is a mean zero Gaussian process with covariance kernel $\Sigma_1(s,t) = \lim_{\alpha\to\infty} \alpha^2 P g\left(\cdot, \theta^{\#} + \frac{s}{\alpha}\right) g\left(\cdot, \theta^{\#} + \frac{t}{\alpha}\right)'$. The *joint covariance kernel of* $\mathcal{Z}_{0,1}(h)$ *and* $h'G'WU_0$ *is given by*

$$\Sigma(s,t) = \lim_{\alpha\to\infty} P \left[\begin{array}{c} \alpha g\left(\cdot, \theta^{\#} + \frac{s}{\alpha}\right) \\ s'G'W\left(\pi\left(\cdot,\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right) \end{array} \right] \left[\begin{array}{cc} \alpha g\left(\cdot,\theta^{\#}+\frac{t}{\alpha}\right)' & \left(\pi\left(\cdot,\theta^{\#}\right)-\pi\left(\theta^{\#}\right)\right)'WGt \end{array} \right]$$

Note that in the case of smooth misspecified models, the asymptotic distribution in Theorem 5 reduces down to the one in Theorem 1 of Hall and Inoue (2003) since then $\pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}(h) = h'Z_0'W\pi\left(\theta^{\#}\right)$, where $Z_0'W\pi\left(\theta^{\#}\right)$ is a mean zero Gaussian random variable with covariance matrix $P\left(\frac{\partial}{\partial\theta}\pi\left(\cdot,\theta^{\#}\right)-G\right)'W\pi\left(\theta^{\#}\right)\pi\left(\theta^{\#}\right)'W\left(\frac{\partial}{\partial\theta}\pi\left(\cdot,\theta^{\#}\right)-G\right)$.

The next theorem states that the rate adaptive bootstrap is consistent for globally misspecified GMM models with directionally differentiable moments.

Theorem 6. *Suppose Assumption 3 is satisfied in addition to the assumptions in Theorem 5,* $\hat{G} \xrightarrow{p} G$*, and* $\hat{H}_j \xrightarrow{p} H_j$ *for* $j = 1\ldots m$*. Then,*

$$n^{1/2}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \arg\min_{h\in\mathbb{R}^d} \left\{\pi\left(\theta^{\#}\right)'W\mathcal{Z}_{0,1}(h) + h'G'WU_0 + \frac{1}{2}h'\bar{H}h\right\}$$

Asymptotic distribution under estimated weighting matrix

Theorem 7. *Suppose* $\pi\left(\theta^{\#}\right) = c$ *for a vector of fixed constants* $c \neq 0$ *and that Assumptions 1- 2 are satisfied for* $\gamma = 1/2$ *and* $\rho = 1$*, and* $\pi(\cdot,\theta)$ *is Lipschitz continuous in* θ *with a stochastically bounded Lipschitz constant.*

If Assumption 4 is satisfied, then $\hat{\theta}_n - \theta^{\#} = o_P(1)$ *and*

$$\sqrt{n}\left(\hat{\theta}_n - \theta^{\#}\right) \rightsquigarrow \arg\min_{h\in\mathbb{R}^d}\left\{\pi\left(\theta^{\#}\right)'W\mathcal{Z}_{0,1}(h) + h'G'WU_0 + h'G'\Phi_0\pi\left(\theta^{\#}\right) + \frac{1}{2}h'\bar{H}h\right\}$$

where $\Phi_0\pi\left(\theta^{\#}\right) \sim N\left(0, P\left(\phi\left(\cdot,\theta_1^{\#}\right) - \phi\left(\theta_1^{\#}\right)\right)\pi\left(\theta^{\#}\right)\pi\left(\theta^{\#}\right)'\left(\phi\left(\cdot,\theta_1^{\#}\right)-\phi\left(\theta_1^{\#}\right)\right)'\right)$*. The joint covariance kernel of* $\mathcal{Z}_{0,1}(h)$*,* $h'G'WU_0$*, and* $h'G'\Phi_0\pi\left(\theta^{\#}\right)$ *is given by*

$$\Omega(s,t) = \lim_{\alpha\to\infty} P\left[\begin{array}{c} \alpha g\left(\cdot,\theta^{\#}+\frac{s}{\alpha}\right) \\ s'G'W\left(\pi\left(\cdot,\theta^{\#}\right)-\pi\left(\theta^{\#}\right)\right) \\ s'G'\left(\phi\left(\cdot,\theta_1^{\#}\right)-\phi\left(\theta_1^{\#}\right)\right)\pi\left(\theta^{\#}\right) \end{array}\right]\left[\begin{array}{c} \alpha g\left(\cdot,\theta^{\#}+\frac{t}{\alpha}\right)' \\ \left(\pi\left(\cdot,\theta^{\#}\right)-\pi\left(\theta^{\#}\right)\right)'WGt \\ \pi\left(\theta^{\#}\right)'\left(\phi\left(\cdot,\theta_1^{\#}\right)-\phi\left(\theta^{\#}\right)\right)'Gt \end{array}\right]'$$

If $W_n - W = o_p\left(n^{-1/2}\right)$*, then* $\hat{\theta}_n - \theta^{\#} = o_P(1)$ *and*

$$\sqrt{n}\left(\hat{\theta}_n - \theta^{\#}\right) \rightsquigarrow \arg\min_{h\in\mathbb{R}^d}\left\{\pi\left(\theta^{\#}\right)'W\mathcal{Z}_{0,1}(h) + h'G'WU_0 + \frac{1}{2}h'\bar{H}h\right\}$$

Note that in the case of smooth misspecified models, the asymptotic distribution in Theorem 7 reduces down to the one in Theorem 2 of Hall and Inoue (2003) since then $\pi\left(\theta^{\#}\right)'W\mathcal{Z}_{0,1}(h) = h'Z_0'W\pi\left(\theta^{\#}\right)$, where $Z_0'W\pi\left(\theta^{\#}\right)$ is a mean zero Gaussian random variable with covariance matrix $P\left(\frac{\partial}{\partial\theta}\pi\left(\cdot,\theta^{\#}\right)-G\right)'W\pi\left(\theta^{\#}\right)\pi\left(\theta^{\#}\right)'W\left(\frac{\partial}{\partial\theta}\pi\left(\cdot,\theta^{\#}\right)-G\right)$.

Theorem 8. *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 globally misspecified models where Assumptions 1 and 2 are satisfied for $\gamma = 1/2$ and $\rho = 1$, if $W_n - W = o_p\left(n^{-1/2}\right)$ and $W_n^* - W_n = o_p^*\left(n^{-1/2}\right)$, then*

$$n^{1/2}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \arg\min_{h \in \mathbb{R}^d} \left\{h'G'WU_0 + \pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}(h) + \frac{1}{2}h'\bar{H}h\right\}$$

If instead Assumption 4 is satisfied,

$$n^{1/2}\left(\hat{\theta}_n^* - \hat{\theta}_n\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \arg\min_{h \in \mathbb{R}^d} \left\{h'G'WU_0 + \pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}(h) + h'G'\Phi_0\pi\left(\theta^{\#}\right) + \frac{1}{2}h'\bar{H}h\right\}$$

7.2 Proofs for Theorems

Proof for Theorem 1

The consistency argument is a direct application of Theorem 5.7 in Van der Vaart (2000) since the arguments in the proof of Theorem 2.6 in Newey and McFadden (1994) imply that $\sup_{\theta \in \Theta}\left|\hat{Q}_n(\theta) - Q(\theta)\right| = o_p(1)$. Next we show that $n^{1/3}\left(\hat{\theta}_n - \theta^{\#}\right) = O_P(1)$. Define $\hat{\mathcal{G}}_n(\theta) = \sqrt{n}\left(P_n - P\right)g(\cdot,\theta)$, $\hat{g}(\theta) = P_n g(\cdot,\theta)$, and $g(\theta) = Pg(\cdot,\theta)$. Then $\hat{\pi}(\theta) = g(\theta) + \hat{\pi}\left(\theta^{\#}\right) + \hat{\eta}_n(\theta)$, where $\hat{\eta}_n(\theta) = \frac{1}{\sqrt{n}}\hat{\mathcal{G}}_n(\theta)$. Recall that $\hat{Q}_n(\theta) = \frac{1}{2}\hat{\pi}(\theta)' W\hat{\pi}(\theta)$. Write $\hat{Q}_n(\theta) - \hat{Q}_n\left(\theta^{\#}\right) = Q_1(\theta) + \hat{Q}_2(\theta) + \hat{Q}_3(\theta)$, where

$$Q_1(\theta) = \frac{1}{2}g(\theta)' Wg(\theta) + g(\theta)' W\pi\left(\theta^{\#}\right), \quad \hat{Q}_3(\theta) = \pi\left(\theta^{\#}\right)' W\hat{\eta}_n(\theta)$$
$$\hat{Q}_2(\theta) = \frac{1}{2}\hat{\eta}_n(\theta)' W\hat{\eta}_n(\theta) + g(\theta)' W\left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right) + g(\theta)' W\hat{\eta}_n(\theta) + \left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)' W\hat{\eta}_n(\theta).$$

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

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

The 1st, 3rd, and 4th terms in $\hat{Q}_2(\theta)$ are all of the form $o_P(1)\,\hat{\eta}_n(\theta)$, 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

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

Therefore, $\forall\epsilon > 0$, $\exists M_n = O_P(1)$ such that $|Q_1(\theta) + \hat{Q}_2(\theta) + \hat{Q}_3(\theta)| \leq \epsilon\|\theta - \theta^{\#}\|^2 + n^{-2/3}M_n^2$. Note that $\left|Q\left(\hat{\theta}_n\right) - Q\left(\theta^{\#}\right)\right| \leq \left|Pg\left(\cdot,\hat{\theta}_n\right)' WPg\left(\cdot,\hat{\theta}_n\right)\right| \leq \left\|Pg\left(\cdot,\hat{\theta}_n\right)\right\|^2$. Pick an ϵ such that $\|Pg(\cdot,\theta)\|^2 \leq -2\epsilon \left\|\theta-\theta^{\#}\right\|^2$ for θ in a neighborhood of $\theta^{\#}$. When $\hat{\theta}_n$ lies in this neighborhood,

$$\begin{aligned}
-o_p\left(n^{-2/3}\right) \leq \inf_{\theta\in\Theta}\hat{Q}_n(\theta)-\hat{Q}_n\left(\hat{\theta}_n\right) &\leq \left|\hat{Q}_n\left(\hat{\theta}_n\right)-\hat{Q}_n\left(\theta^{\#}\right)\right| \\
&\leq \left|\hat{Q}_n\left(\hat{\theta}_n\right)-\hat{Q}_n\left(\theta^{\#}\right)\right| + \left|Q\left(\hat{\theta}_n\right)-Q\left(\theta^{\#}\right)\right| \\
&\leq \epsilon\left\|\hat{\theta}_n-\theta^{\#}\right\|^2 + n^{-2/3}M_n^2 + \left\|Pg\left(\cdot,\hat{\theta}_n\right)\right\|^2 \\
&\leq \epsilon\left\|\hat{\theta}_n-\theta^{\#}\right\|^2 + n^{-2/3}M_n^2 - 2\epsilon\left\|\hat{\theta}_n-\theta^{\#}\right\|^2 \\
\Longrightarrow \left\|\hat{\theta}_n-\theta^{\#}\right\| &\leq \epsilon^{-1/2}n^{-1/3}M_n^2 + o_p\left(n^{-1/3}\right) = O_p\left(n^{-1/3}\right)
\end{aligned}$$

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

Next, $\hat{h}=n^{1/3}\left(\hat{\theta}_n-\theta^{\#}\right)=\arg\min_h n^{2/3}\hat{Q}_n\left(\theta^{\#}+n^{-1/3}h\right)$. Note that $\theta^{\#}$ being in the interior of Θ ensures that $\theta^{\#}+n^{-1/3}h$ will belong in Θ for n large enough. It will follow from the argmin continuous mapping theorem that $\hat{h} \rightsquigarrow \arg\min_h \pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1/2}(h)+\frac{1}{2}h'\bar{H}h$ if we can show that

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

as a process indexed by h in the space of locally bounded functions $\mathbf{B}_{\text{loc}}\left(\mathbb{R}^d\right)$ equipped with the topology of uniform convergence on compacta. Since $Q_1\left(\theta^{\#}+n^{-1/3}h\right)=Q_1\left(\theta^{\#}\right)+n^{-1/3}h'\frac{\partial Q_1\left(\theta^{\#}\right)}{\partial\theta}+\frac{1}{2}n^{-2/3}h'\frac{\partial^2 Q_1\left(\theta^{\#}\right)}{\partial\theta\partial\theta'}h+o\left(n^{-2/3}\right)$, $n^{2/3}Q_1\left(\theta^{\#}+n^{-1/3}h\right)=\frac{1}{2}h'\bar{H}h+o(1)$.

