MISSPECIFICATION-ROBUST BOOTSTRAP FOR MOMENT CONDITION MODELS

by

SeoJeong Lee

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Economics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2012

Date of final oral examination: 04/25/2012

The dissertation is approved by the following members of the Final Oral Committee:
Bruce E. Hansen, Professor of Economics, Chair
Jack R. Porter, Professor of Economics
Kenneth D. West, Professor of Economics
Xiaoxia Shi, Assistant Professor of Economics
Zhengjun Zhang, Associate Professor of Statistics

To my wife, Nary.

ACKNOWLEDGMENTS

This dissertation could not have been completed without the support of many people to whom I am indebted. First and foremost, I would like to thank my advisor Bruce Hansen for his continuous support and encouragement from the very beginning of this project. He remained confident even when I was in a quandary about the direction of my research. He not only greatly influenced my perspective on econometrics, but also has been a great mentor. I am deeply indebted to Jack Porter for his guidance and encouragement. His intuition clarified my own thinking about the result obtained. I am also very grateful to Ken West. He always wanted me to go back to fundamentals, and he taught me how to present the matter in an insightful manner. I would like to thank Xiaoxia Shi and Andrés Aradillas-López for their helpful comments. I would like to thank Zhengjun Zhang for serving as a member of the final oral committee outside of the Economics department.

Many friends and colleagues helped me to develop my research topic. In particular, I would like to thank all of the econometrics reading group members, Chu-An Liu, Enrique Pinzón-García, Jen-Che Liao, Shengjie Hong, Ying-Ying Lee, Jin Yan, Nelson Ramirez-Rondan and Jing Tao, as well as Korean friends at UW-Madison, Seok-Ki Kim, Heewoo Kang, Young Wook Lee, Jaeyoon Lee, Gihwan Yi, Yuseob Lee and Wontae Han. Comments from Mai Seki greatly improved the presentation of my research. Seunggyu Sim, Kyu-Chul Jung, Sang-Yoon Lee, and Chang-Koo Chi were my role models and mentors.

I would also like to thank my advisor at Seoul National University, Jae-Young Kim. I was first interested in econometrics when I took his class as an undergraduate student and I could successfully finish my master's degree in economics under his guidance. The topic of my research paper at Seoul National University was the starting point of this dissertation.

Finally, I would like to express my deep appreciation to my family. My parents devoted enormous effort to help me pursue my goal and finally obtain my Ph.D. degree. My most sincere thanks goes to my loving wife Nary. When we first arrived in Madison four years ago, we were just married. Since then, she has been a great

wife and a great friend getting through all the hardships together. Needless to say, I would not have been able to write this without her support. I, therefore, dedicate this dissertation to her.

CONTENTS

Contents iv

List of Tables vi

List of Figures vii

Abstract viii

1 Asymptotic Refinements of a Misspecification-Robust Bootstrap for Generalized Method of Moments Estimators 1

1.1 Introduction 1

1.2 Why We Care About Misspecification 4

1.3 Outline of the Results 9

1.4 Estimators and Test Statistics 13

1.5 The Misspecification-Robust Bootstrap Procedure 17

1.6 Main Result 19

1.7 The Recentered Bootstrap under Misspecification 22

1.8 Monte Carlo Experiments 25

1.9 Conclusion 37

1.10 Appendix: Lemmas and Proofs 37

1.11 Technical Appendix 58

2 Asymptotic Refinements of a Misspecification-Robust Bootstrap for Empirical Likelihood Estimators 69

2.1 Introduction 69

2.2 Why the Misspecification-Robust Bootstrap Works with the Empirical Likelihood Estimators 71

2.3 Estimators and Test Statistics 73

2.4 The Misspecification-Robust Bootstrap Procedure 79

2.5 Main Result 80

2.6 Monte Carlo Experiments 83
2.7 Appendix: Lemmas and Proofs 88
2.8 Technical Appendix 134

3 Bootstrapping GMM estimators under Local Misspecification 137
3.1 Asymptotic Distribution of GMM 137
3.2 Conventional Bootstrap Methods 139
3.3 Conclusion 150

References 154

LIST OF TABLES

1.1 Tables II and V of Imbens and Lancaster (1994) 6
1.2 Comparison of the Asymptotic and Bootstrap Critical Values 9
1.3 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. 29
1.4 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. 35
1.5 Table IV of Imbens and Lancaster (1994) 59

2.1 Coverage Probabilities of 90% and 95% Confidence Intervals for θ_0 under Model Misspecifications 87

3.1 Moments of the Sampling Distribution of T_n and $T_{HH,n}$ 148

LIST OF FIGURES

1.1 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) 31
1.2 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 34
1.3 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) 36

3.1 Distribution of T_n and $T_{HH,n}$ when DGP 1 151
3.2 Distribution of T_n and $T_{HH,n}$ when DGP 2 152
3.3 Coverage Probabilities of the Asymptotic and the HH Bootstrap Confidence Intervals: 153

ABSTRACT

This dissertation consists of three independent essays in econometric theory.

In the first chapter, I propose a nonparametric iid bootstrap that achieves asymptotic refinements for t tests and confidence intervals (CI's) based on the generalized method of moments (GMM) estimators even when the model is misspecified. In addition, my bootstrap does not require recentering the bootstrap moment function, which has been considered as a critical procedure for bootstrapping GMM. The elimination of the recentering combined with a robust covariance matrix renders the bootstrap robust to misspecification. Regardless of whether the assumed model is correctly specified or not, the misspecification-robust bootstrap achieves the same sharp magnitude of refinements as the conventional bootstrap methods which establish asymptotic refinements in the absence of misspecification using recentering. The key procedure is to use a misspecification-robust variance estimator for GMM in constructing the sample and the bootstrap versions of the t statistic. Two examples of overidentified and possibly misspecified moment condition models are provided: (i) Combining data sets, and (ii) invalid instrumental variables. Monte Carlo simulation results are provided as well.

In the second chapter, I propose a nonparametric iid bootstrap for the empirical likelihood (EL) estimators, including the exponentially tilted empirical likelihood estimator. My bootstrap achieves sharp asymptotic refinements for t tests and CI's regardless of whether the assumed moment condition model is correctly specified or not. This result is new, because asymptotic refinements of bootstrapping for the EL estimators have not been established in the literature even under correct model specifications. Monte Carlo simulation results are provided.

In the third chapter, I examine first-order validity and asymptotic refinements of the bootstrap methods for GMM estimators, when the moment condition model is locally misspecified. Local misspecification implies that the moment condition is misspecified for any finite sample size, but the misspecification vanishes as the sample size grows. I find that the conventional bootstrap methods are still first-order valid, but they do not achieve asymptotic refinements anymore.

1 ASYMPTOTIC REFINEMENTS OF A MISSPECIFICATION-ROBUST BOOTSTRAP FOR GENERALIZED METHOD OF MOMENTS ESTIMATORS

1.1 Introduction

This paper proposes a nonparametric iid bootstrap that achieves asymptotic refinements for t tests and confidence intervals (CI's) based on the generalized method of moments (GMM) estimators, without recentering the bootstrap moment function and without assuming correct model specification. The recentering has been considered as critical to get refinements of the bootstrap for overidentified models, but my bootstrap achieves the same refinements without recentering. In addition, the conventional bootstrap is valid only when the model is correctly specified, while I eliminate the assumption without affecting the ability of achieving asymptotic refinements of the bootstrap. 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 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, and jeopardizes 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 observe that the conditional asymptotic covariance matrix of the bootstrap GMM estimator under misspecification is different from the standard one. The conditional asymptotic covariance matrix is consistently estimable by using the result of Hall and Inoue (2003), and I construct the t statistic of which distribution is asymptotically standard normal even under misspecification.

Hall and Inoue (2003) shows that the asymptotic distributions of GMM estimators under misspecification are different from those of the standard GMM theory.[1] In particular, the asymptotic covariance matrix has additional non-zero terms in the presence of misspecification. Hall and Inoue's formulas for the asymptotic covariance matrix encompass the case of correct specification as a special case. The variance estimator using their formula is denoted by the Hall-Inoue variance estimator, hereinafter. Imbens (1997) also describes the asymptotic covariance matrices of GMM estimators robust to misspecification by using a just-identified formulation of overidentified GMM. However, his description is general, rather than being 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 in the population. The distribution of the proposed bootstrap t statistic mimics that of the sample t statistic which is asymptotically pivotal regardless of misspecification. The sample t statistic is constructed using the Hall-Inoue variance estimator. Thus, the proposed bootstrap is referred to as

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

the misspecification-robust (MR) bootstrap. In contrast, the conventional first-order asymptotics as well as the recentered bootstrap would not work under misspecification, because the conventional t statistic is not asymptotically pivotal anymore.

I note that the MR bootstrap is not for the J test. 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 imposes the null hypothesis of the J test because it eliminates the misspecification in the bootstrap world by recentering. In contrast, the MR bootstrap does not eliminate the misspecification 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 in the absence of misspecification, it is important to conduct the J test and report the result that the model is not rejected. However, even a significant J test statistic would not invalidate the estimation results if possible misspecification of the model is assumed and the validity of 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 1.2 discusses theoretical and empirical implications of misspecified models and explains the advantage of using the MR bootstrap t tests and CI's. Section 1.3 outlines the main result. Section 1.4 defines the estimators and test statistics. Section 1.5 defines the nonparametric iid MR bootstrap for iid data. Section 1.6 states the assumptions and establishes asymptotic refinements of the MR bootstrap. Section 1.7 provides a heuristic explanation of why the recentered bootstrap does not work under misspecification. Section 3.1 presents examples and Monte Carlo simulation results. Section 1.9 concludes the paper. Section 1.10 contains Lemmas and proofs. A detailed calculations of the examples in the following Sections are in Section 1.11.

1.2 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 Imbens and Lancaster (1994), Jondeau et al. (2004), Parker and Julliard (2005), 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.[2]

A significant J statistic means that the test rejects the null hypothesis of correct model specification. For 2SLS estimators, this implies that at least one of the instruments is invalid. 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 et al. (2007), Marmer and Otsu (2009), 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, inferences about the parameters should take into account a possible misspecification in the model. Otherwise, such inferences would be misleading.

For the maximum likelihood (ML) estimators, White (1982) provides a theory of the quasi-maximum likelihood when the assumed probability distribution is misspecified, which includes the standard ML theory as a special case. For GMM, Hall and Inoue (2003) describes the asymptotic distribution of GMM estimators

[2] In the 2SLS framework, the Sargan test is often reported, which is a special case of the J test.

	Intercept θ_0	Edu θ_1	$Age-35$ θ_2	$(Age-35)^2$ θ_3	J test $\chi^2(5)$
ML	1.44^*	$-.009$	$-.002$	$-.002$	-
	(.317)	(.093)	(.015)	(.002)	
GMM	1.86^*	$-.109$	$-.003$	$-.003^*$	11.4
	(.268)	(.084)	(.002)	(.0003)	[.044]

Note: Standard errors in parentheses. p-value in bracket.
$*$: significant at 1% level

Table 1.1: Tables II and V of Imbens and Lancaster (1994)

under misspecification. In particular, Hall and Inoue's asymptotic covariance matrix encompasses the standard GMM covariance matrix in the absence of misspecification as a special case, under the situations considered in this paper.

Example: Combining Micro and Macro Data

Imbens and Lancaster (1994) suggests an econometric procedure that uses nearly exact information on the marginal distribution of economic variables to improve accuracy of estimation. As an application, the authors estimate the following probit model for employment: For an individual i,

$$\begin{aligned} P(L_i = 1|Age_i, Edu_i) &= \Phi(\mathbf{x}_i'\theta) \\ &= \Phi(\theta_0 + \theta_1 \cdot Edu_i + \theta_2 \cdot (Age_i - 35) + \theta_3 \cdot (Age_i - 35)^2), \end{aligned} \tag{1.1}$$

with $\mathbf{x}_i = (1, Edu_i, Age_i - 35, (Age_i - 35)^2)'$ and $\Phi(\cdot)$ is the standard normal cdf. L_i is labor market status ($L_i = 1$ when employed), Edu_i is education level in five categories, and Age_i is age in years. The sample is a micro data set on Dutch labor market histories and the number of observations is 347. Typically, the probit model is estimated by the ML estimator. The first row of Table 1.1 presents the ML point estimates and the standard errors. None of the coefficients are statistically significant except for that of the intercept.

To reduce the standard errors of the estimators, the authors use additional information on the population from the national statistic. By using the statistical yearbooks for the Netherlands which contain 2.355 million observations, they calculated the probability of being employed given the age category (denoted by p_k where the index for the age category $k = 1, 2, 3, 4, 5$) and the probability of being in a particular age category (denoted by q_k). These probabilities are considered as the true population parameters.

The authors suggest to use GMM estimators with the moment function that utilizes the information from the aggregate statistic. The second row of Table 1.1 reports the two-step efficient GMM point estimates and the standard errors. Now the coefficient θ_3 is statistically significant at 1% level and the authors argue:

> *...The standard deviation on the coefficients θ_2 and θ_3, which capture the age-dependency of the employment probability decrease by a factor 7...Age is not ancillary anymore and knowledge about its marginal distribution is informative about θ.*

Although they could successfully improve the accuracy of the estimators by combining two data sets, their argument has a potential problem. The last column of Table 1.1 reports the J test statistic and its p-value. Since the p-value is 4.4%, the model is marginally rejected at 5% level. The problem is that, if the model is truly misspecified, the reported GMM standard errors are inconsistent because the conventional standard errors are only consistent under correct model specification. Then the authors' argument about the coefficient estimates may be flawed. This problem could be avoided if the standard errors which are consistent even under misspecification were used. The formulas for the misspecification-robust standard errors for the GMM estimators are available in Section 4.[3]

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 and Ω^{-1} be a weight matrix. According to Hall and Inoue (2003), (i) the probability

[3]Since the original data sets used in Imbens and Lancaster (1994) are not available, I could not calculate the robust standard errors. Instead, I provide a supporting simulation result with a simple hypothetical model that utilizes additional population information in estimation in Section 1.8

limit of $\hat{\theta}$ is the pseudo-true value that depends on the weight matrix such that

$$\theta_0(\Omega^{-1}) = \arg\min_{\theta} Eg(X_i, \theta)'\Omega^{-1}Eg(X_i, \theta), \tag{1.2}$$

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

$$\sqrt{n}(\hat{\theta} - \theta_0(\Omega^{-1})) \to_d N(0, \Sigma_{MR}), \tag{1.3}$$

where Σ_{MR} is the asymptotic covariance matrix under misspecification that is different from Σ_C, the asymptotic covariance 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, if any, 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), White (1982), 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 standard error of the GMM estimator as honest as possible. (1.3) implies that the conventional t tests and CI's are invalid under misspecification, because the conventional standard errors are based on the estimate of Σ_C. Misspecification-robust standard errors are calculated using the estimate of Σ_{MR}.

Unless one has a complete confidence on the model specification, the robust Hall-Inoue variance estimators for GMM should be seriously considered. By using the robust variance estimators, the resulting asymptotic t tests and CI's are robust to misspecification. The MR bootstrap t tests and CI's improve upon these misspecification-robust asymptotic t tests and CI's in terms of the magnitude of errors in test rejection probability and CI coverage probability. A summary on the advantage of the MR bootstrap over the existing asymptotic and bootstrap t tests

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

†: 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 1.2: Comparison of the Asymptotic and Bootstrap Critical Values

and CI's is given in Table 1.2.

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, $J_n(\theta)$. Suppose that θ is a scalar for notational brevity. 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.[4] 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 (1997), under regularity conditions, $H_n(z, F)$ and $H^*_{n_b}(z, F_n)$ allow Edgeworth expansion of the form

$$\begin{aligned} H_n(z,F) &= H_\infty(z,F) + n^{-1/2}q_1(z,F) + n^{-1}q_2(z,F) + o(n^{-1}), && (1.4)\\ H^*_{n_b}(z,F_n) &= H^*_\infty(z,F_n) + n_b^{-1/2}q_1(z,F_n) + n_b^{-1}q_2(z,F_n) + o_p(n_b^{-1}) && (1.5) \end{aligned}$$

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

$$P(|T(\chi_n)| > z_{\alpha/2}) = \alpha + O(n^{-1}), \quad P(\theta_0 \in CI) = 1 - \alpha + O(n^{-1}), \qquad (1.6)$$

where $CI = [\hat{\theta} \pm z_{\alpha/2}\sqrt{\hat{\Sigma}/n}]$. In other words, the error in the rejection probability

[4] 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.

and coverage probability of the asymptotic two-sided t test and CI is $O(n^{-1})$.

For the bootstrap t test and CI, subtract (1.4) from (1.5), use the fact that q_1 is even, and set $n_b = n$ to show, under H_0,

$$P(|T(\chi_n)| > z^*_{|T|,\alpha}) = \alpha + o(n^{-1}), \quad P(\theta_0 \in CI^*) = 1 - \alpha + o(n^{-1}) \tag{1.7}$$

where $CI^* = [\hat{\theta} \pm z^*_{|T|,\alpha}\sqrt{\hat{\Sigma}/n}]$. The elimination of the leading terms in (1.4) and (1.5) is the source of asymptotic refinements of bootstrapping the asymptotically pivotal statistics (Beran (1988); Hall (1997)).

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,[5] $T_C(\chi^*_{n_b}) = (\hat{\theta}^* - \hat{\theta})/\sqrt{\hat{\Sigma}^*_C/n_b}$, 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 GMM variance 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

[5] 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.

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.

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 1.7.

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,

$$P(|T_{MR}(\chi_n)| > z^*_{|T_{MR}|,\alpha}) = \alpha + O(n^{-2}), \quad P(\theta_0 \in CI^*_{MR}) = 1 - \alpha + O(n^{-2}). \quad (1.8)$$

This rate is sharp. The further reduction in the error from $o(n^{-1})$ of (1.7) to $O(n^{-2})$

of (1.8) 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.

1.4 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 $g : \Theta \to \mathbb{R}^{L_g}$ 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,

$$J(\theta, \Omega^{-1}) = Eg(X_i, \theta)'\Omega^{-1}Eg(X_i, \theta), \tag{1.9}$$

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 covariance matrix of the two-step GMM estimator cannot be simplified to the efficient one under correct specification.

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

$$\min_{\theta\in\Theta} J_n(\theta, I_{L_g}) = \left(n^{-1}\sum_{i=1}^{n} g(X_i,\theta)\right)'\left(n^{-1}\sum_{i=1}^{n} g(X_i,\theta)\right). \tag{1.10}$$

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

$$\min_{\theta\in\Theta} J_n(\theta, W_n(\hat{\theta}_{(1)})) \equiv \left(n^{-1}\sum_{i=1}^{n} g(X_i,\theta)\right)' W_n(\hat{\theta}_{(1)})\left(n^{-1}\sum_{i=1}^{n} g(X_i,\theta)\right), \tag{1.11}$$

where[6]

$$W_n(\theta) = \left(n^{-1}\sum_{i=1}^{n}(g(X_i,\theta)-g_n(\theta))(g(X_i,\theta)-g_n(\theta))'\right)^{-1}, \tag{1.12}$$

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 $G(X_i,\theta) = (\partial/\partial\theta')g(X_i,\theta)$,

$$G_{0(j)} = EG(X_i,\theta_{0(j)}), \quad G_{0(j)}^{(2)} = E\left[\frac{\partial}{\partial\theta'} vec\left\{G(X_i,\theta_{0(j)})\right\}\right], \tag{1.13}$$

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

$$G_n(\theta) = n^{-1}\sum_{i=1}^{n} G(X_i,\theta), \quad G_n^{(2)}(\theta) = n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\theta'} vec\left\{G(X_i,\theta)\right\}, \tag{1.14}$$

[6] 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.

$G_{n(j)} = G_n(\hat{\theta}_{(j)})$, 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

$$\sqrt{n}\begin{pmatrix} (g_n(\theta_{0(1)}) - g_{0(1)}) \\ (G_n(\theta_{0(1)}) - G_{0(1)})'g_{0(1)} \end{pmatrix} \to_d N\left(\mathbf{0}, \underset{(L_g+L_\theta)\times(L_g+L_\theta)}{\Omega_1}\right), \tag{1.15}$$

and

$$\sqrt{n}\begin{pmatrix} (g_n(\theta_{0(2)}) - g_{0(2)}) \\ (G_n(\theta_{0(2)}) - G_{0(2)})'Wg_{0(2)} \\ (W_n - W)g_{0(2)} \end{pmatrix} \to_d N\left(\mathbf{0}, \underset{(2L_g+L_\theta)\times(2L_g+L_\theta)}{\Omega_2}\right). \tag{1.16}$$

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

$$\sqrt{n}(\hat{\theta}_{(j)} - \theta_{0(j)}) \to_d N(0, \Sigma_{MR(j)}), \tag{1.17}$$

where $\Sigma_{MR(j)} = H^{-1}_{0(j)}V_jH^{-1}_{0(j)}$, for $j = 1, 2$,

$$\begin{aligned} V_1 &= \begin{bmatrix} G'_{0(1)} & I_{L_\theta} \end{bmatrix} \Omega_1 \begin{bmatrix} G'_{0(1)} & I_{L_\theta} \end{bmatrix}', \\ V_2 &= \begin{bmatrix} G'_{0(2)}W & I_{L_\theta} & G'_{0(2)} \end{bmatrix} \Omega_2 \begin{bmatrix} G'_{0(2)}W & I_{L_\theta} & G'_{0(2)} \end{bmatrix}'. \end{aligned} \tag{1.18}$$

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

$$\Sigma_{C(1)} = (G'_0G_0)^{-1}G'_0\Omega_CG_0(G'_0G_0)^{-1}, \quad \Sigma_{C(2)} = (G'_0\Omega_C^{-1}G_0)^{-1}, \tag{1.19}$$

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

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

$$\begin{aligned} V_{n(1)} &= \begin{bmatrix} G'_{n(1)} & I_{L_\theta} \end{bmatrix} \Omega_{n(1)} \begin{bmatrix} G'_{n(1)} & I_{L_\theta} \end{bmatrix}', \\ V_{n(2)} &= \begin{bmatrix} G'_{n(2)}W_n & I_{L_\theta} & G'_{n(2)} \end{bmatrix} \Omega_{n(2)} \begin{bmatrix} G'_{n(2)}W_n & I_{L_\theta} & G'_{n(2)} \end{bmatrix}', \end{aligned} \tag{1.20}$$

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

$$\begin{aligned}
\Omega_{n(1)} &= n^{-1}\sum_{i=1}^{n}\begin{pmatrix} g(X_i,\hat{\theta}_{(1)}) - g_{n(1)} \\ (G(X_i,\hat{\theta}_{(1)}) - G_{n(1)})' g_{n(1)} \end{pmatrix}\begin{pmatrix} g(X_i,\hat{\theta}_{(1)}) - g_{n(1)} \\ (G(X_i,\hat{\theta}_{(1)}) - G_{n(1)})' g_{n(1)} \end{pmatrix}', \quad (1.21) \\
\Omega_{n(2)} &= n^{-1}\sum_{i=1}^{n}\begin{pmatrix} g(X_i,\hat{\theta}_{(2)}) - g_{n(2)} \\ (G(X_i,\hat{\theta}_{(2)}) - G_{n(2)})' W_n g_{n(2)} \\ W_i g_{n(2)} \end{pmatrix}\begin{pmatrix} g(X_i,\hat{\theta}_{(2)}) - g_{n(2)} \\ (G(X_i,\hat{\theta}_{(2)}) - G_{n(2)})' W_n g_{n(2)} \\ W_i g_{n(2)} \end{pmatrix}',
\end{aligned}$$

where[7]

$$W_i = -W_n \cdot \left((g(X_i,\hat{\theta}_{(1)}) - g_n(\hat{\theta}_{(1)}))(g(X_i,\hat{\theta}_{(1)}) - g_n(\hat{\theta}_{(1)}))' - W_n^{-1}\right) \cdot W_n. \quad (1.22)$$

The diagonal elements of the covariance estimator $\hat{\Sigma}_{MR(j)}$ for $j = 1, 2$ are the Hall-Inoue variance estimators. In practice, the estimation of the misspecification-robust covariance 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, $\theta_{0(j)}$, 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

$$T_{MR(j)} = \frac{\hat{\theta}_{(j),k} - \theta_{0(j),k}}{\sqrt{(\hat{\Sigma}_{MR(j)})_{kk}/n}}, \quad (1.23)$$

where $j = 1$ for the one-step GMM estimator and $j = 2$ for the two-step GMM estimator.[8] $T_{MR(j)}$ 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, because $\hat{\Sigma}_{C(j)} \neq \hat{\Sigma}_{MR(j)}$ in general even under correct specification, for $j = 1, 2$. Note that $\hat{\Sigma}_{C(j)}$ is a consistent estimator for $\Sigma_{C(j)}$, the asymptotic covariance matrix under correct specification for $j = 1, 2$.

