Econometric Theory, **32**, 2016, 187–242.
doi:10.1017/S0266466614000826

COMPARISON OF INFERENTIAL METHODS IN PARTIALLY IDENTIFIED MODELS IN TERMS OF ERROR IN COVERAGE PROBABILITY

FEDERICO A. BUGNI
Duke University

This paper considers the problem of coverage of the elements of the identified set in a class of partially identified econometric models with a prespecified probability. In order to conduct inference in partially identified econometric models defined by moment (in)equalities, the literature has proposed three methods: bootstrap, subsampling, and asymptotic approximation. The objective of this paper is to compare these methods in terms of the rate at which they achieve the desired coverage level, i.e., in terms of the rate at which the *error in the coverage probability* (ECP) converges to zero.

Under certain conditions, we show that the ECP of the bootstrap and the ECP of the asymptotic approximation converge to zero at the same rate, which is a faster rate than that of the ECP of subsampling methods. As a consequence, under these conditions, the bootstrap and the asymptotic approximation produce inference that is more precise than subsampling. A Monte Carlo simulation study confirms that these results are relevant in finite samples.

1. INTRODUCTION

This paper contributes to the literature on inference in partially identified econometric models defined by moment (in)equalities (i.e., inequalities and equalities). Consider an economic model with a parameter θ belonging to a parameter space Θ, whose main prediction is that the true value of θ, denoted by θ_0, satisfies a finite number of unconditional moment (in)equalities. This model is partially identified, i.e., the restrictions of the model do not necessarily restrict θ_0 to a single value, but rather they constrain it to belong to a certain set, called the identified set, denoted by Θ_I. The partially identified literature discusses several examples

This paper is a subset of the results of my job market paper and was previously circulated under the title: "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Elements of the Identified Set". I am indebted to my advisors, Joel Horowitz, Rosa Matzkin, and Elie Tamer for their guidance and support. I would like to thank the co-editor and two anonymous referees for several comments and suggestions that have significantly improved this paper. I am also thankful for the many helpful comments received from participants at the various conferences and seminars where this work has been presented. Erik Vogt and Takuya Ura provided excellent research assistance. Any and all errors are my own. Address correspondence to Federico Bugni, Department of Economics, Duke University, 213 Social Sciences, Box 90097, Durham, NC, 27708; phone: 919-660-1887; e-mail: federico.bugni@duke.edu.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

of economic models that satisfy this structure, such as selection problems, missing data, or multiplicity of equilibria (see, e.g., Manski, 1995; Tamer, 2003).

In this paper, we are interested in constructing a confidence set, denoted by $C_n(1-\alpha)$, that (asymptotically) covers each of the elements of the identified set with a minimum prespecified probability of $(1-\alpha)$, referred to as the desired confidence level.[1] In precise terms, the coverage objective is given by

$$\inf_{\theta \in \Theta_I} \liminf_{n \to \infty} P(\theta \in C_n(1-\alpha)) \geq (1-\alpha). \tag{1.1}$$

If possible, it is desirable to consider a confidence set that satisfies the coverage objective in equation (1.1) with equality. The reason for this is that if a confidence set does not satisfy the coverage objective with equality, then it might be possible to consider a smaller confidence set that still satisfies the coverage objective and has an associated hypothesis test with strictly higher asymptotic power.[2]

The literature has considered several inferential procedures to conduct inference for partially identified models defined by unconditional moment (in)equalities. In their seminal contribution, Chernozhukov, Hong, and Tamer (2007) (henceforth, referred to as CHT) develop the criterion function approach and consider the use of subsampling and asymptotic approximations in order to implement their inference. Other contributions to this include Andrews, Berry, and Jia-Barwick (2004), Imbens and Manski (2004), Galichon and Henry (2006, 2013), Beresteanu and Molinari (2008), Romano and Shaikh (2008), Rosen (2008), Andrews and Guggenberger (2009), Stoye (2009), Andrews and Soares (2010), Bugni (2010), Canay (2010), Romano and Shaikh (2010), Andrews and Jia-Barwick (2012), Bontemps, Magnac, and Maurin (2012), Bugni, Canay, and Guggenberger (2012), Pakes, Porter, Ho, and Ishii (2014), and Romano, Shaikh, and Wolf (2014). More recently, there have been several contributions to inference in partially identified models defined by *conditional* moment (in)equalities, which include Kim (2008), Ponomareva (2010), Armstrong (2012, 2014), Chetverikov (2012), Andrews and Shi (2013), and Chernozhukov, Lee, and Rosen (2013).

The contribution of this paper is to compare three existing inferential procedures used to construct confidence sets that satisfy equation (1.1) in terms of the rate at which they achieve their desired coverage level. In particular, we compare inferential procedures based on the bootstrap, the asymptotic approximation, and subsampling in terms of the rate of convergence of the *error in the coverage probability* (henceforth, referred to as ECP), which is defined as the difference between the exact coverage level, $P(\theta \in C_n(1-\alpha))$, and the desired coverage level, $(1-\alpha)$. By definition, the rate of convergence of the ECP represents the rate at which the limiting coverage objective is achieved. If two inferential methods, A and B, achieve an asymptotic coverage level of $(1-\alpha)$, but A has a faster rate of convergence of the ECP than B, then, for all sufficiently large sample sizes, the coverage level of A will be closer to $(1-\alpha)$ than that of B. Therefore, among inferential methods that achieve the same limiting coverage level, the rate

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

of convergence of the ECP is an excellent way to compare the precision of the inference.

Our results are as follows. Under certain conditions, we show that the bootstrap procedure and the asymptotic approximation have the same rate of convergence of the ECP, which is smaller than the rate of convergence of the ECP obtained by using subsampling methods. From this result we can establish that, for all sufficiently large sample sizes, inference based on the bootstrap or the asymptotic approximation should be more precise than inference based on subsampling. Furthermore, our results also suggest that inference based on the bootstrap or asymptotic approximation should be of similar precision, i.e., the bootstrap does not provide asymptotic refinements.[3]

These results are analogous to those obtained in Bugni (2010) for a different inferential problem, namely, the problem of constructing a coverage set that covers the identified set itself with a minimum prespecified probability. The difference in the coverage objectives produces differences in the inferential methods, implying that our findings do not trivially follow from those in Bugni (2010). Besides the fact that the coverage objectives are different, Bugni (2010) also requires a much stronger set of assumptions than the ones considered in this paper. In particular, in order to obtain any result regarding the rate of convergence of the ECP, Bugni (2010) requires that the moment (in)equality model satisfies a special structure called "conditional separability", which can be restrictive in empirically relevant situations.[4] No assumptions of this nature are required in the present paper to obtain analogous results.

There are several papers in the literature that also compare existing inferential methods in partially identified moment (in)equality models in terms of their asymptotic properties. In order to better understand our contribution relative to the literature, we now describe some of these references in further detail. In this literature, inferential methods are typically constructed by comparing a test function with an associated critical value. Similar to Andrews and Soares (2010), the present paper fixes the test function and compares methods to compute critical values in terms of their asymptotic properties. However, instead of using asymptotic power as the comparison criterion, we use the rate of convergence of the ECP. In other words, instead of focusing on the ability to reject local sequences of alternative hypotheses, we concentrate on the rate at which each inferential method achieves the desired coverage objective under the null hypothesis. Unlike Andrews and Soares (2010) or the present paper, Canay (2010) fixes the critical value and compares several test functions in terms of their asymptotic power. More recently, Bugni et al. (2012) compare both test functions and critical values in terms of their robustness to local misspecification.

There is an important qualification to our findings. Since the important contributions of Imbens and Manski (2004) and Andrews and Guggenberger (2009), the literature on inference in partially identified econometric models has stressed the importance of deriving asymptotic results that hold uniformly in the space of parameters and probability distributions, as opposed to results that hold

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

pointwise for each parameter and probability distribution. The asymptotic coverage results obtained in this paper are admittedly pointwise in nature. Proving the uniform validity of the rate of convergence of the ECP would require showing that several key results in statistics hold uniformly in a relevant space of probability distributions.[5] To the best of our knowledge, these results are not currently available and their derivation represents a formidable additional task. For this reason, we consider this to be out of the scope of this paper.

The rest of the paper is organized as follows. Section 2 introduces our assumptions and provides an example of an econometric model where these assumptions are satisfied. In Sections 3, 4, and 5, we consider an inferential procedure, show its consistency in level, and analyze the rate of convergence of its ECP. Section 3 considers the bootstrap, Section 4 considers the asymptotic approximation, and Section 5 considers two subsampling procedures. Section 6 presents results of Monte Carlo simulations and Section 7 concludes the paper. The Appendix of the paper collects all the proofs and several intermediate results.

Throughout the paper, we use the following notation. For any $s \in \mathbb{N}$, $\mathbf{0}_s$ and $\mathbf{1}_s$ denote a column vector of size $s \times 1$ composed of zeros and ones, respectively, and $\mathbf{I}_s$ denotes the identity matrix of size $s \times s$. Also, for any $S \subset \Theta$, $Int(S)$ and ∂S denote the interior and boundary of S relative to the topology defined by Θ.

2. SETUP

We begin this section by relating the coverage objective in equation (1.1) to a relevant inference problem. In economic models, we are typically interested in testing whether a particular parameter value, denoted by $\bar{\theta}$, is a possible candidate for the true parameter value, θ_0. In other words, we are interested in testing

$$H_0 : \theta_0 = \bar{\theta} \quad \text{vs.} \quad H_1 : \theta_0 \neq \bar{\theta}. \tag{2.1}$$

In a partially identified model, every element in the identified set is a possible candidate for the true parameter value. As a consequence, the hypothesis testing problem in equation (2.1) can be equivalently reexpressed as follows:

$$H_0 : \bar{\theta} \in \Theta_I \quad \text{vs.} \quad H_1 : \bar{\theta} \notin \Theta_I. \tag{2.2}$$

Let $\alpha \in (0, 1)$ denote the significance level of the hypothesis test. By the duality between hypothesis tests and confidence sets, the inference problem in equation (2.1) (or, equivalently, equation (2.2)) can be addressed by constructing a confidence set $C_n(1-\alpha)$ that satisfies equation (1.1). In particular, the decision of rejecting any parameter value $\bar{\theta}$ outside of $C_n(1-\alpha)$ as a candidate for the true parameter value θ_0 implies a probability of type I error that, in the limit, does not exceed α.

We now introduce the assumptions that will be used throughout the paper. We begin with the assumptions about the probability space.

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

Assumption A1. Let $Z : \Omega \to \mathbb{R}^{d_Z}$ be a random vector in the probability space $(\Omega, \mathcal{B}, P)$. We observe an independent and identically distributed (henceforth, denoted by i.i.d.) sample of size $n \in \mathbb{N}$, denoted by $\mathcal{X}_n \equiv \{Z_i\}_{i=1}^n$.

Assumption A2. The parameter space, denoted by Θ, is a compact and convex subset of $\mathbb{R}^{d_\theta}$.

Assumption A3. The identified set is given by

$$\Theta_I \equiv \{ \theta \in \Theta : E[m(Z,\theta)] \geq \mathbf{0}_J \}, \tag{2.3}$$

where $m(\cdot,\theta) : \mathbb{R}^{d_Z} \to \mathbb{R}^J$ is a measurable function for every $\theta \in \Theta$.

Assumption A4. $V[m_j(Z,\theta)] \in (0,\infty)$ for every $j = 1,\ldots,J$.

Assumption A5. $m(Z,\theta)$ has finite fourth absolute moments.

Assumption A6. $m(Z,\theta)$ satisfies Cramer's condition, i.e.,

$$\limsup_{\|t\|\to\infty} E[\exp(it'm(Z,\theta))] < 1.$$

We comment briefly on some of these assumptions. Assumption A1 requires that the data are i.i.d. The consistency in level of the various inferential methods is based on the law of large numbers, the central limit theorem, and the law of iterated logarithm, whereas the derivation of the rate of convergence of the ECP requires additional results such as the Berry–Esseen theorem. The findings of this paper could be extended to non-i.i.d. settings provided that these asymptotic results hold and, of course, that the inferential procedures are adequately adapted to these settings.

Equation (2.3) in Assumption A3 defines the identified set as the intersection of finitely many unconditional moment inequalities. Of course, moment equalities are also allowed by combining pairs of moment inequalities. The current setup can also admit conditional moment inequalities as long as the conditioning covariates have finite support (see Example 2.1 for an illustration). Finally, notice that Assumption A3 allows the identified set to be empty, which would imply that the econometric model is misspecified.

Assumption A4 implies that the random vector $m(Z,\theta)$ has finite second moments, allowing us to apply the law of large numbers, the central limit theorem, and the law of iterated logarithm, which are the key to establish the consistency in level result. Assumption A5 strengthens the previous assumption and is only required to derive the rate of convergence of the ECP.

Assumption A6 imposes additional smoothness on the distribution of $m(Z,\theta)$. For an excellent discussion of this condition see, e.g., Hall (1992) or Horowitz (2001). As we will soon show, Assumptions A1–A5 are sufficient to establish the

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

rates of convergence of the ECP for the bootstrap and the asymptotic approximation. In order to establish an analogous result for subsampling, we employ an Edgeworth expansion for random sampling without replacement, developed by Babu and Singh (1985). Assumptions A5–A6 are sufficient conditions to establish such an expansion.

The confidence sets considered in this paper are constructed using the *criterion function approach* developed by CHT. The first step in this approach is to define a nonnegative function of the parameter space, referred to as *criterion function* and denoted by $Q(\theta)$, that satisfies

$$Q : \Theta \to \mathbb{R}_+ \quad \text{and} \quad Q(\theta) = 0 \iff \theta \in \Theta_I. \tag{2.4}$$

Let the properly normalized sample analog of $Q(\theta)$ be denoted by $Q_n(\theta)$. In order to satisfy equation (1.1), the criterion function approach proposes the following confidence set

$$C_n(1-\alpha) \equiv \{\theta \in \Theta : Q_n(\theta) \leq \hat{c}_n(\theta, 1-\alpha)\}, \tag{2.5}$$

where $\hat{c}_n(\theta, 1-\alpha)$ is a consistent approximation to the $(1-\alpha)$-quantile of the asymptotic distribution of $Q_n(\theta)$. Therefore, the criterion function approach translates the problem of constructing a confidence set into a problem of approximating the quantiles of the asymptotic distribution of $Q_n(\theta)$. This approximation problem is nonstandard precisely because the econometric model is partially identified. As we have mentioned in Section 1, we consider three methods to conduct this approximation: the bootstrap, the asymptotic approximation, and subsampling.

In order to implement the criterion function approach, the researcher needs to choose the criterion function. In the context of the moment (in)equalities model described in equation (2.3), the criterion function is as follows:

$$Q(\theta) = G(\{[E[m_j(Z,\theta)]]_-\}_{i=1}^J),$$

where $[x]_- \equiv |x| \times 1(x < 0)$ and $G : \mathbb{R}_+^J \to \mathbb{R}$ is such that $G(y) = 0$ if and only if $y = \mathbf{0}_J$. This implies that the criterion function is completely determined by the choice of the function G and, moreover, there is a wide range of possible choices for G. In order to obtain desirable asymptotic properties, it is convenient to impose restrictions on this choice. With this objective in mind, we consider the following assumptions.

Assumption CF1. The population criterion function is $Q(\theta) = G(\{[E[m_j(Z,\theta)]]_-\}_{j=1}^J)$, where $G : \mathbb{R}_+^J \to \mathbb{R}$ is a nonstochastic and nonnegative function that is strictly increasing in every coordinate, weakly convex, continuous, homogeneous of degree $\beta > 0$, and satisfies $G(y) = 0$ if and only if $y = \mathbf{0}_J$.

Assumption CF2. The population criterion function is $Q(\theta) = G(\{[E[m_j(Z,\theta)]]_-\}_{j=1}^J)$, where $G : \mathbb{R}_+^J \to \mathbb{R}$ is one of the following two functions:

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

(a) $G(x) = \sum_{j=1}^{J} \varpi_j x_j$ or (b) $G(x) = \max\{\varpi_j x_j\}_{j=1}^{J}$, where $\varpi \in \mathbb{R}_+^J$ is an arbitrary vector of positive constants.

Several comments are in order. Notice that the criterion function is a choice of the researcher and, in this sense, neither of these assumptions should be considered restrictive. Also, note that Assumption CF2 is a special case of Assumption CF1. While Assumption CF1 is sufficient to derive all of the formal results in the paper, i.e., consistency in level and rates of convergence of the ECP, the linearity implied by Assumption CF2 allows us to derive a sharper rate of convergence of the ECP for both the bootstrap and the asymptotic approximation. Finally, both of these assumptions imply that the properly normalized sample analog criterion function is given by:

$$Q_n(\theta) \;=\; G(\{[\sqrt{n}\bar{m}_{n,j}(\theta)]_-\}_{j=1}^{J}), \tag{2.6}$$

where $\bar{m}_{n,j}(\theta) \equiv n^{-1}\sum_{i=1}^{n} m_j(Z_i,\theta)$ for every $j = 1,\ldots,J$.

Remark 2.1. Assumptions CF1 and CF2 require the function G to be nonstochastic. In particular, the sample moment conditions in equation (2.6) are not allowed to be standardized, i.e., divided by sample standard deviation. As a result of this, the sample criterion function is not invariant to changes in the scale of the moment conditions, which is an important limitation relative to the existing literature.

In the electronic supplement to this paper (Bugni, 2014), we verify that it is possible to extend all of the results in this paper to standardized sample moment conditions provided that Assumption A5 is strengthened from finite fourth absolute moments to slightly more than sixth absolute moments. In addition to requiring stronger assumptions, dealing with standardization also requires the use of significantly longer technical arguments than those used in this paper. For the sake of brevity, we have decided to leave the issue of standardization out of the scope of the present paper.[6]

We conclude the section by considering a well-known economic example of a partially identified model, where we verify all of our assumptions.

Example 2.1 (Missing Data)
Suppose that our model predicts that $E[Y - f(X,\theta)|W] = 0$, where f is a known (measurable) function, $\theta \in \Theta \in \mathbb{R}^{d_\theta}$ is the parameter of interest, Y is a binary explained variable, X is a vector of explanatory variables, and W is a vector of conditioning exogenous variables. Typical examples of this setup are linear index models such as probit or logit. As it is usual in these models, Θ is assumed to be a compact and convex set, thus imposing Assumption A2.

The identification problem is caused because certain observations of the explained variable are missing (or censored). Let U denote the binary variable that equals one if Y is unobserved and zero otherwise. Also, assume that the support of

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

W is given by finitely many values $\{w_k\}_{k=1}^K$. Under these conditions, the identified set is given by:

$$\Theta_I = \left\{ \theta \in \Theta : \left\{ \begin{array}{c} E[(Y(1-U)+U-f(X,\theta)) \times 1[W=w_k]] \geq 0 \\ E[(f(X,\theta)-Y(1-U)) \times 1[W=w_k]] \geq 0 \end{array} \right\}_{k=1}^{K} \right\}. \quad \textbf{(2.7)}$$

Let us denote $Z \equiv (Y(1-U), U, W)$. We assume to observe an i.i.d. sample of $\{Z_i\}_{i=1}^n$, which implies Assumption A1. Notice that equation (2.7) is a particular case of equation (2.3) (for $J \equiv 2K$) if we define

$$\begin{aligned} m_{2k-1}(Z,\theta) &\equiv (Y(1-U)+U-f(X,\theta)) \times 1[W=w_k], \\ m_{2k}(Z,\theta) &\equiv (f(X,\theta)-Y(1-U)) \times 1[W=w_k], \end{aligned}$$

for $k = 1, \ldots, K$, and this verifies Assumption A3. Assumption A4 holds as well provided that

$$\begin{aligned} &V[Y(1-U)+U-f(X,\theta)|W=w_k] \in (0,\infty) \text{ and} \\ &\quad V[f(X,\theta)-Y(1-U)|W=w_k] \in (0,\infty) \end{aligned}$$

for all $k = 1, \ldots, K$. Since Y and W are discrete, Assumptions A5 and A6 are satisfied provided that $E[|f(X,\theta)|^4] < \infty$ and that $\{f(X,\theta)|W=w_k\}_{k=1}^K$ is continuously distributed for all $\theta \in \Theta$. Finally, Assumptions CF1 and CF2 are both satisfied if we choose the following criterion function

$$Q(\theta) = \frac{1}{2K} \sum\nolimits_{j=1}^{2K} [E[m_j(Z,\theta)]]_-.$$

Remark 2.2. In order to derive results on the rate of convergence of the ECP, Bugni (2010) requires the econometric model to satisfy the so-called "conditional separability" condition. Provided that $f(x,\theta)$ is a nontrivial function of x, Example 2.1 can be shown to satisfy the conditional separability condition if and only if the vector of explanatory variables X is a nonstochastic function of the vector of conditioning variables W. In this sense, the conditional separability condition in Bugni (2010) imposes severe constrains to the stochastic properties of the econometric model. In particular, it implies that the vector of explanatory variables X needs to be: (a) exogenous to the moment conditions and (b) discretely distributed with, at most, K support points.

In contrast, Example 2.1 has been shown to satisfy the assumptions of the present paper without requiring any conditional separability, i.e., we allow the vector of explanatory variables X to be stochastic, even after conditioning on the vector of covariates W. By not requiring this condition, we can capture the following empirically relevant scenarios. First, we allow the explanatory variables X to be endogenous to the moment conditions, while the conditioning covariates W represent an exogenous set of instruments. Second, we allow the explanatory variables X to have any kind of distribution (discrete, continuous, or even mixed), while only the conditioning covariates W are required to have finite

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

support. In this sense, the framework in this paper is considerably more general than that in Bugni (2010).

3. BOOTSTRAP

In this section, we consider a bootstrap procedure to construct confidence sets for each element of the identified set. This procedure can be understood as the bootstrap analog of the inferential method introduced by CHT and the Generalized Moment Selection (GMS) procedure of Andrews and Soares (2010).

As discussed in the previous section, we construct confidence sets as in equation (2.5), i.e.,

$$\hat{C}_n^B(1-\alpha) \;=\; \{\theta \in \Theta : Q_n(\theta) \leq \hat{c}_n^*(\theta, 1-\alpha)\}, \tag{3.1}$$

where $Q_n(\theta)$ is as in equation (2.6) and $\hat{c}_n^*(\theta, 1-\alpha)$ is a bootstrap approximation to the $(1-\alpha)$-quantile of its asymptotic distribution. The steps to compute $\hat{c}_n^*(\theta, 1-\alpha)$ are as follows:

1. Choose $\{\tau_n\}_{n\geq 1}$ to be a positive sequence such that $\tau_n/\sqrt{n} = o(1)$ and $\sqrt{\ln \ln n}/\tau_n = o(1)$, almost surely (henceforth, denoted by a.s.).
2. Repeat the following step many times. Construct a sample of size n by sampling randomly with replacement from the data. Denote these observations by $\{Z_i^*\}_{i=1}^n$ and let $\bar{m}_{n,j}^*(\theta) \equiv n^{-1}\sum_{i=1}^n m_j(Z_i^*, \theta)$ for $j = 1, \ldots, J$. Compute

$$Q_n^*(\theta) \;\equiv\; G(\{[\sqrt{n}(\bar{m}_{n,j}^*(\theta) - \bar{m}_{n,j}(\theta))]_- \times 1[\sqrt{n}\bar{m}_{n,j}(\theta) \leq \tau_n]\}_{j=1}^J), \tag{3.2}$$

3. $\hat{c}_n^*(\theta, 1-\alpha)$ is the $(1-\alpha)$-quantile of the distribution of $Q_n^*(\theta)$, simulated with arbitrary accuracy in the previous step.

In order to implement this procedure, we need to specify a sequence $\{\tau_n\}_{n\geq 1}$ that satisfies certain rate restrictions. The restrictions on the rate of the sequence $\{\tau_n\}_{n\geq 1}$ provide little guidance on how to choose this sequence in practice. For example, $\tau_n = \ln \ln n$, $\tau_n = \ln n$, and $\tau_n = n^\varphi$ for any $\varphi \in (0, 0.5)$ are all valid choices. As we will show later, this procedure is consistent in level and has the same rate of convergence regardless of the specific choice of the sequence $\{\tau_n\}_{n\geq 1}$. Thus, our asymptotic analysis does not provide a criterion for an "optimal" choice of $\{\tau_n\}_{n\geq 1}$. The experience drawn from Monte Carlo simulations seems to indicate that the finite sample performance of our inferential method does not depend critically on this choice.

The key to the consistency in level of this inference method is the bootstrap analog criterion function defined in equation (3.2). In particular, it is essential to consistency that we introduce (a) the recentering term (i.e., subtracting the sample average from the bootstrap sample average) and (b) the indicator function terms (i.e., $1[\sqrt{n}\bar{m}_{n,j}(\theta) \leq \tau_n]$ for $j = 1, \ldots, J$).[7] Due to these changes, our bootstrap procedure differs qualitatively from a bootstrap version of the subsampling

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

scheme proposed by CHT, which we refer to as "naive bootstrap". We show in Section A.2 of the Appendix that such a proposal would actually be inconsistent due to a "parameter on the boundary" problem (see Andrews, 2000).

Remark 3.1. With the exception of the standardization (see Remark 2.1), this bootstrap approximation coincides with the bootstrap version of the GMS procedure proposed by Andrews and Soares (2010) for a particular choice of the test function (modified method of moments) and a particular GMS function ($\varphi = \varphi_1$).

We now provide the main asymptotic results for the bootstrap approximation. These results are based on two representation theorems, which are stated and proved in the Appendix. In a first step, we show that $Q_n(\theta)$ has a certain asymptotic representation (Theorem A.1). In a second step, we establish that, conditional on the sample, the bootstrap approximation has an analogous asymptotic representation (Theorem A.2). Based on the similarities between these two representations, we can establish that the bootstrap approximation is consistent in level and, more importantly, we can derive the rate of convergence of its ECP. We begin with the consistency in level result.

THEOREM 3.1 (Consistency in level—Bootstrap). *Assume Assumptions A1–A3, CF1, and that θ satisfies Assumption A4.*

1. *If $\theta \in \partial\Theta_I$, then, for any $\alpha \in (0, 0.5)$, $\lim_{n\to\infty} P(\theta \in \hat{C}_n^B(1-\alpha)) = 1-\alpha$,*
2. *If $\theta \in Int(\Theta_I)$, then, for any $\alpha \in [0, 1]$, $\lim_{n\to\infty} P(\theta \in \hat{C}_n^B(1-\alpha)) = 1$.*

Theorem 3.1 describes the limiting coverage for elements in the identified set and implies that for any $\alpha \in (0, 0.5)$, $\hat{C}_n^B(1-\alpha)$ satisfies the coverage objective in equation (1.1) with equality.[8] Notice that Theorem 3.1 only requires finite second moments in Assumption A4, i.e., finite fourth moments in Assumption A5 are not necessary. We now move on to the main result of the section.

THEOREM 3.2 (Rates of Convergence of the ECP—Bootstrap). *Assume Assumptions A1–A3, CF1, and that θ satisfies Assumptions A4–A5.*

1. *If $\theta \in \partial\Theta_I$, then, for any $\alpha \in (0, 0.5)$, $P(\theta \in \hat{C}_n^B(1-\alpha)) = (1-\alpha) + O(n^{-1/2}\sqrt{\ln n})$. Moreover, if we add Assumption CF2, then the $O(n^{-1/2}\sqrt{\ln n})$ term can be replaced with $O(n^{-1/2})$.*
2. *If $\theta \in Int(\Theta_I)$, then, for any $\alpha \in [0, 1]$, $P(\theta \in \hat{C}_n^B(1-\alpha)) = 1 + O(n^{-1})$.*

Theorem 3.2 provides a rate at which the convergence result in Theorem 3.1 occurs. The coverage of interior points converges to one at a relatively fast rate, while the coverage of boundary points converges to the desired coverage level at a rate of $n^{1/2}/\sqrt{\ln n}$ (or $n^{1/2}$, under Assumption CF2).

It should be noticed that the rates of convergence in Theorem 3.2 do not depend on the threshold sequence $\{\tau_n\}_{n\geq 1}$. The intuition behind this result is as follows. The sequence $\{\tau_n\}_{n\geq 1}$ is used to determine which moment inequalities are binding (or not) by means of the indicator function terms $1[\sqrt{n}\bar{m}_{n,j}(\theta) \leq \tau_n]$ for all

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

$j = 1, \ldots, J$. According to our asymptotic analysis, this task can be completed at a relatively fast rate (in particular, faster than $n^{1/2}$). As a consequence, the particular choice of the sequence $\{\tau_n\}_{n\geq 1}$ has no effect on the rate of convergence of the ECP, as long as it satisfies the conditions in step 1.

To conclude the section, we draw a parallel between our results and those in Bugni (2010). It should be noted that Bugni (2010) deals with a different coverage problem, namely the construction of a confidence set for the identified set (rather than each of its elements). While the coverage problems are different, the actual rates of convergence of the ECP of the bootstrap approximation are the same. As explained in Remark 2.2, the substantial difference between the two results lies in the set of assumptions used. In particular, in order to derive the rates of convergence, Bugni (2010) requires a very strong set of conditions which include, for example, a special structure for the moment conditions in equation (2.3) referred to as "conditional separability". In contrast, the rates of convergence in Theorem 3.2 impose no additional assumptions on the structure of the moment inequalities.

4. ASYMPTOTIC APPROXIMATION

In this section, we consider an asymptotic approximation procedure to construct confidence sets for each element of the identified set. This inferential method is analogous to the one first considered by CHT and to the GMS procedure of Andrews and Soares (2010).