It remains to show that $n^{2/3}\left(\hat{Q}_2+\hat{Q}_3\right)\left(\theta^{\#}+n^{-1/3}h\right) \rightsquigarrow \pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1/2}(h)$. First note that Assumptions 2(iv) and (v) imply that the Lindeberg condition is satisfied. Then the Lindeberg-Feller CLT implies that $S_n(h) \equiv n^{2/3}\hat{\eta}_n\left(\theta^{\#}+n^{-1/3}h\right)=n^{1/6}\hat{\mathcal{G}}_n\left(\theta^{\#}+n^{-1/3}h\right)$ converges in finite dimensional distribution to a mean zero Gaussian process $\mathcal{Z}_{0,1/2}(h)$ with covariance kernel $\Sigma_{1/2}(s,t)=\lim_{\alpha\to\infty}\alpha Pg(\cdot,\theta^{\#}+\frac{s}{\alpha})g(\cdot,\theta^{\#}+\frac{t}{\alpha})'$.

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, and for every $j=1,\ldots,m$,

$$n^{2/3}E\sup_{\mathscr{D}(n)}\left|P_n d_j-Pd_j\right|=o(1) \tag{7.1}$$

where $\mathscr{D}(n)=\left\{d_j\left(\cdot,\theta^{\#},h_1,h_2\right)=g_j\left(\cdot;\theta^{\#}+n^{-1/3}h_1\right)-g_j\left(\cdot;\theta^{\#}+n^{-1/3}h_2\right)\right.$ such that $\max(\|h_1\|,\|h_2\|)\leq M$ and $\|h_1-h_2\|\leq\delta_n\}$. Note that $\mathscr{D}(n)$ has envelope function $D_n=2G_{R(n)}$ where $R(n)=Mn^{-1/3}$.

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

$$n^{2/3}E\sup_{\mathscr{D}(n)}\left|P_nd_j-Pd_j\right| \leq E\sqrt{n^{1/3}P_nD_n^2}\,J\left(\frac{n^{1/3}\sup_{\mathscr{D}(n)}P_nd_j^2}{n^{1/3}P_nD_n^2}\right)$$

$$\leq \sqrt{\eta} J(1) + \sqrt{E n^{1/3} P_n D_n^2} \sqrt{E J^2 \left(\min \left(1, \frac{1}{\eta} n^{1/3} \sup_{\mathscr{D}(n)} P_n d_j^2 \right) \right)}.$$

To show that this is $o(1)$ for each fixed $\eta > 0$, first, note that by Assumption 2(vi), $E n^{1/3} P_n D_n^2 = 4 n^{1/3} E G_{R(n)}^2 = O(n^{1/3} R(n)) = O(1)$ since $R(n) = M n^{-1/3}$. The proof will then be complete if $n^{1/3} \sup_{\mathscr{D}(n)} P_n d_j^2 = o_p(1)$. Next, for each $K > 0$ write $E \sup_{\mathscr{D}(n)} P_n d_j^2 \leq E P_n \sup_{\mathscr{D}(n)} d_j^2 1\{D_n > K\} + K E \sup_{\mathscr{D}(n)} P_n |d_j| \leq E P_n D_n^2 1\{D_n > K\} + K \sup_{\mathscr{D}(n)} P |d_j| + K E \sup_{\mathscr{D}(n)} |P_n |d_j| - P |d_j||$. By Assumption 2(vii), $E P_n D_n^2 1\{D_n > K\} < \eta n^{-1/3}$ for large enough K. By Assumption 2(viii) and the definition of $\mathscr{D}(n)$, $K \sup_{\mathscr{D}(n)} P |d_j| = O(n^{-1/3} \delta_n) = o(n^{-1/3})$. By Assumption 2(vi) and the maximal inequality in Section 3.1 of Kim and Pollard (1990), $K E \sup_{\mathscr{D}(n)} |P_n |d_j| - P |d_j|| < K n^{-\frac{1}{2}} J(1) \sqrt{P D_n^2} = O(n^{-2/3}) = o(n^{-1/3})$. Therefore, $E n^{1/3} \sup_{\mathscr{D}(n)} P_n d_j^2 = o(1)$.

We have shown that $S_n(h) \rightsquigarrow \mathcal{Z}_{0,1/2}(h)$, which implies that $n^{2/3} \hat{Q}_3 \left(\theta^{\#} + n^{-1/3} h \right) \rightsquigarrow \pi \left(\theta^{\#} \right)' W \mathcal{Z}_{0,1/2}(h)$. Since the 1st, 3rd and 4th terms in $n^{2/3} \hat{Q}_2 \left(\theta^{\#} + n^{-1/3} h \right)$ are all of the form $o_P(1) n^{2/3} \hat{\eta}_n \left(\theta^{\#} + n^{-1/3} h \right)$, they all converge in probability to 0. For the 2nd term there,

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

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

$$n^{2/3} \left(Q_1 + \hat{Q}_2 + \hat{Q}_3 \right) \left(\theta^{\#} + n^{-1/3} h \right) \rightsquigarrow \pi \left(\theta^{\#} \right)' W \mathcal{Z}_{0,1/2}(h) + \frac{1}{2} h' \bar{H} h.$$

Lemma 2.6 in Kim and Pollard (1990) implies that the Gaussian process $-\mathcal{Z}_{0,1/2}(h)$ has a unique maximum, which implies that $\mathcal{Z}_{0,1/2}(h)$ has a unique minimum. In combination with the fact that $\frac{1}{2} h' \bar{H} h$ is a convex function of h, there is a unique $\hat{h} = n^{1/3} \left(\hat{\theta}_n - \theta^{\#} \right)$ that minimizes $\pi \left(\theta^{\#} \right)' W \mathcal{Z}_{0,1/2}(h) + \frac{1}{2} h' \bar{H} h$. The result follows from the argmin continuous mapping theorem (Theorem 2.7 in Kim and Pollard (1990)). ■

Proof for Theorems 2 and 6

Equation 4.2 implies that for $\hat{h}^* = n^{\gamma} \left(\hat{\theta}_n^* - \hat{\theta}_n \right)$,

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

Assumptions 2(iv) and (v) imply the Lindeberg condition is satisfied, so by the Lindeberg-Feller CLT, $S_n(h) \equiv n^{\gamma\rho}\sqrt{n}\left(P_n - P\right)\left(\pi\left(\cdot, \theta^\# + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta^\#\right)\right)$ converges in finite dimensional distribution to a mean zero Gaussian process $\mathcal{Z}_{0,\rho}\left(h\right)$ with covariance kernel $\Sigma_\rho(s,t) = \lim_{\alpha\to\infty} \alpha^{2\rho} P g\left(\cdot, \theta^\# + \frac{s}{\alpha}\right) g\left(\cdot, \theta^\# + \frac{t}{\alpha}\right)'$.

We already showed in Theorem 1 that $S_n(h)$ is stochastically equicontinuous in h for $\rho = 1/2$, $\gamma = 1/3$, and we already showed in Theorem 5 that $S_n(h)$ is stochastically equicontinuous in h for $\rho = 1$, $\gamma = 1/2$.

Therefore, $S_n\left(h\right) \rightsquigarrow \mathcal{Z}_{0,\rho}\left(h\right)$ as a process indexed by h in $\mathbf{B}_{\text{loc}}\left(\mathbb{R}^d\right)$ equipped with the topology of uniform convergence on compacta. Theorem 3.6.13 in van der Vaart and Wellner (1996) or Theorem 2.6 in Kosorok (2007) then implies that the bootstrapped process $n^{\gamma\rho}\sqrt{n}\left(P_n^* - P_n\right)\left(\pi\left(\cdot, \theta^\# + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta^\#\right)\right)$ is consistent for the same limiting process as $S_n\left(h\right)$:

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

Under Assumption 2, $\mathcal{G}_R \equiv \left\{\pi_j\left(\cdot, \theta\right) - \pi_j\left(\cdot, \theta^\#\right) : \|\theta - \theta^\#\| \leq R, j = 1 \ldots m\right\}$ for R near zero are uniformly manageable classes (and therefore Donsker classes) that satisfy for all $j = 1 \ldots m$, $P\left(\pi_j\left(\cdot, \theta\right) - \pi_j\left(\cdot, \theta^\#\right)\right)^2 \to 0$ for $\theta \to \theta^\#$. By Lemma 3.3.5 of van der Vaart and Wellner (1996), $n^{\gamma\rho}\sqrt{n}\left(P_n - P\right)\left(\pi\left(\cdot, \theta + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta\right)\right)$ is stochastically equicontinuous in θ, which implies

$$\begin{aligned}
&n^{\gamma\rho}\sqrt{n}\sup_{h\in\mathbb{R}^d}\left\|\left(P_n - P\right)\left(\pi\left(\cdot, \hat{\theta}_n + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \hat{\theta}_n\right) - \left(\pi\left(\cdot, \theta^\# + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta^\#\right)\right)\right)\right\| \\
&= o_p\left(1 + n^\gamma\left\|\hat{\theta}_n - \theta^\#\right\|\right) = o_p(1)
\end{aligned}$$

Under the envelope integrability Assumption 3, Lemma 4.2 in Wellner and Zhan (1996) implies that

$$\begin{aligned}
&n^{\gamma\rho}\sqrt{n}\sup_{h\in\mathbb{R}^d}\left\|\left(P_n^* - P_n\right)\left(\pi\left(\cdot, \hat{\theta}_n + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \hat{\theta}_n\right) - \left(\pi\left(\cdot, \theta^\# + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \theta^\#\right)\right)\right)\right\| \\
&= o_p^*\left(1 + n^\gamma\left\|\hat{\theta}_n - \theta^\#\right\|\right) = o_p^*(1)
\end{aligned}$$

In combination with the fact that $\hat{\pi}\left(\hat{\theta}_n\right) \xrightarrow{p} \pi\left(\theta^\#\right)$,

$$\hat{\pi}\left(\hat{\theta}_n\right)' W n^{\gamma\rho}\sqrt{n}\left(P_n^* - P_n\right)\left(\pi\left(\cdot, \hat{\theta}_n + \frac{h}{n^\gamma}\right) - \pi\left(\cdot, \hat{\theta}_n\right)\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} \pi\left(\theta^\#\right)' W \mathcal{Z}_{0,\rho}\left(h\right)$$

For the second term, note that since $\frac{\sqrt{n}n^{\gamma\rho}}{n^{2\gamma}} = 1$, $\hat{G} \xrightarrow{p} G$, $\hat{H}_j \xrightarrow{p} H_j$ for $j = 1 \ldots m$, and $\hat{\pi}\left(\hat{\theta}_n\right) \xrightarrow{p} \pi\left(\theta^\#\right)$,

$$\frac{\sqrt{n}n^{\gamma\rho}}{2n^{2\gamma}} h'\left(\hat{G}'W\hat{G} + \sum_{j=1}^m \sum_{k=1}^m W_{jk}\hat{\pi}_k\left(\hat{\theta}_n\right)\hat{H}_j\right) h \xrightarrow{p} \frac{1}{2} h'\left(G'WG + \sum_{j=1}^m \sum_{k=1}^m W_{jk}\pi_k\left(\theta^\#\right) H_j\right) h \equiv \frac{1}{2} h'\bar{H} h$$

When $\gamma = 1/3$ and $\rho = 1/2$, $\frac{n^{\gamma\rho}}{n^{\gamma}} = o(1)$, which implies that the third term is $o_p^*(1)$:

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

It then follows from a bootstrapped version of the argmin continuous mapping theorem (see Lemma 14.2 in Hong and Li (2020) for proof)

$$\hat{h}^* \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \arg\min_{h\in\mathbb{R}^d}\left\{\pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,\rho}\left(h\right) + \frac{1}{2}h'\bar{H}h\right\}$$

For misspecified nonsmooth models with $\gamma = 1/2$ and $\rho = 1$, $\frac{n^{\gamma\rho}}{n^{\gamma}} = 1$, so the third term also contributes to the asymptotic distribution.