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

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

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 $j=1,2$.

1.5 The Misspecification-Robust Bootstrap Procedure

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

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

$$\min_{\theta\in\Theta} J_n^*(\theta, I_{L_g}) = \left(n^{-1}\sum_{i=1}^n g(X_i^*,\theta)\right)'\left(n^{-1}\sum_{i=1}^n g(X_i^*,\theta)\right), \tag{1.24}$$

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

$$\min_{\theta\in\Theta} J_n^*(\theta, W_n^*(\hat{\theta}^*_{(1)})) = \left(n^{-1}\sum_{i=1}^n g(X_i^*,\theta)\right)' W_n^*(\hat{\theta}^*_{(1)})\left(n^{-1}\sum_{i=1}^n g(X_i^*,\theta)\right), \tag{1.25}$$

where

$$W_n^*(\theta) = \left(n^{-1}\sum_{i=1}^n (g(X_i^*,\theta) - g_n^*(\theta))(g(X_i^*,\theta) - g_n^*(\theta))'\right)^{-1}, \tag{1.26}$$

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

$$G_n^*(\theta) = n^{-1}\sum_{i=1}^n \frac{\partial}{\partial\theta'}g(X_i^*,\theta), \quad G_n^{(2)*}(\theta) = n^{-1}\sum_{i=1}^n \frac{\partial}{\partial\theta'}vec\left\{\frac{\partial}{\partial\theta'}g(X_i^*,\theta)\right\}, \quad (1.27)$$

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

The bootstrap version of the robust covariance matrix estimator $\hat{\Sigma}_{MR(j)}$ is $\hat{\Sigma}_{MR(j)}^* = H_{n(j)}^{*-1}V_{n(j)}^* H_{n(j)}^{*-1\prime}$ for $j=1,2$, where

$$\begin{aligned} V_{n(1)}^* &= \left[\begin{array}{cc} G_{n(1)}^{*\prime} & I_{L_g}\end{array}\right]\Omega_{n(1)}^*\left[\begin{array}{cc} G_{n(1)}^{*\prime} & I_{L_g}\end{array}\right]', \\ V_{n(2)}^* &= \left[\begin{array}{ccc} G_{n(2)}^{*\prime}W_n^* & I_{L_g} & G_{n(2)}^{*\prime}\end{array}\right]\Omega_{n(2)}^*\left[\begin{array}{ccc} G_{n(2)}^{*\prime}W_n^* & I_{L_g} & G_{n(2)}^{*\prime}\end{array}\right]', \end{aligned} \quad (1.28)$$

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

$$\begin{aligned} \Omega_{n(1)}^* &= n^{-1}\sum_{i=1}^n \left(\begin{array}{c} g(X_i^*,\hat{\theta}_{(1)}^*) - g_{n(1)}^* \\ (G(X_i^*,\hat{\theta}_{(1)}^*) - G_{n(1)}^*)'g_{n(1)}^*\end{array}\right)\left(\begin{array}{c} g(X_i^*,\hat{\theta}_{(1)}^*) - g_{n(1)}^* \\ (G(X_i^*,\hat{\theta}_{(1)}^*) - G_{n(1)}^*)'g_{n(1)}^*\end{array}\right)', \quad (1.29) \\ \Omega_{n(2)}^* &= n^{-1}\sum_{i=1}^n \left(\begin{array}{c} g(X_i^*,\hat{\theta}_{(2)}^*) - g_{n(2)}^* \\ (G(X_i^*,\hat{\theta}_{(2)}^*) - G_{n(2)}^*)'W_n^*g_{n(2)}^* \\ W_i^*g_{n(2)}^*\end{array}\right)\left(\begin{array}{c} g(X_i^*,\hat{\theta}_{(2)}^*) - g_{n(2)}^* \\ (G(X_i^*,\hat{\theta}_{(2)}^*) - G_{n(2)}^*)'W_n^*g_{n(2)}^* \\ W_i^*g_{n(2)}^*\end{array}\right)', \end{aligned}$$

where

$$W_i^* = -W_n^*\cdot\left((g(X_i^*,\hat{\theta}_{(1)}^*) - g_n^*(\hat{\theta}_{(1)}^*))(g(X_i^*,\hat{\theta}_{(1)}^*) - g_n^*(\hat{\theta}_{(1)}^*))' - W_n^{*-1}\right)\cdot W_n^*. \quad (1.30)$$

The MR bootstrap t statistic is

$$T_{MR(j)}^* = \frac{\hat{\theta}_{(j),k}^* - \hat{\theta}_{(j),k}}{\sqrt{(\hat{\Sigma}_{MR(j)}^*)_{kk}/n}}, \quad (1.31)$$

for $j = 1, 2$.[9] 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 $j = 1, 2$,

$$CI^*_{MR(j)} = \left[\hat{\theta}_{(j),k} \pm z^*_{|T_{MR(j)}|,\alpha}\sqrt{(\hat{\Sigma}_{MR(j)})_{kk}/n}\right]. \tag{1.32}$$

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 robust covariance matrix estimator, $\hat{\Sigma}^*_{MR(j)}$, is used to construct the bootstrap t statistic. In the recentered bootstrap, the conventional covariance matrix estimator of Hansen (1982) is used.

1.6 Main Result

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

[9] $T^*_{MR(j)} \equiv T_{MR(j)}(\chi^*_n)$. I suppress the dependence of $T^*_{MR(j)}$ on χ^*_n for notational brevity.

$(\partial^m/\partial\theta^m)g(X_i,\theta)$ 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 1.1. *$X_i, i = 1, 2, ...$ are iid.*

Assumption 1.2. *(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 1.3. *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 1.4. *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 1.1 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 1.2-1.3 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 1.2-1.3 are not primitive, but they lead to simpler results as in Andrews (2002). Assumption 1.4 is the standard Cramér condition for iid sample, that is needed to get Edgeworth expansions.

Asymptotic Refinements of the Misspecification-Robust Bootstrap

Theorem 1.1 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 ofHall 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 1.1. *Suppose Assumptions 1.1-1.4 hold. Under* $H_0 : \theta_k = \theta_{0(j),k}$*, for* $j = 1, 2$,

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

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 1.1. 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 (1997). 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.

1.7 The Recentered Bootstrap under Misspecification

In this section, I discuss about the validity of the recentered bootstrap under misspecification. Let θ be a scalar for notational brevidy. 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 of the Hall-Horowitz bootstrap GMM estimator, $\Sigma_{R(j)|F_n}$. By Glivenko-Cantelli theorem, $F_n(z)$ converges to $F(z)$ uniformly in $z \in \mathbf{R}$, and thus, $\Sigma_{R(j)|F_n} \to_p \Sigma_{R(j)}$ almost surely, where $\Sigma_{R(j)}$ is the (unconditional) asymptotic variance of the distribution of $\sqrt{n}(\hat{\theta}^*_{R(j)} - \hat{\theta}_{(j)})$. The formulas are given by

$$
\begin{aligned}
\Sigma_{R(1)} &= (G'_{0(1)}G_{0(1)})^{-1}G'_{0(1)}\Omega_{R(1)}G_{0(1)}(G'_{0(1)}G_{0(1)})^{-1}, \qquad (1.33)\\
\Sigma_{R(2)} &= (G'_{0(2)}W_RG_{0(2)})^{-1}G'_{0(2)}W_R\Omega_{R(2)}W_RG_{0(2)}(G'_{0(2)}W_RG_{0(2)})^{-1},\\
\Omega_{R(1)} &= E(g(X_i,\theta_{0(1)}) - g_{0(1)})(g(X_i,\theta_{0(1)}) - g_{0(1)})',\\
\Omega_{R(2)} &= E(g(X_i,\theta_{0(2)}) - g_{0(2)})(g(X_i,\theta_{0(2)}) - g_{0(2)})',\\
W_R &= \left[E(g(X_i,\theta_{0(1)}) - g_{0(2)})(g(X_i,\theta_{0(1)}) - g_{0(2)})'\right]^{-1}.
\end{aligned}
$$

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 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.

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,

$$\begin{aligned}
\Sigma^*_{CR(1)} &= (G^{*\prime}_{n(1)} G^*_{n(1)})^{-1} G^{*\prime}_{n(1)} \Omega^*_{R,n(1)} G^*_{n(1)} (G^{*\prime}_{n(1)} G^*_{n(1)})^{-1}, \qquad (1.34)\\
\Sigma^*_{CR(2)} &= (G^{*\prime}_{n(2)} \Omega^{*-1}_{R,n(2)} G^*_{n(2)})^{-1},\\
\Omega^*_{R,n(1)} &= n^{-1} \sum_{i=1}^{n} (g(X_i^*, \hat{\theta}^*_{(1)}) - g_n(\hat{\theta}_{(1)}))(g(X_i^*, \hat{\theta}^*_{(1)}) - g_n(\hat{\theta}_{(1)}))',\\
\Omega^*_{R,n(2)} &= n^{-1} \sum_{i=1}^{n} (g(X_i^*, \hat{\theta}^*_{(2)}) - g_n(\hat{\theta}_{(2)}))(g(X_i^*, \hat{\theta}^*_{(2)}) - g_n(\hat{\theta}_{(2)}))'.
\end{aligned}$$

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,

$$\begin{aligned}
\Sigma_{CR(1)} &= (G'_{0(1)} G_{0(1)})^{-1} G'_{0(1)} \Omega_{R(1)} G_{0(1)} (G'_{0(1)} G_{0(1)})^{-1} = \Sigma_{R(1)}, \qquad (1.35)\\
\Sigma_{CR(2)} &= (G'_{0(2)} \Omega^{-1}_{R(2)} G_{0(2)})^{-1} \neq \Sigma_{R(2)}.
\end{aligned}$$

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 et al. (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.

1.8 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 numbers 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

$$CI_C = \left[\hat{\theta} \pm z_{\alpha/2}\sqrt{\hat{\Sigma}_C/n}\right], \tag{1.36}$$

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

$$CI_{MR} = \left[\hat{\theta} \pm z_{\alpha/2}\sqrt{\hat{\Sigma}_{MR}/n}\right]. \tag{1.37}$$

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

$$CI^*_{HH} = \left[\hat{\theta} \pm z^*_{|T_{HH}|,\alpha}\sqrt{\hat{\Sigma}_C/n}\right], \tag{1.38}$$

$$CI^*_{BN} = \left[\hat{\theta} \pm z^*_{|T_{BN}|,\alpha}\sqrt{\hat{\Sigma}_C/n}\right], \tag{1.39}$$

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:

$$CI^*_{MR} = \left[\hat{\theta} \pm z^*_{|T_{MR}|,\alpha}\sqrt{\hat{\Sigma}_{MR}/n}\right], \tag{1.40}$$

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 1.1.

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

$$g(X_i, \theta) = \begin{pmatrix} Y_i \\ Z_i - \theta \end{pmatrix}. \tag{1.41}$$

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 (1.10) is given by $\hat{\theta}_{(1)} = \bar{Z}$. The two-step GMM estimator solving (1.11) and the pseudo-true value are given by

$$\hat{\theta}_{(2)} = \bar{Z} - \frac{\widehat{Cov}(Y_i, Z_i)}{\widehat{Var}(Y_i)}\bar{Y} \rightarrow_p \theta_{0(2)} = EZ_i - \frac{Cov(Y_i, Z_i)}{Var(Y_i)}EY_i, \tag{1.42}$$

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.[10]

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.

[10] The pseudo-true value may equal to the true value regardless of misspecification. Schennach (2007) provides an example that the pseudo-true value is invariant to misspecification, and thus, is the same with the true value.

Let the true data generating process (DGP) be

$$\begin{pmatrix} Y_i \\ Z_i \end{pmatrix} \sim N\left(\begin{pmatrix} \delta \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \tag{1.43}$$

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 variances $\Sigma_{C(2)}$ and $\Sigma_{MR(2)}$ are[11]

$$\Sigma_{C(2)} = 1 - \rho^2, \qquad \Sigma_{MR(2)} = (1 - \rho^2)(1 + \delta^2). \tag{1.44}$$

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

$$\frac{\Sigma_{MR(2)}}{\Sigma_{C(2)}} = 1 + \delta^2 \to \infty \quad \text{as } \delta \to \infty. \tag{1.45}$$

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.

[11] A detailed calculation is in the technical appendix.

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 1.3: 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.

Table 1.3 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.

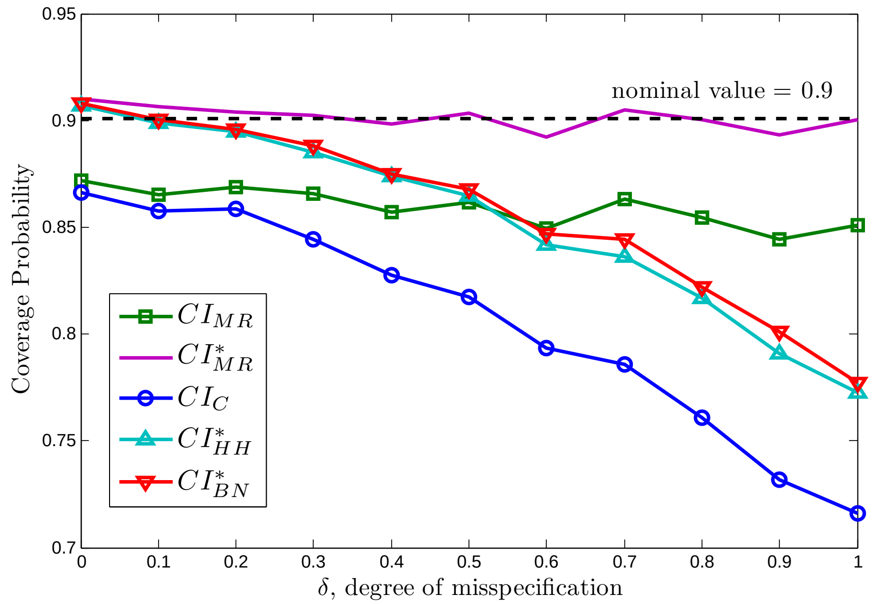


Figure 1.1: Coverage Probabilities of 90% Confidence Intervals for $\theta_{0(2)}$ based on the Two-step GMM Estimator, $\hat{\theta}_{(2)}$, when $\rho = 0.5$ and $n = 25$ in Example 1 (r=5,000, B=1,000)

For different values of δ, Figure 1.1 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.

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

$$\hat{\beta}_{IV_k} = (\sum_{i=1}^{n} z_{ki}x_i)^{-1}\sum_{i=1}^{n} z_{ki}y_i \to_p \beta_{IV_k} = \beta_0 + (Ez_{ki}x_i)^{-1}Ez_{ki}e_i, \tag{1.46}$$

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

$$g(X_i, \beta) = \begin{pmatrix} z_{1i}(y_i - x_i\beta) \\ z_{2i}(y_i - x_i\beta) \end{pmatrix}, \tag{1.47}$$

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 (1.11) by using W_n as a weight matrix instead of using the identity matrix.[12] 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 for $\Sigma_{MR(2)}$, the asymptotic variance for the two-step GMM estimator described in Section 1.4, 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,[13] and that the calculation of the asymptotic variance is very complicated under misspecification.

Let the DGP be

$$\begin{aligned} &y_i = x_i\beta_0 + e_i; \quad x_i = z_{1i}\gamma_1 + z_{2i}\gamma_2 + e_i + \varepsilon_i, \quad z_{2i} = z_{2i}^0 + 0.5\delta e_i + u_i; \\ &(z_{1i}, z_{2i}^0)' \sim N(\mathbf{0}, I_2), \quad e_i \sim N(0,2), \quad \varepsilon_i \sim N(0,1), \quad u_i \sim N(0,1), \end{aligned} \tag{1.48}$$

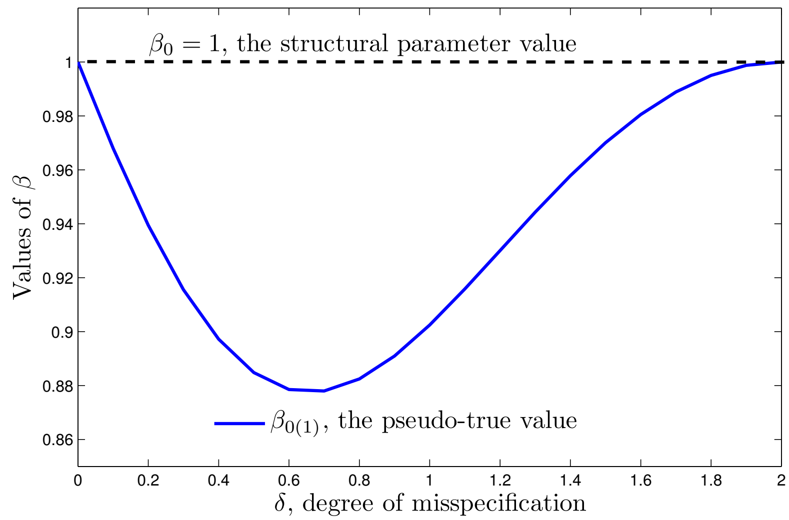
[12] A detailed calculation of $\hat{\beta}_{(1)}$ and its probability limit is in the technical appendix.

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

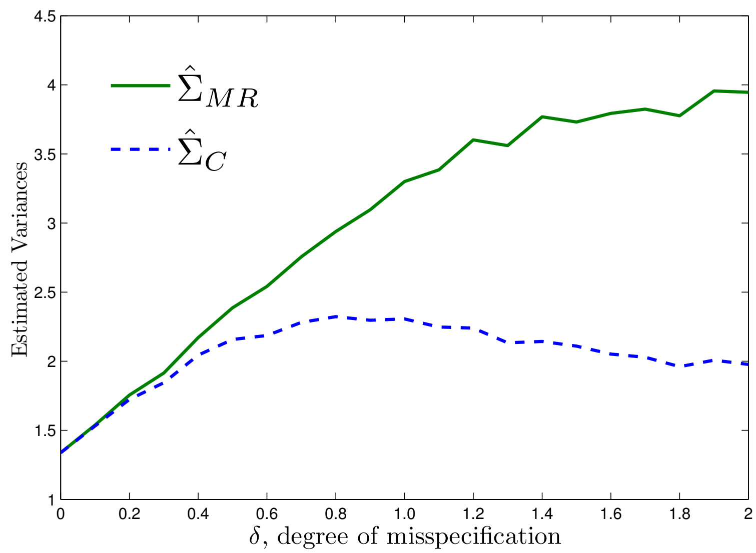
where I_2 is a 2×2 identity matrix and $(z_{1i}, z_{2i}^0)'$, e_i, ε_i, and u_i are iid. This DGP satisfies $Ex_ie_i \neq 0$, $Ez_{1i}e_i = 0$, and $Ez_{2i}e_i = \delta$, where δ measures a degree of misspecification. Therefore, the instrument z_{1i} is valid, while z_{2i} may not. The probability limit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

$$\beta_{0(1)} = \beta_0 + \frac{(2+0.5\delta^2)\gamma_2 + \delta}{\gamma_1^2(2+0.5\delta^2) + ((2+0.5\delta^2)\gamma_2 + \delta)^2} \cdot \delta = \beta_0 + O(\delta^{-1}). \tag{1.49}$$

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 1.2(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 (1.49), 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 1.2(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 $\delta = 2$.



(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 1.2: 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 1.4: 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.

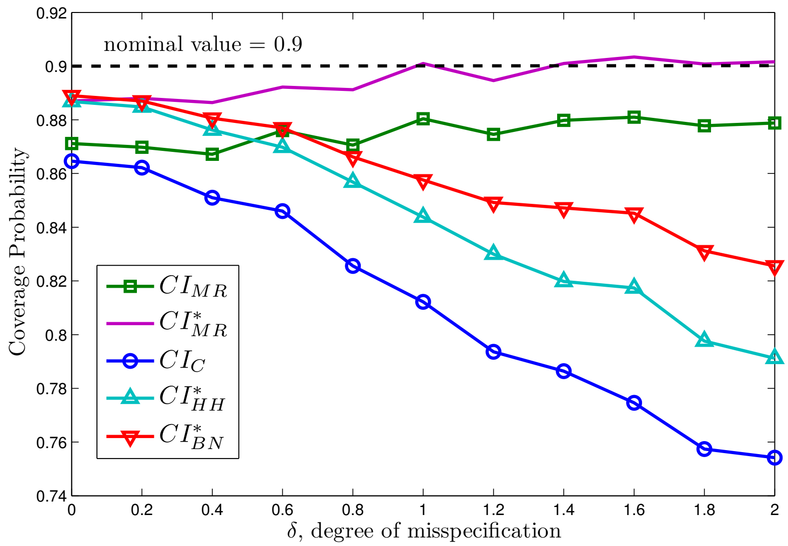


Figure 1.3: Coverage Probabilities of 90% Confidence Intervals for $\beta_{0(1)}$ based on the One-step GMM Estimator, $\hat{\beta}_{(1)}$, $n = 50$ in Example 2 (r=5,000, B=1,000)

Table 1.4 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, 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 1.3 shows the coverage probabilities of the CI's over different degrees of misspecification. Again, the ability of achieving asymptotic refinements of the bootstrap 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.

1.9 Conclusion

This paper gives an alternative bootstrap procedure for GMM that achieves a sharp rate of asymptotic refinements regardless of misspecification. The existing bootstrap procedures for GMM achieve the same rate of asymptotic refinements only for correctly specified models 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 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.

1.10 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.

Lemmas

Lemma 1.2 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 of Andrews (2002) are denoted by AL1, AL2, AL6, and AL7, respectively. In addition, Lemma 5 of Andrews (2002) is denoted by AL5 without modification.

Lemma 1.2.

(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 1.3-1.4 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 1.3. *Suppose Assumptions 1-3 hold. Then, for all* $c \in [0, 1/2)$ *and all* $a \geq 0$*,*

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

Lemma 1.4. *Suppose Assumptions 1-3 hold. Then, for all* $c \in [0, 1/2)$ *and all* $a \geq 0$*,*

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

Lemmas 1.5-1.6 are the bootstrap versions of Lemmas 1.3-1.4, 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 1.5. *Suppose Assumptions 1-3 hold. Then, for all* $c \in [0, 1/2)$ *and all* $a \geq 0$*,*

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

Lemma 1.6. *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 1.7. *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 1.8. *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 1.9. *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,

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

Proof of Theorem 1.1

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 1.7 and 1.9. 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 1.7 and 1.9. 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.*

Proofs of Lemmas

Proof of Lemma 1.2

(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.*

Proof of Lemma 1.3

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,

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

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

$$\lim_{n\to\infty} n^a P(\|\hat{\theta}_{(1)} - \theta_{0(1)}\| > \varepsilon) = 0. \tag{1.50}$$

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

$$\hat{\theta}_{(1)} - \theta_{0(1)} = -\left(\frac{\partial^2}{\partial\theta\partial\theta'} J_n(\tilde{\theta})\right)^{-1} \frac{\partial}{\partial\theta} J_n(\theta_{0(1)}) \tag{1.51}$$

where

$$\frac{\partial}{\partial\theta} J_n(\theta_{0(1)}) = \left\{G'_{0(1)}(g_n(\theta_{0(1)}) - g_{0(1)}) + (G_n(\theta_{0(1)}) - G_{0(1)})' g_n(\theta_{0(1)})\right\}, \tag{1.52}$$

$$\frac{\partial^2}{\partial\theta\partial\theta'} J_n(\theta) \equiv 2\tilde{H}_n(\theta, I_{L_g}) = 2\left\{(g_n(\theta)' \otimes I_{L_\theta})\, G_n^{(2)}(\theta) + G_n(\theta)' G_n(\theta)\right\}, \tag{1.53}$$

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:

$$\lim_{n\to\infty} n^a P\left(\left\|\tilde{H}_n(\tilde{\theta}, I_{L_g}) - \tilde{H}_n(\theta_{0(1)}, I_{L_g})\right\| > \varepsilon\right) = 0, \tag{1.54}$$

$$\lim_{n\to\infty} n^a P\left(\left\|\tilde{H}_n(\theta_{0(1)}, I_{L_g}) - H_{0(1)}\right\| > \varepsilon\right) = 0, \tag{1.55}$$

$$\lim_{n\to\infty} n^a P\left(\left\|G_n(\theta_{0(1)}) - G_{0(1)}\right\| > n^{-c}\right) = 0, \tag{1.56}$$

$$\lim_{n\to\infty} n^a P\left(\left\|g_n(\theta_{0(1)}) - g_{0(1)}\right\| > n^{-c}\right) = 0. \tag{1.57}$$

To show (1.54), we apply the triangle and Cauchy-Schwarz inequalities multiple times,

$$\begin{aligned}
& \left\| \left(g_n(\tilde{\theta})' \otimes I_{L_\theta}\right) G_n^{(2)}(\tilde{\theta}) - \left(g_n(\theta_{0(1)})' \otimes I_{L_\theta}\right) G_n^{(2)}(\theta_{0(1)}) \right. \\
& \left. + G_n(\tilde{\theta})' G_n(\tilde{\theta}) - G_n(\theta_{0(1)})' G_n(\theta_{0(1)}) \right\| \\
\leq\ & \|G_n^{(2)}(\tilde{\theta}) - G_n^{(2)}(\theta_{0(1)})\| \left(\|g_n(\tilde{\theta}) - g_n(\theta_{0(1)})\| + \|g_n(\theta_{0(1)})\|\right) \\
& + \|G_n^{(2)}(\theta_{0(1)})\| \|g_n(\tilde{\theta}) - g_n(\theta_{0(1)})\| \\
& + \|G_n(\tilde{\theta}) - G_n(\theta_{0(1)})\| \left(\|G_n(\tilde{\theta}) - G_n(\theta_{0(1)})\| + 2\|G_n(\theta_{0(1)})\|\right) \\
\leq\ & \|\tilde{\theta} - \theta_{0(1)}\| \left\{ C_{\partial f,n}(C_{g,n} + C_{\partial f,n}) \|\tilde{\theta} - \theta_{0(1)}\| \right. \\
& \left. + C_{g,n} \|G_n^{(2)}(\theta_{0(1)})\| + 2\|G_n(\theta_{0(1)})\| + \|g_n(\theta_{0(1)})\| \right\},
\end{aligned} \tag{1.58}$$

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 (1.50) 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 (1.54).

