Econometric Theory, **0**, 2024, 1–51.
doi:10.1017/S0266466623000385

RATE-ADAPTIVE BOOTSTRAP FOR POSSIBLY MISSPECIFIED GMM

HAN HONG
Stanford University

JESSIE LI
University of California, Santa Cruz

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.

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; Honoré and Hu, 2004b) and simulated method of moments that are based on frequency simulators (McFadden, 1989; 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. For example, Phillips (2015) points out that quantile regression is always misspecified for a model with unit-root nonstationary regressors. In this paper, we derive the rate of convergence and the limit distribution for the GMM estimator based on nonsmooth moment functions in the misspecified case. The study of misspecification is not only important for estimation and inference of model parameters and for model testing and selection,

We thank three anonymous referees, Denis Chetverikov, Ivan Fernandez-Val, Jean-Jacques Forneron, Hiroaki Kaido, Peter Phillips, Zhongjun Qu, Yinchu Zhu, and participants in conferences and seminars for helpful comments. Address correspondence to Jessie Li, Department of Economics, University of California, Santa Cruz, Santa Cruz, CA, USA; e-mail: jeqli@ucsc.edu.

© The Author(s), 2024. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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, Caner, and Fang (2012), Guggenberger (2012), Lee (2014), Hansen and Lee (2021),Bonhomme and Weidner (2022), Giurcanu and Presnell (2018), Armstrong and Kolesár (2021), and Cheng, Liao, and Shi (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, Pinkse, and Wan (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 data generating process (DGP) (Lehmann and Romano, 2006). However, other resampling methods such as subsampling (Politis, Romano, and Wolf, 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 and Nagasawa (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 nonstandard 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 and Nagasawa (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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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. We acknowledge that our rate-adaptive bootstrap is not uniformly valid because it cannot consistently estimate the asymptotic distribution for locally misspecified models where the population moments are drifting toward zero at the $\sqrt{n}$ rate. The difficulty lies in not being able to consistently estimate the drift constant which appears in the asymptotic distribution. For procedures that handle local misspecification for smooth GMM models, we refer readers to the important work by Bonhomme and Weidner (2022), Armstrong and Kolesár (2021), and references therein.

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's (2003) misspecification-robust (MR) estimator of the asymptotic variance of GMM to develop an MR bootstrap procedure. We investigate their procedure in Appendix A.4's Monte Carlo study and find that our procedure has similar performance to theirs when the moments are smooth. Giurcanu and Presnell (2018) 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 (2022) 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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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 the smooth case under either correct specification or 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; Pakes and Pollard, 1989), and dynamic censored regression (Honoré and Hu, 2004a). While Section 2 studies the one-step GMM estimator under a fixed weighting matrix W, Section 4 studies the two-step GMM estimator computed using an estimated weighting matrix W_n and proposes a rate-adaptive bootstrap for consistent inference. 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 a quantile regression model with misspecified nonsmooth 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 $\mathcal{X}_n = (X_1, X_2, \ldots, X_n)$ of independent draws from a probability measure P on a sample space $\mathcal{X}$. Define the empirical measure $P_n \equiv \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, where 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 $P\left(|X_n^*| > \epsilon | \mathcal{X}_n\right) = o_P(1)$ for all $\epsilon > 0$. Also, define $M_n^* = O_P^*(1)$ (hence also $O_P(1)$) if $\lim_{m\to\infty} \limsup_{n\to\infty} P\left(P\left(M_n^* > m | \mathcal{X}_n\right) > \epsilon\right) \to 0 \forall \epsilon > 0$.

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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 $\pi\left(\theta^{\#}\right) = 0$, $\sqrt{n}\left(\hat{\theta}_n - \theta^{\#}\right) \rightsquigarrow \left(G'WG\right)^{-1} G'WN\left(0, P\pi\left(\cdot,\theta^{\#}\right)\pi\left(\cdot,\theta^{\#}\right)'\right)$, where $G = \frac{\partial}{\partial\theta}\pi\left(\theta^{\#}\right)$.

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

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

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 nonstandard, 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

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

$\mathcal{Z}_0(h)$ 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.

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

For $g(\cdot,\theta) = \pi(\cdot,\theta) - \pi\left(\cdot,\theta^{\#}\right)$, the covariance kernel of $\mathcal{Z}_0(h)$ is

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

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 W is

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

Here, $\hat{\pi}_n\left(\hat{\theta}_n\right) = P_n\pi\left(\cdot,\hat{\theta}_n\right)$, $\hat{G}$ 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 and Nagasawa (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.

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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 $j = 1, \ldots, m$.
2. Repeat for B bootstrap iterations: draw a bootstrap sample $X_1^*, \ldots, X_n^*$ and compute

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

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

A $1-\alpha$ 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 parameters:

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

We can also compute a confidence interval for $\rho\left(\theta^{\#}\right)$, where $\rho : \Theta \mapsto \mathbb{R}$, by using the percentiles of the empirical distribution of $\rho\left(\hat{\theta}_n^*\right) - \rho\left(\hat{\theta}_n\right)$:

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

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 $\hat{Q}_n(\theta) \equiv \frac{1}{2}P_n\pi(\cdot,\theta)' WP_n\pi(\cdot,\theta)$ and $Q(\theta) \equiv \frac{1}{2}P\pi(\cdot,\theta)' WP\pi(\cdot,\theta)$, where W is positive definite, suppose the following conditions are satisfied for some $\rho \in \left\{\frac{1}{2}, 1\right\}$ and $\gamma = \frac{1}{2(2-\rho)}$:

(i) $\hat{Q}_n\left(\hat{\theta}_n\right) \le \inf_{\theta\in\Theta}\hat{Q}_n(\theta) + o_P\left(n^{-2\gamma}\right)$.
(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) $\sup_{\theta\in\Theta} P|\pi(\cdot,\theta)| < \infty$.

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

Assumption 2. Let $g(\cdot,\theta) \equiv \pi(\cdot,\theta) - \pi\left(\cdot,\theta^{\#}\right)$ satisfy the following conditions for some $\rho \in \left\{\frac{1}{2}, 1\right\}$ 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 $G_R(\cdot) \equiv \sup_{g_j \in \mathcal{G}_R} \left|g_j(\cdot,\theta)\right|$.
(iii) $Pg(\cdot,\theta)$ is twice differentiable at $\theta^{\#}$ with full-rank 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 $\|\theta_1 - \theta_2\| \to 0$.
(ix) $\bar{H} = G'WG + \sum_{j=1}^m \sum_{k=1}^m W_{jk}\pi_k\left(\theta^{\#}\right) H_j$ is positive definite.

Assumption 1 is needed to show consistency of $\hat{\theta}_n$ for $\theta^{\#}$, while Assumption 2 is needed to derive its asymptotic distribution. Manageable classes are defined in Definition 4.1 of Pollard (1989), and an example is all euclidean classes. A manageable class for a constant envelope is a universal Donsker class in the sense of Dudley (1987). Uniform manageable classes are manageable classes for which a uniform upper bound exists in the maximal inequalities for the corresponding empirical processes. As discussed after Corollary 3.2 of Kim and Pollard (1990), we need to assume $\mathcal{G}_R$ are uniformly manageable in order to demonstrate stochastic equicontinuity of certain processes that appear in the expansion of the objective function. We demonstrate stochastic equicontinuity by applying the maximal inequalities in Lemma 3.1 of Kim and Pollard (1990) over the classes $\mathcal{G}_R$ for all values of R near zero, rather than a particular value of R.

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 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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

misspecification of moments in Chernozhukov and Hansen's (2005)) 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 nonstandard asymptotic distribution.

The moment conditions for IVQR are nonsmooth, in particular non-directionally differentiable, because $\pi(\cdot,\theta) = (\tau - 1(y_i \leq q(d_i, w_i, \theta)))z_i$, where y_i is the dependent variable, d_i is a vector of endogenous regressors, w_i is a vector of exogenous regressors, z_i is a vector of instruments, and $q(\cdot)$ is the quantile response function, 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, and $F_{y|x,z}$ is absolutely continuous. For $\pi(\theta) = E(\tau - F_{y|x,z}(q(x'\theta^{\#})))z$, the Jacobian is $G = \frac{\partial}{\partial\theta}\pi(\theta^{\#}) = -Ef_{y|x,z}(q(x'\theta^{\#}))zq'(x'\theta^{\#})x'$ and the jth element of the Hessian is $H_j = \frac{\partial^2}{\partial\theta\partial\theta'}\pi_j(\theta^{\#}) = -Ef'_{y|x,z}(q(x'\theta^{\#}))z_j(q'(x'\theta^{\#}))^2 xx' + Ef_{y|x,z}(q(x'\theta^{\#}))z_j q''(x'\theta^{\#})xx'$. We will assume that the assumptions in Chernozhukov and Hansen (2005) needed to ensure that G and H_j are well defined are satisfied.