We showed in Theorem 5 $\begin{pmatrix} \hat{\pi}\left(\theta^{\#}\right)' W n^{\gamma\rho}\sqrt{n}\left(P_n - P\right)\left(\pi\left(\cdot, \theta^{\#} + \frac{h}{n^{\gamma}}\right) - \pi\left(\cdot, \theta^{\#}\right)\right) \\ h'\hat{G}'W\sqrt{n}\left(P_n - P\right)\pi\left(\cdot, \theta^{\#}\right) \end{pmatrix} \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}}$ $\begin{pmatrix} \pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}\left(h\right) \\ h'G'WU_0 \end{pmatrix}$. Under Assumption 2, $\mathcal{G}_R \equiv \left\{\pi_j\left(\cdot, \theta\right) - \pi_j\left(\cdot, \theta^{\#}\right) : \|\theta - \theta^{\#}\| \leq R, j = 1\ldots m\right\}$ for R near zero are uniformly manageable classes (and therefore Donsker classes) that satisfy for all $j = 1\ldots m$, $P\left(\pi_j\left(\cdot, \theta\right) - \pi_j\left(\cdot, \theta^{\#}\right)\right)^2 \to 0$ for $\theta \to \theta^{\#}$. By Lemma 3.3.5 of van der Vaart and Wellner (1996),

$$\left\|\sqrt{n}\left(P_n - P\right)\left(\pi\left(\cdot, \hat{\theta}_n\right) - \pi\left(\cdot, \theta^{\#}\right)\right)\right\| = o_p\left(1 + \sqrt{n}\left\|\hat{\theta}_n - \theta^{\#}\right\|\right) = o_p(1)$$

Under the envelope integrability Assumption 3, Lemma 4.2 in Wellner and Zhan (1996) implies that the process is bootstrap equicontinuous.

$$\left\|\sqrt{n}\left(P_n^* - P_n\right)\left(\pi\left(\cdot, \hat{\theta}_n\right) - \pi\left(\cdot, \theta^{\#}\right)\right)\right\| = o_p^*\left(1 + \sqrt{n}\left\|\hat{\theta}_n - \theta^{\#}\right\|\right) = o_p^*(1)$$

Therefore,

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

And it follows from a bootstrapped version of the argmin continuous mapping theorem (see Lemma 14.2 in Hong and Li (2020) for proof)

$$\hat{h}^* \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \arg\min_{h\in\mathbb{R}^d}\left\{\pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}\left(h\right) + h'G'WU_0 + \frac{1}{2}h'\bar{H}h\right\}$$

Under correct model specification, $\pi\left(\theta^{\#}\right) = 0$, so the first term $\pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}\left(h\right)$ disappears and

$$\begin{aligned} \hat{h}^* \quad &\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \quad \arg\min_{h\in\mathbb{R}^d}\left\{\frac{1}{2}h'G'WGh + h'G'WU_0\right\} \\ &= \quad -\left(G'WG\right)^{-1} G'WN\left(0, P\pi\left(\cdot, \theta^{\#}\right)\pi\left(\cdot, \theta^{\#}\right)'\right) \end{aligned}$$

For smooth models that are misspecified, $\hat{\pi}\left(\hat{\theta}_n\right)' Wn\left(P_n^* - P_n\right)\left(\pi\left(\cdot, \hat{\theta}_n + \frac{h}{n^{1/2}}\right) - \pi\left(\cdot, \hat{\theta}_n\right)\right) \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \pi\left(\theta^{\#}\right)' W \mathcal{Z}_{0,1}(h) = h' Z_0' W \pi\left(\theta^{\#}\right)$, where $Z_0' W \pi\left(\theta^{\#}\right)$ is a mean zero Gaussian random variable with covariance matrix $P\left(\frac{\partial}{\partial\theta}\pi\left(\cdot, \theta^{\#}\right) - G\right)' W \pi\left(\theta^{\#}\right) \pi\left(\theta^{\#}\right)' W\left(\frac{\partial}{\partial\theta}\pi\left(\cdot, \theta^{\#}\right) - G\right)$.

Furthermore, the joint distribution of $Z_0' W \pi\left(\theta^{\#}\right)$ and U_0 is given by

$$\left(\begin{array}{c} U_0 \\ Z_0' W \pi\left(\theta^{\#}\right) \end{array}\right) \sim N\left(0, \left(\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right)\right)$$

$$\begin{aligned}
\Sigma_{11} &= P\left(\pi\left(\cdot, \theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)\left(\pi\left(\cdot, \theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)' \\
\Sigma_{12} &= P\left(\pi\left(\cdot, \theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)\pi\left(\theta^{\#}\right)' W\left(\frac{\partial}{\partial\theta}\pi\left(\cdot, \theta^{\#}\right) - G\right) \\
\Sigma_{21} &= P\left(\frac{\partial}{\partial\theta}\pi\left(\cdot, \theta^{\#}\right) - G\right)' W \pi\left(\theta^{\#}\right)\left(\pi\left(\cdot, \theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)' \\
\Sigma_{22} &= P\left(\frac{\partial}{\partial\theta}\pi\left(\cdot, \theta^{\#}\right) - G\right)' W \pi\left(\theta^{\#}\right) \pi\left(\theta^{\#}\right)' W\left(\frac{\partial}{\partial\theta}\pi\left(\cdot, \theta^{\#}\right) - G\right)
\end{aligned}$$

Therefore, the asymptotic distribution is given by

$$\begin{aligned}
\hat{h}^* &\overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \arg\min_{h\in\mathbb{R}^d}\left\{h' Z_0' W \pi\left(\theta^{\#}\right) + h' G' W U_0 + \frac{1}{2} h' \bar{H} h\right\} \\
&= -\bar{H}^{-1} N\left(0, G' W \Sigma_{11} W G + \Sigma_{22} + G' W \Sigma_{12} + \Sigma_{21} W G\right)
\end{aligned}$$

■

Proof for Theorem 3

The consistency argument is a direct application of Theorem 5.7 in Van der Vaart (2000) since the arguments in the proof of Theorem 2.6 in Newey and McFadden (1994) imply that $\sup_{\theta\in\Theta}\left|\hat{Q}_n(\theta) - Q(\theta)\right| = o_p(1)$. Next, write $\hat{Q}_n(\theta) - \hat{Q}_n\left(\theta^{\#}\right) = Q_1(\theta) + \hat{Q}_2(\theta) + \hat{Q}_3(\theta) + \hat{Q}_4(\theta) + \hat{Q}_5(\theta) + \hat{Q}_6(\theta)$, where

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

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 $\hat{\eta}_n(\theta)$, and in turn $\hat{Q}_6(\theta) = o_P(1)\,\hat{\eta}_n(\theta)$ 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 $o_P(1)\,\hat{\eta}_n(\theta)$, 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 $\forall\epsilon > 0$,

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

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

$$\begin{aligned}
\hat{Q}_4(\theta) &= \hat{Q}_4\left(\theta^{\#}\right) + \left(\theta - \theta^{\#}\right)'\frac{\partial\hat{Q}_4\left(\theta^{\#}\right)}{\partial\theta} + \frac{1}{2}\left(\theta - \theta^{\#}\right)'\frac{\partial^2\hat{Q}_4\left(\theta^{\#}\right)}{\partial\theta\partial\theta'}\left(\theta - \theta^{\#}\right) + o_p\left(\|\theta - \theta^{\#}\|^2\right) \\
&= \frac{1}{2}\left(\theta - \theta^{\#}\right)'\left(\frac{\partial^2\hat{Q}_4\left(\theta^{\#}\right)}{\partial\theta\partial\theta'} + o_p(1)\right)\left(\theta - \theta^{\#}\right) + O_p\left(\frac{\|\theta - \theta^{\#}\|}{n^{\gamma}}\right) \\
&\leq \epsilon\|\theta - \theta^{\#}\|^2 + n^{-2\gamma}M_{n,4}^2.
\end{aligned}$$

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 $\left|Q\left(\hat{\theta}_n\right) - Q\left(\theta^{\#}\right)\right| \leq \left|Pg\left(\cdot, \hat{\theta}_n\right)' W Pg\left(\cdot, \hat{\theta}_n\right)\right| \leq \left\|Pg\left(\cdot, \hat{\theta}_n\right)\right\|^2$. Pick an ϵ such that $\|Pg(\cdot, \theta)\|^2 \leq -2\epsilon\left\|\theta - \theta^{\#}\right\|^2$ for θ in a neighborhood of $\theta^{\#}$. When $\hat{\theta}_n$ lies in this neighborhood,

$$\begin{aligned}
-o_p\left(n^{-2/3}\right) \leq \inf_{\theta\in\Theta}\hat{Q}_n(\theta) - \hat{Q}_n\left(\hat{\theta}_n\right) &\leq \left|\hat{Q}_n\left(\hat{\theta}_n\right) - \hat{Q}_n\left(\theta^{\#}\right)\right| \\
&\leq \left|\hat{Q}_n\left(\hat{\theta}_n\right) - \hat{Q}_n\left(\theta^{\#}\right)\right| + \left|Q\left(\hat{\theta}_n\right) - Q\left(\theta^{\#}\right)\right| \\
&\leq \epsilon\left\|\hat{\theta}_n - \theta^{\#}\right\|^2 + n^{-2/3}M_n^2 + \left\|Pg\left(\cdot, \hat{\theta}_n\right)\right\|^2 \\
&\leq \epsilon\left\|\hat{\theta}_n - \theta^{\#}\right\|^2 + n^{-2/3}M_n^2 - 2\epsilon\left\|\hat{\theta}_n - \theta^{\#}\right\|^2 \\
\implies \left\|\hat{\theta}_n - \theta^{\#}\right\| &\leq \epsilon^{-1/2}n^{-1/3}M_n^2 + o_p\left(n^{-1/3}\right) = O_p\left(n^{-1/3}\right)
\end{aligned}$$

We already showed in Theorem 1 that $n^{2/3}Q_1\left(\theta^{\#} + n^{-1/3}h\right) = \frac{1}{2}h'\bar{H}h + o(1)$, $n^{2/3}\hat{Q}_3\left(\theta^{\#} + n^{-1/3}h\right) \rightsquigarrow \pi\left(\theta^{\#}\right)' W\mathcal{Z}_0(h)$, and $n^{2/3}\hat{Q}_2\left(\theta^{\#} + n^{-1/3}h\right) = o_p(1)$. Furthermore, if $W_n - W = O_p\left(n^{-1/3}\right)$,

$$\begin{aligned}
n^{2/3}\hat{Q}_4\left(\theta^{\#} + n^{-1/3}h\right) &= \frac{1}{2}n^{2/3}g\left(\theta^{\#} + n^{-1/3}h\right)'(W_n - W)\,g\left(\theta^{\#} + n^{-1/3}h\right) \\
&\quad + n^{2/3}g\left(\theta^{\#} + n^{-1/3}h\right)'(W_n - W)\,\pi\left(\theta^{\#}\right)
\end{aligned}$$

$$
\begin{aligned}
&= n^{2/3} O_p\left(\frac{\left\|n^{-1/3}h\right\|^2}{n^{1/3}}\right) \\
&+ \left(n^{1/3}\left\{g\left(\theta^{\#}\right)' + h'G'n^{-1/3}\right\} + o_p\left(1\right)\right) n^{1/3}\left(W_n - W\right)\pi\left(\theta^{\#}\right) \\
&= h'G'n^{1/3}\left(W_n - W\right)\pi\left(\theta^{\#}\right) + o_p\left(1\right) \rightsquigarrow h'G'\mathcal{W}_0 \\
n^{2/3}\hat{Q}_5\left(\theta^{\#} + n^{-1/3}h\right) &= n^{2/3} g\left(\theta^{\#} + n^{-1/3}h\right)'\left(W_n - W\right)\left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right) \\
&+ g\left(\theta^{\#} + n^{-1/3}h\right)'\left(W_n - W\right) n^{2/3}\hat{\eta}_n\left(\theta^{\#} + n^{-1/3}h\right) \\
&+ \left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)'\left(W_n - W\right) n^{2/3}\hat{\eta}_n\left(\theta^{\#} + n^{-1/3}h\right) \\
&= n^{2/3} O_p\left(\frac{\left\|n^{-1/3}h\right\|}{n^{5/6}}\right) + O_p\left(\frac{\left\|n^{-1/3}h\right\|}{n^{1/3}}\right) O_p\left(1\right) + O_p\left(n^{-5/6}\right) O_p\left(1\right) \\
&= o_p(1) \\
n^{2/3}\hat{Q}_6\left(\theta^{\#} + n^{-1/3}h\right) &= \pi\left(\theta^{\#}\right)'\left(W_n - W\right) n^{2/3}\hat{\eta}_n\left(\theta^{\#} + n^{-1/3}h\right) \\
&+ \frac{1}{2}\hat{\eta}_n\left(\theta^{\#} + n^{-1/3}h\right)'\left(W_n - W\right) n^{2/3}\hat{\eta}_n\left(\theta^{\#} + n^{-1/3}h\right) \\
&= O_p\left(n^{-1/3}\right) O_p\left(1\right) + O_p\left(n^{-2/3}\right) O_p\left(n^{-1/3}\right) O_p\left(1\right) \\
&= o_p(1)
\end{aligned}
$$

By assumption,

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

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

$$
n^{1/3}\left(\hat{\theta}_n - \theta^{\#}\right) \rightsquigarrow \arg\min_{h \in \mathbb{R}^d}\left\{\pi\left(\theta^{\#}\right)' W \mathcal{Z}_{0,1/2}\left(h\right) + h'G'\mathcal{W}_0 + \frac{1}{2} h'\bar{H}h\right\}
$$

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

$$
\begin{aligned}
n^{2/3}\hat{Q}_4\left(\theta^{\#} + n^{-1/3}h\right) &= n^{2/3} o_p\left(\frac{\left\|n^{-1/3}h\right\|^2}{n^{1/3}}\right) + n^{2/3} o_p\left(\frac{\left\|n^{-1/3}h\right\|}{n^{1/3}}\right) \\
&= o_p\left(\frac{1}{n^{1/3}}\right) + o_p(1) = o_p(1)
\end{aligned}
$$

■

Proof for Theorems 4 and 8

Equation 4.1 implies that for $\hat{h}^* = n^{\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$,

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

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

We already showed in Theorem 2 that

$$\hat{\pi}\left(\hat{\theta}_n\right)'Wn^{\gamma\rho}\sqrt{n}\left(P_n^*-P_n\right)\left(\pi\left(\cdot,\hat{\theta}_n+\frac{h}{n^{\gamma}}\right)-\pi\left(\cdot,\hat{\theta}_n\right)\right)\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}}\pi\left(\theta^{\#}\right)'W\mathcal{Z}_{0,\rho}\left(h\right)$$