As in the previous section, we consider a confidence set of the form:

$$\hat{C}_n^{AA}(1-\alpha) = \{\theta \in \Theta : Q_n(\theta) \leq \hat{c}_n^{AA}(\theta, 1-\alpha)\}, \tag{4.1}$$

where $Q_n(\theta)$ is as in equation (2.6) and $\hat{c}_n^{AA}(\theta, 1-\alpha)$ is an asymptotic approximation to the $(1-\alpha)$-quantile of its asymptotic distribution. The steps to compute $\hat{c}_n^{AA}(\theta, 1-\alpha)$ are as follows:

1. Choose $\{\tau_n\}_{n\geq 1}$ to be a positive sequence such that $\tau_n/\sqrt{n} = o(1)$ and $\sqrt{\ln \ln n}/\tau_n = o(1)$ a.s.
2. Repeat the following step many times. Construct an i.i.d. sample $\{\zeta_i\}_{i=1}^n$ with $\zeta_i \sim N(0,1)$ for all $i = 1, \ldots, n$, and compute

$$Q_n^{AA}(\theta) \equiv G\left(\left\{\left[n^{-1/2}\sum_{i=1}^n \zeta_i(m_j(Z_i,\theta) - \bar{m}_{n,j}(\theta))\right]_- \times 1[\sqrt{n}\bar{m}_{n,j}(\theta) \leq \tau_n]\right\}_{j=1}^J\right). \tag{4.2}$$

3. $\hat{c}_n^{AA}(\theta, 1-\alpha)$ is the $(1-\alpha)$-quantile of the distribution of $Q_n^{AA}(\theta)$, simulated with arbitrary accuracy in the previous step.

Notice that this procedure is analogous to the bootstrap approximation considered in Section 3. The only difference is that the recentered bootstrap sample

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

average (i.e., $\sqrt{n}(\bar{m}^*_{n,j}(\theta) - \bar{m}_{n,j}(\theta))$) is replaced by a conditional normal random variable (i.e., $n^{-1/2}\sum_{i=1}^n \zeta_i(m_j(Z_i,\theta) - \bar{m}_{n,j}(\theta))$). Since the recentered bootstrap sample average has a conditionally asymptotic normal distribution (see, e.g., Bickel and Freedman, 1981), it is intuitive that these two approximation methods share the asymptotic properties. Furthermore, just like in the bootstrap procedure, the asymptotic approximation requires the choice of a threshold sequence $\{\tau_n\}_{n\geq 1}$ and involves the use of computation by simulation.

The consistency in level of the asymptotic approximation can be established using the same arguments as in the bootstrap approximation (see Theorem A.7 in the Appendix). The following result characterizes the rate of convergence of the ECP of the asymptotic approximation.

THEOREM 4.1 (Rates of Convergence of the ECP—AA). *Assume Assumptions A1–A3, CF1, and that θ satisfies Assumptions A4–A5.*

1. *If $\theta \in \partial\Theta_I$, then, for any $\alpha \in (0, 0.5)$, $P(\theta \in \hat{C}_n^{AA}(1-\alpha)) = (1-\alpha) + O(n^{-1/2}\sqrt{\ln n})$. Moreover, if we add Assumption CF2, then the $O(n^{-1/2}\sqrt{\ln n})$ term can be replaced with $O(n^{-1/2})$.*
2. *If $\theta \in Int(\Theta_I)$, then, for any $\alpha \in [0,1]$, $P(\theta \in \hat{C}_n^{AA}(1-\alpha)) = 1 + O(n^{-1})$.*

Theorem 4.1 reveals that the rate of convergence of the ECP is exactly like the one derived for the bootstrap procedure. As a consequence, the precision of the inference based on the bootstrap or the asymptotic approximation should be of similar quality, i.e., the bootstrap provides no asymptotic refinements. Theorem A.1 in the Appendix shows that the asymptotic distribution of our test statistic is not asymptotically pivotal and, thus, asymptotic refinements are not expected.[9] Furthermore, neither of the two procedures has a computational advantage, as both approximations involve the use of simulations. In summary, the bootstrap and the asymptotic approximation are similar in terms of asymptotic properties and computational complexity.

5. SUBSAMPLING

In this section, we analyze the asymptotic properties of two subsampling procedures to construct confidence sets for each element of the identified set. The first subsampling procedure, called subsampling 1, is a subsampling analog of the bootstrap procedure in Section 3. The second subsampling procedure, called subsampling 2, is a standard subsampling approximation to $Q_n(\theta)$.

As in the rest of the paper, we consider confidence sets of the form:

$$\hat{C}^{SS_j}_{b_n,n}(1-\alpha) \;=\; \{\theta \in \Theta : Q_n(\theta) \leq \hat{c}^{SS_j}_{b_n,n}(\theta, 1-\alpha)\}, \quad \text{for } j = 1,2, \tag{5.1}$$

where $Q_n(\theta)$ is as in equation (2.6) and $\hat{c}^{SS_j}_{b_n,n}(\theta, 1-\alpha)$ (with $j = 1,2$) is a subsampling approximation to the $(1-\alpha)$-quantile of its asymptotic distribution

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

using a subsampling size equal to b_n. The superscript of the subsampling critical value, SS_1 or SS_2, indicates whether we are considering subsampling 1 or 2, respectively.

5.1. Subsampling 1

This method was proposed by CHT. In this procedure, $\hat{c}^{SS_1}_{b_n,n}(\theta, 1-\alpha)$ is computed as follows:

1. Choose $\{b_n\}_{n\geq 1}$ to be a positive sequence such that $b_n \to \infty$ and $b_n/n = o(1)$ and choose $\{\tau_n\}_{n\geq 1}$ to be a positive sequence such that $\tau_n/\sqrt{n} = o(1)$ and $\sqrt{\ln \ln n}/\tau_n = o(1)$ a.s.
2. Repeat the following step many times. Construct a subsample of size b_n by sampling randomly without replacement from the data. Denote these observations by $\{Z^{SS}_i\}^{b_n}_{i=1}$ and let $\bar{m}^{ss}_{n,b_n,j}(\theta) \equiv b_n^{-1}\sum^{b_n}_{i=1} m_j(Z^{SS}_i, \theta)$ for $j = 1, \ldots, J$. Compute

$$Q^{SS_1}_{b_n,n}(\theta) \equiv G(\{[\sqrt{b_n}(\bar{m}^{ss}_{n,b_n,j}(\theta) - \bar{m}_{n,j}(\theta))]_- \times 1[\sqrt{n}\bar{m}_{n,j}(\theta) \leq \tau_n]\}^J_{j=1}). \tag{5.2}$$

3. $\hat{c}^{SS_1}_{b_n,n}(\theta, 1-\alpha)$ is the $(1-\alpha)$-quantile of the distribution of $Q^{SS_1}_{b_n,n}(\theta)$, simulated with arbitrary accuracy in the previous step.

Subsampling 1 is the result of considering the bootstrap procedure of Section 3 and replacing the use of bootstrap sampling (i.e., random samples of size n with replacement) with subsampling (i.e., random samples of size b_n without replacement). The consistency in level of subsampling 1 can be established in the same way as with the bootstrap approximation (see Theorem A.11 in the Appendix). The following result characterizes the rate of convergence of the ECP of subsampling 1.

THEOREM 5.1 (Rates of Convergence of the ECP—SS1). *Assume Assumptions 1–3, CF1, and that θ satisfies Assumptions A4–A6.*

1. *If $\theta \in \partial\Theta_I$, then, for any $\alpha \in (0, 0.5)$, $P(\theta \in \hat{C}^{SS_1}_{b_n,n}(1-\alpha)) = (1-\alpha) + O(b_n^{-1/2} + b_n/n)$.*
2. *If $\theta \in Int(\Theta_I)$, then, for any $\alpha \in [0, 1]$, $P(\theta \in \hat{C}^{SS_1}_{b_n,n}(1-\alpha)) = 1 + O(n^{-1})$.*

Theorem 5.1 describes the rate of convergence for elements of the identified set. As in the case with bootstrap or asymptotic approximation, the coverage of interior points converges to one at a relatively fast rate, which does not lead to an interesting comparison of the ECP across inferential procedures. On the other hand, the coverage of boundary points converges to the desired coverage level at a rate of $(b_n^{-1/2} + b_n/n)^{-1}$, which is different from the one obtained for the bootstrap or the asymptotic approximation. These differences will lead to

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

interesting comparisons of the ECP across the inference procedures. For this reason, we focus the rest of the section on the rate of convergence of the ECP for boundary points.[10] The derivation of the rate of convergence at boundary points relies in an Edgeworth expansion for subsampling statistics derived in Babu and Singh (1985).[11] Notice also that the rate of convergence of the ECP depends on the subsampling size b_n but does not depend on the threshold sequence $\{\tau_n\}_{n\geq 1}$. The lack of dependence on the threshold sequence is not a surprise given our findings for the bootstrap or the asymptotic approximation. As we have explained, the goal of the threshold sequence is to determine which moment inequalities are binding and which are not. This task can be achieved at a relatively fast rate and, in particular, faster than $(b_n^{-1/2} + b_n/n)^{-1}$.

Given that the rate of convergence at boundary points depends on the subsampling size, the researcher using subsampling 1 will be interested in choosing the subsampling size to maximize this rate. If so, the researcher should choose a subsampling size of $b_n = O(n^{2/3})$, which results in an ECP converging to zero at a rate of $n^{1/3}$. Recall from previous sections that the ECP of the bootstrap and the asymptotic approximation converges to zero at a rate of $n^{1/2}/\sqrt{\ln n}$ (or, $n^{1/2}$ under Assumption CF2). Provided that the rate of convergence of the ECP of subsampling 1 is sharp, i.e., it cannot be improved, this implies that for all sufficiently large sample sizes, inference based on the bootstrap or on asymptotic approximation should be more precise than inference based on subsampling 1.

The Appendix provides conditions under which the rate of convergence of the ECP of subsampling 1 is sharp. In order to understand the intuition behind these conditions, we now describe the Edgeworth expansion for subsampling 1 based on results in Babu and Singh (1985). Under our assumptions, we show in the Appendix that any parameter $\theta \in \partial\Theta_I$ satisfies the following equation

$$\underbrace{P(Q_{b_n,n}^{SS_1}(\theta) \leq h | \mathcal{X}_n)}_{\text{Subsampling 1 approx.}} - \underbrace{P(Q_n(\theta) \leq h)}_{\text{Exact distribution}}$$
$$= b_n^{-1/2}\Psi_1(h) + (b_n/n)\Psi_2(h) + o_p(b_n^{-1/2} + b_n/n),$$

uniformly in relevant values of h, where Ψ_1 and Ψ_2 are known real-valued functions (see Lemma A.2 in the Appendix). Up to a negligible term, the previous equation implies that the difference between the approximation of subsampling 1 and the exact distribution has two leading terms converging to zero at rates of $b_n^{1/2}$ and n/b_n, respectively. Provided that the expressions $\Psi_1(h)$ and $\Psi_2(h)$ have the same sign (i.e., both positive or both negative) then the ECP for subsampling 1 cannot converge to zero faster than $(b_n^{-1/2} + b_n/n)^{-1}$. We also show in the Appendix that $\Psi_2(h)$ has a known sign (see Lemma A.3) whereas the sign of $\Psi_1(h)$ depends on population parameters. Under plausible assumptions about these parameters, it is then possible that $\Psi_1(h)$ and $\Psi_2(h)$ share the sign, thus ensuring that the ECP cannot converge to zero faster than $(b_n^{-1/2} + b_n/n)^{-1}$. Our findings are summarized in the following result.

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

LEMMA 5.1. *Assume Assumptions A1–A3, CF1, that $\theta \in \partial\Theta_I$ satisfies Assumptions 4–6, and that $\alpha \in (0, 0.5)$. Then, for a nonstochastic sequence $\{a_n\}_{n\geq 1}$ with $a_n = O(b_n^{-1/2} + \sqrt{b_n/n})$,*

$$|P(\theta \in \hat{C}^{SS_2}_{b_n,n}(1-\alpha)) - (1-\alpha)| \leq a_n. \tag{5.3}$$

Furthermore, under certain additional conditions, the result in equation (5.3) is impossible for any nonstochastic sequence $\{a_n\}_{n\geq 1}$ with $a_n = o(b_n^{-1/2} + \sqrt{b_n/n})$.

According to Lemma 5.1, there are conditions under which the ECP of subsampling 1 converges to zero exactly at a rate of $(b_n^{-1/2} + b_n/n)^{-1}$, which cannot be faster than $n^{1/3}$. As a corollary, for all sufficiently large sample sizes, the bootstrap or the asymptotic approximation are more precise than subsampling 1.

5.2. Subsampling 2

This method was first considered by CHT and Romano and Shaikh (2008). In this procedure, $\hat{c}^{SS_2}_{b_n,n}(\theta, 1-\alpha)$ is computed as follows:

1. Choose $\{b_n\}_{n\geq 1}$ to be a positive sequence such that $b_n \to \infty$ and $(\ln \ln n) b_n/n = o(1)$ a.s.
2. Repeat the following step many times. Construct a subsample of size b_n by sampling randomly without replacement from the data. Denote these observations by $\{Z_i^{SS}\}_{i=1}^{b_n}$ and let $\bar{m}^{ss}_{n,b_n,j}(\theta) \equiv b_n^{-1} \sum_{i=1}^{b_n} m_j(Z_i^{SS}, \theta)$ for $j = 1, \ldots, J$. Compute

$$Q^{SS_2}_{b_n,n}(\theta) \equiv G(\{[\sqrt{b_n}\bar{m}^{ss}_{n,b_n,j}(\theta)]_-\}_{j=1}^J). \tag{5.4}$$

3. $\hat{c}^{SS_2}_{b_n,n}(\theta, 1-\alpha)$ is the $(1-\alpha)$-quantile of the distribution of $Q^{SS_2}_{b_n,n}(\theta)$, simulated with arbitrary accuracy in the previous step.

The main difference between subsampling 1 and subsampling 2 is the definition of the subsampling analog criterion function, i.e., equations (5.2) and (5.4). In particular, subsampling 1 defines the criterion function with the recentering and the indicator function terms, whereas subsampling 2 defines the criterion function without these terms. The other difference is that subsampling 2 requires slightly stronger restrictions on $\{b_n\}_{n\geq 1}$ than subsampling 1. These stronger conditions are used to obtain results on the rate of convergence of the ECP for the subsampling criterion function without the recentering terms.

The consistency in level of subsampling 2 can be established using the same arguments as in the bootstrap approximation (see Theorem A.16 in the Appendix). The following result characterizes the rate of convergence of the ECP of subsampling 2.

THEOREM 5.2 (Rates of Convergence of the ECP—SS2). *Assume Assumptions A1-A3, CF2, and that θ satisfies Assumptions A4–A6.*

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

1. *If* $\theta \in \partial\Theta_I$, *then, for any* $\alpha \in (0, 0.5)$, $P(\theta \in \hat{C}^{SS_2}_{b_n,n}(1-\alpha)) = (1-\alpha) + O(b_n^{-1/2} + \sqrt{(\ln\ln n) b_n/n})$.
2. *If* $\theta \in Int(\Theta_I)$, *then, for any* $\alpha \in [0, 1]$, $P(\theta \in \hat{C}^{SS_2}_{b_n,n}(1-\alpha)) = 1 + o(b_n^{-1/2})$.

Theorem 5.2 provides the rate of convergence for elements of the identified set. As with the rest of the inferential methods, the coverage of interior points converges to one at a relatively fast rate and so we focus on boundary points. The coverage for points on the boundary converges to $(1-\alpha)$ at rate of $(b_n^{-1/2} + \sqrt{(\ln\ln n) b_n/n})^{-1}$, which is admittedly a nonsharp rate. Nevertheless, since $b_n/n = o(1)$, the rate of convergence of the ECP of subsampling 2 is slower than the one obtained for subsampling 1. The following result addresses the possible nonsharpness of the rate of convergence derived in Theorem 5.2.

LEMMA 5.2. *Assume Assumptions A1–A3, CF1, that* $\theta \in \partial\Theta_I$ *satisfies Assumptions A4-A6, and that* $\alpha \in (0, 0.5)$. *Then, for a nonstochastic sequence* $\{a_n\}_{n\geq 1}$ *with* $a_n = O(b_n^{-1/2} + \sqrt{(\ln\ln n) b_n/n})$,

$$|P(\theta \in \hat{C}^{SS_2}_{b_n,n}(1-\alpha)) - (1-\alpha)| \leq a_n. \tag{5.5}$$

Furthermore, under certain additional conditions, the result in equation (5.5) is impossible for any nonstochastic sequence $\{a_n\}_{n\geq 1}$ *with* $a_n = o(b_n^{-1/2} + \sqrt{b_n/n})$.

As mentioned earlier, the rate of convergence of the ECP of subsampling 2 described in Theorem 5.2 is not sharp. Nevertheless, Lemma 5.2 reveals that, in general, the rate of convergence of the ECP of subsampling 2 cannot be faster than $(b_n^{-1/2} + \sqrt{b_n/n})^{-1}$, which is already slower than the rate of convergence of the ECP of subsampling 1. In other words, subsampling 2 cannot be a more precise inferential method than subsampling 1, which was already shown to be less precise than the bootstrap or the asymptotic approximation.

6. MONTE CARLO SIMULATIONS

We now present Monte Carlo simulations to evaluate the finite sample performance of these inferential methods. For a parameter space $\Theta = [-B, B]$ with $B > 0$, we consider the following designs

$$\begin{aligned} &\text{design 1:} \quad \Theta_{I,1} = \{\, \theta \in \Theta : E[Z_1] \leq \theta \cap E[Z_2] \leq \theta \,\}. \\ &\text{design 2:} \quad \Theta_{I,2} = \{\, \theta \in \Theta : E[Z_1] \leq \theta \leq E[Z_1] \,\}, \end{aligned}$$

where $E[Z_1] = E[Z_2] = 0$. The econometric model in each design can be expressed as the moment (in)equality model in equation (2.3), where $Z = (Z_1, Z_2) : \Omega \to \mathbb{R}^2$ and $m(z, \theta) : \mathbb{R}^2 \times \Theta \to \mathbb{R}^2$ is given by

$$\begin{aligned} &\text{design 1:} \quad m_1(z,\theta) = \theta - z_1, \quad m_2(z,\theta) = \theta - z_2, \\ &\text{design 2:} \quad m_1(z,\theta) = \theta - z_1, \quad m_2(z,\theta) = z_2 - \theta. \end{aligned}$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

In the first design, $\Theta_{I,1} = [0, B]$, i.e., the parameter of interest is not point identified and, in the second design, $\Theta_{I,2} = \{0\}$, i.e., the parameter of interest is point identified. In both designs, $\theta = 0$ is the only point in the boundary of the identified set. Our results suggest that the procedures considered in this paper only exhibit interesting differences for parameters on the boundary of the identified set. For this reason, we focus our simulation study on the empirical coverage rate of the boundary point $\theta = 0$.[12]

The data $\mathcal{X}_n = \{(Z_{1,i}, Z_{2,i})\}_{i=1}^n$ are an i.i.d. sample of size $n = 100$ from a distribution denoted by F (we consider four different distributions that are described below).[13] To implement our inference, we use the criterion function $Q(\theta) = (1/2)\sum_{j=1}^2 [E[m_j(Z, \theta)]]_-$, which satisfies Assumptions CF1 and CF2. In order to compute the relevant quantile, the second step of each approximation method is simulated $R = 2,000$ times. The bootstrap, the asymptotic approximation, and subsampling 1 depend on the threshold parameter τ_n and both subsampling procedures depend on the subsampling size b_n. We list all the values considered for these parameters in Table 1.

We briefly comment on these parameter choices. We consider four different data distributions F. The first three are a bivariate normal distribution with zero, positive, and negative correlation, respectively. These distributions have finite moments of all orders and thus satisfy all the assumptions in the paper. The fourth bivariate distribution has infinite fourth absolute moments, violating Assumption A4. This distribution will be used to understand the sensitivity of our results to the presence of fat tails.

We consider eight choices of the threshold parameter τ_n: $\ln\ln n$, $\sqrt{\ln n}$, $\ln n$, and $(\sqrt{n})^{c_\tau}$ for $c_\tau \in \{1/6, 1/3, 1/2, 2/3, 5/6\}$. This list includes $\tau_n = \sqrt{\ln n}$, which is the recommendation in Andrews and Soares (2010, p. 131) based on their simulation results. For $n = 100$, these choices imply a wide range of values between $\ln\ln 100 \approx 1.53$ and $100^{5/12} \approx 6.8$. We also consider eight choices for subsampling size b_n: the integer part of n^{c_b} for $c_b \in \{0.3, 0.4, 0.5, 0.6, 2/3, 0.7, 0.8, 0.9\}$. For $n = 100$, these choices also imply a wide range of values between 4 and 63. Asymptotically, all of these choices should produce consistent inference.

We conduct simulations for all the methods considered in this paper, i.e., the bootstrap, the asymptotic approximation, subsampling 1, and subsampling 2.

TABLE 1. Parameters used in the Monte Carlo simulations

F	$F_1: (Y_1, Y_2) \sim N(\mathbf{0}_2, \mathbf{I}_2)$
	$F_2: (Y_1, Y_2) \sim N(\mathbf{0}_2, (1, 0.5; 0.5, 1))$
	$F_3: (Y_1, Y_2) \sim N(\mathbf{0}_2, (1, -0.5; -0.5, 1))$
	$F_4: Y_1 \sim t_3, Y_2 \sim t_3, Y_1 \perp Y_2$
n	100
R	2,000
τ_n	$\ln\ln n$, $\sqrt{\ln n}$, $\ln n$, and $(\sqrt{n})^{c_\tau}$ for $c_\tau \in \{1/6, 1/3, 1/2, 2/3, 5/6\}$
b_n	Integer part of n^{c_b} for $c_b \in \{0.3, 0.4, 0.5, 0.6, 2/3, 0.7, 0.8, 0.9\}$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

In addition, we have also included the "intersection bounds" inference method proposed by Chernozhukov et al. (2013), which we refer to as the CLR method. This inferential procedure can be applied to partially identified moment (in)equality models and has several desirable features that make it one of the best currently available methods in the literature.[14] For these reasons, we include the CLR method as a benchmark of quality in our simulation study.

Table 2 describes the empirical coverage rate of $\theta = 0$ in design 1 for all the inferential methods. The methods are labeled as follows: B denotes bootstrap, AA denotes asymptotic approximation, SS1 denotes subsampling 1, SS2 denotes subsampling 2, and CLR denotes the CLR method. The coverage rates in Table 2 are the outcome of 10,000 simulations with a desired coverage rate of $(1-\alpha) = 95\%$. According to our theoretical results, all of these coverage rates should converge to the desired coverage level.

The results can be described as follows. First, the CLR method produces empirical coverage rates that are extremely close to the desired coverage level. This is another indication that the CLR method constitutes a high standard for our comparison.

TABLE 2. Empirical coverage rates (in %) of $\theta = 0$ for design 1 with $n = 100$ and $(1-\alpha) = 95\%$

				$b_n = 4$	$b_n = 6$	$b_n = 10$	$b_n = 16$	$b_n = 22$	$b_n = 25$	$b_n = 40$	$b_n = 63$
		B	AA	SS1							
	$\tau_n = 1.46$	94.60	94.56	94.18	93.95	93.39	92.51	91.50	91.03	87.40	78.58
	$\tau_n = 1.52$	94.60	94.56	94.18	93.95	93.39	92.51	91.51	91.04	87.45	78.74
	$\tau_n = 2.14$	94.97	94.90	94.56	94.34	93.82	92.96	92.04	91.61	88.19	79.55
F_1	$\tau_n = 2.15$	94.98	94.91	94.57	94.35	93.83	92.97	92.05	91.62	88.20	79.55
	$\tau_n = 3.16$	95.09	95.03	94.67	94.47	93.98	93.15	92.26	91.83	88.47	79.90
	$\tau_n = 4.60$	95.09	95.03	94.67	94.47	93.98	93.15	92.26	91.83	88.48	79.91
	$\tau_n = 4.64$	95.09	95.03	94.67	94.47	93.98	93.15	92.26	91.83	88.48	79.91
	$\tau_n = 6.81$	95.09	95.03	94.67	94.47	93.98	93.15	92.26	91.83	88.48	79.91
		CLR		SS2							
		95.00		97.74	98.19	98.79	99.41	99.69	99.74	99.99	100

				$b_n = 4$	$b_n = 6$	$b_n = 10$	$b_n = 16$	$b_n = 22$	$b_n = 25$	$b_n = 40$	$b_n = 63$
		B	AA	SS1							
	$\tau_n = 1.46$	94.68	94.70	94.38	94.30	94.00	93.35	92.53	92.07	89.37	82.95
	$\tau_n = 1.52$	94.68	94.71	94.39	94.31	94.01	93.36	92.54	92.08	89.38	82.95
	$\tau_n = 2.14$	94.69	94.73	94.41	94.34	94.04	93.40	92.59	92.13	89.45	83.15
F_2	$\tau_n = 2.15$	94.69	94.73	94.41	94.34	94.04	93.40	92.59	92.13	89.45	83.15
	$\tau_n = 3.16$	94.69	94.73	94.41	94.34	94.04	93.40	92.59	92.13	89.45	83.16
	$\tau_n = 4.60$	94.69	94.73	94.41	94.34	94.04	93.40	92.59	92.13	89.45	83.16
	$\tau_n = 4.64$	94.69	94.73	94.41	94.34	94.04	93.40	92.59	92.13	89.45	83.16
	$\tau_n = 6.81$	94.69	94.73	94.41	94.34	94.04	93.40	92.59	92.13	89.45	83.16

Table continues on overleaf

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

TABLE 2. *continued*

		CLR		SS2							
		95.05		97.88	98.29	98.87	99.48	99.71	99.76	100	100
				$b_n = 4$	$b_n = 6$	$b_n = 10$	$b_n = 16$	$b_n = 22$	$b_n = 25$	$b_n = 40$	$b_n = 63$
		B	AA	SS1							
F_3	$\tau_n = 1.46$	93.02	93.01	92.71	92.42	91.94	90.87	89.74	89.19	84.34	73.33
	$\tau_n = 1.52$	93.17	93.14	92.85	92.57	92.09	90.98	89.87	89.33	84.46	73.45
	$\tau_n = 2.14$	94.46	94.38	94.12	93.96	93.43	92.36	91.17	90.62	85.62	74.44
	$\tau_n = 2.15$	94.46	94.38	94.12	93.96	93.43	92.37	91.18	90.63	85.63	74.44
	$\tau_n = 3.16$	94.81	94.72	94.43	94.27	93.79	92.71	91.54	90.98	86.10	74.78
	$\tau_n = 4.60$	94.82	94.74	94.45	94.29	93.81	92.74	91.56	91.00	86.11	74.80
	$\tau_n = 4.64$	94.82	94.74	94.45	94.29	93.81	92.74	91.56	91.00	86.11	74.80
	$\tau_n = 6.81$	94.82	94.74	94.45	94.29	93.81	92.74	91.56	91.00	86.11	74.80
		CLR		SS2							
		94.74		96.99	97.54	98.20	98.78	99.33	99.53	99.97	100
				$b_n = 4$	$b_n = 6$	$b_n = 10$	$b_n = 16$	$b_n = 22$	$b_n = 25$	$b_n = 40$	$b_n = 63$
		B	AA	SS1							
F_4	$\tau_n = 1.46$	94.11	93.62	94.15	94.85	93.96	92.36	90.77	90.04	85.44	74.93
	$\tau_n = 1.52$	94.21	93.72	94.24	94.92	94.06	92.47	90.91	90.18	85.62	75.13
	$\tau_n = 2.14$	94.75	94.36	94.77	95.36	94.59	93.13	91.70	91.00	86.76	76.63
	$\tau_n = 2.15$	94.75	94.36	94.77	95.36	94.59	93.13	91.70	91.00	86.76	76.64
	$\tau_n = 3.16$	95.16	94.81	95.23	95.71	94.99	93.64	92.35	91.70	87.63	77.96
	$\tau_n = 4.60$	95.41	95.06	95.46	95.92	95.24	93.95	92.69	92.02	88.06	78.53
	$\tau_n = 4.64$	95.41	95.06	95.46	95.92	95.24	93.95	92.69	92.02	88.06	78.53
	$\tau_n = 6.81$	95.45	95.11	95.50	95.95	95.29	94.00	92.73	92.08	88.16	78.68
		CLR		SS2							
		95.42		98.06	98.77	99.15	99.56	99.75	99.83	100	100

Second, we turn to the bootstrap and the asymptotic approximation. Across all choices of the threshold parameter τ_n, these two methods provide very similar empirical coverage probabilities, which are also very close to the desired coverage level. The similarities between the bootstrap and the asymptotic approximation are in line with our theoretical results regarding the rates of convergence of the ECP. We also find reassuring that these inferential methods produce results that are of similar quality to the CLR method.

In the context of these simulations, the quality of the bootstrap and the asymptotic approximation does not seem to be too sensitive to the choice of the threshold parameter τ_n. In our approximation schemes, the threshold parameter is used to determine which moment conditions are binding and which are not. Our results indicate that the bootstrap and the asymptotic approximation are able to correctly detect the binding moment conditions for all the choices of τ_n under consideration.

