Asymptotic Refinements of a Misspecification-Robust Bootstrap for Generalized Method of Moments Estimators

SeoJeong Lee*†

Department of Economics

University of Wisconsin-Madison

This version: July 2011

Abstract

I propose an alternative bootstrap procedure for the generalized method of moments (GMM) estimators that achieves asymptotic refinements for t tests and confidence intervals. I extend the results in the existing literature by establishing the same magnitude of asymptotic refinements *without* recentering the bootstrap moment function and *without* assuming correct specification of the moment condition model. As a result, the proposed bootstrap is robust to model misspecification, while the conventional bootstrap is not. The key procedure is to use a misspecification-robust variance estimator for GMM in constructing the t statistic and confidence intervals. Two examples of overidentified and possibly misspecified moment condition models are provided: (i) Combining data sets, and (ii) using invalid instrumental variables. Monte Carlo simulation results are provided as well.

KEYWORDS: nonparametric iid bootstrap, asymptotic refinement, Edgeworth expansion, generalized method of moments, model misspecification.

JEL Classification: C14, C15, C31, C33

*I am very grateful to Bruce Hansen and Jack Porter for their guidance and helpful comments. I also thank Ken West, Andrés Aradillas-López, and Xiaoxia Shi, as well as seminar participants at the University of Wisconsin-Madison and the 2011 North American Meeting of the Econometric Society for their discussions and suggestions.

†Address: Department of Economics, 1180 Observatory Drive, Madison, WI 53706

Email: slee279@wisc.edu

Homepage: https://sites.google.com/site/misspecified/

1 Introduction

This paper proposes an alternative bootstrap procedure for the generalized method of moments (GMM) estimators that achieves a sharp magnitude of asymptotic refinements for t tests and confidence intervals (CI's) without recentering the bootstrap moment function and without assuming correct model specification. Recentering the bootstrap moment function has been considered as critical to get refinements of the bootstrap for overidentified models, and the validity of the conventional asymptotic and bootstrap theories are established under the correct model assumption. Thus, the contribution of this paper may look too good to be true at first glance, but it becomes apparent once we realize that those two eliminations are in fact closely related, because the recentering procedure makes the bootstrap non-robust to misspecification.

Bootstrap critical values and CI's have been considered as alternatives to first-order asymptotic theory of GMM estimators of Hansen (1982), which has been known to provide poor approximations of finite sample distributions of test statistics. Hahn (1996) proves that the bootstrap distribution consistently approximates the distribution of GMM estimators. Hall and Horowitz (1996) shows that the bootstrap critical values provide higher-order improvements over the asymptotic critical values of t tests and the test of overidentifying restrictions (henceforth J test) of GMM estimators. The bootstrap procedure proposed by Hall and Horowitz (1996) is denoted by the Hall-Horowitz bootstrap throughout the paper. Andrews (2002) proposes a k-step bootstrap procedure that achieves the same higher-order improvements but which is computationally more attractive than the original Hall-Horowitz bootstrap. Brown and Newey (2002) suggests an alternative bootstrap procedure using the empirical likelihood (EL) probability. Hereinafter, the bootstrap procedure proposed by Brown and Newey (2002) is denoted by the Brown-Newey bootstrap.

In the existing bootstrap methods for GMM estimators, the key procedure is recentering so that the moment condition is satisfied in the sample. The Hall-Horowitz bootstrap analytically recenters the bootstrap moment function with respect to the sample mean of the moment function. Andrews (2002) and Horowitz (2003) also use the same recentering procedure as the Hall-Horowitz bootstrap. The Brown-Newey bootstrap recenters the bootstrap moment condition by employing the EL probability in resampling the bootstrap sample. Thus, both the Hall-Horowitz bootstrap and the

Brown-Newey bootstrap can be referred as *the recentered bootstrap.*

Horowitz (2001) explains why recentering is important when applying the bootstrap to overidentified moment condition models, where the dimension of a moment function is greater than that of a parameter. In such models, the sample mean of the moment function evaluated at the estimator is not necessarily equal to zero, though it converges in probability to zero if the model is correctly specified. In principle, the bootstrap considers the sample and the estimator as if they were the population and the true parameter, respectively. This implies that the bootstrap version of the moment condition, that the sample mean of the moment function evaluated at the estimator should equal to zero, does not hold when the model is overidentified.

A naive approach to bootstrapping for overidentified GMM is to apply the standard bootstrap procedure as is done for just-identified models, without any additional correction, such as the recentering procedure. However, it turns out that this naive bootstrap fails to achieve asymptotic refinements for t tests and CI's or first-order validity for the J test. Hall and Horowitz (1996) and Brown and Newey (2002) explain that the bootstrap and sample versions of test statistics would have different asymptotic distributions without recentering, because of the violation of the moment condition in the sample.

Although they address that the failure of the naive bootstrap is due to the misspecification in the sample, they do not further investigate the conditional asymptotic distribution of the bootstrap GMM estimator under misspecification. Instead, they eliminate the misspecification problem by recentering. In contrast, I use Hall and Inoue (2003)'s asymptotic theory for GMM under misspecification[1] to construct the t statistic which is robust to misspecification, rather than eliminating the misspecification problem.

Hall and Inoue (2003) shows that the asymptotic distributions of GMM estimators under misspecification are different from those of the standard GMM theory. In particular, their formulas for the asymptotic variance matrices of GMM estimators encompass the case of correct specification as a special case, but they are different in general. The variance estimator using their formula is denoted by the Hall-Inoue variance estimator, hereinafter. Imbens (1997) also describes the formula for the variance matrix of GMM estimators robust to misspecification by using a just-identified formulation of overidentified GMM. However, his description is general, rather than

[1] Hall and Inoue (2003) does not deal with bootstrapping, however.

specific to the misspecification problem defined in this paper.

I propose a bootstrap procedure that uses the Hall-Inoue variance estimators in constructing the sample and the bootstrap t statistics. The procedure ensures that both t statistics satisfy the asymptotic pivotal condition without recentering. The proposed bootstrap achieves asymptotic refinements, a reduction in the error of test rejection probability and CI coverage probability by a factor of n^{-1} for symmetric two-sided t tests and symmetric percentile-t CI's, over the asymptotic counterparts. The magnitude of the error is $O(n^{-2})$, which is sharp. This is the same magnitude of error shown in Andrews (2002), that uses the Hall-Horowitz bootstrap procedure for independent and identically distributed (iid) data with slightly stronger assumptions than those of Hall and Horowitz (1996).

Moreover, the proposed bootstrap procedure does not require the assumption of correct model specification because the distribution of the bootstrap t statistic mimics that of the sample t statistic which is studentized by the Hall-Inoue variance estimator. The sample t statistic is robust to misspecification, by construction. Thus, the bootstrap procedure is robust to misspecification and is referred to as the misspecification-robust (MR) bootstrap. In contrast, the conventional asymptotic theory as well as the recentered bootstrap would not work if the model is misspecified.

I note that the MR bootstrap is not for the J test. In order to get the bootstrap distribution of the J statistic, the bootstrap should be implemented under the null hypothesis that the model is correctly specified. The recentered bootstrap implicitly imposes the null hypothesis of the J test. In contrast, the MR bootstrap does not impose the null hypothesis of the J test and thus, it does not mimic the distribution of the sample J statistic under the null. Since the conventional asymptotic and bootstrap t tests and CI's are valid only if the model is correct, it is important to conduct the J test and report the result. However, the J test would be less important if possible misspecification of the model is assumed and the validity of the asymptotic and bootstrap t tests and CI's is established under such assumption, as is done in this paper.

The remainder of the paper is organized as follows. Section 2 discusses theoretical and empirical implications of misspecified model and explains the advantage of using the MR bootstrap t tests and CI's in practice. Section 3 outlines the main result. Section 4 defines the estimators and test statistics. Section 5 defines the nonparametric iid MR bootstrap for iid data. Section 6 states the assumptions and establishes asymptotic refinements of the MR bootstrap. Section 7 provides a heuristic explanation of why the recentered bootstrap does not work under misspecification. Section 8 presents examples and Monte Carlo simulation results. Section 9 concludes the paper. Appendix A contains Lemmas and proofs of the results.

2 Why We Care About Misspecification

Empirical studies in the economics literature often report a significant J statistic along with GMM estimates, standard errors, and CI's. Such examples include Parker and Julliard (2005), Jondeau, Le Bihan, and Galles (2004), and Agüero and Marks (2008), among others. Significant J statistics are also quite common in the instrumental variables literature using two-stage least squares (2SLS) estimators, where 2SLS estimator is a special case of GMM estimator.

A significant J statistic means that the test rejects the null hypothesis of correct model specification or the validity of the instruments. The problem is that, even if models are likely to be misspecified, inferences are made using the asymptotic theory for correctly specified models and the estimates are interpreted as structural parameters that have economic implications. Various authors justify this by noting that the J test over-rejects the correct null in small samples.

On the other hand, comparing and evaluating the relative fit of competing models have been an important research topic. Vuong (1989), Rivers and Vuong (2002), and Kitamura (2003) suggest a test of the null hypothesis that tests whether two possibly misspecified models provide equivalent approximation to the true model in terms of the Kullback-Leibler information criteria (KLIC). Recent studies such as Chen, Hong, and Shum (2007), Marmer and Otsu (2010), and Shi (2011) generalize and modify the test in broader settings. Hall and Pelletier (2011) shows that the limiting distribution of the Rivers-Vuong test statistic is non-standard that may not be consistently estimable unless both models are misspecified. In this framework, therefore, all competing models are misspecified and the test selects a less misspecified model. For applications of the Rivers-Vuong test, see French and Jones (2004), Gowrisankaran and Rysman (2009) and Bonnet and Dubois (2010).

Either for the empirical studies that report a significant J statistic, or for a model selected by the Rivers-Vuong test, inference about the parameters should be made based on the asymptotic theory for GMM appropriate for misspecified models, developed by Hall and Inoue (2003). Otherwise, such inferences would be misleading. Since Hall and Inoue's asymptotic theory encompasses a correctly specified model as a special case for the situations considered in this paper, it is robust to misspecification and there is little cost of using it.

When the model is misspecified, $Eg(X_i, \theta) \neq 0$ for all θ, where θ is a parameter of interest and $g(X_i, \theta)$ is a known moment function. Let $\hat{\theta}$ be the GMM estimator. According to Hall and Inuoe (2003), (i) the probability limit of $\hat{\theta}$ is the pseudo-true value that depends on the weight matrix such that

$$\theta_0(\Omega^{-1}) = \arg\min_{\theta} Eg(X_i, \theta)'\Omega^{-1}Eg(X_i, \theta), \tag{2.1}$$

and (ii) the asymptotic distribution of the GMM estimator is

$$\sqrt{n}(\hat{\theta} - \theta_0(\Omega^{-1})) \rightarrow_d N(0, \Sigma_{MR}), \tag{2.2}$$

where Σ_{MR} is the asymptotic variance matrix under misspecification that is different from Σ_C, the asymptotic variance matrix under correct specification. If the model is correctly specified, then $\theta_0(\Omega^{-1})$ and Σ_{MR} simplify to θ_0 and Σ_C, respectively.

The pseudo-true value can be interpreted as the best approximation to the true value given the weight matrix. The dependence of the pseudo-true value on the weight matrix may make the interpretation of the estimand unclear. Nevertheless, the literature on estimation under misspecification considers the pseudo-true value as a valid estimand, see Sawa (1978) and Schennach (2007) for more discussions. Other pseudo-true values that minimize the generalized empirical likelihood without using a weight matrix, have better interpretations but comparing different pseudo-true values is beyond the scope of this paper.

Although we cannot fix any potential bias in the pseudo-true value, we can report the estimated asymptotic variance of the GMM estimator as honest as possible. The second consequence (ii) above implies that the conventional standard errors and CI's are inconsistent under misspecification, while consistent standard errors and CI's are available using Hall and Inoue's asymptotic theory. Since Hall and Inoue provides the formulas for the asymptotic variance of the GMM estimator under possible misspecification, the variance estimators using their formulas are consistent for the true

Critical Value† / CI‡	Correct Model: First-order Validity	Correct Model: Asymptotic Refinements	Misspecified Model: First-order Validity	Misspecified Model: Asymptotic Refinements
Conventional Asymptotic	O	X	X	X
Naive Bootstrap	O	X	X	X
Recentered Bootstrap	O	O	X	X
Hall-Inoue Asymptotic	O	X	O	X
MR Bootstrap§	O	O	O	O

†: The critical values are for t tests.

‡: The bootstrap CI's are the percentile-t intervals.

§: MR bootstrap denotes the misspecification-robust bootstrap proposed by the author.

Table 1: Comparison of the Asymptotic and Bootstrap Critical Values

asymptotic variance regardless of misspecification.

Once we admit that the models used in the empirical literature are possibly misspecified, the asymptotic theory for GMM under misspecification should be seriously considered in practice. Furthermore, the advantage of the MR bootstrap critical values and CI's over the existing asymptotic and bootstrap ones is clearly demonstrated, as is shown in Table 1.

3 Outline of the Results

In this section, I outline the misspecification-robust (MR) bootstrap. The idea of the MR bootstrap procedure can be best understood in the same framework with Hall and Horowitz (1996) and Brown and Newey (2002), as is described below.

Suppose that the random sample is $\chi_n = \{X_i : i \leq n\}$ from a probability distribution P. Let F be the corresponding cumulative distribution function (cdf). The empirical distribution function (edf) is denoted by F_n. The GMM estimator, $\hat{\theta}$, minimizes a sample criterion function, $J_n(\theta)$. Let $\hat{\Sigma}$ be a consistent estimator of the asymptotic variance of $\sqrt{n}(\hat{\theta} - plim(\hat{\theta}))$.

I also define the bootstrap sample. Let $\chi^*_{n_b} = \{X^*_i : i \leq n_b\}$ be a sample of random vectors from the empirical distribution P^* conditional on χ_n with the edf F_n. In this section, I distinguish n and n_b, which helps understanding the concept of the conditional asymptotic distribution.[2] I set $n = n_b$ from the following section. Define $J^*_{n_b}(\theta)$ and $\hat{\Sigma}^*$ as $J_n(\theta)$ and $\hat{\Sigma}$ are defined, but with $\chi^*_{n_b}$ in place of χ_n. The bootstrap GMM estimator $\hat{\theta}^*$ minimizes $J^*_{n_b}(\theta)$.

Consider a symmetric two-sided test of the null hypothesis $H_0 : \theta = \theta_0$ with level α. The t statistic under H_0 is $T(\chi_n) = (\hat{\theta} - \theta_0)/\sqrt{\hat{\Sigma}/n}$, a functional of χ_n. One rejects the null hypothesis if $|T(\chi_n)| > z$ for a critical value z. I also consider a $100(1-\alpha)\%$ CI for θ_0, $[\hat{\theta} \pm z\sqrt{\hat{\Sigma}/n}]$. For the asymptotic test or the asymptotic CI, set $z = z_{\alpha/2}$, where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal distribution. For the bootstrap test or the symmetric percentile-t interval, set $z = z^*_{|T|,\alpha}$, where $z^*_{|T|,\alpha}$ is the $1-\alpha$ quantile of the distribution of $|T(\chi^*_{n_b})| \equiv |\hat{\theta}^* - \hat{\theta}|/\sqrt{\hat{\Sigma}^*/n_b}$.

Let $H_n(z, F) = P(T(\chi_n) \leq z|F)$ and $H^*_{n_b}(z, F_n) = P(T(\chi^*_{n_b}) \leq z|F_n)$. According to Hall (1992), under regularity conditions, $H_n(z, F)$ and $H^*_{n_b}(z, F_n)$ allow Edgeworth expansion of the form

$$H_n(z,F) = H_\infty(z,F) + n^{-1/2}q_1(z,F) + n^{-1}q_2(z,F) + o(n^{-1}), \tag{3.1}$$

$$H^*_{n_b}(z,F_n) = H^*_\infty(z,F_n) + n_b^{-1/2}q_1(z,F_n) + n_b^{-1}q_2(z,F_n) + o_p(n_b^{-1}) \tag{3.2}$$

uniformly over z, where $q_1(z, F)$ is an even function of z for each F, $q_2(z, F)$ is an odd function of z for each F, $q_2(z, F_n) \to q_2(z, F)$ almost surely as $n \to \infty$ uniformly over z, $H_\infty(z, F) = \lim_{n\to\infty} H_n(z, F)$ and $H^*_\infty(z, F_n) = \lim_{n_b\to\infty} H^*_{n_b}(z, F_n)$. If $T(\cdot)$ is asymptotically pivotal, then $H_\infty(z, F) = H^*_\infty(z, F_n) = \Phi(z)$ where Φ is the standard normal cdf, because $H_\infty(z, F)$ and $H^*_\infty(z, F_n)$ do not depend on the underlying cdf.

Using (3.1) and the fact that q_1 is even, it can be shown that under H_0,

$$P(|T(\chi_n)| > z_{\alpha/2}) = \alpha + O(n^{-1}), \quad P(\theta_0 \in CI) = 1 - \alpha + O(n^{-1}), \tag{3.3}$$

where $CI = [\hat{\theta} \pm z_{\alpha/2}\sqrt{\hat{\Sigma}/n}]$. In other words, the error in the rejection probability and coverage probability of the asymptotic two-sided t test and CI is $O(n^{-1})$.

For the bootstrap t test and CI, subtract (3.1) from (3.2), use the fact that q_1 is

[2] n_b is the resample size and should be distinguished from the number of bootstrap replication (or resampling), often denoted by B. See Bickel and Freedman (1981) for further discussion.

even, and set $n_b = n$ to show, under H_0,

$$P(|T(\chi_n)| > z^*_{|T|,\alpha}) = \alpha + o(n^{-1}), \quad P(\theta_0 \in CI^*) = 1 - \alpha + o(n^{-1}) \tag{3.4}$$

where $CI^* = [\hat{\theta} \pm z^*_{|T|,\alpha}\sqrt{\hat{\Sigma}/n}]$. The elimination of the leading terms in (3.1) and (3.2) is the source of asymptotic refinements of bootstrapping the asymptotically pivotal statistics (Beran, 1988; Hall, 1992).

First suppose that the model is correctly specified, $Eg(X_i, \theta_0) = 0$ for unique θ_0, where $E[\cdot]$ is the expectation with respect to the cdf F. The conventional t statistic $T_C(\chi_n) = (\hat{\theta} - \theta_0)/\sqrt{\hat{\Sigma}_C/n}$, where $\hat{\Sigma}_C$ is the standard GMM variance estimator, is asymptotically pivotal. However, a naive bootstrap t statistic without recentering,[3] $T_C(\chi^*_{n_b}) = (\hat{\theta}^* - \hat{\theta})/\sqrt{\hat{\Sigma}^*_C/n_b}$, is not asymptotically pivotal because the moment condition under F_n is misspecified, $E_{F_n} g(X_i^*, \hat{\theta}) = n^{-1}\sum_{i=1}^n g(X_i, \hat{\theta}) \neq 0$ almost surely when the model is overidentified. If the moment condition is misspecified, the conventional variance estimator of the GMM estimator is no longer consistent, according to Hall and Inoue (2003). Note that the bootstrap moment condition is evaluated at $\hat{\theta}$, where $\hat{\theta}$ is considered as the true value given F_n.

The recentered bootstrap makes the bootstrap moment condition hold so that the recentered bootstrap t statistic is asymptotically pivotal. For instance, the Hall-Horowitz bootstrap uses a recentered moment function $g^*(X_i^*, \theta) = g(X_i^*, \theta) - n^{-1}\sum_{i=1}^n g(X_i, \hat{\theta})$ so that $E_{F_n} g^*(X_i^*, \hat{\theta}) = 0$ almost surely. The Brown-Newey bootstrap uses the EL distribution function $\hat{F}_{EL}(z) = n^{-1}\sum_{i=1}^n \hat{p}_i \mathbf{1}(X_i \leq z)$ in resampling, where $\hat{p}_i$ is the EL probability and $\mathbf{1}(\cdot)$ is an indicator function, instead of using F_n, so that $E_{\hat{F}_{EL}} g(X_i^*, \hat{\theta}) = 0$ almost surely.