Consistency of W_n for W implies that

$$\hat{\pi}\left(\hat{\theta}_n\right)'W_nn^{\gamma\rho}\sqrt{n}\left(P_n^*-P_n\right)\left(\pi\left(\cdot,\hat{\theta}_n+\frac{h}{n^{\gamma}}\right)-\pi\left(\cdot,\hat{\theta}_n\right)\right)\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}}\pi\left(\theta^{\#}\right)'W\mathcal{Z}_{0,\rho}\left(h\right)$$

We also showed in Theorem 2 that

$$\frac{\sqrt{n}n^{\gamma\rho}}{2n^{2\gamma}}h'\left(\hat{G}'W\hat{G}+\sum_{j=1}^{m}\sum_{k=1}^{m}W_{jk}\hat{\pi}_k\left(\hat{\theta}_n\right)\hat{H}_j\right)h\overset{p}{\to}\frac{1}{2}h'\left(G'WG+\sum_{j=1}^{m}\sum_{k=1}^{m}W_{jk}\pi_k\left(\theta^{\#}\right)H_j\right)h\equiv\frac{1}{2}h'\bar{H}h$$

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

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

When $W_n-W=O_p\left(n^{-1/3}\right)$ and $W_n^*-W_n=O_p^*\left(n^{-1/3}\right)$, we assumed $\begin{pmatrix}\pi\left(\theta^{\#}\right)'Wn^{2/3}\left(P_n^*-P_n\right)g\left(\cdot,\theta^{\#}+n^{-1/3}h\right)\\ h'G'n^{1/3}\left(W_n^*-W_n\right)\pi\left(\theta^{\#}\right)\end{pmatrix}\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}}\begin{pmatrix}\pi\left(\theta^{\#}\right)'W\mathcal{Z}_{0,1/2}\left(h\right)\\ h'G'\mathcal{W}_0\end{pmatrix}$. Under the uniform manageability Assumption 2 and the envelope integrability Assumption 3, we can invoke Lemma 4.2 in Wellner and Zhan (1996) to show bootstrap equicontinuity. Therefore,

$$\begin{pmatrix}\hat{\pi}_n\left(\hat{\theta}_n\right)'W_nn^{2/3}\left(P_n^*-P_n\right)\left(\pi\left(\cdot,\hat{\theta}_n+\frac{h}{n^{1/3}}\right)-\pi\left(\cdot,\hat{\theta}_n\right)\right)\\ h'G'n^{1/3}\left(W_n^*-W_n\right)\pi\left(\hat{\theta}_n\right)\end{pmatrix}\underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}}\begin{pmatrix}\pi\left(\theta^{\#}\right)'W\mathcal{Z}_{0,1/2}\left(h\right)\\ h'G'\mathcal{W}_0\end{pmatrix}$$

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

$$\hat{h}^* \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \arg\min_{h\in\mathbb{R}^d} \left\{ \pi\left(\theta^\#\right)' W \mathcal{Z}_{0,1/2}\left(h\right) + h'G'\mathcal{W}_0 + \frac{1}{2}h'\bar{H}h \right\}$$

For misspecified nonsmooth models with $\rho = 1$, $\gamma = 1/2$, we already showed in Theorem 7

$$\begin{pmatrix} \pi\left(\theta^\#\right)' W_n n\left(P_n - P\right) g\left(\cdot, \theta^\# + n^{-1/2}h\right) \\ h'G'W_n\sqrt{n}\left(P_n - P\right)\pi\left(\cdot, \theta^\#\right) \\ h'G'n^{1/2}\sqrt{n}\left(W_n - W\right)\pi\left(\theta^\#\right) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \pi\left(\theta^\#\right)' W \mathcal{Z}_{0,1}\left(h\right) \\ h'G'WU_0 \\ h'G'\Phi_0\pi\left(\theta^\#\right) \end{pmatrix}$$

Under Assumption 4, the bootstrapped weighting matrix can be written as $\sqrt{n}\left(W_n^* - W_n\right) = \sqrt{n}\left(P_n^* - P_n\right)\phi\left(\cdot, \theta_1^\#\right) + o_p^*(1)$. Under the uniform manageability Assumption 2 and the envelope integrability Assumption 3, we can invoke Lemma 4.2 in Wellner and Zhan (1996) to show bootstrap equicontinuity. Therefore,

$$\begin{pmatrix} \hat{\pi}\left(\hat{\theta}_n\right)' W_n n\left(P_n^* - P_n\right)\left(\pi\left(\cdot, \hat{\theta}_n + \frac{h}{\sqrt{n}}\right) - \pi\left(\cdot, \hat{\theta}_n\right)\right) \\ h'\hat{G}'W_n\sqrt{n}\left(P_n^* - P_n\right)\pi\left(\cdot, \hat{\theta}_n\right) \\ h'\hat{G}'n^{1/2}\sqrt{n}\left(W_n^* - W_n\right)\pi\left(\hat{\theta}_n\right) \end{pmatrix} \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \begin{pmatrix} \pi\left(\theta^\#\right)' W \mathcal{Z}_{0,1}\left(h\right) \\ h'G'WU_0 \\ h'G'\Phi_0\pi\left(\theta^\#\right) \end{pmatrix}$$

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

$$\hat{h}^* \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \arg\min_{h\in\mathbb{R}^d} \left\{ h'G'WU_0 + \pi\left(\theta^\#\right)' W \mathcal{Z}_{0,1}\left(h\right) + h'G'\Phi_0\pi\left(\theta^\#\right) + \frac{1}{2}h'\bar{H}h \right\}.$$

For misspecified smooth models where $\rho = 1$ and $\gamma = 1/2$, $\pi\left(\theta^\#\right)' W \mathcal{Z}_{0,1}\left(h\right) = h'Z_0'W\pi\left(\theta^\#\right)$, and the joint distribution of U_0, $Z_0'W\pi\left(\theta^\#\right)$, and $\Phi_0\pi\left(\theta^\#\right)$ is given by

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

Then the asymptotic distribution is given by

$$\hat{h}^* \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \arg\min_{h\in\mathbb{R}^d} \left\{ h'G'WU_0 + h'Z_0'W\pi\left(\theta^\#\right) + h'G'\Phi_0\pi\left(\theta^\#\right) + \frac{1}{2}h'\bar{H}h \right\} = N\left(0, \bar{H}^{-1}\Omega_W\bar{H}^{-1}\right)$$

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

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

$$\begin{aligned} \hat{h}^* & \overset{\mathbb{P}}{\underset{\mathbb{W}}{\rightsquigarrow}} \arg\min_{h\in\mathbb{R}^d} \left\{ \frac{1}{2}h'G'WGh + h'G'WU_0 \right\} \\ & = \left(G'WG\right)^{-1} G'WN\left(0, P\pi\left(\cdot, \theta^\#\right)\pi\left(\cdot, \theta^\#\right)'\right) \end{aligned}$$

■

Proof for Theorem 5

The consistency argument 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 $\hat{Q}_n(\theta) - \hat{Q}_n\left(\theta^{\#}\right) = Q_1(\theta) + \hat{Q}_2(\theta) + \hat{Q}_3(\theta)$, where

$$Q_1(\theta) = \frac{1}{2} g(\theta)' W g(\theta) + g(\theta)' W \pi\left(\theta^{\#}\right), \quad \hat{Q}_3(\theta) = \pi\left(\theta^{\#}\right)' W \hat{\eta}_n(\theta)$$
$$\hat{Q}_2(\theta) = \frac{1}{2}\hat{\eta}_n(\theta)' W \hat{\eta}_n(\theta) + g(\theta)' W\left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right) + g(\theta)' W \hat{\eta}_n(\theta) + \left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)' W \hat{\eta}_n(\theta).$$

where $\hat{\eta}_n(\theta) = (P_n - P) g(\cdot, \theta)$, $\hat{g}(\theta) = P_n g(\cdot, \theta)$, and $g(\theta) = P g(\cdot, \theta)$. Note that the Taylor expansion of $Q_1(\theta)$ around $\theta^{\#}$ is $Q_1(\theta) = Q_1\left(\theta^{\#}\right) + \left(\theta - \theta^{\#}\right)' \frac{\partial Q_1\left(\theta^{\#}\right)}{\partial \theta} + \frac{1}{2}\left(\theta - \theta^{\#}\right)' \frac{\partial^2 Q_1\left(\theta^{\#}\right)}{\partial \theta \partial \theta'}\left(\theta - \theta^{\#}\right) + o\left(\|\theta - \theta^{\#}\|^2\right) = \frac{1}{2}\left(\theta - \theta^{\#}\right)'\left(\bar{H} + o(1)\right)\left(\theta - \theta^{\#}\right)$ since $\frac{\partial Q_1\left(\theta^{\#}\right)}{\partial \theta} = G' W g\left(\theta^{\#}\right) + G' W \pi\left(\theta^{\#}\right) = 0$ and $\frac{\partial^2 Q_1\left(\theta^{\#}\right)}{\partial \theta \partial \theta'} = \bar{H}$. Next apply a modified version of Kim and Pollard (1990) Lemma 4.1 with $\gamma = 1/2$, $\rho = 1$, [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

$$|g(\theta)' W\left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)| = O_p\left(\frac{\|\theta - \theta^{\#}\|}{\sqrt{n}}\right) \leq \epsilon\|\theta - \theta^{\#}\|^2 + n^{-1} M_{n,22}^2.$$

Therefore, $\forall \epsilon > 0$, $\exists M_n = O_P(1)$ such that $|Q_1(\theta) + \hat{Q}_2(\theta) + \hat{Q}_3(\theta)| \leq \epsilon\|\theta - \theta^{\#}\|^2 + n^{-1} M_n^2$.

Note that $\left|Q\left(\hat{\theta}_n\right) - Q\left(\theta^{\#}\right)\right| \leq \left|P g\left(\cdot, \hat{\theta}_n\right)' W P g\left(\cdot, \hat{\theta}_n\right)\right| \leq \left\|P g\left(\cdot, \hat{\theta}_n\right)\right\|^2$. Pick an ϵ such that $\|P g(\cdot, \theta)\|^2 \leq -2\epsilon\left\|\theta - \theta^{\#}\right\|^2$ for θ in a neighborhood of $\theta^{\#}$. When $\hat{\theta}_n$ lies in this neighborhood,

$$\begin{aligned}
-o_p\left(n^{-1}\right) \leq \inf_{\theta \in \Theta} \hat{Q}_n(\theta) - \hat{Q}_n\left(\hat{\theta}_n\right) &\leq \left|\hat{Q}_n\left(\hat{\theta}_n\right) - \hat{Q}_n\left(\theta^{\#}\right)\right| \\
&\leq \left|\hat{Q}_n\left(\hat{\theta}_n\right) - \hat{Q}_n\left(\theta^{\#}\right)\right| + \left|Q\left(\hat{\theta}_n\right) - Q\left(\theta^{\#}\right)\right| \\
&\leq \epsilon\left\|\hat{\theta}_n - \theta^{\#}\right\|^2 + n^{-1} M_n^2 + \left\|P g\left(\cdot, \hat{\theta}_n\right)\right\|^2 \\
&\leq \epsilon\left\|\hat{\theta}_n - \theta^{\#}\right\|^2 + n^{-1} M_n^2 - 2\epsilon\left\|\hat{\theta}_n - \theta^{\#}\right\|^2 \\
\Longrightarrow \left\|\hat{\theta}_n - \theta^{\#}\right\| &\leq \epsilon^{-1/2} n^{-1/2} M_n^2 + o_p\left(n^{-1/2}\right) = O_p\left(n^{-1/2}\right)
\end{aligned}$$

[1] The main revisions to Lemma 4.1 of Kim and Pollard (1990) are redefining $A(n, j) = (j-1) n^{-\gamma} \leq |\theta| \leq j n^{-\gamma}$, bounding the jth summand in $P(M_n > m)$ by $n^{4\gamma} P \sup_{|\theta| < j n^{-\gamma}} |P_n g(\cdot, \theta) - P g(\cdot, \theta)|^2 \big/ \left[\eta(j-1)^2 + m^2\right]^2$, where the numerator is further bounded by $n^{4\gamma}\left(n^{-1} C' j n^{-\gamma(2\rho)}\right) = C' j$.

Therefore, $\hat{\theta}_n - \theta^\# = O_p\left(n^{-1/2}\right)$.

Next, $\hat{h} = n^{1/2}\left(\hat{\theta}_n - \theta^\#\right) = \arg\min_h n\hat{Q}_n\left(\theta^\# + n^{-1/2}h\right)$. Note that $\theta^\#$ being in the interior of Θ ensures that $\theta^\# + n^{-1/2}h$ will belong in Θ for n large enough. It will follow from the argmin continuous mapping theorem that $\hat{h} \rightsquigarrow \underset{h\in\mathbb{R}^d}{\arg\min}\left\{\pi\left(\theta^\#\right)' W\mathcal{Z}_{0,1}(h) + h'G'WU_0 + \frac{1}{2}h'\bar{H}h\right\}$ if we can show that

$$n\left(\hat{Q}_n\left(\theta^\# + n^{-1/2}h\right) - \hat{Q}_n\left(\theta^\#\right)\right) \rightsquigarrow \pi\left(\theta^\#\right)' W\mathcal{Z}_{0,1}(h) + h'G'WU_0 + \frac{1}{2}h'\bar{H}h$$

as a process indexed by h in the space of locally bounded functions $\mathbf{B}_{\text{loc}}\left(\mathbb{R}^d\right)$ equipped with the topology of uniform convergence on compacta. Since $Q_1\left(\theta^\# + n^{-1/2}h\right) = Q_1\left(\theta^\#\right) + n^{-1/2}h'\frac{\partial Q_1\left(\theta^\#\right)}{\partial\theta} + \frac{1}{2}n^{-1}h'\frac{\partial^2 Q_1\left(\theta^\#\right)}{\partial\theta\partial\theta'}h + o\left(n^{-1}\right)$, $nQ_1\left(\theta^\# + n^{-1/2}h\right) = \frac{1}{2}h'\bar{H}h + o(1)$.