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

TABLE 3. Empirical coverage rates (in %) of $\theta = 0$ for design 2 with $n = 100$ and $(1-\alpha) = 95\%$

				$b_n = 4$	$b_n = 6$	$b_n = 10$	$b_n = 16$	$b_n = 22$	$b_n = 25$	$b_n = 40$	$b_n = 63$
		B	AA	SS1							
F_1	$\tau_n = 1.46$	94.20	94.20	93.71	93.35	93.04	92.04	91.11	90.44	86.96	78.47
	$\tau_n = 1.52$	94.27	94.27	93.79	93.45	93.14	92.15	91.23	90.56	87.07	78.60
	$\tau_n = 2.14$	94.75	94.75	94.34	94.02	93.75	92.82	91.92	91.25	87.89	79.33
	$\tau_n = 2.15$	94.75	94.75	94.34	94.02	93.75	92.82	91.93	91.26	87.91	79.36
	$\tau_n = 3.16$	94.91	94.92	94.54	94.20	93.94	93.03	92.12	91.48	88.13	79.57
	$\tau_n = 4.60$	94.91	94.92	94.54	94.20	93.94	93.03	92.12	91.48	88.13	79.58
	$\tau_n = 4.64$	94.91	94.92	94.54	94.20	93.94	93.03	92.12	91.48	88.13	79.58
	$\tau_n = 6.81$	94.91	94.92	94.54	94.20	93.94	93.03	92.12	91.48	88.13	79.58
		CLR		SS2							
		94.73		97.71	98.11	98.84	99.31	99.65	99.76	99.99	100
				$b_n = 4$	$b_n = 6$	$b_n = 10$	$b_n = 16$	$b_n = 22$	$b_n = 25$	$b_n = 40$	$b_n = 63$
		B	AA	SS1							
F_2	$\tau_n = 1.46$	92.75	92.83	92.17	91.97	91.31	90.29	89.06	88.32	84.03	73.40
	$\tau_n = 1.52$	92.89	92.99	92.34	92.14	91.48	90.47	89.20	88.49	84.15	73.52
	$\tau_n = 2.14$	94.08	94.09	93.45	93.31	92.73	91.64	90.48	89.70	85.29	74.63
	$\tau_n = 2.15$	94.11	94.12	93.48	93.34	92.75	91.66	90.49	89.72	85.30	74.63
	$\tau_n = 3.16$	94.58	94.56	94.02	93.84	93.26	92.24	91.04	90.22	85.63	74.96
	$\tau_n = 4.60$	94.60	94.58	94.04	93.86	93.28	92.26	91.06	90.24	85.63	74.96
	$\tau_n = 4.64$	94.60	94.58	94.04	93.86	93.28	92.26	91.06	90.24	85.63	74.96
	$\tau_n = 6.81$	94.60	94.58	94.04	93.86	93.28	92.26	91.06	90.24	85.63	74.96
		CLR		SS2							
		94.50		96.75	97.25	98.03	98.81	99.24	99.41	99.95	100
				$b_n = 4$	$b_n = 6$	$b_n = 10$	$b_n = 16$	$b_n = 22$	$b_n = 25$	$b_n = 40$	$b_n = 63$
		B	AA	SS1							
F_3	$\tau_n = 1.46$	94.67	94.65	94.40	94.20	93.82	93.19	92.47	91.99	89.39	83.42
	$\tau_n = 1.52$	94.67	94.66	94.41	94.21	93.83	93.20	92.48	92.00	89.40	83.44
	$\tau_n = 2.14$	94.69	94.68	94.43	94.24	93.86	93.23	92.51	92.04	89.46	83.51
	$\tau_n = 2.15$	94.69	94.68	94.43	94.24	93.86	93.23	92.51	92.04	89.46	83.51
	$\tau_n = 3.16$	94.69	94.68	94.43	94.24	93.86	93.23	92.51	92.04	89.46	83.51
	$\tau_n = 4.60$	94.69	94.68	94.43	94.24	93.86	93.23	92.51	92.04	89.46	83.51
	$\tau_n = 4.64$	94.69	94.68	94.43	94.24	93.86	93.23	92.51	92.04	89.46	83.51
	$\tau_n = 6.81$	94.69	94.68	94.43	94.24	93.86	93.23	92.51	92.04	89.46	83.51
		CLR		SS2							
		94.88		97.72	98.28	98.94	99.36	99.68	99.74	99.98	100

Table continues on overleaf

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

TABLE 3. *continued*

				$b_n = 4$	$b_n = 6$	$b_n = 10$	$b_n = 16$	$b_n = 22$	$b_n = 25$	$b_n = 40$	$b_n = 63$
		B	AA	SS1							
F_4	$\tau_n = 1.46$	93.61	93.34	93.70	94.24	93.37	91.93	90.52	89.75	85.17	75.15
	$\tau_n = 1.52$	93.69	93.42	93.78	94.30	93.44	92.00	90.59	89.83	85.23	75.33
	$\tau_n = 2.14$	94.35	94.16	94.45	94.86	94.11	92.80	91.55	90.81	86.46	77.01
	$\tau_n = 2.15$	94.35	94.16	94.45	94.86	94.11	92.80	91.56	90.83	86.48	77.02
	$\tau_n = 3.16$	95.04	94.78	95.09	95.46	94.74	93.56	92.40	91.70	87.50	78.41
	$\tau_n = 4.60$	95.31	95.06	95.34	95.69	95.00	93.88	92.74	92.07	87.97	79.00
	$\tau_n = 4.64$	95.31	95.06	95.34	95.69	95.00	93.88	92.74	92.07	87.98	79.03
	$\tau_n = 6.81$	95.35	95.10	95.38	95.73	95.04	93.92	92.78	92.10	88.04	79.12
		CLR		SS2							
		95.13		98.10	98.80	99.21	99.47	99.77	99.85	99.99	100

Finally, we consider the subsampling approximations. Relative to the previous inferential methods, both subsampling approximations provide empirical coverage probabilities that are farther away from the desired coverage level. In particular, if we consider larger subsampling sizes, subsampling 1 seems to suffer from undercoverage problems while subsampling 2 seems to suffer from overcoverage problems. The fact that subsampling methods perform worse than the bootstrap or the asymptotic approximation is in agreement with our theoretical results regarding the rates of convergence of the ECP.

In principle, subsampling 1 depends on the threshold parameter τ_n and on the subsampling size b_n. Our results indicate that the empirical coverage rate of subsampling 1 is mostly invariant to τ_n but it is greatly affected by the choice of b_n. In turn, the empirical coverage rate of subsampling 2 only depends on the subsampling size b_n and it is also greatly affected by this choice. For both subsampling procedures, the absolute value of the ECP is minimized by choosing the smallest value of b_n.

Table 3 provides analogous results for design 2. The findings are very similar to those of design 1 and in line with our asymptotic results. As before, the CLR method produces excellent results. The bootstrap and the asymptotic approximations are as good as the CLR method and are not very sensitive to the choice of the threshold parameter τ_n. Finally, both subsampling approximations are relatively less precise than the other methods and are very sensitive to the choice of the subsampling size b_n.

7. CONCLUSION

This paper contributes to the literature on inference in partially identified econometric models defined by moment (in)equalities. In particular, we consider

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

the problem of covering the elements of the identified set with a prespecified probability. The existing literature has proposed three methods to implement this inferential problem: bootstrap, subsampling, and asymptotic approximation. All of these inferential procedures satisfy their coverage objective with equality, i.e., the exact coverage level converges to the desired coverage level of $(1-\alpha)$ as the sample size goes to infinity. The contribution of this paper is to compare these inferential methods in terms of the rate at which they achieve the desired coverage level, i.e., in terms of the rate of convergence of the ECP. This is a very relevant criterion to compare the precision of the inference: if two inferential methods, A and B, achieve an asymptotic coverage level of $(1-\alpha)$, but A has a faster rate of convergence of the ECP than B, then, for all sufficiently large sample sizes, the exact coverage level of A will be closer to $(1-\alpha)$ than that of B.

We derive the rate of convergence of the ECP of the bootstrap, the asymptotic approximation, and two versions of subsampling. The bootstrap and the asymptotic approximation are shown to have an ECP that converges to zero at a rate of $n^{1/2}$. Under certain conditions, we show that both versions of subsampling have an ECP that converges to zero at a rate of $n^{1/3}$ or slower. As a consequence, under these conditions, the bootstrap and the asymptotic approximation produce inference that is more precise than subsampling. A Monte Carlo simulation study confirms the relevance of our asymptotic results in finite samples.

NOTES

1. There is a separate literature of inference in partially identified models that considers a different coverage objective. In particular, this literature proposes a confidence set that (asymptotically) covers the identified set itself with a minimum prespecified probability of $(1-\alpha)$. This literature includes Chernozhukov et al. (2007), Beresteanu and Molinari (2008), Romano and Shaikh (2010), and Bugni (2010).

2. Under our assumptions, all the inferential procedures considered in this paper satisfy the coverage objective with equality.

3. This is an expected result given that the statistic of interest is not asymptotically pivotal. In order to obtain asymptotic refinements, one could consider a computationally intensive procedure called prepivoting, introduced by Beran (1987, 1988).

4. We illustrate this in Example 2.1, where we consider an empirically relevant missing data model that satisfies all the assumptions in this paper but does not satisfy the "conditional separability" assumption in Bugni (2010).

5. In particular, we would need to develop uniform versions of classical results in statistics (e.g., the Berry–Esseen theorem and the law of iterated logarithm), as well as specific technical results in the literature (e.g., Babu and Singh, 1985, Thm. 1; Davydov, Lifshits, and Smorodina, 1995, Thm. 11.1).

6. I thank an anonymous referee for suggesting the consideration of this important problem.

7. Previous versions of this paper used a different expression for equation (3.2). In particular, our previous version of equation (3.2) differed from the current one in two ways: (a) the indicator function terms were $1[|\sqrt{n}\bar{m}_{n,j}(\theta)| \leq \tau_n]$ instead of $1[\sqrt{n}\bar{m}_{n,j}(\theta) \leq \tau_n]$ and (b) the entire expression was multiplied by an additional indicator function term $1[\theta \in \hat{\Theta}_I(\tau_n)]$, where $\hat{\Theta}_I(\tau_n)$ was a conservative estimator of the identified set. These differences were originally included to improve the finite sample properties of the approximation but have no impact on the asymptotic findings developed in this paper. Following the suggestion of an anonymous referee, we decided to remove them to enhance the comparability of our results to the literature.

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

8. Restricting the significance level α to $(0, 0.5)$ avoids possible discontinuities in the limiting distribution of our test statistic $Q_n(\theta)$. This is not restrictive in practice as α is usually chosen to be a small number (typically, 1%, 5%, or 10%).

9. See Hall (1992) or Horowitz (2001) for excellent explanations regarding the relationship between the asymptotic pivotality of the statistic under consideration and the asymptotic refinements of the bootstrap approximation.

10. I thank an anonymous referee for suggesting this interpretation of the results.

11. The rates of convergence of the ECP for the bootstrap or asymptotic approximations are not based on Edgeworth expansions. Instead, they rely on other results, such as the Berry–Esseen theorem and the central limit theorem and, consequently, Assumption A6 is not necessary to derive Theorems 3.2 or 4.1.

12. In the case of design 1, we also considered empirical coverage rate for hypothesis tests for parameter values in the interior of the identified set and obtained results that are in line with the theoretical findings: interior points are covered with very high frequency by all inferential procedures. For reasons of brevity, these additional results are omitted from the paper.

13. We have also conducted simulations with $n = 200$, 500, and $1,000$, and obtained qualitatively similar conclusions.

14. The CLR method allows for partially identified models defined by a continuum of population (in)equalities, which can be applied to a wide range of relevant economic models, including moment (in)equalities. This method distinguishes relevant (in)equalities from irrelevant ones using a threshold sequence that is data-driven and adaptive to the smoothness of the (in)equalities. This is the so-called "Adaptive Inequality Selection" property and it constitutes an important advantage relative to the rest of the existing literature. In addition, the CLR approach to inference for conditional moment (in)equalities is conceptually different from other ideas in the literature. Rather than translating conditional moment conditions into a collection of unconditional ones, the CLR method considers the extremum of precision corrected-estimators of the conditional moment conditions. Finally, the method is generally applicable as it allows for sieves or kernel-type of nonparametric estimators.

REFERENCES

Andrews, D.W.K. (2000) Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* 68, 399–405.

Andrews, D.W.K., S. Berry, & P. Jia-Barwick (2004) Confidence Regions for Parameters in Discrete Games with Multiple Equilibria with an Application to Discount Chain Store Location. Mimeo, Yale University and M.I.T.

Andrews, D.W.K. & P. Guggenberger (2009) Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. *Econometric Theory* 25, 669–709.

Andrews, D.W.K. & P. Jia-Barwick (2012) Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica* 80, 2805–2826.

Andrews, D.W.K. & X. Shi (2013) Inference based on conditional moment inequalities. *Econometrica* 81, 609–666.

Andrews, D.W.K. & G. Soares (2010) Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* 78, 119–157.

Armstrong, T.B. (2012) Asymptotically Exact Inference in Conditional Moment Inequality Models. Mimeo, Yale University.

Armstrong, T.B. (2014) Weighted KS statistics for inference on conditional moment inequalities. *Journal of Econometrics* 181, 92–116.

Babu, G.J. & K. Singh (1985) Edgeworth expansions for sampling without replacement from finite populations. *Journal of Multivariate Analysis* 17, 261–278.

Beran, R. (1987) Prepivoting to reduce level error of confidence sets. *Biometrika* 74, 457–468.

Beran, R. (1988) Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* 83, 687–697.

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

Beresteanu, A. & F. Molinari (2008) Asymptotic properties for a class of partially identified models. *Econometrica* 76, 763–814.

Bhattacharya, R.N. & R.R. Rao (1976) *Normal Approximation and Asymptotic Expansions*. Wiley.

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

Bontemps, C., T. Magnac, & E. Maurin (2012) Set identified linear models. *Econometrica* 80, 1129–1155.

Bugni, F.A. (2010) Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica* 78, 735–753.

Bugni, F.A. (2014) Supplemental Material to "Comparison of inferential methods in partially identified models in terms of error in coverage probability". Accepted for electronic publication in *Econometric Theory* on August 11, 2014.

Bugni, F.A., I.A. Canay, & P. Guggenberger (2012) Distortions of asymptotic confidence size in locally misspecified moment inequality models. *Econometrica* 80, 1741–1768.

Canay, I.A. (2010) E.L. inference for partially identified models: Large deviations optimality and bootstrap validity. *Journal of Econometrics* 156, 408–425.

Chernozhukov, V., H. Hong, & E. Tamer (2007) Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75, 1243–1284.

Chernozhukov, V., S. Lee, & A.M. Rosen (2013) Intersection bounds: Estimation and inference. *Econometrica* 81, 667–737.

Chetverikov, D. (2012) Adaptive Test of Conditional Moment Inequalities. Mimeo, MIT.

Davydov, Y.A., M.A. Lifshits, & N.V. Smorodina (1995) *Local Properties of Distributions of Stochastic Functionals*. Translations of Mathematical Monographs, vol. 173. American Mathematical Society.

Galichon, A. & M. Henry (2006) Inference in Incomplete Models. Mimeo, Ecole Polytechnique, Paris - Department of Economic Sciences and Pennsylvania State University.

Galichon, A. & M. Henry (2013) Dilation bootstrap. *Journal of Econometrics* 177, 109–115.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*. Springer Series in Statistics. Springer.

Horowitz, J.L. (2001) The bootstrap. In J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, vol. 4, chap. 52, pp. 3159–3228. Elsevier Science B.V.

Hüsler, J., R. Liu, & K. Singh (2002) A formula for the tail probability of a multivariate normal distribution and its applications. *Journal of Multivariate Analysis* 82, 422–430.

Imbens, G. & C.F. Manski (2004) Confidence Intervals for Partially Identified Parameters. *Econometrica* 72, 1845–1857.

Joag-Dev, K. & F. Proschan (1983) Negative association of random variables with applications. *Annals of Statistics* 11, 286–295.

Kim, K. (2008) Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities. Mimeo, Michigan State University.

Manski, C.F. (1995) *Identification Problems in the Social Sciences*. Harvard University Press.

Pakes, A., J. Porter, K. Ho, & J. Ishii (2014) Moment Inequalities and Their Application. http://dx.doi.org/10.1920/wp.cem.2007.1607.

Politis, D.N., J.P. Romano, & M. Wolf (1999) *Subsampling*. Springer.

Ponomareva, M. (2010) Inference in Models Defined by Conditional Moment Inequalities with Continuous Covariates. Mimeo, Northern Illinois University.

Romano, J.P. & A.M. Shaikh (2008) Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference* 138, 2786–2807.

Romano, J.P. & A.M. Shaikh (2010) Inference for the identified set in partially identified econometric models. *Econometrica* 78, 169–211.

Romano, J.P., A.M. Shaikh, & M. Wolf (2014) A practical two-step method for testing moment inequalities. *Econometrica* 82, 1979–2002.

Rosen, A.M. (2008) Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *Journal of Econometrics* 146, 107–117.

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

Stoye, J. (2009) More on confidence intervals for partially identified parameters. *Econometrica* 77, 299–1315.

Tamer, E. (2003) Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies* 70, 147–165.

APPENDIX

A.1. Additional Notation

Throughout this appendix, "SLLN" refers to the strong law of large numbers, "CLT" refers to the central limit theorem, "CMT" refers to the continuous mapping theorem, "LIL" refers to the law of iterated logarithm. Also, "s.t." abbreviates "such that", and "RHS" and " LHS" abbreviate "right hand side" and "left hand side", respectively. Unless it is specified, limits (i.e., lim, lim inf, and lim sup) are computed as $n \to \infty$.

For any $\theta \in \Theta$ and $j = 1, \ldots, J$, we define $v_n(m_{j,\theta}) \equiv \sqrt{n}(\bar{m}_j(\theta) - E[m_j(Z,\theta)])$ and $v_n(m_\theta) \equiv \{v_n(m_{j,\theta})\}_{j=1}^J$, and, analogously, we define $v_n^*(m_{j,\theta}) \equiv \sqrt{n}(\bar{m}_j^*(\theta) - \bar{m}_j(\theta))$ and $v_n^*(m_\theta) \equiv \{v_n^*(m_{j,\theta})\}_{j=1}^J$.

We metrize matrices with the Frobenius norm, i.e., $\forall s_1, s_2 \in \mathbb{N}$, $M \in \mathbb{R}^{s_1 \times s_2}$, $\|M\| \equiv (\sum_{i=1}^{s_1} \sum_{j=1}^{s_2} M_{[i,j]}^2)^{1/2}$. For any $s \in \mathbb{N}$, $\mathcal{C}_s$ denotes the space of Borel measurable convex sets in $\mathbb{R}^s$. For any $H : A_1 \to A_2$ and any $S \subset A_2$, $H^{-1}(S) \equiv \{x \in A_1 : H(x) \in S\}$. For any $\varepsilon > 0$ and any $S \subset \mathbb{R}^s$, $S^\varepsilon \equiv \{x \in \mathbb{R}^s : \exists x' \in S \ s.t. \ \|x - x'\| \leq \varepsilon\}$. For any positive semidefinite $\Sigma \in \mathbb{R}^{s \times s}$ and any Borel measurable set A, $\Phi_\Sigma(A) \equiv P(Z \in A)$ with $Z \sim N(\mathbf{0}_s, \Sigma)$. If $\Sigma \in \mathbb{R}^{s \times s}$ is positive definite, $\phi_\Sigma(x)$ denotes the density of $Z \sim N(\mathbf{0}_s, \Sigma)$.

A.2. Inconsistency of the Naive Bootstrap

The bootstrap procedure considered in Section 3 differs qualitatively from replacing the subsampling scheme in CHT with the traditional bootstrap. In their Remark 5.2, CHT propose approximating the $(1-\alpha)$-quantile of the distribution of interest by simulating the following subsampling statistic:

$$Q_{b_n,n}^{SS_2}(\theta) \;=\; G\left(\left\{\left[\sqrt{b_n}\bar{m}_{b_n,n,j}^{SS}(\theta)\right]_-\right\}_{j=1}^J\right),$$

where $\{Z_i^{SS}\}_{i=1}^{b_n}$ is a random sample of size b_n extracted without replacement from $\mathcal{X}_n$ (i.e., subsampling) and $\bar{m}_{b_n,n,j}^{SS}(\theta) \equiv b_n^{-1} \sum_{i=1}^{b_n} m_j(Z_i^{SS}, \theta)$ for $j = 1, \ldots, J$. This approximation coincides with our subsampling 2. Our results in Section 5.2 show that, under mild conditions, this subsampling approximation leads to correct inference in level. From this result, it is tempting to (naively) replace the subsampling procedure with the bootstrap, which produces the following statistic:

$$Q_n^{*,Naive}(\theta) \;=\; G(\{[\sqrt{n}\bar{m}_{n,j}^*(\theta)]_-\}_{j=1}^J),$$

where $\{Z_i^*\}_{i=1}^n$ is a random sample of size n extracted with replacement from $\mathcal{X}_n$ (i.e., bootstrap) and $\bar{m}_{n,j}^*(\theta) \equiv n^{-1} \sum_{i=1}^n m_j(Z_i^*, \theta)$ for $j = 1, \ldots, J$. We refer to this bootstrap scheme as *naive bootstrap*, because it is the result of replacing one resampling scheme

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

with another one. Unlike the subsampling approximation, the naive bootstrap can produce inconsistent inference in level. This problem is related to the "parameter on the boundary problem" initially described in Andrews (2000).

For the sake of completeness, we illustrate the inconsistency of the naive bootstrap using a very simple example. Let $\Theta = [-B, B]$ for $B > 0$ and $\Theta_I = \{\theta \in \Theta : E[Z] - \theta \geq 0\}$. Furthermore, assume that $E[Z] = 0$, and so $\Theta_I = [-B, 0]$. The normalized sample analog criterion function is given by $Q_n(\theta) = G([\sqrt{n}(\bar{Z}_n - \theta)]_-)$. It is easy to verify that $Q_n(\theta) \xrightarrow{d} G([\zeta]_-)1[\theta = 0]$ with $\zeta \sim N(0,1)$. We are interested in conducting inference on $\theta = 0$, i.e., our objective is to approximate the $(1-\alpha)$-quantile of $G([\zeta]_-)$ with $\zeta \sim N(0,1)$.

Consider first an approximation using the naive bootstrap at $\theta = 0$. The naive bootstrap statistic is given by:

$$Q_n^{*,Naive}(0) = G([\sqrt{n}\bar{Z}_n^*]_-) = G([\sqrt{n}(\bar{Z}_n^* - \bar{Z}_n) + \sqrt{n}(\bar{Z}_n - E[Z])]_-),$$

where $\bar{Z}_n$ and $\bar{Z}_n^*$ are the sample average and the bootstrap sample average, respectively. For any $c > 0$, define the events $A_1 \equiv \{\limsup\{\sqrt{n}(\bar{Z}_n - E[Z]) > c\}\}$, $A_2 \equiv \{\{\sqrt{n}(\bar{Z}_n^* - \bar{Z}_n)|\mathcal{X}_n\} \xrightarrow{d} N(0,1)\}$, and $A \equiv A_1 \cap A_2$. We now show that the event A implies that the naive bootstrap is inconsistent. Suppose that A occurs. On the one hand, A_1 implies that there is a subsequence $\{k_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ s.t. $\sqrt{n}(\bar{Z}_{k_n} - E[Z]) > c$. Along this subsequence,

$$Q_{k_n}^{*,Naive}(0) \leq G\left(\left[\sqrt{k_n}\left(\bar{Z}_{k_n}^* - \bar{Z}_{k_n}\right) + c\right]_-\right).$$

On the other hand, A_2 implies that the conditional distribution of the RHS converges weakly to $G([\zeta + c]_-)$ with $\zeta \sim N(0,1)$. Since $c > 0$, the $(1-\alpha)$-quantile of $G([\zeta + c]_-)$ is strictly smaller than the $(1-\alpha)$-quantile of $G([\zeta]_-)$, which implies that the naive bootstrap leads to undercoverage along the subsequence $\{k_n\}_{n\geq 1}$. In order to establish that the naive bootstrap is inconsistent, it suffices to show that $P(A) > 0$. The LIL implies that $P(A_1) = 1$, Bickel and Freedman (1981, Thm. 2.1) implies that $P(A_2) = 1$ and, together, these imply that $P(A) = 1$.

In contrast, the bootstrap procedure proposed by this paper would consider the following statistic:

$$Q_n^*(0) \;=\; G([\sqrt{n}(\bar{Z}_n^* - \bar{Z}_n)]_- \times 1[\sqrt{n}\bar{Z}_n \leq \tau_n]).$$

The results in the paper show that this bootstrap approximation leads to correct inference in level. In fact, using similar arguments to those used to analyze the naive bootstrap, it is easy to show that the conditional distribution of $Q_n^*(0)$ converges weakly to $G([\zeta]_-)$ with $\zeta \sim N(0,1)$ a.s.

A.3. Results for the Sample Statistic

THEOREM A.1. *Assume Assumptions A1–A3, CF1, and that θ satisfies Assumption A4. Let $\rho \equiv rank(V[m(Z,\theta)])$.*

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

1. *If* $\theta \in \partial\Theta_I$, *then* $Q_n(\theta) = H(\sqrt{n}\bar{Y}_n) + \delta_n$, *where*
   (a) *for any positive sequence* $\{\varepsilon_n\}_{n\geq 1}$ *with* $\varepsilon_n = O(n^{-1/2})$, $P(|\delta_n| > \varepsilon_n) = o(n^{-1/2})$.
   (b) $\bar{Y}_n : \Omega_n \to \mathbb{R}^\rho$ *is a zero mean sample average of n i.i.d. observations from a distribution denoted by* Υ *with a nonsingular variance-covariance matrix* $V = \mathbf{I}_\rho$ *and finite third absolute moments. If Assumption A5 is added, then this distribution has finite fourth absolute moments.*
   (c) $H : \mathbb{R}^\rho \to \mathbb{R}$ *is continuous, nonnegative, weakly convex, and homogeneous of degree* β. $H(y) = 0$ *implies for some nonzero vector* $b \in \mathbb{R}^\rho$, $b'y \leq 0$. *For any* $\mu > 0$, *any* $|h| \geq \mu > 0$, *any* $C > 0$, *any positive sequence* $\{g_n\}_{n\geq 1}$, *and any positive sequence* $\{\varepsilon_n\}_{n\geq 1}$ *with* $\varepsilon_n = o(1)$, $\{H^{-1}(\{h\}^{\varepsilon_n}) \cap \|y\| \leq C\sqrt{g_n}\} \subseteq \{H^{-1}(\{h\})\}^{\eta_n}$ *where* $\eta_n = O(\sqrt{g_n}\varepsilon_n)$, *uniformly in h.*
   (d) *If we add Assumption CF2, then for any* $\mu > 0$, *any* $|h| \geq \mu > 0$, *and any sequence* $\{\varepsilon_n\}_{n\geq 1}$ *with* $\varepsilon_n = o(1)$, $\{H^{-1}(\{h\}^{\varepsilon_n})\} \subseteq \{H^{-1}(\{h\})\}^{\gamma_n}$ *where* $\gamma_n \leq O(\varepsilon_n)$, *uniformly in h.*
2. *If* $\theta \in Int(\Theta_I)$, *then* $\liminf\{Q_n(\theta) = 0\}$ *a.s.*

Proof. Part 1. Define $\delta_n \equiv Q_n(\theta) - G(\{[\upsilon_n(m_{j,\theta})]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J)$. Since $\theta \in \Theta_I$, $\delta_n \geq 0$ and so, it suffices to show that $P(\delta_n > \varepsilon_n) = o(n^{-1/2})$. For every $n \in \mathbb{N}$, let $d_n = n^{1/4+\delta/2}$ for some $\delta \in (0, 1/8)$. By elementary arguments,

$$P(\delta_n > \varepsilon_n) = P(\delta_n > \varepsilon_n \cap \{\|\upsilon_n(m_\theta)\| > d_n\}) + P(\delta_n > \varepsilon_n \cap \{\|\upsilon_n(m_\theta)\| \leq d_n\})$$
$$\leq \sum_{j=1}^J P(|\upsilon_n(m_{j,\theta})| > d_n/\sqrt{J}) + P(\delta_n > \varepsilon_n \cap \{\|\upsilon_n(m_\theta)\| \leq d_n\}). \qquad \textbf{(A.1)}$$

We now show that the two terms on the RHS of equation (A.1) are $o(n^{-1/2})$. We begin with the first term. Fix $j = 1, \ldots, J$ arbitrarily. By Chebyshev's inequality, $V[m_j(Z,\theta)] \geq P(|\upsilon_n(m_{j,\theta})| > d_n)d_n^2$ and, since $\delta > 0$, it follows that $P(|\upsilon_n(m_{j,\theta})| > d_n) = o(n^{-1/2})$. We now consider the second term on the RHS of equation (A.1). Suppose that $\{\|\upsilon_n(m_\theta)\| \leq d_n\}$ occurs. Let $S \equiv \{j \in \{1, \ldots, J\} : E[m_j(Z,\theta)] = 0\}$, $\bar{S} \equiv \{1, \ldots, J\}/S$, and $\eta \equiv \min_{j\in\bar{S}} E[m_j(Z,\theta)] > 0$. Then,

$$Q_n(\theta) = G(\{[\upsilon_n(m_{j,\theta})]_-\}_{j\in S}, \{[\upsilon_n(m_{j,\theta}) + \sqrt{n}E[m_j(Z,\theta)]]_-\}_{j\in\bar{S}})$$
$$\leq G(\{[\upsilon_n(m_{j,\theta})]_-\}_{j\in S}, \{[-d_n + \sqrt{n}\eta]_-\}_{j\in\bar{S}})$$
$$\leq G(\{[\upsilon_n(m_{j,\theta})]_-\}_{j\in S}, \{0\}_{j\in\bar{S}})$$
$$= G(\{[\upsilon_n(m_{j,\theta})]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J),$$

where the first inequality holds by definition of η and by assuming $\{\|\upsilon_n(m_\theta)\| \leq d_n\}$, while the second inequality holds for sufficiently large $n \in \mathbb{N}$ due to $\delta < 1/8$. Therefore, $\liminf P(\delta_n > \varepsilon_n \cap \{\|\upsilon_n(m_\theta)\| \leq d_n\}) = 0$, which implies that $P(\delta_n > \varepsilon_n \cap \{\|\upsilon_n(m_\theta)\| \leq d_n\}) = o(n^{-1/2})$.