The MR bootstrap uses the original *non-recentered* moment function in implementing the bootstrap and resamples according to the edf F_n. This is similar to the naive bootstrap. The distinction is that the MR bootstrap uses the Hall-Inoue variance estimator in constructing the sample and the bootstrap versions of the t statistic instead of using the conventional GMM variance estimator. The sample t statistic is $T_{MR}(\chi_n) = (\hat{\theta} - \theta_0)/\sqrt{\hat{\Sigma}_{MR}/n}$, where $\hat{\Sigma}_{MR}$ is a consistent estimator of Σ_{MR} and Σ_{MR} is the asymptotic variance of the GMM estimator regardless of misspecification. Then, $T_{MR}(\chi_n)$ is asymptotically pivotal.

[3] A naive bootstrap for GMM is constructing $\hat{\theta}^*$ and $\hat{\Sigma}^*$ in the same way we construct $\hat{\theta}$ and $\hat{\Sigma}$, using the bootstrap sample $\chi^*_{n_b}$ in place of χ_n.

The MR bootstrap t statistic is $T_{MR}(\chi^*_{n_b}) = (\hat{\theta}^* - \hat{\theta})/\sqrt{\hat{\Sigma}^*_{MR}/n_b}$, where $\hat{\Sigma}^*_{MR}$ uses the same formula as $\hat{\Sigma}_{MR}$ with $\chi^*_{n_b}$ in place of χ_n. Then, $\hat{\Sigma}^*_{MR}$ is consistent for the conditional asymptotic variance of the bootstrap GMM estimator, $\Sigma_{MR|F_n}$, almost surely, even if the bootstrap moment condition is not satisfied. As a result, $T_{MR}(\chi^*_{n_b})$ is asymptotically pivotal. Therefore, the MR bootstrap achieves asymptotic refinements without recentering under correct specification.

Now suppose that the model is misspecified in the population, $Eg(X_i, \theta) \neq 0$ for all θ. The advantage of the MR bootstrap is that the assumption of correct model is not required for both the sample and the bootstrap t statistics. Since $T_{MR}(\chi_n)$ and $T_{MR}(\chi^*_{n_b})$ are constructed by using the Hall-Inoue variance estimator, they are asymptotically pivotal regardless of model misspecification. Thus, the ability of achieving asymptotic refinements of the MR bootstrap is not affected.

The conclusion changes dramatically for the recentered bootstrap, however. First of all, the conventional t statistic $T_C(\chi_n)$ is no longer asymptotically pivotal and this invalidates the use of the asymptotic t test and CI's. Moreover, since the recentered bootstrap mimics the distribution of $T_C(\chi_n)$ under correct specification, the recentered bootstrap t test and CI's are not even first-order valid. The conditional and unconditional distributions of the recentered bootstrap t statistic is described in Section 7.

Let $z^*_{|T_{MR}|,\alpha}$ be the $1-\alpha$ quantile of the distribution of $|T_{MR}(\chi^*_{n_b})|$ and let $CI^*_{MR} = [\hat{\theta} \pm z^*_{|T_{MR}|,\alpha}\sqrt{\hat{\Sigma}_{MR}/n}]$. Using the MR bootstrap without assuming the correct model, I show that, under H_0,

$$P(|T_{MR}(\chi_n)| > z^*_{|T_{MR}|,\alpha}) = \alpha + O(n^{-2}), \quad P(\theta_0 \in CI^*_{MR}) = 1 - \alpha + O(n^{-2}). \quad (3.5)$$

This rate is sharp. The further reduction in the error from $o(n^{-1})$ of (3.4) to $O(n^{-2})$ of (3.5) is based on the argument given in Hall (1988). Andrews (2002) shows the same sharp bound using the Hall-Horowitz bootstrap and assuming the correct model.

4 Estimators and Test Statistics

Given an $L_g \times 1$ vector of moment conditions $g(X_i, \theta)$, where θ is $L_\theta \times 1$, and $L_g \geq L_\theta$, define a correctly specified and a misspecified model as follows: The model is *correctly specified* if there exists a unique value θ_0 in $\Theta \subset \mathbb{R}^{L_\theta}$ such that $Eg(X_i, \theta_0) = 0$, and the model is *misspecified* if there exists no θ in $\Theta \subset \mathbb{R}^{L_\theta}$ such that $Eg(X_i, \theta) = 0$. That is, $Eg(X_i, \theta) = g(\theta)$ where $g : \Theta \to \mathbb{R}^{L_\theta}$ such that $\|g(\theta)\| > 0$ for all $\theta \in \Theta$, if the model is misspecified. Assume that the model is possibly misspecified.

The (pseudo-)true parameter θ_0 minimizes the population criterion function,

$$J(\theta, \Omega^{-1}) = Eg(X_i, \theta)'\Omega^{-1}Eg(X_i, \theta), \tag{4.1}$$

where Ω^{-1} is a weight matrix. Since the model is possibly misspecified, the moment condition and the population criterion may not equal to zero for any $\theta \in \Theta$. In this case, the minimizer of the population criterion depends on Ω^{-1} and is denoted by $\theta_0(\Omega^{-1})$. We call $\theta_0(\Omega^{-1})$ the pseudo-true value. The dependence vanishes when the model is correctly specified.

Consider two forms of GMM estimator. The first one is a one-step GMM estimator using the identity matrix I_{L_g} as a weight matrix, which is the common usage. The second one is a two-step GMM estimator using a weight matrix constructed from the one-step GMM estimator. Under correct specifications, the common choice of the weight matrix is an asymptotically optimal one. However, the optimality is not established under misspecification because the asymptotic variance matrix of the two-step GMM estimator cannot be simplified to the efficient variance matrix under correct specifications.

The one-step GMM estimator, $\hat{\theta}_{(1)}$, solves

$$\min_{\theta\in\Theta} J_n(\theta, I_{L_g}) = \left(n^{-1}\sum_{i=1}^{n} g(X_i, \theta)\right)' \left(n^{-1}\sum_{i=1}^{n} g(X_i, \theta)\right). \tag{4.2}$$

The two-step GMM estimator, $\hat{\theta}_{(2)}$ solves

$$\min_{\theta\in\Theta} J_n(\theta, W_n(\hat{\theta}_{(1)})) \equiv \left(n^{-1}\sum_{i=1}^{n} g(X_i, \theta)\right)' W_n(\hat{\theta}_{(1)}) \left(n^{-1}\sum_{i=1}^{n} g(X_i, \theta)\right), \tag{4.3}$$

where[4]

$$W_n(\theta) = \left(n^{-1}\sum_{i=1}^{n}(g(X_i,\theta) - g_n(\theta))(g(X_i,\theta) - g_n(\theta))'\right)^{-1}, \tag{4.4}$$

and $g_n(\theta) = n^{-1}\sum_{i=1}^{n} g(X_i,\theta)$. Suppress the dependence of W_n on θ and write $W_n \equiv W_n(\hat{\theta}_{(1)})$. Under regularity conditions, the GMM estimators are consistent: $\hat{\theta}_{(1)}$ converges to a pseudo-true value $\theta_0(I) \equiv \theta_{0(1)}$, and $\hat{\theta}_{(2)}$ converges to a pseudo-true value $\theta_0(W) \equiv \theta_{0(2)}$. Under misspecification, $\theta_{0(1)} \neq \theta_{0(2)}$ in general. The probability limit of the weight matrix W_n is $W = \left\{E[(g(X_i,\theta_{0(1)}) - g_{0(1)})(g(X_i,\theta_{0(1)}) - g_{0(1)})']\right\}^{-1}$, where $g_{0(j)} = Eg(X_i,\theta_{0(j)})$ for $j = 1, 2$.

To further simplify notation, let

$$G_{0(j)} = E\left[\frac{\partial}{\partial\theta'}g(X_i,\theta_{0(j)})\right], \quad G^{(2)}_{0(j)} = E\left[\frac{\partial}{\partial\theta'}vec\left\{\frac{\partial}{\partial\theta'}g(X_i,\theta_{0(j)})\right\}\right], \tag{4.5}$$

and an $L_\theta \times L_\theta$ matrix $H_{0(j)} = G'_{0(j)}\Omega^{-1}G_{0(j)} + (g'_{0(j)}\Omega^{-1} \otimes I_{L_\theta})G^{(2)}_{0(j)}$, where $\Omega^{-1} = I_{L_g}$ for $j = 1$ and $\Omega^{-1} = W$ for $j = 2$. Let

$$G_n(\theta) = n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\theta'}g(X_i,\theta), \quad G^{(2)}_n(\theta) = n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\theta'}vec\left\{\frac{\partial}{\partial\theta'}g(X_i,\theta)\right\}, \tag{4.6}$$

$G_{n(j)} = G_n(\hat{\theta}_{(j)})$, and $H_{n(j)} = G'_{n(j)}\Omega^{-1}G_{n(j)} + (g'_{n(j)}\Omega^{-1} \otimes I_{L_\theta})G^{(2)}_{n(j)}$, where $\Omega^{-1} = I_{L_g}$ for $j = 1$ and $\Omega^{-1} = W_n$ for $j = 2$. Let Ω_1 and Ω_2 denote positive-definite matrices such that

$$\sqrt{n}\begin{pmatrix} (g_n(\theta_{0(1)}) - g_{0(1)}) \\ (G_n(\theta_{0(1)}) - G_{0(1)})'g_{0(1)} \end{pmatrix} \to_d N\left(\mathbf{0}, \underset{(L_g+L_\theta)\times(L_g+L_\theta)}{\Omega_1}\right), \tag{4.7}$$

and

$$\sqrt{n}\begin{pmatrix} (g_n(\theta_{0(2)}) - g_{0(2)}) \\ (G_n(\theta_{0(2)}) - G_{0(2)})'Wg_{0(2)} \\ (W_n - W)g_{0(2)} \end{pmatrix} \to_d N\left(\mathbf{0}, \underset{(2L_g+L_\theta)\times(2L_g+L_\theta)}{\Omega_2}\right). \tag{4.8}$$

[4] One may consider an $L_g \times L_g$ nonrandom positive-definite symmetric matrix for the one-step GMM estimator or the *uncentered* weight matrix, $W_n(\theta) = (n^{-1}\sum_{i=1}^{n} g(X_i,\theta)g(X_i,\theta)')^{-1}$, for the two-step GMM estimator. This does not affect the main result of the paper, though the resulting pseudo-true values are different. In practice, however, the uncentered weight matrix may not behave well under misspecification, because the elements of the uncentered weight matrix include bias terms of the moment function. See Hall (2000) for more discussion on the issue.

To obtain the misspecification-robust asymptotic variance matrix for the GMM estimator, I use Theorems 1 and 2 of Hall and Inoue (2003). Then,

$$\sqrt{n}(\hat{\theta}_{(j)} - \theta_{0(j)}) \to_d N(0, \Sigma_{MR(j)}), \tag{4.9}$$

where $\Sigma_{MR(j)} = H_{0(j)}^{-1} V_j H_{0(j)}^{-1}$, for $j = 1, 2$,

$$\begin{aligned} V_1 &= \begin{bmatrix} G'_{0(1)} & I_{L_\theta} \end{bmatrix} \Omega_1 \begin{bmatrix} G'_{0(1)} & I_{L_\theta} \end{bmatrix}', \\ V_2 &= \begin{bmatrix} G'_{0(2)}W & I_{L_\theta} & G'_{0(2)} \end{bmatrix} \Omega_2 \begin{bmatrix} G'_{0(2)}W & I_{L_\theta} & G'_{0(2)} \end{bmatrix}'. \end{aligned} \tag{4.10}$$

Under correct specifications, $\Sigma_{MR(1)}$ and $\Sigma_{MR(2)}$ reduce to the standard asymptotic variance matrices of the GMM estimators, $\Sigma_{C(1)}$ and $\Sigma_{C(2)}$ respectively, where

$$\Sigma_{C(1)} = (G_0'G_0)^{-1}G_0'\Omega_C G_0(G_0'G_0)^{-1}, \quad \Sigma_{C(2)} = (G_0'\Omega_C^{-1}G_0)^{-1}, \tag{4.11}$$

and $\Omega_C = E[g(X_i, \theta_0)g(X_i, \theta_0)']$.

A consistent estimator of $\Sigma_{MR(j)}$ is $\hat{\Sigma}_{MR(j)} = H_{n(j)}^{-1} V_{n(j)} H_{n(j)}^{-1\prime}$ for $j = 1, 2$, where

$$\begin{aligned} V_{n(1)} &= \begin{bmatrix} G'_{n(1)} & I_{L_\theta} \end{bmatrix} \Omega_{n(1)} \begin{bmatrix} G'_{n(1)} & I_{L_\theta} \end{bmatrix}', \\ V_{n(2)} &= \begin{bmatrix} G'_{n(2)}W_n & I_{L_\theta} & G'_{n(2)} \end{bmatrix} \Omega_{n(2)} \begin{bmatrix} G'_{n(2)}W_n & I_{L_\theta} & G'_{n(2)} \end{bmatrix}', \end{aligned} \tag{4.12}$$

and $\Omega_{n(j)}$ is a consistent estimator of Ω_j, with the population moments replaced by the sample moments. In particular,

$$\begin{aligned} \Omega_{n(1)} &= n^{-1}\sum_{i=1}^{n} \begin{pmatrix} g(X_i, \hat{\theta}_{(1)}) - g_{n(1)} \\ (G(X_i, \hat{\theta}_{(1)}) - G_{n(1)})'g_{n(1)} \end{pmatrix} \begin{pmatrix} g(X_i, \hat{\theta}_{(1)}) - g_{n(1)} \\ (G(X_i, \hat{\theta}_{(1)}) - G_{n(1)})'g_{n(1)} \end{pmatrix}', \\ \Omega_{n(2)} &= n^{-1}\sum_{i=1}^{n} \begin{pmatrix} g(X_i, \hat{\theta}_{(2)}) - g_{n(2)} \\ (G(X_i, \hat{\theta}_{(2)}) - G_{n(2)})'W_n g_{n(2)} \\ W_i g_{n(2)} \end{pmatrix} \begin{pmatrix} g(X_i, \hat{\theta}_{(2)}) - g_{n(2)} \\ (G(X_i, \hat{\theta}_{(2)}) - G_{n(2)})'W_n g_{n(2)} \\ W_i g_{n(2)} \end{pmatrix}', \end{aligned} \tag{4.13}$$

where[5]

$$W_i = -W_n \cdot \Big((g(X_i, \hat{\theta}_{(1)}) - g_n(\hat{\theta}_{(1)}))(g(X_i, \hat{\theta}_{(1)}) - g_n(\hat{\theta}_{(1)}))' - W_n^{-1} \Big) \cdot W_n. \tag{4.14}$$

[5]Note that $W_n - W = -W(W_n^{-1} - W^{-1})W_n$.

In practice, the estimation of the MR variance matrices does not involve much complication. What we need to calculate additionally is the second derivative of the moment function.

Let θ_k, $\theta_{0(j),k}$, and $\hat{\theta}_{(j),k}$ denote the kth elements of θ, θ_0, and $\hat{\theta}_{(j)}$ respectively. Let $(\hat{\Sigma}_{MR(j)})_{kk}$ denote the (k,k)th element of $\hat{\Sigma}_{MR(j)}$. The t statistic for testing the null hypothesis $H_0 : \theta_k = \theta_{0(j),k}$ is

$$T_{MR(j)} = \frac{\hat{\theta}_{(j),k} - \theta_{0(j),k}}{\sqrt{(\hat{\Sigma}_{MR(j)})_{kk}/n}}, \tag{4.15}$$

where $j = 1$ for the one-step GMM estimator and $j = 2$ for the two-step GMM estimator.[6] $T_{MR(j)}$ is misspecification-robust because it has an asymptotic $N(0,1)$ distribution under H_0, without assuming the correct model. $T_{MR(j)}$ is different from the conventional t statistic constructed from the conventional variance estimator, $\Sigma_{C(j)}$, even under correct specification, because $\hat{\Sigma}_{C(j)} \neq \hat{\Sigma}_{MR(j)}$, $j = 1, 2$, in general.

The MR bootstrap described in the next section achieves asymptotic refinements over the misspecification-robust asymptotic t test and CI, rather than the conventional non-robust ones. Define the misspecification-robust asymptotic t test and CI as follows. The symmetric two-sided t test with asymptotic significance level α rejects H_0 if $|T_{MR(j)}| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal distribution. The corresponding CI for $\theta_{0(j),k}$ with asymptotic confidence level $100(1-\alpha)\%$ is $CI_{MR(j)} = [\hat{\theta}_{(j),k} \pm z_{\alpha/2}\sqrt{(\hat{\Sigma}_{MR(j)})_{kk}/n}]$, $j = 1, 2$. The error in the rejection probability of the t test with $z_{\alpha/2}$ and coverage probability of $CI_{MR(j)}$ is $O(n^{-1})$: Under H_0, $P\left(|T_{MR(j)}| > z_{\alpha/2}\right) = \alpha + O(n^{-1})$ and $P\left(\theta_{0(j),k} \in CI_{MR(j)}\right) = 1 - \alpha + O(n^{-1})$, for $j = 1, 2$.

5 The Misspecification-Robust Bootstrap Procedure

The nonparametric iid bootstrap is implemented by sampling $X_1^*, \cdots, X_n^*$ randomly with replacement from the sample $X_1, \cdots, X_n$.

[6] $T_{MR(j)} \equiv T_{MR(j)}(\chi_n)$. I suppress the dependence of $T_{MR(j)}$ on χ_n for notational brevity.

The bootstrap one-step GMM estimator, $\hat{\theta}^*_{(1)}$ solves:

$$\min_{\theta\in\Theta} J_n^*(\theta, I_{L_g}) = \left(n^{-1}\sum_{i=1}^{n} g(X_i^*,\theta)\right)' \left(n^{-1}\sum_{i=1}^{n} g(X_i^*,\theta)\right), \tag{5.1}$$

and the bootstrap two-step GMM estimator $\hat{\theta}^*_{(2)}$ solves

$$\min_{\theta\in\Theta} J_n^*(\theta, W_n^*(\hat{\theta}^*_{(1)})) = \left(n^{-1}\sum_{i=1}^{n} g(X_i^*,\theta)\right)' W_n^*(\hat{\theta}^*_{(1)}) \left(n^{-1}\sum_{i=1}^{n} g(X_i^*,\theta)\right), \tag{5.2}$$

where

$$W_n^*(\theta) = \left(n^{-1}\sum_{i=1}^{n}(g(X_i^*,\theta) - g_n^*(\theta))(g(X_i^*,\theta) - g_n^*(\theta))'\right)^{-1}, \tag{5.3}$$

and $g_n^*(\theta) = n^{-1}\sum_{i=1}^{n} g(X_i^*,\theta)$. Suppress the dependence of W_n^* on θ and write $W_n^* \equiv W_n^*(\hat{\theta}^*_{(1)})$. To further simplify notation, let

$$G_n^*(\theta) = n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\theta'}g(X_i^*,\theta), \quad G_n^{(2)*}(\theta) = n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\theta'}vec\left\{\frac{\partial}{\partial\theta'}g(X_i^*,\theta)\right\}, \tag{5.4}$$

$G^*_{n(j)} = G_n^*(\hat{\theta}^*_{(j)})$, and $H^*_{n(j)} = G^{*\prime}_{n(j)}\Omega^{-1}G^*_{n(j)} + (g^{*\prime}_{n(j)}\Omega^{-1}\otimes I_{L_\theta})G^{(2)*}_{n(j)}$, where $\Omega^{-1} = I_{L_g}$ for $j=1$ and $\Omega^{-1} = W_n^*$ for $j=2$.