It remains to show that $n\left(\hat{Q}_2 + \hat{Q}_3\right)\left(\theta^\# + n^{-1/2}h\right) \rightsquigarrow \pi\left(\theta^\#\right)' W\mathcal{Z}_{0,1}(h) + h'G'WU_0$. Since the 1st, 3rd and 4th terms in $n\hat{Q}_2\left(\theta^\# + n^{-1/2}h\right)$ are all of the form $o_P(1)\, n\hat{\eta}_n\left(\theta^\# + n^{-1/2}h\right)$, they all converge in probability to 0. For the 2nd term, we can Taylor expand $g\left(\theta^\# + n^{-1/2}h\right)$ around $\theta^\#$:

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

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

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

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

Since $h'G'W\sqrt{n}\left(\hat{\pi}\left(\theta^\#\right) - \pi\left(\theta^\#\right)\right)$ 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, and for every $j = 1, \ldots, m$,

$$nE\sup_{\mathscr{D}(n)}\left|P_n d_j - Pd_j\right| = o(1) \tag{7.2}$$

where $\mathscr{D}(n) = \{d_j\left(\cdot,\theta^\#, h_1, h_2\right) = g_j\left(\cdot,\theta^\# + n^{-1/2}h_1\right) - g_j\left(\cdot,\theta^\# + n^{-1/2}h_2\right)$ such that $\max(\|h_1\|,\|h_2\|) \le M$ and $\|h_1 - h_2\| \le \delta_n\}$. Note that $\mathscr{D}(n)$ 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 $nP_nD_n^2 \le \eta$ for each $\eta > 0$, and applying the Cauchy-Schwarz inequality,

$$
\begin{aligned}
nE\sup_{\mathscr{D}(n)}|P_n d_j - P d_j| \le & E\sqrt{nP_n D_n^2}J\left(\frac{n\sup_{\mathscr{D}(n)}P_n d_j^2}{nP_n D_n^2}\right) \\
\le & \sqrt{\eta}J(1)+\sqrt{EnP_nD_n^2}\sqrt{EJ^2\left(\min\left(1,\frac{1}{\eta}n\sup_{\mathscr{D}(n)}P_nd_j^2\right)\right)}.
\end{aligned}
$$

To show that this is $o(1)$ for each fixed $\eta > 0$, first, note that by Assumption 2(vi), $EnP_nD_n^2 = 4nEG_{R(n)}^2 = O\left(nR(n)^2\right) = O(1)$ since $R(n) = Mn^{-1/2}$. The proof will then be complete if $n\sup_{\mathscr{D}(n)}P_nd_j^2 = o_p(1)$.

For each $K > 0$ write $E\sup_{\mathscr{D}(n)}P_nd_j^2 \le EP_n\sup_{\mathscr{D}(n)}d_j^2 1\{D_n > K\} + KE\sup_{\mathscr{D}(n)}P_n|d_j| \le EP_nD_n^2 1\{D_n > K\} + K\sup_{\mathscr{D}(n)}P|d_j| + KE\sup_{\mathscr{D}(n)}|P_n|d_j| - P|d_j||$. By Assumption 2(vii), $EP_nD_n^2 1\{D_n > K\} < \eta n^{-1}$ for large enough K. By Assumption 2(viii) and the definition of $\mathscr{D}(n)$, $K\sup_{\mathscr{D}(n)}P|d_j| = O(n^{-1}\delta_n) = o(n^{-1})$. Under the assumption that $g(\cdot,\theta)$ is Lipschitz in θ, so that $D_n = O_p\left(n^{-1/2}\delta_n\right)$, use the maximal inequality in Section 3.1 of Kim and Pollard (1990) to show $KE\sup_{\mathscr{D}(n)}|P_n|d_j| - P|d_j|| < Kn^{-\frac{1}{2}}J(1)\sqrt{PD_n^2} = O(n^{-1}\delta_n) = o(n^{-1})$. Therefore, $En\sup_{\mathscr{D}(n)}P_nd_j^2 = o(1)$. It follows that

$$
\begin{pmatrix} \pi\left(\theta^{\#}\right)' W n\hat{\eta}_n\left(\theta^{\#}+n^{-1/2}h\right) \\ h'G'W\sqrt{n}\left(\hat{\pi}\left(\theta^{\#}\right)-\pi\left(\theta^{\#}\right)\right)\end{pmatrix} \rightsquigarrow \begin{pmatrix} \pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}(h) \\ h'G'WU_0 \end{pmatrix}
$$

as a process indexed by h in the product space of locally bounded functions $\left\{\mathbf{B}_{\text{loc}}\left(\mathbb{R}^d\right)\right\}^2$ equipped with the topology of uniform convergence on compacta. By Slutsky's Theorem,

$$
n\left(\hat{Q}_1+\hat{Q}_2+\hat{Q}_3\right)\left(\theta^{\#}+n^{-1/2}h\right) \rightsquigarrow \pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}(h) + h'G'WU_0 + \frac{1}{2}h'\bar{H}h.
$$

Lemma 2.6 in Kim and Pollard (1990) implies that the Gaussian process $\pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}(h)$ has a unique minimum. In combination with the fact that $h'G'WU_0+\frac{1}{2}h'\bar{H}h$ is a convex function of h, there is a unique $\hat{h} = n^{1/2}\left(\hat{\theta}_n - \theta^{\#}\right)$ that minimizes $\pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}(h) + h'G'WU_0 + \frac{1}{2}h'\bar{H}h$. The result follows from the argmin continuous mapping theorem (Theorem 2.7 in Kim and Pollard (1990)). ■

Proof for Theorem 7

The consistency argument is the same as in Theorem 3. Next, write $\hat{Q}_n(\theta) - \hat{Q}_n\left(\theta^{\#}\right) = Q_1(\theta) + \hat{Q}_2(\theta) + \hat{Q}_3(\theta) + \hat{Q}_4(\theta) + \hat{Q}_5(\theta) + \hat{Q}_6(\theta)$, where

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

$$
\begin{aligned}
\hat{Q}_2(\theta) =& \frac{1}{2}\hat{\eta}_n(\theta)' W\hat{\eta}_n(\theta) + g(\theta)' W\left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right) + g(\theta)' W\hat{\eta}_n(\theta) + \left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)' W\hat{\eta}_n(\theta)\\
\hat{Q}_4(\theta) =& \frac{1}{2}g(\theta)'(W_n - W)g(\theta) + g(\theta)'(W_n - W)\pi\left(\theta^{\#}\right)\\
\hat{Q}_5(\theta) =& g(\theta)'(W_n - W)\left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)\\
&+ g(\theta)'(W_n - W)\hat{\eta}_n(\theta) + \left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)'(W_n - W)\hat{\eta}_n(\theta)\\
\hat{Q}_6(\theta) =& \pi\left(\theta^{\#}\right)'(W_n - W)\hat{\eta}_n(\theta) + \frac{1}{2}\hat{\eta}_n(\theta)'(W_n - W)\hat{\eta}_n(\theta)
\end{aligned}
$$

We already showed in Theorem 5 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 $\hat{Q}_6(\theta) = o_P(1)\hat{\eta}_n(\theta)$ implies that $\forall\epsilon > 0$, $\exists M_{n,6} = O_P(1)$ such that

$$
|\hat{Q}_6(\theta)| \leq \epsilon\|\theta - \theta^{\#}\|^2 + n^{-1}M_{n,6}^2.
$$

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 $\forall\epsilon > 0$,

$$
|g(\theta)'(W_n - W)\left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)| = o_p\left(\frac{\|\theta - \theta^{\#}\|}{\sqrt{n}}\right) \leq \epsilon\|\theta - \theta^{\#}\|^2 + n^{-1}M_{n,52}^2.
$$

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 3 suggest that $\hat{\theta}_n - \theta^{\#} = O_p\left(n^{-1/2}\right)$.

We already showed in Theorem 5 that $nQ_1\left(\theta^{\#} + n^{-1/2}h\right) = \frac{1}{2}h'\bar{H}h + o(1)$, and $n\hat{Q}_2\left(\theta^{\#} + n^{-1/2}h\right) + n\hat{Q}_3\left(\theta^{\#} + n^{-1/2}h\right) \rightsquigarrow \pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}(h) + h'G'W'U_0$. Furthermore, if $W_n - W = O_p\left(n^{-1/2}\right)$,

$$
\begin{aligned}
n\hat{Q}_4\left(\theta^{\#} + n^{-1/2}h\right) &= \frac{1}{2}ng\left(\theta^{\#} + n^{-1/2}h\right)'(W_n - W)g\left(\theta^{\#} + n^{-1/2}h\right)\\
&+ ng\left(\theta^{\#} + n^{-1/2}h\right)'(W_n - W)\pi\left(\theta^{\#}\right)\\
&= nO_p\left(\frac{\left\|n^{-1/2}h\right\|^2}{n^{1/2}}\right) + \left(n^{1/2}\left\{g\left(\theta^{\#}\right)' + h'G'n^{-1/2}\right\} + o_p(1)\right)n^{1/2}(W_n - W)\pi\left(\theta^{\#}\right)\\
&= h'G'n^{1/2}(W_n - W)\pi\left(\theta^{\#}\right) + o_p(1)\\
&\rightsquigarrow h'G'\Phi_0\pi\left(\theta^{\#}\right)\\
n\hat{Q}_5\left(\theta^{\#} + n^{-1/2}h\right) &= ng\left(\theta^{\#} + n^{-1/2}h\right)'(W_n - W)\left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)\\
&+ g\left(\theta^{\#} + n^{-1/2}h\right)'(W_n - W)n\hat{\eta}_n\left(\theta^{\#} + n^{-1/2}h\right)\\
&+ \left(\hat{\pi}\left(\theta^{\#}\right) - \pi\left(\theta^{\#}\right)\right)'(W_n - W)n\hat{\eta}_n\left(\theta^{\#} + n^{-1/2}h\right)\\
&= nO_p\left(\frac{\left\|n^{-1/2}h\right\|}{n}\right) + O_p\left(\frac{\left\|n^{-1/2}h\right\|}{n^{1/2}}\right)O_p(1) + O_p\left(n^{-1}\right)O_p(1)\\
&= O_p\left(n^{-1/2}\right) + O_p\left(n^{-1}\right) + O_p\left(n^{-1}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= o_p(1) \\
n\hat{Q}_6\left(\theta^{\#} + n^{-1/2}h\right) &= \pi\left(\theta^{\#}\right)'\left(W_n - W\right) n\hat{\eta}_n\left(\theta^{\#} + n^{-1/2}h\right) \\
&+ \frac{1}{2}\hat{\eta}_n\left(\theta^{\#} + n^{-1/2}h\right)'\left(W_n - W\right) n\hat{\eta}_n\left(\theta^{\#} + n^{-1/2}h\right) \\
&= O_p\left(n^{-1/2}\right) O_p\left(1\right) + O_p\left(n^{-1}\right) O_p\left(n^{-1/2}\right) O_p\left(1\right) \\
&= o_p(1)
\end{aligned}
$$

The joint Lindeberg condition is satisfied by Assumption 4: for each $\epsilon > 0$ and $t \in \mathbb{R}^d$,

$$
\lim_{n\to\infty} P \left\| \begin{pmatrix} \sqrt{n} g\left(\cdot, \theta^{\#} + \frac{t}{\sqrt{n}}\right) \\ \pi\left(\cdot, \theta^{\#}\right) \\ \text{vech}\left(\phi\left(\cdot, \theta_1^{\#}\right)\right) \end{pmatrix} \right\|^2 1\left\{ \left\| \begin{pmatrix} \sqrt{n} g\left(\cdot, \theta^{\#} + \frac{t}{\sqrt{n}}\right) \\ \pi\left(\cdot, \theta^{\#}\right) \\ \text{vech}\left(\phi\left(\cdot, \theta_1^{\#}\right)\right) \end{pmatrix} \right\| > \epsilon\sqrt{n} \right\} = 0
$$