The objective is now to show that $G(\{[\upsilon_n(m_{j,\theta})]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J) = H(\sqrt{n}\bar{Y}_n)$, where H and $\bar{Y}_n$ satisfy certain properties. Let $\Sigma \equiv V[m(Z,\theta)]$ and, thus, $\rho \equiv rank(\Sigma)$. As a preliminary step, we now show how to define the following objects: (a) the matrix $B \in \mathbb{R}^{J\times\rho}$ that is required in the definition of the function H and (b) the

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

sample of observations $\{Y_i\}_{i=1}^n$ whose average is $\bar{Y}_n$. By definition of the rank of a matrix, there are $(J-\rho)$ coordinates of $m(Z_i,\theta)$ that can be expressed as a linear combination of the remaining ρ coordinates of $m(Z_i,\theta)$ for all $i=1,\ldots,n$ (a.s.). We refer to these ρ coordinates of $m(Z_i,\theta)$ as the "fundamental" coordinates. Without loss of generality, we can rearrange $m(Z_i,\theta)$ s.t. the fundamental coordinates are the last ρ ones. This implies that there is a matrix $A \in \mathbb{R}^{(J-\rho)\times\rho}$ s.t.

$$m(Z_i,\theta) = [A', \mathbf{I}_\rho]'\{m_j(Z_i,\theta)\}_{j=1}^{\rho} \text{ for all } i=1,\ldots,n, \ a.s. \tag{A.2}$$

Let Σ_ρ denote the variance-covariance matrix of $\{m_j(Z_i,\theta)\}_{j=1}^{\rho}$, which is necessarily positive definite and let $\Sigma_\rho^{-1/2}$ denote the inverse of its square root. Then, we define $Y_i \equiv \Sigma_\rho^{-1/2}\{m_j(Z_i,\theta) - E[m_j(Z_i,\theta)]\}_{j=1}^{\rho}$ for all $i=1,\ldots,n$ and the matrix $B \equiv [A', \mathbf{I}_\rho]'\Sigma_\rho^{1/2} \in \mathbb{R}^{J\times\rho}$.

The previous paragraph defined a sample $\{Y_i\}_{i=1}^n$. From its definition, we now verify that this sample satisfies all the desired properties. Since $\{Z_i\}_{i=1}^n$ is an i.i.d. sample, it then follows that $\{Y_i\}_{n\geq 1}$ is also an i.i.d. sample from a distribution denoted by Υ in the statement. It is easy to verify that $E[Y_i] = \mathbf{0}_\rho$, $V[Y_i] = \mathbf{I}_\rho$, and $E[||m(Z,\theta)||^k] < \infty$ implies $E[||Y_i||^k] < \infty$. Finally, it also follows from the definition of $\{Y_i\}_{i=1}^n$ that $BY_i = m(Z_i,\theta) - E[m(Z_i,\theta)]$ for all $i=1,\ldots,n$ and, thus, $\upsilon_n(m_\theta) = B\sqrt{n}\bar{Y}_n$.

We now turn to the definition of the function $H:\mathbb{R}^\rho \to \mathbb{R}$. If we let $B_j \in \mathbb{R}^{1\times\rho}$ denote the j^{th} row of B, we define $H(y) \equiv G(\{[B_j y]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J)$. We now show that this function has all desired properties. It is easy to verify that it is continuous, nonnegative, weakly convex, and homogeneous of degree β. Since the matrix B has rank ρ, $B_j \neq \mathbf{0}_\rho$ for all $j=1,\ldots,J$. If $\theta \in \partial\Theta_I$, then $\exists j = 1,\ldots,J$ s.t. $E[m_j(Z,\theta)] = 0$ and, thus, if we set $b' \equiv B_j$, $H(y) = 0$ implies that for $b \neq \mathbf{0}_\rho$, $b'y \leq 0$.

We now show that for any $\mu > 0$ and any h s.t. $|h| \geq \mu$, $\{H^{-1}(\{h\}^{\varepsilon_n}) \cap \|y\| \leq C\sqrt{g_n}\} \subseteq \{H^{-1}(\{h\})\}^{\delta_n}$ where $\delta_n = O(\varepsilon_n\sqrt{g_n})$, uniformly in h. To this purpose, consider $\tilde{y} \in H^{-1}(\{h\}^{\varepsilon_n})$ with $\|\tilde{y}\| \leq C\sqrt{g_n}$. We need to show that $\exists y \in H^{-1}(\{h\})$ s.t. $\|\tilde{y} - y\| \leq O(\varepsilon_n\sqrt{g_n})$, uniformly in h. Consider $y = \tilde{y}(h/\tilde{h})^{1/\beta}$. By homogeneity of degree β, $H(y) = h$. Also, $||\tilde{y} - y|| \leq ||\tilde{y}|| \times |1 - (\tilde{h}/h)^{-1/\beta}| \leq C\sqrt{g_n}\max\{1-(\tilde{h}/h)^{-1/\beta}, (\tilde{h}/h)^{-1/\beta} - 1\}$, where $|\tilde{h} - h| \leq \varepsilon_n$. By the mean value theorem, $\forall h$ with $|h| \geq \mu > 0$ and $\varepsilon_n = o(1)$, $\max\{1-(\tilde{h}/h)^{-1/\beta}, (\tilde{h}/h)^{-1/\beta} - 1\} = O(|\tilde{h} - h|) = O(\varepsilon_n)$, uniformly in h. Thus, $||\tilde{y} - y|| \leq O(\sqrt{g_n}\varepsilon_n)$, uniformly in h.

Finally, if we add Assumption CF2, then any $\mu > 0$, any $|h| \geq \mu > 0$ and any sequence $\varepsilon_n = o(1)$, $\{H^{-1}(\{h\}^{\varepsilon_n})\} \subseteq \{H^{-1}(\{h\})\}^{\gamma_n}$ where $\gamma_n = O(\varepsilon_n)$, uniformly in h. Consider $y_A \in H^{-1}([h_B - \varepsilon_n, h_B + \varepsilon_n])$. By definition, this means that $\exists h_A$ s.t. $\|h_A - h_B\| \leq C_1\varepsilon_n$ for a constant C_1, and $H(y_A) = h_A$. To conclude this, it suffices to show that $\exists y_B \in \mathbb{R}^\rho$ s.t. $\|y_A - y_B\| \leq C_2\varepsilon_n$, where C_2 does not depend on h, and $H(y_B) = h_B$. We only consider the case when $G(x) = \sum_{j=1}^J w_j x_j$, as the arguments for $G(x) = \max_{j=1,\ldots,J}\{w_j x_j\}$ are analogous. For $y \in \mathbb{R}^\rho$, let $S_1(y) = \{j \in \{1,\ldots,J\} : E[m_j(Z,\theta)] = 0 \cap B_j y < 0\}$ and $S_0(y) = \{j \in \{1,\ldots,J\} : E[m_j(Z,\theta)] = 0 \cap B_j y = 0\}$. By definition, $y_A \in \mathbb{R}^\rho$ satisfies $B_j y = 0$ and $\sum_{j\in S_1(y_A)} B_j y = h_A$ for all $j \in S_0(y_A)$. Thus, $y_A \in \mathbb{R}^\rho$ satisfies a system of equations of the form $[K_1', K_2']'y = [h_A, \mathbf{0}']'$, where $K_1 = \Sigma_{j\in S_1(y_A)} B_j$, $K_2 = \{B_j\}_{j\in S_0(y_A)}$, and the $\mathbf{0}$ has the cardinality of $S_0(y_A)$. Notice that the matrix $[K_1', K_2']'$ does not depend on h_A. Consider the homogenous system $K_2 y = \mathbf{0}$. The matrix

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

K_2 may or may not have full rank, but it can always be reduced to a system $K_3 y = \mathbf{0}$, where K_3 has full rank and the corresponding $\mathbf{0}$ has dimension equal to $rank(K_3)$. Since $h_A > 0$, $[K_1', K_3']'$ has full rank. If this rank is ρ, then $y_A = [[K_1', K_3']']^{-1}[h_A, \mathbf{0}']'$ and we consider $y_B \equiv [[K_1', K_3']']^{-1}[h_B, \mathbf{0}']'$. If the rank is less than ρ, we can find K_4 so that $[K_1', K_3', K_4']'$ has rank ρ, set c s.t. $K_4 y_A = c$. Then, $y_A = [[K_1', K_3', K_4']']^{-1}[h_A, \mathbf{0}', c']'$ and we consider $y_B \equiv [[K_1', K_3', K_4']']^{-1}[h_B, \mathbf{0}', c']'$. By construction, $||y_A - y_B|| \leq C_2 \varepsilon_n$, where C_2 does not depend on h. If we verify that $H(y_B) = h_B$, then C_2 is precisely the aforementioned constant. By construction and a continuity argument, $S_1(y_A) = S_1(y_B)$ and $\sum_{j \in S_1(y_A)} B_j y_A = h_A$ imply that $\sum_{j \in S_1(y_B)} B_{j,k} y_B = h_B$. Also, by construction, $S_0(y_A) = S_0(y_B)$. Finally, by a continuity argument, for all $j = 1, \ldots, J$ s.t. $E[m_j(Z, \theta)] = 0$ and $B_j y_A > 0$, it follows that $B_j y_B > 0$. As a consequence, $H(y_B) = h_B$.

Part 2. If $\theta \in Int(\Theta_I)$, then $E[m_j(Z, \theta)] > 0$ for all $j = 1, \ldots, J$. Define $\eta \equiv \min_{j=1,\ldots,J}\{E[m_j(Z, \theta)]\} > 0$. Consider any positive sequence $\{\varepsilon_n\}_{n \geq 1}$ s.t. $\sqrt{\ln \ln n}/\varepsilon_n = o(1)$ and $\varepsilon_n / \sqrt{n} = o(1)$. Suppose that $\{\|\upsilon_n(m_\theta)\| \leq \varepsilon_n\}$ occurs. Then, for all large enough $n \in \mathbb{N}$, $Q_n(\theta) \leq G(\{[(-\varepsilon_n/\sqrt{n} + \eta)\sqrt{n}]_-\}_{j=1}^J) = 0$. Therefore, $\liminf\{\|\upsilon_n(m_\theta)\| \leq \varepsilon_n\} \subseteq \liminf\{Q_n(\theta) = 0\}$. By the LIL, $P(\liminf\{\|\upsilon_n(m_\theta)\| \leq \varepsilon_n\}) = 1$ and, thus, the result follows. ■

A.4. Results for the Bootstrap

THEOREM A.2. *Assume Assumptions A1–A3, CF1, and that θ satisfies Assumption A4. Let $\rho \equiv rank(V[m(Z, \theta)])$.*

1. *If $\theta \in \partial\Theta_I$, then $Q_n^*(\theta) = H(\sqrt{n}\bar{Y}_n^*) + \delta_n^*$, where*
   (a) *for any positive sequence $\{\varepsilon_n\}_{n \geq 1}$ with $\varepsilon_n = O(n^{-1/2})$, $P(|\delta_n^*| > \varepsilon_n | \mathcal{X}_n) = o(n^{-1/2})$ a.s.*
   (b) *$\{\bar{Y}_n^* | \mathcal{X}_n\} : \Omega_n \to \mathbb{R}^\rho$ is a zero (conditional) mean sample average of n i.i.d. observations from a distribution with a (conditional) variance-covariance matrix $\hat{V}$ which is nonsingular a.s. and finite (conditional) third absolute moments a.s. If Assumption A5 is added, then $||\hat{V} - \mathbf{I}_\rho|| \leq O_p(n^{-1/2})$ and the distribution has finite (conditional) fourth absolute moments a.s.*
   (c) *$H : \mathbb{R}^\rho \to \mathbb{R}$ is the same function as in Theorem A.1.*
2. *If $\theta \in Int(\Theta_I)$, then $\liminf\{Q_n^*(\theta) = 0\}$ a.s.*

Proof. Part 1. Since $\theta \in \partial\Theta_I \subseteq \Theta_I$ then $E[m_j(Z, \theta)] \geq 0$ for all $j = 1, \ldots, J$. Define $\delta_n^* \equiv Q_n^*(\theta) - G(\{[\upsilon_n^*(m_{j,\theta})]_- 1[E[m_j(Z, \theta)] = 0]\}_{j=1}^J)$. Let $S \equiv \{j \in \{1, \ldots, J\} : E[m_j(Z, \theta)] = 0\}$, $\bar{S} \equiv \{1, \ldots, J\}/S$, and $\eta \equiv \min_{j \in \bar{S}} E[m_j(Z, \theta)] > 0$. Consider the following derivation:

$$\{\|\upsilon_n(m_\theta)\| \leq \tau_n\} \subseteq \cap_{j=1}^J \{\sqrt{n}|(\bar{m}_{n,j}(\theta) - E[m_j(Z, \theta)])| \leq \tau_n\}$$
$$\subseteq \{\cap_{j \in S}\{\sqrt{n}\bar{m}_{n,j}(\theta) \leq \tau_n\}\} \cap \{\cap_{j \in \bar{S}}\{\sqrt{n}\bar{m}_{n,j}(\theta) > \tau_n\}\} = \{\delta_n^* = 0\},$$

where the second inclusion holds for large enough $n \in \mathbb{N}$ a.s., i.e., $\liminf\{\{\|\upsilon_n(m_\theta)\| \leq \tau_n\} \subseteq \{\delta_n^* = 0\}\}$ a.s. Therefore, $\liminf\{\|\upsilon_n(m_\theta)\| \leq \tau_n\} \subseteq \liminf\{\delta_n^* = 0\}$ a.s. By the LIL, $P(\liminf\{\|\upsilon_n(m_\theta)\| \leq \tau_n\}) = 1$, which then implies that $P(|\delta_n^*| > 0 | \mathcal{X}_n) = o(n^{-1/2})$ a.s.

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

To complete the argument, we need to verify that $G(\{[v_n^*(m_{j,\theta})]_{-}1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J) = H(\sqrt{n}\bar{Y}_n^*)$ where H and $\bar{Y}_n^*$ satisfy certain properties. Let the matrices $A \in \mathbb{R}^{(J-\rho)\times\rho}$, $\Sigma_\rho \in \mathbb{R}^{\rho\times\rho}$, and $B \equiv [A', \mathbf{I}_\rho]'\Sigma_\rho^{1/2} \in \mathbb{R}^{J\times\rho}$ be as in proof of Theorem A.1. Since bootstrap samples are randomly drawn from the original random sample, the coordinates of $\{m(Z_i^*,\theta)\}_{i=1}^n$ can be arranged into the same ρ fundamental and $(J-\rho)$ nonfundamental coordinates described in Theorem A.1. In particular, the bootstrap sample has to satisfy equation (A.2). Then, by following the arguments in Theorem A.1, we define $Y_i^* \equiv \Sigma_\rho^{-1/2}\{m_j(Z_i^*,\theta) - \bar{m}_{n,j}(\theta)\}_{j=1}^\rho$ for all $i = 1,\ldots,n$, which can be shown to satisfy $BY_i^* = m(Z_i^*,\theta) - \bar{m}_n(\theta)$ for all $i = 1,\ldots,n$. By elementary properties of i.i.d. bootstrap sampling, $E[m(Z_i^*,\theta)|\mathcal{X}_n] = \bar{m}_n(\theta)$ and $V[m(Z_i^*,\theta)|\mathcal{X}_n] = \hat{\Sigma}$ where $\hat{\Sigma}$ denotes the sample variance of $m(Z,\theta)$. By equation (A.2), $\hat{\Sigma} = [A', \mathbf{I}_\rho]'\hat{\Sigma}_\rho[A', \mathbf{I}_\rho]$, where $\hat{\Sigma}_\rho$ denotes the sample variance of $\{m_j(Z,\theta)\}_{j=1}^\rho$. By construction, $\{Y_i^*\}_{n\geq 1}$ is an i.i.d. sample with $E[Y_i^*|\mathcal{X}_n] = \mathbf{0}_\rho$, $V[Y_i^*|\mathcal{X}_n] = \hat{V}$ where $\hat{V} - \mathbf{I}_\rho = \Sigma_\rho^{-1/2}(\hat{\Sigma}_\rho - \Sigma_\rho)\Sigma_\rho^{-1/2}$. By the SLLN, $\hat{\Sigma}_\rho - \Sigma_\rho = o(1)$ a.s. and then the CMT implies that $\hat{V} - \mathbf{I}_\rho = o(1)$ a.s., and so $\hat{V}$ is nonsingular a.s. By a similar argument, $E(||m(Z,\theta)||^k) < \infty$ implies that $E(||Y_i^*||^k|\mathcal{X}_n) < \infty$ a.s. Finally, if we add Assumption A5, the CLT implies that $\sqrt{n}(\hat{\Sigma} - \Sigma) = O_p(1)$ with $\Sigma = V[m(Z,\theta)]$ and, so, $||\hat{V} - \mathbf{I}_\rho|| \leq O_p(n^{-1/2})$. Combining all these findings, we conclude that $G(\{[v_n^*(m_{j,\theta})]_{-}1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J) = H(\sqrt{n}\bar{Y}_n^*)$, where H is the function in Theorem A.1 and $\bar{Y}_n^*$ is a random variable that satisfies all the desired properties.

Part 2. If $\theta \in Int(\Theta_I)$, then $E[m_j(Z,\theta)] > 0$ for all $j = 1,\ldots,J$. Define $\eta \equiv \min_{j=1,\ldots,J} E[m_j(Z,\theta)] > 0$. Consider the following derivation:

$$\begin{aligned}\{\|v_n(m_\theta)\| \leq \tau_n\} &\subseteq \cap_{j=1}^J\{\sqrt{n}|\bar{m}_{n,j}(\theta) - E[m_j(Z,\theta)]| \leq \tau_n\}\\ &\subseteq \cap_{j=1}^J\{\bar{m}_{n,j}(\theta) \geq \eta/2\}\\ &\subseteq \cap_{j=1}^J\{\sqrt{n}\bar{m}_{n,j}(\theta) > \tau_n\} \subseteq \{Q_n^*(\theta) = 0\},\end{aligned}$$

where the second inclusion holds for large enough $n \in \mathbb{N}$ a.s. Therefore, $\liminf\{\|v_n(m_\theta)\| \leq \tau_n\} \subseteq \liminf\{Q_n^*(\theta) = 0\}$. By the LIL, $P(\liminf\{\|v_n(m_\theta)\| \leq \tau_n\}) = 1$, which implies that $\liminf\{Q_n^*(\theta) = 0\}$ a.s. ■

THEOREM A.3 (Consistency on Any Set Excluding Zero—Bootstrap). *Assume Assumptions A1-A3, CF1, and that θ satisfies Assumption A4.*

1. *If $\theta \in \partial\Theta_I$, then, $\forall \mu, \varepsilon > 0$,*

$$\lim P\left(\sup_{|h|\geq\mu}|P(Q_n^*(\theta) \leq h|\mathcal{X}_n) - \lim_{m\to\infty} P(Q_m(\theta) \leq h)| \leq \varepsilon\right) = 1.$$

2. *If $\theta \in Int(\Theta_I)$, then*

$$P\left(\liminf\left\{\sup_{h\in\mathbb{R}}|P(Q_n^*(\theta) \leq h|\mathcal{X}_n) - \lim_{m\to\infty} P(Q_m(\theta) \leq h)| = 0\right\}\right) = 1.$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

Proof. Part 1. Suppose that $\theta \in \partial\Theta_I$. By Theorem A.1, the CLT, and the CMT, $Q_n(\theta) \xrightarrow{d} H(\zeta)$, where H is function in Theorem A.1 and $\zeta \sim N(\mathbf{0}_\rho, \mathbf{I}_\rho)$. By Theorem A.2, Bickel and Freedman (1981, Thm. 2.2), and the CMT, $\{Q_n^*(\theta)|\mathcal{X}_n\} \xrightarrow{d} H(\zeta)$ a.s. We now show that the $P(H(\zeta) \leq h)$ is continuous $h \neq 0$. For $h < 0$, the result holds because H is nonnegative. For $h > 0$, the result follows from Davydov et al. (1995, Thm. 11.1, part (i)) and that H is weakly convex and lower semicontinuous. Combining these arguments, it follows that, for any $h \neq 0$, $\lim_{n\to\infty} |P(Q_n^*(\theta) \leq h|\mathcal{X}_n) - \lim_{m\to\infty} P(Q_m(\theta) \leq h)| = 0$ a.s. By continuity of the limiting distribution everywhere except at zero, the convergence can be shown to hold uniformly for any $|h| \geq \mu$, i.e.,

$$P\left(\lim_{n\to\infty} \sup_{|h|\geq\mu} |P(Q_n^*(\theta) \leq h|\mathcal{X}_n) - \lim_{m\to\infty} P(Q_m(\theta) \leq h)| = 0\right) = 1.$$

By using elementary arguments, the previous equation implies the result.

Part 2. Suppose that $\theta \in Int(\Theta_I)$. By Theorem A.1, $\lim_{m\to\infty} P(Q_m(\theta) \leq h) = 1[h \geq 0]$. By Theorem A.2, $\liminf\{P(Q_n^*(\theta) \leq h|\mathcal{X}_n) = 1[h \geq 0]\}$ a.s. The result follows from combining these two findings. ■

Proof of Theorem 3.1. Part 1. Suppose that $\theta \in \partial\Theta_I$. For any $\mu > 0$, consider the following derivation:

$$\begin{aligned}
&|P(\theta \in \hat{C}_n^B(1-\alpha)) - (1-\alpha)| \\
&\leq \left\{ \begin{array}{l} |P(Q_n(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)) - P(H(\zeta) \leq \hat{c}_n^B(\theta, 1-\alpha))| \\ \quad + |P(H(\zeta) \leq \hat{c}_n^B(\theta, 1-\alpha)) - P(Q_n^*(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)|\mathcal{X}_n)| \\ \quad + |P(Q_n^*(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)|\mathcal{X}_n) - (1-\alpha)| \end{array} \right\} \\
&\leq \left\{ \begin{array}{l} \sup_{|h|\geq\mu} |P(Q_n^*(\theta) \leq h|\mathcal{X}_n) - P(H(\zeta) \leq h)| \\ \quad + \sup_{|h|\geq\mu} |P(Q_n(\theta) \leq h) - P(H(\zeta) \leq h)| \\ \quad + 2 \times 1[|\hat{c}_n^B(\theta, 1-\alpha)| < \mu] + |P(Q_n^*(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)|\mathcal{X}_n) - (1-\alpha)| \end{array} \right\}, \quad \textbf{(A.3)}
\end{aligned}$$

where H is the function in Theorem A.1 and $\zeta \sim N(\mathbf{0}_\rho, \mathbf{I}_\rho)$. The last expression on the RHS of equation (A.3) is the sum of four terms. We show that $\exists\mu > 0$ s.t. these terms are $o_p(1)$. By Theorem A.3 and its proof (which proves that $\lim_{m\to\infty} P(Q_m(\theta) \leq h) = P(H(\zeta) \leq h)$), the first term is $o_p(1)$. Using arguments analogous to those in the proof of Theorem A.3, the second term is $o(1)$. We now turn to the last two terms on the RHS of equation (A.3). By Theorem A.1, $H(\zeta) \leq 0$ implies $\exists b \neq \mathbf{0}_\rho$, s.t. $b'\zeta \leq 0$. Since $\zeta \sim N(\mathbf{0}_\rho, \mathbf{I}_\rho)$, $b'\zeta \sim N(\mathbf{0}_\rho, b'b)$ with $b'b > 0$, and, therefore, $P(H(\zeta) \leq 0) \leq P(b'\zeta \leq 0) \leq 0.5 < (1-\alpha)$. For any $p \in (0,1)$, let $c_\infty(p)$ denote the p-quantile of $H(\zeta)$. By the argument in Theorem A.3, $\forall h > 0$, $P(H(\zeta) \leq h)$ is continuous, and so, $\forall\alpha \in (0, 0.5)$, $P(H(\zeta) \leq c_\infty(1-\alpha)) = (1-\alpha)$, which implies that $c_\infty(1-\alpha) > 0$. Since $P(H(\zeta) \leq h)$ is continuous for $h > 0$, $\exists\varepsilon > 0$ so that $c_\infty(1-\alpha-\varepsilon) \geq \mu$ for some $\mu > 0$. Now, for $d \in (0, \varepsilon)$, consider the following derivation:

$$\begin{aligned}
&\{\sup_{|h|\geq\mu} |P(Q_n^*(\theta) \leq h|\mathcal{X}_n) - P(H(\zeta) \leq h)| \leq d\} \\
&\quad \subseteq \{|P(Q_n^*(\theta) \leq c_\infty(1-\alpha-\varepsilon)|\mathcal{X}_n) - (1-\alpha) + \varepsilon| \leq d\} \\
&\quad \subseteq \{P(Q_n^*(\theta) \leq c_\infty(1-\alpha-\varepsilon)|\mathcal{X}_n) < (1-\alpha) \leq P(Q_n^*(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)|\mathcal{X}_n)\} \\
&\quad \subseteq \{\mu < \hat{c}_n^B(\theta, 1-\alpha)\}, \quad \textbf{(A.4)}
\end{aligned}$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

where the first inclusion holds because $c_\infty(1-\alpha-\varepsilon) \geq \mu$ and the continuity of $P(H(\zeta) \leq h)$ at $h = c_\infty(1-\alpha-\varepsilon)$, the second inclusion holds because $d < \varepsilon$, and the last inclusion holds by the monotonicity of the CDF and that $c_\infty(1-\alpha-\varepsilon) \geq \mu$. By Theorem A.3 and equation (A.4), $2 \times 1[|\hat{c}_n^B(\theta, 1-\alpha)| < \mu] = o_p(1)$. For the same choice of μ, consider the following derivation:

$$
\begin{aligned}
&\{ \sup_{|h|\geq\mu} |P(Q_n^*(\theta) \leq h|\mathcal{X}_n) - P(H(\zeta) \leq h)| \leq \varepsilon/2\} \\
&\subseteq \{|P(Q_n^*(\theta) \leq c_\infty(1-\alpha+\varepsilon/2)|\mathcal{X}_n) - (1-\alpha) - \varepsilon/2| \leq \varepsilon/2\} \\
&\subseteq \{(1-\alpha) \leq P(Q_n^*(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)|\mathcal{X}_n) \leq P(Q_n^*(\theta) \\
&\leq c_\infty(1-\alpha+\varepsilon/2)|\mathcal{X}_n) \leq (1-\alpha)+\varepsilon\} \\
&\subseteq \{|P(Q_n^*(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)|\mathcal{X}_n) - (1-\alpha)| \leq \varepsilon\}, \qquad \textbf{(A.5)}
\end{aligned}
$$

where the first inclusion holds because $c_\infty(1-\alpha+\varepsilon/2) \geq c_\infty(1-\alpha-\varepsilon) \geq \mu$ and the continuity of $P(H(\zeta) \leq h)$ at $h = c_\infty(1-\alpha+\varepsilon/2)$, and the remaining ones are elementary. By Theorem A.3 and equation (A.5), $|P(Q_n^*(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)|\mathcal{X}_n) - (1-\alpha)| = o_p(1)$.

By collecting our arguments from equation (A.3), $|P(\theta \in \hat{C}_n^B(1-\alpha)) - (1-\alpha)| = o_p(1)$. Since $P(\theta \in \hat{C}_n^B(1-\alpha))$ is nonstochastic, $o_p(1)$ can be replaced by $o(1)$.

Part 2. Suppose that $\theta \in Int(\Theta_I)$. By Theorem A.1, $\liminf\{Q_n(\theta) = 0\}$ a.s. By definition, $\{Q_n^*(\theta) = 0\} \subseteq \{\hat{c}_n^B(\theta, 1-\alpha) = 0\}$, and, thus, Theorem A.2 implies that $\liminf\{\hat{c}_n^B(\theta, 1-\alpha) = 0\}$ a.s. Then, by elementary derivations, $\liminf\{Q_n(\theta) = \hat{c}_n^B(\theta, 1-\alpha)\}$ a.s., which implies that $P(\theta \in \hat{C}_n^B(1-\alpha)) = P(Q_n(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)) \to 1$. ■

LEMMA A.1. *Assume Assumptions A1–A3, CF1, and that θ satisfies Assumptions A4–A5. Let H be the function in Theorem A.1, $\{\varepsilon_n\}_{n\geq 1}$ be a positive sequence with $\varepsilon_n = o(1)$, and $\{g_n\}_{n\geq 1}$ be s.t. $g_n \to \infty$ and $\exp(-g_n/2) g_n^{(\rho-3)/2} \varepsilon_n^{-1} = O(1)$.*

1. *Let $\vartheta \sim N(\mathbf{0}_\rho, \Xi)$, where $\Xi \in \mathbb{R}^{\rho\times\rho}$ is nonsingular. Then, $\forall \mu > 0$,*

$$\sup_{|h|\geq\mu} |P(H(\vartheta) \in (h-\varepsilon_n, h+\varepsilon_n])| \leq O(\varepsilon_n\sqrt{g_n}). \qquad \textbf{(A.6)}$$

   Moreover, if we add Assumption CF2, then the $O(\varepsilon_n\sqrt{g_n})$ on the RHS of equation (A.6) can be replaced with $O(\varepsilon_n)$.

2. *Let $\{\vartheta_n|\mathcal{X}_n\} \sim N(\mathbf{0}_\rho, \Xi_n)$ where $\Xi_n \in \mathbb{R}^{\rho\times\rho}$ is conditionally nonstochastic and $\|\Xi_n - \Xi\| = o_p(1)$ for a nonsingular $\Xi \in \mathbb{R}^{\rho\times\rho}$. Then, $\forall \mu > 0$,*

$$\sup_{|h|\geq\mu} |P(H(\vartheta_n) \in (h-\varepsilon_n, h+\varepsilon_n]|\mathcal{X}_n)| \leq O_p(\varepsilon_n\sqrt{g_n}). \qquad \textbf{(A.7)}$$

   Moreover, if we add Assumption CF2, then the $O_p(\varepsilon_n\sqrt{g_n})$ on the RHS of equation (A.7) can be replaced with $O_p(\varepsilon_n)$.