For (1.55), apply the triangle and Cauchy-Schwarz inequalities to get

$$\begin{aligned}
& \|(g_n(\theta_{0(1)})' \otimes I_{L_\theta}) G_n^{(2)}(\theta_{0(1)}) - (g_{0(1)}' \otimes I_{L_\theta}) G_{0(1)}^{(2)}\| \\
\leq\ & \|G_n^{(2)}(\theta_{0(1)}) - G_{0(1)}^{(2)}\| \cdot \|g_n(\theta_{0(1)})\| + \|G_{0(1)}^{(2)}\| \cdot \|g_n(\theta_{0(1)}) - g_{0(1)}\|,
\end{aligned} \tag{1.59}$$

and

$$\begin{aligned}
& \|G_n(\theta_{0(1)})' G_n(\theta_{0(1)}) - G_{0(1)}' G_{0(1)}\| \\
\leq\ & \|G_n(\theta_{0(1)}) - G_{0(1)}\| \cdot (\|G_n(\theta_{0(1)}) - G_{0(1)}\| + 2\|G_{0(1)}\|).
\end{aligned}$$

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 (1.56) 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 (1.57) 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.*

Proof of Lemma 1.4

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 1.3 show that

$$\lim_{n\to\infty} n^a P(\|\hat{\theta}_{(2)} - \theta_{0(2)}\| > \varepsilon) = 0. \tag{1.60}$$

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

$$\hat{\theta}_{(2)} - \theta_{0(2)} \;=\; -\left(\frac{\partial^2}{\partial\theta\partial\theta'}J_n(\tilde{\theta},W_n)\right)^{-1}\frac{\partial}{\partial\theta}J_n(\theta_{0(2)},W_n), \tag{1.61}$$

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

$$\begin{aligned}
\frac{\partial}{\partial\theta}J_n(\theta_{0(2)},W_n) &= G_n(\theta_{0(2)})'W_n(g_n(\theta_{0(2)}) - g_{0(2)}) & (1.62)\\
&\quad + \left(G_n(\theta_{0(2)}) - G_{0(2)}\right)' W g_{0(2)} + G_n(\theta_{0(2)})'(W_n - W)g_{0(2)}, \\
\frac{\partial^2}{\partial\theta\partial\theta'}J_n(\theta,W_n) &= 2\tilde{H}_n(\theta,W_n) & (1.63)\\
&= 2\left\{(g_n(\theta)'W_n \otimes I_{L_\theta})\, G_n^{(2)}(\theta) + G_n(\theta)'W_nG_n(\theta)\right\},
\end{aligned}$$

and $\tilde{\theta}$ is between $\hat{\theta}_{(2)}$ and $\theta_{0(2)}$ and may differ across rows. Note that (1.62) includes additional terms that are zero under correct specification. Thus, in order to show

$$\lim_{n\to\infty} n^a P\left(\left\|\frac{\partial}{\partial\theta}J_n(\theta_{0(2)},W_n)\right\| > n^{-c}\right) = 0, \tag{1.64}$$

we need

$$\begin{aligned}
&\lim_{n\to\infty} n^a P\left(\left\|g_n(\theta_{0(2)}) - g_{0(2)}\right\| > n^{-c}\right) = 0, & (1.65)\\
&\lim_{n\to\infty} n^a P\left(\left\|G_n(\theta_{0(2)}) - G_{0(2)}\right\| > n^{-c}\right) = 0, & (1.66)\\
&\lim_{n\to\infty} n^a P(\|W_n(\hat{\theta}_{(1)}) - W\| > n^{-c}) = 0. & (1.67)
\end{aligned}$$

Note that (1.66) and (1.67) are required for possibly misspecified models.[14]

(1.65) and (1.66) hold by AL1(a) with $h(X_i) = g_i(\theta_{0(2)}) - g_{0(2)}$ or $h(X_i) = G_i(\theta_{0(2)}) - G_{0(2)}$. (1.67) follows from

$$\lim_{n\to\infty} n^a P(\|W_n(\hat{\theta}_{(1)})^{-1} - W_n(\theta_{0(1)})^{-1}\| > n^{-c}) = 0, \text{ and} \tag{1.68}$$
$$\lim_{n\to\infty} n^a P(\|W_n(\theta_{0(1)})^{-1} - W^{-1}\| > n^{-c}) = 0. \tag{1.69}$$

To show (1.68), observe that

$$\begin{aligned} & \|W_n(\hat{\theta}_{(1)})^{-1} - W_n(\theta_{0(1)})^{-1}\| \\ = \; & \|n^{-1}\sum_{i=1}^{n}(g_i(\hat{\theta}_{(1)})g_i(\hat{\theta}_{(1)})' - g_i(,\theta_{0(1)})g_i(\theta_{0(1)})')\| + \|g_n(\theta_{0(1)})g_n(\theta_{0(1)})' - g_{n(1)}g_{n(1)}'\|. \end{aligned} \tag{1.70}$$

For the first term of the right-hand side of (1.70), we apply the mean value expansion and the Cauchy-Schwarz inequality to get

$$\begin{aligned} & \|n^{-1}\sum_{i=1}^{n}(g_i(\hat{\theta}_{(1)})g_i(\hat{\theta}_{(1)})' - g_i(\theta_{0(1)})g_i(\theta_{0(1)})')\| \\ \leq \; & 2n^{-1}\sum_{i=1}^{n}\sup_{\theta\in N_{0(1)}}\|G_i(\theta)\|\|g_i(\theta)\| \cdot \|\hat{\theta}_{(1)} - \theta_{0(1)}\|. \end{aligned} \tag{1.71}$$

For the second term of (1.70), we apply the Cauchy-Schwarz inequality,

$$\begin{aligned} & \|g_n(\theta_{0(1)})g_n(\theta_{0(1)})' - g_{n(1)}g_{n(1)}'\| \\ = \; & \|(g_n(\theta_{0(1)}) - g_n(\hat{\theta}_{(1)}))(g_n(\theta_{0(1)}) + g_n(\hat{\theta}_{(1)}))'\| \\ \leq \; & n^{-1}\sum_{i=1}^{n}\|g_i(\theta_{0(1)}) - g_i(\hat{\theta}_{(1)})\|n^{-1}\sum_{i=1}^{n}\|g_i(\theta_{0(1)}) + g_i(\hat{\theta}_{(1)})\| \\ \leq \; & \|\hat{\theta}_{(1)} - \theta_{0(1)}\|C_{g,n}(2n^{-1}\sum_{i=1}^{n}\|g_i(\theta_{0(1)})\| + \|\hat{\theta}_{(1)} - \theta_{0(1)}\|C_{g,n}). \end{aligned} \tag{1.72}$$

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 1.3 proves (1.68).

[14] Andrews (2002) proves (1.67) by replacing n^{-c} with ε under correct specification.

(1.69) 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

$$\begin{aligned} \|W_n(\theta_{0(1)})^{-1} - W^{-1}\| &\leq \left\| n^{-1}\sum_{i=1}^{n} g_i(\theta_{0(1)})g_i(\theta_{0(1)})' - Eg_i(\theta_{0(1)})g_i(\theta_{0(1)})' \right\| \\ &\quad + \Big(2\|g_{0(1)}\| + \|g_n(\theta_{0(1)}) - g_{0(1)}\|\Big)\, \|g_n(\theta_{0(1)}) - g_{0(1)}\|. \end{aligned} \tag{1.73}$$

Lastly, the Lemma follows from

$$\lim_{n\to\infty} n^a P\left(\left\|\tilde{H}_n(\tilde{\theta}, W_n) - \tilde{H}_n(\theta_{0(2)}, W)\right\| > \varepsilon\right) = 0, \tag{1.74}$$

$$\lim_{n\to\infty} n^a P\left(\left\|\tilde{H}_n(\theta_{0(2)}, W) - H_{0(2)}\right\| > \varepsilon\right) = 0, \tag{1.75}$$

that can be shown by multiple applications of AL1 and the results (1.67) and (1.60). *Q.E.D.*

Proof of Lemma 1.5

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 1.3 and write

$$\begin{aligned} J_n(\theta) - J_n(\hat{\theta}_{(1)}) &= J(\theta) - J(\theta_{0(1)}) + J_n(\theta) - J_n(\hat{\theta}_{(1)}) \\ &\quad - J(\theta) + J(\hat{\theta}_{(1)}) + J(\theta_{0(1)}) - J(\hat{\theta}_{(1)}) \\ &\geq J(\theta) - J(\theta_{0(1)}) - |J_n(\theta) - J_n(\hat{\theta}_{(1)}) - J(\theta) + J(\hat{\theta}_{(1)})| \\ &\quad - |J(\hat{\theta}_{(1)}) - J(\theta_{0(1)})|. \end{aligned} \tag{1.76}$$

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 1.3. Taking $\lambda < M/2$ proves the claim.

Thus, we have

$$\begin{aligned} & n^a P(P^*(\|\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)}\| > \varepsilon) > n^{-a}) \\ \leq\ & n^a P(P^*(J_n(\hat{\theta}^*_{(1)}) - J_n^*(\hat{\theta}^*_{(1)}) + J_n^*(\hat{\theta}^*_{(1)}) - J_n(\hat{\theta}_{(1)}) > \delta) > n^{-a}) \\ \leq\ & n^a P(P^*(J_n(\hat{\theta}^*_{(1)}) - J_n^*(\hat{\theta}^*_{(1)}) + J_n^*(\hat{\theta}_{(1)}) - J_n(\hat{\theta}_{(1)}) > \delta) > n^{-a}) \\ \leq\ & n^a P\left(P^*\left(\sup_{\theta\in\Theta} |J_n^*(\theta) - J_n(\theta)| > \delta/2\right) > n^{-a}\right) \to 0, \end{aligned} \tag{1.77}$$

since $\hat{\theta}^*_{(1)}$ is the minimizer of $J_n^*(\theta)$. To verify the last conclusion of (1.77), we apply the triangle and Cauchy-Schwarz inequalities,

$$\begin{aligned} |J_n^*(\theta) - J_n(\theta)| &= |g_n^*(\theta)' g_n^*(\theta) - g_n(\theta)' g_n(\theta)| \\ &\leq \|g_n^*(\theta) - g_n(\theta)\|^2 \\ &\quad +2\left(\|g_n(\theta) - Eg(X_i,\theta)\| + \|Eg(X_i,\theta)\|\right)\|g_n^*(\theta) - g_n(\theta)\| \\ &= \|g_n^*(\theta) - E^* g_i^*(\theta)\|^2 \\ &\quad +2\left(\|g_n(\theta) - Eg(X_i,\theta)\| + \|Eg(X_i,\theta)\|\right)\|g_n^*(\theta) - E^* g_i^*(\theta)\|, \end{aligned} \tag{1.78}$$

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,

$$\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)} = -\left(\frac{\partial^2}{\partial\theta\partial\theta'} J_n^*(\tilde{\theta}^*)\right)^{-1} \frac{\partial}{\partial\theta} J_n^*(\hat{\theta}_{(1)}), \tag{1.79}$$

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 1.3 with some modifications for the bootstrap version.

First, we prove

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|\frac{\partial}{\partial\theta}J_n^*(\hat\theta_{(1)})\right\| > n^{-c}\right) > n^{-a}\right) = 0, \tag{1.80}$$

where

$$\frac{\partial}{\partial\theta}J_n^*(\hat\theta_{(1)}) = G_n(\hat\theta_{(1)})'\left(g_n^*(\hat\theta_{(1)}) - g_n(\hat\theta_{(1)})\right) + \left(G_n^*(\hat\theta_{(1)}) - G_n(\hat\theta_{(1)})\right)' g_n^*(\hat\theta_{(1)}), \tag{1.81}$$

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,

$$\lim_{n\to\infty} n^a P(P^*(\|G_n(\hat\theta_{(1)})\| > \varepsilon) > n^{-a}), \tag{1.82}$$
$$\lim_{n\to\infty} n^a P(P^*(\|g_n^*(\hat\theta_{(1)})\| > \varepsilon) > n^{-a}), \tag{1.83}$$
$$\lim_{n\to\infty} n^a P(P^*(\|g_n^*(\hat\theta_{(1)}) - g_n(\hat\theta_{(1)})\| > n^{-c}) > n^{-a}), \tag{1.84}$$
$$\lim_{n\to\infty} n^a P(P^*(\|G_n^*(\hat\theta_{(1)}) - G_n(\hat\theta_{(1)})\| > n^{-c}) > n^{-a}). \tag{1.85}$$

For (1.82), note that $\|G_n(\hat\theta_{(1)})\| \le \|G_n(\theta_{0(1)})\| + \|G_n(\hat\theta_{(1)}) - G_n(\theta_{0(1)})\|$ holds by the triangle inequality and claim

$$\lim_{n\to\infty} n^a P(P^*(\|G_n(\theta_{0(1)})\| > \varepsilon) > n^{-a}) = 0, \tag{1.86}$$
$$\lim_{n\to\infty} n^a P(P^*(\|G_n(\hat\theta_{(1)}) - G_n(\theta_{0(1)})\| > \varepsilon) > n^{-a}) = 0. \tag{1.87}$$

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,

$$\begin{aligned} & n^a P(P^*(\|G_n(\theta_{0(1)})\| > \varepsilon) > n^{-a}) \qquad (1.88)\\ = \; & n^a P(1\{\|G_n(\theta_{0(1)})\| > \varepsilon\} > n^{-a}, \|G_n(\theta_{0(1)})\| > \varepsilon) \\ & + n^a P(1\{\|G_n(\theta_{0(1)})\| > \varepsilon\} > n^{-a}, \|G_n(\theta_{0(1)})\| \le \varepsilon) \\ \le \; & n^a P(\|G_n(\theta_{0(1)})\| > \varepsilon) \to 0, \end{aligned}$$

by AL1(b). (1.87) can be shown similarly by applying AL1(a). By (1.86) and (1.87), the first result (1.82) is proved. To show the second result (1.83), apply the triangle inequality and Assumption 2 to get

$$\begin{aligned} \|g_n^*(\hat{\theta}_{(1)})\| &\leq \|g_n^*(\theta_{0(1)})\| + \|g_n^*(\theta_{0(1)}) - g_n^*(\hat{\theta}_{(1)})\| \\ &\leq \|g_n^*(\theta_{0(1)})\| + C_{g,n}^*\|\hat{\theta}_{(1)} - \theta_{0(1)}\|, \end{aligned} \tag{1.89}$$

where $C_{g,n}^* = n^{-1}\sum_{i=1}^n C_g(X_i^*)$. By applying AL6(d) and Lemma 1.3, we have the result (1.83). For the third and the last result, we apply the triangle inequality and Assumptions 2-3,

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

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 1.3 and Assumption 3. This proves the result (1.84) and (1.85).

Next, we claim

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|\tilde{H}_n^*(\tilde{\theta}^*, I_{L_g}) - \tilde{H}_n^*(\theta_{0(1)}, I_{L_g})\right\| > \varepsilon\right) > n^{-a}\right) = 0, \tag{1.90}$$

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|\tilde{H}_n^*(\theta_{0(1)}, I_{L_g}) - H_{0(1)}\right\| > \varepsilon\right) > n^{-a}\right) = 0, \tag{1.91}$$

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 1.3 prove (1.90) and (1.91) 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 1.3 and (1.77). By combining (1.80), (1.90), and (1.91), the Lemma follows. *Q.E.D.*

Proof of Lemma 1.6

We first show that

$$\lim_{n\to\infty} n^a P(P^*(\|W_n^*(\hat{\theta}^*_{(1)}) - W\| > n^{-c}) > n^{-a}) = 0. \quad (1.92)$$

This follows from

$$\lim_{n\to\infty} n^a P(P^*(\|W_n^*(\hat{\theta}^*_{(1)})^{-1} - W_n^*(\theta_{0(1)})^{-1}\| > n^{-c}) > n^{-a}) = 0, \quad (1.93)$$
$$\lim_{n\to\infty} n^a P(P^*(\|W_n^*(\theta_{0(1)})^{-1} - W^{-1}\| > n^{-c}) > n^{-a}) = 0. \quad (1.94)$$

To obtain (1.93), we use the same argument as that in the proof of Lemma 1.4 and the triangle inequality to show

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

where

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

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 1.3 and 1.5 to get (1.93). The proof of (1.94) is analogous to that of (1.69) 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 1.3 and Lemma 1.5 with (1.92) show that

$$\lim_{n\to\infty} n^a P(P^*(\|\hat{\theta}^*_{(2)} - \hat{\theta}_{(2)}\| > \varepsilon) > n^{-a}) = 0. \quad (1.95)$$

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,

$$\hat{\theta}^*_{(2)} - \hat{\theta}_{(2)} = -\left(\frac{\partial^2}{\partial\theta\partial\theta'} J^*_n(\tilde{\theta}^*, W^*_n)\right)^{-1} \frac{\partial}{\partial\theta} J^*_n(\hat{\theta}_{(2)}, W^*_n), \tag{1.96}$$

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

$$\begin{aligned} \frac{\partial}{\partial\theta} J^*_n(\hat{\theta}_{(2)}, W^*_n) &= G^*_n(\hat{\theta}_{(2)})' W^*_n g^*_n(\hat{\theta}_{(2)}) \quad (1.97) \\ &= G^*_n(\hat{\theta}_{(2)})' W^*_n \left(g^*_n(\hat{\theta}_{(2)}) - g_n(\hat{\theta}_{(2)})\right) + \left(G^*_n(\hat{\theta}_{(2)}) - G_n(\hat{\theta}_{(2)})\right)' W g_n(\hat{\theta}_{(2)}) \\ &\quad + G^*_n(\hat{\theta}_{(2)})'(W^*_n - W) g_n(\hat{\theta}_{(2)}) + G_n(\hat{\theta}_{(2)})'(W - W_n) g_n(\hat{\theta}_{(2)}), \end{aligned}$$

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,

$$\begin{aligned} \|W^*_n\| &\leq \|W\| + \|W^*_n - W\|, \quad (1.98) \\ \|G^*_n(\hat{\theta}_{(2)})\| &\leq \|G^*_n(\theta_{0(2)})\| + \|G^*_n(\hat{\theta}_{(2)}) - G^*_n(\theta_{0(2)})\| \quad (1.99) \\ &\leq \|G^*_n(\theta_{0(2)})\| + C^*_{\partial f,n} \|\hat{\theta}_{(2)} - \theta_{0(2)}\|, \\ \|g^*_n(\hat{\theta}_{(2)}) - g_n(\hat{\theta}_{(2)})\| &\leq \|g^*_n(\theta_{0(2)}) - g_n(\theta_{0(2)})\| \quad (1.100) \\ &\quad + \|g^*_n(\hat{\theta}_{(2)}) - g^*_n(\theta_{0(2)})\| + \|g_n(\hat{\theta}_{(2)}) - g_n(\theta_{0(2)})\| \\ &\leq \|g^*_n(\theta_{0(2)}) - g_n(\theta_{0(2)})\| + \|\hat{\theta}_{(2)} - \theta_{0(2)}\|(C^*_{g,n} + C_{g,n}). \end{aligned}$$

We apply AL6(a) with $h(X_i) = g_i(\theta_{0(2)})$, AL6(d), Lemma 1.4, and (1.95) to show that

$$\lim_{n\to\infty} n^a P(P^* \left\| G^*_n(\hat{\theta}_{(2)})' W^*_n \left(g^*_n(\hat{\theta}_{(2)}) - g_n(\hat{\theta}_{(2)})\right)\right\| > n^{-c}) > n^{-a}) = 0. \tag{1.101}$$

Similar arguments apply to the remaining terms and we conclude that

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|\frac{\partial}{\partial\theta} J^*_n(\hat{\theta}_{(2)}, W^*_n)\right\| > n^{-c}\right) > n^{-a}\right) = 0. \tag{1.102}$$

Now, the Lemma follows from

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|\tilde{H}_n^*(\tilde{\theta}^*, W_n^*) - \tilde{H}_n^*(\theta_{0(2)}, W)\right\| > \varepsilon\right) > n^{-a}\right) = 0, \quad (1.103)$$
$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|\tilde{H}_n^*(\theta_{0(2)}, W) - H_{0(2)}\right\| > \varepsilon\right) > n^{-a}\right) = 0, \quad (1.104)$$

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 1.5, 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 1.4 and (1.95). *Q.E.D.*

Proof of Lemma 1.7

(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

$$0 = \frac{\partial J_n(\theta_{0(1)})}{\partial\theta} + \frac{\partial^2 J_n(\theta_{0(1)})}{\partial\theta\partial\theta'}\delta_n + \frac{1}{2}\frac{\partial^3 J_n(\theta_{0(1)})}{\partial\theta\partial\theta_i\partial\theta_j}\delta_{ni}\delta_{nj} \quad (1.105)$$
$$+ \cdots + \frac{1}{(d_1-1)!}\frac{\partial^{d_1} J_n(\theta_{0(1)})}{\partial\theta\partial\theta_i\cdots\partial\theta_\kappa}\delta_{ni}\cdots\delta_{n\kappa} + \zeta_n, \quad (1.106)$$

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

$$\zeta_n = \frac{1}{(d_1-1)!}\left(\frac{\partial^{d_1} J_n(\bar{\theta}_n)}{\partial\theta\partial\theta_i\cdots\partial\theta_\kappa} - \frac{\partial^{d_1} J_n(\theta_{0(1)})}{\partial\theta\partial\theta_i\cdots\partial\theta_\kappa}\right)\delta_{ni}\cdots\delta_{n\kappa}, \quad (1.107)$$

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, (1.106) 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 (1.106) 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

$$\hat{\theta}_{(1)} - \theta_{0(1)} \equiv \delta_n = \Lambda_1(R_n + e_n). \tag{1.108}$$

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 1.3 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 1.3 that $\|e_n\| \leq n^{-d_1 c}$ with probability $1 - o(n^{-a})$. Therefore, by the mean value theorem for some $\tilde{M} < \infty$,

$$n^a P\left(\|(\hat{\theta}_{(1)} - \theta_{0(1)}) - \Lambda_1(R_n)\| > n^{-d_1 c}\right) \leq n^a P\left(\tilde{M}\|e_n\| > n^{-d_1 c}\right) = o(1), \tag{1.109}$$

as $n \to \infty$. In order to apply AL5(a) with $\xi_n = n^{1/2}\zeta_n$, we need $d_1 c \geq a + 1/2$ 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 (1.109) and AL5(a),

$$\lim_{n\to\infty} \sup_z n^a \left|P\left(n^{1/2}(\hat{\theta}_{(1)} - \theta_{0(1)}) \leq z\right) - P\left(n^{1/2}\Lambda_1(R_n) \leq z\right)\right| = 0. \tag{1.110}$$

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, $\partial_1 J_n(\hat{\theta}_{(2)}, \hat{\theta}_{(1)})/\partial\theta = 0$ 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 $\partial_1 J_n(\hat{\theta}_{(2)}, \hat{\theta}_{(1)})/\partial\theta$ through order d_1 about $(\theta, \tilde{\theta}) = (\theta_{0(2)}, \theta_{0(1)})$ [15] that with probability $1-o(n^{-a})$

$$0 = \frac{\partial_1 J_n(\theta_{0(2)}, \theta_{0(1)})}{\partial\theta} + Q_n^2 \eta_n + \frac{1}{2} Q_n^3 \eta_{ni}\eta_{nj} + \cdots + \frac{1}{(d_1-1)!} Q_n^{d_1} \eta_{ni}\eta_{nj}...\eta_{n\kappa} + \nu_n, \quad (1.111)$$

where $N\{i,j,...,\kappa\} = d_1 - 1$, Q_n^m is the mth order derivative of $\partial_1 J_n(\cdot,\cdot)/\partial\theta$ 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 (1.111) and its inverse exists and is bounded with probability $1-o(n^{-a})$ by AL1. Using arguments similar to those used in proving (2.53), we apply the implicit function theorem to obtain

$$\hat{\theta}_{(2)} - \theta_{0(2)} = \Lambda_2\left(S_n, \nu_n, \Lambda_1(R_n + e_n)\right) \quad (1.112)$$

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 1.3 and Lemma 1.4, $\|\eta_n\| < n^{-c}$ and thus, $\|\nu_n\| < n^{-d_1 c}$ with probability $1-o(n^{-a})$. By the triangle inequality and the mean value theorem,

$$\begin{aligned} & \|\Lambda_2(S_n, \nu_n, \Lambda_1(R_n + e_n)) - \Lambda_2(S_n, 0, 0)\| \quad (1.113) \\ \leq\ & \|\Lambda_2(S_n, \nu_n, \Lambda_1(R_n + e_n)) - \Lambda_2(S_n, 0, \Lambda_1(R_n + e_n))\| \\ & + \|\Lambda_2(S_n, 0, \Lambda_1(R_n + e_n)) - \Lambda_2(S_n, 0, \Lambda_1(R_n))\| + \|\Lambda_2(S_n, 0, \Lambda_1(R_n)) - \Lambda_2(S_n, 0, 0)\| \\ \leq\ & M_1\|\nu_n\| + M_2\|e_n\| + M_3\|R_n - R\| \end{aligned}$$

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

[15] 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.