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 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; 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 $\pi(\cdot,\theta) = \left(y_i - \frac{1}{S}\sum_{s=1}^{S} 1(h(x_i'\theta) + \eta_{is} > 0)\right)z_i$, 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 $\pi(\theta) = E(y - F_{\eta|x,z}(h(x'\theta)))z$, the Jacobian is $G = \frac{\partial}{\partial\theta}\pi(\theta^{\#}) = -Ef_{\eta|x,z}(h(x'\theta^{\#}))zh'(x'\theta^{\#})x'$ and the jth element of the Hessian is $H_j = \frac{\partial^2}{\partial\theta\partial\theta'}\pi_j(\theta^{\#}) = -Ef'_{\eta|x,z}(h(x'\theta^{\#}))z_j(h'(x'\theta^{\#}))^2 xx' + Ef_{\eta|x,z}(h(x'\theta^{\#}))z_j h''(x'\theta^{\#})xx'$. We will assume that the assumptions in McFadden (1989) needed to ensure that G and H_j are well defined are satisfied.

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

We verify in the Appendix that the value of ρ that satisfies Assumption 2 is $\rho = 1/2$.

Example 3. *Dynamic censored regression:* Honoré and Hu (2004a) consider the estimation of a panel data censored regression model with lagged dependent variables: $y_{it} = \max\{0, y_{it-1}\theta + \alpha_i + \epsilon_{it}\}$ where $\{\epsilon_{it}\}_{t=1}^{T}$ is a sequence of i.i.d. random variables conditional on (y_{i0}, α_i). They show that the GMM estimator of θ 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. The stacked moments are $\pi(\theta) = [\pi_2(\theta), \ldots, \pi_T(\theta)]'$, where for each $t = 2, \ldots, T$, $\pi_t(\theta) = E[\max\{0, y_{it} - y_{it-1}\theta\} - y_{it-1}] = E[1(y_{it} > y_{it-1}\theta)(y_{it} - y_{it-1}\theta) - y_{it-1}]$. 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 for $t = 2, \ldots, T$ are $H_t = E\left[y_{it-1}f_{y_{it}|y_{it-1}}\left(y_{it-1}\theta^{\#}\right)\right]$.

Even though $\pi(\cdot,\theta)$ is nonsmooth, the $\sqrt{n}$ rate of convergence arises because $\pi(\cdot,\theta)$ remains directionally differentiable. We check in the Appendix that the value of ρ that satisfies Assumption 2 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 and 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

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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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 and Nagasawa (2020) because Cattaneo and Nagasawa (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 a relatively mild envelope (square) integrability assumption (their Assumption A.5).

Assumption 3. For some $\rho \in \left\{\frac{1}{2}, 1\right\}$ 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 $\epsilon_n \to 0$ and any compact set $K \subset \mathbb{R}^d$,

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}\sup_{t\ge\lambda} t^2 P\left\{\sup_{h\in K,\,\|\theta-\theta^{\#}\|\le\epsilon_n}\left\|\frac{m_n(\cdot,\theta,h)-m_n(\cdot,\theta^{\#},h)}{1+n^{\gamma}\|\theta-\theta^{\#}\|}\right\| > t\right\} = 0.$$

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

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

Using the notation of Assumption A.5 of Wellner and Zhan (1996), Assumption 3(i) is using $B(\theta)(h) = m_n(\cdot,\theta,h)$ and $\mathcal{H} = K$. Assumption 3(ii) is using $B(\theta)(h) = \pi(\cdot,\theta)$ and $\mathcal{H}$ is a finite set whose cardinality is the dimension of θ.

Strong sufficient conditions for Assumption 3 are that the envelopes are uniformly bounded. For all sufficiently large n such that $\epsilon_n \to 0$ and any compact $K \subset \mathbb{R}^d$, there exists some constants $C_1 > 0$ and $C_2 > 0$ such that $\sup_{h\in K,\,\|\theta-\theta^{\#}\|\le\epsilon_n}\left\|\frac{m_n(\cdot,\theta,h)-m_n(\cdot,\theta^{\#},h)}{1+n^{\gamma}\|\theta-\theta^{\#}\|}\right\| \le C_1$, and $\sup_{\|\theta-\theta^{\#}\|\le\epsilon_n}\left\|\frac{\pi(\cdot,\theta)-\pi(\cdot,\theta^{\#})}{1+\sqrt{n}\|\theta-\theta^{\#}\|}\right\| \le C_2$.

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 and 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 and 2 are satisfied for* $\gamma = 1/2, \rho = 1$,

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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

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

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

4. THE CASE OF AN ESTIMATED WEIGHTING MATRIX

We now consider the case of an estimated weighting matrix. First, we show that a 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: $\theta_1^\# = \arg\min_{\theta\in\Theta} \pi(\theta)' W_1 \pi(\theta)$. For example, we may choose $W\left(\theta_1^\#\right)$ to be the inverse of the variance–covariance matrix of the population moments evaluated at $\theta_1^\#$: $W\left(\theta_1^\#\right) = \left(E\left[\pi\left(\cdot,\theta_1^\#\right)\pi\left(\cdot,\theta_1^\#\right)'\right] - \pi\left(\theta_1^\#\right)\pi\left(\theta_1^\#\right)'\right)^{-1}$. Different choices of W_1 and $W\left(\theta_1^\#\right)$ typically lead to different values of $\theta^\#$, but we suppress the dependence of $\theta^\#$ on the weighting matrices for notational simplicity.

The estimated weighting matrix $W_n\left(\hat{\theta}_1\right)$ will depend on the one-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 two-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 and 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 $\bar{\mathcal{Z}}_0(h) \equiv \pi\left(\theta^\#\right)' W \mathcal{Z}_{0,1/2}(h)$,

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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

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

When W_n converges to W at the cubic-root rate, we assume that $n^{1/3}\left(W_n - W\right) \pi\left(\theta^{\#}\right)$ converges in distribution to some tight random vector $\mathcal{W}_0$. W_n's cubic-root rate of convergence arises when the moments are nonsmooth (non-directionally differentiable) and misspecified because then the one-step GMM estimator $\hat{\theta}_1$ converges at the cubic-root rate and determines the asymptotics of W_n. More details are in Appendix A.3.

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

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

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:

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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:

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

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

THEOREM 4. *Suppose Assumptions 1 and 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,*

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

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

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

$$\Sigma_{13} = P\left(\pi\left(\cdot,\theta^\#\right) - \pi\left(\theta^\#\right)\right)\pi\left(\theta^\#\right)'\left(\phi\left(\cdot,\theta_1^\#\right) - \phi\left(\theta_1^\#\right)\right)',$$

$$\Sigma_{31} = \Sigma_{13}',$$

$$\Sigma_{23} = P\left(\frac{\partial}{\partial\theta}\pi\left(\cdot,\theta^\#\right) - G\right)' W\pi\left(\theta^\#\right)\pi\left(\theta^\#\right)'\left(\phi\left(\cdot,\theta_1^\#\right) - \phi\left(\theta_1^\#\right)\right)',$$

$$\Sigma_{32} = \Sigma_{23}',$$

$$\Sigma_{33} = P\left(\phi\left(\cdot,\theta_1^\#\right) - \phi\left(\theta_1^\#\right)\right)\pi\left(\theta^\#\right)\pi\left(\theta^\#\right)'\left(\phi\left(\cdot,\theta_1^\#\right) - \phi\left(\theta_1^\#\right)\right)'.$$

(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* $W_n^* - W_n = o_P^*\left(n^{-1/3}\right)$,

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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 $\begin{pmatrix} \pi\left(\theta^\#\right)' W n^{2/3} (P_n - P) g\left(\cdot, \theta^\# + n^{-1/3} h\right) \\ h' G' n^{1/3} (W_n - W) \pi\left(\theta^\#\right) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \pi\left(\theta^\#\right)' W \mathcal{Z}_{0,1/2}(h) \\ h' G' \mathcal{W}_0 \end{pmatrix}$
in $\left\{\mathbf{B}_{loc}\left(\mathbb{R}^d\right)\right\}^2$ *for some tight random vector* $\mathcal{W}_0$*, 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 $\pi(\cdot,\theta) = [1(y_i \le \theta) - \tau; y_i - \theta]'$, let the population moments be

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

The model cannot 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:

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

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

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

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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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

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

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

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

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 three 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 $\tau \in \{0.1, 0.3, 0.5\}$, $n \in \{200, 800, 1,600, 3,200, 6,400\}$, $B = 1,000$ bootstrap iterations, and $R = 1,000$ 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 well 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.

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

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

TABLE 1. Rate-adaptive versus standard bootstrap, fixed weighting matrix.