Proof. Part 1. First, consider h s.t. $h \leq \mu$. Since H is nonnegative and $\varepsilon_n = o(1)$ then, for large enough $n \in \mathbb{N}$, $h + \varepsilon_n < 0$, and so $P(H(\vartheta) \leq h + \varepsilon_n) = 0$. Next, consider h s.t. $h \geq \mu$. Since $\varepsilon_n = o(1)$, then, for large enough $n \in \mathbb{N}$, $h - \varepsilon_n > 0$. By elementary derivations, $\forall C > 0$,

$$
\begin{aligned}
&P(H(\vartheta) \in (h-\varepsilon_n, h+\varepsilon_n]) \\
&\quad \leq P(\|\vartheta\| > C\sqrt{g_n}) + P(\vartheta \in H^{-1}((h-\varepsilon_n, h+\varepsilon_n]) \cap \|\vartheta\| \leq C\sqrt{g_n}). \qquad \textbf{(A.8)}
\end{aligned}
$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

We show that both terms on the RHS of equation (A.8) are $O(\varepsilon_n\sqrt{g_n})$, uniformly in $h \geq \mu$. We begin with the first term. By Hüsler, Liu, and Singh (2002, Thm. 1), $P(\|\vartheta\| > C\sqrt{g_n}) \leq O(\exp(-g_n/2)g_n^{\rho/2-1}) = O(\varepsilon_n\sqrt{g_n})$, uniformly in Ξ and in $h \geq \mu$. For the second term, consider the following derivation:

$$\begin{aligned}\{\vartheta \in H^{-1}((h-\varepsilon_n, h+\varepsilon_n]) \cap \|\vartheta\| \leq C\sqrt{g_n}\} &\subseteq \{\vartheta \in \{H^{-1}(\{h\})\}^{\gamma_n}\} \\ &= \{\xi \in \Xi^{-1/2}\{H^{-1}(\{h\})\}^{\gamma_n}\} \\ &\subseteq \{\xi \in \{\Xi^{-1/2}H^{-1}(\{h\})\}^{\eta_n}\} \\ &\subseteq \{\xi \in \{\partial[\Xi^{-1/2}H^{-1}((-\infty,h])]\}^{\eta_n}\}, \end{aligned} \tag{A.9}$$

where $\xi = \Xi^{-1/2}\vartheta \sim N(\mathbf{0}_\rho, \mathbf{I}_\rho)$, $\gamma_n = O(\varepsilon_n\sqrt{g_n})$, and $\eta_n = O(\varepsilon_n\sqrt{g_n})$, both uniformly in $h \geq \mu$. The first inclusion holds by definition of $\{h\}^{\varepsilon_n} \subseteq (h-\varepsilon_n, h+\varepsilon_n]$ and Theorem A.1, the following equality holds by setting $\xi \equiv \Xi^{-1/2}\vartheta \sim N(\mathbf{0}_\rho, \mathbf{I}_\rho)$, the second inclusion holds by $\Xi^{-1/2}\{H^{-1}(\{h\})\}^{\gamma_n} \subseteq \{\Xi^{-1/2}H^{-1}(\{h\})\}^{\eta_n}$ due to the submultiplicative property of the matrix norm, and the final inclusion holds by the continuity of H at $h > 0$ and the submultiplicative property of the norm, which guarantee that $H^{-1}(\{h\}) \subseteq \partial H^{-1}((-\infty,h])$ and $\Xi^{-1/2}\partial H^{-1}((-\infty,h]) = \partial\Xi^{-1/2}H^{-1}((-\infty,h])$. By the weak quasiconvexity of H, $\Xi^{-1/2}H^{-1}((-\infty,h]) \in \mathcal{C}_\rho$ and, thus, the second term on the RHS of equation (A.8) is bounded by $\sup_{A\in\mathcal{C}_\rho}\Phi_{\mathbf{I}_\rho}(\{\partial A\}^{\eta_n})$. By Bhattacharya and Rao (1976, Cor. 3.2) (with $s = 0$), this last expression is $O(\eta_n) = O(\varepsilon_n\sqrt{g_n})$, uniformly in $h \geq \mu$.

If we add Assumption CF2, Theorem A.1 implies that $\{H^{-1}((h-\varepsilon_n, h+\varepsilon_n]\} \subseteq \{H^{-1}(\{h\})\}^{\eta_n}$ where $\eta_n = O(\varepsilon_n)$, uniformly in $h \geq \mu$. Therefore, we bound the LHS of equation (A.8) by $\{\vartheta \in \{H^{-1}(\{h\})\}^{\gamma_n}\}$ without the need of conditioning on $\{\|\vartheta\| \leq C\sqrt{g_n}\}$. The desired result now follows from the previous arguments.

Part 2. Since $\Xi \in \mathbb{R}^{\rho\times\rho}$ is nonsingular, $\exists\varepsilon > 0$ s.t. any matrix $M \in \mathbb{R}^{\rho\times\rho}$ with $\|M - \Xi\| \leq \varepsilon$ is nonsingular. By arguments similar to those in part 1, $\forall C > 0$,

$$\begin{aligned} &P(H(\vartheta_n) \in (h-\varepsilon_n, h+\varepsilon_n]|\mathcal{X}_n) \\ &\leq \Phi_{\mathbf{I}_\rho}(\{\partial[\tilde{\Xi}_n^{-1/2}H^{-1}((-\infty,h])]\}^{\eta_n}) + P(||\tilde{\vartheta}_n|| > C\sqrt{g_n}|\mathcal{X}_n \cap \{\|\Xi_n - \Xi\| \leq \varepsilon\}) \\ &\quad + 1\{\|\Xi_n - \Xi\| > \varepsilon\} \\ &\leq \sup_{A\in\mathcal{C}_\rho}\Phi_{\mathbf{I}_\rho}(\{\partial A\}^{\eta_n}) + \sup_{\{\|M-\Xi\|\leq\varepsilon\}} P(\|\zeta_M\| > C\sqrt{g_n}) + 1\{\|\Xi_n - \Xi\| > \varepsilon\}, \end{aligned} \tag{A.10}$$

where $\tilde{\Xi}_n \equiv \Xi_n \times 1[||\Xi_n - \Xi|| \leq \varepsilon] + \Xi \times 1[||\Xi_n - \Xi|| > \varepsilon]$, $\zeta_M \sim N(\mathbf{0}_\rho, \mathbf{M})$, and $\eta_n = O(\varepsilon_n\sqrt{g_n})$, uniformly in $h \geq \mu$. We now show that the three terms on RHS of equation (A.10) are $O_p(\varepsilon_n\sqrt{g_n})$. By Bhattacharya and Rao (1976, Cor. 3.2) (with $s = 0$), the first term is $O(\varepsilon_n\sqrt{g_n})$, by Hüsler et al. (2002, Thm. 1), the second term is $O(\varepsilon_n\sqrt{g_n})$, and since $\|\Xi_n - \Xi\| = o_p(1)$, the third term is $O_p(\varepsilon_n\sqrt{g_n})$. The analysis when we add Assumption CF2 is analogous to the one in part 1 and is therefore omitted. ■

THEOREM A.4 (Rates of Convergence—Bootstrap). *Assume Assumptions A1–A3, CF1, and that θ satisfies Assumption A4.*

1. *If $\theta \in \partial\Theta_I$ and satisfies Assumption A5 then, $\forall\mu, \varepsilon > 0$,*

$$\sup_{|h|\geq\mu} |P(Q_n^*(\theta) \leq h|\mathcal{X}_n) - P(Q_n(\theta) \leq h)| \leq O_p(n^{-1/2}\sqrt{\ln n}). \tag{A.11}$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

Moreover, if we add Assumption CF2, then the $O_p(n^{-1/2}\sqrt{\ln n})$ on the RHS of equation (A.11) can be replaced with $O_p(n^{-1/2})$.

2. *If $\theta \in Int(\Theta_I)$, then*

$$P\left(\liminf\left\{\sup_{h\in\mathbb{R}}|P(Q_n^*(\theta)\le h|\mathcal{X}_n)-P(Q_n(\theta)\le h)|=0\right\}\right)=1.$$

Proof. Part 1. Suppose that $\theta \in \partial\Theta_I$. For any fixed $\mu > 0$, consider the following derivation:

$$\begin{aligned}&\sup_{|h|\ge\mu}|P(Q_n^*(\theta)\le h|\mathcal{X}_n)-P(Q_n(\theta)\le h)|\\&\quad\le\left\{\begin{array}{l}\sup_{|h|\ge\mu}|P(Q_n(\theta)\le h)-P(H(\zeta)\le h)|\\ \quad+\sup_{|h|\ge\mu}|P(Q_n^*(\theta)\le h|\mathcal{X}_n)-P(H(\hat{\zeta})\le h|\mathcal{X}_n)|\\ \quad+\sup_{|h|\ge\mu}|P(H(\zeta)\le h)-P(H(\hat{\zeta})\le h|\mathcal{X}_n)|\end{array}\right\},\end{aligned}\tag{A.12}$$

where H is the function in Theorem A.1, $\zeta \sim N(\mathbf{0}_\rho, \mathbf{I}_\rho)$, $\hat{\zeta} \sim N(\mathbf{0}_\rho, \hat{V})$, and $||\hat{V} - \mathbf{I}_\rho|| = O_p(n^{-1/2})$. The RHS of equation (A.12) is a sum of three terms. We show that the first term is $O(n^{-1/2}\sqrt{\ln n})$, the second term is $O_p(n^{-1/2}\sqrt{\ln n})$, and the third term is $O_p(n^{-1/2})$.

We begin with the first term. For $h \le \mu$, $|P(H(\zeta) \le h) - P(Q_n(\theta) \le h)| = 0$ because Q_n and $H(\zeta)$ are nonnegative. For any $h \ge \mu$ and by choosing $\varepsilon_n = n^{-1/2}$, we have the following derivation:

$$\begin{aligned}&\sup_{|h|\ge\mu}P(Q_n(\theta)\le h)-P(H(\zeta)\le h)\\&\quad\le\sup_{|h|\ge\mu}\{P(H(\sqrt{n}\bar{Y}_n)\le h+\varepsilon_n)-P(H(\zeta)\le h+\varepsilon_n)\\&\qquad+P(H(\zeta)\in(h-\varepsilon_n,h+\varepsilon_n])+P(|\delta_n|>\varepsilon_n)\}\\&\quad\le\sup_{A\in\mathcal{C}_\rho}|P(\sqrt{n}\bar{Y}_n\in A)-\Phi_{\mathbf{I}_\rho}(A)|\\&\qquad+\sup_{|h|\ge\mu}P(H(\zeta)\in(h-\varepsilon_n,h+\varepsilon_n])+P(|\delta_n|>\varepsilon_n),\end{aligned}\tag{A.13}$$

where δ_n is the random variable described in Theorem A.1. The RHS of equation (A.13) is a sum of three terms. The first term is $O(n^{-1/2})$ by the Berry–Esseen theorem, the second term is $O(n^{-1/2}(\ln(n^{1+2\gamma}))^{1/2}) = O(n^{-1/2}\sqrt{\ln n})$ by part 1 of Lemma A.1 and $g_n = \ln(n^{1+2\gamma})$ for any $\gamma > 0$, and the third term is $o(n^{-1/2})$ by Theorem A.1. Combining the three terms, the RHS is $O(n^{-1/2}\sqrt{\ln n})$. By repeating this argument for $P(Q_n(\theta) > h)$ (instead of $P(Q_n(\theta) \le h)$), we can conclude that the first term on the RHS of equation (A.12) is $O(n^{-1/2}\sqrt{\ln n})$. If we add Assumption CF2, then the same argument and Lemma A.1 imply that the RHS of equation (A.12) is $O(n^{-1/2})$.

The argument for the second term on the RHS of equation (A.12) is analogous to the one used for the first term, except that it is based on the Berry–Esseen theorem, part 2 of Lemma A.1, and Theorem A.2. Finally, the third term on the RHS of equation (A.12) is bounded above by $\int_{\mathbb{R}^\rho}|\phi_{\hat{V}}(x)dx - \phi_{\mathbf{I}_\rho}(x)|dx$, which can be shown to be $O(||\hat{V} - \mathbf{I}_\rho||) = O_p(n^{-1/2})$ by the mean value theorem and properties of the normal distribution.

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

Part 2. If $\theta \in Int(\Theta_I)$, Theorems A.1 and A.2 imply that $\liminf\{Q_n(\theta) = 0 \cap Q_n^*(\theta) = 0\}$ a.s. The result then follows from the fact that $\{Q_n(\theta) = 0 \cap Q_n^*(\theta) = 0\}$ implies that $P(Q_n^*(\theta) \leq h|\mathcal{X}_n) = P(Q_n(\theta) \leq h)$ for all $h \in \mathbb{R}$. ■

Proof of Theorem 3.2. Part 1. Suppose that $\theta \in \partial\Theta_I$. For any fixed $K, \mu > 0$ and $n \in \mathbb{N}$, consider the following derivation:

$$\begin{aligned}
&\{|P(\theta \in \hat{C}_n^B(1-\alpha)) - (1-\alpha)| > Kn^{-1/2}\sqrt{\ln n}\} \\
&\quad = \{|P(Q_n(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)) - (1-\alpha)| > Kn^{-1/2}\sqrt{\ln n}\} \\
&\quad \subseteq \left\{ \begin{array}{l} \sup_{|h|\geq\mu}\{|P(Q_n(\theta) \leq h) - P(Q_n^*(\theta) \leq h|\mathcal{X}_n)| > (K/2)n^{-1/2}\sqrt{\ln n}\} \\ \quad \cup\{\hat{c}_n^B(\theta, 1-\alpha) < \mu\} \cup \{|P(Q_n^*(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)|\mathcal{X}_n) - (1-\alpha)| \\ \quad > (K/2)n^{-1/2}\sqrt{\ln n}\} \end{array} \right\},
\end{aligned}$$

and, therefore,

$$\begin{aligned}
&P(|P(\theta \in \hat{C}_n^B(1-\alpha)) - (1-\alpha)| > Kn^{-1/2}\sqrt{\ln n}) \\
&\quad \leq \left\{ \begin{array}{l} P(\sup_{|h|\geq\mu}|P(Q_n(\theta) \leq h) - P(Q_n^*(\theta) \leq h|\mathcal{X}_n)| > (K/2)n^{-1/2}\sqrt{\ln n}) \\ \quad + P(\hat{c}_n^B(\theta, 1-\alpha) < \mu) + P(|P(Q_n^*(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)|\mathcal{X}_n) - (1-\alpha)| \\ \quad > (K/2)n^{-1/2}\sqrt{\ln n}) \end{array} \right\}. \quad \textbf{(A.14)}
\end{aligned}$$

The RHS of equation (A.14) is a sum of three terms. Fix $\delta > 0$ arbitrarily. We now show that for a particular choice of K and $\mu > 0$, the RHS of equation (A.14) is less than δ. We begin with the first term. By Theorem A.4, $\exists K_1 > 0$ s.t. $\forall K \geq K_1$ and $\forall n \in \mathbb{N}$, the first term is less than $\delta/3$. We now consider the second term. For all $p \in [0, 1]$, let $c_\infty(p)$ denote the p-quantile of $H(\zeta)$, where H is the function in Theorem A.2 and $\zeta \sim N(\mathbf{0}_\rho, \mathbf{I}_\rho)$. By arguments in Theorem 3.1, $\exists \varepsilon, \mu > 0$ with $c_\infty(1-\alpha-\varepsilon) \geq \mu$. We fix $\mu > 0$ to this value. By arguments in Theorem 3.1, $P(\hat{c}_n^B(\theta, 1-\alpha) \geq \mu) = 1 + o(1)$ and, so, $\exists N \in \mathbb{N}$ s.t. $\forall n \geq N$, the second term is less than $\delta/3$. Finally, we consider the third term on the RHS of equation (A.14). For any fixed $K > 0$ and $n \in \mathbb{N}$, consider the following derivation:

$$\begin{aligned}
&P\left(\sup_{|h|\geq\mu}|P(Q_n^*(\theta) \leq h|\mathcal{X}_n) - P(H(\zeta) \leq h)| \leq (K/4)n^{-1/2}\sqrt{\ln n}\right) \\
&\quad \leq P\left(\left| \begin{array}{l} P(Q_n^*(\theta) \leq c_\infty(1-\alpha+(K/4)n^{-1/2}\sqrt{\ln n})|\mathcal{X}_n) \\ \quad -P(H(\zeta) \leq c_\infty(1-\alpha+(K/4)n^{-1/2}\sqrt{\ln n})) \end{array} \right| \leq (K/4)n^{-1/2}\sqrt{\ln n}\right) \\
&\quad = P\left(\begin{array}{l} (1-\alpha) \leq P(Q_n^*(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)|\mathcal{X}_n) \\ \quad \leq P(Q_n^*(\theta) \leq c_\infty(1-\alpha+(K/4)n^{-1/2}\sqrt{\ln n})|\mathcal{X}_n) \\ \quad \leq (1-\alpha) + (K/2)n^{-1/2}\sqrt{\ln n} \end{array} \right) \\
&\quad \leq P(|P(Q_n^*(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)|\mathcal{X}_n) - (1-\alpha)| \leq (K/2)n^{-1/2}\sqrt{\ln n}), \quad \textbf{(A.15)}
\end{aligned}$$

where the first inequality holds by $c_\infty(1-\alpha-\varepsilon) \geq \mu$ and the monotonicity of the CDF, the first equality holds by definition of quantile, and the final inequality holds by elementary arguments. By arguments used in the proof of Theorem A.4, $\forall \delta > 0$, $\exists K > 0$ s.t. $\forall n \in \mathbb{N}$ the first line in equation (A.15) is greater than or equal to $(1-\delta/3)$. By equation (A.15), it follows that $\forall \delta > 0, \exists K_2 > 0$ s.t. $\forall K \geq K_2$ and $\forall n \in \mathbb{N}$, $P(|P(Q_n^*(\theta) \leq \hat{c}_n^B(\theta, 1-\alpha)|\mathcal{X}_n) - (1-\alpha)| > (K/2)n^{-1/2}\sqrt{\ln n}) < \delta/3$. Then, by setting $K = \max\{K_1, K_2\}$, it follows that

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

$\exists N \in \mathbb{N}$ and $K, \mu > 0$ s.t. $\forall n \geq N$, the RHS of equation (A.14) is less than δ. By appropriate choice of $K > 0$, the result can be extended $\forall n \in \mathbb{N}$.

This implies that $P(\theta \in \hat{C}_n^B(1-\alpha)) = (1-\alpha) + O_p(n^{-1/2}\sqrt{\ln n})$. Since this event is nonstochastic, it follows that $P(\theta \in \hat{C}_n^B(1-\alpha)) = (1-\alpha) + O(n^{-1/2}\sqrt{\ln n})$. Under Assumption CF2, the same argument can be repeated with $O(n^{-1/2})$ instead of $O(n^{-1/2}\sqrt{\ln n})$.

Part 2. If $\theta \in Int(\Theta_I)$, then $E[m_j(Z,\theta)] > 0$ for all $j = 1, \ldots, J$. Define $\eta \equiv \min_{j=1,\ldots,J} E[m_j(Z,\theta)] > 0$. Consider the following derivation:

$$\{\theta \notin \hat{C}_n^B(1-\alpha)\} = \{Q_n(\theta) > \hat{c}_n^B(\theta, 1-\alpha)\} \subseteq \cup_{j=1}^J \{\sqrt{n}\bar{m}_{n,j}(\theta) < 0\}$$
$$\subseteq \cup_{j=1}^J \{v_n(m_{j,\theta}) < -\sqrt{n}\eta\}.$$

Thus, $P(\theta \notin \hat{C}_n^B(1-\alpha)) \leq \sum_{i=1}^J P(v_n(m_{j,\theta}) < -\sqrt{n}\eta)$. By Chebyshev's inequality, $P(v_n(m_{j,\theta}) < -\sqrt{n}\eta) = O(n^{-1})$ for all $j = 1, \ldots, J$. ■

A.5. Results for Asymptotic Approximation

THEOREM A.5. *Assume Assumptions A1–A3, CF1, and that θ satisfies Assumption refass:A4. Let $\rho \equiv rank(V[m(Z,\theta)])$.*

1. *If $\theta \in \partial\Theta_I$, then $Q_n^{AA}(\theta) = H(\vartheta_n) + \delta_n^{AA}$, where*
   (a) *for any positive sequence $\{\varepsilon_n\}_{n\geq 1}$ with $\varepsilon_n = O(n^{-1/2})$, $P(|\delta_n^{AA}| > \varepsilon_n | \mathcal{X}_n) = o(n^{-1/2})$ a.s.*
   (b) *$\{\vartheta_n | \mathcal{X}_n\} : \Omega \to \mathbb{R}^\rho$ is a zero (conditional) mean sample average of n i.i.d. observations from a distribution with a (conditional) variance-covariance matrix $\hat{V}$ that is nonsingular a.s. and finite (conditional) third absolute moments a.s. If Assumption A5 is added, then $||\hat{V} - \mathbf{I}_\rho|| \leq O_p(n^{-1/2})$ and the distribution has finite (conditional) fourth absolute moments.*
   (c) *$H : \mathbb{R}^\rho \to \mathbb{R}$ is the same function as in Theorem A.1.*
2. *If $\theta \in Int(\Theta_I)$, then $\liminf\{Q_n^{AA}(\theta) = 0\}$ a.s.*

Proof. This proof follows closely the arguments used to prove Theorem A.2. In part 1(a), we define $\delta_n^{AA}(\theta) = Q_n^{AA}(\theta) - G(\{[v_n^{AA}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J)$, where $v_n^{AA}(m_{j,\theta}) \equiv n^{-1/2}\sum_{i=1}^n \zeta_i(m_j(Z_i,\theta) - \bar{m}_{n,j}(\theta))$ for $j = 1, \ldots, n$. Conditional on the sample, $v_n^{AA}(m_\theta)$ is a linear combination of normal random variables and, thus, $\{v_n^{AA}(m_\theta)|\mathcal{X}_n\} \sim N(\mathbf{0}_J, \hat{\Sigma})$, where $\hat{\Sigma}$ denotes the sample variance of $m(Z,\theta)$. The remainder of the argument follows from the one used to prove Theorem A.2. ■

THEOREM A.6 (Consistency on Any Set Excluding Zero—AA). *Assume Assumptions A1–A3, CF1, and that θ satisfies Assumption A4.*

1. *If $\theta \in \partial\Theta_I$, then, $\forall \mu, \varepsilon > 0$,*

$$\lim P\left(\sup_{|h|\geq\mu} \left| P(Q_n^{AA}(\theta) \leq h|\mathcal{X}_n) - \lim_{m\to\infty} P(Q_m(\theta) \leq h) \right| \leq \varepsilon \right) = 1.$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

2. *If* $\theta \in Int(\Theta_I)$, *then*

$$P\left(\liminf\left\{\sup_{h\in\mathbb{R}}\left|P(Q_n^{AA}(\theta)\le h|\mathcal{X}_n)-\lim_{m\to\infty}P(Q_m(\theta)\le h)\right|=0\right\}\right)=1.$$

Proof. This proof follows closely the arguments used to prove Theorem A.3. The only difference is that Theorem A.2 and Bickel and Freedman (1981, Thm. 2.2) are replaced by Theorem A.5 and $\{v_n^{AA}(m_\theta)|\mathcal{X}_n\}\sim N(\mathbf{0}_J,\hat{\Sigma})$, respectively. ■

THEOREM A.7 (Consistency in Level—AA). *Assume Assumptions A1–A3, CF1, and* θ *satisfies Assumption A4.*

1. *If* $\theta\in\partial\Theta_I$, *then,* $\forall\alpha\in(0,0.5)$, $\lim P(\theta\in\hat{C}_n^{AA}(1-\alpha))=1-\alpha$.
2. *If* $\theta\in Int(\Theta_I)$, *then,* $\forall\alpha\in[0,1]$, $\lim P(\theta\in\hat{C}_n^{AA}(1-\alpha))=1$.

Proof. This proof follows closely the arguments used to prove Theorem 3.1. ■

THEOREM A.8 (Rates of Convergence—AA). *Assume Assumptions A1–A3, CF1, and that* θ *satisfies Assumption A4.*

1. *If* $\theta\in\partial\Theta_I$ *and satisfies Assumption A5, then* $\forall\mu>0$,

$$\sup_{|h|\ge\mu}|P(Q_n^{AA}(\theta)\le h|\mathcal{X}_n)-P(Q_n(\theta)\le h)|\le O_p(n^{-1/2}\sqrt{\ln n}). \tag{A.16}$$

Moreover, if we add Assumption CF2, then the $O_p(n^{-1/2}\sqrt{\ln n})$ *on the RHS of equation (A.16) can be replaced with* $O_p(n^{-1/2})$.
2. *If* $\theta\in Int(\Theta_I)$, *then*

$$P\left(\liminf\left\{\sup_{h\in\mathbb{R}}\left|P(Q_n^{AA}(\theta)\le h|\mathcal{X}_n)-P(Q_n(\theta)\le h)\right|=0\right\}\right)=1.$$

Proof. This proof follows closely the arguments used to prove Theorem A.4. ■

Proof of Theorem 4.1. This proof follows closely the arguments used to prove Theorem 3.2. ■

A.6. Results for Subsampling Approximations

A.6.1. Additional Assumptions In order to provide sharp rate of convergence results for our subsampling procedures, we require the following additional technical assumptions.

Assumption A7. $m_j(Z,\theta)$ is absolutely continuous with respect to Lebesgue measure for all $j=1,\ldots,J$.

Assumption A8. This assumption requires first introducing some notation. Let $\rho\equiv rank(V[m(Z,\theta)])$, the function $H:\mathbb{R}^\rho\to\mathbb{R}_+$, and the distribution Υ be as in Theorem A.1. Then, let $c_\infty(\theta,1-\alpha)$ denote the $(1-\alpha)$-quantile of $H(\zeta)$ with $\zeta\sim N(\mathbf{0}_\rho,\mathbf{I}_\rho)$ and, for $Y\sim\Upsilon$, let $K_1:\mathcal{C}_\rho\to\mathbb{R}$ denote the following function:

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

$$K_1(S) \equiv - \sum_{t \in \{\ell \in \mathbb{N}^\rho : \sum_{j=1}^\rho \ell_j = 3\}} \frac{1}{\prod_{j=1,\ldots,\rho} t_j!} E\left[\prod_{j=1,\ldots,\rho} (Y_j - E[Y_j])^{t_j} \right] \int_{x \in S} D^t \phi_{\mathbf{I}_\rho}(x) dx,$$

where, for any $t \in \mathbb{N}^\rho$, $D^t \phi_{\mathbf{I}_\rho}(x)dx$ denotes the $(\sum_{j=1}^\rho t_j)$-th partial derivative of $\phi_{\mathbf{I}_\rho}(x)$, t_1 times with respect to x_1, t_2 times with respect to x_2, and so on. This assumption requires that

$K_1(H^{-1}((-\infty, c_\infty(\theta, 1-\alpha)])) > 0$ for $\rho = 1$,
$K_1(H^{-1}((-\infty, c_\infty(\theta, 1-\alpha)])) < 0$ for $\rho > 1$.

We briefly comment on Assumptions A7–A8. Like Assumption A6, Assumption A7 can be interpreted as a smoothness condition on the distribution of $m(Z, \theta)$. In fact, Hall (1992, p. 65) shows that Assumption A7 implies Assumption A6. In combination with Assumption A5, Assumption A7 allows us to establish an Edgeworth expansion for the subsampling distribution, which is the key in obtaining the convergence rate of the ECP of the subsampling approximation. The role of AssumptionA8 is to determine the sign of one of the leading terms in this Edgeworth expansion. In turn, this restriction allows us to establish a sharp rate of convergence of the ECP of the subsampling approximation. This result is important for this paper, as it provides sufficient conditions under which the rate of convergence of the subsampling approximation is strictly slower than that of the bootstrap or the asymptotic approximation, regardless of the choice of subsampling size.

Remark A.1. Assumption A8 is admittedly a high-level condition that is hard to interpret in practice. We now provide a description of this condition. By definition, $K_1 : \mathcal{C}_\rho \to \mathbb{R}$ is a function of convex sets in $\mathbb{R}^\rho$ that depends on: (a) integrals of partial derivatives of the standard normal distribution and (b) the matrix of third centered moments of the distribution of $Y \sim \Upsilon$, which is a standardized zero-mean version of $m(Z, \theta)$ (see the definition in Theorem A.1). Given a parameter value $\theta \in \Theta$, Assumption A8 imposes that the function K_1 applied to the convex set $H^{-1}((-\infty, c_\infty(\theta, 1-\alpha)])$ has to have a specific sign (positive if $\rho = 1$ or negative if $\rho > 1$). Given the structure of the model and the choice of $\theta \in \Theta$, Assumption A8 can then be understood as a restriction on the sign of a complicated function of the matrix of third centered moments of the standardized zero-mean version of $m(Z, \theta)$.