$o(1)$ and by AL5,

$$\lim_{n\to\infty}\sup_z n^a \left|P\left(n^{1/2}(\hat{\theta}_{(2)} - \theta_{0(2)}) \le z\right) - P\left(n^{1/2}\Lambda_2(S_n, 0, 0) \le z\right)\right| = 0. \qquad (1.114)$$

For $T_{MR(j)}$, we use the fact that the covariance 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 1.3 and 1.4, 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

$$0 = \frac{\partial J_n^*(\hat{\theta}_{(1)})}{\partial\theta} + \frac{\partial^2 J_n(\hat{\theta}_{(1)})}{\partial\theta\partial\theta'}\delta_n^* + \cdots + \frac{1}{(d_1-1)!}\frac{\partial^{d_1} J_n^*(\hat{\theta}_{(1)})}{\partial\theta\partial\theta_i...\partial\theta_\kappa}\delta_{ni}^*...\delta_{n\kappa}^* + \zeta_n^*, \qquad (1.115)$$

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 (2.53). Since all the terms in the expansion are the same with (1.106) by replacing R_n and $\theta_{0(1)}$ with R_n^* and $\hat{\theta}_{(1)}$, we can write

$$\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)} \equiv \delta^*_n = \Lambda_1(R^*_n + e^*_n) \tag{1.116}$$

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 1.5 yields $\lim_{n\to\infty} n^a P\left(P^*\left(\|e^*_n\| > n^{-d_1 c}\right) > n^{-a}\right) = 0$ and thus,

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|(\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)}) - \Lambda_1(R^*_n)\| > n^{-d_1 c}\right) > n^{-a}\right) = 0. \tag{1.117}$$

By AL5(b),

$$\lim_{n\to\infty} n^a P\left(\sup_z \left| P^*(n^{1/2}(\hat{\theta}^*_{(1)} - \hat{\theta}_{(1)}) \le z) - P^*(n^{1/2}\Lambda_1(R^*_n) \le z) \right| > n^{-a}\right) = 0. \tag{1.118}$$

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 1.5-1.6 instead of Lemmas 1.3-1.4. *Q.E.D.*

Proof of Lemma 1.8

Since X_i's are iid by Assumption 1, we set $\gamma = 0$ and replace $0 \le \xi < 1/2 - \gamma$ 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.*

Proof of Lemma 1.9

By Lemma 1.7 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.[16] To see this, observe that the asymptotic variances of $n^{1/2}A(S_n)$ and $T_{MR(j)}$ are the same by Lemma 1.7(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 1.7(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

$$\sqrt{n}(\hat{\theta}^*_{(j)} - \hat{\theta}_{(j)}) \to_d N(0, \Sigma_{MR(j)|F_n}), \tag{1.119}$$

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 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 1.8 with $c = 0$ to ensure that the coefficients $\nu^*_{n,a}$ are well

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

behaved. *Q.E.D.*

1.11 Technical Appendix

1. Example: Imbens and Lancaster (1994)

The probit model is

$$\begin{aligned} P(L_i = 1|Age_i, Edu_i) &= \Phi(\mathbf{x}_i'\theta) \\ &= \Phi(\theta_0 + \theta_1 \cdot Edu_i + \theta_2 \cdot (Age_i - 35) + \theta_3 \cdot (Age_i - 35)^2), \end{aligned}$$

with $\mathbf{x}_i = (1, Edu_i, Age_i - 35, (Age_i - 35)^2)'$ and $\Phi(\cdot)$ is the standard normal cdf. L_i is labor market status ($L_i = 1$ when employed), Edu_i is education level in five categories, and Age_i is age in years. Typically, the probit model is estimated by the maximum likelihood (ML) estimator. The log-likelihood function is

$$\log(\theta) = \sum_{i=1}^{n} L_i \log \Phi(\mathbf{x}_i'\theta) + (1 - L_i)\log(1 - \Phi(\mathbf{x}_i'\theta)),$$

with the first-order condition (FOC)

$$0 = \sum_{i=1}^{n} \frac{L_i - \Phi(\mathbf{x}_i'\hat{\theta}_{ML})}{\Phi(\mathbf{x}_i'\hat{\theta}_{ML})(1 - \Phi(\mathbf{x}_i'\hat{\theta}_{ML}))} \phi(\mathbf{x}_i'\hat{\theta}_{ML}) \cdot x_{ij},$$

where $\phi(\cdot)$ is the standard normal pdf and x_{ij} is the jth element of $\mathbf{x}_i$.

By using the statistical yearbooks for The Netherlands which contains 2.355 million observations, they calculated the probability of being employed given the age category (denoted by p_k where the index for the age category $k = 1, 2, 3, 4, 5$) and the probability of being in a particular age category (denoted by q_k). These probabilities are shown in Table 1.5 and are considered as the true population parameters.

The moment function is $g(L_i, \mathbf{x}_i, \theta) = (g_1(L_i, \mathbf{x}_i, \theta)', g_2(\mathbf{x}_i, \theta))'$, with

$$\begin{aligned} g_1(L_i, \mathbf{x}_i, \theta) &= \frac{L_i - \Phi(\mathbf{x}_i'\theta)}{\Phi(\mathbf{x}_i'\theta)(1 - \Phi(\mathbf{x}_i'\theta))}\phi(\mathbf{x}_i'\theta) \cdot x_{ij}, \quad \text{for } j = 1, 2, 3, 4, \\ g_2(\mathbf{x}_i, \theta) &= 1(Age_i \in C_k) \cdot (p_k - \Phi(\mathbf{x}_i'\theta)), \quad \text{for } k = 1, 2, 3, 4, 5. \end{aligned}$$

The first four moment restrictions $g_1(L_i, \mathbf{x}_i, \theta)$ are from the FOC of the ML estimator. The last five moment restrictions $g_2(\mathbf{x}_i, \theta)$ are from the true conditional employment probability.

Age category (C_k) Age	C_1 25-29	C_2 30-34	C_3 35-39	C_4 40-44	C_5 45-49
p_k	0.911	0.933	0.932	0.932	0.891
q_k	0.258	0.227	0.185	0.168	0.160

Table 1.5: Table IV of Imbens and Lancaster (1994)

2. Example 1: Combining Data Sets

The GMM estimators

The FOCs for the one-step and the two-step GMM estimators are $G_n(\hat{\theta}_{(1)})'g_n(\hat{\theta}_{(1)}) = 0$ and $G_n(\hat{\theta}_{(2)})'W_n(\hat{\theta}_{(1)})g_n(\hat{\theta}_{(1)}) = 0$, where $G_n(\theta) = (0, -1)'$, $g_n(\theta) = (\bar{Y}, \bar{Z} - \theta)'$, and

$$W_n(\theta)^{-1} = \begin{pmatrix} \widehat{Var}(Y_i) & \widehat{Cov}(Y_i, Z_i) \\ \widehat{Cov}(Y_i, Z_i) & \widehat{Var}(Z_i) \end{pmatrix}.$$

The one-step GMM estimator $\hat{\theta}_{(1)} = \bar{Z}$ and the two-step GMM estimator is

$$\hat{\theta}_{(2)} = \bar{Z} - \frac{\widehat{Cov}(Y_i, Z_i)}{\widehat{Var}(Y_i)}\bar{Y}.$$

The probability limit of $\hat{\theta}_{(2)}$ is the pseudo-true value,

$$\theta_{0(2)} = EZ_i - \frac{Cov(Y_i, Z_i)}{Var(Y_i)} EY_i = -\rho\delta.$$

Hall and Horowitz (1996), Andrews (2002), and Brown and Newey (2002) propose the two-step GMM estimator that uses a uncentered weight matrix, and thus their estimator is different from $\hat{\theta}_{(2)}$. Nevertheless, the two estimators are asymptotically equivalent under correct specification and the recentered bootstrap achieves asymptotic refinements over the CI based on $\hat{\theta}_{(2)}$.

Under misspecification, using different weight matrix implies the resulting pseudo-true values are different. Thus, the recentered bootstrap theory should be taylored to use the centered weight matrix in the sample and in the bootstrap. However, this modification is suggested by neither Hall and Horowitz (1996) nor Brown and Newey (2002), and further investigation of this modification is not the objective of this paper. Therefore, I just follow the original recipe of the recentered bootstrap, except that I use the centered weight matrix for the sample two-step GMM estimator. In the Monte Carlo simulation not reported here, I found that the result are similar when using the uncentered weight matrix in the sample for the recentered bootstrap.

Asymptotic variance under correct specification, $\Sigma_{C(2)}$

To find $\Sigma_{C(2)}$, a relevant question is: If a researcher uses the conventional variance estimator $\hat{\Sigma}_{C(2)}$ based on the above GMM estimator, what is she estimating under misspecification? To answer the question, $\Sigma_{C(2)}$ should be interpreted as the probability limit of $\hat{\Sigma}_{C(2)}$, evaluated at the pseudo-true value $\theta_{0(2)}$, which is the probability limit of the GMM estimator.

First, we calculate $\Sigma_{C(2)} = (G_0'\Omega_C^{-1}G_0)^{-1}$. Observe that $G_0 = (0\ -1)'$ and

$$\begin{aligned}
\Omega_C &= Eg(X_i,\theta_{0(2)})g(X_i,\theta_{0(2)})' \\
&= \begin{pmatrix} Var(Y_i) + (EY_i)^2 & Cov(Y_i,Z_i) + EY_i(EZ_i - \theta_{0(2)}) \\ Cov(Y_i,Z_i) + EY_i(EZ_i - \theta_{0(2)}) & Var(Z_i) + (EZ_i - \theta_{0(2)})^2 \end{pmatrix}. \\
&= \begin{pmatrix} 1+\delta^2 & \rho(1+\delta^2) \\ \rho(1+\delta^2) & 1+\rho^2\delta^2 \end{pmatrix}.
\end{aligned}$$

The inverse matrix is given by

$$\Omega_C^{-1} = \frac{1}{D_C}\begin{pmatrix} Var(Z_i) + (EZ_i - \theta_{0(2)})^2 & -Cov(Y_i,Z_i) - EY_i(EZ_i - \theta_{0(2)}) \\ -Cov(Y_i,Z_i) - EY_i(EZ_i - \theta_{0(2)}) & Var(Y_i) + (EY_i)^2 \end{pmatrix},$$

where

$$\begin{aligned}
D_C &= (Var(Y_i) + (EY_i)^2)(Var(Z_i) + (EZ_i - \theta_{0(2)})^2) \\
&\quad -(Cov(Y_i,Z_i) + EY_i(EZ_i - \theta_{0(2)}))^2 \\
&= (1+\delta^2)(1-\rho^2).
\end{aligned}$$

Thus,

$$\Sigma_{C(2)} = (G_0'\Omega_C^{-1}G_0)^{-1} = \frac{D_C}{Var(Y_i) + (EY_i)^2} = 1 - \rho^2.$$

Asymptotic variance robust to misspecification, $\Sigma_{MR(2)}$

Observe that $g_n(\theta_{0(2)}) - g_{0(2)} = (\bar{Y} - EY_i, \bar{Z} - EZ_i)'$ and $(G_n(\theta_{0(2)}) - G_{0(2)})'Wg_{0(2)} = 0$. Let

$$W \equiv \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix},$$

where $w_{12} = w_{21}$. To find the asymptotic distribution of $(W_n - W)g_{0(2)}$, observe that $W_n - W = -W(W_n^{-1} - W^{-1})W_n$ and

$$\sqrt{n}\,vech(W_n^{-1} - W^{-1}) = \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} (Y_i - \bar{Y})^2 - Var(Y_i) \\ (Y_i - \bar{Y})(Z_i - \bar{Z}) - Cov(Y_i, Z_i) \\ (Z_i - \bar{Z})^2 - Var(Z_i) \end{pmatrix} \to_d N(0, \Omega_W),$$

where Ω_W is 3×3 matrix such that

$$\begin{aligned}
\Omega_W(1,1) &= E(Y_i - EY_i)^4 - Var(Y_i)^2, \\
\Omega_W(2,2) &= E(Y_i - EY_i)^2(Z_i - EZ_i)^2 - Cov(Y_i, Z_i)^2, \\
\Omega_W(3,3) &= E(Z_i - EZ_i)^4 - Var(Z_i)^2, \\
\Omega_W(1,2) &= E(Y_i - EY_i)^3(Z_i - EZ_i) - Var(Y_i)Cov(Y_i, Z_i), \\
\Omega_W(1,3) &= E(Y_i - EY_i)^2(Z_i - EZ_i)^2 - Var(Y_i)Var(Z_i), \\
\Omega_W(2,3) &= E(Y_i - EY_i)(Z_i - EZ_i)^3 - Var(Z_i)Cov(Y_i, Z_i), \\
\Omega_W(2,1) &= \Omega_W(1,2), \\
\Omega_W(3,1) &= \Omega_W(1,3), \\
\Omega_W(3,2) &= \Omega_W(2,3).
\end{aligned}$$

Let

$$\begin{aligned}
f_1 &= (Y_i - EY_i)^2 - Var(Y_i), \\
f_2 &= (Y_i - EY_i)(Z_i - EZ_i) - Cov(Y_i, Z_i), \\
f_3 &= (Z_i - EZ_i)^2 - Var(Z_i).
\end{aligned}$$

Then, $\sqrt{n}(W_n - W)g_{0(2)}$ can be written as

$$\begin{aligned}\sqrt{n}(W_n - W)g_{0(2)} &= -\sqrt{n}W(W_n^{-1} - W^{-1})W_n g_{0(2)} \\ &\simeq -\sqrt{n}W(W_n^{-1} - W^{-1})W g_{0(2)} \\ &\simeq \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} W_{1,i} \\ W_{2,i}\end{pmatrix},\end{aligned}$$

where $W_{1,i} = a_1 f_1 + a_2 f_2 + a_3 f_3$ and $W_{2,i} = b_1 f_1 + b_2 f_2 + b_3 f_3$ ($a_k, b_k,\ k = 1, 2, 3$ are defined below). Note that the replacement of W_n with W between the RHS of the first and the second line of the above equations does not change the asymptotic distribution of $\sqrt{n}(W_n - W)g_{0(2)}$, while calculations become simpler. Similarly, the replacement of $vech(W_n^{-1} - W^{-1})$ with $(f_1, f_2, f_3)'$ between the RHS of the second and the last line is justified.

Let $D = Var(Y_i)Var(Z_i) - Cov(Y_i, Z_i)^2$, $g_1 = g_{0(2)}(1) = EY_i - \mu_Y$ and $g_2 = g_{0(2)}(2) = EZ_i - \theta_{0(2)}$. Then,

$$\begin{aligned}a_1 &= -w_{11}^2 g_1 - w_{11}w_{12}g_2 = -\frac{Var(Z_i)}{Var(Y_i)}\frac{\delta}{D}, \\ a_2 &= -2w_{11}w_{12}g_1 - (w_{12}^2 + w_{11}w_{22})g_2 = \frac{Cov(Y_i, Z_i)}{Var(Y_i)}\frac{\delta}{D}, \\ a_3 &= -w_{12}^2 g_1 - w_{11}w_{12}g_2 = \left(\frac{Var(Z_i)}{Var(Y_i)} - 1\right)Cov(Y_i, Z_i)^2\frac{\delta}{D^2}, \\ b_1 &= -w_{12}w_{11}g_1 - w_{12}^2 g_2 = a_2, \\ b_2 &= -(w_{11}w_{22} + w_{12}^2)g_1 - 2w_{12}w_{22}g_2 = -\frac{\delta}{D}, \\ b_3 &= -w_{12}w_{22}g_1 - w_{22}^2 g_2 = 0.\end{aligned}$$

Now we are ready to find

$$\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} W_{1,i} \\ W_{2,i}\end{pmatrix} \rightarrow_d N\left(\begin{pmatrix} 0 \\ 0\end{pmatrix}, \begin{pmatrix} \sigma_{W_1}^2 & \sigma_{W_1 W_2} \\ \sigma_{W_1 W_2} & \sigma_{W_2}^2\end{pmatrix}\right),$$

where

$$\begin{aligned}
\sigma^2_{W_1} &= a_1^2\Omega_W(1,1) + a_2^2\Omega_W(2,2) + a_3^2\Omega_W(3,3) + 2a_1a_2\Omega_W(1,2) \\
&\quad +2a_1a_3\Omega_W(1,3) + 2a_2a_3\Omega_W(2,3) \\
\sigma^2_{W_2} &= b_1^2\Omega_W(1,1) + b_2^2\Omega_W(2,2) + 2b_1b_2\Omega_W(1,2) \\
\sigma_{W_1W_2} &= a_1b_1\Omega_W(1,1) + (a_1b_2 + a_2b_1)\Omega_W(1,2) + a_2b_2\Omega_W(2,2) \\
&\quad +a_3b_1\Omega_W(3,1) + a_3b_2\Omega_W(3,2).
\end{aligned}$$

In order to find the covariances of $g_n(\theta_{0(2)}) - g_{0(2)}$ and $(W_n - W)g_{0(2)}$, observe that $W_{1,i}$ and $W_{2,i}$ are mean zero processes and

$$\begin{aligned}
\sigma_{Y,W_1} &\equiv E[(Y_i - EY_i)W_{1,i}] \\
&= a_1E(Y_i - EY_i)^3 + a_2E(Y_i - EY_i)^2(Z_i - EZ_i) + a_3E(Y_i - EY_i)(Z_i - EZ_i)^2, \\
\sigma_{Y,W_2} &\equiv E[(Y_i - EY_i)W_{2,i}] \\
&= b_1E(Y_i - EY_i)^3 + b_2E(Y_i - EY_i)^2(Z_i - EZ_i) + b_3E(Y_i - EY_i)(Z_i - EZ_i)^2, \\
\sigma_{Z,W_1} &\equiv E[(Z_i - EZ_i)W_{1,i}] \\
&= a_1E(Y_i - EY_i)^2(Z_i - EZ_i) + a_2E(Y_i - EY_i)(Z_i - EZ_i)^2 + a_3E(Z_i - EZ_i)^3, \\
\sigma_{Z,W_2} &\equiv E[(Z_i - EZ_i)W_{2,i}] \\
&= b_1E(Y_i - EY_i)^2(Z_i - EZ_i) + b_2E(Y_i - EY_i)(Z_i - EZ_i)^2 + b_3E(Z_i - EZ_i)^3.
\end{aligned}$$

Now we have

$$\sqrt{n}\begin{pmatrix} g_n(\theta_{0(2)}) - g_{0(2)} \\ (G_n(\theta_{0(2)}) - G_{0(2)})'Wg_{0(2)} \\ (W_n - W)g_{0(2)} \end{pmatrix} = \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} Y_i - EY_i \\ Z_i - EZ_i \\ 0 \\ W_{1,i} \\ W_{2,i} \end{pmatrix} \to_d N(\mathbf{0}, \Omega_2),$$

where

$$\Omega_2 = \begin{pmatrix} Var(Y_i) & Cov(Y_i, Z_i) & 0 & \sigma_{Y,W_1} & \sigma_{Y,W_2} \\ Cov(Y_i, Z_i) & Var(Z_i) & 0 & \sigma_{Z,W_1} & \sigma_{Z,W_2} \\ 0 & 0 & 0 & 0 & 0 \\ \sigma_{Y,W_1} & \sigma_{Z,W_1} & 0 & \sigma^2_{W_1} & \sigma_{W_1W_2} \\ \sigma_{Y,W_2} & \sigma_{Z,W_2} & 0 & \sigma_{W_1W_2} & \sigma^2_{W_2} \end{pmatrix}.$$

Since $H_{0(2)} = G_0' W G_0 = Var(Y_i)/D$ and

$$\begin{aligned} V_2 &= \begin{pmatrix} G'_{0(2)}W & 1 & G'_{0(2)} \end{pmatrix} \Omega_2 \begin{pmatrix} G'_{0(2)}W & 1 & G'_{0(2)} \end{pmatrix}' \\ &= \begin{pmatrix} \frac{Cov(Y_i,Z_i)}{D} & -\frac{Var(Y_i)}{D} & 1 & 0 & -1 \end{pmatrix} \Omega_2 \begin{pmatrix} \frac{Cov(Y_i,Z_i)}{D} & -\frac{Var(Y_i)}{D} & 1 & 0 & -1 \end{pmatrix}' \\ &= \frac{Var(Y_i)}{D}(1 + 2\sigma_{Z,W_2}) - 2\frac{Cov(Y_i, Z_i)}{D}\sigma_{Y,W_2} + \sigma^2_{W_2}, \end{aligned}$$

we finally have,

$$\Sigma_{MR} = H^{-1}_{0(2)} V_2 H^{-1}_{0(2)} = \frac{D}{Var(Y_i)}(1+2\sigma_{Z,W_2}) - 2\frac{D \cdot Cov(Y_i, Z_i)}{Var(Y_i)^2}\sigma_{Y,W_2} + \frac{D^2}{Var(Y_i)^2}\sigma^2_{W_2}.$$

Since X_i is a bivariate normal random vector,

$$\begin{aligned} E(Y_i - EY_i)^4 &= 3Var(Y_i)^2, \\ E(Y_i - EY_i)^2(Z_i - EZ_i)^2 &= Var(Y_i)Var(Z_i) + 2Cov(Y_i, Z_i)^2, \\ E(Z_i - EZ_i)^4 &= 3Var(Z_i)^2, \\ E(Y_i - EY_i)^3(Z_i - EZ_i) &= 3Var(Y_i)Cov(Y_i, Z_i), \\ E(Y_i - EY_i)(Z_i - EZ_i)^3 &= 3Var(Z_i)Cov(Y_i, Z_i). \end{aligned}$$

Since

$$E(Y_i-EY_i)^3 = E(Y_i-EY_i)^2(Z_i-EZ_i) = E(Y_i-EY_i)(Z_i-EZ_i)^2 = E(Z_i-EZ_i)^3 = 0,$$

the covariances of $g_n(\theta_{0(2)}) - g_{0(2)}$ and $(W_n - W)g_{0(2)}$ are zeros. Then,

$$\begin{aligned}\Sigma_{MR} &= \frac{D}{Var(Y_i)}(1+2\sigma_{Z,W_2}) - 2\frac{D\cdot Cov(Y_i,Z_i)}{Var(Y_i)^2}\sigma_{Y,W_2} + \frac{D^2}{Var(Y_i)^2}\sigma^2_{W_2}\\ &= \frac{D}{Var(Y_i)^2}(Var(Y_i)+\delta^2) = (1-\rho^2)(1+\delta^2).\end{aligned}$$

The MR bootstrap GMM estimators $\hat{\theta}^*_{(j)}$, $j = 1, 2$.