	Rate-adaptive bootstrap			Standard bootstrap		
τ	0.1	0.3	0.5	0.1	0.3	0.5
$n = 200$	0.949	0.946	0.952	0.900	0.917	0.949
	(0.330)	(0.298)	(0.279)	(0.277)	(0.277)	(0.277)
$n = 800$	0.950	0.955	0.949	0.864	0.920	0.947
	(0.180)	(0.156)	(0.140)	(0.139)	(0.139)	(0.139)
$n = 1,600$	0.959	0.950	0.948	0.845	0.909	0.954
	(0.134)	(0.113)	(0.099)	(0.098)	(0.098)	(0.098)
$n = 3200$	0.957	0.960	0.955	0.828	0.889	0.936
	(0.101)	(0.083)	(0.070)	(0.070)	(0.070)	(0.070)
$n = 6,400$	0.952	0.946	0.947	0.797	0.894	0.951
	(0.076)	(0.061)	(0.049)	(0.049)	(0.049)	(0.049)

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

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

The bootstrapped weighting matrix is

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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 $\hat{F}_y^*\left(\hat{\theta}_1^*\right) = \frac{1}{n}\sum_{i=1}^n 1\left(y_i^* \le \hat{\theta}_1^*\right)$. 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 three columns of Table 2 show the empirical coverage frequencies and average interval lengths 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 $B = 1{,}000$ bootstrap iterations and $R = 1{,}000$ Monte Carlo simulations. There is some slight under-coverage 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 under-coverage 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.

5.2. Quantile Regression

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

$$y_i = \alpha_0 + \beta_0 d_i + u_i, \qquad \begin{pmatrix} u_i \\ d_i \\ w_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \delta \\ 0 & 1 & 0.5 \\ \delta & 0.5 & 1 \end{pmatrix}\right).$$

It follows then that

$$\begin{aligned} u_i | d_i, w_i &\sim N\left(\begin{pmatrix} 0 & \delta \end{pmatrix}\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}^{-1}\begin{pmatrix} d_i \\ w_i \end{pmatrix}, 1 - \begin{pmatrix} 0 & \delta \end{pmatrix}\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ \delta \end{pmatrix}\right) \\ &= N\left(\left(-\frac{2}{3}d_i + \frac{4}{3}w_i\right)\delta, 1 - \frac{4}{3}\delta^2\right) \\ y_i | d_i, w_i &\sim N\left(\alpha_0 + \beta_0 d_i + \left(-\frac{2}{3}d_i + \frac{4}{3}w_i\right)\delta, 1 - \frac{4}{3}\delta^2\right). \end{aligned}$$

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

TABLE 2. Rate-adaptive versus standard bootstrap, estimated weighting matrix.

	Rate-adaptive bootstrap			Standard bootstrap		
τ	0.1	0.3	0.5	0.1	0.3	0.5
$n = 200$	0.967	0.968	0.944	0.691	0.844	0.951
	(1.211)	(0.686)	(0.285)	(0.341)	(0.329)	(0.310)
$n = 800$	0.984	0.979	0.951	0.633	0.765	0.957
	(0.702)	(0.369)	(0.143)	(0.175)	(0.172)	(0.160)
$n = 1,600$	0.982	0.949	0.941	0.606	0.755	0.965
	(0.521)	(0.293)	(0.101)	(0.124)	(0.122)	(0.114)
$n = 3,200$	0.985	0.960	0.963	0.575	0.708	0.976
	(0.398)	(0.235)	(0.071)	(0.089)	(0.088)	(0.081)
$n = 6,400$	0.977	0.950	0.951	0.562	0.674	0.963
	(0.314)	(0.189)	(0.050)	(0.064)	(0.063)	(0.057)

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

$$\begin{aligned}\pi(\theta) &= E\left[\left(\frac{1}{2} - 1(y_i \le \alpha + \beta d_i)\right) z_i\right] \\ &= E\left[\left(\frac{1}{2} - F_{y|d,w}(\alpha + \beta d_i)\right) z_i\right] \\ &= E\left[\left(\frac{1}{2} - \Phi\left(\frac{\alpha - \alpha_0 + (\beta - \beta_0) d_i + \left(\frac{2}{3} d_i - \frac{4}{3} w_i\right)\delta}{\sqrt{1 - \frac{4}{3}\delta^2}}\right)\right) z_i\right].\end{aligned}$$

At the true parameter values, the population moments are

$$\pi(\theta_0) = E\left[\left(\frac{1}{2} - \Phi\left(\frac{\left(\frac{2}{3} d_i - \frac{4}{3} w_i\right)\delta}{\sqrt{1 - \frac{4}{3}\delta^2}}\right)\right) z_i\right].$$

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 sample moments are

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

The bootstrapped sample moments using the multinomial bootstrap are

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

The population Jacobian and Hessians are for $\tilde{d}_i = (1, d_i)'$,

$$G(\theta) = -E\left[f_{y|d,w}(\alpha + \beta d_i) z_i \tilde{d}_i'\right], \qquad H_1(\theta) = -E\left[f'_{y|d,w}(\alpha + \beta d_i)\tilde{d}_i \tilde{d}_i'\right],$$
$$H_2(\theta) = -E\left[f'_{y|d,w}(\alpha + \beta d_i) d_i \tilde{d}_i \tilde{d}_i'\right], \qquad H_3(\theta) = -E\left[f'_{y|d,w}(\alpha + \beta d_i) w_i \tilde{d}_i \tilde{d}_i'\right].$$

Their estimates are

$$\hat{G} = -\frac{1}{nh}\sum_{i=1}^{n} K_h\left(y_i - \hat{\alpha}_n - \hat{\beta}_n d_i\right) z_i \tilde{d}_i', \qquad \hat{H}_1 = -\frac{1}{nh^2}\sum_{i=1}^{n} K_h'\left(y_i - \hat{\alpha}_n - \hat{\beta}_n d_i\right)\tilde{d}_i \tilde{d}_i',$$
$$\hat{H}_2 = -\frac{1}{nh^2}\sum_{i=1}^{n} K_h'\left(y_i - \hat{\alpha}_n - \hat{\beta}_n d_i\right) d_i \tilde{d}_i \tilde{d}_i', \qquad \hat{H}_3 = -\frac{1}{nh^2}\sum_{i=1}^{n} K_h'\left(y_i - \hat{\alpha}_n - \hat{\beta}_n d_i\right) w_i \tilde{d}_i \tilde{d}_i',$$

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

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

Table 3 compares 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 $B = 2{,}000$ bootstrap iterations, $R = 5{,}000$ Monte Carlo simulations, and $\delta = 0.1$ and $h = 0.3$. The standard bootstrap produces shorter intervals and undercovers for all values of n, while the rate-adaptive bootstrap achieves coverage close to the nominal level for both parameters.

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

TABLE 3. Rate-adaptive versus standard bootstrap, fixed weighting matrix $W = I$.

	n	200	800	1,600	3,200	6,400
	α_0	0.967	0.954	0.957	0.957	0.957
Rate-adaptive		(0.362)	(0.188)	(0.136)	(0.099)	(0.072)
	β_0	0.969	0.951	0.950	0.954	0.955
		(0.380)	(0.195)	(0.139)	(0.100)	(0.072)
	α_0	0.861	0.915	0.919	0.915	0.910
Standard		(0.280)	(0.177)	(0.125)	(0.088)	(0.062)
	β_0	0.852	0.909	0.913	0.912	0.921
		(0.282)	(0.184)	(0.130)	(0.092)	(0.065)

$\theta^{\#} = \arg\min_{\theta} \pi(\theta)' W\left(\theta_1^{\#}\right)\pi(\theta)$. $W\left(\theta_1^{\#}\right)$ is the inverse of the variance–covariance matrix of the population moments

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

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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

TABLE 4. Rate-adaptive versus standard bootstrap, estimated weighting matrix.

	n	200	800	1,600	3,200	6,400
	α_0	0.968	0.949	0.951	0.949	0.947
Rate-adaptive		(0.369)	(0.195)	(0.143)	(0.106)	(0.079)
	β_0	0.969	0.951	0.954	0.955	0.956
		(0.371)	(0.194)	(0.142)	(0.105)	(0.078)
	α_0	0.889	0.916	0.913	0.886	0.878
Standard		(0.278)	(0.177)	(0.126)	(0.089)	(0.063)
	β_0	0.864	0.918	0.904	0.901	0.877
		(0.262)	(0.176)	(0.126)	(0.089)	(0.063)

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

Table 4 compares 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 $\tilde{\theta}_2 = \arg\min_{\theta\in\Theta}\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)' W_n^*\left(\tilde{\theta}_1\right)\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)$, where the weighting matrix depends on $\tilde{\theta}_1 = \arg\min_{\theta\in\Theta}\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)'$ $\left(\hat{\pi}_n^*(\theta) - \hat{\pi}_n\left(\hat{\theta}_n\right)\right)$. We use $B = 2{,}000$ bootstrap iterations, $R = 5{,}000$ Monte Carlo simulations, and $\delta = 0.1$ and $h = 0.3$. The standard bootstrap produces shorter intervals and undercovers for all values of n, while the rate-adaptive bootstrap achieves coverage close to the nominal level.