In order to make this concrete, consider the following identified set $\Theta_I = \{\theta \in \Theta : E[Z] - \theta \geq 0\}$, where $E[Z] = 0$ and $V[Z] = 1$. In this case, $\partial\Theta_I = \{0\}$. In this setup, $\rho = J = 1$, $H(y) \equiv [y]_- 1[E[Z] = \theta]$, $Z \sim \Upsilon$, $\mathcal{C}_1$ is the set of convex intervals in $\mathbb{R}$, and, $\forall x_L, x_H \in \mathbb{R} \cup \{\pm\infty\}$ s.t. $x_L < x_H$,

$$K_1([x_L, x_H]) \equiv - \frac{1}{3!} E[Z^3] \int_{x_L}^{x_H} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)(3x - x^3)dx.$$

Now suppose that we are interested in conducting inference for $\theta = 0$ (note that $\theta \in \partial\Theta_I$ is the only relevant case for Assumption A8, see Lemma A.4 and Theorem A.13). When $\theta = 0$, $c_\infty(0, 1-\alpha)$ is the $(1-\alpha)$-quantile of $[\zeta]_-$ with $\zeta \sim N(0, 1)$. For any $\alpha < 0.5$, $H^{-1}((-\infty, c_\infty(\theta, 1-\alpha)]) = [-c_\infty(\theta, 1-\alpha), \infty)$ with $c_\infty(\theta, 1-\alpha) > 0$ and, thus,

$$\begin{aligned} K_1(H^{-1}((-\infty, c_\infty(\theta, 1-\alpha)])) &= -\frac{1}{3!} E[Z^3] \int_{-c_\infty(\theta,1-\alpha)}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)(3x - x^3)dx \\ &= (\phi(c_\infty(\theta, 1-\alpha))/3!) \times (c_\infty(\theta, 1-\alpha)^2 - 1)E[Z^3]. \end{aligned}$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

In this example, Assumption A8 occurs if and only if $(c_\infty(\theta, 1-\alpha)^2 - 1)E[Z^3] > 0$. In particular, $\alpha \leq 0.3$ implies that $c_\infty(\theta, 1-\alpha) > 1$ and, therefore, Assumption A8 is satisfied as long as Z has positive skewness.

A.6.2. Results for Subsampling 1

THEOREM A.9. *Assume Assumptions A1–A3, CF1, and that θ satisfies Assumption A4. Let $\rho \equiv rank(V[m(Z,\theta)])$.*

1. *If $\theta \in \partial\Theta_I$, then $Q^{SS_1}_{b_n,n}(\theta) = H(\sqrt{b_n}\bar{Y}^{SS}_{b_n,n}) + \delta^{SS_1}_{b_n,n}$, where*
   (a) $\liminf\{P(\delta^{SS_1}_{b_n,n} = 0|\mathcal{X}_n) \geq 1[||v_n(m_\theta)|| \leq \tau_n]\}$ *and* $\liminf\{\delta^{SS_1}_{b_n,n} = 0\}$ *a.s.*
   (b) $\{\tilde{Y}^{SS}_{b_n,n}|\mathcal{X}_n\} : \Omega_n \to \mathbb{R}^\rho$ *is a zero (conditional) mean sample average of b_n observations sampled without replacement from a distribution with a (conditional) variance-covariance matrix $\hat{V}$ that is nonsingular a.s. and finite (conditional) third absolute moments a.s. If Assumption A5 is added, then $||\hat{V} - \mathbf{I}_\rho|| \leq O_p(n^{-1/2})$ and the distribution has finite (conditional) fourth absolute moments a.s.*
   (c) $H : \mathbb{R}^\rho \to \mathbb{R}$ *is the same function as in Theorem A.1.*
2. *If $\theta \in Int(\Theta_I)$, then* $\liminf\{Q^{SS_1}_{b_n,n}(\theta) = 0\}$ *a.s.*

Proof. This proof follows closely the arguments used to prove Theorem A.2. We only point out the differences, which occur in parts 1(a) and 1(b).

We begin with part 1(a). Define $\delta^{SS_1}_{b_n,n} = Q^{SS_1}_{b_n,n}(\theta) - G(\{[v^{SS}_{b_n,n}(m_{j,\theta})] - 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J)$. Let $S \equiv \{j \in \{1,\ldots,J\} : E[m_j(Z,\theta)] = 0\}$, $\bar{S} \equiv \{1,\ldots,J\}/S$, and $\eta \equiv \min_{j\in\bar{S}} E[m_j(Z,\theta)] > 0$. By the argument as in the proof of Theorem A.2, $\liminf\{\{\|v_n(m_\theta)\| \leq \tau_n\} \subseteq \{\delta^{SS_1}_{b_n,n} = 0\}\}$. This has two implications. First, it implies that $\liminf\{P(\|v_n(m_\theta)\| \leq \tau_n|\mathcal{X}_n) \leq P(\delta^{SS_1}_{b_n,n} = 0|\mathcal{X}_n)\}$ and since $\{\|v_n(m_\theta)\| \leq \tau_n\}$ is nonstochastic conditional on $\mathcal{X}_n$, it follows that $\liminf\{P(\delta^{SS_1}_{b_n,n} = 0|\mathcal{X}_n) \geq 1[||v_n(m_\theta)|| \leq \tau_n]\}$. Second, it implies that $\liminf\{\|v_n(m_\theta)\| \leq \tau_n\} \subseteq \liminf\{\delta^{SS_1}_{b_n,n} = 0\}$. By the LIL, $P(\liminf\{\|v_n(m_\theta)\| \leq \tau_n\}) = 1$ and so $\liminf\{\delta^{SS_1}_{b_n,n} = 0\}$ a.s.

We now consider part 1(b). The difference between this result and the one in Theorem A.2 lies in the resampling method. While both use the realized sample as their source to generate random observations, subsampling constructs samples of size b_n without replacement while the bootstrap constructs samples of size n with replacement.

We now cover some details of this proof that will become necessary in the proof of Lemma A.2. Let the matrices $\Sigma_\rho \in \mathbb{R}^{\rho\times\rho}$, $B \in \mathbb{R}^{J\times\rho}$, and the sample $\{Y_i\}_{i=1}^n$ be as in proof of Theorem A.1, i.e., $Y_i \equiv \Sigma_\rho^{-1/2}\{m_j(Z_i,\theta) - \bar{m}_{n,j}(\theta)\}_{j=1}^\rho$ for all $i = 1,\ldots,n$. Let $\{Z_i^{SS}\}_{i=1}^{b_n}$ be a sample of b_n observations sampled without replacement from $\{Z_i\}_{i=1}^n$ (subsampling). Then, we define $Y_i^{SS} \equiv \Sigma_\rho^{-1/2}\{m_j(Z_i^{SS},\theta)\}_{j=1}^\rho$ and $\tilde{Y}_i^{SS} \equiv \Sigma_\rho^{-1/2}\{m_j(Z_i^{SS},\theta) - \bar{m}_{n,j}(\theta)\}_{j=1}^\rho$ for all $i = 1,\ldots,n$. These can be shown to satisfy $BY_i^{SS} = m(Z_i^{SS},\theta)$ and $B\tilde{Y}_i^{SS} = m(Z_i^{SS},\theta) - \bar{m}_n(\theta)$ for all $i = 1,\ldots,n$. By these

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

definitions, the sample $\{\tilde{Y}_i^{SS}\}_{i=1}^n$ and its average $\tilde{Y}_{b_n,n}^{SS} \equiv {b_n}^{-1} \sum_{i=1}^{b_n} \tilde{Y}_i^{SS}$ can be shown to have all the desired properties in part 1(b) by using arguments that are similar to those in the proof of Theorem A.2. ■

THEOREM A.10 (Consistency on Any Set Excluding Zero—SS1). *Assume Assumptions A1-A3, CF1, and that θ satisfies Assumption A4.*

1. *If $\theta \in \partial\Theta_I$, then, $\forall \mu, \varepsilon > 0$,*

$$\lim P\left(\sup_{|h|\geq\mu} \left| P(Q_{b_n,n}^{SS_1}(\theta) \leq h|\mathcal{X}_n) - \lim_{m\to\infty} P(Q_m(\theta) \leq h) \right| \leq \varepsilon \right) = 1.$$

2. *If $\theta \in Int(\Theta_I)$, then,*

$$P\left(\liminf \left\{ \sup_{h\in\mathbb{R}} \left| P(Q_{b_n,n}^{SS_1}(\theta) \leq h|\mathcal{X}_n) - \lim_{m\to\infty} P(Q_m(\theta) \leq h) \right| = 0 \right\}\right) = 1.$$

Proof. This proof follows closely the arguments used to prove Theorem A.3. The only difference is that Theorem A.2 and Bickel and Freedman (1981, Thm. 2.2) are replaced by Theorem A.9 and Politis, Romano, and Wolf (1999, Thm. 2.2.1), respectively. ■

THEOREM A.11 (Consistency in Level—SS1). *Assume Assumptions A1–A3, CF1, and θ satisfies Assumption A4.*

1. *If $\theta \in \partial\Theta_I$, then, $\forall \alpha \in (0, 0.5)$, $\lim P(\theta \in \hat{C}_{b_n,n}^{SS_1}(1-\alpha)) = 1-\alpha$.*
2. *If $\theta \in Int(\Theta_I)$, then, $\forall \alpha \in [0, 1]$, $\lim P(\theta \in \hat{C}_{b_n,n}^{SS_1}(1-\alpha)) = 1$.*

Proof. This proof follows closely the arguments used to prove Theorem 3.1. ■

LEMMA A.2. *Assume Assumptions A1–A3, CF1, and that θ satisfies Assumptions A4 and A6. Then, the conditional distribution of $\tilde{Y}_{b_n,n}^{SS}$ in Theorem A.9 satisfies the following representation*

$$P\left(\sqrt{b_n}\tilde{Y}_{b_n,n}^{SS} \in S|\mathcal{X}_n\right) = \Phi_{\mathbf{I}_\rho}(S) + K_1(S)b_n^{-1/2} + K_2(S)b_n/n + o_p\left(b_n^{-1/2} + b_n/n\right),$$

uniformly in $S \in \mathcal{C}_\rho$, where $K_1 : \mathcal{C}_\rho \to \mathbb{R}$ and $K_2 : \mathcal{C}_\rho \to \mathbb{R}$ satisfy $\sup_{S\in\mathcal{C}_\rho} |K_1(S)| < \infty$ and $\sup_{S\in\mathcal{C}_\rho} |K_2(S)| < \infty$.

Proof. Let $\{Y_i^{SS}\}_{i=1}^{b_n}$ and $\{\tilde{Y}_i^{SS}\}_{i=1}^{b_n}$ be the random samples constructed by subsampling from the original data sample in $\{Y_i\}_{i=1}^n$ in Theorem A.9. Recall that $\tilde{Y}_{b_n,n}^{SS} \equiv \bar{Y}_{b_n,n}^{SS} - \bar{Y}_n$, where $\tilde{Y}_{b_n,n}^{SS} \equiv {b_n}^{-1} \sum_{i=1}^{b_n} \tilde{Y}_i^{SS}$, $\bar{Y}_{b_n,n}^{SS} \equiv {b_n}^{-1} \sum_{i=1}^{b_n} Y_i^{SS}$, and $\bar{Y}_n \equiv n^{-1} \sum_{i=1}^n Y_i$.

For any $S \in \mathcal{C}_\rho$, let $S_n(S) = \{x \in \mathbb{R}^\rho : (1 - b_n/n)^{1/2} x \in S\}$. Consider the following derivation:

$$\begin{aligned} P\left(\sqrt{b_n}\tilde{Y}_{b_n,n}^{SS} \in S|\mathcal{X}_n\right) &= P\left(\sqrt{b_n}(\bar{Y}_{b_n,n}^{SS} - \bar{Y}_n) \in S|\mathcal{X}_n\right) \\ &= P\left(\sqrt{b_n}(1 - b_n/n)^{-1/2}(\bar{Y}_{b_n,n}^{SS} - \bar{Y}_n) \in S_n(S)|\mathcal{X}_n\right) \end{aligned}$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

$$
\begin{aligned}
&= \Phi_{\mathbf{I}_\rho}(S_n(S)) + b_n^{-1/2} K_1(S_n(S)) + o_p(b_n^{-1/2}) \\
&= \Phi_{\mathbf{I}_\rho}(S) + (b_n/n) K_2(S) + b_n^{-1/2} K_1(S) \\
&\quad + o_p(b_n^{-1/2} + b_n/n), \qquad \textbf{(A.17)}
\end{aligned}
$$

uniformly in $S \in \mathcal{C}_\rho$, where for $\zeta \sim N(\mathbf{0}_\rho, \mathbf{I}_\rho)$ and $\forall S \in \mathcal{C}_\rho$, $K_1(S)$ and $K_2(S)$ are defined by

$$
K_1(S) \equiv - \sum_{t \in \{\ell \in \mathbb{N}^\rho : \sum_{j=1}^\rho \ell_j = 3\}} \frac{1}{\prod_{j=1,\ldots,\rho} t_j!} E\left[\prod_{j=1,\ldots,\rho} (Y_j - E[Y_j])^{t_j} \right] \int_{x \in S} D^t \phi_{\mathbf{I}_\rho}(x) dx,
$$

$$
K_2(S) \equiv P(\zeta \in S) E[1 - \zeta'\zeta | \zeta \in S]. \qquad \textbf{(A.18)}
$$

The first and second equalities in equation (A.17) are given by the definitions of $\tilde{Y}_{b_n,n}^{SS}$ and $S_n(S)$. The third equality is given by a combination of Babu and Singh (1985, Thm. 1) and arguments in Bhattacharya and Rao (1976, Thm. 3.1). Specifically, Babu and Singh (1985, Thm. 1) provide an Edgeworth expansion for averages of samples extracted without replacement from a finite population. By rearranging terms, the leading terms of this expansion are functions of sample moments and integrals based on the normal distribution. Replacing sample moments by population moments introduces an error term that is $O_p(n^{-1/2}) = o_p(b_n^{-1/2})$, uniformly in $S \in \mathcal{C}_\rho$, and giving rise to the formula for $K_1(\cdot)$ in equation (A.18). The fourth equality in equation (A.17) holds because for every $S \in \mathcal{C}_\rho$, a Taylor expansion and change of variables imply

$$
\begin{aligned}
\Phi_{\mathbf{I}_\rho}(S_n(S)) &= \Phi_{\mathbf{I}_\rho}(S) + (b_n/n) K_2(S) + o(b_n/n) \\
b_n^{-1/2} K_1(S_n(S)) &= b_n^{-1/2} K_1(S) + o(b_n^{-1/2}),
\end{aligned}
$$

both uniformly in $S \in \mathcal{C}_\rho$. Since the normal distribution has finite absolute moments of all orders, it follows that $\sup_{S \in \mathcal{C}_\rho} |K_1(S)| < \infty$ and $\sup_{S \in \mathcal{C}_\rho} |K_2(S)| < \infty$. ■

THEOREM A.12 (Rates of Convergence—SS1). *Assume Assumptions A1–A3, CF1, and that θ satisfies Assumption A4.*

1. *If $\theta \in \partial\Theta_I$ and satisfies Assumptions A5–A6, then, $\forall \mu > 0$,*

$$
\sup_{|h| \geq \mu} |P(Q_{b_n,n}^{SS_1}(\theta) \leq h | \mathcal{X}_n) - P(Q_n(\theta) \leq h)| \leq O_p(b_n^{-1/2} + b_n/n).
$$

2. *If $\theta \in Int(\Theta_I)$, then*

$$
P\left(\liminf \left\{ \sup_{h \in \mathbb{R}} \left| P(Q_{b_n,n}^{SS_1}(\theta) \leq h | \mathcal{X}_n) - P(Q_n(\theta) \leq h) \right| = 0 \right\} \right) = 1.
$$

Proof. Part 1. For any fixed $\mu > 0$, consider the following derivation:

$$
\begin{aligned}
&\sup_{|h| \geq \mu} |P(Q_{b_n,n}^{SS_1}(\theta) \leq h | \mathcal{X}_n) - P(Q_n(\theta) \leq h)| \\
&\quad \leq \sup_{|h| \geq \mu} |P(Q_{b_n,n}^{SS_1}(\theta) \leq h | \mathcal{X}_n) - P(H(\zeta) \leq h)| \\
&\qquad + \sup_{|h| \geq \mu} |P(H(\zeta) \leq h) - P(Q_n(\theta) \leq h)|, \qquad \textbf{(A.19)}
\end{aligned}
$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

where H is function in Theorem A.1 and $\zeta \sim N(\mathbf{0}_\rho, \mathbf{I}_\rho)$. The RHS of equation (A.19) is a sum of two terms. By arguments in the proof of Theorem A.4, the second term is $O(n^{-1/2}\sqrt{\ln n})$. We now focus on the first term. For any positive sequence $\{\varepsilon_n\}_{n\geq 1}$ s.t. $\varepsilon_n = O(n^{-1/2})$, consider the following derivation:

$$\begin{aligned}
&\sup_{|h|\geq\mu} |P(Q^{SS_1}_{b_n,n}(\theta) \leq h|\mathcal{X}_n) - P(H(\zeta) \leq h)| \\
&\quad \leq \left\{ \begin{array}{l} \sup_{|h|\geq\mu} |P(H(\sqrt{b_n}\bar{Y}^{SS}_{b_n,n}) \leq h+\varepsilon_n|\mathcal{X}_n) - P(H(\zeta) \leq h+\varepsilon_n)| \\ \quad + \sup_{|h|\geq\mu} P(H(\zeta) \in (h-\varepsilon_n, h+\varepsilon_n]) + P(|\delta^{SS_1}_{b_n,n}| > \varepsilon_n|\mathcal{X}_n) \end{array} \right\} \\
&\quad \leq \sup_{S\in\mathcal{C}_\rho} |P(\sqrt{b_n}\bar{Y}^{SS}_{b_n,n} \in S|\mathcal{X}_n) - \Phi_{\mathbf{I}_\rho}(S)| \\
&\qquad + \sup_{|h|\geq\mu} \Phi_{\mathbf{I}_\rho}(H^{-1}((h-\varepsilon_n, h+\varepsilon_n])) + P(|\delta^{SS_1}_{b_n,n}| > \varepsilon_n|\mathcal{X}_n).
\end{aligned}$$

The RHS is a sum of three terms. The first term is $O_p(b_n^{-1/2} + b_n/n)$ by Lemma A.2, the second term is $O(n^{-1/2}\sqrt{\ln n})$ by setting $g_n = \ln(n^{1+2\gamma})$ for any $\gamma > 0$, and, according to Theorem A.9, the third term is bounded above by $1[||v_n(m_\theta)|| > \tau_n]$ which, in turn, is $o_p(n^{-1/2})$ by the LIL. Combining these arguments with $n^{-1/2}\sqrt{\ln n} = O(b_n^{-1/2} + b_n/n)$, the RHS of equation (A.19) is $O_p(b_n^{-1/2} + b_n/n)$.

Part 2. This part follows from arguments that are analogous to the ones used to prove part 2 of Theorem A.4. ■

Proof of Theorem 5.1. This proof follows closely the arguments used to prove Theorem 3.2. ■

LEMMA A.3. *For any $\gamma \in (0, 0.28)$, let h_L and h_H be s.t. $P(H(\zeta) \leq h_L) = 0.72$ and $P(H(\zeta) \leq h_H) = 1-\gamma$. Also, let $\Lambda(\gamma) \equiv \{S \in \mathcal{C}_\rho : \exists h \in [h_L, h_H] : S = \{y \in \mathbb{R}^\rho : H(y) \leq h\}\}$, where H is defined as in Theorem A.1, $\rho \equiv rank(V[m(Z,\theta)])$, and $\zeta \sim N(\mathbf{0}_\rho, \mathbf{I}_\rho)$. Then,*

$$\begin{aligned}
&\inf_{S\in\Lambda(\gamma)} K_2(S) > 0 \quad \textit{for } \rho = 1, \\
&\sup_{S\in\Lambda(\gamma)} K_2(S) < 0 \quad \textit{for } \rho > 1,
\end{aligned}$$

where K_2 is the function defined in Lemma A.2.

Proof. Case 1: $\rho = 1$. Consider the following derivation:

$$\begin{aligned}
\inf_{S\in\Lambda(\gamma)} K_2(S) &= \inf_{\{\{y_1,y_2\}\in\mathbb{R}_+\cup\{\infty\}:[-y_1,y_2]\in\Lambda(\gamma)\}} K_2([-y_1, y_2]) \\
&= \inf_{\{\{y_1,y_2\}\in\mathbb{R}_+\cup\{\infty\}:[-y_1,y_2]\in\Lambda(\gamma)\}} \{K_2([0, \min\{y_1,y_2\}]) + K_2([0, \max\{y_1,y_2\}])\} \\
&\geq \inf_{\eta_L\leq y\leq\eta_H} K_2([0,y]) + \inf_{\eta_L\leq y} K_2([0,y]) \\
&= \min\{K_2([0,\eta_L]), K_2([0,\eta_H])\}, \qquad \textbf{(A.20)}
\end{aligned}$$

where η_L, $\eta_H \in (0,\infty)$ are s.t. $P(\zeta \leq \eta_L) = 0.72$ and $P(|\zeta| \leq \eta_H) = 1-\gamma$. The first equality in equation (A.20) holds because H is homogenous of degree $\beta \geq 1$ and,

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

thus, $y \in S$ implies that $\lambda y \in S$ for all $\lambda \in [0,1]$, the second equality holds because K_2 is s.t. $K_2([-y_1, y_2]) = K_2([0, \min\{y_1, y_2\}]) + K_2([0, \max\{y_1, y_2\}])$. To prove the next weak inequality in equation (A.20) it suffices to show that $\min\{y_1, y_2\} \geq \eta_L$ and $\max\{y_1, y_2\} \leq \eta_H$. To see that $\min\{y_1, y_2\} \geq \eta_L$, notice that $[-y_1, y_2] \in \Lambda(\gamma)$ implies that $0.72 \leq P(\zeta \in [-y_1, y_2]) \leq P(\zeta \leq \min\{y_1, y_2\})$ which, together with the definition of η_L, implies the result. To see that $\min\{y_1, y_2\} \leq \eta_H$, notice that $[-y_1, y_2] \in \Lambda(\gamma)$ implies that $1 - \gamma \geq P(\zeta \in [-y_1, y_2]) \geq P(|\zeta| \leq \min\{y_1, y_2\})$ which, together with the definition of η_H, implies the result. Finally, the last equality in equation (A.20) holds because $K_2([0, y])$ is a nonnegative function that is increasing in $y \in [0,1]$, decreasing for $y \in [1, \infty]$, and s.t. $K_2([0, \infty]) = 0$. We can set $C \equiv \min\{K_2([0, \eta_L]), K_2([0, \eta_H])\} > 0$, where the strict inequality follows from the fact that $K_2([0, y]) > 0$ for all $y \in (0, \infty)$. Then, equation (A.20) implies that $\exists C > 0$ s.t. $\inf_{S \in \Lambda(\gamma)} K_2(S) \geq C > 0$.

Case 2: $\rho > 1$. Fix $S \in \Lambda(\gamma)$ arbitrarily. Let $\zeta_\rho \sim N(\mathbf{0}_\rho, \mathbf{I}_\rho)$ and $\forall (\rho, \pi) \in \mathbb{N} \times [0,1]$, let $c(\pi, \rho)$ be (uniquely) defined by $P(\zeta_\rho' \zeta_\rho \leq c(\pi, \rho)) = \pi$. According to this definition,

$$
\begin{aligned}
&P(\zeta_\rho \in \{S \cap \{x \in \mathbb{R}^\rho : x'x > c(P(\zeta_\rho \in S), \rho)\}\}) \\
&= P(\zeta_\rho \in S) - P(\zeta_\rho \in \{S \cap \{x \in \mathbb{R}^\rho : x'x \leq c(P(\zeta_\rho \in S), \rho)\}\}) \\
&= P(\zeta_\rho' \zeta_\rho \leq c(P(\zeta_\rho \in S), \rho)) - P(\zeta_\rho \in \{S \cap \{x \in \mathbb{R}^\rho : x'x \leq c(P(\zeta_\rho \in S), \rho)\}\}) \\
&= P(\zeta_\rho \in \{\{S\}^c \cap \{x \in \mathbb{R}^\rho : x'x \leq c(P(\zeta_\rho \in S), \rho)\}\}), \qquad \textbf{(A.21)}
\end{aligned}
$$

where the second equality follows from the fact that $P(\zeta_\rho' \zeta_\rho \leq c(P(\zeta_\rho \in S), \rho)) = P(\zeta_\rho \in S)$. Based on this, consider the following derivation:

$$
\begin{aligned}
K_2(S) &= \int_{x \in S} (1 - x'x) d\Phi_{\mathbf{I}_\rho}(x) \\
&\leq \int_{x \in \{S \cap \{x'x \leq c(P(\vartheta_\rho \in S), \rho)\}\}} (1 - x'x) d\Phi_{\mathbf{I}_\rho}(x) \\
&\quad + \int_{x \in \{S \cap \{x'x > c(P(\vartheta_\rho \in S), \rho)\}\}} (1 - c(P(\vartheta_\rho \in S), \rho)) d\Phi_{\mathbf{I}_\rho}(x) \\
&= \int_{x \in \{S \cap \{x'x \leq c(P(\vartheta_\rho \in S), \rho)\}\}} (1 - x'x) d\Phi_{\mathbf{I}_\rho}(x) \\
&\quad + \int_{x \in \{\{S\}^c \cap \{x'x \leq c(P(\vartheta_\rho \in S), \rho)\}\}} (1 - c(P(\vartheta_\rho \in S), \rho)) d\Phi_{\mathbf{I}_\rho}(x) \\
&\leq P(\zeta_\rho \in S)(1 - E[\zeta_\rho' \zeta_\rho | \zeta_\rho' \zeta_\rho \leq c(P(\zeta_\rho \in S), \rho)]), \qquad \textbf{(A.22)}
\end{aligned}
$$

where the derivations use the definition of K_2 and equation (A.21). Since $S \in \Lambda(\gamma)$, $c(\pi, \rho)$ is increasing in π for every ρ, and since $E[\zeta_\rho' \zeta_\rho | \zeta_\rho' \zeta_\rho \leq c(0.72, \rho)]$ is increasing in ρ and greater than one for $\rho = 2$, then equation (A.22) implies that $\forall \rho \geq 2$, $\exists C > 0$, $\sup_{S \in \Lambda(\gamma)} K_2(S) \leq -C < 0$. ■

LEMMA A.4. *Assume Assumptions A1–A3, CF1, that $\theta \in \partial\Theta_I$ satisfies Assumptions A4, A6, and A8, and let $\rho = rank(V[m(Z, \theta)])$. For any $\alpha \in (0, 0.5)$, $\exists \eta, C > 0$ s.t.*

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

1. *If $\rho = 1$, then*

$$P\Bigg(\inf_{h\in[c_\infty(\theta,1-\alpha-\eta),c_\infty(\theta,1-\alpha+\eta)]}(P(Q^{SS_1}_{b_n,n}(\theta)\le h|\mathcal{X}_n)-P(Q_n(\theta)\le h)) \le C(b_n^{-1/2}+b_n/n)\Bigg)=o(1). \tag{A.23}$$

If $\rho > 1$, then

$$P\Bigg(\sup_{h\in[c_\infty(\theta,1-\alpha-\eta),c_\infty(\theta,1-\alpha+\eta)]}(P(Q^{SS_1}_{b_n,n}(\theta) \le h|\mathcal{X}_n)-P(Q_n(\theta)\le h)) \ge -C(b_n^{-1/2}+b_n/n)\Bigg)=o(1). \tag{A.24}$$

2. $P(\hat{c}^{SS_1}_{b_n,n}(\theta,1-\alpha)\notin[c_\infty(\theta,1-\alpha)-\eta, c_\infty(\theta,1-\alpha)+\eta])=o(1)$.

Proof. Part 1. Suppose that $\theta\in\partial\Theta_I$ and fix $\mu>0$ arbitrarily. For any h, consider the following derivation:

$$P(Q^{SS_1}_{b_n,n}(\theta)\le h|\mathcal{X}_n)-P(Q_n(\theta)\le h) = \left\{\begin{array}{l} P(Q^{SS_1}_{b_n,n}(\theta)\le h|\mathcal{X}_n)-P(H(\sqrt{b_n}\bar{Y}^{SS}_{b_n,n})\le h|\mathcal{X}_n) \\ \quad +P(H(\zeta)\le h)-P(Q_n(\theta)\le h) \\ \quad +P(H(\sqrt{b_n}\bar{Y}^{SS}_{b_n,n})\le h|\mathcal{X}_n)-P(H(\zeta)\le h) \end{array}\right\}, \tag{A.25}$$

where H is function in Theorem A.1 and $\zeta\sim N(\mathbf{0}_\rho,\mathbf{I}_\rho)$. The RHS of equation (A.25) is a sum of three terms. The first term is $o_p(n^{-1/2})$ (uniformly in $h\in\mathbb{R}$) by Theorem A.9 and the second term is $O_p(n^{-1/2}\sqrt{\ln n})$ (uniformly in $|h|\ge\mu$) by arguments used in the proof of Theorem A.4. By handling the third term with the representation in Lemma A.2 and combining all these results, it follows that

$$P(Q^{SS_1}_{b_n,n}(\theta)\le h|\mathcal{X}_n)-P(Q_n(\theta)\le h)=K_1(H^{-1}((-\infty,h]))b_n^{-1/2} + K_2(H^{-1}((-\infty,h]))b_n/n+v_n(h), \tag{A.26}$$

where $\sup_{|h|\ge\mu}|v_n(h)|=o_p(b_n^{-1/2}+b_n/n)$, and, to deduce this, we have used that $(n^{-1/2}\sqrt{\ln n})/(b_n^{-1/2}+b_n/n)=o(1)$.