The MR bootstrap one-step GMM estimator is $\hat{\theta}^*_{(1)} = \bar{Z}^*$. Then, we construct a (centered) weight matrix W_n^* such that

$$\begin{aligned}W_n^* &= \left(n^{-1}(g(X_i^*,\hat{\theta}^*_{(1)}) - g_n^*(\hat{\theta}^*_{(1)}))(g(X_i^*,\hat{\theta}^*_{(1)}) - g_n^*(\hat{\theta}^*_{(1)}))'\right)^{-1}\\ &= \begin{pmatrix} \widehat{Var}(Y_i^*) & \widehat{Cov}(Y_i^*,Z_i^*)\\ \widehat{Cov}(Y_i^*,Z_i^*) & \widehat{Var}(Z_i^*)\end{pmatrix}^{-1}.\end{aligned}$$

In this example, the weight matrix does not depend on $\hat{\theta}^*_{(1)}$. The FOC for the MR bootstrap two-step GMM estimator is $G_n^*(\hat{\theta}^*_{(2)})'W_n^* g_n^*(\hat{\theta}^*_{(2)}) = 0$. Since $G_n^*(\hat{\theta}^*_{(2)})' = (0, -1)$,

$$\hat{\theta}^*_{(2)} = \bar{Z}^* - \frac{\widehat{Cov}(Y_i^*,Z_i^*)}{\widehat{Var}(Y_i^*)}\bar{Y}^*.$$

The Hall-Horowitz bootstrap GMM estimators $\hat{\theta}^*_{HH(j)}$, $j = 1, 2$.

The HH bootstrap one-step estimator solves $G_n^*(\hat{\theta}^*_{HH(1)})'(g_n^*(\hat{\theta}^*_{HH(1)}) - g_n(\hat{\theta}_{(1)})) = 0$ and is given by $\hat{\theta}^*_{HH(1)} = \bar{Z}^*$. Then, we construct a weight matrix $\widetilde{W}_n^*$ with the recentered moment function such that

$$\begin{aligned}\tilde{W}_n^* &= \left(n^{-1}(g(X_i^*,\hat{\theta}^*_{HH(1)}) - g_n(\hat{\theta}_{(2)}))(g(X_i^*,\hat{\theta}^*_{HH(1)}) - g_n(\hat{\theta}_{(2)}))'\right)^{-1}\\ &= \begin{pmatrix} \widehat{Var}(Y_i^*) + (\bar{Y}^*-\bar{Y})^2 & \widehat{Cov}(Y_i^*,Z_i^*) - (\bar{Y}^*-\bar{Y})(\bar{Z}-\hat{\theta}_{(2)})\\ \widehat{Cov}(Y_i^*,Z_i^*) - (\bar{Y}^*-\bar{Y})(\bar{Z}-\hat{\theta}_{(2)}) & \widehat{Var}(Z_i^*) + (\bar{Z}-\hat{\theta}_{(2)})^2\end{pmatrix}^{-1}\\ &\equiv \begin{pmatrix} \tilde{w}^*_{11} & \tilde{w}^*_{12}\\ \tilde{w}^*_{21} & \tilde{w}^*_{22}\end{pmatrix}.\end{aligned}$$

The FOC for the HH bootstrap two-step GMM estimator is $G_n^*(\hat{\theta}_{HH(2)}^*)'\tilde{W}_n^*(g_n^*(\hat{\theta}_{HH(2)}^*)-g_n(\hat{\theta}_{(2)}))=0$. Then,

$$\hat{\theta}_{HH(2)}^* = \bar{Z}^* + \frac{\bar{w}_{21}^*}{\bar{w}_{22}^*}(\bar{Y}^* - \bar{Y}) - (\bar{Z} - \hat{\theta}_{(2)}).$$

Note that the bootstrap two-step GMM estimator using the original moment function $\hat{\theta}_{(2)}^*$, and using the recentered moment function $\hat{\theta}_{HH(2)}^*$, are different.

The Brown-Newey bootstrap GMM estimators $\hat{\theta}_{BN}^*$

For the Brown-Newey bootstrap, we resample the bootstrap sample $X_{EL,1}^*, ...X_{EL,n}^*$ from the empirical likelihood probability distribution ($\hat{p}_i$ attached to each sample observation), rather then the edf. The BN bootstrap one-step GMM estimator is $\hat{\theta}_{BN(1)}^* = \bar{Z}_{EL}^*$. Then, we construct a uncentered weight matrix $\tilde{\tilde{W}}_n^*$ with the original moment function such that

$$\begin{aligned}\tilde{\tilde{W}}_n^* &= \left(n^{-1}(g(X_{EL,i}^*, \hat{\theta}_{BN(1)}^*))(g(X_{EL,i}^*, \hat{\theta}_{BN(1)}^*))'\right)^{-1} \\ &= \begin{pmatrix} \widehat{Var}(Y_{EL,i}^*) + \bar{Y}_{EL}^{*2} & \widehat{Cov}(Y_{EL,i}^*, Z_{EL,i}^*) \\ \widehat{Cov}(Y_{EL,i}^*, Z_{EL,i}^*) & \widehat{Var}(Z_{EL,i}^*) \end{pmatrix}^{-1}.\end{aligned}$$

The FOC for the BN bootstrap two-step GMM estimator is $G_n^*(\hat{\theta}_{BN(2)}^*)'\tilde{\tilde{W}}_n^* g_n^*(\hat{\theta}_{BN(2)}^*) = 0$. Then,

$$\hat{\theta}_{BN(2)}^* = \bar{Z}_{EL}^* - \frac{\widehat{Cov}(Y_{EL,i}^*, Z_{EL,i}^*)}{\widehat{Var}(Y_{EL,i}^*) + \bar{Y}_{EL}^{*2}}\bar{Y}_{EL}^*.$$

3. Example 2: Invalid Instrumental Variables

The GMM estimator and the pseudo-true value

The FOC for $\hat\beta_{(1)}$ is $G_n(\hat\beta_{(1)})'W_n g_n(\hat\beta_{(1)}) = 0$, where

$$\begin{aligned} G_n(\beta) &= \begin{pmatrix} -n^{-1}\sum z_{1i}x_i \\ -n^{-1}\sum z_{2i}x_i) \end{pmatrix}, \\ g_n(\beta) &= \begin{pmatrix} n^{-1}\sum z_{1i}y_i - n^{-1}\sum z_{1i}x_i\beta \\ n^{-1}\sum z_{2i}y_i - n^{-1}\sum z_{2i}x_i\beta \end{pmatrix}, \\ W_n &= \left(n^{-1}\sum_{i=1}^n \begin{pmatrix} z_{1i}^2 & z_{1i}z_{2i} \\ z_{1i}z_{2i} & z_{2i}^2 \end{pmatrix} \right)^{-1}. \end{aligned}$$

Thus, the one-step GMM estimator is

$$\hat\beta_{(1)} = \frac{GW1 \cdot n^{-1}\sum_{i=1}^n z_{1i}y_i + GW2 \cdot n^{-1}\sum_{i=1}^n z_{2i}y_i}{GW1 \cdot n^{-1}\sum_{i=1}^n z_{1i}x_i + GW2 \cdot n^{-1}\sum_{i=1}^n z_{2i}x_i},$$

where

$$\begin{aligned} GW1 &= n^{-1}\sum_{i=1}^n z_{1i}x_i \cdot n^{-1}\sum_{i=1}^n z_{2i}^2 - n^{-1}\sum_{i=1}^n z_{2i}x_i \cdot n^{-1}\sum_{i=1}^n z_{1i}z_{2i}, \\ GW2 &= -n^{-1}\sum_{i=1}^n z_{1i}x_i \cdot n^{-1}\sum_{i=1}^n z_{1i}z_{2i} + n^{-1}\sum_{i=1}^n z_{2i}x_i \cdot n^{-1}\sum_{i=1}^n z_{1i}^2. \end{aligned}$$

The probability limit of $\hat\beta_{(1)}$ is

$$\beta_{0(1)} = \frac{Ez_{1i}x_iEz_{2i}^2Ez_{1i}y_i + Ez_{2i}x_iEz_{1i}^2Ez_{2i}y_i}{Ez_{2i}^2(Ez_{1i}x_i)^2 + Ez_{1i}^2(Ez_{2i}x_i)^2} = \beta_0 + \frac{Ez_{2i}x_iEz_{1i}^2}{Ez_{2i}^2(Ez_{1i}x_i)^2 + Ez_{1i}^2(Ez_{2i}x_i)^2} \cdot \delta,$$

provided that $Ez_{1i}z_{2i} = 0$, $Ez_{1i}e_i = 0$ and $Ez_{2i}e_i = \delta$.

2 ASYMPTOTIC REFINEMENTS OF A MISSPECIFICATION-ROBUST BOOTSTRAP FOR EMPIRICAL LIKELIHOOD ESTIMATORS

2.1 Introduction

The goal of this paper is to establish asymptotic refinements of a misspecification-robust bootstrap critical values for t tests and confidence intervals (CI's) based on the empirical likelihood (EL) estimators. The term *misspecification-robust* implies that the bootstrap t tests and CI's yield correct rejection and coverage probabilities regardless of whether the model is correctly specified or not. The model is *misspecified* when there is no such parameter that satisfies the assumed moment restrictions. This type of misspecification occurs only if the model is *overidentified*, that is, the number of moment restrictions is greater than the number of parameters to be estimated.

For overidentified moment condition models, generalized method of moments (GMM) of Hansen (1982) is traditionally used to get point estimates, to make inferences, and to construct CI's. However, GMM estimators have been known to have relatively large finite sample bias. In addition, simulation studies show first-order asymptotic approximation to the distribution of test statistics based on GMM estimators perform poorly in finite sample. To improve upon first-order asymptotic approximation for GMM, Hall and Horowitz (1996) and Andrews (2002) suggest to use the bootstrap critical values and CI's. These bootstrap tests and CI's achieve asymptotic refinements over the first-order asymptotic critical values and CI's, which means it has smaller errors in the test rejection probability and CI coverage probability.

On the other hand, generalized empirical likelihood (GEL) estimators (Newey and Smith (2004)) have been considered as alternatives to GMM estimators. GEL circumvent the estimation of the optimal weight matrix, which has been considered as a significant source of poor finite sample performance of the two-step efficient GMM. GEL includes the EL estimator of Owen (1988), Owen (1990), Qin and Lawless (1994), and Imbens (1997), the exponential tilting (ET) of Kitamura and Stutzer

(1997) and Imbens et al. (1998), and the continuously updating (CU) estimator of Hansen et al. (1996). By investigating the higher-order asymptotic properties, Newey and Smith (2004) shows that EL possesses theoretical advantages over the other GEL estimators. Thus, the EL estimator is a favorable alternative to GMM estimators and it is natural to consider bootstrap t tests and CI's based on the EL estimator to further improve its finite sample property.

Few published papers explicitly deal with bootstrapping for the EL estimator, however. Brown and Newey (2002) and Allen et al. (2011) employ the EL probability in resampling the bootstrap sample for GMM estimators, not for the EL estimator. Canay (2010) proposes a bootstrap method that uses EL probability for moment inequality models. The resulting bootstrap confidence region for the empirical likelihood ratio statistic is shown to be valid, but neither the validity nor asymptotic refinements for t tests based on the EL estimator are not established.

This paper proposes a bootstrap procedure for the critical values for t tests and CI's based on the EL estimator, that achieves asymptotic refinements robust to misspecification. In other words, the refinements result holds both under the correct model and the misspecified model. In addition, I establish the same asymptotic refinements of the misspecification-robust bootstrap for the exponentially tilted empirical likelihood (ETEL) estimator of Schennach (2007). The reason that I also consider the ETEL estimator is that the original EL estimator shows a questionable behavior under misspecification. It is not $\sqrt{n}$-consistent without a strong assumption on the support of the moment function, henceforth referred to as the uniform boundedness condition. The ETEL estimator is designed to behave well under misspecification without the uniform boundedness condition and it shares the same favorable higher-order asymptotic properties with the EL estimator under correct model specification. Thus, the EL estimators considered in this paper includes both the original EL estimator and the ETEL estimator.

The remainder of the paper is organized as follows. Section 2.2 provides a heuristic explanation on why the misspecification-robust bootstrap works with the EL estimators. Section 2.3 defines the estimators and the t statistic. Section 2.4 describes the nonparametric iid misspecification-robust bootstrap procedure for the

EL estimators. Section 2.5 states the assumptions and establishes the asymptotic refinements of the misspecification-robust bootstrap. Section 2.6 presents an example and the results of the Monte Carlo simulation. Appendix contains Lemmas and proofs of the results. Technical Appendix contains the calculation of the pseudo-true value in the example of Section 2.6.

2.2 Why the Misspecification-Robust Bootstrap Works with the Empirical Likelihood Estimators

Construction of the asymptotically pivotal statistic is a critical condition to get asymptotic refinements of the bootstrap: See Beran (1988), Hall (1997), Hall and Horowitz (1996), Horowitz (2001), and Brown and Newey (2002) among others. That is, the sample test statistic is required to be asymptotically pivotal and the bootstrap test statistic is required to be asymptotically pivotal conditional on the sample. Since the t statistic is the one of interest, we need to construct the t statistic to converge in distribution to a standard normal distribution, both in the sample and the bootstrap sample.

Suppose that $\chi_n = \{X_i : i \leq n\}$ is an independent and identically distributed (iid) random sample. Let F be the corresponding cumulative distribution function (cdf). The empirical distribution function (edf) is denoted by F_n. Let θ be a parameter of interest and $g(X_i, \theta)$ be a moment function. Let $\hat{\theta}$ be either the EL estimator with uniform boundedness condition, $\sup_{\theta \in \Theta, x \in \chi} \|g(x, \theta)\| < \infty$, or the ETEL estimator. Then $\hat{\theta}$ is $\sqrt{n}$-consistent regardless of whether the model is misspecified or not. Let $\hat{\Sigma}$ be a consistent estimator of the asymptotic variance of $\sqrt{n}(\hat{\theta} - \theta_0)$. The formula for the misspecification-robust estimator $\hat{\Sigma}$ is available in the following section.

The (pseudo-)true value $\theta_0 \equiv plim(\hat{\theta})$ uniquely maximizes the corresponding objective function. If the moment condition $Eg(X_i, \theta_0) = 0$ holds, then the model is correctly specified. If $Eg(X_i, \theta) \neq 0$ for all θ in the parameter space, then the model is misspecified. This can happen if the model is overidentified. Throughout the paper, I assume that the model is possibly misspecified and overidentified. Note that this type of misspecification is different from the one of White (1982). In his quisi-maximum likelihood (QML) framework, what is misspecified is the underlying probability distribution. In addition, the QML theory deals with just-identified models, where the number of parameters is equal to the number of moment restrictions. For bootstrapping QML estimators, see Gonçalves and White (2004).

I also define the bootstrap sample. Let $\chi^*_{n_b} = \{X^*_i : i \leq n_b\}$ be a sample of random vectors conditional on χ_n with the edf F_n. In this section, I distinguish the number of sample n and the number of bootstrap sample n_b, which helps understanding the concept of the conditional asymptotic distribution. Define the bootstrap EL or ETEL estimator $\hat{\theta}^*$ and the bootstrap covariance estimator $\hat{\Sigma}^*$ as their sample versions are defined, but with $\chi^*_{n_b}$ in place of χ_n.

Construct the sample and the bootstrap t statistics

$$T(\chi_n) \equiv \frac{\hat{\theta} - \theta_0}{\sqrt{\hat{\Sigma}/n}}, \quad T(\chi^*_{n_b}) \equiv \frac{\hat{\theta}^* - \hat{\theta}}{\sqrt{\hat{\Sigma}^*/n_b}},$$

respectively. By writing the t statistics as $T(\chi_n)$ and $T(\chi^*_{n_b})$, I emphasize that the two versions of the t statistic have the same formula except that the sample t statistic is based on the sample χ_n and the bootstrap t statistic is based on the bootstrap sample $\chi^*_{n_b}$.

The correctly specified population moment condition is $Eg(X_i, \theta_0) = 0$, but I assume that this may not hold. By construction, $T(\chi_n) \rightarrow_d N(0, 1)$ regardless of whether the population moment condition holds or not. For the bootstrap t statistic $T(\chi^*_{n_b})$, the corresponding bootstrap moment condition is $E^*g(X^*_i, \hat{\theta}) = 0$, where E^* is expectation with respect to the distribution of the bootstrap sample conditional on the sample. This bootstrap moment condition does not hold in general, because

$$E^*g(X^*_i, \hat{\theta}) = n^{-1}\sum_{i=1}^{n} g(X_i, \hat{\theta}) \neq 0,$$

when the model is overidentified. The above equality holds because the distribution of the bootstrap sample is the edf F_n. Thus, the model is misspecified in the sample, and this happens even if $Eg(X_i, \theta_0) = 0$ holds. By fixing n and taking $n_b \to \infty$, we get the conditional asymptotic distribution of $T(\chi^*_{n_b})$. Since $\hat{\theta}$ and $\hat{\Sigma}$ are considered as the true values given the edf F_n, we have $T(\chi^*_{n_b}) \to_{n_b} N(0,1)$ conditional on χ_n as $n_b \to \infty$. Therefore, $T(\chi^*_{n_b})$ is asymptotically pivotal conditional on the sample, regardless of whether the bootstrap moment condition as well as the population moment condition is correctly specified or not.

A natural question is whether we can use the EL- or ETEL-estimated distribution function estimator $\hat{F}$ instead of F_n in resampling. This is possible only when the population moment condition is correctly specified. By construction, $\hat{F}$ satisfies $E^*g(X_i^*, \hat{\theta}) = 0$, so that the bootstrap moment condition is *always* correctly specified. For instance, Brown and Newey (2002) argues that using the EL-estimated distribution function $\hat{F}_{EL}(z) \equiv \sum_i \mathbf{1}(X_i \leq z)\hat{p}_i$, where $\hat{p}_i$ is the EL probability, in place of F_n in resampling would improve efficiency of bootstrapping for GMM. Their argument relies on the fact that $\hat{F}_{EL}$ is a consistent estimator of the true distribution function F, which holds only under correct model specifications.

If the population moment condition happens to be misspecified, then neither $\hat{F}_{EL}$ nor the ETEL-estimated distribution function $\hat{F}_{ETEL}$ is consistent for F. To see why, note that $E^*g(X_i^*, \hat{\theta}) = 0$ holds in large sample, while $Eg(X_i, \theta_0) \neq 0$. In contrast, the edf F_n is consistent for F regardless of whether the population moment condition holds or not by Glivenko-Cantelli Theorem. Therefore, $\hat{F}_{EL}$ or $\hat{F}_{ETEL}$ cannot be used in place of F_n in the misspecification-robust bootstrap.

2.3 Estimators and Test Statistics

Let $g(X_i, \theta)$ be a moment function where θ is a parameter of interest. Let $G^{(j)}(X_i, \theta)$ denote the vectors of partial derivatives with respect to θ of order j of $g(X_i, \theta)$. In particular, $G^{(1)}(X_i, \theta) \equiv G(X_i, \theta) \equiv (\partial/\partial\theta')g(X_i, \theta)$ is a $L_g \times L_\theta$ matrix and $G^{(2)}(X_i, \theta) \equiv (\partial/\partial\theta')vec\{G(X_i, \theta)\}$ is a $L_g L_\theta \times L_\theta$ matrix, where $vec\{\cdot\}$ is the vectorization of a matrix. To simplify notation, write $g_i = g(X_i, \theta)$, $G_i^{(j)} = G^{(j)}(X_i, \theta)$,

$\hat{g}_i = g(X_i, \hat{\theta})$, and $\hat{G}_i^{(j)} = G^{(j)}(X_i, \hat{\theta})$ for $j = 1, ..., d+1$, where $\hat{\theta}$ is either the EL or the ETEL estimator.

Empirical Likelihood Estimator

The EL estimator is given by

$$\hat{\theta} = \arg\min_{\theta} -n^{-1}\sum_{i=1}^{n} \ln n\hat{p}_i(\theta), \quad \hat{p}_i(\theta) = \frac{1}{n(1-\hat{\lambda}(\theta)'g_i)},$$

where

$$\hat{\lambda}(\theta) = \arg\max_{\lambda\in\mathbf{R}^{L_g}} -n^{-1}\sum_{i=1}^{n} \ln(1-\lambda' g_i).$$

Equivalently, $\hat{\theta}$ maximizes the objective function

$$\ln \hat{L}_{EL}(\theta) = -n^{-1}\sum_{i=1}^{n} \ln(1-\hat{\lambda}(\theta)'g_i), \tag{2.1}$$

where $\hat{\lambda}(\theta)$ is such that

$$n^{-1}\sum_{i=1}^{n} \frac{g_i}{1-\hat{\lambda}'g_i} = 0. \tag{2.2}$$

The first-order conditions (FOC's) are

$$\underset{L_\theta\times 1}{0} = n^{-1}\sum_{i=1}^{n} \frac{\hat{G}_i'\hat{\lambda}}{1-\hat{\lambda}'\hat{g}_i}, \quad \underset{L_g\times 1}{0} = n^{-1}\sum_{i=1}^{n} \frac{\hat{g}_i}{1-\hat{\lambda}'\hat{g}_i},$$

and the FOC's hold regardless of model misspecifications. By using the standard asymptotic theory of just-identified GMM estimators (or Z-estimators), we can find the asymptotic distribution of the EL estimator robust to misspecification. Assume appropriate regularity conditions that includes the uniform boundedness condition on $g(X_i, \theta)$, $\sup_{\theta\in\Theta, x\in\chi} \|g(x,\theta)\| < \infty$.

Let $\beta = (\lambda', \theta')'$ and $\psi(X_i, \beta)$ be a $(L_g + L_\theta) \times 1$ vector such that

$$\psi(X_i, \beta) \equiv \begin{bmatrix} \psi_1(X_i, \beta) \\ \psi_2(X_i, \beta) \end{bmatrix} = \begin{bmatrix} (1 - \lambda' g_i)^{-1} G_i' \lambda \\ (1 - \lambda' g_i)^{-1} g_i \end{bmatrix}.$$

Then, $\hat{\beta} = (\hat{\lambda}', \hat{\theta}')'$ is given by the solution to $n^{-1} \sum_i^n \psi(X_i, \hat{\beta}) = 0$ and the following Proposition holds:

Proposition 2.1. *Suppose regularity conditions hold. Then,*

$$\sqrt{n}(\hat{\beta} - \beta_0) \to_d N(0, \Gamma_\psi^{-1} \Psi (\Gamma_\psi')^{-1}),$$

where $\Gamma_\psi = E(\partial/\partial\beta')\psi(X_i, \beta_0)$ *and* $\Psi = E\psi(X_i, \beta_0)\psi(X_i, \beta_0)'$.

$\beta_0 = (\lambda_0', \theta_0')'$ is the pseudo-true value that solves the population version of the FOC's:

$$\underset{L_\theta \times 1}{0} = E \frac{G_i(\theta_0)' \lambda_0}{1 - \lambda_0' g_i(\theta_0)}, \quad \underset{L_g \times 1}{0} = E \frac{g_i(\theta_0)}{1 - \lambda_0' g_i(\theta_0)}.$$

The Jacobian matrix is given by

$$\frac{\partial \psi(X_i, \beta)}{\partial \beta'} = \begin{bmatrix} \frac{G_i'}{1 - \lambda' g_i} + \frac{G_i' \lambda g_i'}{(1 - \lambda' g_i)^2} & \frac{(\lambda' \otimes I_{L_\theta}) G_i^{(2)}}{1 - \lambda' g_i} + \frac{G_i' \lambda \lambda' G_i}{(1 - \lambda' g_i)^2} \\ \frac{g_i g_i'}{(1 - \lambda' g_i)^2} & \frac{G_i}{1 - \lambda' g_i} + \frac{g_i \lambda' G_i}{(1 - \lambda' g_i)^2} \end{bmatrix}.$$

Γ_ψ and Ψ are estimated by

$$\hat{\Gamma}_\psi = n^{-1} \sum_i^n \frac{\partial \psi(X_i, \hat{\beta})}{\partial \beta'} \quad \text{and} \quad \hat{\Psi} = n^{-1} \sum_i^n \psi(X_i, \hat{\beta}) \psi(X_i, \hat{\beta})',$$

respectively. The lower right $L_\theta \times L_\theta$ submatrix of $\Gamma_\psi^{-1} \Psi (\Gamma_\psi')^{-1}$, denoted by Σ, is the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$. Let $\hat{\Sigma}$ be the corresponding submatrix of the covariance estimator $\hat{\Gamma}_\psi^{-1} \hat{\Psi} (\hat{\Gamma}_\psi')^{-1}$.