6. CONCLUSION

We have demonstrated that globally misspecified GMM estimators with nonsmooth (specifically non-directionally differentiable) moments have a cubic-root 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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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.

A. Appendix

A.1. Additional Results for Misspecified GMM with Directionally Differentiable Moments

Asymptotic distribution under fixed weighting matrix.

THEOREM A.1. *Suppose* $\pi\left(\theta^{\#}\right)=c$ *for a vector of fixed constants* $c\neq 0$ *and that Assumptions 1 and 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*

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

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* $\Sigma_1(s,t)=\lim_{\alpha\to\infty}\alpha^2Pg\left(\cdot,\theta^{\#}+\frac{s}{\alpha}\right)g\left(\cdot,\theta^{\#}+\frac{t}{\alpha}\right)'$. *The joint covariance kernel of* $\mathcal{Z}_{0,1}(h)$ *and* $h'G'WU_0$ *is given by*

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

Note that in the case of smooth misspecified models, the asymptotic distribution in Theorem A.1 reduces down to the one in Theorem 1 of Hall and Inoue (2003) since then $\pi\left(\theta^{\#}\right)'W\mathcal{Z}_{0,1}(h)$ can be replaced by $h'Z_0'W\pi\left(\theta^{\#}\right)$, 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 A.2. *Suppose Assumption 3 is satisfied in addition to the assumptions in Theorem A.1,* $\hat{G}\xrightarrow{p}G$, *and* $\hat{H}_j\xrightarrow{p}H_j$ *for* $j=1,\ldots,m$. *Then,*

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

Asymptotic distribution under estimated weighting matrix.

THEOREM A.3. *Suppose $\pi\left(\theta^{\#}\right)=c$ for a vector of fixed constants $c\neq 0$ and that Assumptions 1 and 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

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

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 A.3 reduces down to the one in Theorem 2 of Hall and Inoue (2003) since then $\pi\left(\theta^{\#}\right)'W\mathcal{Z}_{0,1}(h)$ can be replaced by $h'Z_0'W\pi\left(\theta^{\#}\right)$, 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 A.4. *Suppose Assumptions 1–3 are satisfied, $\hat{G}\xrightarrow{p}G$, and $\hat{H}_j\xrightarrow{p}H_j$ for $j=1,\dots,m$. 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*

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

If instead Assumption 4 is satisfied,

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

A.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 equation array in the proof of Theorem 2.6 in Newey and McFadden (1994) in combination with Assumptions 1(iii) and (iv) imply

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

$\sup_{\theta\in\Theta}\left|\hat{Q}_n(\theta)-Q(\theta)\right| = o_P(1)$. Next, we show that $n^{1/3}\left(\hat{\theta}_n-\theta^{\#}\right)=O_P(1)$. Define $\hat{\mathcal{G}}_n(\theta)=\sqrt{n}(P_n-P)\,g(\cdot,\theta)$, $\hat{g}(\theta)=P_n g(\cdot,\theta)$, and $g(\theta)=Pg(\cdot,\theta)$. Then $\hat{\pi}_n(\theta)=g(\theta)+\hat{\pi}_n\left(\theta^{\#}\right)+\hat{\eta}_n(\theta)$, where $\hat{\eta}_n(\theta)=\frac{1}{\sqrt{n}}\hat{\mathcal{G}}_n(\theta)$. Recall that $\hat{Q}_n(\theta)=\frac{1}{2}\hat{\pi}_n(\theta)'W\hat{\pi}_n(\theta)$. Write $\hat{Q}_n(\theta)-\hat{Q}_n\left(\theta^{\#}\right)=Q_1(\theta)+\hat{Q}_2(\theta)+\hat{Q}_3(\theta)$, where

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

Apply Kim and Pollard (1990, Lem. 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

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

The first, third, and fourth 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 second term in $\hat{Q}_2(\theta)$, for n large enough, $\forall\epsilon>0$, $\exists M_{n,22}=O_P(1)$ such that

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

Therefore, $\forall\epsilon>0$, $\exists M_n=O_P(1)$ such that $|\hat{Q}_2(\theta)+\hat{Q}_3(\theta)|\le\epsilon\|\theta-\theta^{\#}\|^2+n^{-2/3}M_n^2$.

Because $Q_1(\theta)$ achieves the minimal value of 0 at $\theta^{\#}$, the Taylor expansion of $Q_1(\theta)$ around $\theta^{\#}$ is $Q_1(\theta)=Q_1\left(\theta^{\#}\right)+\left(\theta-\theta^{\#}\right)'\frac{\partial Q_1(\theta^{\#})}{\partial\theta}+\frac{1}{2}\left(\theta-\theta^{\#}\right)'\frac{\partial^2 Q_1(\theta^{\#})}{\partial\theta\partial\theta'}\left(\theta-\theta^{\#}\right)+o\left(\|\theta-\theta^{\#}\|^2\right)=\frac{1}{2}\left(\theta-\theta^{\#}\right)'\left(\bar{H}+o(1)\right)\left(\theta-\theta^{\#}\right)$ since $\frac{\partial Q_1(\theta^{\#})}{\partial\theta}=G'Wg\left(\theta^{\#}\right)+G'W\pi\left(\theta^{\#}\right)=0$ and $\frac{\partial^2 Q_1(\theta^{\#})}{\partial\theta\partial\theta'}=\bar{H}$. Because $\bar{H}$ is positive definite, there exists $C>0$ and a small enough neighborhood of $\theta^{\#}$ such that $Q_1(\theta)\ge C\left\|\theta-\theta^{\#}\right\|^2$. By consistency of $\hat{\theta}_n$ for $\theta^{\#}$, with probability approaching 1, $Q_1\left(\hat{\theta}_n\right)\ge C\left\|\hat{\theta}_n-\theta^{\#}\right\|^2$. Then,

$$Q_1\left(\hat{\theta}_n\right)+\hat{Q}_2\left(\hat{\theta}_n\right)+\hat{Q}_3\left(\hat{\theta}_n\right)=\hat{Q}_n\left(\hat{\theta}_n\right)-\hat{Q}_n\left(\theta^{\#}\right)\le\hat{Q}_n\left(\hat{\theta}_n\right)-\inf_{\theta\in\Theta}\hat{Q}_n(\theta)\le o_P\left(n^{-2/3}\right).$$

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

$$o_P\left(n^{-2/3}\right)\ge Q_1\left(\hat{\theta}_n\right)-\epsilon\left\|\hat{\theta}_n-\theta^{\#}\right\|^2-n^{-2/3}M_n^2$$
$$\ge(C-\epsilon)\left\|\hat{\theta}_n-\theta^{\#}\right\|^2-n^{-2/3}M_n^2$$
$$\Longrightarrow\left\|\hat{\theta}_n-\theta^{\#}\right\|^2\le(C-\epsilon)^{-1}n^{-2/3}M_n^2+o_P\left(n^{-2/3}\right)=O_P\left(n^{-2/3}\right).$$

It follows that $n^{1/3}\left(\hat{\theta}_n-\theta^{\#}\right)=O_P(1)$.

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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 $Q_1\left(\theta^{\#} + n^{-1/3}h\right) = Q_1\left(\theta^{\#}\right) + n^{-1/3}h'\frac{\partial Q_1(\theta^{\#})}{\partial\theta} + \frac{1}{2}n^{-2/3}h'\frac{\partial^2 Q_1(\theta^{\#})}{\partial\theta\partial\theta'}h + o\left(n^{-2/3}\right)$, $n^{2/3}Q_1\left(\theta^{\#} + n^{-1/3}h\right) = \frac{1}{2}h'\bar{H}h + o(1)$.

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) implies 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 $j = 1,\ldots,m$,

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

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 $\left.\max(\|h_1\|,\|h_2\|) \le M \text{ and } \|h_1 - h_2\| \le \delta_n\right\}$. Note that $\mathscr{D}(n)$ has envelope function $D_n = 2G_{R(n)}$ where $R(n) = Mn^{-1/3}$.