The bootstrap version of the Hall-Inoue variance estimator $\hat{\Sigma}_{MR(j)}$ is $\hat{\Sigma}^*_{MR(j)} = H^{*-1}_{n(j)}V^*_{n(j)}H^{*-1\prime}_{n(j)}$ for $j=1,2$, where

$$\begin{aligned} V^*_{n(1)} &= \begin{bmatrix} G^{*\prime}_{n(1)} & I_{L_g} \end{bmatrix} \Omega^*_{n(1)} \begin{bmatrix} G^{*\prime}_{n(1)} & I_{L_g} \end{bmatrix}', \\ V^*_{n(2)} &= \begin{bmatrix} G^{*\prime}_{n(2)}W_n^* & I_{L_g} & G^{*\prime}_{n(2)} \end{bmatrix} \Omega^*_{n(2)} \begin{bmatrix} G^{*\prime}_{n(2)}W_n^* & I_{L_g} & G^{*\prime}_{n(2)} \end{bmatrix}', \end{aligned} \tag{5.5}$$

and $\Omega^*_{n(j)}$ is constructed by replacing the sample moments in $\Omega_{n(j)}$ with the bootstrap sample moments. In particular,

$$\Omega^*_{n(1)} = n^{-1}\sum_{i=1}^{n}\begin{pmatrix} g(X_i^*,\hat{\theta}^*_{(1)}) - g^*_{n(1)} \\ (G(X_i^*,\hat{\theta}^*_{(1)}) - G^*_{n(1)})' g^*_{n(1)} \end{pmatrix}\begin{pmatrix} g(X_i^*,\hat{\theta}^*_{(1)}) - g^*_{n(1)} \\ (G(X_i^*,\hat{\theta}^*_{(1)}) - G^*_{n(1)})' g^*_{n(1)} \end{pmatrix}', \quad (5.6)$$

$$\Omega^*_{n(2)} = n^{-1}\sum_{i=1}^{n}\begin{pmatrix} g(X_i^*,\hat{\theta}^*_{(2)}) - g^*_{n(2)} \\ (G(X_i^*,\hat{\theta}^*_{(2)}) - G^*_{n(2)})' W_n^* g^*_{n(2)} \\ W_i^* g^*_{n(2)} \end{pmatrix}\begin{pmatrix} g(X_i^*,\hat{\theta}^*_{(2)}) - g^*_{n(2)} \\ (G(X_i^*,\hat{\theta}^*_{(2)}) - G^*_{n(2)})' W_n^* g^*_{n(2)} \\ W_i^* g^*_{n(2)} \end{pmatrix}',$$

where

$$W_i^* = -W_n^* \cdot \left((g(X_i^*,\hat{\theta}^*_{(1)}) - g_n^*(\hat{\theta}^*_{(1)}))(g(X_i^*,\hat{\theta}^*_{(1)}) - g_n^*(\hat{\theta}^*_{(1)}))' - W_n^{*-1} \right) \cdot W_n^*. \quad (5.7)$$

The MR bootstrap t statistic is

$$T^*_{MR(j)} = \frac{\hat{\theta}^*_{(j),k} - \hat{\theta}_{(j),k}}{\sqrt{(\hat{\Sigma}^*_{MR(j)})_{kk}/n}}, \quad (5.8)$$

for $j = 1, 2$.[7] Let $z^*_{|T_{MR(j)}|,\alpha}$ denote the $1-\alpha$ quantile of $|T^*_{MR(j)}|$, $j = 1, 2$. Following Andrews (2002), we define $z^*_{|T_{MR(j)}|,\alpha}$ to be a value that minimizes $|P^*(|T^*_{MR(j)}| \leq z) - (1-\alpha)|$ over $z \in \mathbf{R}$, since the distribution of $|T^*_{MR(j)}|$ is discrete. The symmetric two-sided bootstrap t test of $H_0 : \theta_k = \theta_{0(j),k}$ versus $H_1 : \theta_k \neq \theta_{0(j),k}$ rejects if $|T_{MR(j)}| > z^*_{|T_{MR(j)}|,\alpha}$, $j = 1, 2$, and this test is of asymptotic significance level α. The $100(1-\alpha)\%$ symmetric percentile-t interval for $\theta_{0(j),k}$ is, for $j = 1, 2$,

$$CI^*_{MR(j)} = \left[\hat{\theta}_{(j),k} \pm z^*_{|T_{MR(j)}|,\alpha} \sqrt{(\hat{\Sigma}_{MR(j)})_{kk}/n} \right]. \quad (5.9)$$

The MR bootstrap t statistic differs from the recentered bootstrap t statistic. First, the MR bootstrap GMM estimator, unlike the Hall-Horowitz bootstrap, is calculated from the original moment function with the bootstrap sample. Second, the Hall-Inoue variance matrix estimator, $\hat{\Sigma}^*_{MR(j)}$, is used to construct the bootstrap t statistic. In the recentered bootstrap, the conventional variance matrix estimator of Hansen (1982) is used.

[7] $T^*_{MR(j)} \equiv T_{MR(j)}(\chi_n^*)$. I suppress the dependence of $T^*_{MR(j)}$ on χ_n^* for notational brevity.

6 Main Result

6.1 Assumptions

The assumptions are analogous to those of Hall and Horowitz (1996) and Andrews (2002). The main difference is that I do not assume correct model specification. If the model is misspecified, then the probability limits of the one-step and the two-step GMM estimators are different. Thus, we need to distinguish $\theta_{0(1)}$ from $\theta_{0(2)}$, the probability limit of $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(2)}$, respectively. The assumptions are modified to hold for both pseudo-true values. If the model happens to be correctly specified, then the pseudo-true values become identical.

Let $f(X_i, \theta)$ denote the vector containing the unique components of $g(X_i, \theta)$ and $g(X_i, \theta)g(X_i, \theta)'$, and their derivatives through order $d_1 \geq 6$ with respect to θ. Let $(\partial^m/\partial\theta^m)g(X_i, \theta)$ and $(\partial^m/\partial\theta^m)f(X_i, \theta)$ denote the vectors of partial derivatives with respect to θ of order m of $g(X_i, \theta)$ and $f(X_i, \theta)$, respectively.

Assumption 1. *$X_i, i = 1, 2, ...$ are iid.*

Assumption 2. *(a) Θ is compact and $\theta_{0(1)}$ and $\theta_{0(2)}$ are interior points of Θ.*
(b) $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(2)}$ minimize $J_n(\theta, I_{L_g})$ and $J_n(\theta, W_n)$ over $\theta \in \Theta$, respectively; $\theta_{0(1)}$ and $\theta_{0(2)}$ are the pseudo-true values that uniquely minimize $J(\theta, I_{L_g})$ and $J(\theta, W)$ over $\theta \in \Theta$, respectively; for some function $C_g(x)$, $\|g(x, \theta_1) - g(x, \theta_2)\| < C_g(x)\|\theta_1 - \theta_2\|$ for all x in the support of X_1 and all $\theta_1, \theta_2 \in \Theta$; and $EC_g^{q_1}(X_1) < \infty$ and $E\|g(X_1, \theta)\|^{q_1} < \infty$ for all $\theta \in \Theta$ for all $0 < q_1 < \infty$.

Assumption 3. *The followings hold for $j = 1, 2$.*
(a) Ω_j is positive definite.
(b) $H_{0(j)}$ is nonsingular and $G_{0(j)}$ is full rank L_θ.
(c) $g(x, \theta)$ is $d = d_1 + d_2$ times differentiable with respect to θ on $N_{0(j)}$, where $N_{0(j)}$ is some neighborhood of $\theta_{0(j)}$, for all x in the support of X_1, where $d_1 \geq 6$ and $d_2 \geq 5$.
(d) There is a function $C_{\partial f}(X_1)$ such that $\|(\partial^m/\partial\theta^m)f(X_1, \theta) - (\partial^m/\partial\theta^m)f(X_1, \theta_{0(j)})\| \leq C_{\partial f}(X_1)\|\theta - \theta_{0(j)}\|$ for all $\theta \in N_{0(j)}$ for all $m = 0, ..., d_2$.
(e) $EC_{\partial f}^{q_2}(X_1) < \infty$ and $E\|(\partial^m/\partial\theta^m)f(X_1, \theta_{0(j)})\|^{q_2} \leq C_f < \infty$ for all $m = 0, ..., d_2$ for some constant C_f (that may depend on q_2) and all $0 < q_2 < \infty$. (f) $f(X_1, \theta_{0(j)})$ is once differentiable with respect to X_1 with uniformly continuous first derivative.

Assumption 4. *For $t \in \mathbf{R}^{dim(f)}$ and $j = 1, 2$, $\limsup_{\|t\| \to \infty} \left|E\left(\exp(it'f(X_1, \theta_{0(j)}))\right)\right| < 1$, where $i = \sqrt{-1}$.*

Assumption 1 says that we restrict our attention to iid sample. Hall and Horowitz (1996) and Andrews (2002) deal with dependent data. I focus on iid sample and nonparametric iid bootstrap to emphasize the role of the Hall-Inoue variance estimator in implementing the MR bootstrap and to avoid the complications arising when constructing blocks to deal with dependent data. For example, the Hall-Horowitz bootstrap needs an additional correction factor as well as the recentering procedure for the bootstrap t statistic with dependent data. The correction factor is required to properly mimic the dependence between the bootstrap blocks in implementing the MR bootstrap. I do not investigate this issue further in this paper.

Assumptions 2-3 are similar to Assumptions 2-3 of Andrews (2002), except that I eliminate the correct model assumption. In particular, I relax Assumption 2 of Hall and Horowitz (1996) and Assumption 2(b)(i) of Andrews (2002). The moment conditions in Assumptions 2-3 are not primitive, but they lead to simpler results as in Andrews (2002). Assumption 4 is the standard Cramér condition for iid sample, that is needed to get Edgeworth expansions.

6.2 Asymptotic Refinements of the Misspecification-Robust Bootstrap

Theorem 1 shows that the MR bootstrap symmetric two-sided t test has rejection probability that is correct up to $O(n^{-2})$, and the same magnitude of convergence holds for the MR bootstrap symmetric percentile-t interval. This result extends the results of Theorem 3 of Hall and Horowitz (1996) and Theorem 2(c) of Andrews (2002), because their results hold only under correctly specified models. In other words, the following Theorem establishes that the MR bootstrap achieves the same magnitude of asymptotic refinements with the existing bootstrap procedures, without assuming the correct model and without the recentering procedure.

Theorem 1. *Suppose Assumptions 1-4 hold. Under $H_0 : \theta_k = \theta_{0(j),k}$, for $j = 1, 2$,*

$$P(|T_{MR(j)}| > z^*_{|T_{MR(j)}|,\alpha}) = \alpha + O(n^{-2}) \quad \textit{or} \quad P(\theta_{0(j),k} \in CI^*_{MR(j)}) = 1 - \alpha + O(n^{-2}),$$

where $z^*_{|T_{MR(j)}|,\alpha}$ is the $1-\alpha$ quantile of the distribution of $|T^*_{MR(j)}|$.

Since $P\left(|T_{MR(j)}| > z_{\alpha/2}\right) = \alpha + O(n^{-1})$, the bootstrap critical value has a reduction in the error of rejection probability by a factor of n^{-1} for symmetric two-sided t

tests. The symmetric percentile-t interval is formulated by the symmetric two-sided t test, and the CI also has a reduction in the error of coverage probability by a factor of n^{-1}.

We note that asymptotic refinements for the J test are not established in Theorem 1. The MR bootstrap is implemented with a misspecified moment condition in the sample, $E^* g(X_i^*, \hat{\theta}) \neq 0$, where E^* is the expectation over the bootstrap sample. Thus, the distribution of the MR bootstrap J statistic does not consistently approximate that of the sample J statistic under the null hypothesis, which is $Eg(X_i, \theta_0) = 0$. Though it is typical to report the J test result in practice, the test itself has little relevance in this context since the Theorem holds without the assumption of $Eg(X_i, \theta_0) = 0$.

The proof of the Theorem proceeds by showing that the misspecification-robust t statistic studentized by the Hall-Inoue variance estimator can be approximated by a smooth function of sample moments. Once we establish that the approximation is close enough, then we can use the result of Edgeworth expansions for a smooth function in Hall (1992). The proof extensively follows those of Hall and Horowitz (1996) and Andrews (2002). The differences are that I allow for distinct probability limits of the one-step and the two-step GMM estimators, and that no special bootstrap version of the test statistic is needed for the MR bootstrap. Indeed, the recentering creates more complication than it seems even under correct specification, because $\hat{\theta}_{(1)} \neq \hat{\theta}_{(2)}$ in general, which in turn implies that there are two (pseudo-)true values in the bootstrap world. This issue is not explicitly explained in Hall and Horowitz (1996) and Andrews (2002). Therefore, the idea of the proof given in this paper is more straightforward than theirs.

7 The Recentered Bootstrap under Misspecification

In this section, I discuss about the validity of the recentered bootstrap under misspecification. Consider the conventional t statistic $T_{C(j)}(\chi_n) = (\hat{\theta}_{(j)} - \theta_{0(j)})/\sqrt{\hat{\Sigma}_{C(j)}/n}$ for $j = 1, 2$, where $\hat{\Sigma}_{C(j)}$ is the conventional GMM variance estimator of Hansen (1982). Since $\hat{\Sigma}_{C(j)}$ is inconsistent for the true asymptotic variance, $T_{C(j)}(\chi_n)$ is not asymptotically pivotal under misspecification. Therefore, the resulting asymptotic t test and CI would have incorrect rejection probability and coverage probability. Since the asymptotic pivotal condition of the sample and the bootstrap versions of the test statistic is critical to get asymptotic refinements, it is obvious that any bootstrap method would not provide refinements as long as we use the conventional t statistic.

Since the recentered bootstrap depends on the assumption of correct model in achieving asymptotic refinements, it is inappropriate to use the recentered bootstrap if the model is possibly misspecified. Nevertheless, I provide a heuristic description of the conditional and unconditional asymptotic distributions of the Hall-Horowitz bootstrap t statistics under misspecification.

Let $\hat{\theta}^*_{R(j)}$ be the Hall-Horowitz bootstrap GMM estimator with the recentered moment function. By standard consistency arguments, it can be shown that $\hat{\theta}^*_{R(j)} \to_p \hat{\theta}_{(j)}$ conditional on the sample. Since the model is correctly specified in the sample, we apply standard asymptotic normality arguments as in Newey and McFadden (1994) to get the conditional asymptotic variance matrix of the Hall-Horowitz bootstrap GMM estimator, $\Sigma_{R(j)|F_n}$. By Glivenko-Cantelli theorem, F_n converges to F uniformly in z, and thus, $\Sigma_{R(j)|F_n} \to_p \Sigma_{R(j)}$ almost surely, where $\Sigma_{R(j)}$ is the (unconditional) asymptotic variance matrix of the distribution of $\sqrt{n}(\hat{\theta}^*_{R(j)} - \hat{\theta}_{(j)})$. The formulas are given by

$$\begin{aligned}
\Sigma_{R(1)} &= (G'_{0(1)}G_{0(1)})^{-1}G'_{0(1)}\Omega_{R(1)}G_{0(1)}(G'_{0(1)}G_{0(1)})^{-1}, \qquad (7.1)\\
\Sigma_{R(2)} &= (G'_{0(2)}W_R G_{0(2)})^{-1}G'_{0(2)}W_R\Omega_{R(2)}W_R G_{0(2)}(G'_{0(2)}W_R G_{0(2)})^{-1},\\
\Omega_{R(1)} &= E(g(X_i,\theta_{0(1)}) - g_{0(1)})(g(X_i,\theta_{0(1)}) - g_{0(1)})',\\
\Omega_{R(2)} &= E(g(X_i,\theta_{0(2)}) - g_{0(2)})(g(X_i,\theta_{0(2)}) - g_{0(2)})',\\
W_R &= \left[E(g(X_i,\theta_{0(1)}) - g_{0(2)})(g(X_i,\theta_{0(1)}) - g_{0(2)})'\right]^{-1}.
\end{aligned}$$

The above formulas describe the asymptotic variance of the Hall-Horowitz bootstrap GMM estimators under misspecification. One of the fundamental reasons for the failure of the Hall-Horowitz bootstrap is that the probability limits of the preliminary and the two-step GMM estimators are different. In particular, $\Sigma_{R(2)}$ cannot be further simplified to the variance matrix of the efficient two-step GMM estimator, because W_R and $\Omega_{R(2)}$ do not cancel each other out. In contrast, $g_{0(j)} = 0$ for $j = 1, 2$, and $\theta_{0(1)} = \theta_{0(2)}$ when the model is correctly specified. Then, $\Sigma_{R(j)}$ simplifies to $\Sigma_{C(j)}$, the conventional variance matrix.

In order to construct the Hall-Horowitz bootstrap t statistic, we need the bootstrap variance estimator, $\hat{\Sigma}^*_{CR(j)}$. It is constructed by using the recentered moment function

$g(X_i^*, \theta) - g_n(\hat{\theta}_{(j)})$ and following the standard GMM formula. In particular,

$$\begin{aligned}
\Sigma^*_{CR(1)} &= (G^{*\prime}_{n(1)} G^*_{n(1)})^{-1} G^{*\prime}_{n(1)} \Omega^*_{R,n(1)} G^*_{n(1)} (G^{*\prime}_{n(1)} G^*_{n(1)})^{-1}, \\
\Sigma^*_{CR(2)} &= (G^{*\prime}_{n(2)} \Omega^{*-1}_{R,n(2)} G^*_{n(2)})^{-1}, \\
\Omega^*_{R,n(1)} &= n^{-1} \sum_{i=1}^{n} (g(X_i^*, \hat{\theta}^*_{(1)}) - g_n(\hat{\theta}_{(1)}))(g(X_i^*, \hat{\theta}^*_{(1)}) - g_n(\hat{\theta}_{(1)}))', \\
\Omega^*_{R,n(2)} &= n^{-1} \sum_{i=1}^{n} (g(X_i^*, \hat{\theta}^*_{(2)}) - g_n(\hat{\theta}_{(2)}))(g(X_i^*, \hat{\theta}^*_{(2)}) - g_n(\hat{\theta}_{(2)}))'.
\end{aligned} \tag{7.2}$$

By standard consistency arguments, we can show $G^*_{n(j)} \to_p G_{0(j)}$ and $\Omega^*_{R,n(j)} \to_p \Omega_{R(j)}$ almost surely for $j = 1, 2$. Let $\Sigma_{CR(j)}$ be the (unconditional) probability limit of $\Sigma^*_{CR(j)}$. Then,

$$\begin{aligned}
\Sigma_{CR(1)} &= (G'_{0(1)} G_{0(1)})^{-1} G'_{0(1)} \Omega_{R(1)} G_{0(1)} (G'_{0(1)} G_{0(1)})^{-1} = \Sigma_{R(1)}, \\
\Sigma_{CR(2)} &= (G'_{0(2)} \Omega^{-1}_{R(2)} G_{0(2)})^{-1} \neq \Sigma_{R(2)}.
\end{aligned} \tag{7.3}$$

Thus, studentizing the Hall-Horowitz bootstrap t statistic with $\Sigma^*_{CR(2)}$ hoping that $\Sigma^*_{CR(2)}$ is consistent for the asymptotic variance of the Hall-Horowitz bootstrap GMM estimator would not work under misspecifications.

Finally, Results 1 and 2 describe the asymptotic distribution of the Hall-Horowitz bootstrap t statistics.

Result 1 $T^*_{R,n(1)} \equiv \frac{\hat{\theta}^*_{R(1)} - \hat{\theta}_{(1)}}{\sqrt{\hat{\Sigma}^*_{CR(1)}/n}} \xrightarrow[n\to\infty]{d} N(0, 1)$, conditional on the sample almost surely.

Result 2 $T^*_{R,n(2)} \equiv \frac{\hat{\theta}^*_{R(2)} - \hat{\theta}_{(2)}}{\sqrt{\hat{\Sigma}^*_{CR(2)}/n}} \xrightarrow[n\to\infty]{d} N(0, \frac{\Sigma_{R(2)}}{\Sigma_{CR(2)}})$, conditional on the sample almost surely.