Therefore, by the Lindeberg-Feller CLT and stochastic equicontinuity arguments similar to those in Theorem 5,

$$
\begin{pmatrix} \pi\left(\theta^{\#}\right)' W_n n\left(P_n - P\right) g\left(\cdot, \theta^{\#} + n^{-1/2}h\right) \\ h'G'W_n\sqrt{n}\left(P_n - P\right)\pi\left(\cdot, \theta^{\#}\right) \\ h'G'n^{1/2}\sqrt{n}\left(P_n - P\right)\phi\left(\cdot, \theta_1^{\#}\right)\pi\left(\theta^{\#}\right) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}\left(h\right) \\ h'G'WU_0 \\ h'G'\Phi_0\pi\left(\theta^{\#}\right) \end{pmatrix}
$$

as a process indexed by h in the product space of locally bounded functions $\left\{\mathbf{B}_{\text{loc}}\left(\mathbb{R}^d\right)\right\}^3$ equipped with the topology of uniform convergence on compacta. By Assumption 4, $\sqrt{n}\left(W_n - W\right) = \sqrt{n}\left(P_n - P\right)\phi\left(\cdot, \theta_1^{\#}\right) + o_p(1)$; therefore,

$$
\begin{pmatrix} \pi\left(\theta^{\#}\right)' W_n n\left(P_n - P\right) g\left(\cdot, \theta^{\#} + n^{-1/2}h\right) \\ h'G'W_n\sqrt{n}\left(P_n - P\right)\pi\left(\cdot, \theta^{\#}\right) \\ h'G'n^{1/2}\sqrt{n}\left(W_n - W\right)\pi\left(\theta^{\#}\right) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}\left(h\right) \\ h'G'WU_0 \\ h'G'\Phi_0\pi\left(\theta^{\#}\right) \end{pmatrix}
$$

By Slutsky's theorem and the argmin continuous mapping theorem,

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

If $W_n - W = o_p\left(n^{-1/2}\right)$,

$$
\begin{aligned}
n\hat{Q}_4\left(\theta^{\#} + n^{-1/2}h\right) &= no_p\left(\frac{\left\|n^{-1/2}h\right\|^2}{\sqrt{n}}\right) + no_p\left(\frac{\left\|n^{-1/2}h\right\|}{n^{1/2}}\right) \\
&= o_p\left(\frac{1}{\sqrt{n}}\right) + o_p(1) = o_p(1)
\end{aligned}
$$

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

7.3 More Details for Examples

7.3.1 IV Quantile Regression

We show that the classes $\mathcal{G}_R \equiv \left\{\pi_j\left(\cdot,\theta\right)-\pi_j\left(\cdot,\theta^{\#}\right):\left\|\theta-\theta^{\#}\right\|\leq R, j=1,\ldots,m\right\}$ have envelope functions which decay at the linear rate:

$$\begin{aligned}G_R\left(\cdot\right) &= \sup_{\left\|\theta-\theta^{\#}\right\|\leq R}\left|\pi_j\left(\cdot,\theta\right)-\pi_j\left(\cdot,\theta^{\#}\right)\right| \\ &= \sup_{\left\|\theta-\theta^{\#}\right\|\leq R}\left|z_{ij}\left(1\left(y_i\leq q\left(x_i'\theta^{\#}\right)\right)-1\left(y_i\leq q\left(x_i'\theta\right)\right)\right)\right|\end{aligned}$$

$$\begin{aligned}PG_R^2 &\leq E\left[\left|z_{ij}\right|^2 E\left[\sup_{\left\|\theta-\theta^{\#}\right\|\leq R}\left|1\left(y_i\leq q\left(x_i'\theta^{\#}\right)\right)-1\left(y_i\leq q\left(x_i'\theta\right)\right)\right|\middle| x_i,z_i\right]\right] \\ &\leq E\left[\left|z_{ij}\right|^2\sum_{\omega\in\{-1,1\}^d}\left(P\left(q\left(x_i'\left(\theta^{\#}-\omega R\right)\right)\leq y_i\leq q\left(x_i'\theta^{\#}\right)\middle| x_i,z_i\right)\right)\right] \\ &+ E\left[\left|z_{ij}\right|^2\sum_{\omega\in\{-1,1\}^d}P\left(q\left(x_i'\theta^{\#}\right)\leq y_i\leq q\left(x_i'\left(\theta^{\#}+\omega R\right)\right)\middle| x_i,z_i\right)\right] \\ &\leq E\left[\left|z_{ij}\right|^2\sup_{\theta\in\Theta}\sum_{\omega\in\{-1,1\}^d}2f_{y|x,z}\left(q\left(x_i'\theta\right)\right)q'\left(x_i'\theta\right)x_i'\omega R\right]=O(R)\end{aligned}$$

For the second inequality, we applied an upper bound to the expectation of supremum by considering all possible ways of adding or subtracting R from each coordinate of $\theta^{\#}$. For the third inequality, we applied mean-value expansions to the probabilities followed by a sup bound.

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 $\hat{\theta}_n=\arg\min_\theta\hat{\pi}_n\left(\theta\right)'W_n\left(\hat{\theta}_1\right)\hat{\pi}_n\left(\theta\right)$ depends on the 1-step GMM estimator $\hat{\theta}_1=\arg\min_\theta\hat{\pi}_n\left(\theta\right)'W_1\hat{\pi}_n\left(\theta\right)$ whose probability limit is $\theta_1^{\#}=\arg\min_\theta\pi\left(\theta\right)'W_1\pi\left(\theta\right)$. The pseudo-true parameters are given by $\theta^{\#}=\arg\min_\theta\pi\left(\theta\right)'W\left(\theta_1^{\#}\right)\pi\left(\theta\right)$, where $W\left(\theta_1^{\#}\right)$ is the inverse of the variance-covariance matrix of the population moments

$$\begin{aligned}W\left(\theta_1^{\#}\right) &= \left(E\left[\pi\left(\cdot,\theta_1^{\#}\right)\pi\left(\cdot,\theta_1^{\#}\right)'\right]-\pi\left(\theta_1^{\#}\right)\pi\left(\theta_1^{\#}\right)'\right)^{-1} \\ &= \left(E\left[\left(\tau-1\left(y_i\leq q\left(x_i'\theta_1^{\#}\right)\right)\right)^2 z_iz_i'\right]-\pi\left(\theta_1^{\#}\right)\pi\left(\theta_1^{\#}\right)'\right)^{-1} \\ &= \left(E\left[E\left[\left(\tau-1\left(y_i\leq q\left(x_i'\theta_1^{\#}\right)\right)\right)^2\middle| x_i,z_i\right]z_iz_i'\right]-\pi\left(\theta_1^{\#}\right)\pi\left(\theta_1^{\#}\right)'\right)^{-1}\end{aligned}$$

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

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

$$\begin{aligned} E\left[\left(\tau - 1\left(y_i \leq q\left(x_i'\theta_1^{\#} \right) \right) \right)^2 \middle| z_i \right] &= (\tau-1)^2 F_{y|x,z}\left(q\left(x_i'\theta_1^{\#} \right) \right) + \tau^2 \left(1 - F_{y|x,z}\left(q\left(x_i'\theta_1^{\#} \right) \right) \right) \\ &= \tau^2 + (1-2\tau) F_{y|x,z}\left(q\left(x_i'\theta_1^{\#} \right) \right) \end{aligned}$$

Note that in the case of correct specification, $W\left(\theta_1^{\#} \right)$ reduces down to $\left(\tau(1-\tau) E\left[z_i z_i' \right] \right)^{-1}$ since $F_{y|x,z}\left(q\left(x_i'\theta_1^{\#} \right) \right) = \tau$.

The estimated weighting matrix is

$$\begin{aligned} W_n\left(\hat{\theta}_1 \right) &= \left(\frac{1}{n} \sum_{i=1}^n \left(\tau^2 + (1-2\tau) \hat{F}_{y|x,z}\left(q\left(x_i'\hat{\theta}_1 \right) \right) \right) z_i z_i' - \hat{\pi}_n\left(\hat{\theta}_1 \right) \hat{\pi}_n\left(\hat{\theta}_1 \right)' \right)^{-1} \\ &= \left(\tau^2 \frac{1}{n} \sum_{i=1}^n z_i z_i' + (1-2\tau) \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n 1\left(y_j \leq q\left(x_i'\hat{\theta}_1 \right) \right) z_i z_i' - \hat{\pi}_n\left(\hat{\theta}_1 \right) \hat{\pi}_n\left(\hat{\theta}_1 \right)' \right)^{-1} \end{aligned}$$

$$\hat{\pi}_n\left(\hat{\theta}_1 \right) = \frac{1}{n} \sum_{i=1}^n \left(\tau - 1\left(y_i \leq q\left(x_i'\hat{\theta}_1 \right) \right) \right) z_i$$

Suppose there exists D_n such that the following mean value expansion around $\theta_1^{\#}$ holds:

$$\left(W_n\left(\hat{\theta}_1 \right) - W\left(\theta_1^{\#} \right) \right) \pi\left(\theta^{\#} \right) = \left(W_n\left(\theta_1^{\#} \right) - W\left(\theta_1^{\#} \right) \right) \pi\left(\theta^{\#} \right) + D_n'\left(\hat{\theta}_1 - \theta_1^{\#} \right) + o_p(1)$$

D_n can be interpreted as an approximate derivative (with respect to θ_1) of $W_n(\theta_1)\pi\left(\theta^{\#} \right)$ evaluated at $\theta_1^{\#}$ that is consistent for the population derivative: $D_n \xrightarrow{p} D_0 \equiv \left. \frac{\partial W(\theta_1)\pi\left(\theta^{\#} \right)}{\partial \theta_1} \right|_{\theta_1 = \theta_1^{\#}}$.

If the 1-step GMM estimator $\hat{\theta}_1$ has an influence function representation, then there is also an influence function representation for $W_n\left(\hat{\theta}_1 \right)$. However, $\hat{\theta}_1$ has an influence function representation only in the case of correct specification in which case we can use the simpler estimated weighting matrix $W_n = \left(\tau(1-\tau) \frac{1}{n} \sum_i z_i z_i' \right)^{-1}$ as in Chernozhukov and Hansen (2005). In the case of misspecification so that $n^{1/3}\left(\hat{\theta}_1 - \theta_1^{\#} \right) \rightsquigarrow \mathcal{J}$, the estimated weighting matrix is cubic-root consistent because of the dominant effect of $\hat{\theta}_1$.

$$\begin{aligned} & n^{1/3}\left(W_n\left(\hat{\theta}_1 \right) - W\left(\theta_1^{\#} \right) \right) \pi\left(\theta^{\#} \right) \\ &= \frac{n^{1/3}}{\sqrt{n}} \underbrace{\sqrt{n}\left(W_n\left(\theta_1^{\#} \right) - W\left(\theta_1^{\#} \right) \right) \pi\left(\theta^{\#} \right)}_{O_p(1)} + D_n' n^{1/3}\left(\hat{\theta}_1 - \theta_1^{\#} \right) + o_p(1) \end{aligned}$$

$$
\begin{aligned}
&= D_n' n^{1/3}\left(\hat{\theta}_1 - \theta_1^{\#}\right) + o_p(1) \\
&\rightsquigarrow D_0'\mathcal{J} \equiv \mathcal{W}_0
\end{aligned}
$$

7.3.2 Simulated Method of Moments

The classes $\mathcal{G}_R \equiv \left\{\pi_j\left(\cdot,\theta\right) - \pi_j\left(\cdot,\theta^{\#}\right) : \left\|\theta-\theta^{\#}\right\| \le R, j = 1,\ldots,m\right\}$ have envelope functions

$$
\begin{aligned}
G_R\left(\cdot\right) &= \sup_{\left\|\theta-\theta^{\#}\right\|\le R}\left|\pi_j\left(\cdot,\theta\right) - \pi_j\left(\cdot,\theta^{\#}\right)\right| \\
&= \sup_{\left\|\theta-\theta^{\#}\right\|\le R}\left|z_{ij}\frac{1}{S}\sum_{s=1}^{S}\left(1\left(h\left(x_i'\theta^{\#}\right)+\eta_{is}>0\right) - 1\left(h\left(x_i'\theta\right)+\eta_{is}>0\right)\right)\right|
\end{aligned}
$$

Using similar arguments as in the previous example,

$$
\begin{aligned}
PG_R^2 &\le E\left[\left|z_{ij}\right|^2 E\left[\sup_{\left\|\theta-\theta^{\#}\right\|\le R}\frac{1}{S}\sum_{s=1}^{S}\left|1\left(h\left(x_i'\theta^{\#}\right)+\eta_{is}>0\right) - 1\left(h\left(x_i'\theta\right)+\eta_{is}>0\right)\right| x_i, z_i\right]\right] \\
&\le E\left[\left|z_{ij}\right|^2\frac{1}{S}\sum_{s=1}^{S}\sum_{\omega_s\in\{-1,1\}^d} P\left(h\left(x_i'\left(\theta^{\#}-\omega_s R\right)\right) \le -\eta_{is} \le h\left(x_i'\theta^{\#}\right)\Big| x_i, z_i\right)\right] \\
&+ E\left[\left|z_{ij}\right|^2\frac{1}{S}\sum_{s=1}^{S}\sum_{\omega_s\in\{-1,1\}^d} P\left(h\left(x_i'\theta^{\#}\right) \le -\eta_{is} \le h\left(x_i'\left(\theta^{\#}+\omega_s R\right)\right)\Big| x_i, z_i\right)\right] \\
&\le E\left[\left|z_{ij}\right|^2\frac{1}{S}\sum_{s=1}^{S}\sup_{\theta\in\Theta}\sum_{\omega_s\in\{-1,1\}^d} 2f_{\eta|x,z}\left(h\left(x_i'\theta\right)\right)h'\left(x_i'\theta\right)x_i'\omega_s R\right] = O(R)
\end{aligned}
$$