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

$$\begin{aligned} n^{2/3}E\sup_{\mathscr{D}(n)}\left|P_nd_j - Pd_j\right| &\le E\sqrt{n^{1/3}P_nD_n^2}J\left(\frac{n^{1/3}\sup_{\mathscr{D}(n)}P_nd_j^2}{n^{1/3}P_nD_n^2}\right)\\ &\le \sqrt{\eta}J(1) + \sqrt{En^{1/3}P_nD_n^2}\sqrt{EJ^2\left(\min\left(1,\frac{1}{\eta}n^{1/3}\sup_{\mathscr{D}(n)}P_nd_j^2\right)\right)}. \end{aligned}$$

To show that this is $o(1)$ for each fixed $\eta > 0$, first, note that by Assumption 2(vi), $En^{1/3}P_nD_n^2 = 4n^{1/3}EG_{R(n)}^2 = O(n^{1/3}R(n)) = O(1)$ since $R(n) = Mn^{-1/3}$. The proof will then be complete if $n^{1/3}\sup_{\mathscr{D}(n)}P_nd_j^2 = o_p(1)$. Next, for each $K > 0$, write

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

$E\sup_{\mathscr{D}(n)} P_n d_j^2 \le EP_n \sup_{\mathscr{D}(n)} d_j^2 1\{D_n > K\} + KE\sup_{\mathscr{D}(n)} P_n |d_j| \le EP_n D_n^2 1\{D_n > K\} + K\sup_{\mathscr{D}(n)} P|d_j| + KE\sup_{\mathscr{D}(n)} \left|P_n|d_j| - P|d_j|\right|$. By Assumption 2(vii), $EP_n D_n^2 1\{D_n > K\} < \eta n^{-1/3}$ 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 Lemma 3.1 of Kim and Pollard (1990), $KE\sup_{\mathscr{D}(n)} \left|P_n|d_j| - P|d_j|\right| < Kn^{-\frac{1}{2}}J(1)\sqrt{PD_n^2} = O(n^{-2/3}) = o(n^{-1/3})$. Therefore, $En^{1/3}\sup_{\mathscr{D}(n)} P_n d_j^2 = o(1)$.

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 first, third, and fourth 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 second term there,

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

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 h 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)). □

Proofs for Theorems 2 and A.2. Equation 4.2 implies that for $\hat{h}^* = n^\gamma\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$,

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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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 A.1 that $S_n(h)$ is stochastically equicontinuous in h for $\rho = 1$, $\gamma = 1/2$.

Therefore, $S_n(h) \rightsquigarrow \mathcal{Z}_{0,\rho}(h)$ 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 $S_n(h)$:

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

We already showed $n^{\gamma\rho}\sqrt{n}\left(P_n - P\right)\left(\pi\left(\cdot,\theta + \frac{h}{n^\gamma}\right) - \pi(\cdot,\theta)\right)$ is stochastically equicontinuous in θ, which implies that for any compact $K \subset \mathbb{R}^d$,

$$n^{\gamma\rho}\sqrt{n}\sup_{h\in K}\left\|\left(P_n - P\right)\left(\pi\left(\cdot,\hat{\theta}_n + \frac{h}{n^\gamma}\right) - \pi\left(\cdot,\hat{\theta}_n\right) - \left(\pi\left(\cdot,\theta^\# + \frac{h}{n^\gamma}\right) - \pi\left(\cdot,\theta^\#\right)\right)\right)\right\|$$
$$= o_P\left(1 + n^\gamma\left\|\hat{\theta}_n - \theta^\#\right\|\right) = o_P(1).$$

Under the envelope integrability Assumption 3, Lemma 4.2 in Wellner and Zhan (1996) implies that for any compact $K \subset \mathbb{R}^d$,

$$n^{\gamma\rho}\sqrt{n}\sup_{h\in K}\left\|\left(P_n^* - P_n\right)\left(\pi\left(\cdot,\hat{\theta}_n + \frac{h}{n^\gamma}\right) - \pi\left(\cdot,\hat{\theta}_n\right) - \left(\pi\left(\cdot,\theta^\# + \frac{h}{n^\gamma}\right) - \pi\left(\cdot,\theta^\#\right)\right)\right)\right\|$$
$$= o_P^*\left(1 + n^\gamma\left\|\hat{\theta}_n - \theta^\#\right\|\right) = o_P^*(1).$$

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

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

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. 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 $\hat{\pi}_n\left(\hat{\theta}_n\right) \xrightarrow{p} \pi\left(\theta^\#\right)$,

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

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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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

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

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

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

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

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,

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

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

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

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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

For smooth models that are misspecified, $\hat{\pi}_n\left(\hat{\theta}_n\right)' Wn\left(P_n^* - P_n\right)\left(\pi\left(\cdot,\hat{\theta}_n + \frac{h}{\sqrt{n}}\right) - \pi\left(\cdot,\hat{\theta}_n\right)\right) \underset{\mathbb{W}}{\overset{\mathbb{P}}{\rightsquigarrow}} h'Z_0'W\pi\left(\theta^\#\right)$, 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

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

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

Therefore, the asymptotic distribution is given by

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

Proof for Theorem 3. The consistency argument is a direct application of Theorem 5.7 in van der Vaart (2000) since the equation array in the proof of Theorem 2.6 in Newey and McFadden (1994) in combination with Assumption 1(iii) and (iv) and $W_n - W = o_p(1)$ imply $\sup_{\theta\in\Theta}\left|\hat{Q}_n(\theta) - Q(\theta)\right| = o_P(1)$. 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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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

Next, recall that Kim and Pollard (1990, Lem. 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 second and third 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 first term in $\hat{Q}_5(\theta)$ can also be bounded by, for some $M_{n,52} = O_P(1)$ and $\forall\epsilon > 0$,

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

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

$$\begin{aligned}
\hat{Q}_4(\theta) &= \hat{Q}_4\left(\theta^{\#}\right) + \left(\theta - \theta^{\#}\right)'\frac{\partial\hat{Q}_4(\theta^{\#})}{\partial\theta} + \frac{1}{2}\left(\theta - \theta^{\#}\right)'\frac{\partial^2\hat{Q}_4(\theta^{\#})}{\partial\theta\partial\theta'}\left(\theta - \theta^{\#}\right) + o_p\left(\|\theta - \theta^{\#}\|^2\right) \\
&= \frac{1}{2}\left(\theta - \theta^{\#}\right)'\left(G'(W_n - W)G + \sum_{j=1}^{m}\sum_{k=1}^{m}(W_{n,jk} - W_{jk})\pi_k\left(\theta^{\#}\right)H_j + o_p(1)\right)\left(\theta - \theta^{\#}\right) \\
&\quad + O_P\left(\frac{\|\theta - \theta^{\#}\|}{n^{\gamma}}\right) \\
&\leq \epsilon\|\theta - \theta^{\#}\|^2 + n^{-2\gamma}M_{n,4}^2.
\end{aligned}$$

Then $\forall\epsilon > 0$, there exists $M_n = O_P(1)$ such that $|\hat{Q}_2(\theta) + \hat{Q}_3(\theta) + \hat{Q}_4(\theta) + \hat{Q}_5(\theta) + \hat{Q}_6(\theta)| \leq \epsilon\|\theta - \theta^{\#}\|^2 + n^{-2/3}M_n^2$.

We already showed that there exists some $C > 0$ such that almost surely $Q_1\left(\hat{\theta}_n\right) \geq C\left\|\hat{\theta}_n - \theta^{\#}\right\|^2$. Then,

$$Q_1\left(\hat{\theta}_n\right) + \hat{Q}_2\left(\hat{\theta}_n\right) + \hat{Q}_3\left(\hat{\theta}_n\right) + \hat{Q}_4\left(\hat{\theta}_n\right) + \hat{Q}_5\left(\hat{\theta}_n\right) + \hat{Q}_6\left(\hat{\theta}_n\right) \leq \hat{Q}_n\left(\hat{\theta}_n\right) - \inf_{\theta\in\Theta}\hat{Q}_n(\theta) \leq o_P\left(n^{-2/3}\right).$$

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

$$\begin{aligned}
o_P\left(n^{-2/3}\right) &\geq Q_1\left(\hat{\theta}_n\right) - \epsilon\left\|\hat{\theta}_n - \theta^{\#}\right\|^2 - n^{-2/3}M_n^2 \\
&\geq (C - \epsilon)\left\|\hat{\theta}_n - \theta^{\#}\right\|^2 - n^{-2/3}M_n^2 \\
\implies \left\|\hat{\theta}_n - \theta^{\#}\right\|^2 &\leq (C - \epsilon)^{-1}n^{-2/3}M_n^2 + o_P\left(n^{-2/3}\right) = O_P\left(n^{-2/3}\right).
\end{aligned}$$

It follows that $n^{1/3}\left(\hat{\theta}_n - \theta^{\#}\right) = O_P(1)$.