Now, consider the Brown-Newey bootstrap. The Brown-Newey bootstrap uses the *original* moment function. The difference between the naive and the Brown-Newey bootstrap is that we use $\hat{F}_{EL}$ based on the EL probabilities in place of the edf F_n. According to Chen, Hong, and Shum (2007), $\hat{F}_{EL}$ is consistent for the pseudo-true cdf F_δ, which is different from the true cdf F, under misspecification. This implies that the Brown-Newey bootstrap resampling procedure does not mimic the true data generating process asymptotically. In addition, Schennach (2007) shows that the asymptotic behavior of the EL probability is problematic if the moment function

$g(X_i, \theta)$ is not bounded in absolute terms. Brown and Newey (2002) does not have this bound in its assumptions. Thus, a further investigation is needed to use the EL probability in implementing the bootstrap.

8 Monte Carlo Experiments

In this section, I compare the actual coverage probabilities of the asymptotic and bootstrap CI's under correct specification and misspecification for different number of samples. Since the actual rejection probability of the t test is the coverage probability subtracted from one, I only report the coverage probabilities.

The conventional asymptotic CI with coverage probability $100(1-\alpha)\%$ is

$$CI_C = \left[\hat{\theta} \pm z_{\alpha/2}\sqrt{\hat{\Sigma}_C/n}\right], \tag{8.1}$$

where $z_{\alpha/2}$ is the $1-\alpha/2$th quantile of a standard normal distribution. The misspecification-robust asymptotic CI using the Hall-Inoue variance estimator with coverage probability $100(1-\alpha)\%$ is

$$CI_{MR} = \left[\hat{\theta} \pm z_{\alpha/2}\sqrt{\hat{\Sigma}_{MR}/n}\right]. \tag{8.2}$$

The only difference between this CI and the conventional CI is the choice of the variance estimator. Under correct model specification, both asymptotic CI's have coverage probability $100(1-\alpha)\%$ asymptotically and the error in the coverage probability is $O(n^{-1})$. Under misspecification, CI_{MR} is still first-order valid, but CI_C is not.

The Hall-Horowitz and the Brown-Newey bootstrap CI's with coverage probability $100(1-\alpha)\%$ are given by

$$CI^*_{HH} = \left[\hat{\theta} \pm z^*_{|T_{HH}|,\alpha}\sqrt{\hat{\Sigma}_C/n}\right], \tag{8.3}$$

$$CI^*_{BN} = \left[\hat{\theta} \pm z^*_{|T_{BN}|,\alpha}\sqrt{\hat{\Sigma}_C/n}\right], \tag{8.4}$$

where $z^*_{|T_{HH}|,\alpha}$ is the $1-\alpha$th quantile of the Hall-Horowitz bootstrap distribution of the t statistic and $z^*_{|T_{BN}|,\alpha}$ is the $1-\alpha$th quantile of the Brown-Newey bootstrap distribution of the t statistic. Both the recentered bootstrap CI's achieve asymptotic refinements over CI_C under correct specification. However, they are first-order invalid under misspecification.

The MR bootstrap CI with coverage probability $100(1-\alpha)\%$ is:

$$CI^*_{MR} = \left[\hat{\theta} \pm z^*_{|T_{MR}|,\alpha} \sqrt{\hat{\Sigma}_{MR}/n} \right], \tag{8.5}$$

where $z^*_{|T_{MR}|,\alpha}$ is the $1-\alpha$th quantile of the MR bootstrap distribution of the t statistic. This CI achieves asymptotic refinements over CI_{MR} regardless of misspecification by Theorem 1.

8.1 Example 1: Combining Data Sets

Suppose that we observe $X_i = (Y_i, Z_i)' \in \mathbb{R}^2$, $i = 1, ...n$, and we have an econometric model based on Z_i with moment function $g_1(Z_i, \theta)$, where θ is a parameter of interest. Also, suppose that we know the mean (or other population information) of Y_i. If Y_i and Z_i are correlated, we can exploit the known information on EY_i to get more accurate estimates of θ. This situation is common in survey sampling: A sample survey consists of a random sample from some population and aggregate statistics from the same population. Imbens and Lancaster (1994) and Hellerstein and Imbens (1999) show how to efficiently combine data sets and make an inference. For more examples, see Imbens (2002) and Section 3.10 of Owen (2001).

Let $g_1(Z_i, \theta) = Z_i - \theta$, so that the parameter of interest is the mean of Z_i. Without the knowledge on EY_i, the natural estimator is the method of moments (MOM) estimator, which is the sample mean of Z_i: $\hat{\theta}_{MOM} = \bar{Z} \equiv n^{-1} \sum_{i=1}^n Z_i$. If an additional information, $EY_i = 0$, is available, then we form the moment function as

$$g(X_i, \theta) = \begin{pmatrix} Y_i \\ Z_i - \theta \end{pmatrix}. \tag{8.6}$$

Since the number of moment restrictions ($L_g = 2$) is greater than that of the parameter ($L_\theta = 1$), the model is overidentified and we can use GMM estimators to estimate θ. If the assumed mean of Y is not true, i.e., $EY_i \neq 0$, then the model is misspecified because there is no θ that satisfies $Eg(X_i, \theta) = 0$.

The one-step GMM estimator solving (4.2) is given by $\hat{\theta}_{(1)} = \bar{Z}$. The two-step

GMM estimator solving (4.3) and the pseudo-true value are given by

$$\hat{\theta}_{(2)} = \bar{Z} - \frac{\widehat{Cov}(Y_i, Z_i)}{\widehat{Var}(Y_i)}\bar{Y} \rightarrow_p \theta_{0(2)} = EZ_i - \frac{Cov(Y_i, Z_i)}{Var(Y_i)}EY_i, \tag{8.7}$$

where $\widehat{Var}(Y_i) = n^{-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$ and $\widehat{Cov}(Y_i, Z_i) = n^{-1}\sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z})$. Note that the pseudo-true value reduces to $\theta_{0(2)} = EZ_i$ when $EY_i = 0$, i.e., the model is correctly specified.[8]

Without considering a possible misspecification in the model, the conventional asymptotic variance of $\hat{\theta}_{(2)}$ is $\Sigma_{C(2)} = (G_0'\Omega_C^{-1}G_0)^{-1}$. If we admit a possibility that the model is misspecified, the misspecification-robust asymptotic variance of $\hat{\theta}_{(2)}$ is $\Sigma_{MR(2)}$, where the formula for $\Sigma_{MR(2)}$ is given in the previous section.

Let the true data generating process (DGP) be

$$\begin{pmatrix} Y_i \\ Z_i \end{pmatrix} \sim N\left(\begin{pmatrix} \delta \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \tag{8.8}$$

where $0 < \rho < 1$ is a correlation between Y_i and Z_i, and $(Y_i, Z_i)'$ is iid. Thus, the assumed mean of Y_i, zero, may not equal to the true value, δ. As δ gets larger, the degree of misspecification becomes larger. The pseudo-true value is $\theta_{0(2)} = -\rho\delta$.

The asymptotic variance matrices $\Sigma_{C(2)}$ and $\Sigma_{MR(2)}$ are[9]

$$\Sigma_{C(2)} = 1 - \rho^2, \qquad \Sigma_{MR(2)} = (1 - \rho^2)(1 + \delta^2). \tag{8.9}$$

If the model is correctly specified, then using the additional information reduces the variance of the estimator by ρ^2, because the asymptotic variance of the MOM estimator $\bar{Z}$ is $Var(Z_i) = 1$. However, this reduction does not occur when the additional information is misspecified, and furthermore, the conventional variance estimator is inconsistent for the true asymptotic variance of the estimator. In contrast, the Hall-Inoue variance estimator is consistent for the true asymptotic variance regardless of misspecification. As the degree of misspecification becomes larger, the ratio of $\Sigma_{MR(2)}$

[8] This holds for this particular example. Schennach (2007) provides an example that the pseudo-true value is invariant to misspecification, and thus, is the same with the true value.

[9] A detailed calculation of $\Sigma_{C(2)}$ and $\Sigma_{MR(2)}$ as well as the GMM estimators is available at the author's personal webpage, https://sites.google.com/site/misspecified/

to $\Sigma_{C(2)}$ increases:

$$\frac{\Sigma_{MR(2)}}{\Sigma_{C(2)}} = 1 + \delta^2 \to \infty \quad \text{as } \delta \to \infty. \tag{8.10}$$

This implies that the t statistic constructed with the conventional variance estimator $\hat{\Sigma}_C$ does not converge in distribution to standard normal: the asymptotic variance of the conventional t statistic departs from 1 to infinity, as $\delta \to \infty$. Therefore, t tests or confidence intervals based on the conventional t statistic would yield incorrect rejection probability or coverage probability under misspecification.

Comment Imbens and Lancaster (1994) provides an empirical example that estimates a probit model for employment combining a micro dataset on Dutch labor market with aggregate information from national statistics. Interestingly, the J test rejects the null hypothesis of correct model, suggesting that their model might be misspecified.

Table 2 shows coverage probabilities of 90% and 95% CI's based on the two-step GMM estimator, $\hat{\theta}_{(2)}$, when $\rho = 0.5$. For a correctly specified model ($\delta = 0$), the coverage probability of the CI is the number of events that the CI contains the true value, $\theta_0 = 0$, divided by the number of Monte Carlo repetition, r. The simulation results show that the bootstrap CI's, CI^*_{MR}, CI^*_{HH}, and CI^*_{BN}, achieve asymptotic refinements over the asymptotic CI's. When the model is correctly specified, the actual and the nominal levels of the (asymptotic) J test are about the same at 1%.

For misspecified models ($\delta = 0.6$ or 1), the coverage probability of the CI is the number of events that the CI contains the pseudo-true value, $\theta_{0(2)}$, divided by r. CI^*_{MR} clearly demonstrates asymptotic refinements over CI_{MR} regardless of misspecification. In contrast, the conventional asymptotic and bootstrap CI's are first-order invalid. When $n = 25$, the asymptotic J test rejects the null about 53.2% of the Monte Carlo repetition for moderately misspecified model ($\delta = 0.6$) and about 97.2% of the Monte Carlo repetition for largely misspecified model ($\delta = 1$). Note that the degree of misspecification can be arbitrarily large, and it makes the coverage probabilities of CI_C, CI^*_{HH}, and CI^*_{BN} arbitrarily close to zero.

For different values of δ, Figure 1 shows the coverage probabilities of the CI's when $n = 25$. The figure supports the arguments made throughout the paper: Asymptotic refinements of the MR bootstrap and the first-order invalidity of the conventional asymptotic and bootstrap CI's.

Degree of Misspecification	Nominal Value	$n = 25$ 0.90	$n = 25$ 0.95	$n = 100$ 0.90	$n = 100$ 0.95
$\delta = 0$ (correct specification)	CI_{MR}	0.871	0.926	0.895	0.944
	CI^*_{MR}	0.910	0.956	0.901	0.950
	CI_C	0.866	0.925	0.893	0.944
	CI^*_{HH}	0.907	0.952	0.900	0.949
	CI^*_{BN}	0.908	0.953	0.897	0.949
	J test, 1% level (Rejection Prob.)	1.0%		1.0%	
$\delta = 0.6$ (moderate misspecification)	CI_{MR}	0.850	0.907	0.881	0.938
	CI^*_{MR}	0.892	0.942	0.895	0.945
	CI_C	0.793	0.862	0.824	0.892
	CI^*_{HH}	0.842	0.909	0.835	0.904
	CI^*_{BN}	0.847	0.913	0.834	0.903
	J test, 1% level (Rejection Prob.)	53.2%		99.9%	
$\delta = 1$ (large misspecification)	CI_{MR}	0.851	0.911	0.891	0.941
	CI^*_{MR}	0.901	0.952	0.902	0.951
	CI_C	0.716	0.792	0.745	0.820
	CI^*_{HH}	0.773	0.857	0.755	0.836
	CI^*_{BN}	0.777	0.855	0.754	0.831
	J test, 1% level (Rejection Prob.)	97.2%		100%	

Table 2: Coverage Probabilities of 90% and 95% Confidence Intervals for $\theta_{0(2)}$ based on the Two-step GMM Estimator, $\hat{\theta}_{(2)}$, when $\rho = 0.5$ in Example 1, where the number of Monte Carlo repetition (r) = 5,000, the number of bootstrap replication (B) = 1,000.

8.2 Example 2: Invalid Instrumental Variables

Suppose that there is an endogeneity in the linear model $y_i = x_i\beta_0 + e_i$, where $y_i, x_i \in \mathbb{R}$ and $Ex_ie_i \neq 0$. The OLS estimator $\hat{\beta}_{OLS}$ is inconsistent for β_0 because $\hat{\beta}_{OLS} \rightarrow_p \beta_{OLS} = \beta_0 + (Ex_i^2)^{-1}Ex_ie_i$, where the second term on the right-hand side is not equal to zero. Consider two instruments, z_{1i} and z_{2i}. By using one of the instrument, z_{ki}, $k = 1$ or 2, the IV estimator and its probability limit are

$$\hat{\beta}_{IV_k} = (\sum_{i=1}^n z_{ki}x_i)^{-1}\sum_{i=1}^n z_{ki}y_i \rightarrow_p \beta_{IV_k} = \beta_0 + (Ez_{ki}x_i)^{-1}Ez_{ki}e_i, \tag{8.11}$$

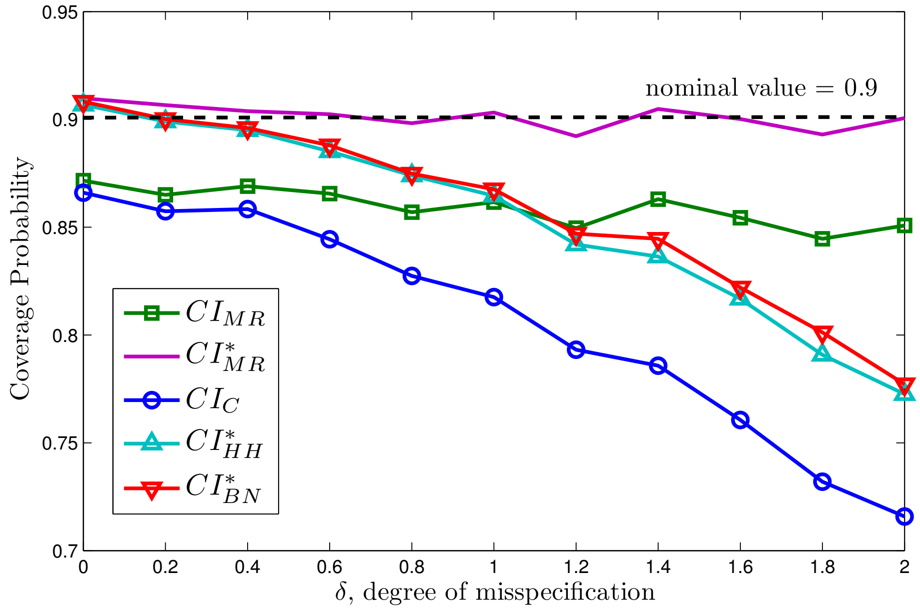


Figure 1: Coverage Probabilities of 90% Confidence Intervals for $\theta_{0(2)}$ based on the Two-step GMM Estimator, $\hat{\theta}_{(2)}$, when $\rho = 0.5$ and $n = 25$ in Example 1 (r=5,000, B=1,000)

and $\beta_{IV_k} = \beta_0$ when $Ez_{ki}e_i = 0$. If the instrument is invalid, i.e., $Ez_{ki}e_i \neq 0$, then β_{IV_k} is biased.

Now consider using both instruments in estimating β by GMM. The moment function is

$$g(X_i, \beta) = \begin{pmatrix} z_{1i}(y_i - x_i\beta) \\ z_{2i}(y_i - x_i\beta) \end{pmatrix}, \tag{8.12}$$

where $X_i = (y_i, x_i, z_{1i}, z_{2i})'$. This moment function is correctly specified when $Eg(X_i, \beta_0) = 0$ holds, which is implied by the validity of the instruments $Ez_{1i}e_i = Ez_{2i}e_i = 0$. In practice, a commonly used weight matrix is $W_n = (n^{-1}\sum_{i=1}^{n} \mathbf{z}_i\mathbf{z}_i')^{-1}$, where $\mathbf{z}_i = (z_{1i}, z_{2i})'$. The one-step GMM estimator $\hat{\beta}_{(1)}$ solves (4.3) by using W_n as a weight matrix instead of using the identity matrix.[10] Then $\hat{\beta}_{(1)}$ is a weighted average of the two instrumental variable estimators, $\hat{\beta}_{IV1}$ and $\hat{\beta}_{IV2}$. Let $\hat{\Sigma}_{MR}$ be the Hall-Inoue variance estimator and let $\hat{\Sigma}_C$ be the conventional variance estimator for $\hat{\beta}_{(1)}$.

The asymptotic variance $\lim_{n\to\infty} \hat{\Sigma}_{MR}$ can be calculated by using the formula

[10] A detailed calculation of $\hat{\beta}_{(1)}$ and its probability limit is available at the author's personal webpage, https://sites.google.com/site/misspecified/

for $\Sigma_{MR(2)}$, the asymptotic variance for the two-step GMM estimator described in Section 4, because $\sqrt{n}vech(W_n - W)$ converges to a normal distribution. Maasoumi and Phillips (1982) and Newey and McFadden (1994) address that the conventional variance estimator is inconsistent for the true asymptotic variance,[11] and that the calculation of the asymptotic variance is very complicated under misspecification.

Let the DGP be

$$y_i = x_i\beta_0 + e_i; \quad x_i = z_{1i}\gamma_1 + z_{2i}\gamma_2 + e_i + \varepsilon_i, \quad z_{2i} = z_{2i}^0 + 0.5\delta e_i + u_i; \tag{8.13}$$
$$(z_{1i}, z_{2i}^0)' \sim N(\mathbf{0}, I_2), \quad e_i \sim N(0,2), \quad \varepsilon_i \sim N(0,1), \quad u_i \sim N(0,1),$$

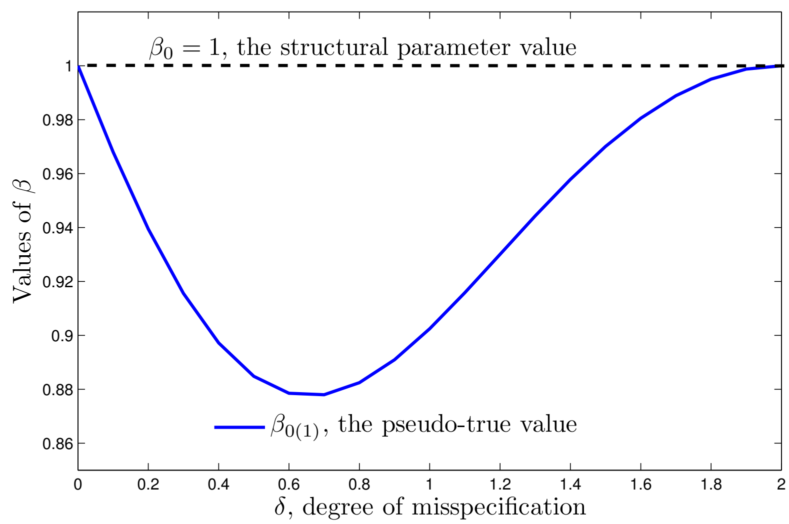
where I_2 is a 2×2 identity matrix and $(z_{1i}, z_{2i}^0)'$, e_i, ε_i, and u_i are iid. This DGP satisfies $Ex_ie_i \neq 0$, $Ez_{1i}e_i = 0$, and $Ez_{2i}e_i = \delta$, where δ measures a degree of misspecification. Therefore, the instrument z_{1i} is valid, while z_{2i} may not. The probability limit of $\hat{\beta}_{(1)}$ is

$$\beta_{0(1)} = \beta_0 + \frac{(2+0.5\delta^2)\gamma_2 + \delta}{\gamma_1^2(2+0.5\delta^2) + ((2+0.5\delta^2)\gamma_2 + \delta)^2} \cdot \delta = \beta_0 + O(\delta^{-1}). \tag{8.14}$$

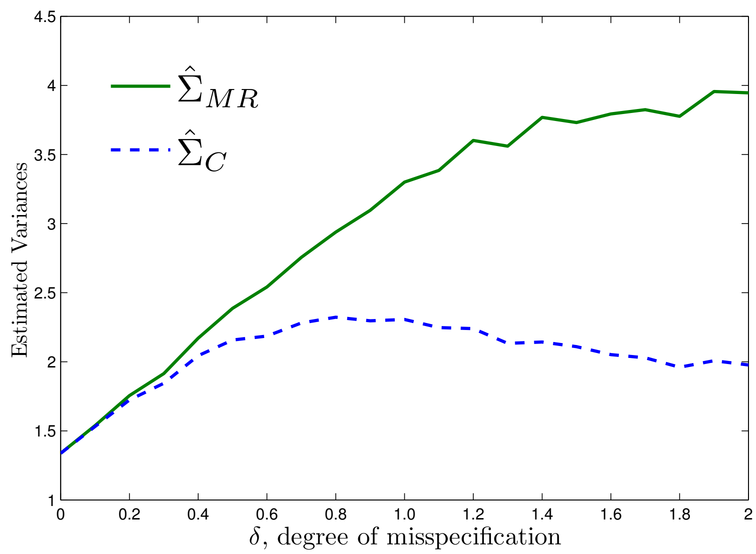
When the model is correctly specified ($\delta = 0$), then $\beta_{0(1)} = \beta_0$. Otherwise, $\beta_{0(1)} \neq \beta_0$. Note that $\beta_{0(1)} \to \beta_0$ as $\delta \to \infty$ according to the above formula. This is because the weight on the misspecified moment restriction, $Ez_{2i}e_i = 0$, converges to zero as the degree of misspecification grows. Thus, larger misspecification does not necessarily imply larger potential bias in the pseudo-true value. For example, Figure 2(a) compares the pseudo-true value with the structural parameter β_0, when $\beta_0 = 1$, $\gamma_1 = 1$, and $\gamma_2 = -0.5$. In fact, if $\gamma_2 = -\delta(2+0.5\delta^2)^{-1}$ in (8.14), then $\beta_{0(1)} = \beta_0$ holds. However, Σ_{MR} and Σ_C are different in general even if $\beta_{0(1)} = \beta_0$. Figure 2(b) shows that the values of the Hall-Inoue variance estimator and the conventional variance estimator are different under misspecification for $n = 100,000$. $\hat{\Sigma}_{MR}$ is almost twice as large as $\hat{\Sigma}_C$ at $\delta = 2$.