Just as in the previous example, the pseudo-true parameters are given by $\theta^{\#} = \arg\min_{\theta}\pi\left(\theta\right)' W\left(\theta_1^{\#}\right)\pi\left(\theta\right)$, where $W\left(\theta_1^{\#}\right)$ is the inverse of the variance-covariance matrix of the population moments:

$$
\begin{aligned}
W\left(\theta_1^{\#}\right) &= \left(E\left[\pi\left(\cdot,\theta_1^{\#}\right)\pi\left(\cdot,\theta_1^{\#}\right)'\right] - \pi\left(\theta_1^{\#}\right)\pi\left(\theta_1^{\#}\right)'\right)^{-1} \\
&= \left(E\left[\left(y_i - \frac{1}{S}\sum_{s=1}^{S}1\left(h\left(x_i'\theta_1^{\#}\right)+\eta_{is}>0\right)\right)^2 z_i z_i'\right] - \pi\left(\theta_1^{\#}\right)\pi\left(\theta_1^{\#}\right)'\right)^{-1}
\end{aligned}
$$

The estimated weighting matrix is

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

Suppose there exists D_n such that the following mean value expansion around $\theta_1^{\#}$ holds:

$$\left(W_n\left(\hat{\theta}_1\right) - W\left(\theta_1^{\#}\right)\right)\pi\left(\theta^{\#}\right) = \left(W_n\left(\theta_1^{\#}\right) - W\left(\theta_1^{\#}\right)\right)\pi\left(\theta^{\#}\right) + D_n'\left(\hat{\theta}_1 - \theta_1^{\#}\right) + o_p(1)$$

D_n can be interpreted as an approximate derivative (with respect to θ_1) of $W_n(\theta_1)\pi\left(\theta^{\#}\right)$ evaluated at $\theta_1^{\#}$ that is consistent for the population derivative: $D_n \xrightarrow{p} D_0 \equiv \left.\frac{\partial W(\theta_1)\pi\left(\theta^{\#}\right)}{\partial\theta_1}\right|_{\theta_1=\theta_1^{\#}}$. If the 1-step GMM estimator $\hat{\theta}_1$ has an influence function representation, then there is also an influence function representation for $W_n\left(\hat{\theta}_1\right)$. However, $\hat{\theta}_1$ has an influence function representation only in the case of correct specification in which case we can use $W_n\left(\hat{\theta}_1\right) = \left(\frac{1}{n}\sum_{i=1}^n\left(y_i - \frac{1}{S}\sum_{s=1}^S 1\left(h\left(x_i'\hat{\theta}_1\right) + \eta_{is} > 0\right)\right)^2 z_i z_i'\right)^{-1}$. In the case of misspecification so that $n^{1/3}\left(\hat{\theta}_1 - \theta_1^{\#}\right) \rightsquigarrow \mathcal{J}$, the estimated weighting matrix is cubic-root consistent because of the dominant effect of $\hat{\theta}_1$.

$$\begin{aligned}
&n^{1/3}\left(W_n\left(\hat{\theta}_1\right) - W\left(\theta_1^{\#}\right)\right)\pi\left(\theta^{\#}\right)\\
&= \frac{n^{1/3}}{\sqrt{n}}\underbrace{\sqrt{n}\left(W_n\left(\theta_1^{\#}\right) - W\left(\theta_1^{\#}\right)\right)\pi\left(\theta^{\#}\right)}_{O_p(1)} + D_n' n^{1/3}\left(\hat{\theta}_1 - \theta_1^{\#}\right) + o_p(1)\\
&= D_n' n^{1/3}\left(\hat{\theta}_1 - \theta_1^{\#}\right) + o_p(1)\\
&\rightsquigarrow D_0'\mathcal{J} \equiv \mathcal{W}_0
\end{aligned}$$

7.3.3 Dynamic Censored Regression

The classes $\mathcal{G}_R \equiv \left\{\pi_j(\cdot,\theta) - \pi_j\left(\cdot,\theta^{\#}\right) : \left\|\theta - \theta^{\#}\right\| \le R, j = 1,\ldots,m\right\}$ have envelope functions

$$\begin{aligned}
G_R(\cdot) &= \sup_{\left\|\theta-\theta^{\#}\right\|\le R}\left|\pi_j(\cdot,\theta) - \pi_j\left(\cdot,\theta^{\#}\right)\right|\\
&= \sup_{\left\|\theta-\theta^{\#}\right\|\le R}\left|\max\{0, y_{it} - y_{it-1}\theta\} - \max\left\{0, y_{it} - y_{it-1}\theta^{\#}\right\}\right|
\end{aligned}$$

Because each moment condition $\pi_j(\cdot,\theta)$ is Lipschitz in θ, PG_R^2 will be $O\left(R^2\right)$ if $E\sup_{1\le t\le T}\left|y_{it}\right|^2 < \infty$ [2]. For $y_{i\cdot} \equiv \left[\begin{array}{ccc} y_{i2} & ,\ldots, & y_{iT}\end{array}\right]'$ and $y_{i\cdot-} \equiv \left[\begin{array}{ccc} y_{i1} & ,\ldots, & y_{iT-1}\end{array}\right]'$, $W\left(\theta_1^{\#}\right)$ is the inverse of the variance-covariance matrix of the population moments

$$\begin{aligned}
W\left(\theta_1^{\#}\right) &= \left(E\left[\pi\left(\cdot,\theta_1^{\#}\right)\pi\left(\cdot,\theta_1^{\#}\right)'\right] - \pi\left(\theta_1^{\#}\right)\pi\left(\theta_1^{\#}\right)'\right)^{-1}\\
&= \left(E\left[\left(\max\left\{0, y_{i\cdot} - y_{i\cdot-}\theta_1^{\#}\right\} - y_{i\cdot-}\right)\left(\max\left\{0, y_{i\cdot} - y_{i\cdot-}\theta_1^{\#}\right\} - y_{i\cdot-}\right)'\right] - \pi\left(\theta_1^{\#}\right)\pi\left(\theta_1^{\#}\right)'\right)^{-1}
\end{aligned}$$

[2] We thank an anonymous referee for pointing this out.

The estimated weighting matrix is

$$W_n\left(\hat{\theta}_1\right) = \left(\frac{1}{n}\sum_{i=1}^{n}\left(\max\left\{0, y_{i\cdot} - y_{i\cdot -}\hat{\theta}_1\right\} - y_{i\cdot -}\right)\left(\max\left\{0, y_{i\cdot} - y_{i\cdot -}\hat{\theta}_1\right\} - y_{i\cdot -}\right)' - \hat{\pi}_n\left(\hat{\theta}_1\right)\hat{\pi}_n\left(\hat{\theta}_1\right)'\right)^{-1}$$

$$\hat{\pi}_n\left(\hat{\theta}_1\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\max\left\{0, y_{i\cdot} - y_{i\cdot -}\hat{\theta}_1\right\} - y_{i\cdot -}\right)$$

If the 1-step GMM estimator $\hat{\theta}_1$ has an influence function representation $\sqrt{n}\left(\hat{\theta}_1 - \theta_1^{\#}\right) = \sqrt{n}\left(P_n - P\right)\kappa\left(\cdot, \theta_1^{\#}\right) + o_p(1)$, then there is also an influence function representation for the estimated weighting matrix, which we now derive. Suppose there exists Δ_n such that the following mean value expansion around $\theta_1^{\#}$ holds:

$$\begin{aligned}
&\sqrt{n}\; vech\left(W_n\left(\hat{\theta}_1\right) - W\left(\theta_1^{\#}\right)\right) \\
&= \sqrt{n}\; vech\left(W_n\left(\theta_1^{\#}\right) - W\left(\theta_1^{\#}\right)\right) + \Delta_n'\sqrt{n}\left(\hat{\theta}_1 - \theta_1^{\#}\right) + o_p(1) \\
&= -vech\left(W\left(\theta_1^{\#}\right)\sqrt{n}\left(P_n - P\right)\psi\left(\cdot, \theta_1^{\#}\right)W\left(\theta_1^{\#}\right)\right) + \Delta_0'\sqrt{n}\left(P_n - P\right)\kappa\left(\cdot, \theta_1^{\#}\right) + o_p(1) \\
&= vech\left(\sqrt{n}\left(P_n - P\right)\phi\left(\cdot, \theta_1^{\#}\right)\right) + o_p(1)
\end{aligned}$$

Δ_n can be interpreted as an approximate derivative of $vech\left(W_n\left(\theta_1\right)\right)$ evaluated at $\theta_1^{\#}$ that is consistent for the derivative of the population weighting matrix: $\Delta_n \xrightarrow{p} \Delta_0 \equiv \left.\frac{\partial vech(W(\theta_1))}{\partial\theta_1}\right|_{\theta_1 = \theta_1^{\#}}$.

We can obtain the expression for $\sqrt{n}\left(P_n - P\right)\psi\left(\cdot, \theta_1^{\#}\right)$ using U-statistic projection arguments:

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\pi\left(\cdot, \theta_1^{\#}\right)\pi\left(\cdot, \theta_1^{\#}\right)' - E\left[\pi\left(\cdot, \theta_1^{\#}\right)\pi\left(\cdot, \theta_1^{\#}\right)'\right]\right) - \left(\hat{\pi}_n\left(\theta_1^{\#}\right)\hat{\pi}_n\left(\theta_1^{\#}\right)' - \pi\left(\theta_1^{\#}\right)\pi\left(\theta_1^{\#}\right)'\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\underbrace{\left(\max\left\{0, y_{i\cdot} - y_{i\cdot -}\theta_1^{\#}\right\} - y_{i\cdot -}\right)\left(\max\left\{0, y_{i\cdot} - y_{i\cdot -}\theta_1^{\#}\right\} - y_{i\cdot -}\right)'}_{\delta(y_{i\cdot}, y_{i\cdot})} - E\left[\delta\left(y_{i\cdot}, y_{i\cdot}\right)\right]\right\} \\
&- \frac{\sqrt{n}}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\underbrace{\underbrace{\left(\max\left\{0, y_{i\cdot} - y_{i\cdot -}\theta_1^{\#}\right\} - y_{i\cdot -}\right)\left(\max\left\{0, y_{j\cdot} - y_{j\cdot -}\theta_1^{\#}\right\} - y_{j\cdot -}\right)'}_{\delta(y_{i\cdot}, y_{j\cdot})} - E\left[\delta\left(y_{i\cdot}, y_{j\cdot}\right)\right]}_{g(y_{i\cdot}, y_{j\cdot})}\right\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\delta\left(y_{i\cdot}, y_{i\cdot}\right) - E\left[\delta\left(y_{i\cdot}, y_{i\cdot}\right)\right]\right) \\
&- \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left[g\left(y_{i\cdot}, y_{j\cdot}\right)\middle| y_{i\cdot}\right] + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}E\left[g\left(y_{i\cdot}, y_{j\cdot}\right)\middle| y_{j\cdot}\right]\right) + o_p(1)
\end{aligned}$$

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

where the second to last equality follows from the fact that $\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} g\left(y_{i\cdot}, y_{j\cdot}\right)$ is a non-degenerate V-statistic which has the same asymptotic distribution as the non-degenerate U-statistic $\frac{2}{n(n-1)}\sum_{1\le i<j\le n} g\left(y_{i\cdot}, y_{j\cdot}\right)$ if $E\left[\|g\left(y_{i\cdot}, y_{i\cdot}\right)\|\right] < \infty$ and $E\left[\|g\left(y_{i\cdot}, y_{j\cdot}\right)\|^2\right] < \infty$. A discussion of this asymptotic equivalence result can be found in Section 8.2 of Newey and McFadden (1994), Section 6.4 of Serfling (1980), and Appendix A of Zhou et al. (2021).