By Theorem 3.1, $c_\infty(1-\alpha)>0$. Then, by arguments in Theorem A.3, $\exists\eta>0$ s.t. $[c_\infty(1-\alpha-\eta),c_\infty(1-\alpha+\eta)]\subseteq\{s\in\mathbb{R}:|s|\ge\mu\}$. Since $K_1(H^{-1}((-\infty,h]))$ is continuous in $h\in\{s\in\mathbb{R}:|s|\ge\mu\}$ and since $K_1(H^{-1}((-\infty,c_\infty(1-\alpha)]))>0$ for $\rho=1$ and <0 for $\rho>1$, $\exists\eta_1,C_1>0$ s.t. $\inf_{[c_\infty(\theta,1-\alpha-\eta_1),c_\infty(\theta,1-\alpha+\eta_1)]}K_1(H^{-1}((-\infty,h]))\ge C_1>0$ for $\rho=1$ and $\sup_{[c_\infty(\theta,1-\alpha-\eta_1),c_\infty(\theta,1-\alpha+\eta_1)]}K_1(H^{-1}((-\infty,h]))\le -C_1<0$ for $\rho>1$. By Lemma A.3, $\exists\eta_2,C_2>0$ s.t. $\inf_{[c_\infty(\theta,1-\alpha-\eta_2),c_\infty(\theta,1-\alpha+\eta_2)]}K_2(H^{-1}((-\infty,h]))\ge C_2>0$ for $\rho=1$, and $\sup_{[c_\infty(\theta,1-\alpha-\eta_2),c_\infty(\theta,1-\alpha+\eta_2)]}K_2(H^{-1}((-\infty,h]))\le -C_2<0$ for $\rho>1$. Define $\eta\equiv\min\{\eta_1,\eta_2\}>0$ and $C\equiv\min\{C_1,C_2\}/2>0$. If $\rho=1$,

$$\inf_{h\in[c_\infty(\theta,1-\alpha-\eta),c_\infty(\theta,1-\alpha+\eta)]}(K_1(H^{-1}((-\infty,h]))b_n^{-1/2} + K_2(H^{-1}((-\infty,h]))b_n/n)\ge 2C(b_n^{-1/2}+b_n/n). \tag{A.27}$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

If $\rho > 1$,

$$\sup_{h\in[c_\infty(\theta,1-\alpha-\eta),c_\infty(\theta,1-\alpha+\eta)]} (K_1(H^{-1}((-\infty,h]))b_n^{-1/2} + K_2(H^{-1}((-\infty,h]))b_n/n) \leq -2C(b_n^{-1/2} + b_n/n).$$

In the case when $\rho = 1$, consider the following derivation:

$$\begin{aligned}
&\left\{\inf_{h\in[c_\infty(\theta,1-\alpha-\eta),c_\infty(\theta,1-\alpha+\eta)]}[P(Q^{SS_1}_{b_n,n}(\theta)\leq h|\mathcal{X}_n) - P(Q_n(\theta)\leq h)] \leq C(b_n^{-1/2}+b_n/n)\right\}\\
&= \left\{\inf_{h\in[c_\infty(\theta,1-\alpha-\eta),c_\infty(\theta,1-\alpha+\eta)]}[K_1(H^{-1}((-\infty,h]))b_n^{-1/2} + K_2(H^{-1}((-\infty,h]))b_n/n + v_n(h)] \leq C(b_n^{-1/2}+b_n/n)\right\}\\
&\subseteq \left\{\inf_{h\in[c_\infty(\theta,1-\alpha-\eta),c_\infty(\theta,1-\alpha+\eta)]}[K_1(H^{-1}((-\infty,h]))b_n^{-1/2} + K_2(H^{-1}((-\infty,h]))b_n/n] - \sup_{|h|\geq\mu}|v_n(h)| \leq C(b_n^{-1/2}+b_n/n)\right\}\\
&\subseteq \left\{\sup_{|h|\geq\mu}|v_n(h)| \geq C(b_n^{-1/2}+b_n/n)\right\},
\end{aligned}$$

where the first inclusion uses equation (A.26), and the remaining steps follow from equation (A.27). Since $\sup_{|h|\geq\mu}|v_n(h)| = o_p(b_n^{-1/2}+b_n/n)$, equation (A.23) follows. If $\rho > 1$, an analogous argument can be used to deduce equation (A.24).

Part 2. Let $\eta > 0$ be the value chosen in the previous step which, by definition, implies that $[c_\infty(1-\alpha-\eta), c_\infty(1-\alpha+\eta)] \subseteq \{s\in\mathbb{R} : |s|\geq\mu\}$. Now consider the following derivation for $\varepsilon \in (0,\eta)$,

$$\begin{aligned}
&\sup_{|h|\geq\mu}\{|P(Q^{SS_1}_{b_n,n}(\theta)\leq h|\mathcal{X}_n) - P(H(\zeta)\leq h)| \leq \varepsilon\}\\
&\quad\subseteq \{|P(Q^{SS_1}_{b_n,n}(\theta)\leq c_\infty(1-\alpha-\eta)|\mathcal{X}_n) - (1-\alpha) + \eta| \leq \varepsilon\} \cap \{|P(Q^{SS_1}_{b_n,n}(\theta)\\
&\qquad\leq c_\infty(1-\alpha+\eta)|\mathcal{X}_n) - (1-\alpha) - \eta| \leq \varepsilon\}\\
&\quad\subseteq \{P(Q^{SS_1}_{b_n,n}(\theta)\leq c_\infty(1-\alpha-\eta)|\mathcal{X}_n) \leq (1-\alpha) - \eta + \varepsilon\} \cap \{(1-\alpha) + \eta - \varepsilon\\
&\qquad\leq P(Q^{SS_1}_{b_n,n}(\theta)\leq c_\infty(1-\alpha+\eta)|\mathcal{X}_n)\}\\
&\quad\subseteq \{P(Q^{SS_1}_{b_n,n}(\theta)\leq c_\infty(1-\alpha-\eta)|\mathcal{X}_n) < (1-\alpha)\} \cap \{(1-\alpha) < P(Q^{SS_1}_{b_n,n}(\theta)\\
&\qquad\leq c_\infty(1-\alpha+\eta)|\mathcal{X}_n)\}\\
&\quad\subseteq \{\hat{c}^{SS_1}_{b_n,n}(\theta,1-\alpha) \in [c_\infty(1-\alpha-\eta), c_\infty(1-\alpha+\eta)]\}.
\end{aligned}$$

This result and Theorem A.10 imply that $P(\hat{c}^{SS_1}_{b_n,n}(\theta,1-\alpha) \notin [c_\infty(1-\alpha-\eta), c_\infty(1-\alpha+\eta)]) = o(1)$. ■

THEOREM A.13. *Assume Assumptions A1–A3, CF1, and that $\theta \in \partial\Theta_I$ satisfies Assumptions A4–A6 and A7–A8. For any $\alpha \in (0, 0.5)$, $\exists C > 0$ and $\exists N \in \mathbb{N}$ s.t. $\forall n \geq N$:*

$$|P(\theta \in \hat{C}^{SS_1}_{b_n,n}(1-\alpha)) - (1-\alpha)| > C(b_n^{-1/2}+b_n/n). \tag{A.28}$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

In particular, if we use $\rho \equiv rank(V[m(Z,\theta)])$, $\exists C > 0$ *and* $\exists N \in \mathbb{N}$ *s.t.* $\forall n \geq N$,

$$P(\theta \in \hat{C}^{SS_1}_{b_n,n}(1-\alpha)) - (1-\alpha) < -C(b_n^{-1/2} + b_n/n) \quad \text{if } \rho = 1,$$
$$P(\theta \in \hat{C}^{SS_1}_{b_n,n}(1-\alpha)) - (1-\alpha) > C(b_n^{-1/2} + b_n/n) \quad \text{if } \rho > 1. \tag{A.29}$$

Proof. Notice that equation (A.28) follows immediately from equation (A.29). We only show equation (A.29) for $\rho = 1$, as the result for $\rho > 1$ is completely analogous. By Lemma A.4, $\forall \varepsilon > 0$, $\exists \eta, C > 0$ and $N_1 \in \mathbb{N}$ s.t. $\forall n \geq N_1$,

$$P\left(\inf_{h \in [c_\infty(\theta, 1-\alpha-\eta), c_\infty(\theta, 1-\alpha+\eta)]} (P(Q^{SS_1}_{b_n,n}(\theta) \leq h | \mathcal{X}_n) - P(Q_n(\theta) \leq h)) > 2C(b_n^{-1/2} + b_n/n)\right) \geq 1 - \varepsilon/4,$$
$$P(\hat{c}^{SS_1}_{b_n,n}(\theta, 1-\alpha) \notin [c_\infty(1-\alpha-\eta), c_\infty(1-\alpha+\eta)]) \leq \varepsilon/4.$$

As a consequence, we have the following derivation:

$$\begin{aligned}
1 - \varepsilon/4 &\leq P\left(\inf_{h \in [c_\infty(\theta, 1-\alpha-\eta), c_\infty(\theta, 1-\alpha+\eta)]} (P(Q^{SS_1}_{b_n,n}(\theta) \leq h | \mathcal{X}_n) \right. \\
&\qquad \left. - P(Q_n(\theta) \leq h)) > 2C(b_n^{-1/2} + b_n/n)\right) \\
&\leq \left\{ \begin{array}{l} P(P(Q^{SS_1}_{b_n,n}(\theta) \leq \hat{c}^{SS_1}_{b_n,n}(\theta, 1-\alpha) | \mathcal{X}_n) - P(Q_n(\theta) \leq \hat{c}^{SS_1}_{b_n,n}(\theta, 1-\alpha)) \\ \quad > 2C(b_n^{-1/2} + b_n/n)) + P(\hat{c}^{SS_1}_{b_n,n}(\theta, 1-\alpha) \notin [c_\infty(1-\alpha-\eta), c_\infty(1-\alpha+\eta)]) \end{array} \right\} \\
&\leq P(q^{SS_1}_{b_n,n}(\theta, 1-\alpha) - P(\theta \in \hat{C}^{SS_1}_{b_n,n}(1-\alpha)) > 2C(b_n^{-1/2} + b_n/n)) + \varepsilon/4,
\end{aligned}$$

where $q^{SS_1}_{b_n,n}(\theta, 1-\alpha) \equiv P(Q^{SS_1}_{b_n,n}(\theta) \leq \hat{c}^{SS_1}_{b_n,n}(\theta, 1-\alpha) | \mathcal{X}_n)$. From this, it follows that $\forall \varepsilon > 0$, $\exists \eta, C > 0$ and $N_1 \in \mathbb{N}$ s.t. $\forall n \geq N_1$

$$P(q^{SS_1}_{b_n,n}(\theta, 1-\alpha) - P(\theta \in \hat{C}^{SS_1}_{b_n,n}(1-\alpha)) > 2C(b_n^{-1/2} + b_n/n)) \geq 1 - \varepsilon/2. \tag{A.30}$$

We now show that:

$$P(|q^{SS_1}_{b_n,n}(1-\alpha) - (1-\alpha)| > n^{-1/2}) = o(1). \tag{A.31}$$

Set $k_n \equiv n!/(b_n!(n-b_n)!)$, i.e., the total number of subsamples of size b_n without replacement from a sample of size n. For any sample realization, the subsample distribution is a discrete distribution with masspoints at the possible values that can be obtained by sampling without replacement. Denote these subsamples by $s = 1, \ldots, k_n$, i.e., $\{m(Z_i^s, \theta)\}_{i=1}^{b_n}$ refers to the s^{th} subsample. By properties of the combinatorial formula, $\exists N_2$ s.t., $\forall n \geq N_2$, $k_n^{-1} \leq n^{-1/2}$. Define the following event:

$$A_n \equiv \left\{ \bigcap_{\substack{S_1, S_2 \subseteq \{1, \ldots, J\} \\ S_1 \neq \emptyset, S_2 \neq \emptyset}} \bigcap_{\substack{d_2, d_1 = 1, \ldots, k_n \\ d_1 \neq d_2}} \left\{ \begin{array}{l} G(\{\sqrt{b_n}[b_n^{-1} \sum_{i=1}^{b_n} m_{j_1}(Z_i^{d_1}, \theta) \\ \quad - \bar{m}_{n,j_1}(\theta)]\}_{j_1 \in S_1}) \\ \neq G(\{\sqrt{b_n}[b_n^{-1} \sum_{i=1}^{b_n} m_{j_2}(Z_i^{d_2}, \theta) \\ \quad - \bar{m}_{n,j_2}(\theta)]\}_{j_2 \in S_2}) \end{array} \right\} \right\}.$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

By definition, $\{q^{SS_1}_{b_n,n}(1-\alpha) \geq 1-\alpha > 0\}$. Then, if A_n occurs, the $(1-\alpha)$-quantile is achieved at a particular realization of $Q^{SS_1}_{b_n,n}(\theta)$ which, by definition, implies that there is a specific subsample of size b_n of $\{m(Z_i,\theta)\}_{i=1}^n$, denoted by $\{m(Z_i^d,\theta)\}_{i=1}^{b_n}$, s.t.

$$
\begin{aligned}
Q^{SS_1}_{b_n,n}(\theta) &= G(\{[\sqrt{b_n}(\bar{m}^{SS}_{n,j}(\theta) - \bar{m}_{n,j}(\theta))]_- 1[\sqrt{n}\bar{m}_{n,j}(\theta) \leq \tau_n]\})_{j=1}^J \\
&= G(\{\sqrt{b_n}[b_n^{-1}\sum\nolimits_{i=1}^{b_n} m_j(Z_i^d,\theta) - \bar{m}_{n,j}(\theta)]\}_{j\in S_1}),
\end{aligned}
$$

where the first equality holds by definition and the second equality holds for a set $S_1 \subseteq \{1,\ldots,J\}$ with $S_1 \neq \emptyset$ of indices j for which $\{\sqrt{n}\bar{m}_{n,j}(\theta) \leq \tau_n\}$ and $\sqrt{b_n}(\bar{m}^{SS}_{n,j}(\theta) - \bar{m}_{n,j}(\theta)) > 0$. Clearly, if A_n occurs, then there cannot be another subsample of size b_n without replacement out of the n observations that achieves the same value and, thus, $|q^{SS_1}_{b_n,n}(1-\alpha) - (1-\alpha)| \leq k_n^{-1} \leq n^{-1/2}$. As a consequence, for $n \geq N_2$, the event $\{|q^{SS_1}_{b_n,n}(1-\alpha) - (1-\alpha)| > n^{-1/2} \cap A_n\}$ does not occur and, so, $P(|q^{SS_1}_{b_n,n}(1-\alpha) - (1-\alpha)| > n^{-1/2} \cap A_n) = o(1)$. On the other hand, $P(|q^{SS_1}_{b_n,n}(1-\alpha) - (1-\alpha)| > n^{-1/2} \cap A_n^c) \leq P(A_n^c)$. In order to complete the proof of equation (A.31), it suffices to show that $P(A_n^c) = o(1)$. We now show that $P(A_n^c) = 0$ for all $n \in \mathbb{N}$. To prove this, it suffices to show that $P(B_n(s_1,s_2,S_1,S_2)) = 0$ for any $S_1, S_2 \subseteq \{1,\ldots,J\}$, $S_1 \neq \emptyset$, $S_2 \neq \emptyset$, and $s_2, s_1 = 1,\ldots,k_n$ s.t. $s_1 \neq s_2$, where the event $B_n(s_1,s_2,S_1,S_2)$ is defined as follows:

$$
B_n(s_1,s_2,S_1,S_2) \equiv \left\{ \begin{array}{l} G(\{\sqrt{b_n}(b_n^{-1}\sum_{i=1}^{b_n} m_{j_1}(Z_i^{s_1},\theta) - \bar{m}_{n,j_1}(\theta))\}_{j_1\in S_1}) \\ \quad = G(\{\sqrt{b_n}(b_n^{-1}\sum_{i=1}^{b_n} m_{j_2}(Z_i^{s_2},\theta) - \bar{m}_{n,j_2}(\theta))\}_{j_2\in S_2}) \end{array} \right\}.
$$

Notice that the indices of the subsamples s_1 and s_2 that are involved in the definition of $B_n(s_1,s_2,S_1,S_2)$ are fixed but, nevertheless, $B_n(s_1,s_2,S_1,S_2)$ is a random event because the sample is random. By definition of subsampling, $\{m(Z_i^{s_1},\theta)\}_{i=1}^{b_n}$ and $\{m(Z_i^{s_2},\theta)\}_{i=1}^{b_n}$ are composed of different observations. Let ℓ denote an arbitrary index of observation that belongs to $\{m(Z_i^{s_1},\theta)\}_{i=1}^{b_n}$ (and, thus, not to $\{m(Z_i^{s_2},\theta)\}_{i=1}^{b_n}$). Then,

$$
\begin{aligned}
&B_n(s_1,s_2,S_1,S_2) \\
&\quad = \left\{ \begin{array}{l} G(\{\sqrt{b_n}m_{j_1}(Z_\ell,\theta)(b_n^{-1} - n^{-1}) + \sqrt{b_n}(b_n^{-1}\sum_{i=1,i\neq\ell}^{b_n} m_{j_1}(Z_i^{s_1},\theta) \\ \quad\quad - n^{-1}\sum_{u=1,i\neq\ell}^{n} m_{j_1}(Z_u,\theta))\}_{j_1\in S_1}) \\ = G(\{-\sqrt{b_n}m_{j_2}(Z_\ell,\theta)n^{-1} + \sqrt{b_n}(b_n^{-1}\sum_{i=1,i\neq\ell}^{b_n}(m_{j_2}(Z_i^{s_2},\theta)) \\ \quad\quad - n^{-1}\sum_{u=1,i\neq\ell}^{n}(m_h(Z_{j_2},\theta)))\}_{j_2\in S_2}) \end{array} \right\}.
\end{aligned}
$$

By the i.i.d. assumption,

$$
\begin{aligned}
&P(B_n(s_1,s_2,S_1,S_2)) \\
&\quad = \int P\left(\begin{array}{l} G(\{\sqrt{b_n}m_{j_1}(Z_\ell,\theta)(b_n^{-1} - n^{-1}) + M(j_1,s_1,\ell,\{z_i\}_{i\neq\ell})\}_{j_1\in S_1}) \\ \quad = G(\{-\sqrt{b_n}m_{j_2}(Z_\ell,\theta)n^{-1} + M(j_2,s_2,\ell,\{z_i\}_{i\neq\ell})\}_{j_2\in S_2}) \end{array} \right) dP(\{Z_i\}_{i\neq\ell} \\
&\quad = \{z_i\}_{i\neq\ell}),
\end{aligned}
$$

where we have used $M(j,s,\ell,\{z_i\}_{i\neq\ell}) \equiv \sqrt{b_n}(b_n^{-1}\sum_{i=1,i\neq\ell}^{b_n}(m_j(z_i^s,\theta)) - n^{-1}\sum_{u=1,i\neq\ell}^{n}(m_j(z_u,\theta)))$ for all $s = s_1, s_2$, ℓ, j, and $\{z_i\}_{i\neq\ell}$. For any $S_1, S_2 \subseteq \{1,\ldots,J\}$,

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

$S_1 \neq \emptyset$, $S_2 \neq \emptyset$, $C \in \mathbb{R}^J$, consider the following set:

$$\{m \in \mathbb{R}^J : G(\{\sqrt{b_n}(b_n^{-1} - n^{-1})m_{j_1} + C_{j_1}\}_{j_1 \in S_1}) = G(\{-\sqrt{b_n}n^{-1}m_{j_2} + C_{j_2}\}_{j_2 \in S_2})\}. \tag{A.32}$$

For any arbitrary $g \in S_1 \neq \emptyset$ and $\{\tilde{m}_i\}_{i \neq g} \in \mathbb{R}^{J-1}$, there is only one value of $\tilde{m}_g \in \mathbb{R}$ s.t. $\tilde{m}$ that belongs to the set in equation (A.32) as G is strictly increasing in every coordinate. Since the choice of S_1 was arbitrary, the set in equation (A.32) has zero Lebesgue measure in at least one of its coordinates. Since $m_j(Z, \theta)$ is assumed to be absolutely continuous with respect to Lebesgue measure for all $j = 1, \ldots, J$, it follows that $P(B_n(s_1, s_2, S_1, S_2)) = 0$ for any $S_1, S_2 \subseteq \{1, \ldots, J\}$, $S_1 \neq \emptyset$, $S_2 \neq \emptyset$, and any $s_2, s_1 = 1, \ldots, k_n$ with $s_1 \neq s_2$. Therefore, $P(A_n^c) = 0$ for all $n \in \mathbb{N}$.

Combining all the findings, consider the following derivation. For every $\varepsilon > 0$, $\exists C > 0$ and $\exists N = \max\{N_1, N_2\}$ s.t. $\forall n \geq N$,

$$\begin{aligned}
1 - \varepsilon &\leq P\left(q_{b_n,n}^{SS_1}(1-\alpha) - P\left(\theta \in \hat{C}_{b_n,n}^{SS_1}(1-\alpha)\right) > 2C\left(b_n^{-1/2} + b_n/n\right)\right) \\
&\quad + P\left(\left|q_{b_n,n}^{SS_1}(1-\alpha) - (1-\alpha)\right| \leq n^{-1/2}\right) - 1 \\
&\leq P\left(\left\{q_{b_n,n}^{SS_1}(1-\alpha) - P\left(\theta \in \hat{C}_{b_n,n}^{SS_1}(1-\alpha)\right) > 2C(b_n^{-1/2} + b_n/n)\right\}\right. \\
&\quad \left. \cap \left\{\left|q_{b_n,n}^{SS_1}(1-\alpha) - (1-\alpha)\right| \leq n^{-1/2}\right\}\right) \\
&\leq P\left((1-\alpha) - P(\theta \in \hat{C}_{b_n,n}^{SS_1}(1-\alpha)) > 2C(b_n^{-1/2} + b_n/n)\right. \\
&\quad \left. - n^{-1/2} > C\left(b_n^{-1/2} + b_n/n\right)\right) \\
&\leq P\left((1-\alpha) - P\left(\theta \in \hat{C}_{b_n,n}^{SS_1}(1-\alpha)\right) > C(b_n^{-1/2} + b_n/n)\right).
\end{aligned}$$

Since the event inside the probability is nonrandom, then it must occur $\forall n \geq N$. ■

Proof of Lemma 5.1. For any nonstochastic sequence $\{a_n\}_{n \geq 1}$ with $a_n = O(b_n^{-1/2} + \sqrt{b_n/n})$, equation (5.3) holds by part 1 of Theorem 5.1.

For the second part, consider any nonstochastic sequence $\{a_n\}_{n \geq 1}$ with $a_n = o(b_n^{-1/2} + \sqrt{b_n/n})$. We now show that equation (5.3) is impossible under additional conditions. According to the statement, we are considering inference under Assumptions A1–A3, CF1, for some $\theta \in \partial\Theta_I$ that satisfies Assumptions A4–A6, and for $\alpha \in (0, 0.5)$. As additional conditions, suppose that θ also satisfies Assumptions A7–A8. According to the discussion in Section A.6.1, these additional conditions are compatible with the rest of the assumptions (recall that Assumption A7 implies Assumption A6 and Assumption A8 represents a restriction on the sign of third moments of a standardized zero-mean version of $m(Z, \theta)$). Under all of these conditions, equation (A.28) of Theorem 5.1 implies that $\exists C > 0$ and $\exists N \in \mathbb{N}$ s.t. $\forall n \geq N$, $|P(\theta \in \hat{C}_{b_n,n}^{SS_1}(1-\alpha)) - (1-\alpha)| > C(b_n^{-1/2} + b_n/n)$. If equation (5.3) were true, it would then follow that $\exists C > 0$ s.t. $O(a_n/(b_n^{-1/2} + b_n/n)) > C$, contradicting the fact that $a_n = o(b_n^{-1/2} + b_n/n)$. ■

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

A.6.3. Results for Subsampling 2

THEOREM A.14. *Assume Assumptions A1–A3, CF2, and that θ satisfies Assumption 4. Let $\rho \equiv rank(V[m(Z,\theta)])$.*

1. *If $\theta \in \partial\Theta_I$, then $Q_{b_n,n}^{SS_2}(\theta) = H(\sqrt{b_n}\bar{Y}_{b_n,n}^{SS}) + \delta_{b_n,n}^{SS_2}$, where*
   (a) *for some $C > 0$, $\sqrt{b_n}P(|\delta_{b_n,n}^{SS_2}| > C\sqrt{(\ln\ln n)b_n/n}|\mathcal{X}_n) = o(1)$ a.s.*
   (b) *$\{\bar{Y}_{b_n,n}^{SS}|\mathcal{X}_n\} : \Omega_n \to \mathbb{R}^\rho$ is a zero (conditional) mean sample average of b_n observations sampled without replacement from a distribution with a (conditional) variance-covariance matrix $\hat{V}$ that is nonsingular a.s. and finite (conditional) third absolute moments a.s. If Assumption A5 is added, then $||\hat{V} - \mathbf{I}_\rho|| \leq O_p(n^{-1/2})$ and the distribution has finite (conditional) fourth absolute moments a.s.*
   (c) *$H : \mathbb{R}^\rho \to \mathbb{R}$ is the same function as in Theorem A.1.*
2. *If $\theta \in Int(\Theta_I)$, then $\sqrt{b_n}P(Q_{b_n,n}^{SS_2}(\theta) \neq 0|\mathcal{X}_n) = o(1)$ a.s.*

Proof. Part 1. Define $\delta_{b_n,n}^{SS_2} \equiv Q_{b_n,n}^{SS_2}(\theta) - G(\{[\upsilon_{b_n,n}^{SS}(m_{j,\theta})] - 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J)$, where $\upsilon_{b_n,n}^{SS}(m_{j,\theta}) \equiv \sqrt{b_n}(\bar{m}_{b_n,n,j}^{SS}(\theta) - \bar{m}_{n,j}(\theta))$. Then, parts 1(b) and (c) follow from Theorem A.9. In order to complete the proof, it suffices to show part 1(a).

For any $\delta \in (0, 1/2)$, let $S \equiv \{j \in \{1,\ldots,J\} : E[m_j(Z,\theta)] = 0\}$, $\bar{S} \equiv \{1,\ldots,J\}/S$, $\eta \equiv \min_{j\in\bar{S}} E[m_j(Z,\theta)] > 0$, $\lambda \equiv \max_{j=1,\ldots,J}(1+\delta)\sqrt{2V[m_j(Z,\theta)]}$, and let A_n be defined as follows:

$$A_n \equiv \left\{|\upsilon_{b_n,n}^{SS}(m_{j,\theta})| \leq b_n^{(1-\delta)/2}\right\}_{j=1}^J \cap \left\{|\upsilon_n(m_{j,\theta})| \leq \lambda\sqrt{\ln\ln n}\right\}_{j=1}^J\}. \tag{A.33}$$

For any $C > 0$, consider the following derivation:

$$\begin{aligned}\sqrt{b_n}P(|\delta_{b_n,n}^{SS_2}| > C\sqrt{(\ln\ln n)b_n/n}|\mathcal{X}_n) &\leq \sqrt{b_n}P(\{A_n\}^c|\mathcal{X}_n)\\ &+ \sqrt{b_n}P(\{|\delta_{b_n,n}^{SS_2}| > C\sqrt{(\ln\ln n)b_n/n}\} \cap A_n|\mathcal{X}_n).\end{aligned} \tag{A.34}$$

The proof is completed by showing that the two terms on the RHS of equation (A.34) are $o(1)$ a.s.

We begin with the first term on the RHS of equation (A.34). Consider the following derivation:

$$\begin{aligned}\sqrt{b_n}P(\{A_n\}^c|\mathcal{X}_n) &= \sum_{j=1}^J \sqrt{b_n}P(|\upsilon_{b_n,n}^{SS}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n)\\ &\quad + \sum_{j=1}^J \sqrt{b_n}P(|\upsilon_n(m_{j,\theta})| > \lambda\sqrt{\ln\ln n}|\mathcal{X}_n)\\ &= \sum_{j=1}^J \sqrt{b_n}P(|\upsilon_{b_n,n}^{SS}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n)\\ &\quad + \sum_{j=1}^J \sqrt{b_n}1(|\upsilon_n(m_{j,\theta})| > \lambda\sqrt{\ln\ln n}).\end{aligned}$$

Fix $j = 1,\ldots,J$ and $\varepsilon > 0$ arbitrarily. By the LIL, $\liminf\{|\upsilon_n(m_{j,\theta})| \leq \lambda\sqrt{\ln\ln n}\}_{j=1}^J$ a.s. and so $P(\lim\sqrt{b_n}1(|\upsilon_n(m_{j,\theta})| > \lambda\sqrt{\ln\ln n}) = 0) = 1$. Next, consider the following derivation:

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

$$
\begin{aligned}
&P(\liminf\{P(|v^{SS}_{b_n,n}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n)\sqrt{b_n} \leq \varepsilon\}) \\
&\quad \geq \left\{ \begin{array}{c} P(\liminf\{P(|v^{SS}_{b_n,n}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n)b_n^{(\delta-1)} \leq (V[m_j(X,\theta)]+\varepsilon)\}) \\ +P(\liminf\{b_n^{(\delta-1/2)}(V[m_j(X,\theta)]+\varepsilon) \leq \varepsilon\}) - 1 \end{array} \right\} \\
&\quad = P(\liminf\{P(|v^{SS}_{b_n,n}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n)b_n^{(\delta-1)} \leq (V[m_j(X,\theta)]+\varepsilon)\}) \\
&\quad \geq P(\liminf\{\hat{V}[m_j(X,\theta)] \leq V[m_j(X,\theta)]+\varepsilon\}) = 1,
\end{aligned}
$$

where the first inequality is elementary, the first equality follows from $b_n^{(\delta-1/2)} = o(1)$, the second inequality follows from an inequality shown later in equation (A.35), and the final equality follows from the SLLN.