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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 $W_n - W = O_P\left(n^{-1/3}\right)$,

$$\begin{aligned}
n^{2/3}\hat{Q}_4\left(\theta^{\#}+n^{-1/3}h\right) &= \frac{1}{2}n^{2/3}g\left(\theta^{\#}+n^{-1/3}h\right)'(W_n-W)\,g\left(\theta^{\#}+n^{-1/3}h\right)\\
&\quad + n^{2/3}g\left(\theta^{\#}+n^{-1/3}h\right)'(W_n-W)\,\pi\left(\theta^{\#}\right)\\
&= n^{2/3}O_P\left(\frac{\left\|n^{-1/3}h\right\|^2}{n^{1/3}}\right)\\
&\quad + \left(n^{1/3}\left\{g\left(\theta^{\#}\right)' + h'G'n^{-1/3}\right\} + o_P(1)\right)n^{1/3}(W_n-W)\,\pi\left(\theta^{\#}\right)\\
&= h'G'n^{1/3}(W_n-W)\,\pi\left(\theta^{\#}\right) + o_P(1) \rightsquigarrow h'G'\mathcal{W}_0,\\
n^{2/3}\hat{Q}_5\left(\theta^{\#}+n^{-1/3}h\right) &= n^{2/3}g\left(\theta^{\#}+n^{-1/3}h\right)'(W_n-W)\left(\hat{\pi}_n\left(\theta^{\#}\right)-\pi\left(\theta^{\#}\right)\right)\\
&\quad + g\left(\theta^{\#}+n^{-1/3}h\right)'(W_n-W)\,n^{2/3}\hat{\eta}_n\left(\theta^{\#}+n^{-1/3}h\right)\\
&\quad + \left(\hat{\pi}_n\left(\theta^{\#}\right)-\pi\left(\theta^{\#}\right)\right)'(W_n-W)\,n^{2/3}\hat{\eta}_n\left(\theta^{\#}+n^{-1/3}h\right)\\
&= n^{2/3}O_P\left(\frac{\left\|n^{-1/3}h\right\|}{n^{5/6}}\right) + O_P\left(\frac{\left\|n^{-1/3}h\right\|}{n^{1/3}}\right)O_P(1) + O_P\left(n^{-5/6}\right)O_P(1)\\
&= o_P(1),\\
n^{2/3}\hat{Q}_6\left(\theta^{\#}+n^{-1/3}h\right) &= \pi\left(\theta^{\#}\right)'(W_n-W)\,n^{2/3}\hat{\eta}_n\left(\theta^{\#}+n^{-1/3}h\right)\\
&\quad + \frac{1}{2}\hat{\eta}_n\left(\theta^{\#}+n^{-1/3}h\right)'(W_n-W)\,n^{2/3}\hat{\eta}_n\left(\theta^{\#}+n^{-1/3}h\right)\\
&= O_P\left(n^{-1/3}\right)O_P(1) + O_P\left(n^{-2/3}\right)O_P\left(n^{-1/3}\right)O_P(1)\\
&= o_P(1).
\end{aligned}$$

By assumption,

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

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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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

$$n^{2/3}\hat{Q}_4\left(\theta^{\#}+n^{-1/3}h\right) = n^{2/3}o_P\left(\frac{\left\|n^{-1/3}h\right\|^2}{n^{1/3}}\right) + n^{2/3}o_P\left(\frac{\left\|n^{-1/3}h\right\|}{n^{1/3}}\right)$$
$$= o_P\left(n^{-1/3}\right) + o_P(1) = o_P(1). \qquad \square$$

Proofs for Theorems 4 and A.4. Equation 4.1 implies that for $\hat{h}^* = n^{\gamma}\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$,

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

We already showed in Theorem 2 that

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

Consistency of W_n for W implies that

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

We also showed in Theorem 2 that

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

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 $o_P^*(1)$:

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

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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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

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

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

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

For misspecified nonsmooth models with $\rho = 1$, $\gamma = 1/2$, we already showed in Theorem A.3

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

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,

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

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

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

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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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:

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

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

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

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

and $\hat{\eta}_n(\theta) = (P_n - P)g(\cdot,\theta)$, $\hat{g}(\theta) = P_n g(\cdot,\theta)$, and $g(\theta) = Pg(\cdot,\theta)$. 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

$$|\hat{Q}_3(\theta)| \le \epsilon\|\theta - \theta^\#\|^2 + n^{-1}M_{n,3}^2.$$

The first, third, and fourth 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 second term in $\hat{Q}_2(\theta)$, for n large enough, $\forall\epsilon > 0$, $\exists M_{n,22} = O_P(1)$ such that

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

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

Because $Q_1(\theta)$ achieves the minimal value of 0 at $\theta^\#$, the Taylor expansion of $Q_1(\theta)$ around $\theta^\#$ is $Q_1(\theta) = Q_1\left(\theta^\#\right) + \left(\theta - \theta^\#\right)'\frac{\partial Q_1(\theta^\#)}{\partial\theta} + \frac{1}{2}\left(\theta - \theta^\#\right)'\frac{\partial^2 Q_1(\theta^\#)}{\partial\theta\partial\theta'}\left(\theta - \theta^\#\right) + o\left(\|\theta - \theta^\#\|^2\right) = \frac{1}{2}\left(\theta - \theta^\#\right)'\left(\bar{H} + o(1)\right)\left(\theta - \theta^\#\right)$ since $\frac{\partial Q_1(\theta^\#)}{\partial\theta} = G'Wg\left(\theta^\#\right) +$

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

$G'W\pi\left(\theta^{\#}\right)=0$ and $\frac{\partial^2 Q_1(\theta^{\#})}{\partial\theta\partial\theta'}=\bar{H}$. Because $\bar{H}$ is positive definite, there exists $C>0$ and a small enough neighborhood of $\theta^{\#}$ such that $Q_1(\theta)\geq C\left\|\theta-\theta^{\#}\right\|^2$. By consistency of $\hat{\theta}_n$ for $\theta^{\#}$, with probability approaching 1, $Q_1\left(\hat{\theta}_n\right)\geq C\left\|\hat{\theta}_n-\theta^{\#}\right\|^2$. Then,

$$Q_1\left(\hat{\theta}_n\right)+\hat{Q}_2\left(\hat{\theta}_n\right)+\hat{Q}_3\left(\hat{\theta}_n\right)=\hat{Q}_n\left(\hat{\theta}_n\right)-\hat{Q}_n\left(\theta^{\#}\right)\leq\hat{Q}_n\left(\hat{\theta}_n\right)-\inf_{\theta\in\Theta}\hat{Q}_n(\theta)\leq o_P\left(n^{-1}\right).$$

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

$$\begin{aligned}o_P\left(n^{-1}\right)&\geq Q_1\left(\hat{\theta}_n\right)-\epsilon\left\|\hat{\theta}_n-\theta^{\#}\right\|^2-n^{-1}M_n^2\\&\geq(C-\epsilon)\left\|\hat{\theta}_n-\theta^{\#}\right\|^2-n^{-1}M_n^2\\\implies\left\|\hat{\theta}_n-\theta^{\#}\right\|^2&\leq(C-\epsilon)^{-1}n^{-1}M_n^2+o_P\left(n^{-1}\right)=O_P\left(n^{-1}\right).\end{aligned}$$

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

Next, $\hat{h}=\sqrt{n}\left(\hat{\theta}_n-\theta^{\#}\right)=\arg\min_h n\hat{Q}_n\left(\theta^{\#}+n^{-1/2}h\right)$. 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 $\hat{h}\rightsquigarrow\arg\min_{h\in\mathbb{R}^d}\left\{\pi\left(\theta^{\#}\right)'W\mathcal{Z}_{0,1}(h)+h'G'WU_0+\frac{1}{2}h'\bar{H}h\right\}$ 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 $Q_1\left(\theta^{\#}+n^{-1/2}h\right)=Q_1\left(\theta^{\#}\right)+n^{-1/2}h'\frac{\partial Q_1(\theta^{\#})}{\partial\theta}+\frac{1}{2}n^{-1}h'\frac{\partial^2 Q_1(\theta^{\#})}{\partial\theta\partial\theta'}h+o\left(n^{-1}\right)$, $nQ_1\left(\theta^{\#}+n^{-1/2}h\right)=\frac{1}{2}h'\bar{H}h+o(1)$.

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 first, third, and fourth 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 second term, we can Taylor expand $g\left(\theta^{\#}+n^{-1/2}h\right)$ around $\theta^{\#}$:

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

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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

the Lindeberg–Feller CLT implies that $S_n(h) \equiv \begin{pmatrix} \pi\left(\theta^{\#}\right)' W n\hat{\eta}_n\left(\theta^{\#}+n^{-1/2}h\right) \\ h'G'W\sqrt{n}\left(\hat{\pi}_n\left(\theta^{\#}\right)-\pi\left(\theta^{\#}\right)\right) \end{pmatrix}$ 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 $\Sigma_1(s,t) = \lim_{\alpha\to\infty} \alpha^2 Pg\left(\cdot,\theta^{\#}+\frac{s}{\alpha}\right) g\left(\cdot,\theta^{\#}+\frac{t}{\alpha}\right)'$, 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 $h'G'W\sqrt{n}\left(\hat{\pi}_n\left(\theta^{\#}\right)-\pi\left(\theta^{\#}\right)\right)$ 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 $j = 1,\ldots,m$,

$$nE\sup_{\mathscr{D}(n)} \left|P_n d_j - P d_j\right| = o(1), \tag{A.2}$$

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 Lemma 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,

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

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

For each $K > 0$, write $E\sup_{\mathscr{D}(n)} P_n d_j^2 \le EP_n \sup_{\mathscr{D}(n)} d_j^2 1\{D_n > K\} + KE\sup_{\mathscr{D}(n)} P_n|d_j| \le EP_nD_n^2 1\{D_n > K\} + K\sup_{\mathscr{D}(n)} P|d_j| + KE\sup_{\mathscr{D}(n)} \left|P_n|d_j| - P|d_j|\right|$. 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 Lemma 3.1 of Kim and Pollard (1990) to show $KE\sup_{\mathscr{D}(n)} \left|P_n|d_j| - P|d_j|\right| < Kn^{-\frac{1}{2}}J(1)\sqrt{PD_n^2} = O(n^{-1}\delta_n) = o(n^{-1})$.