Table 3 shows coverage probabilities of 90% and 95% CI's based on the one-step GMM estimator, $\hat{\beta}_{(1)}$, when $\beta_0 = 1$, $\gamma_1 = 1$, and $\gamma_2 = -0.5$. Although asymptotic refinements of CI_{MR}^* do not depend on a particular choice of parameter values, the actual amount of refinements can differ according to the DGP, the sample size,

[11] The asymptotic variance formula of Hall and Inoue (2003) encompasses that of Maasoumi and Phillips (1982) as a special case.



(a) Comparison of The Pseudo-True Value and the Structural Parameter Value



(b) Comparison of The Estimated Variances, $\hat{\Sigma}_{MR}$ and $\hat{\Sigma}_{C}$ when $n = 100,000$

Figure 2: The Pseudo-True Value and The Hall-Inoue Variance Estimates under Different Degrees of Misspecification; $\beta_0 = 1$, $\gamma_1 = 1$, $\gamma_2 = -0.5$ in Example 2

Degree of Misspecification	Nominal Value	$n = 25$ 0.90	$n = 25$ 0.95	$n = 100$ 0.90	$n = 100$ 0.95
$\delta = 0$ (correct specification)	CI_{MR}	0.829	0.875	0.888	0.934
	CI^*_{MR}	0.868	0.917	0.900	0.944
	CI_C	0.816	0.862	0.886	0.932
	CI^*_{HH}	0.862	0.912	0.901	0.946
	CI^*_{BN}	0.867	0.918	0.901	0.946
	J test, 1% level (Rejection Prob.)	7.1%		6.4%	
$\delta = 1$ (moderate misspecification)	CI_{MR}	0.847	0.890	0.884	0.935
	CI^*_{MR}	0.881	0.924	0.897	0.948
	CI_C	0.784	0.836	0.818	0.884
	CI^*_{HH}	0.825	0.876	0.839	0.907
	CI^*_{BN}	0.856	0.905	0.847	0.914
	J test, 1% level (Rejection Prob.)	59.7%		98.9%	
$\delta = 2$ (large misspecification)	CI_{MR}	0.848	0.906	0.884	0.938
	CI^*_{MR}	0.892	0.943	0.894	0.948
	CI_C	0.732	0.812	0.747	0.832
	CI^*_{HH}	0.800	0.869	0.765	0.854
	CI^*_{BN}	0.859	0.919	0.779	0.872
	J test, 1% level (Rejection Prob.)	94.6%		100%	

Table 3: Coverage Probabilities of 90% and 95% Confidence Intervals for $\beta_{0(1)}$ based on the One-step GMM Estimator, $\hat{\beta}_{(1)}$ in Example 2, where the number of Monte Carlo repetition (r) = 5,000, the number of bootstrap replication (B) = 1,000.

and the choice of parameter values. The simulation results show that the bootstrap CI's, CI^*_{MR}, CI^*_{HH}, and CI^*_{BN}, achieve asymptotic refinements over the asymptotic CI's when the model is correctly specified, but the bootstrap does not completely remove the error in the coverage probability. The J test over-rejects the correct null hypothesis. Interestingly, the errors of CI^*_{MR} are smaller when there is a larger misspecification. The conventional asymptotic and bootstrap CI's are first-order invalid under misspecification.

Figure 3 shows the coverage probabilities of the CI's over different degrees of misspecification. Again, the ability of achieving asymptotic refinements of the bootstrap

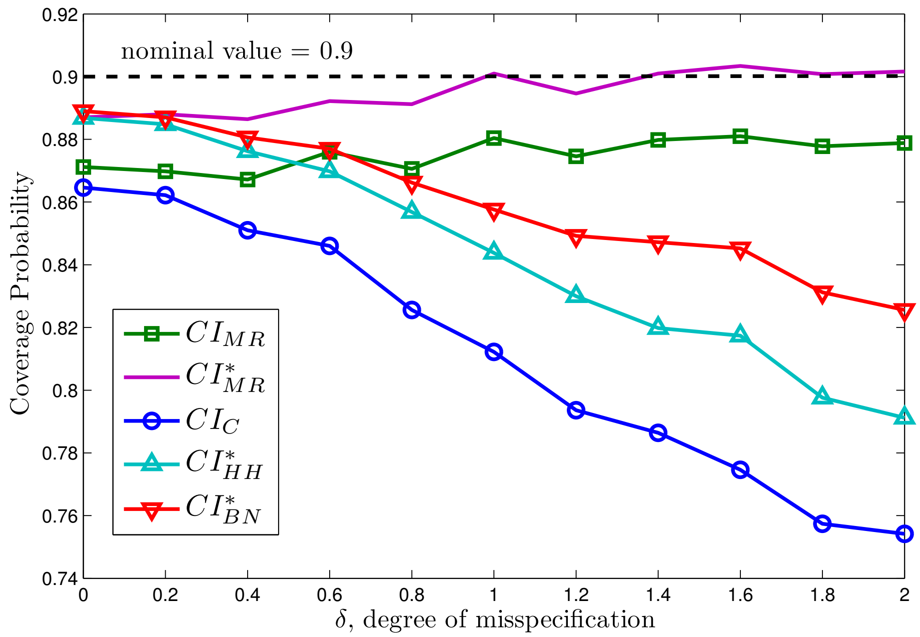


Figure 3: Coverage Probabilities of 90% Confidence Intervals for $\beta_{0(1)}$ based on the One-step GMM Estimator, $\hat{\beta}_{(1)}$, $n = 50$ in Example 2 (r=5,000, B=1,000)

CI's is clearly demonstrated at $\delta = 0$, and CI^*_{MR} maintain the ability regardless of misspecification. As the sample size grows, the invalidity of the conventional asymptotic and bootstrap CI's becomes clearer, while the gap between the asymptotic and bootstrap CI's becomes smaller.

9 Conclusion

This paper gives an alternative bootstrap procedure for GMM that achieves the sharp rate of asymptotic refinements under both correctly specified and misspecified moment condition models. The existing bootstrap procedures for GMM achieve the same rate of asymptotic refinements only under the assumption of correct model by using an additional correction, the recentering procedure. The proposed misspecification-robust bootstrap procedure requires neither the assumption of correct model nor the recentering. The use of the misspecification-robust Hall-Inoue variance estimator in constructing the sample and bootstrap versions of the test statistic is critical in implementing the bootstrap for overidentified and possibly misspecified models. Possible extensions of this paper would be to apply the MR bootstrap to the generalized empirical likelihood (GEL) estimators.

References

Agüero, J. M. and M. S. Marks (2008): "Motherhood and Female Labor Force Participation: Evidence from Infertility Shocks," *The American Economic Review* 98, 500-504.

Andrews, D. W. K. (2002): "Higher-Order Improvements of a Computationally Attractive k-step Boostrap for Extremum Estimators,"*Econometrica*, 70, 119-162.

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

Bhattacharya, R. N. (1987): "Some Aspects of Edgeworth Expansions in Statistics and Probability," in *New Perspectives in Theoretical and Applied Statistics*, ed. by M. L. Puri, J.P. Vilaploma, and W. Wertz. New York: Wiley, 157-170.

Bhattacharya, R. N. and J. K. Ghosh (1978): "On the Validity of the Formal Edgeworth Expansion," *Annals of Statistics*, 6, 434-451.

Bickel, P. J. and D. A. Freedman (1981): "Some Asymptotic Theory for the Bootstrap," *Annals of Statistics*, 9, 1196-1217.

Brown, B. W. and W. K. Newey (2002): "Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference," *Journal of Business and Economic Statistics* 20, 507-517.

Bonnet, C. and P. Dubois (2010): "Inference on Vertical Contracts Between Manufacturers and Retailers Allowing for Nonlinear Pricing and Resale Price Maintenance," *The RAND Journal of Economics* 41, 139-164.

Chen, X., H. Hong and M. Shum (2007): "Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models," *Journal of Econometrics* 141, 109-140.

French, E. and J. B. Jones (2004): "On the Distribution and Dynamics of Health Care Costs," *Journal of Applied Econometrics* 19, 705-721.

Götze, F. and C. Hipp (1983): "Asymptotic Expansions for Sums of Weakly Dependent Random Vectors," *Wahrscheinlichkeitstheorie verw. Gebiete*, 64, 211-239.

Gowrisankaran, G. and M. Rysman (2009): "Dynamics of Consumer Demand for New Durable Goods," *NBER Working Paper* No. 14737.

Hahn, J. (1996): "A Note on Bootstrapping Generalized Method of Moments Estimators," *Econometric Theory* 12, 187-197.

Hall, A. R. (2000): "Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test," *Econometrica* 68, 1517-1527.

Hall, A. R. and A. Inoue (2003): "The Large Sample Behavior of the Generalized Method of Moments Estimator in Misspecified Models," *Journal of Econometrics* 114, 361-394.

Hall, A. R. and D. Pelletier (2008): "Non-Nested Testing in Models Estimated via Generalized Method of Moments," Unpublished working paper, North Carolina State University.

Hall, P. (1988): "On Symmetric Bootstrap Confidence Intervals," *Journal of the Royal Statistical Society*, Series B, 50, 35-45.

Hall, P. (1992): *The bootstrap and Edgeworth expansion*, New York: Springer-Verlag.

Hall, P. and J. L. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica* 64, 891-916.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1054.

Hellerstein, J. and G. W. Imbens (1999): "Imposing Moment Restrictions From Auxiliary Data by Weighting," *Review of Economics and Statistics* 81, 1-14.

Horowitz, J. L. (2001): "The Bootstrap," in James J. Heckman and E. Leamer eds., *Handbook of Econometrics, Vol. 5*, New York: Elsevier Science.

Horowitz, J. L. (2003): "Bootstrap methods for Markov processes," *Econometrica* 71, 1049-1082.

Imbens, G. W. (1997): "One-step Estimators for Over-Identified Generalized Method of Moments Models," *The Review of Economic Studies* 64, 359-383.

Imbens, G. W. (2002): "Generalized Method of Moments and Empirical Likelihood," *Review of Economics and Statistics* 20, 493-506.

Imbens, G. W. and T. Lancaster (1994): "Combining Micro and Macro Data in Microeconometric Models," *The Review of Economic Studies* 61, 655-680.

Jondeau, E., H. Le Bihan and C. Galles (2004): "Assessing Generalized Method-of-Moments Estimates of the Federal Reserve Reaction Function," *Journal of Business and Economic Statistics* 22, 225-239.

Kitamura, Y. (2003): "A Likelihood-Based Approach to the Analysis of a Class of Nested and Non-Nested Models," Unpublished working paper, University of Pennsylvania.

Maasoumi, E. and P.C.B. Phillips (1982): "On the Behavior of Inconsistent Instrumental Variable Estimators," *Journal of Econometrics* 19, 183-201.

Marmer, V. and T. Otsu (2010): "Optimal Comparison of Misspecified Moment Restriction Models under Chosen measure of Fit," Unpublished working paper, University of British Columbia.

Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in R. Engle and D. McFadden eds., *Handbook of Econometrics, Vol. 4*, North-Holland.

Owen, A. B. (2001): *Empirical Likelihood*, London: Chapman and Hall.

Parker, J. A. and C. Julliard (2005): "Consumption Risk and the Cross Section of Expected Returns," *Journal of Political Economy* 113, 185-222.

Rivers, D. and Q. H. Vuong (2002): "Model Selection Tests for Nonlinear Dynamic Models," *Econometrics Journal* 5, 1-39.

Sawa, T. (1978): "Information Criteria for Discriminating Among Alternative Regression Models," *Econometrica* 46, 1273-1291.

Schennach, S. M. (2007): "Point Estimation with Exponentially Tilted Empirical Likelihood," *Annals of Statistics* 35, 634-672.

Shi, X. (2011): "Size Distortion and Modification of Classical Vuong Tests," Unpublished working paper, University of Wisconsin-Madison.

Vuong, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica* 57, 307-333.

A Appendix: Lemmas and Proofs

The proofs of the Theorem and Lemmas are analogous to those of Hall and Horowitz (1996) and Andrews (2002) by allowing possible model misspecification. Throughout the Appendix, write $g_i(\theta) = g(X_i, \theta)$, $g_i^*(\theta) = g(X_i^*, \theta)$, $G_i(\theta) = G(X_i, \theta)$, $G_i^*(\theta) = G(X_i^*, \theta)$, $f_i(\theta) = f(X_i, \theta)$, and $f_i^*(\theta) = f(X_i^*, \theta)$ for notational brevity.

A.1 Lemmas

Lemma 1 modifies Lemmas 1, 2, 6, and 7 of Andrews (2002) for nonparametric iid bootstrap under possible misspecification. The modified Lemmas 1, 2, 6, and 7 are denoted by AL1, AL2, AL6, and AL7, respectively. In addition, Lemma 5 of Andrews (2002) is denoted by AL5 without modification.

Lemma 1.

(a) *Lemma 1 of Andrews (2002) holds by replacing $\widetilde{X}_i$ and N with X_i and n, respectively, under our Assumption 1.*

(b) *Lemma 2 of Andrews (2002) for $j = 1$ holds under our Assumptions 1-3.*

(c) *Lemma 6 of Andrews (2002) holds by replacing $\widetilde{X}_i$ and N with X_i and n, respectively, and by letting $l = 1$ and $\gamma = 0$, under our Assumption 1.*

(d) *Lemma 7 of Andrews (2002) for $j = 1$ holds by replacing $\widetilde{X}_i$ and N with X_i and n, respectively, and by letting $l = 1$ and $\gamma = 0$, under our Assumptions 1-3.*

Lemmas 2-3 prove that the one-step and two-step GMM estimators are consistent for the (pseudo-)true values, $\theta_{0(1)}$ and $\theta_{0(2)}$, respectively, under possible misspecification.

Lemma 2. *Suppose Assumptions 1-3 hold. Then, for all* $c \in [0, 1/2)$ *and all* $a \geq 0$,

$$\lim_{n\to\infty} n^a P(\|\hat{\theta}_{(1)} - \theta_{0(1)}\| > n^{-c}) = 0.$$

Lemma 3. *Suppose Assumptions 1-3 hold. Then, for all* $c \in [0, 1/2)$ *and all* $a \geq 0$,

$$\lim_{n\to\infty} n^a P(\|\hat{\theta}_{(2)} - \theta_{0(2)}\| > n^{-c}) = 0.$$

Lemmas 4-5 are the bootstrap versions of Lemmas 2-3, respectively, and consistency of the MR bootstrap is established under possible misspecification. Note that the bootstrap GMM estimators are different from the Hall-Horowitz bootstrap GMM estimators, which use the recentered bootstrap moment function.

Lemma 4. *Suppose Assumptions 1-3 hold. Then, for all* $c \in [0, 1/2)$ *and all* $a \geq 0$,

$$\lim_{n\to\infty} n^a P(P^*(\|\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)}\| > n^{-c}) > n^{-a}) = 0.$$

Lemma 5. *Suppose Assumptions 1-3 hold. Then, for all* $c \in [0, 1/2)$ *and all* $a \geq 0$,

$$\lim_{n\to\infty} n^a P(P^*(\|\hat{\theta}^*_{(2)} - \hat{\theta}_{(2)}\| > n^{-c}) > n^{-a}) = 0.$$

We now introduce some additional notation. Let S_n be the vector containing the unique components of $n^{-1}\sum_{i=1}^n \left(f_i(\theta_{0(1)})', f_i(\theta_{0(2)})'\right)'$ on the support of X_i, and $S = ES_n$. Similarly, let S_n^* denote the vector containing the unique components of $n^{-1}\sum_{i=1}^n \left(f_i^*(\hat{\theta}_{(1)})', f_i^*(\hat{\theta}_{(2)})'\right)'$ on the support of X_i, and $S^* = E^* S_n^*$. Note that the definitions of S_n and S_n^* are different from those of Hall and Horowitz (1996) and Andrews (2002), because they do not distinguish $\theta_{0(1)}$ and $\theta_{0(2)}$ by assuming the unique true value θ_0. Under misspecifications, $\theta_{0(1)}$ and $\theta_{0(2)}$ are different and thus, $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(2)}$ have different probability limits. In addition, Hall and Horowitz (1996) and Andrews (2002) define S_n^* by using the recentered moment function.

Lemma 6. *Let* Δ_n *and* Δ_n^* *denote* $n^{1/2}(\hat{\theta}_{(j)} - \theta_{0(j)})$ *and* $n^{1/2}(\hat{\theta}^*_{(j)} - \hat{\theta}_{(j)})$, *or* $T_{MR(j)}$ *and* $T^*_{MR(j)}$ *for* $j = 1, 2$. *For each definition of* Δ_n *and* Δ_n^*, *there is an infinitely differentiable function* $A(\cdot)$ *with* $A(S) = 0$ *and* $A(S^*) = 0$ *such that the following results hold.*

(a) *Suppose Assumptions 1-4 hold with* $d_1 \geq 2a + 2$, *where* $2a$ *is some nonnegative integer. Then,*

$$\lim_{n\to\infty} \sup_z n^a |P(\Delta_n \leq z) - P(n^{1/2} A(S_n) \leq z)| = 0.$$

(b) *Suppose Assumptions 1-4 hold with* $d_1 \geq 2a+2$*, where* $2a$ *is some nonnegative integer. Then,*

$$\lim_{n\to\infty} n^a P\left(\sup_z |P^*(\Delta_n^* \leq z) - P^*(n^{1/2}A(S_n^*) \leq z)| > n^{-a}\right) = 0.$$

We define the components of the Edgeworth expansions of the test statistic $T_{MR(j)}$ and its bootstrap analog $T^*_{MR(j)}$. Let $\Psi_n = n^{1/2}(S_n - S)$ and $\Psi_n^* = n^{1/2}(S_n^* - S^*)$. Let $\Psi_{n,k}$ and $\Psi^*_{n,k}$ denote the kth elements of Ψ_n and Ψ_n^* respectively. Let $\nu_{n,a}$ and $\nu^*_{n,a}$ denote vectors of moments of the form $n^{\alpha(m)}E\prod_{\mu=1}^m \Psi_{n,k_\mu}$ and $n^{\alpha(m)}E^*\prod_{\mu=1}^m \Psi^*_{n,k_\mu}$, respectively, where $2 \leq m \leq 2a+2$, $\alpha(m) = 0$ if m is even, and $\alpha(m) = 1/2$ if m is odd. Let $\nu_a = \lim_{n\to\infty}\nu_{n,a}$. The limit exists under Assumption 1 of Andrews (2002), and thus under our Assumption 1.