Exponentially Tilted Empirical Likelihood Estimator

Schennach (2007) suggests the ETEL estimator, which is designed to be robust to model misspecifications without the uniform boundedness condition, $\sup_{\theta\in\Theta,x\in\mathcal{X}}\|g(x,\theta)\| < \infty$.

The ETEL estimator is given by

$$\hat{\theta} = \arg\min_{\theta} -n^{-1}\sum_{i=1}^{n}\ln n\hat{w}_i(\theta), \quad \hat{w}_i(\theta) = \frac{e^{\hat{\lambda}(\theta)'g_i}}{\sum_{j=1}^{n} e^{\hat{\lambda}(\theta)'g_j}},$$

where

$$\hat{\lambda}(\theta) = \arg\max_{\lambda\in\mathbf{R}^{L_g}} -n^{-1}\sum_{i=1}^{n} e^{\lambda' g_i}.$$

This estimator is a hybrid of the EL estimator and the ET probability. Equivalently, the ETEL estimator $\hat{\theta}$ maximizes the objective function

$$\ln \hat{L}(\theta) = -\ln\left(n^{-1}\sum_{i=1}^{n} e^{\hat{\lambda}(\theta)'(g_i - g_n)}\right), \tag{2.3}$$

where $\hat{\lambda}(\theta)$ is such that

$$n^{-1}\sum_{i=1}^{n} e^{\hat{\lambda}(\theta)'g_i}\cdot g_i = 0, \tag{2.4}$$

and $g_n = n^{-1}\sum_{i=1}^{n} g_i$. In order to describe the asymptotic distribution of the ETEL estimator, Schennach introduces auxiliary parameters to formulate the problem into a just-identified GMM. Let $\beta = (\tau,\kappa',\lambda',\theta')'$, where τ is a scalar and $\kappa\in\mathbf{R}^{L_g}$. By Lemma 9 of Schennach (2007), the ETEL estimator $\hat{\theta}$ is given by the subvector of $\hat{\beta} = (\hat{\tau},\hat{\kappa}',\hat{\lambda}',\hat{\theta}')'$, the solution to

$$n^{-1}\sum_{i}\phi(X_i,\hat{\beta}) = 0, \tag{2.5}$$

where

$$\phi(X_i,\beta) \equiv \begin{bmatrix}\phi_1(X_i,\beta)\\ \phi_2(X_i,\beta)\\ \phi_3(X_i,\beta)\\ \phi_4(X_i,\beta)\end{bmatrix} = \begin{bmatrix} e^{\lambda' g_i} - \tau \\ e^{\lambda' g_i}\cdot g_i \\ (\tau - e^{\lambda' g_i})\cdot g_i + e^{\lambda' g_i}\cdot g_i g_i'\kappa \\ e^{\lambda' g_i} G_i'\left(\kappa + \lambda g_i'\kappa - \lambda\right) + \tau G_i'\lambda \end{bmatrix}.$$

By Theorem 10 of Schennach (2007),

$$\sqrt{n}(\hat{\beta} - \beta_0) \to_d N(0, \Gamma_\phi^{-1}\Phi(\Gamma_\phi')^{-1}),$$

where $\Gamma_\phi = E(\partial/\partial\beta')\phi(X_i, \beta_0)$ and $\Phi = E\phi(X_i, \beta_0)\phi(X_i, \beta_0)'$. The pseudo-true value β_0 solves $E\phi(X_i, \beta_0) = 0$. In particular, $\lambda_0(\theta)$ is the solution to $Ee^{\lambda' g_i} g_i = 0$ and θ_0 is a unique maximizer of the population objective function:

$$\ln L(\theta) = -\ln\left(Ee^{\lambda_0(\theta)'(g_i - Eg_i)}\right).$$

Γ_ϕ and Φ are estimated by

$$\hat{\Gamma}_\phi \equiv n^{-1}\sum_i^n \frac{\partial\phi(X_i, \hat{\beta})}{\partial\beta'} \quad \text{and} \quad \hat{\Phi} \equiv n^{-1}\sum_i^n \phi(X_i, \hat{\beta})\phi(X_i, \hat{\beta})',$$

respectively.

In order to estimate Γ_ϕ, we need an exact formula of $(\partial/\partial\beta')\phi(X_i, \beta)$.[1] Let $G_i^{(2)} = (\partial/\partial\theta')vec\{G_i\}$, a $L_g L_\theta \times L_\theta$ matrix. The partial derivative of $\phi_1(X_i, \beta)$ is given by

$$\frac{\partial\phi_1(X_i, \beta)}{\partial\beta'} = \left(\underset{1\times 1}{-1} \quad \underset{1\times L_g}{\mathbf{0}} \quad \underset{1\times L_g}{e^{\lambda' g_i} g_i'} \quad \underset{1\times L_\theta}{e^{\lambda' g_i}\lambda' G_i} \right).$$

The partial derivative of $\phi_2(X_i, \beta)$ is given by

$$\frac{\partial\phi_2(X_i, \beta)}{\partial\beta'} = \left(\underset{L_g\times 1}{\mathbf{0}} \quad \underset{L_g\times L_g}{\mathbf{0}} \quad \underset{L_g\times L_g}{e^{\lambda' g_i} g_i g_i'} \quad \underset{L_g\times L_\theta}{e^{\lambda' g_i}(G_i + g_i\lambda' G_i)} \right).$$

The partial derivative of $\phi_3(X_i, \beta)$ is given by

$$\frac{\partial\phi_3(X_i, \beta)}{\partial\beta'} = \left(\underset{L_g\times 1}{g_i} \quad \underset{L_g\times L_g}{e^{\lambda' g_i} g_i g_i'} \quad \underset{L_g\times L_g}{e^{\lambda' g_i}(g_i g_i'\kappa g_i' - g_i g_i')} \quad \underset{L_g\times L_\theta}{(\partial/\partial\theta')\phi_3(X_i, \beta)} \right),$$

[1] This formula is not given in Schennach (2007), though the value of Γ under correct specification appears in the technical report of Schennach (2007).

where

$$\frac{\partial \phi_3(X_i,\beta)}{\partial \theta'} = e^{\lambda' g_i}\left\{g_i g_i'\kappa\lambda' + g_i\kappa' - g_i\lambda' + (g_i'\kappa - 1)I_{L_g}\right\}G_i + \tau G_i.$$

The partial derivatives of $\phi_4(X_i,\beta)$ is given by

$$\frac{\partial \phi_4(X_i,\beta)}{\partial \beta'} = \left(\underset{L_\theta\times 1}{G_i'\lambda} \quad \underset{L_\theta\times L_g}{e^{\lambda' g_i}(G_i' + G_i'\lambda g_i')} \quad \underset{L_\theta\times L_g}{(\partial/\partial\lambda')\phi_4(X_i,\beta)} \quad \underset{L_\theta\times L_\theta}{(\partial/\partial\theta')\phi_4(X_i,\beta)} \right),$$

where

$$\begin{aligned}
\frac{\partial \phi_4(X_i,\beta)}{\partial \lambda'} &= e^{\lambda' g_i}G_i'\left\{\kappa g_i' + \lambda g_i'\kappa g_i' - \lambda g_i' + (g_i'\kappa - 1)I_{L_g}\right\} + \tau G_i', \\
\frac{\partial \phi_4(X_i,\beta)}{\partial \theta'} &= e^{\lambda' g_i}\left\{G_i'(\kappa\lambda' + \lambda\kappa' + \lambda g_i'\kappa\lambda' - \lambda\lambda')G_i\right. \\
&\quad \left. + ((\kappa' + \kappa' g_i\lambda' - \lambda')\otimes I_{L_\theta})G_i^{(2)}\right\} + \tau(\lambda'\otimes I_{L_\theta})G_i^{(2)}.
\end{aligned}$$

The lower right $L_\theta \times L_\theta$ submatrix of $\Gamma_\phi^{-1}\Phi(\Gamma_\phi')^{-1}$, denoted by Σ, is the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$. Let $\hat{\Sigma}$ be the corresponding submatrix of the covariance estimator $\hat{\Gamma}_\phi^{-1}\hat{\Phi}(\hat{\Gamma}_\phi')^{-1}$.

The t statistic

Let $\hat{\theta}$ be either the EL estimator that solves (2.1)-(2.2) or the ETEL estimator that solves (2.3)-(2.4). $\hat{\Sigma}$ is the corresponding covariance matrix estimator.

Let θ_r, $\theta_{0,r}$, and $\hat{\theta}_r$ denote the rth elements of θ, θ_0, and $\hat{\theta}$ respectively. Let $\hat{\Sigma}_{rr}$ denote the (r,r)th element of $\hat{\Sigma}$. The t statistic for testing the null hypothesis $H_0 : \theta_r = \theta_{0,r}$ is

$$T(\chi_n) = \frac{\hat{\theta}_r - \theta_{0,r}}{\sqrt{\hat{\Sigma}_{rr}/n}}.$$

$T(\chi_n)$ is misspecification-robust because it has an asymptotic $N(0,1)$ distribution under H_0, without assuming the correct model.

The symmetric two-sided t test with asymptotic significance level α rejects H_0 if $|T(\chi_n)| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal distribution.

The corresponding CI for $\theta_{0,r}$ with asymptotic confidence level $100(1-\alpha)\%$ is $CI_n = [\hat{\theta}_r \pm z_{\alpha/2}\sqrt{\hat{\Sigma}_{rr}/n}]$. The error in the rejection probability of the t test with $z_{\alpha/2}$ and coverage probability of CI_n is $O(n^{-1})$: Under H_0, $P\left(|T(\chi_n)| > z_{\alpha/2}\right) = \alpha + O(n^{-1})$ and $P\left(\theta_{0,r} \in CI_n\right) = 1 - \alpha + O(n^{-1})$.

2.4 The Misspecification-Robust Bootstrap Procedure

The nonparametric iid bootstrap is implemented by sampling $X_1^*, \cdots, X_n^*$ randomly with replacement from the sample $X_1, \cdots, X_n$. I do not employ the EL or ETEL probability in resampling, because the EL or ETEL estimator of the cdf is inconsistent for the true cdf.

The bootstrap EL estimator $\hat{\theta}^*$ is given by the subvector of $\hat{\beta}^* = (\hat{\lambda}^{*\prime}, \hat{\theta}^{*\prime})'$, the solution to

$$n^{-1}\sum_{i}^{n}\psi(X_i^*, \hat{\beta}^*) = 0. \tag{2.6}$$

The bootstrap version of the covariance matrix estimator is $\hat{\Gamma}_\psi^{*-1}\hat{\Psi}^*(\hat{\Gamma}_\psi^{*\prime})^{-1}$, where $\hat{\Gamma}_\psi^* \equiv n^{-1}\sum_i^n(\partial/\partial\beta')\psi(X_i^*, \hat{\beta}^*)$ and $\hat{\Psi}^* \equiv n^{-1}\sum_i^n \psi(X_i^*, \hat{\beta}^*)\psi(X_i^*, \hat{\beta}^*)'$.

The bootstrap ETEL estimator $\hat{\theta}^*$ is given by the subvector of $\hat{\beta}^* = (\hat{\tau}^*, \hat{\kappa}^{*\prime}, \hat{\lambda}^{*\prime}, \hat{\theta}^{*\prime})'$, the solution to

$$n^{-1}\sum_{i}^{n}\phi(X_i^*, \hat{\beta}^*) = 0. \tag{2.7}$$

The bootstrap version of the covariance matrix estimator is $\hat{\Gamma}_\phi^{*-1}\hat{\Phi}^*(\hat{\Gamma}_\phi^{*\prime})^{-1}$, where $\hat{\Gamma}_\phi^* \equiv n^{-1}\sum_i^n(\partial/\partial\beta')\phi(X_i^*, \hat{\beta}^*)$ and $\hat{\Phi}^* \equiv n^{-1}\sum_i^n \phi(X_i^*, \hat{\beta}^*)\phi(X_i^*, \hat{\beta}^*)'$.

Let $\hat{\Sigma}^*$ be the corresponding submatrix of the bootstrap covariance estimator $\hat{\Gamma}_\psi^{*-1}\hat{\Psi}^*(\hat{\Gamma}_\psi^{*\prime})^{-1}$ if $\hat{\beta}^*$ solves (2.6), and $\hat{\Gamma}_\phi^{*-1}\hat{\Phi}^*(\hat{\Gamma}_\phi^{*\prime})^{-1}$ if $\hat{\beta}^*$ solves (2.7). The bootstrap version of the estimators use the same formula with the sample version of the estimators. The only difference between the bootstrap and the sample estimators is that the former is calculated from the bootstrap sample, χ_n^*, in place of the original sample, χ_n. Therefore, there is no additional correction such as the recentering, as in

Hall and Horowitz (1996) and Andrews (2002).

The misspecification-robust bootstrap t statistic is

$$T(\chi_n^*) = \frac{\hat{\theta}_r^* - \hat{\theta}_r}{\sqrt{\hat{\Sigma}_{rr}^*/n}}.$$

Let $z_{|T|,\alpha}^*$ denote the $1-\alpha$ quantile of $|T(\chi_n^*)|$. Following Andrews (2002), we define $z_{|T|,\alpha}^*$ to be a value that minimizes $|P^*(|T(\chi_n^*)| \leq z) - (1-\alpha)|$ over $z \in \mathbf{R}$, since the distribution of $|T(\chi_n^*)|$ is discrete. The symmetric two-sided bootstrap t test of $H_0 : \theta_r = \theta_{0,r}$ versus $H_1 : \theta_r \neq \theta_{0,r}$ rejects if $|T(\chi_n)| > z_{|T|,\alpha}^*$, and this test is of asymptotic significance level α. The $100(1-\alpha)\%$ symmetric percentile t interval for $\theta_{0,r}$ is $CI_n^* = [\hat{\theta}_r \pm z_{|T|,\alpha}^* \sqrt{\hat{\Sigma}_{rr}/n}]$.

In sum, the misspecification-robust bootstrap procedure is as follows:

1. Draw n random observations $X_1^*, ...X_n^*$ with replacement from the original sample, $X_1, ..., X_n$.

2. From the bootstrap sample χ_n^*, calculate $\hat{\theta}^*$ and $\hat{\Sigma}^*$.

3. Construct and save $T(\chi_n^*)$.

4. Repeat steps 1-3 B times and get the distribution of $|T(\chi_n^*)|$, which is discrete.

5. Find $z_{|T|,\alpha}^*$ from the distribution of $|T(\chi_n^*)|$.

6. Use $z_{|T|,\alpha}^*$ in testing $H_0 : \theta_r = \theta_{0,r}$ or in constructing CI_n^*.

2.5 Main Result

Let $f(X_i, \beta)$ be a vector containing the unique components of $\phi(X_i, \beta)$ and its derivatives with respect to the components of β through order d, and $\phi(X_i, \beta)\phi(X_i, \beta)'$ and its derivatives with respect to the components of β through order $d-1$. Let g and $G^{(j)}$ be an element of g_i and $G_i^{(j)}$, respectively, for $j = 1, ..., d+1$. In addition, let g^k be a multiplication of any k-combination of elements of g_i. For instance, if

$g(X_i,\theta) = (g_{i,1}, g_{i,2})'$, a 2×1 vector, then $g^2 = g_{i,1}^2$, $g_{i,1}g_{i,2}$, or $g_{i,2}^2$. $G^{(j)k}$ are defined analogously. Then, $f(X_i,\beta)$ contains terms of the form

$$\alpha \cdot e^{k_\tau \lambda' g_i} \cdot g^{k_0} \cdot G^{k_1} \cdot G^{(2)k_2} \cdots G^{(d+1)k_{d+1}},$$

where $k_\tau = 0,1,2$, $k_0 \le d+3$, $k_j \le d+2-j$, $0 \le \sum_{j=0}^{d+1} k_j \le d+3$, and k_j's are nonnegative integers for $j = 1, ..., d+1$ and where α denotes products of components of β.

Assumption 2.1. *$X_i, i = 1, 2, ...$ are iid.*

Assumption 2.2.

(a) *Θ is compact and θ_0 is an interior point of Θ; $\Lambda(\theta)$ is a compact set such that $\lambda_0(\theta)$ is an interior point of $\Lambda(\theta)$.*

(b) *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 $\theta_1, \theta_2 \in \Theta$; $EC_g^{q_g}(X_1) < \infty$ and $E\|g(X_1,\theta)\|^{q_g} < \infty$ for all $\theta \in \Theta$ for all $0 < q_g < \infty$.*

(c) *For some function $C_\tau(x)$, $|e^{\lambda_1' g(x,\theta_1)} - e^{\lambda_2' g(x,\theta_2)}| < C_\tau(x)\|(\lambda_1', \theta_1') - (\lambda_2', \theta_2')\|$, for all x in the support of X_1 and all $(\lambda_1', \theta_1'), (\lambda_2', \theta_2') \in \Lambda(\theta) \times \Theta$; $EC_\tau^{q_\tau}(X_1) < \infty$ and $Ee^{\lambda' g(X_1,\theta)\cdot q_\tau} < \infty$ for all $(\lambda', \theta') \in \Lambda(\theta) \times \Theta$ for some $q_\tau \ge 2$.*

Assumption 2.3.

(a) *Γ_0 is nonsigular.*

(b) *$g(x,\theta)$ is $d+1$ times differentiable with respect to θ on $N(\theta_0)$, some neighborhood of θ_0, for all x in the support of X_1, where $d \ge 1$.*

(c) *There is a function $C_G(X_1)$ such that $\|G^{(j)}(X_1,\theta) - G^{(j)}(X_1,\theta_0)\| \le C_G(X_1)\|\theta - \theta_0\|$ for all $\theta \in N(\theta_0)$ for $j = 0, 1, ..., d$; $EC_G^{q_G}(X_1) < \infty$ and $E\|G^{(j)}(X_1,\theta_0)\|^{q_G} < \infty$ for $j = 0, 1, ..., d+1$ for all $0 < q_G < \infty$.*

Assumption 2.4. *There exists a nonempty open subset U of $\mathbf{R}^{dim(X_1)}$ with the following properties:*

(a) *the distribution F of X_1 has a nonzero absolutely continuous component with respect to Lebesgue measure on $\mathbf{R}^{dim(X_1)}$ with a positive density on U, and*

(b) *$f(X_1, \beta_0)$ is continuously differentiable on U.*

Assumption 2.1 is that the sample is iid, which is also assumed in Schennach (2007). Assumption 2.2(a)-(b) are similar to Assumption 2(a)-(b) of Andrews (2002). Assumption 2.2(c) is needed especially for the ETEL estimator, and is similar to but slightly stronger than Assumption 3(4)-(6) of Schennach (2007). Assumption 2.3(a), (b), and (d) are usual regularity conditions for well-defined covariance matrix and moment function. Assumption 2.3(c) is similar to Assumption 3(d)-(e) of Andrews (2001), except that I specify the form of $f(X_i, \beta)$ and replace $f(X_i, \beta)$ with the derivatives of the moment function. The standard Cramér condition for Edgeworth expansion holds if Assumption 2.4 holds.

The moment conditions in Assumptions 2.2-2.3 that the statements hold for all $0 < q_g, q_G < \infty$ are slightly stronger than necessary, but yield a simpler result. This is also assumed in Andrews (2002) for the same reason.

Theorem 2.1 shows that the misspecification-robust bootstrap symmetric two-sided t test and percentile t interval achieve asymptotic refinements over the asymptotic test and confidence interval. This result is new, because asymptotic refinements for the EL estimators, including the ETEL, have not been established in the existing literature even under correct model specifications.

Recall that the asymptotic test and CI are correct up to $O(n^{-1})$. Theorem 2.1(a) describes the minimum requirement for asymptotic refinements of the bootstrap. This result is comparable to Theorem 3 of Hall and Horowitz (1996), that establishes the same magnitude of asymptotic refinements for GMM estimators. Theorem 2.1(b) is that the bootstrap test and CI achieve the sharp magnitude of asymptotic refinements under more stronger conditions. This result is comparable to Theorem 2(c) of Andrews

(2002) and Theorem 1.1 of Chapter 1, where the former holds under correct model specifications and the latter is robust to model misspecifications.

The degree of asymptotic refinements varies according to q_τ and d. q_τ determines the set that the moment generating function $M_\theta(\lambda) = Ee^{\lambda' g_i}$ exists. d is the number of differentiability of the moment function, $g(X_i, \theta)$.

Theorem 2.1. *Assume that $T(\chi_n)$, $T(\chi_n^*)$, CI_n, and CI_n^* are based on the ETEL estimator.*

(a) *Suppose Assumptions 2.1-2.4 hold with $q_\tau = 32(1+\zeta)$ for any $\zeta > 0$ and $d = 4$. Under $H_0 : \theta_r = \theta_{0,r}$,*

$$P(|T(\chi_n)| > z^*_{|T|,\alpha}) = \alpha + o(n^{-1}) \quad \text{or} \quad P(\theta_{0,r} \in CI_n^*) = 1 - \alpha + o(n^{-1});$$

(b) *Suppose Assumptions 2.1-2.4 hold with all $q_\tau < \infty$ and $d = 6$. Under $H_0 : \theta_r = \theta_{0,r}$,*

$$P(|T(\chi_n)| > z^*_{|T|,\alpha}) = \alpha + O(n^{-2}) \quad \text{or} \quad P(\theta_{0,r} \in CI_n^*) = 1 - \alpha + O(n^{-2}),$$

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

2.6 Monte Carlo Experiments

An Example: Estimating the mean

Let $X_i = (Y_i, Z_i)'$ be iid sample from two independent normal distributions: $Y_i \sim N(1,1)$ and $Z_i \sim N(1-\delta, 1)$, where $0 \leq \delta < 1$. Let

$$g(X_i, \theta) = \begin{pmatrix} \theta Y_i - 1 \\ \theta Z_i - 1 \end{pmatrix}$$

be a vector-valued function. A moment condition is $Eg(X_i, \theta_0) = 0$. In words, the moment condition implies that Y_i and Z_i have the same nonzero and finite mean

θ_0^{-1}. The parameter δ measures a degree of misspecification. $\delta = 0$ implies no misspecification, that is, the two random variables have the same mean. As δ tends to one, EZ_i decreases to zero and the degree of misspecification becomes larger.

The ETEL estimator is calculated by solving the equation (2.5). The estimator does not have a closed form solution, and should be solved numerically. Nevertheless, the probability limit of the estimator can be calculated analytically for this particular example. By definition, $\lambda_0(\theta) \equiv (\lambda_{0,1}(\theta), \lambda_{0,2}(\theta))'$ is the solution to $Ee^{\lambda' g_i} g_i = 0$. The pseudo-true value θ_0 maximizes the criterion

$$\ln L(\theta) \equiv -\ln Ee^{\lambda_0(\theta)'(g_i - Eg_i)}.$$

By solving the equations, we have

$$(\lambda_{0,1}(\theta), \lambda_{0,2}(\theta)) = \left(\frac{1-\theta_0}{\theta_0^2}, \frac{1-(1-\delta)\theta_0}{\theta_0^2}\right), \quad \theta_0 = \frac{2}{2-\delta}.$$

Note that $\theta_0 = 1$ when $\delta = 0$, i.e., the model is correctly specified. In addition, the corresponding $\lambda_{0,1}$ and $\lambda_{0,2}$ become zeros when $\delta = 0$.

Confidence Intervals

In this section, I describe asymptotic and bootstrap CI's based on the ET and ETEL estimators. For brevity, let $L_\theta = 1$.