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

Therefore, $En \sup_{\mathscr{D}(n)} P_n d_j^2 = o(1)$. It follows that

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

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,

$$n\left(Q_1+\hat{Q}_2+\hat{Q}_3\right)\left(\theta^{\#}+n^{-1/2}h\right) \rightsquigarrow \pi\left(\theta^{\#}\right)' W\mathcal{Z}_{0,1}(h) + h'G'WU_0 + \frac{1}{2}h'\bar{H}h.$$

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 h 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 A.3. 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

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

We already showed in Theorem A.1 that $\forall\epsilon > 0$, there exists $M_n = O_P(1)$ such that $|\hat{Q}_2(\theta) + \hat{Q}_3(\theta)| \le \epsilon\|\theta - \theta^{\#}\|^2 + n^{-1}M_n^2$.

A modified version of Kim and Pollard (1990, Lem. 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

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

The second and third 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 first term in $\hat{Q}_5(\theta)$ can also be bounded by, for

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

some $M_{n,52} = O_P(1)$ and $\forall \epsilon > 0$,

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

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

$$\begin{aligned}
\hat{Q}_4(\theta) &= \hat{Q}_4\left(\theta^{\#}\right) + \left(\theta - \theta^{\#}\right)' \frac{\partial \hat{Q}_4\left(\theta^{\#}\right)}{\partial \theta} + \frac{1}{2}\left(\theta - \theta^{\#}\right)' \frac{\partial^2 \hat{Q}_4\left(\theta^{\#}\right)}{\partial \theta \partial \theta'}\left(\theta - \theta^{\#}\right) + o_p\left(\|\theta - \theta^{\#}\|^2\right) \\
&= \frac{1}{2}\left(\theta - \theta^{\#}\right)'\left(G'(W_n - W)G + \sum_{j=1}^{m}\sum_{k=1}^{m}(W_{n,jk} - W_{jk})\pi_k\left(\theta^{\#}\right)H_j + o_p(1)\right)\left(\theta - \theta^{\#}\right) \\
&\quad + O_P\left(\frac{\|\theta - \theta^{\#}\|}{n^{\gamma}}\right) \\
&\le \epsilon\|\theta - \theta^{\#}\|^2 + n^{-2\gamma}M_{n,4}^2.
\end{aligned}$$

Then $\forall \epsilon > 0$, there exists $M_n = O_P(1)$ such that $|\hat{Q}_2(\theta) + \hat{Q}_3(\theta) + \hat{Q}_4(\theta) + \hat{Q}_5(\theta) + \hat{Q}_6(\theta)| \le \epsilon\|\theta - \theta^{\#}\|^2 + n^{-1}M_n^2$. We already showed that there exists some $C > 0$ such that almost surely $Q_1\left(\hat{\theta}_n\right) \ge C\left\|\hat{\theta}_n - \theta^{\#}\right\|^2$. Then,

$$Q_1\left(\hat{\theta}_n\right) + \hat{Q}_2\left(\hat{\theta}_n\right) + \hat{Q}_3\left(\hat{\theta}_n\right) + \hat{Q}_4\left(\hat{\theta}_n\right) + \hat{Q}_5\left(\hat{\theta}_n\right) + \hat{Q}_6\left(\hat{\theta}_n\right) \le \hat{Q}_n\left(\hat{\theta}_n\right) - \inf_{\theta \in \Theta}\hat{Q}_n(\theta) \le o_P\left(n^{-1}\right).$$

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

$$\begin{aligned}
o_P\left(n^{-1}\right) &\ge Q_1\left(\hat{\theta}_n\right) - \epsilon\left\|\hat{\theta}_n - \theta^{\#}\right\|^2 - n^{-1}M_n^2 \\
&\ge (C - \epsilon)\left\|\hat{\theta}_n - \theta^{\#}\right\|^2 - n^{-1}M_n^2 \\
\implies \left\|\hat{\theta}_n - \theta^{\#}\right\|^2 &\le (C - \epsilon)^{-1}n^{-1}M_n^2 + o_P\left(n^{-1}\right) = O_P\left(n^{-1}\right).
\end{aligned}$$

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

We already showed in Theorem A.1 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 $W_n - W = O_P\left(n^{-1/2}\right)$,

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

$$
\begin{aligned}
&+\left(\sqrt{n}\left\{g\left(\theta^{\#}\right)'+h'G'n^{-1/2}\right\}+o_P(1)\right)\sqrt{n}(W_n-W)\pi\left(\theta^{\#}\right)\\
&=h'G'\sqrt{n}(W_n-W)\pi\left(\theta^{\#}\right)+o_P(1)\\
&\rightsquigarrow h'G'\Phi_0\pi\left(\theta^{\#}\right),
\end{aligned}
$$

$$
\begin{aligned}
n\hat{Q}_5\left(\theta^{\#}+n^{-1/2}h\right)&=ng\left(\theta^{\#}+n^{-1/2}h\right)'(W_n-W)\left(\hat{\pi}_n\left(\theta^{\#}\right)-\pi\left(\theta^{\#}\right)\right)\\
&+g\left(\theta^{\#}+n^{-1/2}h\right)'(W_n-W)n\hat{\eta}_n\left(\theta^{\#}+n^{-1/2}h\right)\\
&+\left(\hat{\pi}_n\left(\theta^{\#}\right)-\pi\left(\theta^{\#}\right)\right)'(W_n-W)n\hat{\eta}_n\left(\theta^{\#}+n^{-1/2}h\right)\\
&=nO_P\left(\frac{\left\|n^{-1/2}h\right\|}{n}\right)+O_P\left(\frac{\left\|n^{-1/2}h\right\|}{\sqrt{n}}\right)O_P(1)+O_P\left(n^{-1}\right)O_P(1)\\
&=O_P\left(n^{-1/2}\right)+O_P\left(n^{-1}\right)+O_P\left(n^{-1}\right)\\
&=o_P(1),\\
n\hat{Q}_6\left(\theta^{\#}+n^{-1/2}h\right)&=\pi\left(\theta^{\#}\right)'(W_n-W)n\hat{\eta}_n\left(\theta^{\#}+n^{-1/2}h\right)\\
&+\frac{1}{2}\hat{\eta}_n\left(\theta^{\#}+n^{-1/2}h\right)'(W_n-W)n\hat{\eta}_n\left(\theta^{\#}+n^{-1/2}h\right)\\
&=O_P\left(n^{-1/2}\right)O_P(1)+O_P\left(n^{-1}\right)O_P\left(n^{-1/2}\right)O_P(1)\\
&=o_P(1).
\end{aligned}
$$

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

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

Therefore, by the Lindeberg–Feller CLT and stochastic equicontinuity arguments similar to those in Theorem A.1,

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

as a process indexed by h in the product space of locally bounded functions $\left\{\mathbf{B}_{\mathrm{loc}}\left(\mathbb{R}^d\right)\right\}^3$ equipped with the topology of uniform convergence on compacta. By Assumption 4,

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

$\sqrt{n}(W_n - W) = \sqrt{n}(P_n - P)\phi\left(\cdot, \theta_1^{\#}\right) + o_P(1)$; therefore,

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

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

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

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

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

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

A.3. More Details for Examples

IV quantile regression. We show that 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 which decay at the linear rate:

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

Using monotonicity of $q(\cdot)$, we can bound the second moment of the envelope function by considering all possible ways of adding or subtracting R from each coordinate of $\theta^{\#}$.

$$\begin{aligned} PG_R^2 &\le E\left[\left|z_{ij}\right|^2 E\left[\sup_{\|\theta-\theta^{\#}\|\le R}\left|1\left(y_i \le q\left(x_i'\theta^{\#}\right)\right) - 1\left(y_i \le q\left(x_i'\theta\right)\right)\right| x_i, z_i\right]\right] \\ &\le E\left[\left|z_{ij}\right|^2 \sum_{\omega\in\{-1,1\}^d}\left(P\left(q\left(x_i'\left(\theta^{\#} - \omega R\right)\right) \le y_i \le q\left(x_i'\theta^{\#}\right) \middle| x_i, z_i\right)\right)\right] \end{aligned}$$

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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

For the third inequality, we applied mean-value expansions to the probabilities since we assumed that $F_{y|x,z}$ is absolutely continuous and $q(\cdot)$ is twice differentiable.