Let $\pi_i(\delta, \nu_a)$ be a polynomial in $\delta = \partial/\partial z$ whose coefficients are polynomials in the elements of ν_a and for which $\pi_i(\delta,\nu_a)\Phi(z)$ is an even function of z when i is odd and is an odd function of z when i is even for $i = 1, ..., 2a$, where $2a$ is an integer. The Edgeworth expansions of $T_{MR(j)}$ and $T^*_{MR(j)}$ depend on $\pi_i(\delta,\nu_a)$ and $\pi_i(\delta,\nu^*_{n,a})$, respectively.

The following Lemma shows that the bootstrap moments $\nu^*_{n,a}$ are close to the population moments ν_a in large samples. The Lemma is an iid version of Lemma 14 of Andrews (2002).

Lemma 7. *Suppose Assumptions 1 and 3 hold with* $d_2 \geq 2a+1$ *for some* $a \geq 0$*. Then, for all* $c \in [0, 1/2)$*,*

$$\lim_{n\to\infty} n^a P(\|\nu^*_{n,a} - \nu_a\| > n^{-c}) = 0.$$

Lemma 8. *For* $j = 1, 2$*, (a) Suppose Assumptions 1-4 hold with* $d_1 \geq 2a+2$*, where* $2a$ *is some nonnegative integer. Then,*

$$\lim_{n\to\infty} n^a \sup_{z\in\mathbf{R}} \left| P(T_{MR(j)} \leq z) - \left[1 + \sum_{i=1}^{2a} n^{-i/2}\pi_i(\delta,\nu_a)\right]\Phi(z)\right| = 0.$$

(b) Suppose Assumptions 1-4 hold with $d_1 \geq 2a+2$ *and* $d_2 \geq 2a+1$*, where* $2a$ *is some nonnegative integer. Then,*

$$\lim_{n\to\infty} n^a P\left(\sup_{z\in\mathbf{R}} \left| P^*(T^*_{MR(j)} \leq z) - \left[1 + \sum_{i=1}^{2a} n^{-i/2}\pi_i(\delta,\nu^*_{n,a})\right]\Phi(z)\right| > n^{-a}\right) = 0.$$

A.2 Proof of Theorem 1

The usage of the Hall-Inoue variance estimators in constructing the sample and bootstrap versions of the t statistic without recentering the bootstrap moment function is taken into account by Lemmas 6 and 8. Once we establish the Edgeworth expansions of $T_{MR(j)}$ and $T^*_{MR(j)}$ for $j = 1, 2$, the proof of the Theorem is the same with that of Theorem 2(c) of Andrews (2002) with his Lemmas 13 and

16 replaced by our Lemmas 6 and 8. His proof relies on the argument of Hall (1988, 1992)'s methods developed for "smooth functions of sample averages," for iid data. *Q.E.D.*

A.3 Proofs of Lemmas

A.3.1 Proof of Lemma 1

(a) Assumption 1 of Andrews (2002) is satisfied if our Assumption 1 holds. Then, Lemma 1 of Andrews (2002) holds.
(b) We use the proof of Lemma 2 of Andrews (2002) which relies on that of Lemma 2 of Hall and Horowitz (1996). Since their proof does not require $Eg(X_i, \theta_0) = 0$, the Lemma holds under our Assumptions 1-3.
(c) Assumption 1 of Andrews (2002) is satisfied if our Assumption 1 holds. Then, Lemma 6 of Andrews (2002) holds for nonparametric iid bootstrap.
(d) We use the proof of Lemma 7 of Andrews (2002) which relies on that of Lemma 8 of Hall and Horowitz (1996). Since their proof does not require $Eg(X_i, \theta_0) = 0$, the Lemma holds for nonparametric iid bootstrap under our Assumptions 1-3. *Q.E.D.*

A.3.2 Proof of Lemma 2

Write $J(\theta) \equiv J(\theta, I_{L_g})$, $J_n(\theta) \equiv J_n(\theta, I_{L_g})$ throughout the proof for notational brevity. We first prove the result with n^{-c} replaced by arbitrary fixed $\varepsilon > 0$. Given $\varepsilon > 0$, $\exists \delta > 0$ such that $\|\theta - \theta_{0(1)}\| > \varepsilon$ implies that $J(\theta) - J(\theta_{0(1)}) \geq \delta > 0$, because $\theta_{0(1)}$ uniquely minimizes $J(\theta)$. Note that $J(\theta_{0(1)})$ may not be zero. Thus, by the triangle inequality,

$$\begin{aligned} n^a P(\|\hat{\theta}_{(1)} - \theta_{0(1)}\| > \varepsilon) &\leq n^a P(J(\hat{\theta}_{(1)}) - J_n(\hat{\theta}_{(1)}) + J_n(\hat{\theta}_{(1)}) - J(\theta_{0(1)}) > \delta) \\ &\leq n^a P(J(\hat{\theta}_{(1)}) - J_n(\hat{\theta}_{(1)}) + J_n(\theta_{0(1)}) - J(\theta_{0(1)}) > \delta) \\ &\leq n^a P\left(\sup_{\theta \in \Theta} |J(\theta) - J_n(\theta)| > \delta/2\right) = o(1). \end{aligned} \tag{A.1}$$

The last conclusion holds by AL2 and the argument in the proof of Theorem 2.6 of Newey and McFadden (1994). This proves

$$\lim_{n \to \infty} n^a P(\|\hat{\theta}_{(1)} - \theta_{0(1)}\| > \varepsilon) = 0. \tag{A.2}$$

Next, we prove the result as stated in the Lemma. The first order condition is $(\partial/\partial\theta) J_n(\hat{\theta}_{(1)}) = G'_{n(1)} g_{n(1)} = 0$ with probability $1 - o(n^{-a})$. By using the population first order condition, $G'_{0(1)} g_{0(1)} = 0$, and by the mean value theorem, with probability $1 - o(n^{-a})$,

$$\hat{\theta}_{(1)} - \theta_{0(1)} = -\left(\frac{\partial^2}{\partial\theta\partial\theta'} J_n(\tilde{\theta})\right)^{-1} \frac{\partial}{\partial\theta} J_n(\theta_{0(1)}) \tag{A.3}$$

where

$$\frac{\partial}{\partial\theta}J_n(\theta_{0(1)}) = \left\{G'_{0(1)}(g_n(\theta_{0(1)}) - g_{0(1)}) + (G_n(\theta_{0(1)}) - G_{0(1)})'g_n(\theta_{0(1)})\right\}, \quad \text{(A.4)}$$

$$\frac{\partial^2}{\partial\theta\partial\theta'}J_n(\theta) \equiv 2\tilde{H}_n(\theta, I_{L_g}) = 2\left\{(g_n(\theta)' \otimes I_{L_\theta})\, G_n^{(2)}(\theta) + G_n(\theta)'G_n(\theta)\right\}, \quad \text{(A.5)}$$

and $\tilde{\theta}$ is between $\hat{\theta}_{(1)}$ and $\theta_{0(1)}$ and may differ across rows. Note that the first and second derivatives of $J_n(\theta)$ include additional terms that do not appear under correct specification, $g_{0(1)} = 0$. Then, combining the following results proves the Lemma:

$$\lim_{n\to\infty} n^a P\left(\left\|\tilde{H}_n(\tilde{\theta}, I_{L_g}) - \tilde{H}_n(\theta_{0(1)}, I_{L_g})\right\| > \varepsilon\right) = 0, \quad \text{(A.6)}$$

$$\lim_{n\to\infty} n^a P\left(\left\|\tilde{H}_n(\theta_{0(1)}, I_{L_g}) - H_{0(1)}\right\| > \varepsilon\right) = 0, \quad \text{(A.7)}$$

$$\lim_{n\to\infty} n^a P\left(\|G_n(\theta_{0(1)}) - G_{0(1)}\| > n^{-c}\right) = 0, \quad \text{(A.8)}$$

$$\lim_{n\to\infty} n^a P\left(\|g_n(\theta_{0(1)}) - g_{0(1)}\| > n^{-c}\right) = 0. \quad \text{(A.9)}$$

To show (A.6), we apply the triangle and Cauchy-Schwarz inequalities multiple times,

$$\begin{aligned}
&\left\|\left(g_n(\tilde{\theta})' \otimes I_{L_\theta}\right) G_n^{(2)}(\tilde{\theta}) - \left(g_n(\theta_{0(1)})' \otimes I_{L_\theta}\right) G_n^{(2)}(\theta_{0(1)}) \right. \quad \text{(A.10)}\\
&\left. + G_n(\tilde{\theta})'G_n(\tilde{\theta}) - G_n(\theta_{0(1)})'G_n(\theta_{0(1)})\right\| \\
\leq\; & \|G_n^{(2)}(\tilde{\theta}) - G_n^{(2)}(\theta_{0(1)})\| \left(\|g_n(\tilde{\theta}) - g_n(\theta_{0(1)})\| + \|g_n(\theta_{0(1)})\|\right) + \|G_n^{(2)}(\theta_{0(1)})\|\|g_n(\tilde{\theta}) - g_n(\theta_{0(1)})\| \\
& + \|G_n(\tilde{\theta}) - G_n(\theta_{0(1)})\| \left(\|G_n(\tilde{\theta}) - G_n(\theta_{0(1)})\| + 2\|G_n(\theta_{0(1)})\|\right) \\
\leq\; & \|\tilde{\theta} - \theta_{0(1)}\| \left\{C_{\partial f,n}(C_{g,n} + C_{\partial f,n})\|\tilde{\theta} - \theta_{0(1)}\| + C_{g,n}\|G_n^{(2)}(\theta_{0(1)})\| + 2\|G_n(\theta_{0(1)})\| + \|g_n(\theta_{0(1)})\|\right\},
\end{aligned}$$

where $C_{g,n} = n^{-1}\sum_{i=1}^n C_g(X_i)$ and $C_{\partial f,n} = n^{-1}\sum_{i=1}^n C_{\partial f}(X_i)$. Using (A.2) and multiple applications of AL1(a) with $h(X_i) = (\partial^j/\partial\theta^j)g_i(\theta_{0(1)})$ for $j = 0, 1, 2$ or $h(X_i) = C_g(X_i)$, or $h(X_i) = C_{\partial f}(X_i)$ proves (A.6).

For (A.7), apply the triangle and Cauchy-Schwarz inequalities to get

$$\begin{aligned}
& \|(g_n(\theta_{0(1)})' \otimes I_{L_\theta})G_n^{(2)}(\theta_{0(1)}) - (g'_{0(1)} \otimes I_{L_\theta})G_{0(1)}^{(2)}\| \quad \text{(A.11)}\\
\leq\; & \|G_n^{(2)}(\theta_{0(1)}) - G_{0(1)}^{(2)}\| \cdot \|g_n(\theta_{0(1)})\| + \|G_{0(1)}^{(2)}\| \cdot \|g_n(\theta_{0(1)}) - g_{0(1)}\|,
\end{aligned}$$

and

$$\|G_n(\theta_{0(1)})'G_n(\theta_{0(1)}) - G'_{0(1)}G_{0(1)}\| \leq \|G_n(\theta_{0(1)}) - G_{0(1)}\| \cdot (\|G_n(\theta_{0(1)}) - G_{0(1)}\| + 2\|G_{0(1)}\|). \quad \text{(A.12)}$$

Then, it follows by AL1(b) with $h(X_i) = (\partial^j/\partial\theta^j)g_i(\theta_{0(1)})$ and by Lemma AL1(a) with $h(X_i) = (\partial^j/\partial\theta^j)g_i(\theta_{0(1)}) - E(\partial^j/\partial\theta^j)g_i(\theta_{0(1)})$ for $j = 0, 1, 2, c = 0$, and $p = q_2$.

The third result (A.8) holds by AL1(a) with $h(X_i) = G_i(\theta_{0(1)}) - G_{0(1)}$, $c = 0$, and $p = q_2$. The last result (A.9) follows from AL1(a) with $h(X_i) = g_i(\theta_{0(1)}) - g_{0(1)}$, $c = c$, and $p = q_1$. *Q.E.D.*

A.3.3 Proof of Lemma 3

We first prove the result with n^{-c} replaced by arbitrary fixed $\varepsilon > 0$. By Theorem 2.6 of Newey and McFadden (1994), $\sup_{\theta\in\Theta} |J_n(\theta, W_n) - J(\theta, W)| \to_p 0$, provided that $W_n \to_p W$. Then, analogous arguments to that of Lemma 2 show that

$$\lim_{n\to\infty} n^a P(\|\hat{\theta}_{(2)} - \theta_{0(2)}\| > \varepsilon) = 0. \tag{A.13}$$

By the mean value expansion of the first-order condition,

$$\hat{\theta}_{(2)} - \theta_{0(2)} = -\left(\frac{\partial^2}{\partial\theta\partial\theta'} J_n(\tilde{\theta}, W_n)\right)^{-1} \frac{\partial}{\partial\theta} J_n(\theta_{0(2)}, W_n), \tag{A.14}$$

with probability $1 - o(n^{-a})$, where

$$\begin{aligned} \frac{\partial}{\partial\theta} J_n(\theta_{0(2)}, W_n) &= G_n(\theta_{0(2)})' W_n (g_n(\theta_{0(2)}) - g_{0(2)}) \qquad \text{(A.15)} \\ &\quad + \left(G_n(\theta_{0(2)}) - G_{0(2)}\right)' W g_{0(2)} + G_n(\theta_{0(2)})'(W_n - W) g_{0(2)}, \\ \frac{\partial^2}{\partial\theta\partial\theta'} J_n(\theta, W_n) &= 2\tilde{H}_n(\theta, W_n) \qquad \text{(A.16)} \\ &= 2\left\{(g_n(\theta)' W_n \otimes I_{L_\theta}) G_n^{(2)}(\theta) + G_n(\theta)' W_n G_n(\theta)\right\}, \end{aligned}$$

and $\tilde{\theta}$ is between $\hat{\theta}_{(2)}$ and $\theta_{0(2)}$ and may differ across rows. Note that (A.15) includes additional terms that are zero under correct specification. Thus, in order to show

$$\lim_{n\to\infty} n^a P\left(\left\|\frac{\partial}{\partial\theta} J_n(\theta_{0(2)}, W_n)\right\| > n^{-c}\right) = 0, \tag{A.17}$$

we need

$$\lim_{n\to\infty} n^a P\left(\left\|g_n(\theta_{0(2)}) - g_{0(2)}\right\| > n^{-c}\right) = 0, \tag{A.18}$$

$$\lim_{n\to\infty} n^a P\left(\left\|G_n(\theta_{0(2)}) - G_{0(2)}\right\| > n^{-c}\right) = 0, \tag{A.19}$$

$$\lim_{n\to\infty} n^a P(\|W_n(\hat{\theta}_{(1)}) - W\| > n^{-c}) = 0. \tag{A.20}$$

Note that (A.19) and (A.20) are required for possibly misspecified models.[12]

(A.18) and (A.19) hold by AL1(a) with $h(X_i) = g_i(\theta_{0(2)}) - g_{0(2)}$ or $h(X_i) = G_i(\theta_{0(2)}) - G_{0(2)}$. (A.20) follows from

$$\lim_{n\to\infty} n^a P(\|W_n(\hat{\theta}_{(1)})^{-1} - W_n(\theta_{0(1)})^{-1}\| > n^{-c}) = 0, \text{ and} \tag{A.21}$$

$$\lim_{n\to\infty} n^a P(\|W_n(\theta_{0(1)})^{-1} - W^{-1}\| > n^{-c}) = 0. \tag{A.22}$$

[12] Andrews (2002) proves (A.20) by replacing n^{-c} with ε under correct specification.

To show (A.21), observe that

$$
\begin{aligned}
& \|W_n(\hat{\theta}_{(1)})^{-1} - W_n(\theta_{0(1)})^{-1}\| && \text{(A.23)} \\
= \;& \|n^{-1}\sum_{i=1}^{n}(g_i(\hat{\theta}_{(1)})g_i(\hat{\theta}_{(1)})' - g_i(,\theta_{0(1)})g_i(\theta_{0(1)})')\| + \|g_n(\theta_{0(1)})g_n(\theta_{0(1)})' - g_{n(1)}g_{n(1)}'\|.
\end{aligned}
$$

For the first term of the right-hand side of (A.23), we apply the mean value expansion and the Cauchy-Schwarz inequality to get

$$
\begin{aligned}
& \|n^{-1}\sum_{i=1}^{n}(g_i(\hat{\theta}_{(1)})g_i(\hat{\theta}_{(1)})' - g_i(\theta_{0(1)})g_i(\theta_{0(1)})')\| && \text{(A.24)} \\
\leq \;& 2n^{-1}\sum_{i=1}^{n}\sup_{\theta\in N_{0(1)}}\|G_i(\theta)\|\|g_i(\theta)\|\cdot\|\hat{\theta}_{(1)} - \theta_{0(1)}\|.
\end{aligned}
$$

For the second term of (A.23), we apply the Cauchy-Schwarz inequality,

$$
\begin{aligned}
& \|g_n(\theta_{0(1)})g_n(\theta_{0(1)})' - g_{n(1)}g_{n(1)}'\| && \text{(A.25)} \\
= \;& \|(g_n(\theta_{0(1)}) - g_n(\hat{\theta}_{(1)}))(g_n(\theta_{0(1)}) + g_n(\hat{\theta}_{(1)}))'\| \\
\leq \;& n^{-1}\sum_{i=1}^{n}\|g_i(\theta_{0(1)}) - g_i(\hat{\theta}_{(1)})\| n^{-1}\sum_{i=1}^{n}\|g_i(\theta_{0(1)}) + g_i(\hat{\theta}_{(1)})\| \\
\leq \;& \|\hat{\theta}_{(1)} - \theta_{0(1)}\|C_{g,n}(2n^{-1}\sum_{i=1}^{n}\|g_i(\theta_{0(1)})\| + \|\hat{\theta}_{(1)} - \theta_{0(1)}\|C_{g,n}).
\end{aligned}
$$

Then, AL1(b) with $h(X_i) = C_g(X_i)$, $h(X_i) = g_i(\theta_{0(1)})$, and $h(X_i) = \sup_{\theta\in N_{0(1)}}\|G_i(\theta)\|\|g_i(\theta)\|$ and Lemma 2 proves (A.21).