Conventional CI based on the EL estimator

I first consider the EL estimator, $\hat{\theta}_{EL}$. According to its asymptotic theory, under correct model specifications, we have

$$\sqrt{n}(\hat{\theta}_{EL} - \theta_0) \to_d N(0, (G_0'\Omega_0^{-1}G_0)^{-1}),$$

where $\Omega_0 = Eg_i(\theta_0)g_i(\theta_0)'$ and θ_0 is the true value that satisfies $Eg_i(\theta_0) = 0$. The asymptotic covariance matrix can be estimated by $\hat{\Sigma}_C = (G_n'\Omega_n^{-1}G_n)^{-1}$, where $G_n = n^{-1}\sum_{i=1}^n \hat{G}_i$ and $\Omega_n = n^{-1}\sum_{i=1}^n \hat{g}_i\hat{g}_i'$. Then, the $100(1-\alpha)\%$ asymptotic CI for

θ_0 is given by

$$\text{Asymp.EL: } \left(\hat{\theta}_{EL} \pm z_{\alpha/2}\sqrt{\hat{\Sigma}_C/n}\right).$$

This conventional CI yields correct coverage probability $1-\alpha$ only if the model is correctly specified, because the covariance estimator $\hat{\Sigma}_C$ is not consistent for the true covariance matrix under misspecification.

I also consider the bootstrap CI based on the EL estimator using the efficient bootstrapping suggested by Brown and Newey (2002), that utilizes the EL probability in resampling. Although their paper is on constructing the bootstrap CI based on the GMM estimator, not the EL estimator, it is natural to apply the procedure to the EL estimator. The key procedure is to use the EL-estimated distribution function, $\hat{F}_{EL}(z) = \sum_{i=1}^n \mathbf{1}(X_i \leq z)\hat{p}_i$, where $\hat{p}_i = n^{-1}(1-\hat{\lambda}_{EL}'\hat{g}_i)^{-1}$, in place of the edf, F_n. The bootstrap procedure is as follows: (i) For a given sample, calculate $(\hat{\lambda}_{EL}', \hat{\theta}_{EL}')'$ and $\hat{p}_i$'s, (ii) Draw the bootstrap sample with replacement from $\hat{F}_{EL}$, (iii) Calculate $T_{EL,n}^* \equiv (\hat{\theta}_{EL}^* - \hat{\theta}_{EL})/\sqrt{\hat{\Sigma}_C^*/n}$, where $\hat{\Sigma}_C^* = (G_n^{*\prime}\Omega_n^{*-1}G_n^*)^{-1}$, $G_n^* = n^{-1}\sum_{i=1}^n \hat{G}_i^*$ and $\Omega_n^* = n^{-1}\sum_{i=1}^n \hat{g}_i^*\hat{g}_i^{*\prime}$, (iv) repeat this B times and get the distribution of $|T_{EL,n}^*|$, (iv) Find the $1-\alpha$ quantile $z_{|T_{EL}|,\alpha}^*$ of the distribution of $|T_{EL,n}^*|$ and construct the symmetric percentile t interval,

$$\text{BN-Boot.EL: } \left(\hat{\theta}_{EL} \pm z_{|T_{EL}|,\alpha}^*\sqrt{\hat{\Sigma}_C/n}\right).$$

This CI would yield correct coverage probability $1-\alpha$ only if the model is correctly specified. Under misspecifications, this CI would not work because (i) $\hat{\Sigma}_C$ is inconsistent for the true covariance matrix under misspecification and (ii) $\hat{F}_{EL}$ is inconsistent for the true distribution F. Moreover, in the simulation described in the following section, some of the EL probabilities had negative values under misspecifications, making the EL weighted resampling hard to be implemented.

Misspecification-Robust CI based on the EL estimator

Based on the same EL estimator, we can construct misspecification-robust asymptotic and bootstrap CI's.

$$\text{MR-Asymp.EL:} \qquad \left(\hat{\theta}_{EL} \pm z_{\alpha/2}\sqrt{\hat{\Sigma}_{EL}/n}\right),$$
$$\text{MR-Boot.EL:} \qquad \left(\hat{\theta}_{EL} \pm z^*_{|T|,\alpha}\sqrt{\hat{\Sigma}_{EL}/n}\right),$$

where all the quantities are defined in the previous sections. These CI's are robust to misspecification. Moreover, the bootstrap symmetric percentile t interval *MR-Boot.EL* achieves asymptotic refinements over the asymptotic CI *MR-Asymp.EL*.

Misspecification-Robust CI based on the ETEL estimator

Finally, I consider the asymptotic and the bootstrap CI's based on the ETEL estimator $\hat{\theta}$:

$$\text{MR-Asymp.ETEL:} \qquad \left(\hat{\theta} \pm z_{\alpha/2}\sqrt{\hat{\Sigma}/n}\right),$$
$$\text{MR-Boot.ETEL:} \qquad \left(\hat{\theta} \pm z^*_{|T|,\alpha}\sqrt{\hat{\Sigma}/n}\right),$$

where all the quantities are defined in the previous sections. These CI's are robust to misspecification. Moreover, the bootstrap symmetric percentile t interval *MR-Boot.ETEL* achieves asymptotic refinements over the asymptotic CI *MR-Asymp.ETEL*, according to Theorem 2.1.

Simulation Result

Table 2.1 presents the coverage probabilities of 90% and 95% CI's for the (pseudo-)true value θ_0 based on the ETEL or EL estimator under model misspecifications. δ denotes a degree of misspecification. As δ gets larger, the model becomes more misspecified. $\delta = 0$ implies that the model is correctly specified. Both the numbers of Monte Carlo repetition (r) and the bootstrap repetition (B) are 1,000. The (actual)

Degree of Misspecification	CI's	$n=25$ 90%	$n=25$ 95%	$n=50$ 90%	$n=50$ 95%
$\delta = 0$ (none)	Asymp.EL	0.877	0.919	0.896	0.949
	BN-Boot.EL	0.899	0.954	0.916	0.962
	MR-Asymp.ETEL	0.880	0.925	0.903	0.950
	MR-Boot.ETEL	0.904	0.962	0.914	0.959
over-ID test, 1% level (Rejection Prob.)		1.5%		1.0%	
$\delta = 0.5$ (moderate)	Asymp.EL	0.834	0.886	0.834	0.897
	BN-Boot.EL	0.858	0.909	0.831	0.902
	MR-Asymp.ETEL	0.876	0.912	0.886	0.932
	MR-Boot.ETEL	0.911	0.945	0.909	0.956
over-ID test, 1% level (Rejection Prob.)		22.2%		47.0%	
$\delta = 0.75$ (large)	Asymp.EL	0.748	0.812	0.719	0.789
	BN-Boot.EL	0.717	0.769	0.574	0.642
	MR-Asymp.ETEL	0.820	0.850	0.872	0.915
	MR-Boot.ETEL	0.866	0.911	0.900	0.938
over-ID test, 1% level (Rejection Prob.)		56.8%		88.4%	

Table 2.1: Coverage Probabilities of 90% and 95% Confidence Intervals for θ_0 under Model Misspecifications

coverage probabilities are given by

$$\text{Coverage Probability of CI} = \frac{\sum_r \mathbf{1}(\theta_0 \in CI)}{r},$$

where CI is either *Asymp.EL*, *MR-Asymp.ETEL*, or *MR-Boot.ETEL*. Since the nominal coverage probabilities are given by 90% and 95%, CI's that yield actual coverage probabilities close to 0.90 and 0.95 are favorable.

Consider the case $\delta = 0$. *MR-Boot.ETEL* performs better than *Asymp.ETEL*, especially when $n = 25$. This confirms Theorem 2.1. Although Theorem 2.1 does not establish asymptotic refinements of *BN-Boot.EL*, the simulation result shows that *BN-Boot.EL* performs better than the asymptotic CI based on the EL estimator *Asymp.EL* when $n = 25$. Asymptotic refinements of the bootstrap CI's become less clear when $n = 50$, for both *MR-Boot.ETEL* and *BN-Boot.EL*. If we use a non-Normal distribution for the data-generating process, then asymptotic refinements of the bootstrap would be clearer than those shown in Table 2.1.

When the model is misspecified, $\delta = 0.5$ or $\delta = 0.75$, the CI's based on the conventional EL estimator, *Asymp.EL* and *BN-Boot.EL*, seem to fail to yield correct coverage probabilities asymptotically. In addition, *BN-Boot.EL* loses asymptotic refinements over *Asymp.EL*. In contrast, the CI's based on the ETEL estimator, *MR-Asymp.ETEL* and *MR-Boot.ETEL*, are robust to misspecification as expected. *MR-Boot.ETEL* does not lose the ability of asymptotic refinements even if the model is misspecified. This supports Theorem 2.1.

One might argue that any misspecification in the model could be detected by using the overidentifying restrictions test. I emphasize that the test does not always reject the misspecified model and in fact, the rejection probability is quite low in small sample. The rejection probabilities of the EL overidentifying restrictions test are reported in Table 2.1: *over-ID test, 1% level (Rejection Prob.)*. The null hypothesis of correct model specifications is rejected at 1% level. For instance, when $n = 25$ and $\delta = 0.5$, the probability that the overidentifying restrictions test correctly rejects the misspecified model is 22.2%. When $n = 25$ and $\delta = 0.75$, the chance of rejecting the misspecified model is 56.8%, even if there is a large misspecification. If one proceeds with the misspecified model using the conventional EL estimator, then inferences and CI's would be invalid.

2.7 Appendix: Lemmas and Proofs

Lemmas

Lemma 2.2 is a nonparametric iid bootstrap version of Lemma 1 of Andrews (2002).

Lemma 2.2. *Suppose Assumption 1 holds.*
(a) Let $h(\cdot)$ be a matrix-valued function that satisfies $Eh(X_i) = 0$ and $E\|h(X_i)\|^p < \infty$ for $p \geq 2$ and $p > 2a/(1-2c)$ for some $c \in [0, 1/2)$ and $a \geq 0$. Then, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} n^a P\left(\left\| n^{-1}\sum_{i=1}^{n} h(X_i)\right\| > n^{-c}\varepsilon\right) = 0.$$

(b) Let $h(\cdot)$ be a matrix-valued function that satisfies $E\|h(X_i)\|^p < \infty$ for $p \geq 2$ and $p > 2a$ for some $a \geq 0$. Then, there exists a constant $K < \infty$ such that

$$\lim_{n\to\infty} n^a P\left(\left\| n^{-1}\sum_{i=1}^{n} h(X_i)\right\| > K\right) = 0.$$

In the following Lemmas, we assume that $a = 1$ or 2, because these values are only required in the proofs of the Theorems. This simplifies the conditions for the Lemmas.

Lemma 2.3. *Suppose Assumptions 1-3 hold with $q_\tau > 2a$. Then, for all $\varepsilon > 0$,*

$$(a) \qquad \lim_{n\to\infty} n^a P\left(\sup_{\theta\in\Theta}\left\| n^{-1}\sum_{i=1}^{n}(g_i(\theta) - Eg_i(\theta))\right\| > \varepsilon\right) = 0,$$

$$(b) \qquad \lim_{n\to\infty} n^a P\left(\sup_{\theta\in\Theta}\left| n^{-1}\sum_{i=1}^{n}\left(e^{\hat{\lambda}(\theta)' g_i(\theta)} - Ee^{\lambda_0(\theta)' g_i(\theta)}\right)\right| > \varepsilon\right) = 0,$$

$$(c) \qquad \lim_{n\to\infty} n^a P\left(\sup_{\theta\in\Theta}\left\|\hat{\lambda}(\theta) - \lambda_0(\theta)\right\| > \varepsilon\right) = 0.$$

Let $\tau_0(\theta) \equiv Ee^{\lambda_0(\theta)' g_i(\theta)}$ and $\kappa_0(\theta) \equiv -(Ee^{\lambda_0(\theta)' g_i(\theta)} g_i(\theta) g_i(\theta)')^{-1}\tau_0(\theta) Eg_i(\theta)$. Analogous to the definition of $\Lambda(\theta)$, define $\mathcal{T}(\theta)$ and $\mathcal{K}(\theta)$ be compact sets such that $\tau_0(\theta) \in int(\mathcal{T}(\theta))$ and $\kappa_0(\theta) \in int(\mathcal{K}(\theta))$. For $\beta \equiv (\tau, \kappa', \lambda', \theta')'$, define a compact set $\mathcal{B} \equiv \mathcal{T}(\theta) \times \mathcal{K}(\theta) \times \Lambda(\theta) \times \Theta$.

Lemma 2.4. *Suppose Assumptions 1-3 hold with $q_\tau > 2a(1+\zeta)$ for any $\zeta > 0$. Then, for all $\varepsilon > 0$,*

$$\lim_{n\to\infty} n^a P\left(\|\hat{\beta} - \beta_0\| > \varepsilon\right) = 0,$$

where $\beta_0 = (\tau_0, \kappa_0', \lambda_0', \theta_0')'$ and $\hat{\beta} = (\hat{\tau}, \hat{\kappa}', \hat{\lambda}', \hat{\theta}')'$.

Lemma 2.5. *Suppose Assumptions 1-3 hold with $q_\tau > \frac{2a}{1-2c}(1+\zeta)$ for any $\zeta > 0$. Then, for all $c \in [0, 1/2)$ and $\varepsilon > 0$,*

$$\lim_{n\to\infty} n^a P\left(\|\hat{\beta} - \beta_0\| > n^{-c}\right) = 0,$$

where $\beta_0 = (\tau_0, \kappa_0', \lambda_0', \theta_0')'$ and $\hat{\beta} = (\hat{\tau}, \hat{\kappa}', \hat{\lambda}', \hat{\theta}')'$.

Lemma 2.6 is identical to Andrews (2002). We restate it for convenience.

Lemma 2.6. (a) *Let $\{A_n : n \geq 1\}$ be a sequence of $L_A \times 1$ random vectors with either (i) uniformly bounded densities over $n > 1$ or (ii) an Edgeworth expansion with coefficients of order $O(1)$ and remainder of order $o(n^{-a})$ for some $a \geq 0$ (i.e., for some polynomials $\pi_i(\delta)$ in $\delta = \partial/\partial z$ whose coefficients are $O(1)$ for $i = 1, ..., 2a$, $\lim_{n\to\infty} n^a \sup_{z \in R^{L_A}} |P(A_n \leq z) - [1 + \sum_{i=1}^{2a} n^{-i/2} \pi_i(\partial/\partial z)]\Phi_{\Sigma_n}(z)| = 0$, where $\Phi_{\Sigma_n}(z)$ is the distribution function of a $N(0, \Sigma_n)$ random variable and the eigenvalues of Σ_n are bounded away from 0 and ∞ for $n \geq 1$). Let $\{\xi_n : n \geq 1\}$ be a sequence of $L_A \times 1$ random vectors with $P(\|\xi_n\| > \vartheta_n) = o(n^{-a})$ for some constants $\vartheta_n = o(n^{-a})$ and some $a \geq 0$. Then,*

$$\lim_{n\to\infty} \sup_{z \in R^{L_A}} n^a |P(A_n + \xi_n \leq z) - P(A_n \leq z)| = 0.$$

(b) *Let $\{A_n^* : n \geq 1\}$ be a sequence of $L_A \times 1$ bootstrap random vectors that possesses an Edgeworth expansion with coefficients of order $O(1)$ and remainder of order $o(n^{-a})$ that holds except if $\{\chi_n : n \geq 1\}$ are in a sequence of sets with probability $o(n^{-a})$ for some $a \geq 0$. (That is, for all $\varepsilon > 0$, $\lim_{n\to\infty} n^a P(n^a \sup_{z \in R_A^L} |P^*(A_n^* \leq z) - [1 + \sum_{i=1}^{2a} n^{-i/2} \pi_i^*(\partial/\partial z)]\Phi_{\Sigma_n^*}(z)| > \varepsilon) = 0$, where $\pi_i^*(\delta)$ are polynomials in $\delta = \partial/\partial z$ whose coefficients, C_n^*, satisfy: for* *all $\rho > 0$, there exists $K_\rho < \infty$ such that $\lim_{n\to\infty} n^a P(P^*(|C_n^*| > K_\rho) > \rho) = 0$ for all $i = 1, ..., 2a$, $\Phi_{\Sigma_n^*}(z)$ is the distribution function of a $N(0, \Sigma_n^*)$ random variable conditional on Σ_n^* and Σ_n^* is a possibly random matrix whose eigenvalues are bounded away from 0 and ∞ with probability $1 - o(n^{-a})$ ad $n \to \infty$.) Let $\{\xi_n^* : n \geq 1\}$ be a sequence of $L_A \times 1$ random vectors with $\lim_{n\to\infty} n^a P(P^*(\|\xi_n^*\| > \vartheta_n) > n^{-a}) = 0$ for some constants $\vartheta_n = o(n^{-a})$. Then,*

$$\lim_{n\to\infty} n^a P\left(\sup_{z\in R^{L_A}} |P^*(A_n^* + \xi_n^* \leq z) - P^*(A_n^* \leq z)| > n^{-a}\right) = 0.$$

Let P^* be the probability distribution of the bootstrap sample conditional on the original sample. Let E^* denote expectation with respect to the distribution of the bootstrap sample conditional on the original sample. Since we consider iid sample and nonparametric iid bootstrap, E^* is taken over the original sample with respect to the empirical distribution function. For example, $E^* X_i^* = n^{-1}\sum_{i=1}^n X_i$.

Lemma 2.7 is a nonparametric iid bootstrap version of Lemma 6 of Andrews (2002).

Lemma 2.7. *Suppose Assumption 1 holds. Let $h(\cdot)$ be a matrix-valued function that satisfies $Eh(X_i) = 0$ and $E\|h(X_i)\|^p < \infty$ for $p \geq 2$ and $p > 4a/(1-2c)$ for some $c \in [0, 1/2)$ and $a \geq 0$. Then,*

$$(a) \qquad \lim_{n\to\infty} n^a P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^n h(X_i^*) - E^* h(X_i^*)\right\| > n^{-c}\right) > n^{-a}\right) = 0,$$

$$(b) \qquad \lim_{n\to\infty} n^a P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^n E^* h(X_i^*)\right\| > n^{-c}\right) > n^{-a}\right) = 0,$$

$$(c) \qquad \lim_{n\to\infty} n^a P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^n h(X_i^*)\right\| > n^{-c}\right) > n^{-a}\right) = 0,$$

$$(d) \qquad \lim_{n\to\infty} n^a P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^n E^* h(X_i^*)\right\| > K\right) > n^{-a}\right) = 0 \text{ and}$$

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^n h(X_i^*)\right\| > K\right) > n^{-a}\right) = 0,$$

for some $K < \infty$, even if $Eh(X_i) \neq 0$ and p only satisfies $p \geq 2$ and $p > 4a$ in part (d).

Lemma 2.8. *Suppose Assumptions 1-3 hold with $q_\tau > 4a$. Then, for all $\varepsilon > 0$,*

$$(a) \qquad \lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}\left\|n^{-1}\sum_{i=1}^{n}(g_i^*(\theta) - E^* g_i^*(\theta))\right\| > \varepsilon\right) > n^{-a}\right) = 0,$$

$$(b) \qquad \lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i=1}^{n}\left(e^{\hat\lambda^*(\theta)'g_i^*(\theta)} - e^{\hat\lambda(\theta)'g_i(\theta)}\right)\right| > \varepsilon\right) > n^{-a}\right) = 0,$$

$$(c) \qquad \lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}\left\|\hat\lambda^*(\theta) - \hat\lambda(\theta)\right\| > \varepsilon\right) > n^{-a}\right) = 0.$$

Lemma 2.9. *Suppose Assumptions 1-3 hold with $q_\tau > 4a(1+\zeta)$ for any $\zeta > 0$. Then, for all $\varepsilon > 0$,*

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|\hat\beta^* - \hat\beta\| > \varepsilon\right) > n^{-a}\right) = 0,$$

where $\hat\beta^* = (\hat\tau^*, \hat\kappa^{*\prime}, \hat\lambda^{*\prime}, \hat\theta^{*\prime})'$ and $\hat\beta = (\hat\tau, \hat\kappa', \hat\lambda', \hat\theta')'$.

Lemma 2.10. *Suppose Assumptions 1-3 hold with $q_\tau > \frac{4a}{1-2c}(1+\zeta)$ for any $\zeta > 0$. Then, for all $c \in [0, 1/2)$ and $\varepsilon > 0$,*

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|\hat\beta^* - \hat\beta\| > n^{-c}\right) > n^{-a}\right) = 0,$$

where $\hat\beta^* = (\hat\tau^*, \hat\kappa^{*\prime}, \hat\lambda^{*\prime}, \hat\theta^{*\prime})'$ and $\hat\beta = (\hat\tau, \hat\kappa', \hat\lambda', \hat\theta')'$.

Lemma 2.11. *(a) Suppose Assumptions 1-3 hold with $q_\tau > 4a(1+\zeta)$ for any $\zeta > 0$. Then, for all $\beta \in N(\beta_0)$, some $K < \infty$, and some $C(X_i)$ that satisfies $\lim_{n\to\infty} n^a P(\|n^{-1}\sum_{i=1}^n C(X_i)\| > K) = 0$,*

$$\|f(X_i, \beta) - f(X_i, \beta_0)\| \leq C(X_i)\|\beta - \beta_0\|.$$

(b) Suppose Assumptions 1-3 hold with $q_\tau > 8a(1+\zeta)$ for some $\zeta > 0$. Then, for all

$\beta \in N(\beta_0)$, *some* $K < \infty$, *and some* $C^*(X_i^*)$ *that satisfies*

$$\lim_{n\to\infty} n^a P(P^*(\|n^{-1}\sum_{i=1}^{n} C^*(X_i^*)\| > K) > n^{-a}) = 0,$$

$$\|f(X_i^*, \beta) - f(X_i^*, \beta_0)\| \leq C^*(X_i^*)\|\beta - \beta_0\|.$$

We now introduce some additional notation. Let S_n be the vector containing the unique components of $n^{-1}\sum_{i=1}^{n} f(X_i, \beta_0)$ on the support of X_i, and $S = ES_n$. Similarly, let S_n^* denote the vector containing the unique components of $n^{-1}\sum_{i=1}^{n} f(X_i^*, \hat{\beta})$ on the support of X_i, and $S^* = E^* S_n^*$.

Lemma 2.12. *(a) Suppose Assumptions 1-3 hold with* $q_\tau > 4a(1+\zeta)$ *for any* $\zeta > 0$. *Then, for all* $\varepsilon > 0$,

$$\lim_{n\to\infty} n^a P\left(\|S_n - S\| > \varepsilon\right) = 0.$$

(b) Suppose Assumptions 1-3 hold with $q_\tau > 8a(1+\zeta)$ *for any* $\zeta > 0$. *Then, for all* $\varepsilon > 0$,

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|S_n^* - S^*\| > \varepsilon\right) > n^{-a}\right) = 0.$$

Lemma 2.13. *Let* Δ_n *and* Δ_n^* *denote* $\sqrt{n}(\hat{\theta} - \theta_0)$ *and* $\sqrt{n}(\hat{\theta}^* - \hat{\theta})$, *or* T_n *and* T_n^*. *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 $q_\tau > \max\left\{\frac{2ad}{d-2a-1}, 4a\right\} \cdot (1+\zeta)$ *for any* $\zeta > 0$, *and* $d \geq 2a+2$, *where* $2a$ *is some positive integer. Then,*

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

(b) Suppose Assumptions 1-4 hold with $q_\tau > \max\left\{\frac{4ad}{d-2a-1}, 8a\right\} \cdot (1+\zeta)$ *for any* $\zeta > 0$,

and $d \geq 2a+2$, where $2a$ is some positive integer. Then,

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

We define the components of the Edgeworth expansions of the test statistic T_n and its bootstrap analog T_n^*. Let $\Psi_n = \sqrt{n}(S_n - S)$ and $\Psi_n^* = \sqrt{n}(S_n^* - S^*)$. Let $\Psi_{n,j}$ and $\Psi_{n,j}^*$ denote the jth elements of Ψ_n and Ψ_n^* respectively. Let $\nu_{n,a}$ and $\nu_{n,a}^*$ denote vectors of moments of the form $n^{\alpha(m)} E\Pi_{\mu=1}^m \Psi_{n,j_\mu}$ and $n^{\alpha(m)} E^*\Pi_{\mu=1}^m \Psi_{n,j_\mu}^*$, 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 existence of the limit is proved in Lemma 2.14.

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_n and T_n^* depend on $\pi_i(\delta, \nu_a)$ and $\pi_i(\delta, \nu_{n,a}^*)$, respectively.