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

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

The last line follows from the fact that conditional on x_i, z_i, $\tau - 1\left(y_i \le q\left(x_i'\theta_1^\#\right)\right)$ 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,

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

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 $F_{y|x,z}\left(q\left(x_i'\theta_1^\#\right)\right) = \tau$.

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

The estimated weighting matrix is

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

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 a subgradient (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 matrix: $D_n\xrightarrow{p}D_0\equiv\left.\frac{\partial W(\theta_1)\pi(\theta^\#)}{\partial\theta_1}\right|_{\theta_1=\theta_1^\#}$. If the one-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 $W_n=\left(\tau(1-\tau)\frac{1}{n}\sum_i z_iz_i'\right)^{-1}$ 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 $\hat{\theta}_1$.

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

Simulated method of moments. 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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

Using similar arguments as in the previous example,

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

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

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

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 a subgradient (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 matrix: $D_n \xrightarrow{p} D_0 \equiv \left.\frac{\partial W(\theta_1)\pi\left(\theta^\#\right)}{\partial\theta_1}\right|_{\theta_1=\theta_1^\#}$. If the one-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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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 $\hat{\theta}_1$.

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

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

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

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

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

The estimated weighting matrix is

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

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

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

If the one-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 a subgradient of $vech\left(W_n\left(\theta_1\right)\right)$ evaluated at $\theta_1^{\#}$ that is consistent for the population derivative 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:

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

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 $E\left[\| vech\left(g\left(y_{i\cdot}, y_{i\cdot}\right)\right)\|\right] < \infty$ and $E\left[\| vech\left(g\left(y_{i\cdot}, y_{j\cdot}\right)\right)\|^2\right] < \infty$. 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, Mentch, and Hooker (2021).

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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 one-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 a subgradient of $vech\left(W_n^*\left(\theta_1\right)\right)$ evaluated at $\theta_1^{\#}$ that is consistent for the population derivative 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 a subgradient of $vech\left(W_n\left(\theta_1\right)\right)$ evaluated at $\theta_1^{\#}$, and it 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).

A.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

$$\pi_1(\cdot, \theta) = y_i, \quad \pi_2(\cdot, \theta) = z_i - \theta.$$

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

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 one-step GMM estimator (using the identity weighting matrix) is $\hat{\theta}_1 = \bar{z}$ and the two-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 one-step GMM estimator is numerically identical to the standard bootstrap one-step GMM estimator. We can see this by noting that $\left(P_n^* - P_n\right)\left(\pi(\cdot,\theta) - \pi\left(\cdot,\hat{\theta}_n\right)\right) = 0$, $H = 0$, $G = [0; -1]$, $G'G = 1$, and therefore

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

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

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

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:

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

TABLE A.1. Empirical coverage frequencies for $\delta = 1$.

n	200	800	1,600	3,200	6,400	9,600
Rate-adaptive	0.948	0.953	0.948	0.952	0.953	0.953
	(0.343)	(0.170)	(0.120)	(0.085)	(0.060)	(0.049)
Standard	0.944	0.952	0.948	0.953	0.952	0.953
	(0.339)	(0.170)	(0.120)	(0.085)	(0.060)	(0.049)
MR	0.944	0.951	0.949	0.952	0.952	0.953
	(0.339)	(0.170)	(0.120)	(0.085)	(0.060)	(0.049)

TABLE A.2. Empirical coverage frequencies for $\delta = 0.1$.

n	200	800	1,600	3,200	6,400	9,600
Rate-adaptive	0.950	0.947	0.953	0.950	0.951	0.952
	(0.240)	(0.121)	(0.085)	(0.060)	(0.043)	(0.035)
Standard	0.950	0.947	0.952	0.950	0.951	0.952
	(0.241)	(0.121)	(0.085)	(0.060)	(0.043)	(0.035)
MR	0.951	0.947	0.952	0.950	0.950	0.952
	(0.241)	(0.121)	(0.085)	(0.060)	(0.043)	(0.035)

$\left[\hat{\theta}_2 - d_{0.975}, \hat{\theta}_2 - d_{0.025}\right]$, 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's (2014) nominal 95% MR bootstrap confidence intervals. We use $B = 5{,}000$ bootstrap iterations and $R = 5{,}000$ Monte Carlo simulations.

From Tables A.1 and A.2, 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 the 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.

References

Armstrong, T. B., & Kolesár, M. (2021). Sensitivity analysis using approximate moment condition models. *Quantitative Economics*, 12(1), 77–108.

Berkowitz, D., Caner, M., & Fang, Y. (2012). The validity of instruments revisited. *Journal of Econometrics*, 166(2), 255–266.

Bickel, P. J., & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9(6), 1196–1217.

Bonhomme, S., & Weidner, M. (2022). Minimizing sensitivity to model misspecification. *Quantitative Economics*, 13(3), 907–954.

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

Cattaneo, J., & Nagasawa, K. (2020). Bootstrap-based inference for cube root asymptotics. *Econometrica*, 88(5), 2203–2219.

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

Cattaneo, M. D., & Jansson, M. (2018). Kernel-based semiparametric estimators: Small bandwidth asymptotics and bootstrap consistency. *Econometrica*, 86(3), 955–995.

Cattaneo, M. D., & Jansson, M. (2022) Average density estimators: Efficiency and bootstrap consistency. *Econometric Theory*, 38(6), 1140–1174.

Cheng, X., Liao, Z., & Shi, R. (2019). On uniform asymptotic risk of averaging GMM estimators. *Quantitative Economics*, 10(3), 931–979.

Chernozhukov, V., & Hansen, C. (2005). An IV model of quantile treatment effects. *Econometrica*, 73(1), 245–261.

Chernozhukov, V., & Hong, H. (2003) An MCMC approach to classical estimation. *Journal of Econometrics*, 115(2), 293–346.

Creel, M., Gao, J., Hong, H., & Kristensen, D. (2015). *Bayesian indirect inference and the ABC of GMM. PubMed Central*. https://doi.org/10.48550/arXiv.1512.07385

Dudley, R. M. (1987). Universal Donsker classes and metric entropy. *Annals of Probability*, 15(4), 1306–1326.

Giurcanu, M., & B. Presnell (2018). Bootstrap inference for misspecified moment condition models. *Annals of the Institute of Statistical Mathematics*, 70(3), 605–630.

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

Hall, A. R., & Inoue, A. (2003). The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics*, 114(2), 361–394.

Hall, P., & Horowitz, J. L. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, 64(4), 891–916.

Hansen, B. E., & Lee, S. (2021). Inference for iterated GMM under misspecification. *Econometrica*, 89(3), 1419–1447.

Hong, H., & Li, J. (2020). The numerical bootstrap. *Annals of Statistics*, 48(1), 397–412.

Honoré, B. E., & Hu, L. (2004a). Estimation of cross sectional and panel data censored regression models with endogeneity. *Journal of Econometrics*, 122(2), 293–316.

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

Jun, S. J., Pinkse, J., & Wan, Y. (2015). Classical Laplace estimation for $\sqrt[3]{n}$-consistent estimators: Improved convergence rates and rate-adaptive inference. *Journal of Econometrics*, 187(1), 201–216.

Kim, J., & Pollard, D. (1990). Cube root asymptotics. *Annals of Statistics*, 18(1), 191–219.

Kosorok, M. R. (2007). *Introduction to empirical processes and semiparametric inference*. Springer.

Lee, S. (2014). Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators. *Journal of Econometrics*, 178, 398–413.

Lehmann, E. L., & Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5), 995–1026.

Newey, W., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. Engle, & D. McFadden (Eds.), *Handbook of econometrics* (Vol. 4, pp. 2113–2241). North-Holland.

Pakes, A., & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5), 1027–1057.

Phillips, P. C. B. (2015). Halbert White Jr. memorial JFEC lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics*, 13(3), 521–555.

Politis, D., Romano, J., & Wolf, M. (1999). *Subsampling*. Springer Series in Statistics. Springer.

Pollard, D. (1989). Asymptotics via empirical processes. *Statistical Science*, 4(4), 341–354.

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

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics* (Vol. 162). John Wiley & Sons.

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press

van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics (Vol. 3). Cambridge University Press).

van der Vaart, A. W., & Wellner, J. (1996). *Weak convergence and empirical processes*. Springer.

Wellner, J. A. , & Zhan, Y. (1996). *Bootstrapping Z-estimators*. University of Washington Department of Statistics Technical Report, 308.

Zhou, Z., Mentch, L., & Hooker, G. (2021). V-statistics and variance estimation. *Journal of Machine Learning Research*, 22(287), 1–48.

https://doi.org/10.1017/S0266466623000385 Published online by Cambridge University Press