(A.22) holds by applications of AL1(a) with $h(X_i) = g_i(\theta_{0(1)})g_i(\theta_{0(1)})' - Eg_i(\theta_{0(1)})g_i(\theta_{0(1)})'$ and $p = q_1/2$, and $h(X_i) = g_i(\theta_{0(1)}) - g_{0(1)}$ and $p = q_1$ since

$$
\begin{aligned}
\|W_n(\theta_{0(1)})^{-1} - W^{-1}\| \;\leq\; & \left\|n^{-1}\sum_{i=1}^{n}g_i(\theta_{0(1)})g_i(\theta_{0(1)})' - Eg_i(\theta_{0(1)})g_i(\theta_{0(1)})'\right\| && \text{(A.26)} \\
& + \left(2\|g_{0(1)}\| + \|g_n(\theta_{0(1)}) - g_{0(1)}\|\right)\|g_n(\theta_{0(1)}) - g_{0(1)}\|.
\end{aligned}
$$

Lastly, the Lemma follows from

$$
\lim_{n\to\infty} n^a P\left(\left\|\tilde{H}_n(\tilde{\theta}, W_n) - \tilde{H}_n(\theta_{0(2)}, W)\right\| > \varepsilon\right) = 0, \quad \text{(A.27)}
$$

$$
\lim_{n\to\infty} n^a P\left(\left\|\tilde{H}_n(\theta_{0(2)}, W) - H_{0(2)}\right\| > \varepsilon\right) = 0, \quad \text{(A.28)}
$$

that can be shown by multiple applications of AL1 and the results (A.20) and (A.13). *Q.E.D.*

A.3.4 Proof of Lemma 4

Write $J(\theta) \equiv J(\theta, I_{L_g})$ and $J_n^*(\theta) \equiv J_n^*(\theta, I_{L_g})$ for notational brevity. First, we prove the result with n^{-c} replaced by a fixed $\varepsilon > 0$. We claim that given $\varepsilon > 0$, $\exists\delta > 0$ independent of n such that $\|\theta - \hat{\theta}_{(1)}\| > \varepsilon$ implies that $J_n(\theta) - J_n(\hat{\theta}_{(1)}) \geq \delta > 0$ with probability $1 - o(n^{-a})$. To see this, note that $\|\hat{\theta}_{(1)} - \theta_{0(1)}\| \leq \varepsilon/2$ with probability $1 - o(n^{-a})$ by Lemma 2 and write

$$\begin{aligned} J_n(\theta) - J_n(\hat{\theta}_{(1)}) &= J(\theta) - J(\theta_{0(1)}) + J_n(\theta) - J_n(\hat{\theta}_{(1)}) \\ &\quad -J(\theta) + J(\hat{\theta}_{(1)}) + J(\theta_{0(1)}) - J(\hat{\theta}_{(1)}) \\ &\geq J(\theta) - J(\theta_{0(1)}) - |J_n(\theta) - J_n(\hat{\theta}_{(1)}) - J(\theta) + J(\hat{\theta}_{(1)})| \\ &\quad -|J(\hat{\theta}_{(1)}) - J(\theta_{0(1)})|. \end{aligned} \tag{A.29}$$

Define $M = \inf_{\theta \in N_\varepsilon(\hat{\theta}_{(1)})^c \cap \Theta} J(\theta) - J(\theta_{0(1)})$, where $N_\varepsilon(\hat{\theta}_{(1)})^c = \{\theta : \|\theta - \hat{\theta}_{(1)}\| > \varepsilon\}$, then $M > 0$ because (i) $J(\theta)$ is uniquely minimized at $\theta_{0(1)}$ and is continuous on Θ, and (ii) we can take a neighborhood around $\theta_{0(1)}$ such that $N_{\varepsilon/4}(\theta_{0(1)}) \subset N_\varepsilon(\hat{\theta}_{(1)})$. By AL2 and the proof of Theorem 2.6 of Newey and McFadden (1994), we have (iii) $\lim_{n\to\infty} n^a P(\sup_{\theta\in\Theta} |J_n(\theta) - J_n(\hat{\theta}_{(1)}) - J(\theta) + J(\hat{\theta}_{(1)})| > \lambda) = 0$ for all $\lambda > 0$ and (iv) $\lim_{n\to\infty} n^a P(|J(\hat{\theta}_{(1)}) - J(\theta_{0(1)})| > \lambda) = 0$ by Lemma 2. Taking $\lambda < M/2$ proves the claim.

Thus, we have

$$\begin{aligned} &n^a P(P^*(\|\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)}\| > \varepsilon) > n^{-a}) \\ \leq\; &n^a P(P^*(J_n(\hat{\theta}^*_{(1)}) - J_n^*(\hat{\theta}^*_{(1)}) + J_n^*(\hat{\theta}^*_{(1)}) - J_n(\hat{\theta}_{(1)}) > \delta) > n^{-a}) \\ \leq\; &n^a P(P^*(J_n(\hat{\theta}^*_{(1)}) - J_n^*(\hat{\theta}^*_{(1)}) + J_n^*(\hat{\theta}_{(1)}) - J_n(\hat{\theta}_{(1)}) > \delta) > n^{-a}) \\ \leq\; &n^a P\left(P^*\left(\sup_{\theta\in\Theta} |J_n^*(\theta) - J_n(\theta)| > \delta/2\right) > n^{-a}\right) \to 0, \end{aligned} \tag{A.30}$$

since $\hat{\theta}^*_{(1)}$ is the minimizer of $J_n^*(\theta)$. To verify the last conclusion of (A.30), we apply the triangle and Cauchy-Schwarz inequalities,

$$\begin{aligned} |J_n^*(\theta) - J_n(\theta)| &= |g_n^*(\theta)' g_n^*(\theta) - g_n(\theta)' g_n(\theta)| \\ &\leq \|g_n^*(\theta) - g_n(\theta)\|^2 + 2\left(\|g_n(\theta) - Eg(X_i,\theta)\| + \|Eg(X_i,\theta)\|\right) \|g_n^*(\theta) - g_n(\theta)\| \\ &= \|g_n^*(\theta) - E^* g_i^*(\theta)\|^2 \\ &\quad +2\left(\|g_n(\theta) - Eg(X_i,\theta)\| + \|Eg(X_i,\theta)\|\right) \|g_n^*(\theta) - E^* g_i^*(\theta)\|, \end{aligned} \tag{A.31}$$

and apply AL2 and AL7.

Next, we prove the result stated in the Lemma. The first-order condition is $(\partial/\partial\theta) J_n^*(\hat{\theta}^*_{(1)}) = G_n^*(\hat{\theta}^*_{(1)})' g_n^*(\hat{\theta}^*_{(1)}) = 0$ with P^* probability $1 - o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$. By the mean value theorem,

$$\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)} = -\left(\frac{\partial^2}{\partial\theta\partial\theta'} J_n^*(\tilde{\theta}^*)\right)^{-1} \frac{\partial}{\partial\theta} J_n^*(\hat{\theta}_{(1)}), \tag{A.32}$$

with P^* probability $1 - o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$, where $\tilde{\theta}^*$ is between $\hat{\theta}^*_{(1)}$ and $\hat{\theta}_{(1)}$ and may differ across rows. The proof follows that of Lemma 2 with some modifications for the bootstrap version.

First, we prove

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|\frac{\partial}{\partial\theta} J_n^*(\hat{\theta}_{(1)})\right\| > n^{-c}\right) > n^{-a}\right) = 0, \tag{A.33}$$

where

$$\frac{\partial}{\partial\theta} J_n^*(\hat{\theta}_{(1)}) = G_n(\hat{\theta}_{(1)})' \left(g_n^*(\hat{\theta}_{(1)}) - g_n(\hat{\theta}_{(1)})\right) + \left(G_n^*(\hat{\theta}_{(1)}) - G_n(\hat{\theta}_{(1)})\right)' g_n^*(\hat{\theta}_{(1)}), \tag{A.34}$$

since the sample first-order condition $G_n(\hat{\theta}_{(1)})' g_n(\hat{\theta}_{(1)}) = 0$ holds. This can be done by combining the following results,

$$\lim_{n\to\infty} n^a P(P^*(\|G_n(\hat{\theta}_{(1)})\| > \varepsilon) > n^{-a}), \tag{A.35}$$

$$\lim_{n\to\infty} n^a P(P^*(\|g_n^*(\hat{\theta}_{(1)})\| > \varepsilon) > n^{-a}), \tag{A.36}$$

$$\lim_{n\to\infty} n^a P(P^*(\|g_n^*(\hat{\theta}_{(1)}) - g_n(\hat{\theta}_{(1)})\| > n^{-c}) > n^{-a}), \tag{A.37}$$

$$\lim_{n\to\infty} n^a P(P^*(\|G_n^*(\hat{\theta}_{(1)}) - G_n(\hat{\theta}_{(1)})\| > n^{-c}) > n^{-a}). \tag{A.38}$$

For (A.35), note that $\|G_n(\hat{\theta}_{(1)})\| \leq \|G_n(\theta_{0(1)})\| + \|G_n(\hat{\theta}_{(1)}) - G_n(\theta_{0(1)})\|$ holds by the triangle inequality and claim

$$\lim_{n\to\infty} n^a P(P^*(\|G_n(\theta_{0(1)})\| > \varepsilon) > n^{-a}) = 0, \tag{A.39}$$

$$\lim_{n\to\infty} n^a P(P^*(\|G_n(\hat{\theta}_{(1)}) - G_n(\theta_{0(1)})\| > \varepsilon) > n^{-a}) = 0. \tag{A.40}$$

To see this, observe that $P^*(\|G_n(\theta_{0(1)})\| > \varepsilon) = 1\{\|G_n(\theta_{0(1)})\| > \varepsilon\}$, where $1\{\cdot\}$ is an indicator function. Then,

$$\begin{aligned} & n^a P(P^*(\|G_n(\theta_{0(1)})\| > \varepsilon) > n^{-a}) \\ = \; & n^a P(1\{\|G_n(\theta_{0(1)})\| > \varepsilon\} > n^{-a}, \|G_n(\theta_{0(1)})\| > \varepsilon) \\ & + n^a P(1\{\|G_n(\theta_{0(1)})\| > \varepsilon\} > n^{-a}, \|G_n(\theta_{0(1)})\| \leq \varepsilon) \\ \leq \; & n^a P(\|G_n(\theta_{0(1)})\| > \varepsilon) \to 0, \end{aligned} \tag{A.41}$$

by AL1(b). (A.40) can be shown similarly by applying AL1(a). By (A.39) and (A.40), the first result (A.35) is proved. To show the second result (A.36), apply the triangle inequality and Assumption 2 to get

$$\begin{aligned} \|g_n^*(\hat{\theta}_{(1)})\| & \leq \|g_n^*(\theta_{0(1)})\| + \|g_n^*(\theta_{0(1)}) - g_n^*(\hat{\theta}_{(1)})\| \\ & \leq \|g_n^*(\theta_{0(1)})\| + C_{g,n}^* \|\hat{\theta}_{(1)} - \theta_{0(1)}\|, \end{aligned} \tag{A.42}$$

where $C^*_{g,n} = n^{-1}\sum_{i=1}^n C_g(X_i^*)$. By applying AL6(d) and Lemma 2, we have the result (A.36). For the third and the last result, we apply the triangle inequality and Assumptions 2-3,

$$\begin{aligned} \|g_n^*(\hat{\theta}_{(1)}) - g_n(\hat{\theta}_{(1)})\| &\leq \|g_n^*(\theta_{0(1)}) - g_n(\theta_{0(1)})\| + \|\hat{\theta}_{(1)} - \theta_{0(1)}\|(C_{g,n} + C^*_{g,n}), \quad \text{(A.43)} \\ \|G_n^*(\hat{\theta}_{(1)}) - G_n(\hat{\theta}_{(1)})\| &\leq \|G_n^*(\theta_{0(1)}) - G_n(\theta_{0(1)})\| + \|\hat{\theta}_{(1)} - \theta_{0(1)}\|(C_{\partial f,n} + C^*_{\partial f,n}), \end{aligned}$$

where $C^*_{\partial f,n} = n^{-1}\sum_{i=1}^n C_{\partial f}(X_i^*)$. Let $h(X_i) = g_i(\theta_{0(1)}) - g_{0(1)}$ or $h(X_i) = G_i(\theta_{0(1)}) - G_{0(1)}$ so that $Eh(X_i) = 0$. Then, $h(X_i^*) = g_i^*(\theta_{0(1)}) - g_{0(1)}$ or $h(X_i^*) = G_i^*(\theta_{0(1)}) - G_{0(1)}$, and $\|g_n^*(\theta_{0(1)}) - g_n(\theta_{0(1)})\| = \|n^{-1}\sum_{i=1}^n h(X_i^*) - E^*h(X_i^*)\|$ or $\|G_n^*(\theta_{0(1)}) - G_n(\theta_{0(1)})\| = \|n^{-1}\sum_{i=1}^n h(X_i^*) - E^*h(X_i^*)\|$. Now, we apply AL6(a). For the second terms on the right-hand side, apply Lemma 2 and Assumption 3. This proves the result (A.37) and (A.38).

Next, we claim

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|\tilde{H}_n^*(\tilde{\theta}^*, I_{L_g}) - \tilde{H}_n^*(\theta_{0(1)}, I_{L_g})\right\| > \varepsilon\right) > n^{-a}\right) = 0, \quad \text{(A.44)}$$

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|\tilde{H}_n^*(\theta_{0(1)}, I_{L_g}) - H_{0(1)}\right\| > \varepsilon\right) > n^{-a}\right) = 0, \quad \text{(A.45)}$$

where $\tilde{H}_n^*(\theta, I_{L_g}) = (g_n^*(\theta)' \otimes I_{L_\theta})G_n^{(2)*}(\theta) + G_n^*(\theta)'G_n^*(\theta)$ and $(\partial^2/\partial\theta\partial\theta')J_n^*(\theta) = 2\tilde{H}_n^*(\theta, I_{L_g})$. Similar arguments with the proof of Lemma 2 prove (A.44) and (A.45) using AL6 in place of AL1. In particular, $\|\tilde{\theta}^* - \theta_{0(1)}\| \leq \|\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)}\| + \|\hat{\theta}_{(1)} - \theta_{0(1)}\|$ by the triangle inequality and we use Lemma 2 and (A.30). By combining (A.33), (A.44), and (A.45), the Lemma follows. *Q.E.D.*

A.3.5 Proof of Lemma 5

We first show that

$$\lim_{n\to\infty} n^a P(P^*(\|W_n^*(\hat{\theta}^*_{(1)}) - W\| > n^{-c}) > n^{-a}) = 0. \quad \text{(A.46)}$$

This follows from

$$\lim_{n\to\infty} n^a P(P^*(\|W_n^*(\hat{\theta}^*_{(1)})^{-1} - W_n^*(\theta_{0(1)})^{-1}\| > n^{-c}) > n^{-a}) = 0, \quad \text{(A.47)}$$

$$\lim_{n\to\infty} n^a P(P^*(\|W_n^*(\theta_{0(1)})^{-1} - W^{-1}\| > n^{-c}) > n^{-a}) = 0. \quad \text{(A.48)}$$

To obtain (A.47), we use the same argument as that in the proof of Lemma 3 and the triangle inequality to show

$$\begin{aligned} \|W_n^*(\hat{\theta}^*_{(1)})^{-1} - W_n^*(\theta_{0(1)})^{-1}\| &\leq C^*\|\hat{\theta}^*_{(1)} - \theta_{0(1)}\| \quad \text{(A.49)} \\ &\leq C^*\left(\|\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)}\| + \|\hat{\theta}_{(1)} - \theta_{0(1)}\|\right), \end{aligned}$$

where

$$C^* = \left\{2n^{-1}\sum_{i=1}^n \sup_{\theta\in N_{0(1)}} \|G_i^*(\theta)\|\|g_i^*(\theta)\| + C^*_{g,n}(2n^{-1}\sum_{i=1}^n \|g_i^*(\theta_{0(1)})\| + \|\hat{\theta}^*_{(1)} - \theta_{0(1)}\|C^*_{g,n})\right\}. \quad \text{(A.50)}$$

Apply AL6(d) with $h(X_i) = C_g(X_i)$, $h(X_i) = g_i(\theta_{0(1)})$, and $h(X_i) = \sup_{\theta \in N_{0(1)}} \|G_i(\theta)\| \|g_i(\theta)\|$ and use Lemmas 2 and 4 to get (A.47). The proof of (A.48) is analogous to that of (A.22) with AL6(c) in place of AL1(a), using the same $h(X_i)$, c, and p.

For the rest of the proof, we write $W_n^* \equiv W_n^*(\hat{\theta}_{(1)}^*)$ and $W_n \equiv W_n(\hat{\theta}_{(1)})$ for notational brevity. Analogous arguments to that of Lemma 2 and Lemma 4 with (A.46) show that

$$\lim_{n\to\infty} n^a P(P^*(\|\hat{\theta}_{(2)}^* - \hat{\theta}_{(2)}\| > \varepsilon) > n^{-a}) = 0. \tag{A.51}$$

The first-order condition is $(\partial/\partial\theta) J_n^*(\hat{\theta}_{(2)}^*, W_n^*) = 0$, with P^* probability $1 - o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$. By the mean value theorem,

$$\hat{\theta}_{(2)}^* - \hat{\theta}_{(2)} = -\left(\frac{\partial^2}{\partial\theta\partial\theta'} J_n^*(\tilde{\theta}^*, W_n^*)\right)^{-1} \frac{\partial}{\partial\theta} J_n^*(\hat{\theta}_{(2)}, W_n^*), \tag{A.52}$$

with P^* probability $1 - o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$, where $\tilde{\theta}^*$ is between $\hat{\theta}_{(2)}^*$ and $\hat{\theta}_{(2)}$ and may differ across rows. Write

$$\begin{aligned} \frac{\partial}{\partial\theta} J_n^*(\hat{\theta}_{(2)}, W_n^*) &= G_n^*(\hat{\theta}_{(2)})' W_n^* g_n^*(\hat{\theta}_{(2)}) \qquad \text{(A.53)} \\ &= G_n^*(\hat{\theta}_{(2)})' W_n^* \left(g_n^*(\hat{\theta}_{(2)}) - g_n(\hat{\theta}_{(2)})\right) + \left(G_n^*(\hat{\theta}_{(2)}) - G_n(\hat{\theta}_{(2)})\right)' W g_n(\hat{\theta}_{(2)}) \\ &\quad + G_n^*(\hat{\theta}_{(2)})'(W_n^* - W) g_n(\hat{\theta}_{(2)}) + G_n(\hat{\theta}_{(2)})'(W - W_n) g_n(\hat{\theta}_{(2)}), \end{aligned}$$

since the sample first-order condition $G_n(\hat{\theta}_{(2)})' W_n g_n(\hat{\theta}_{(2)}) = 0$ holds.