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.

$$W_n^*\left(\hat{\theta}_1^*\right) = \left(\frac{1}{n}\sum_{i=1}^{n}\left(\max\left\{0, y_{i\cdot}^* - y_{i\cdot-}^*\hat{\theta}_1^*\right\} - y_{i\cdot-}^*\right)\left(\max\left\{0, y_{i\cdot}^* - y_{i\cdot-}^*\hat{\theta}_1^*\right\} - y_{i\cdot-}^*\right)' - \hat{\pi}_n\left(\hat{\theta}_1^*\right)\hat{\pi}_n\left(\hat{\theta}_1^*\right)'\right)^{-1}$$

$$\hat{\pi}_n\left(\hat{\theta}_1^*\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\max\left\{0, y_{i\cdot}^* - y_{i\cdot-}^*\hat{\theta}_1^*\right\} - y_{i\cdot-}^*\right)$$

We can show that when the bootstrapped 1-step GMM estimator $\hat{\theta}_1^*$ has the same influence function representation as $\hat{\theta}_1$, $\sqrt{n}\left(\hat{\theta}_1^* - \hat{\theta}_1\right) = \sqrt{n}\left(P_n^* - P_n\right)\kappa\left(\cdot, \theta_1^{\#}\right) + o_p^*(1)$, the bootstrapped weighting matrix $W_n^*\left(\hat{\theta}_1^*\right)$ has the same influence function representation as $W_n\left(\hat{\theta}_1\right)$. Suppose there exists Δ_n^* such that $W_n^*\left(\hat{\theta}_1^*\right)$ has a mean value expansion around $\theta_1^{\#}$. Δ_n^* can be interpreted as an approximate derivative of $vech\left(W_n^*\left(\theta_1\right)\right)$ evaluated at $\theta_1^{\#}$ that is consistent for the derivative of the population weighting matrix: $\Delta_n^* \xrightarrow{p} \Delta_0 \equiv \left.\frac{\partial vech(W(\theta_1))}{\partial\theta_1}\right|_{\theta_1=\theta_1^{\#}}$. Then since Δ_n is an approximate derivative of $vech\left(W_n\left(\theta_1\right)\right)$ evaluated at $\theta_1^{\#}$ that is also consistent for Δ_0, we can write

$$\begin{aligned}
&\sqrt{n}\ vech\left(W_n^*\left(\hat{\theta}_1^*\right) - W_n\left(\hat{\theta}_1\right)\right)\\
&= \sqrt{n}\ vech\left(W_n^*\left(\hat{\theta}_1^*\right) - W_n\left(\theta_1^{\#}\right)\right) - \sqrt{n}\ vech\left(W_n\left(\hat{\theta}_1\right) - W_n\left(\theta_1^{\#}\right)\right)\\
&= \sqrt{n}\ vech\left(W_n^*\left(\theta_1^{\#}\right) - W_n\left(\theta_1^{\#}\right)\right) + \left(\Delta_n^{*\prime}\sqrt{n}\left(\hat{\theta}_1^* - \theta_1^{\#}\right) - \Delta_n'\sqrt{n}\left(\hat{\theta}_1^* - \theta_1^{\#}\right)\right) + o_p^*(1)\\
&= \sqrt{n}\ vech\left(W_n^*\left(\theta_1^{\#}\right) - W_n\left(\theta_1^{\#}\right)\right) + \Delta_0'\sqrt{n}\left(\hat{\theta}_1^* - \hat{\theta}_1\right) + o_p^*(1)\\
&= -vech\left(W\left(\theta_1^{\#}\right)\sqrt{n}\left(P_n^* - P_n\right)\psi\left(\cdot, \theta_1^{\#}\right)W\left(\theta_1^{\#}\right)\right) + \Delta_0'\sqrt{n}\left(P_n^* - P_n\right)\kappa\left(\cdot, \theta_1^{\#}\right) + o_p(1)\\
&= vech\left(\sqrt{n}\left(P_n^* - P_n\right)\phi\left(\cdot, \theta_1^{\#}\right)\right) + o_p^*(1)
\end{aligned}$$

where $\sqrt{n}\left(W_n^*\left(\theta_1^{\#}\right) - W_n\left(\theta_1^{\#}\right)\right) = -W\left(\theta_1^{\#}\right)\sqrt{n}\left(P_n^* - P_n\right)\psi\left(\cdot, \theta_1^{\#}\right)W\left(\theta_1^{\#}\right) + o_p(1)$ follows from the consistency of the multinomial bootstrap for V-statistics of order 2 (see Theorem 3.1 in Bickel and Freedman (1981)) if $\int g\left(y_{1\cdot}, y_{2\cdot}\right)g\left(y_{1\cdot}, y_{2\cdot}\right)' dF\left(y_{1\cdot}\right)dF\left(y_{2\cdot}\right) < \infty$ and $\int g\left(y_{\cdot}, y_{\cdot}\right)g\left(y_{\cdot}, y_{\cdot}\right)' dF\left(y_{\cdot}\right) < \infty$.

7.4 Monte Carlo Simulation for Smooth Misspecified GMM

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 θ. 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 $W_n = \begin{pmatrix} S_y^2 & S_{yz} \\ S_{yz} & S_z^2 \end{pmatrix}^{-1} = \frac{1}{S_y^2 S_z^2 - S_{yz}^2}\begin{pmatrix} S_z^2 & -S_{yz} \\ -S_{yz} & S_y^2 \end{pmatrix}$ is $\hat{\theta}_2 = \bar{z} - \frac{S_{yz}}{S_y^2}\bar{y}$. We would like to compare the performance of our rate-adaptive bootstrap to the standard bootstrap estimators $\tilde{\theta}_1^* = \bar{z}^*$ and $\tilde{\theta}_2^* = \bar{z}^* - \frac{S_{yz^*}}{S_{y^*}^2}\bar{y}^*$.

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 $(P_n^* - P_n)\left(\pi(\cdot,\theta) - \pi\left(\cdot,\hat{\theta}_n\right)\right) = 0$, $H = 0$, $G = [0; -1]$, $G'G = 1$, and therefore

$$\begin{aligned}
\hat{\theta}_1^* &= \arg\min_{\theta\in\Theta}\left\{\frac{1}{2}\left(\theta - \hat{\theta}_1\right)^2 + \left(\theta - \hat{\theta}_1\right)\hat{G}'\left(P_n^* - P_n\right)\pi\left(\cdot,\hat{\theta}_1\right)\right\} \\
&= \arg\min_{\theta\in\Theta}\left\{\frac{1}{2}\left(\theta - \bar{z}\right)^2 - \left(\theta - \bar{z}\right)\left(\bar{z}^* - \bar{z}\right)\right\} \\
&= \bar{z}^*
\end{aligned}$$

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

$$\begin{aligned}
\hat{\theta}_2^* &= \arg\min_{\theta\in\Theta}\left\{\frac{1}{2}\left(\theta - \hat{\theta}_2\right)^2\hat{G}'W_n\hat{G} + \left(\theta - \hat{\theta}_2\right)\hat{G}'W_n\left(P_n^* - P_n\right)\pi\left(\cdot,\hat{\theta}_2\right)\right. \\
&\qquad\left. + \left(\theta - \hat{\theta}_2\right)\hat{G}'\left(W_n^* - W_n\right)\hat{\pi}\left(\hat{\theta}_2\right)\right\} \\
&= \arg\min_{\theta\in\Theta}\left\{\frac{1}{2}\left(\theta - \hat{\theta}_2\right)^2\frac{S_y^2}{S_y^2S_z^2 - S_{yz}^2} + \left(\theta - \hat{\theta}_2\right)\frac{\left(S_{yz}\left(\bar{y}^* - \bar{y}\right) - S_y^2\left(\bar{z}^* - \bar{z}\right)\right)}{S_y^2S_z^2 - S_{yz}^2}\right. \\
&\qquad\left. + \left(\theta - \hat{\theta}_2\right)\frac{\left(S_{yz^*} - S_{y^*}^2\frac{S_{yz}}{S_y^2}\right)\bar{y}}{S_{y^*}^2S_{z^*}^2 - S_{yz^*}^2}\right\} \\
\implies \hat{\theta}_2^* &= \bar{z}^* - \frac{S_{yz}\bar{y}^*}{S_y^2} - \frac{S_y^2S_z^2 - S_{yz}^2}{S_{y^*}^2S_{z^*}^2 - S_{yz^*}^2}\left(\frac{S_{yz^*}}{S_y^2} - \frac{S_{y^*}^2}{S_y^2}\frac{S_{yz}}{S_y^2}\right)\bar{y}
\end{aligned}$$

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

From Tables 7 and 8 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. The coverage frequencies of three methods are very similar because in the smooth case, the asymptotic distribution remains normal so the standard bootstrap will be consistent. Results for symmetric confidence intervals are very similar and available upon request.

Table 7: 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 8: 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)

References

Armstrong, Timothy B and Michal Kolesár, "Sensitivity analysis using approximate moment condition models," *Quantitative Economics*, 2021, *12* (1), 77–108. 2

Berkowitz, Daniel, Mehmet Caner, and Ying Fang, "The validity of instruments revisited," *Journal of Econometrics*, 2012, *166* (2), 255–266. 2

Bickel, Peter J and David A Freedman, "Some asymptotic theory for the bootstrap," *The Annals of Statistics*, 1981, pp. 1196–1217. 43

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

Brown, Bryan W and Whitney K Newey, "Generalized method of moments, efficient bootstrapping, and improved inference," *Journal of Business & Economic Statistics*, 2002, *20* (4), 507–517. 2, 3

Cattaneo, Matias D and Michael Jansson, "Kernel-Based Semiparametric Estimators: Small Bandwidth Asymptotics and Bootstrap Consistency," *Econometrica*, 2018, *86* (3), 955–995. 3

_ and _ , "Average density estimators: Efficiency and bootstrap consistency," *Econometric Theory*, 2021, pp. 1–35. 3

_ , _ , and Kenichi Nagasawa, "Bootstrap-Based Inference for Cube Root Asymptotics," *Econometrica*, 2020, *88* (5), 2203–2219. 2, 6, 9

Cheng, Xu, Zhipeng Liao, and Ruoyao Shi, "On uniform asymptotic risk of averaging GMM estimators," *Quantitative Economics*, 2019, *10* (3), 931–979. 2

Chernozhukov, Victor and Christian Hansen, "An IV model of quantile treatment effects," *Econometrica*, 2005, *73* (1), 245–261. 1, 3, 8, 16, 39

_ and Han Hong, "A MCMC Approach to Classical Estimation," *Journal of Econometrics*, 2003, *115* (2), 293–346. 16

Creel, Michael, Jiti Gao, Han Hong, and Dennis Kristensen, "Bayesian Indirect Inference and the ABC of GMM," *arXiv preprint arXiv:1512.07385*, 2015. 2

der Vaart, Aad W Van, *Asymptotic statistics*, Vol. 3, Cambridge university press, 2000. 23, 28

Giurcanu, Mihai and Brett Presnell, "Bootstrap inference for misspecified moment condition models," *Annals of the Institute of Statistical Mathematics*, 2018, *70* (3), 605–630. 2, 3

Guggenberger, Patrik, "On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption," *Econometric Theory*, 2012, *28* (2), 387–421. 2

Hall, Alastair R and Atsushi Inoue, "The large sample behaviour of the generalized method of moments estimator in misspecified models," *Journal of Econometrics*, 2003, *114* (2), 361–394. 2, 3, 5, 22

Hall, Peter and Joel L Horowitz, "Bootstrap critical values for tests based on generalized-method-of-moments estimators," *Econometrica: Journal of the Econometric Society*, 1996, pp. 891–916. 2, 3

Hansen, Bruce E and Seojeong Lee, "Inference for iterated GMM under misspecification," *Econometrica*, 2021, *89* (3), 1419–1447. 2

Hong, Han and Jessie Li, "The numerical bootstrap," *The Annals of Statistics*, 2020, *48* (1), 397–412. 2, 27

Honore, Bo E and Luojia Hu, "Estimation of cross sectional and panel data censored regression models with endogeneity," *Journal of Econometrics*, 2004, *122* (2), 293–316. 3, 9

Honoré, Bo E and Luojia Hu, "On the performance of some robust instrumental variables estimators," *Journal of Business & Economic Statistics*, 2004, *22* (1), 30–39. 1

Jun, Sung Jae, Joris Pinkse, and Yuanyuan Wan, "Classical Laplace estimation for-consistent estimators: Improved convergence rates and rate-adaptive inference," *Journal of Econometrics*, 2015, *187* (1), 201–216. 2

Kim, Jeankyung and David Pollard, "Cube root asymptotics," *The Annals of Statistics*, 1990, pp. 191–219. 2, 7, 23, 24, 25, 28, 33, 34, 35, 36

Kosorok, Michael R, *Introduction to empirical processes and semiparametric inference*, Springer, 2007. 4, 26

Lee, Seojeong, "Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators," *Journal of Econometrics*, 2014, *178*, 398–413. 2, 3, 43, 44

Lehmann, Erich L and Joseph P Romano, *Testing Statistical Hypotheses*, Springer Science & Business Media, 2006. 2

McFadden, Daniel, "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica: Journal of the Econometric Society*, 1989, pp. 995–1026. 1, 3, 8

Newey, W. and D. McFadden, "Large Sample Estimation and Hypothesis Testing," in R. Engle and D. McFadden, eds., *Handbook of Econometrics, Vol. 4*, North Holland, 1994, pp. 2113–2241. 4, 9, 23, 28, 43

Pakes, A. and D. Pollard, "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 1989, *57* (5), 1027–1057. 1, 3, 8

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

Pollard, David, "New ways to prove central limit theorems," *Econometric Theory*, 1985, *1* (03), 295–313.

_ , "Asymptotics via empirical processes," *Statistical science*, 1989, pp. 341–354. 7, 9

_ , "Asymptotics for least absolute deviation regression estimators," *Econometric Theory*, 1991, *7* (2), 186–199. 34

Serfling, Robert J, *Approximation theorems of mathematical statistics*, Vol. 162, John Wiley & Sons, 1980. 43

van der Vaart, AW and Jon Wellner, *Weak Convergence and Empirical Processes*, Springer, 1996. 26, 27

Wellner, Jon A and Yihui Zhan, "Bootstrapping Z-estimators," *University of Washington Department of Statistics Technical Report*, 1996, *308.* 9, 26, 27, 31, 32

Zhou, Zhengze, Lucas Mentch, and Giles Hooker, "V-statistics and Variance Estimation.," *J. Mach. Learn. Res.*, 2021, *22*, 287–1. 43