Set $k_n \equiv n!/(b_n!(n-b_n)!)$, i.e., the total number of subsamples of size b_n without replacement from a sample of size n. Let us denote these subsamples by $s = 1, \ldots, k_n$, i.e., $\{m(Z_i^s,\theta)\}_{i=1}^{b_n}$ refers to the s^{th} subsample. To complete this argument, consider the following derivation:

$$
\begin{aligned}
&E(v^{SS}_{b_n,n}(m_{j,\theta})^2|\mathcal{X}_n) \\
&\quad = \left(\begin{array}{l} \frac{1}{k_n}\sum_{s=1}^{k_n}\frac{1}{b_n}\sum_{i=1}^{b_n}(m_j(X_{i,s},\theta)-\bar{m}_{n,j}(\theta))^2 \\ \quad +\frac{1}{k_n}\sum_{s=1}^{k_n}\frac{1}{b_n}\sum_{a=1}^{b_n}\sum_{b=1,b\neq a}^{b_n}(m_j(X_{a,s},\theta)-\bar{m}_{n,j}(\theta))(m_j(X_{b,s},\theta)-\bar{m}_{n,j}(\theta)) \end{array} \right) \\
&\quad = \left(\begin{array}{l} \frac{1}{n}\sum_{i=1}^{n}(m_j(X_i,\theta)-\bar{m}_{n,j}(\theta))^2 \\ +\frac{2}{n(n-1)}\sum_{a=1}^{n}\sum_{b=1,b\neq a}^{n}(m_j(X_a,\theta)-\bar{m}_{n,j}(\theta))(m_j(X_b,\theta)-\bar{m}_{n,j}(\theta)) \end{array} \right) \\
&\quad \leq \hat{V}[m_j(X,\theta)],
\end{aligned}
$$

where the first equality holds by expanding squares, the following equality holds by the fact that we are sampling without replacement, and the final inequality holds by verifying that, in the previous line, the first term equals $\hat{V}[m_j(X,\theta)]$ and the second term is nonpositive by the negative associated produced by sampling without replacement (see Joag-Dev and Proschan, 1983, Sect. 3.2(a)). Chebyshev's inequality then implies that

$$
\hat{V}[m_j(X,\theta)] \geq E(v^{SS}_{b_n,n}(m_{j,\theta})^2|\mathcal{X}_n) \geq P(|v^{SS}_{b_n,n}(m_{j,\theta})| > b_n^{(1-\delta)/2}|\mathcal{X}_n)b_n^{(1-\delta)}, \qquad \textbf{(A.35)}
$$

which completes the proof for the first term on the RHS of equation (A.34).

We now consider the second term on the RHS of equation (A.34). If we condition on $\mathcal{X}_n$ and assume that A_n occurs, it follows that

$$
\begin{aligned}
Q^{SS_2}_{b_n,n}(\theta) &= G(\{[v^{SS}_{b_n,n}(m_{j,\theta}) + \sqrt{b_n/n}v_n(m_{j,\theta}) + \sqrt{b_n}E[m_j(Z,\theta)]]_-\}_{j=1}^J) \\
&\leq G(\{[v^{SS}_{b_n,n}(m_{j,\theta}) - \lambda\sqrt{(\ln\ln n)b_n/n} + \sqrt{b_n}E[m_j(Z,\theta)]]_-\}_{j=1}^J) \\
&\leq G\left(\left\{ \begin{array}{l} [v^{SS}_{b_n,n}(m_{j,\theta}) - \lambda\sqrt{(\ln\ln n)b_n/n}]_- 1[E[m_j(Z,\theta)] = 0] + \\ [-b_n^{1/2-\delta} - \lambda\sqrt{(\ln\ln n)b_n/n} + \eta\sqrt{b_n}]_- 1[E[m_j(Z,\theta)] > 0] \end{array} \right\}_{j=1}^J \right) \\
&= G(\{[v^{SS}_{b_n,n}(m_{j,\theta}) - \lambda\sqrt{(\ln\ln n)b_n/n}]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J) \\
&\leq G(\{[v^{SS}_{b_n,n}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J + \lambda \mathbf{1}_J\sqrt{(\ln\ln n)b_n/n}), \qquad \textbf{(A.36)}
\end{aligned}
$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

where the first inequality holds because G is increasing and A_n implies that $\upsilon_n(m_{j,\theta}) \geq -\lambda\sqrt{\ln\ln n}$, the second inequality follows from the fact that A_n implies that $\upsilon^{SS}_{b_n,n}(m_{j,\theta}) > -b_n^{(1-\delta)/2}$, the next equality follows for all large enough n as $\eta, \delta > 0$, and the final inequality holds by elementary arguments. By using similar arguments, we can establish an analogous lower bound for $Q^{SS_2}_{b_n,n}(\theta)$. As a consequence, if A_n occurs,

$$Q^{SS_2}_{b_n,n}(\theta) \in \left[\begin{array}{c} G(\{[\upsilon^{SS}_{b_n,n}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J - \lambda\sqrt{(\ln\ln n)b_n/n}\mathbf{1}_J), \\ G(\{[\upsilon^{SS}_{b_n,n}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J + \lambda\sqrt{(\ln\ln n)b_n/n}\mathbf{1}_J) \end{array}\right].$$

By Assumption CF2, $\forall x \in \mathbb{R}^J$ and $\forall \varepsilon > 0$, $\exists D > 0$ s.t. $|G(x+\varepsilon) - G(x)| \leq D||\varepsilon||$. If we set $C = D\lambda > 0$, then

$$\begin{aligned} |\delta^{SS_2}_{b_n,n}| &= |Q^{SS_2}_{b_n,n}(\theta) - G(\{[\upsilon^{SS}_{b_n,n}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J)| \\ &\leq \max_{r\in\{-1,1\}} \left\{ \left| \begin{array}{c} G(\{[\upsilon^{SS}_{b_n,n}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J + r\lambda\mathbf{1}_J\sqrt{(\ln\ln n)b_n/n}) \\ -G(\{[\upsilon^{SS}_{b_n,n}(m_{j,\theta})]_- 1[E[m_j(Z,\theta)] = 0]\}_{j=1}^J) \end{array} \right| \right\} \\ &\leq D\|\lambda\sqrt{(\ln\ln n)b_n/n} \times \mathbf{1}_J\| = D\lambda\sqrt{(\ln\ln n)b_n/n}. \end{aligned}$$

This completes the argument for the second term on the RHS of equation (A.34).

Part 2. Since $Q^{SS_2}_{b_n,n}(\theta) \geq 0$, it suffices to show that $\sqrt{b_n}P(Q^{SS_2}_{b_n,n}(\theta) > 0|\mathcal{X}_n) = o(1)$. For arbitrary $\delta \in (0, 1/2)$, let $\eta \equiv \min_{j=1,\ldots,J} E[m_j(Z,\theta)] > 0$ and let A_n be defined as in equation (A.33). By elementary arguments,

$$\sqrt{b_n}P(Q^{SS_2}_{b_n,n}(\theta) > 0|\mathcal{X}_n) \leq \sqrt{b_n}P(\{A_n\}^c|\mathcal{X}_n) + \sqrt{b_n}P(\{Q^{SS_2}_{b_n,n}(\theta) > 0\} \cap A_n|\mathcal{X}_n).$$

We can now repeat arguments used in part 1 to show that both terms on the RHS are $o(1)$ a.s. On the one hand, the same argument as in part 1 implies that $\sqrt{b_n}P(\{A_n\}^c|\mathcal{X}_n) = o(1)$ a.s. On the other hand, if A_n occurs, the argument used in equation (A.36) implies that $Q^{SS_2}_{b_n,n}(\theta) \leq G(\{[-b_n^{1/2-\delta} - \lambda\sqrt{(\ln\ln n)b_n/n} + \eta\sqrt{b_n}]_-\}_{j=1}^J)$. Since $\eta, \delta > 0$, the RHS expression is equal to zero for all large enough n, implying that $\sqrt{b_n}P(\{Q^{SS_2}_{b_n,n}(\theta) > 0\} \cap A_n|\mathcal{X}_n) = o(1)$. ■

THEOREM A.15 (Consistency on Any Set Excluding Zero—SS2). *Assume Assumptions A1-A3, CF2, and that θ satisfies Assumption A4. If $\theta \in \Theta_I$ and $\mu, \varepsilon > 0$,*

$$\lim_{n\to\infty} P\left(\sup_{|h|\geq\mu} \left| P(Q^{SS_2}_{b_n,n}(\theta) \leq h|\mathcal{X}_n) - \lim_{m\to\infty} P(Q_m(\theta) \leq h) \right| \leq \varepsilon \right) = 1.$$

Proof. For $\theta \in \partial\Theta_I$, the proof follows using the same arguments used to prove Theorem A.10, except that we replace Theorem A.9 with Theorem A.14.

For $\theta \in Int(\Theta_I)$, consider the following argument. By Theorem A.1, $\lim_{m\to\infty} P(Q_m(\theta) \leq h) = 1[h \geq 0]$. On the other hand, Theorem A.10 reveals that $P(Q^{SS_2}_{b_n,n}(\theta) = 0|\mathcal{X}_n) \to 1$ a.s., and, thus, $\forall h \neq 0$, $\lim_{n\to\infty} |P(Q^{SS_2}_{b_n,n} \leq h|\mathcal{X}_n) - \lim_{m\to\infty} P(Q_m(\theta) \leq h)| = 0$ a.s. By continuity of the limiting distribution everywhere except at zero, the convergence can be shown to hold uniformly $\forall h$ s.t. $|h| \geq \mu$. ■

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

THEOREM A.16 (Consistency in Level—SS2). *Assume Assumptions A1–A3, CF1, and θ satisfies Assumption A4.*

1. *If $\theta \in \partial\Theta_I$, then, $\forall \alpha \in (0, 0.5)$, $\lim P(\theta \in \hat{C}^{SS_2}_{b_n,n}(1-\alpha)) = 1-\alpha$.*
2. *If $\theta \in Int(\Theta_I)$, then, $\forall \alpha \in [0, 1]$, $\lim P(\theta \in \hat{C}^{SS_2}_{b_n,n}(1-\alpha)) = 1$.*

Proof. Part 1 follows from the same arguments used to prove Theorem 3.1. Therefore, we only show part 2.

Part 2. Suppose that $\theta \in Int(\Theta_I)$. By Theorem A.1, $\liminf\{Q_n(\theta) = 0\}$ a.s. Since $Q^{SS_2}_{b_n,n}(\theta) \geq 0$, then $\hat{c}^{SS_2}_{b_n,n}(\theta, 1-\alpha) \geq 0$ and, thus, it follows that $\liminf\{Q_n(\theta) \leq \hat{c}^{SS_2}_{b_n,n}(\theta, 1-\alpha)\}$ a.s., which implies that $P(\theta \in \hat{C}^{SS_2}_{b_n,n}(1-\alpha)) = P(Q_n(\theta) \leq \hat{c}^{SS_2}_{b_n,n}(\theta, 1-\alpha)) = 1 + o(1)$. ■

THEOREM A.17 (Rates of Convergence—SS2). *Assume Assumptions A1–A3, CF2, that θ satisfies Assumption A4–A6, and let $\mu > 0$.*

1. *If $\theta \in \partial\Theta_I$,*

$$\sup_{|h|\geq\mu} |P(Q^{SS_2}_{b_n,n}(\theta) \leq h|\mathcal{X}_n) - P(Q_n(\theta) \leq h)| \leq O_p(b_n^{-1/2} + \sqrt{(\ln\ln n) b_n/n}).$$

2. *If $\theta \in Int(\Theta_I)$,*

$$\sup_{|h|\geq\mu} |P(Q^{SS_2}_{b_n,n}(\theta) \leq h|\mathcal{X}_n) - P(Q_n(\theta) \leq h)| \leq o_p(b_n^{-1/2}).$$

Proof. Part 1 follows from the same arguments used to prove Theorem A.12. Therefore, we only show part 2.

Part 2. Fix $\theta \in Int(\Theta_I)$ and $\mu > 0$, and consider the following derivation:

$$\begin{aligned}
&\sqrt{b_n} \sup_{|h|\geq\mu} |P(Q^{SS_2}_{b_n,n}(\theta) \leq h|\mathcal{X}_n) - P(Q_n(\theta) \leq h)| \\
&= \sqrt{b_n} \sup_{h\geq\mu} |P(Q^{SS_2}_{b_n,n}(\theta) \leq h|\mathcal{X}_n) - P(Q_n(\theta) \leq h)| \\
&= \sqrt{b_n} \sup_{h\geq\mu} |P(Q^{SS_2}_{b_n,n}(\theta) > h|\mathcal{X}_n)| + \sqrt{b_n} \sup_{h\geq\mu} |P(Q_n(\theta) > h)| \\
&\leq \sqrt{b_n} P(Q^{SS_2}_{b_n,n}(\theta) > 0|\mathcal{X}_n) + \sqrt{b_n} P(Q_n(\theta) > 0), \qquad \textbf{(A.37)}
\end{aligned}$$

where the first equality follows from $Q^{SS_2}_{b_n,n}(\theta) \geq 0$ and $Q_n(\theta) \geq 0$, the second equality follows from the triangular inequality and that $\lim_{m\to\infty} P(Q_m(\theta) \leq h) = 1[h \geq 0]$ (shown in Theorem A.1), and the final inequality follows from elementary arguments. The expression on the RHS is a sum of two terms. By Theorem A.14, the first term is $o(1)$, a.s. To complete the proof, set $\eta \equiv \min_{j=1,\ldots,J} E[m_j(Z, \theta)] > 0$ and consider the following derivation:

$$P(Q_n(\theta) > 0) \leq \sum_{i=1}^{J} P(v_n(m_{j,\theta}) < -\sqrt{n}\eta) = O(n^{-1}),$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

where the first inequality follows from $\{Q_n(\theta) > 0\} \subseteq \cup_{j=1}^{J}\{\sqrt{n}\bar{m}_{n,j}(\theta) < 0\} \subseteq \cup_{j=1}^{J}\{v_n(m_{j,\theta}) < -\sqrt{n}\eta\}$ and the rate of convergence follows from Chebyshev's inequality. ■

Proof of Theorem 5.2. This proof follows closely the arguments used to prove Theorem 3.2. ■

Proof of Lemma 5.2. For a nonstochastic sequence $\{a_n\}_{n\geq 1}$ with $a_n = O(b_n^{-1/2} + \sqrt{(\ln\ln n)b_n/n})$, equation (5.5) holds by part 1 of Theorem 5.2.

For the second part, consider any nonstochastic sequence $\{a_n\}_{n\geq 1}$ with $a_n = o(b_n^{-1/2} + \sqrt{b_n/n})$. We now show that equation (5.5) is impossible under additional conditions. According to the statement, we are considering inference under AssumptionsA1–A3, CF2, for some $\theta \in \partial\Theta_I$ that satisfies Assumptions A4–A6, and for $\alpha \in (0, 0.5)$. In order to verify the claim, let us consider a particular example that satisfies these conditions and some additional ones.

For the sake of simplicity, we consider a moment inequality model with one moment inequality (i.e., $J = 1$). Let $\Theta = [-B, B]$ for $B > 0$ and $\Theta_I = \{\theta \in \Theta : E[Z] - \theta \geq 0\}$. Furthermore, assume that $\mathcal{X}_n = \{Z_i\}_{i=1}^n$ are i.i.d. observations of a continuous distribution with respect to Lebesgue measure s.t. $E[Z] = 0$, $V[Z] = 1$, $E[Z^3] \neq 0$, and $E[Z^4] < \infty$. This implies that $\Theta_I = [-B, 0]$, $\partial\Theta_I = \{0\}$ and, so, we are interested in conducting inference at $\theta = 0$. Furthermore, suppose that $\alpha \in (0, 0.3)$ which then implies that $c_\infty(0, 1-\alpha) > 1$, where $c_\infty(0, 1-\alpha)$ denotes the $(1-\alpha)$-quantile of $[\zeta]_-$ with $\zeta \sim N(0, 1)$. Finally, we choose $G(x) = x$.

In this context, it is easy to verify that Assumptions A1–A7 and CF2 are satisfied. Furthermore, the arguments in Remark A.1 imply that Assumption A8 is satisfied if and only if $E[Z^3] > 0$. As we will show, Assumption A8 is not necessary in order to derive the result in the current setup.

In this context, $Q_n(\theta) = G([\sqrt{n}(\bar{Z}_n - \theta)]_-) = [\sqrt{n}(\bar{Z}_n - \theta)]_-$. For any arbitrary $h \geq 0$, consider the following derivation:

$$P(Q_n(0) \leq h) = P([\sqrt{n}\bar{Z}_n]_- \leq h) = P(\sqrt{n}\bar{Z}_n \geq -h) = \Phi(h) + o(b_n^{-1/2}), \tag{A.38}$$

where the last equality holds uniformly in $h \geq 0$ due to the Berry–Esseen theorem.

Now consider subsampling 2 approximation. Notice that $Q_{b_n,n}^{SS_2}(\theta) = G([\sqrt{b_n}(\bar{Z}_{b_n,n}^{SS} - \theta)]_-) = [\sqrt{b_n}(\bar{Z}_{b_n,n}^{SS} - \theta)]_-$, where $\bar{Z}_{b_n,n}^{SS} \equiv b_n^{-1}\sum_{i=1}^{b_n} Z_i^{SS}$ and $\{Z_i^{SS}\}_{i=1}^{b_n}$ is a random sample of size b_n extracted without replacement from $\mathcal{X}_n$ (i.e., subsampling). For any arbitrary $h \geq 0$ and conditional on $\mathcal{X}_n$, consider the following derivation:

$$\begin{aligned} P(Q_{b_n,n}^{SS_2}(0) \leq h|\mathcal{X}_n) &= P([\sqrt{b_n}\bar{Z}_{b_n,n}^{SS}]_- \leq h|\mathcal{X}_n) \\ &= P(\sqrt{b_n}(\bar{Z}_{b_n,n}^{SS} - \bar{Z}_n) \geq -\sqrt{b_n}\bar{Z}_n - h|\mathcal{X}_n) \\ &= \Phi(h + \sqrt{b_n}\bar{Z}_n) + \tilde{K}_1(h + \sqrt{b_n}\bar{Z}_n)b_n^{-1/2} \\ &\quad + \tilde{K}_2(h + \sqrt{b_n}\bar{Z}_n)b_n/n + o_p(b_n^{-1/2} + b_n/n), \end{aligned} \tag{A.39}$$

where the equality in the last line holds uniformly in $h \geq 0$ by Lemma A.2 and uses $\tilde{K}_1(s) \equiv E[Z^3](s^2 - 1)\phi(s)/3!$ and $\tilde{K}_2(s) \equiv s\phi(s)$ ($\tilde{K}_1$ and $\tilde{K}_2$ are related to K_1 and

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

K_2 in Lemma A.2 for $S = [-\sqrt{b_n}\bar{Z}_n - h, \infty))$. Since $\sqrt{b_n}\bar{Z}_n = O_p(\sqrt{b_n/n}) = o_p(1)$, a Taylor expansion implies that

$$\begin{aligned}
&\tilde{K}_1(h+\sqrt{b_n}\bar{Z}_n)b_n^{-1/2} + \tilde{K}_2(h+\sqrt{b_n}\bar{Z}_n)b_n/n \\
&\quad = \tilde{K}_1(h)b_n^{-1/2} + \tilde{K}_2(h)b_n/n + o_p(b_n^{-1/2} + b_n/n) \\
&\Phi(h+\sqrt{b_n}\bar{Z}_n) = \Phi(h) + \sqrt{b_n}\bar{Z}_n\phi(h) + o_p(\sqrt{b_n/n}), \qquad \textbf{(A.40)}
\end{aligned}$$

and both hold uniformly in $h \in \mathbb{R}$. By combining equations (A.38), (A.39), and (A.40), we can conclude that:

$$\begin{aligned}
&P(Q^{SS_2}_{b_n,n}(0) \le h|\mathcal{X}_n) - P(Q_n(0) \le h) \\
&\quad = \sqrt{b_n}\bar{Z}_n\phi(h) + \tilde{K}_1(h)b_n^{-1/2} + \tilde{K}_2(h)b_n/n + o_p(\sqrt{b_n/n}) + o_p(b_n^{-1/2} + b_n/n) \\
&\quad = \sqrt{b_n}\bar{Z}_n\phi(h) + \tilde{K}_1(h)b_n^{-1/2} + \tilde{K}_2(h)b_n/n + \upsilon_n(h), \qquad \textbf{(A.41)}
\end{aligned}$$

where $\sup_{h\ge 0}|\upsilon_n(h)| = o_p(\sqrt{b_n/n}) + o_p(b_n^{-1/2} + b_n/n) = o_p(b_n^{-1/2} + \sqrt{b_n/n})$. There are two possible cases.

Case 1. $E[Z^3] > 0$. In this case, $\tilde{K}_1(c_\infty(0, 1-\alpha)) > 0$, which can be shown to imply that Assumption A8 is satisfied. By arguments in Theorem A.3, $\exists \eta \in (0, 1-\alpha)$ s.t. $c_\infty(0, 1-\alpha-\eta) > 1$ and $c_\infty(0, 1-\alpha+\eta) < \infty$. By definition of these functions, $\exists C_1, C_2 > 0$ s.t. $0 < C_1 < \inf_{h\in[c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} \tilde{K}_1(h)$, $0 < \inf_{h\in[c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} \tilde{K}_2(h)$, $\sup_{h\in[c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} \tilde{K}_2(h) < \infty$, and $0 < C_2 < \inf_{h\in[c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} \phi(h)$. If we define $C \equiv \min\{C_1, C_2/2\} > 0$, it follows that $\exists N$ s.t. $\forall n \ge N$,

$$\inf_{h\in[c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} |\sqrt{b_n/n}\phi(h) + \tilde{K}_1(h)b_n^{-1/2} + \tilde{K}_2(h)b_n/n| \ge C(b_n^{-1/2} + \sqrt{b_n/n}), \qquad \textbf{(A.42)}$$

where the inequality uses that $b_n/n = o(\sqrt{b_n/n})$.

We now proceed by contradiction, i.e., suppose that there is a nonstochastic sequence $\{a_n\}_{n\ge 1}$ with $a_n = o(b_n^{-1/2} + \sqrt{b_n/n})$ s.t.

$$|P(0 \in \hat{C}^{SS_2}_{b_n,n}(1-\alpha)) - (1-\alpha)| \le a_n. \qquad \textbf{(A.43)}$$

Suppose that $\hat{c}^{SS_2}_{b_n,n}(0, 1-\alpha) \in [c_\infty(0, 1-\alpha-\eta), c_\infty(0, 1-\alpha+\eta)]$ and $|q^{SS_2}_{b_n,n}(0, 1-\alpha) - (1-\alpha)| \le n^{-1/2}$, where $q^{SS_2}_{b_n,n}(0, 1-\alpha) \equiv P(Q^{SS_2}_{b_n,n}(0) \le \hat{c}^{SS_2}_{b_n,n}(0, 1-\alpha)|\mathcal{X}_n)$. Then, consider the following derivation:

$$\begin{aligned}
a_n &\ge |P(0 \in \hat{C}^{SS_2}_{b_n,n}(1-\alpha)) - (1-\alpha)| \\
&\ge |P(Q_n(0) \le \hat{c}^{SS_2}_{b_n,n}(0, 1-\alpha)) - P(Q^{SS_2}_{b_n,n}(0) \le \hat{c}^{SS_2}_{b_n,n}(0, 1-\alpha)|\mathcal{X}_n) \\
&\quad | - |q^{SS_2}_{b_n,n}(0, 1-\alpha) - (1-\alpha)| \\
&\ge \inf_{h\in[c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} |P(Q^{SS_2}_{b_n,n}(0) \le h|\mathcal{X}_n) - P(Q_n(0) \le h)| - n^{-1/2} \\
&\ge \inf_{h\in[c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} |\sqrt{b_n}\bar{Z}_n\phi(h) + \tilde{K}_1(h)b_n^{-1/2} + \tilde{K}_2(h)b_n/n| \\
&\quad - \sup_{h\ge 0}|\upsilon_n(h)| - n^{-1/2},
\end{aligned}$$

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

where the first inequality uses equation (A.43), the second inequality uses that $P(0 \in \hat{C}^{SS_2}_{b_n,n}(1-\alpha)) = P(Q_n(0) \leq \hat{c}^{SS_2}_{b_n,n}(0,1-\alpha))$, the third inequality uses that $\hat{c}^{SS_2}_{b_n,n}(0,1-\alpha) \in [c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]$ and $|q^{SS_2}_{b_n,n}(0,1-\alpha)-(1-\alpha)| \leq n^{-1/2}$, and the fourth inequality uses equation (A.41). From here, we conclude that:

$$\begin{aligned}
&\{\hat{c}^{SS_2}_{b_n,n}(0,1-\alpha) \in [c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]\} \cap \{|q^{SS_2}_{b_n,n}(0,1-\alpha)-(1-\alpha)| \leq n^{-1/2}\} \\
&\quad \subseteq \left\{ \inf_{h \in [c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} |\sqrt{b_n}\bar{Z}_n \phi(h) + \tilde{K}_1(h) b_n^{-1/2} + \tilde{K}_2(h) b_n/n| - a_n - n^{-1/2} \right. \\
&\qquad \left. \leq \sup_{h \geq 0} |v_n(h)| \right\}. \qquad \textbf{(A.44)}
\end{aligned}$$

By repeating arguments used in Lemma A.4 and Theorem 5.1, it follows that

$$\begin{aligned}
&\lim P(\{|q^{SS_2}_{b_n,n}(0,1-\alpha)-(1-\alpha)| \leq n^{-1/2}\} \\
&\quad \cap \{\hat{c}^{SS_2}_{b_n,n}(0,1-\alpha) \in [c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]\}) = 1. \qquad \textbf{(A.45)}
\end{aligned}$$

By combining equations (A.44) and (A.45), we deduce that:

$$\begin{aligned}
&\lim P\left(\inf_{h \in [c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} |\sqrt{b_n}\bar{Z}_n \phi(h) + \tilde{K}_1(h) b_n^{-1/2} + \tilde{K}_2(h) b_n/n| \right. \\
&\qquad \left. - a_n - n^{-1/2} \leq \sup_{h \geq 0} |v_n(h)| \right) = 1. \qquad \textbf{(A.46)}
\end{aligned}$$

To conclude, consider the following argument:

$$\begin{aligned}
&P\left(\inf_{h \in [c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} |\sqrt{b_n}\bar{Z}_n \phi(h) + \tilde{K}_1(h) b_n^{-1/2} + \tilde{K}_2(h) b_n/n| \right. \\
&\qquad \left. - a_n - n^{-1/2} \leq \sup_{h \geq 0} |v_n(h)| \right) \\
&\leq \left\{ \begin{array}{l} P\left(\inf_{h \in [c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} |\sqrt{b_n/n}\phi(h) + \tilde{K}_1(h) b_n^{-1/2} \right. \\ \qquad \left. + \tilde{K}_2(h) b_n/n| - a_n - n^{-1/2} \leq \sup_{h \geq 0} |v_n(h)| \right) + P(\sqrt{n}\bar{Z}_n \leq 1) \end{array} \right\} \\
&\leq P\left(C(b_n^{-1/2} + \sqrt{b_n/n}) - a_n - n^{-1/2} \leq \sup_{h \geq 0} |v_n(h)| \right) + P(\sqrt{n}\bar{Z}_n \leq 1) \\
&\leq P\left(C/2(b_n^{-1/2} + \sqrt{b_n/n}) \leq \sup_{h \geq 0} |v_n(h)| \right) \\
&\quad + P(\sqrt{n}\bar{Z}_n \leq 1) \leq P(\zeta \leq 1) + o(1), \qquad \textbf{(A.47)}
\end{aligned}$$

where $\zeta \sim N(0,1)$. The first inequality holds by elementary arguments, the second inequality holds by equation (A.42), the third inequality holds by $a_n + n^{-1/2} = o(b_n^{-1/2} + \sqrt{b_n/n})$, and the fourth inequality holds by the CLT, $E[Z] = 0$, and $\sup_{h \geq 0} |v_n(h)| = o_p(b_n^{-1/2} + \sqrt{b_n/n})$. Since $P(\zeta \leq 1) < 1$, this is a contradiction to equation (A.46).

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press

Case 2. $E[Z^3] < 0$. In this case, $\tilde{K}_1(c_\infty(0, 1-\alpha)) < 0$, which can be shown to imply that Assumption A8 is not satisfied. As before, $\exists \eta \in (0, 1-\alpha)$ s.t. $c_\infty(0, 1-\alpha-\eta) > 1$ and $c_\infty(0, 1-\alpha+\eta) < \infty$. By definition of these functions, we have that $\exists C_1, C_2 > 0$ s.t. $0 > -C_1 > \sup_{h \in [c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} \tilde{K}_1(h)$, $0 < \inf_{h \in [c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} \tilde{K}_2(h)$, $\sup_{h \in [c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} \tilde{K}_2(h) < \infty$, and $0 > -C_2 > \sup_{h \in [c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} -\phi(h)$. Define $C \equiv \min\{C_1, C_2/2\} > 0$. From here, deduce that $\exists N$ s.t. $\forall n \geq N$,

$$\inf_{h \in [c_\infty(0,1-\alpha-\eta), c_\infty(0,1-\alpha+\eta)]} |-\sqrt{b_n/n}\phi(h) + \tilde{K}_1(h) b_n^{-1/2} - \tilde{K}_2(h) b_n/n| \geq C(b_n^{-1/2} + \sqrt{b_n/n}),$$

where the inequality uses that $b_n/n = o(\sqrt{b_n/n})$. The remainder of the proof is completely analogous to case 1. The only difference occurs in equation (A.47) where, instead of partitioning the sample space according to $\{\sqrt{n}\bar{Z}_n > 1\}$ and $\{\sqrt{n}\bar{Z}_n \leq 1\}$, we partition according to $\{\sqrt{n}\bar{Z}_n < -1\}$ and $\{\sqrt{n}\bar{Z}_n \geq -1\}$. ■

https://doi.org/10.1017/S0266466614000826 Published online by Cambridge University Press