For the fist term of the right-hand side, by the triangle inequality and Assumptions 2-3,

$$\begin{aligned} \|W_n^*\| &\leq \|W\| + \|W_n^* - W\|, \qquad \text{(A.54)} \\ \|G_n^*(\hat{\theta}_{(2)})\| &\leq \|G_n^*(\theta_{0(2)})\| + \|G_n^*(\hat{\theta}_{(2)}) - G_n^*(\theta_{0(2)})\| \\ &\leq \|G_n^*(\theta_{0(2)})\| + C_{\partial f,n}^* \|\hat{\theta}_{(2)} - \theta_{0(2)}\|, \\ \|g_n^*(\hat{\theta}_{(2)}) - g_n(\hat{\theta}_{(2)})\| &\leq \|g_n^*(\theta_{0(2)}) - g_n(\theta_{0(2)})\| + \|g_n^*(\hat{\theta}_{(2)}) - g_n^*(\theta_{0(2)})\| + \|g_n(\hat{\theta}_{(2)}) - g_n(\theta_{0(2)})\| \\ &\leq \|g_n^*(\theta_{0(2)}) - g_n(\theta_{0(2)})\| + \|\hat{\theta}_{(2)} - \theta_{0(2)}\| (C_{g,n}^* + C_{g,n}). \end{aligned}$$

We apply AL6(a) with $h(X_i) = g_i(\theta_{0(2)})$, AL6(d), Lemma 3, and (A.51) to show that

$$\lim_{n\to\infty} n^a P(P^* \left\| G_n^*(\hat{\theta}_{(2)})' W_n^* \left(g_n^*(\hat{\theta}_{(2)}) - g_n(\hat{\theta}_{(2)})\right)\right\| > n^{-c}) > n^{-a}) = 0. \tag{A.55}$$

Similar arguments apply to the remaining terms and we conclude that

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|\frac{\partial}{\partial\theta} J_n^*(\hat{\theta}_{(2)}, W_n^*)\right\| > n^{-c}\right) > n^{-a}\right) = 0. \tag{A.56}$$

Now, the Lemma follows from

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|\tilde{H}_n^*(\tilde{\theta}^*, W_n^*) - \tilde{H}_n^*(\theta_{0(2)}, W)\right\| > \varepsilon\right) > n^{-a}\right) = 0, \tag{A.57}$$

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|\tilde{H}_n^*(\theta_{0(2)}, W) - H_{0(2)}\right\| > \varepsilon\right) > n^{-a}\right) = 0, \tag{A.58}$$

where $\tilde{H}_n^*(\theta, W_n^*) = (g_n^*(\theta)'W_n^* \otimes I_{L_\theta})G_n^{(2)*}(\theta) + G_n^*(\theta)'W_n^*G_n^*(\theta)$ and $(\partial^2/\partial\theta\partial\theta')J_n^*(\theta, W_n^*) = 2\tilde{H}_n^*(\theta, W_n^*)$. The proof is analogous to that given in Lemma 4, by applying the Cauchy-Schwarz inequality and the triangle inequality multiple times. In particular, we use the triangle inequality to get $\|\tilde{\theta}^* - \theta_{0(2)}\| \leq \|\hat{\theta}^*_{(2)} - \hat{\theta}_{(2)}\| + \|\hat{\theta}_{(2)} - \theta_{0(2)}\|$, and apply Lemma 3 and (A.51). *Q.E.D.*

A.3.6 Proof of Lemma 6

(a) The proof mimics that of Proposition 1 of Hall and Horowitz (1996), but the proof differs from theirs by allowing distinct probability limits for the one-step and the two-step GMM estimators. The main problem to be solved is showing that $\hat{\theta}_{(j)} - \theta_{0(j)}$ can be approximated by a function of sample moments. First, let $\delta_n = \hat{\theta}_{(1)} - \theta_{0(1)}$ and δ_{ni} denote the ith component of δ_n. Write $J_n(\theta) \equiv J_n(\theta, I_{L_g})$ for notational brevity. Using the convention of summing over common subscripts, a Taylor expansion of $0 = \partial J_n(\hat{\theta}_{(1)})/\partial\theta$ about $\theta = \theta_{0(1)}$ yields

$$0 = \frac{\partial J_n(\theta_{0(1)})}{\partial\theta} + \frac{\partial^2 J_n(\theta_{0(1)})}{\partial\theta\partial\theta'}\delta_n + \frac{1}{2}\frac{\partial^3 J_n(\theta_{0(1)})}{\partial\theta\partial\theta_i\partial\theta_j}\delta_{ni}\delta_{nj} \tag{A.59}$$

$$+ \cdots + \frac{1}{(d_1-1)!}\frac{\partial^{d_1} J_n(\theta_{0(1)})}{\partial\theta\partial\theta_i\cdots\partial\theta_\kappa}\delta_{ni}\cdots\delta_{n\kappa} + \zeta_n, \tag{A.60}$$

with probability $1 - o(n^{-a})$, where

$$\zeta_n = \frac{1}{(d_1-1)!}\left(\frac{\partial^{d_1} J_n(\bar{\theta}_n)}{\partial\theta\partial\theta_i\cdots\partial\theta_\kappa} - \frac{\partial^{d_1} J_n(\theta_{0(1)})}{\partial\theta\partial\theta_i\cdots\partial\theta_\kappa}\right)\delta_{ni}\cdots\delta_{n\kappa}, \tag{A.61}$$

and $\bar{\theta}_n$ is between $\hat{\theta}_{(1)}$ and $\theta_{0(1)}$ and may differ across rows. Let R_n be the column vector whose elements are the unique components of $\partial^m J_n(\theta_{0(1)})/\partial\theta\partial\theta_i\cdots\partial\theta_\kappa$, $m = 1, ..., d_1 - 1$. Note that $N\{i, ..., \kappa\} = m - 1$ and $i, ..., \kappa = 1, ..., L_\theta$, where $N\{\cdot\}$ is the number of elements in the set. Let R denote almost sure limit of R_n as $n \to \infty$ and e_n be the conformable vector $(\zeta_n', 0, ..., 0)'$ such that the dimension of e_n is the same with that of R_n.

Then, (A.60) can be rewritten as $0 = \Xi(\delta_n, R_n + e_n)$, where $\Xi(\cdot,\cdot)$ is a polynomial and thus, infinitely differentiable with respect to its arguments. Consider a sequence of δ_n and $R_n + e_n$, then $0 = \Xi(\delta_n, R_n + e_n)$ holds for every n and $0 = \Xi(0, R)$ because δ_n and e_n converge to zero as $n \to \infty$. Let $\delta = \theta - \theta_{0(1)}$. If we differentiate Ξ with respect to its first argument and evaluate at $\delta = 0$, we have $(\partial^2/\partial\theta\partial\theta')J_n(\theta_{0(1)})$. $[(\partial^2/\partial\theta\partial\theta')J_n(\theta_{0(1)})]^{-1}$ exists and bounded with probability $1 - o(n^{-a})$ by AL1. Now, we apply the implicit function theorem to (A.60) and get the result that there is a function Λ_1 such that $\Lambda_1(R)=0$, Λ_1 is infinitely differentiable in a neighborhood of R, and

$$\hat{\theta}_{(1)}-\theta_{0(1)}\equiv\delta_n=\Lambda_1(R_n+e_n). \tag{A.62}$$

Each component of R_n is a continuous function of S_n. By AL1(a), for any $\varepsilon>0$, $\|R_n-R\|\leq\varepsilon$ with probability $1-o(n^{-a})$. By multiple applications of AL1(a) and AL1(b), similar arguments with the proof of Lemma 2 show that $\|\zeta_n\|<M\|\hat{\theta}_{(1)}-\theta_{0(1)}\|^{d_1}$ for some $M<\infty$ with probability $1-o(n^{-a})$. It follows from Lemma 2 that $\|e_n\|\leq n^{-d_1c}$ with probability $1-o(n^{-a})$. Therefore, by the mean value theorem for some $\tilde{M}<\infty$,

$$n^aP\left(\|(\hat{\theta}_{(1)}-\theta_{0(1)})-\Lambda_1(R_n)\|>n^{-d_1c}\right)\leq n^aP\left(\tilde{M}\|e_n\|>n^{-d_1c}\right)=o(1), \tag{A.63}$$

as $n\to\infty$. In order to apply AL5(a) with $\xi_n=n^{1/2}\zeta_n$, we need $d_1c\geq a+1/2$ for some $c\in[0,1/2)$ and we need $2a$ to be an integer. Both hold by assumption of the Lemma. By the result (A.63) and AL5(a),

$$\lim_{n\to\infty}\sup_z n^a\left|P\left(n^{1/2}(\hat{\theta}_{(1)}-\theta_{0(1)})\leq z\right)-P\left(n^{1/2}\Lambda_1(R_n)\leq z\right)\right|=0. \tag{A.64}$$

Now write $J_n(\hat{\theta},\tilde{\theta})\equiv J_n(\hat{\theta},W_n(\tilde{\theta}))$ and let $(\partial_1/\partial\theta)J(\cdot,\cdot)$ denote the gradient of $J_n(\cdot,\cdot)$ with respect to its first argument. Then, $\partial_1J_n(\hat{\theta}_{(2)},\hat{\theta}_{(1)})/\partial\theta=0$ with probability $1-o(n^{-a})$ by the first-order condition. Let $\eta_n=[(\hat{\theta}_{(2)}-\theta_{0(2)})',(\hat{\theta}_{(1)}-\theta_{0(1)})']'$, and let η_{ni} be the ith component of η_n. Then, a Taylor series expansion of $\partial_1J_n(\hat{\theta}_{(2)},\hat{\theta}_{(1)})/\partial\theta$ through order d_1 about $(\theta,\tilde{\theta})=(\theta_{0(2)},\theta_{0(1)})$ [13] that with probability $1-o(n^{-a})$

$$0=\frac{\partial_1J_n(\theta_{0(2)},\theta_{0(1)})}{\partial\theta}+Q_n^2\eta_n+\frac{1}{2}Q_n^3\eta_{ni}\eta_{nj}+\cdots+\frac{1}{(d_1-1)!}Q_n^{d_1}\eta_{ni}\eta_{nj}...\eta_{n\kappa}+\nu_n, \tag{A.65}$$

where $N\{i,j,...,\kappa\}=d_1-1$, Q_n^m is the mth order derivative of $\partial_1J_n(\cdot,\cdot)/\partial\theta$ with respect to both of its arguments evaluated at $(\theta_{0(2)},\theta_{0(1)})$, and ν_n is the remainder term of the Taylor series expansion, where $\|\nu_n\|=O(\|\eta_n\|^{d_1})$. Observe that $(\partial_1^2/\partial\theta\partial\theta')J_n(\theta_{0(2)},\theta_{0(1)})$ is the coefficient of $\hat{\theta}_{(2)}-\theta_{0(2)}$ in (A.65) and its inverse exists and is bounded with probability $1-o(n^{-a})$ by AL1. Using arguments similar to those used in proving (A.62), we apply the implicit function theorem to obtain

$$\hat{\theta}_{(2)}-\theta_{0(2)}=\Lambda_2\left(S_n,\nu_n,\Lambda_1(R_n+e_n)\right) \tag{A.66}$$

with probability $1-o(n^{-a})$ for some Λ_2, $\Lambda_2(S,0,0)=0$ and Λ_2 is infinitely differentiable in a neighborhood of $(S,0,0)$. By Lemma 2 and Lemma 3, $\|\eta_n\|<n^{-c}$ and thus, $\|\nu_n\|<n^{-d_1c}$ with

[13] Hall and Horowitz (1996) takes the Taylor expansion around $(\theta_a,\theta_b)=(\theta_0,\theta_0)$, the unique true value. Thus, each term of the expansion can be expressed as a function of $n^{-1}\sum_i^n f(X_i,\theta_0)$. This can be done only under the assumption of correct model specification.

probability $1 - o(n^{-a})$. By the triangle inequality and the mean value theorem,

$$\begin{aligned} & \|\Lambda_2(S_n, \nu_n, \Lambda_1(R_n + e_n)) - \Lambda_2(S_n, 0, 0)\| \quad \text{(A.67)} \\ \leq \; & \|\Lambda_2(S_n, \nu_n, \Lambda_1(R_n + e_n)) - \Lambda_2(S_n, 0, \Lambda_1(R_n + e_n))\| \\ & + \|\Lambda_2(S_n, 0, \Lambda_1(R_n + e_n)) - \Lambda_2(S_n, 0, \Lambda_1(R_n))\| + \|\Lambda_2(S_n, 0, \Lambda_1(R_n)) - \Lambda_2(S_n, 0, 0)\| \\ \leq \; & M_1\|\nu_n\| + M_2\|e_n\| + M_3\|R_n - R\| \end{aligned}$$

for some $M_k < \infty$, $k = 1, 2, 3$. It follows that $n^a P\left(\|(\hat{\theta}_{(2)} - \theta_{0(2)}) - \Lambda_2(S_n, 0, 0)\| > n^{-d_1 c}\right) = o(1)$ and by AL5,

$$\lim_{n\to\infty} \sup_z n^a \left| P\left(n^{1/2}(\hat{\theta}_{(2)} - \theta_{0(2)}) \leq z\right) - P\left(n^{1/2}\Lambda_2(S_n, 0, 0) \leq z\right)\right| = 0. \quad \text{(A.68)}$$

For $T_{MR(j)}$, we use the fact that the variance matrix estimator, $\hat{\Sigma}_{MR(j)}$, is a function of $\hat{\theta}_{(j)}$, $j = 1, 2$, by construction. Write $\hat{\Sigma}_{MR(1)}(\hat{\theta}_{(1)}) \equiv \hat{\Sigma}_{MR(1)}$ and $\hat{\Sigma}_{MR(2)}(\hat{\theta}_{(1)}, \hat{\theta}_{(2)}) \equiv \hat{\Sigma}_{MR(2)}$, so that $T_{MR(1)}(\theta) = n^{1/2}(\theta - \theta_{0(1)})/(\hat{\Sigma}_{MR(1)}(\theta))^{1/2}$ and $T_{MR(2)}(\theta_a, \theta_b) = n^{1/2}(\theta_b - \theta_{0(1)})/(\hat{\Sigma}_{MR(2)}(\theta_a, \theta_b))^{1/2}$, where $\theta = (\theta_a', \theta_b')'$ for $T_{MR(2)}(\cdot, \cdot)$. Then, $T_{MR(1)}(\theta_{0(1)}) = 0$, $T_{MR(2)}(\theta_{0(1)}, \theta_{0(2)}) = 0$ and their derivatives through order $d_1 - 1$ are functions of S_n. To ensure the existence of the derivatives of $T_{MR(j)}$, we need at least $d_1 + 1$ times differentiability of $g_i(\theta)$ with respect to θ because $\Sigma_{MR(j)}$ involves second derivatives of the moment function. By Assumption 3(c), this is satisfied.

Taylor series expansions of $T_{MR(1)}$ about $\theta = \theta_{0(1)}$ through order d_1 yields results of the form $T_{MR(1)} = n^{1/2}[\Lambda_3(S_n, \hat{\theta}_{(1)} - \theta_{0(1)}) + \zeta_n]$, where ζ_n is the remainder term of the expansion, $\|\zeta_n\| = O(\|\hat{\theta}_{(1)} - \theta_{0(1)}\|^{d_1})$, Λ_3 is infinitely differentiable in a neighborhood of $(S, 0)$, and $\Lambda_3(S, 0) = 0$. Since $\|\eta_n\| < n^{-c}$ with probability $1 - o(n^{-a})$ by Lemma 2 and 3, the result follows from AL5. The proof for $T_{MR(2)}$ proceeds similarly.

(b) The proof mimics that of Proposition 2 of Hall and Horowitz (1996). Let R_n^* be the column vector whose elements are the unique components of $\partial^m J_n^*(\hat{\theta}_{(1)})/\partial\theta\partial\theta_i \cdots \partial\theta_\kappa$, $m = 1, ...d_1 - 1$, $N\{i, ..., \kappa\} = m - 1$, and $i, ..., \kappa = 1, ..., L_\theta$. Then, R_n^* is the same with R_n, except that we place X_i^* instead of X_i. Let $\delta_n^* = \hat{\theta}_{(1)}^* - \hat{\theta}_{(1)}$ and let e_n^* be a conformable column vector with zeros for all but its first L_θ elements. Apply a Taylor expansion of the bootstrap first-order condition around $\hat{\theta}_{(1)}^* = \hat{\theta}_{(1)}$ to obtain

$$0 = \frac{\partial J_n^*(\hat{\theta}_{(1)})}{\partial\theta} + \frac{\partial^2 J_n(\hat{\theta}_{(1)})}{\partial\theta\partial\theta'}\delta_n^* + \cdots + \frac{1}{(d_1 - 1)!}\frac{\partial^{d_1} J_n^*(\hat{\theta}_{(1)})}{\partial\theta\partial\theta_i ... \partial\theta_\kappa}\delta_{ni}^* ... \delta_{n\kappa}^* + \zeta_n^*, \quad \text{(A.69)}$$

with P^* probability $1 - o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$, where ζ_n^* is the remainder term. Define Λ as in (A.62). Since all the terms in the expansion are the same with (A.60) by replacing R_n and $\theta_{0(1)}$ with R_n^* and $\hat{\theta}_{(1)}$, we can write

$$\hat{\theta}_{(1)}^* - \hat{\theta}_{(1)} \equiv \delta_n^* = \Lambda_1(R_n^* + e_n^*) \quad \text{(A.70)}$$

with P^* probability $1 - o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$ (That is, for all $\varepsilon > 0$, $\lim_{n\to\infty} n^a P(P^*(\|(\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)}) - \Lambda_1(R^*_n + e^*_n)\| > \varepsilon) > n^{-a}) = 0.$). Observe that $\Lambda_1(R^*) = 0$, where $R^* = E^* R^*_n$. This can be verified by increasing the number of the bootstrap draw given the sample, χ_n, because δ^*_n and e^*_n converge to zero conditional on χ_n. Since $\|\zeta^*_n\| < M^* \|\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)}\|^{d_1}$ for some $M^* < \infty$, Lemma 4 yields $\lim_{n\to\infty} n^a P\left(P^*\left(\|e^*_n\| > n^{-d_1 c}\right) > n^{-a}\right) = 0$ and thus,

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|(\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)}) - \Lambda_1(R^*_n)\| > n^{-d_1 c}\right) > n^{-a}\right) = 0. \tag{A.71}$$

By AL5(b),

$$\lim_{n\to\infty} n^a P\left(\sup_z \left| P^*(n^{1/2}(\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)}) \leq z) - P^*(n^{1/2}\Lambda_1(R^*_n) \leq z) \right| > n^{-a}\right) = 0. \tag{A.72}$$

For the rest of the proof, observe that Δ^*_n has the same form of Δ_n by replacing S_n and $\theta_{0(j)}$ with S^*_n and $\hat{\theta}_{(j)}$, respectively, since Δ^*_n does not involve any recentering procedure as in HH. Therefore, the remainder of the proof proceeds as in the previous proof for part (a) of the Lemma. We use Lemmas 4-5 instead of Lemmas 2-3. *Q.E.D.*

A.3.7 Proof of Lemma 7

Since X_i's are iid by Assumption 1, we set $\gamma = 0$ and replace $0 \leq \xi < 1/2 - \gamma$ with $\forall c \in [0, 1/2)$ in Lemma 14 of Andrews (2002). Since Assumptions 1 and 3 of Andrews (2002) hold under our Assumptions 1 and 3, the Lemma holds by the proof of Lemma 14 of Andrews (2002). *Q.E.D.*

A.3.8 Proof of Lemma 8

By Lemma 6 for $\Delta_n = T_{MR(j)}$ and $\Delta^*_n = T^*_{MR(j)}$, it suffices to show that $n^{1/2}A(S_n)$ and $n^{1/2}A(S^*_n)$ possess Edgeworth expansions with remainder $o(n^{-a})$, where $A(\cdot)$ is an infinitely differentiable real-valued function. The function $A(\cdot)$ is normalized so that the asymptotic variances of $n^{1/2}A(S_n)$ and $n^{1/2}A(S^*_n)$ are one.[14] To see this, observe that the asymptotic variances of $n^{1/2}A(S_n)$ and $T_{MR(j)}$ are the same by Lemma 6(a), and the conditional asymptotic variances of $n^{1/2}A(S^*_n)$ and $T^*_{MR(j)}$ are the same, except if χ_n is in a sequence of sets with probability $o(n^{-a})$ by Lemma 6(b). By Theorem 1 and 2 of Hall and Inoue (2003), the asymptotic variance of $T_{MR(j)}$ is one for $j = 1, 2$. To find the conditional asymptotic variance of $T^*_{MR(j)}$, we use the proof of Theorem 2.1. of Bickel and Freedman (1981). Conditional on χ_n, where χ_n is in a sequence of sets with P probability $1 - o(n^{-a})$, the ordinary central limit theorem and the law of large numbers imply

$$\sqrt{n}(\hat{\theta}^*_{(j)} - \hat{\theta}_{(j)}) \to_d N(0, \Sigma_{MR(j)|F_n}), \tag{A.73}$$

and $\hat{\Sigma}^*_{MR(j)} \to_p \Sigma_{MR(j)|F_n}$ where $\Sigma_{MR(j)|F_n}$ is obtained by replacing the population moments by the sample moments in the formula of $\Sigma_{MR(j)}$. Then, by Slutsky's theorem, $T^*_{MR(j)}$ has the asymptotic

[14]Hall and Horowitz (1996) and Andrews (2002) do this normalization by recentering, but the procedure is implicit.

variance of one for $j = 1, 2$, conditional on χ_n, where χ_n is in a sequence of sets with P probability $1 - o(n^{-a})$.

The rest of the proof is analogous to that of Lemma 16 of Andrews (2002) which uses the results of Bhattacharya (1987) with the properly normalized $n^{1/2}A(\cdot)$ in place of his $n^{1/2}H(\cdot)$. For part (a), we apply Theorem 3.1 of Bhattacharya (1987) with his integer parameter s satisfying $(s-2)/2 = a$ for a assumed in the Lemma and with his $\bar{X} = S_n$. Conditions $(A_1) - (A_4)$ of Bhattacharya (1987) hold by Assumption 3(e), the fact that $A(\cdot)$ is infinitely differentiable and real-valued, and Assumptions 1 and 4. For part (b), the result hold by an analogous argument as for part (a), but with Theorem 3.1 of Bhattacharya (1987) replaced by Theorem 3.3 of Bhattacharya (1987) and using Lemma 7 with $c = 0$ to ensure that the coefficients $\nu^*_{n,a}$ are well behaved. *Q.E.D.*