Lemma 2.14. *(a) Suppose Assumptions 1-3 hold with $q_\tau \geq 4(a+1)(1+\zeta)$ for any $\zeta > 0$. Then, $\nu_{n,a}$ and $\nu_a \equiv \lim_{n\to\infty} \nu_{n,a}$ exist.*
(b) Suppose Assumptions 1-3 hold with $q_\tau > \frac{16a(a+1)}{1-2\gamma}(1+\zeta)$ for any $\zeta > 0$ and $\gamma \in [0, 1/2)$. Then,

$$\lim_{n\to\infty} n^a P\left(\|\nu_{n,a}^* - \nu_a\| > n^{-\gamma}\right) = 0.$$

Lemma 2.15. *(a) Suppose Assumptions 1-4 hold with $q_\tau \geq 4(a+1)(1+\zeta)$ and $q_\tau > \frac{2ad}{d-2a-1}(1+\zeta)$ for any $\zeta > 0$, and $d \geq 2a+2$, where $2a$ is some positive integer. Then,*

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

(b) Suppose Assumptions 1-4 hold with $q_\tau > \max\left\{\frac{4ad}{d-2a-1}, 16a(a+1)\right\} \cdot (1+\zeta)$ for

any $\zeta > 0$, and $d \geq 2a + 2$, where $2a$ is some positive integer. Then,

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

Proof of Theorem 1

Proof of (a). Let $a = 1$. We apply Lemma 2.14 with $\gamma = 0$ and Lemma 2.15. Letting $d = 4$ and $q_\tau = 32(1+\zeta)$ for all $\zeta > 0$ satisfies the required condition. The proof mimics that of Theorem 2 of Andrews (2001). By the evenness of $\pi_i(\delta, \nu_1)\Phi(z)$ and $\pi_i(\delta, \nu_{n,1}^*)\Phi(z)$, for any $\varepsilon > 0$, Lemma 2.15 yields

$$\sup_{z\in\mathbf{R}}\left|P(|T_n| \leq z) - \left[1 + n^{-1}\pi_2(\delta, \nu_1)\right](\Phi(z) - \Phi(-z))\right| = o(n^{-1}),$$
$$P\left(\sup_{z\in\mathbf{R}}\left|P^*(|T_n^*| \leq z) - \left[1 + n^{-1}\pi_2(\delta, \nu_{n,1}^*)\right](\Phi(z) - \Phi(-z))\right| > n^{-1}\varepsilon\right) = o(n^{-1}).$$

By Lemma 2.14 with $\gamma = 0$,

$$\begin{aligned} & P\left(\sup_{z\in\mathbf{R}}|P(|T_n| \leq z) - P^*(|T_n^*| \leq z)| > n^{-1}\varepsilon\right) \\ \leq\ & P\left(n^{-1}|\pi_2(\delta, \nu_1) - \pi_2(\delta, \nu_{n,1}^*)| > n^{-1}\varepsilon\right) + o(n^{-1}) = o(n^{-1}). \end{aligned} \tag{2.8}$$

Then by (2.8),

$$\begin{aligned} & P\left(\left|P^*(|T_n^*| \leq z_{|T|,\alpha}^*) - P(|T_n| \leq z_{|T|,\alpha}^*)\right| > n^{-1}\varepsilon\right) \\ =\ & P\left(|1 - \alpha - F_{|T|}(z_{|T|,\alpha}^*)| > n^{-1}\varepsilon\right) = o(n^{-1}), \end{aligned} \tag{2.9}$$

where $F_{|T|}(\cdot)$ denote the distribution function of $|T_n|$. Using (2.9), we have

$$\begin{aligned} & P(|T_n > z^*_{|T|,\alpha}) \\ = \; & P\left(F_{|T|}(|T_n|) > F_{|T|}(z^*_{|T|,\alpha}), |1-\alpha - F_{|T|}(z^*_{|T|,\alpha})| \leq n^{-1}\varepsilon\right) \\ & +P\left(F_{|T|}(|T_n|) > F_{|T|}(z^*_{|T|,\alpha}), |1-\alpha - F_{|T|}(z^*_{|T|,\alpha})| > n^{-1}\varepsilon\right) \\ \leq \; & P(F_{|T|}(|T_n|) > 1-\alpha - n^{-1}\varepsilon) + P(|1-\alpha - F_{|T|}(z^*_{|T|,\alpha})| > n^{-1}\varepsilon) \\ \leq \; & \alpha + n^{-1}\varepsilon + o(n^{-1}), \end{aligned} \tag{2.10}$$

and similarly,

$$P(|T_n > z^*_{|T|,\alpha}) \geq P(F_{|T|}(|T_n|) > 1-\alpha - n^{-1}\varepsilon) \geq \alpha - n^{-1}\varepsilon. \tag{2.11}$$

Combining (2.10) and (2.11) establishes the present Theorem (a).

Proof of (b). The proof is the same with that of Theorem 2(c) of Andrews (2002) with his Lemmas 13 and 16 replaced by our Lemmas 2.13 and 2.15. The proof relies on the arguments of Hall (1988) and Hall (1997) developed for "smooth functions of sample averages," for iid data. *Q.E.D.*

Proofs of Lemmas

Proof of Lemma 2.2

Proof. Assumption 1 of Andrews (2002) is satisfied if Assumption 1 holds. Then, Lemma 1 of Andrews (2002) holds. *Q.E.D.*

Proof of Lemma 2.3

Proof. The present Lemma (a) is proved by the proof of Lemma 2 of Hall and Horowitz (1996). We apply Lemma 2.2 with $c = 0$ and $p = q_g$ using Assumption 2(b).

Proving the present Lemma (b) and (c) involves several steps. First, we need to show[2] that

$$\lim_{n\to\infty} n^a P\left(\sup_{(\lambda',\theta')'\in\Lambda(\theta)\times\Theta}\left\|n^{-1}\sum_{i=1}^{n} e^{\lambda' g_i(\theta)} - Ee^{\lambda' g_i(\theta)}\right\| > \varepsilon\right) = 0. \tag{2.12}$$

We apply similar arguments to the proof of Lemma 2 of Hall and Horowitz (1996). Then, Lemma 2.2 with $c = 0$ and $p = q_\tau$ gives the result by Assumption 2(c).

Next, we show

$$\lim_{n\to\infty} n^a P\left(\sup_{\theta\in\Theta}\left\|\bar{\lambda}(\theta) - \lambda_0(\theta)\right\| > \varepsilon\right) = 0, \tag{2.13}$$

where $\bar{\lambda}(\theta) = \arg\min_{\lambda\in\Lambda(\theta)} n^{-1}\sum_{i=1}^{n} e^{\lambda' g_i(\theta)}$. We show (2.13) by using the arguments in the proof of Theorem 10 of Schennach (2007). For a given $\varepsilon > 0$, let

$$\eta = \inf_{\theta\in\Theta} \inf_{\substack{\lambda\in\Lambda(\theta)\\ \|\lambda-\lambda_0(\theta)\|>\varepsilon}} \left(Ee^{\lambda' g_i(\theta)} - Ee^{\lambda_0(\theta)' g_i(\theta)}\right),$$

which is positive by the strict convexity of $Ee^{\lambda' g_i(\theta)}$ in λ, $\lambda_0(\theta) \equiv \arg\min_\lambda Ee^{\lambda' g_i(\theta)}$, and the fact that Θ is compact. By the definition of η,

$$P\left(\sup_{\theta\in\Theta}\left(Ee^{\bar{\lambda}(\theta)' g_i(\theta)} - Ee^{\lambda_0(\theta)' g_i(\theta)}\right) \le \eta\right) \le P\left(\sup_{\theta\in\Theta}\|\bar{\lambda}(\theta) - \lambda_0(\theta)\| \le \varepsilon\right). \tag{2.14}$$

[2] One may write $M_\theta(\theta) \equiv Ee^{\lambda' g_i(\theta)}$ and $\hat{M}_\theta(\theta) \equiv n^{-1}\sum_{i=1}^{n} e^{\lambda' g_i(\theta)}$ to emphasize that $Ee^{\lambda' g_i(\theta)}$ and $n^{-1}\sum_{i=1}^{n} e^{\lambda' g_i(\theta)}$ are functions of λ and θ, and can be interpreted as the population and sample moment generating functions, as in Schennach (2007). In the proof, however, I do not use these notations.

Since $n^{-1}\sum_{i=1}^{n} e^{\bar{\lambda}(\theta)' g_i(\theta)} - n^{-1}\sum_{i=1}^{n} e^{\lambda_0(\theta)' g_i(\theta)} < 0$,

$$\begin{aligned}
& \sup_{\theta\in\Theta}\left(Ee^{\bar{\lambda}(\theta)' g_i(\theta)} - Ee^{\lambda_0(\theta)' g_i(\theta)}\right) \\
\leq\ & \sup_{\theta\in\Theta}\left(Ee^{\bar{\lambda}(\theta)' g_i(\theta)} - n^{-1}\sum_{i=1}^{n} e^{\bar{\lambda}(\theta)' g_i(\theta)}\right) + \sup_{\theta\in\Theta}\left(n^{-1}\sum_{i=1}^{n} e^{\bar{\lambda}(\theta)' g_i(\theta)} - n^{-1}\sum_{i=1}^{n} e^{\lambda_0(\theta)' g_i(\theta)}\right) \\
& + \sup_{\theta\in\Theta}\left(n^{-1}\sum_{i=1}^{n} e^{\lambda_0(\theta)' g_i(\theta)} - Ee^{\lambda_0(\theta)' g_i(\theta)}\right) \\
\leq\ & 2 \sup_{(\lambda',\theta')'\in\Lambda(\theta)\times\Theta}\left|n^{-1}\sum_{i=1}^{n} e^{\lambda' g_i(\theta)} - Ee^{\lambda' g_i(\theta)}\right|.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
& n^a P\left(\sup_{\theta\in\Theta}\left(Ee^{\bar{\lambda}(\theta)' g_i(\theta)} - Ee^{\lambda_0(\theta)' g_i(\theta)}\right) \leq \eta\right) \\
\geq\ & n^a P\left(\sup_{(\lambda',\theta')'\in\Lambda(\theta)\times\Theta}\left|n^{-1}\sum_{i=1}^{n} e^{\lambda' g_i(\theta)} - Ee^{\lambda' g_i(\theta)}\right| \leq \eta/2\right) \to 1,
\end{aligned}$$

by (2.12). Using this result, (2.14) implies (2.13). However, $\bar{\lambda}(\theta)$ is the minimizer on $\Lambda(\theta)$. To get the present Lemma (c) for $\hat{\lambda}(\theta) = \arg\min_{\lambda\in L_g} n^{-1}\sum_{i=1}^{n} e^{\lambda' g_i(\theta)}$, we use the argument similar to the proof of Theorem 2.7 of Newey and McFadden (1994). By using the convexity of $n^{-1}\sum_{i=1}^{n} e^{\lambda' g_i(\theta)}$ in λ for any θ, $\bar{\lambda}(\theta) = \hat{\lambda}(\theta)$ in the event that $\bar{\lambda}(\theta) \in int(\Lambda(\theta))$, where $int(\Lambda(\theta))$ is the interior of $\Lambda(\theta)$. Take a closed neighborhood $\bar{N}_\delta(\lambda_0(\theta))$ of radius δ around $\lambda_0(\theta)$ such that $\bar{N}_\delta(\lambda_0(\theta)) \subset int(\Lambda(\theta))$. Then, whenever $\bar{\lambda}(\theta) \in int(\Lambda(\theta))$, $\|\bar{\lambda}(\theta) - \lambda_0(\theta)\| \leq \delta$ for all $\theta \in \Theta$. Thus,

$$\begin{aligned}
& P\left(\sup_{\theta\in\Theta}\left\|\hat{\lambda}(\theta) - \lambda_0(\theta)\right\| > \varepsilon\right) \\
=\ & P\left(\sup_{\theta\in\Theta}\left\|\hat{\lambda}(\theta) - \lambda_0(\theta)\right\| > \varepsilon,\ \sup_{\theta\in\Theta}\left\|\bar{\lambda}(\theta) - \lambda_0(\theta)\right\| \leq \delta\right) \\
& + P\left(\sup_{\theta\in\Theta}\left\|\hat{\lambda}(\theta) - \lambda_0(\theta)\right\| > \varepsilon,\ \sup_{\theta\in\Theta}\left\|\bar{\lambda}(\theta) - \lambda_0(\theta)\right\| > \delta\right) \\
\leq\ & P\left(\sup_{\theta\in\Theta}\left\|\bar{\lambda}(\theta) - \lambda_0(\theta)\right\| > \varepsilon\right) + P\left(\sup_{\theta\in\Theta}\left\|\bar{\lambda}(\theta) - \lambda_0(\theta)\right\| > \delta\right) = o(n^{-a}),
\end{aligned}$$

by (2.13). This proves the present Lemma (c).

The present Lemma (b) can be shown as follows. By the triangle inequality,

$$\begin{aligned} & \left| n^{-1} \sum_{i=1}^{n} e^{\hat{\lambda}(\theta)' g_i(\theta)} - E e^{\lambda_0(\theta)' g_i(\theta)} \right| \\ \leq & \left| n^{-1} \sum_{i=1}^{n} e^{\hat{\lambda}(\theta)' g_i(\theta)} - n^{-1} \sum_{i=1}^{n} e^{\lambda_0(\theta)' g_i(\theta)} \right| + \left| n^{-1} \sum_{i=1}^{n} e^{\lambda_0(\theta)' g_i(\theta)} - E e^{\lambda_0(\theta)' g_i(\theta)} \right|. \end{aligned}$$

Combining the following results gives the desired result.

$$\lim_{n\to\infty} n^a P\left(\sup_{\theta\in\Theta} \left| n^{-1} \sum_{i=1}^{n} e^{\hat{\lambda}(\theta)' g_i(\theta)} - n^{-1} \sum_{i=1}^{n} e^{\lambda_0(\theta)' g_i(\theta)} \right| > \varepsilon\right) = 0, \qquad (2.15)$$

$$\lim_{n\to\infty} n^a P\left(\sup_{\theta\in\Theta} \left| n^{-1} \sum_{i=1}^{n} e^{\lambda_0(\theta)' g_i(\theta)} - E e^{\lambda_0(\theta)' g_i(\theta)} \right| > \varepsilon\right) = 0. \qquad (2.16)$$

Since $\lambda_0(\theta) \in int(\Lambda(\theta))$, (2.16) follows from (2.12). To show (2.15), we apply the triangle inequality and use the fact that $\bar{\lambda}(\theta) = \hat{\lambda}(\theta)$ in the event that $\bar{\lambda}(\theta) \in int(\Lambda(\theta))$:

$$\begin{aligned} & P\left(\sup_{\theta\in\Theta} \left| n^{-1} \sum_{i=1}^{n} e^{\hat{\lambda}(\theta)' g_i(\theta)} - n^{-1} \sum_{i=1}^{n} e^{\lambda_0(\theta)' g_i(\theta)} \right| > \varepsilon\right) \\ = & P\left(\sup_{\theta\in\Theta} \left| n^{-1} \sum_{i=1}^{n} \left(e^{\hat{\lambda}(\theta)' g_i(\theta)} - e^{\lambda_0(\theta)' g_i(\theta)}\right) \right| > \varepsilon,\ \sup_{\theta\in\Theta} \left\| \bar{\lambda}(\theta) - \lambda_0(\theta) \right\| \leq \delta\right) \\ & + P\left(\sup_{\theta\in\Theta} \left| n^{-1} \sum_{i=1}^{n} \left(e^{\hat{\lambda}(\theta)' g_i(\theta)} - e^{\lambda_0(\theta)' g_i(\theta)}\right) \right| > \varepsilon,\ \sup_{\theta\in\Theta} \left\| \bar{\lambda}(\theta) - \lambda_0(\theta) \right\| > \delta\right) \\ \leq & P\left(\sup_{\theta\in\Theta} \left| n^{-1} \sum_{i=1}^{n} \left(e^{\bar{\lambda}(\theta)' g_i(\theta)} - e^{\lambda_0(\theta)' g_i(\theta)}\right) \right| > \varepsilon\right) + P\left(\sup_{\theta\in\Theta} \left\| \bar{\lambda}(\theta) - \lambda_0(\theta) \right\| > \delta\right) \\ \leq & P\left(\sup_{\theta\in\Theta} \left\| \bar{\lambda}(\theta) - \lambda_0(\theta) \right\| n^{-1} \sum_{i=1}^{n} C_\tau(X_i) > \varepsilon\right) + o(n^{-a}) = o(n^{-a}), \end{aligned}$$

by applying Lemma 2.2(b) with $h(X_i) = C_\tau(X_i)$ and $p = q_\tau$ and using the result (2.13). *Q.E.D.*

Proof of Lemma 2.4

Proof. First, we show

$$\lim_{n\to\infty} n^a P\left(\sup_{\theta\in\Theta} |\ln \hat{L}(\theta) - \ln L(\theta)| > \varepsilon\right) = 0. \tag{2.17}$$

Since

$$\begin{aligned}
\ln \hat{L}(\theta) &= -\ln\left(n^{-1}\sum_{i=1}^{n} e^{\hat{\lambda}(\theta)' g_i(\theta)}\right) + \hat{\lambda}(\theta)' g_n(\theta), \text{ and} \\
\ln L(\theta) &= -\ln\left(E e^{\lambda_0(\theta)' g_i(\theta)}\right) + \lambda_0(\theta)' E g_i(\theta),
\end{aligned}$$

(2.17) follows from

$$\begin{aligned}
&\lim_{n\to\infty} n^a P\left(\sup_{\theta\in\Theta} |n^{-1}\sum_{i=1}^{n} e^{\hat{\lambda}(\theta)' g_i(\theta)} - E e^{\lambda_0(\theta)' g_i(\theta)}| > \varepsilon\right) = 0, \\
&\lim_{n\to\infty} n^a P\left(\sup_{\theta\in\Theta} |\hat{\lambda}(\theta)' g_n(\theta) - \lambda_0(\theta)' E g_i(\theta)| > \varepsilon\right) = 0.
\end{aligned}$$

The first result holds by Lemma 2.3(b). To show the second result, we apply Schwarz matrix inequality[3] to get

$$|\hat{\lambda}(\theta)' g_n(\theta) - \lambda_0(\theta)' E g_i(\theta)| \leq \|\hat{\lambda}(\theta) - \lambda_0(\theta)\| \|g_n(\theta)\| + \|\lambda_0(\theta)\| \|g_n(\theta) - E g_i(\theta)\|.$$

By Lemma 2.3(a), (c), Lemma 2.2(b) with $p = q_g$, and the fact that $\lambda_0(\theta)$ exists for all $\theta \in \Theta$ and Θ is compact, the second conclusion follows.

Since $\ln L(\theta)$ is continuous and uniquely maximized at θ_0, $\forall \varepsilon > 0$, $\exists \eta > 0$ such that $\|\theta - \theta_0\| > \varepsilon$ implies that $\ln L(\theta_0) - \ln L(\theta) \geq \eta > 0$. By the triangle inequality,

[3] See Appendix A of Hansen (2012).

the fact that $\hat\theta$ maximizes $\ln \hat L(\theta)$, and (2.17),

$$\begin{aligned} P\left(\|\hat\theta-\theta_0\|>\varepsilon\right) &\le P\left(\ln L(\theta_0)-\ln L(\hat\theta)\le\eta\right)\\ &= P\left(\ln L(\theta_0)-\ln \hat L(\hat\theta)+\ln \hat L(\hat\theta)-\ln L(\hat\theta)\le\eta\right)\\ &\le P\left(\ln L(\theta_0)-\ln \hat L(\theta_0)+\ln \hat L(\hat\theta)-\ln L(\hat\theta)\le\eta\right)\\ &\le P\left(\sup_{\theta\in\Theta}\left|\ln \hat L(\theta)-\ln L(\theta)\right|\le\eta/2\right)=o(n^{-a}). \end{aligned}$$

Thus, we have

$$\lim_{n\to\infty} n^a P\left(\|\hat\theta-\theta_0\|>\varepsilon\right)=0. \tag{2.18}$$

Next, we show

$$\lim_{n\to\infty} n^a P\left(\|\hat\lambda(\hat\theta)-\lambda_0(\theta_0)\|>\varepsilon\right)=0. \tag{2.19}$$

By the triangle inequality,

$$\begin{aligned} P\left(\|\hat\lambda(\hat\theta)-\lambda_0(\theta_0)\|>\varepsilon\right) &\le P\left(\|\hat\lambda(\hat\theta)-\lambda_0(\hat\theta)\|>\varepsilon/2\right)+P\left(\|\lambda_0(\hat\theta)-\lambda_0(\theta_0)\|>\varepsilon/2\right)\\ &\le P\left(\sup_{\theta\in\Theta}\|\hat\lambda(\theta)-\lambda_0(\theta)\|>\varepsilon/2\right)+o(n^{-a})=o(n^{-a}). \end{aligned}$$

The last equality holds by Lemma 2.3(c). To see $P\left(\|\lambda_0(\hat\theta)-\lambda_0(\theta_0)\|>\varepsilon/2\right)=o(n^{-a})$, let $f(\theta,\lambda)\equiv Ee^{\lambda' g_i(\theta)}g_i(\theta)$, a continuously differentiable function on $(\Theta,\Lambda(\theta))$. Since $f(\theta_0,\lambda_0)=0$ and $(\partial/\partial\lambda')f(\theta_0,\lambda_0)=Ee^{\lambda_0' g_i(\theta_0)}g_i(\theta_0)g_i(\theta_0)'$ is invertible by Assumption 3(a), $\lambda_0(\theta)$ is continuous in a neighborhood of θ_0 by the implicit function theorem. By (2.18), $\hat\theta$ is in a neighborhood of θ_0 with probability $1-o(n^{-a})$ and this implies that $\|\lambda_0(\hat\theta)-\lambda_0(\theta_0)\|\to 0$ with probability $1-o(n^{-a})$. Thus, (2.19) is proved.

Next, we show

$$\lim_{n\to\infty} n^a P\left(\|\hat\tau-\tau_0\|>\varepsilon\right)=0. \tag{2.20}$$

Write $\hat\lambda\equiv\hat\lambda(\hat\theta)$ and $\lambda_0\equiv\lambda_0(\theta_0)$. Since $\hat\tau=n^{-1}\sum_{i=1}^n e^{\hat\lambda' g_i(\hat\theta)}$ and $\tau_0=Ee^{\lambda_0' g_i(\theta_0)}$, (2.20) follows from

$$\lim_{n\to\infty} n^a P\left(\left|n^{-1}\sum_{i=1}^n e^{\hat{\lambda}' g_i(\hat{\theta})} - n^{-1}\sum_{i=1}^n e^{\lambda_0' g_i(\theta_0)}\right| > \varepsilon\right) = 0,$$
$$\lim_{n\to\infty} n^a P\left(\left|n^{-1}\sum_{i=1}^n e^{\lambda_0' g_i(\theta_0)} - E e^{\lambda_0' g_i(\theta_0)}\right| > \varepsilon\right) = 0.$$

To show these results, we apply the argument with the proof of Lemma 2(b) that $\hat{\lambda}(\theta)$ is in the interior of $\Lambda(\theta)$ with probability $1 - o(n^{-a})$. Using this fact, Assumption 2(c), Lemma 2.2(b) with $h(X_i) = C_\tau(X_i)$ and $p = q_\tau$, (2.18), and (2.19) prove the first result. The second result holds by Lemma 2.2(a) with $p = q_\tau$.

Finally, we show

$$\lim_{n\to\infty} n^a P\left(\|\hat{\kappa} - \kappa_0\| > \varepsilon\right) = 0. \tag{2.21}$$

Since

$$\begin{aligned}
\hat{\kappa} &= -\left(n^{-1}\sum_{i=1}^n e^{\hat{\lambda}' g_i(\hat{\theta})} g_i(\hat{\theta}) g_i(\hat{\theta})'\right)^{-1} \hat{\tau} g_n(\hat{\theta}), \text{ and} \\
\kappa_0 &= -\left(E e^{\lambda_0' g_i(\theta_0)} g_i(\theta_0) g_i(\theta_0)'\right)^{-1} \tau_0 E g_i(\theta_0),
\end{aligned}$$

(2.21) follows from

$$\lim_{n\to\infty} n^a P\left(\left\|n^{-1}\sum_{i=1}^n e^{\hat{\lambda}' g_i(\hat{\theta})} g_i(\hat{\theta}) g_i(\hat{\theta})' - n^{-1}\sum_{i=1}^n e^{\lambda_0' g_i(\theta_0)} g_i(\theta_0) g_i(\theta_0)'\right\| > \varepsilon\right) = 0, \tag{2.22}$$
$$\lim_{n\to\infty} n^a P\left(\left\|n^{-1}\sum_{i=1}^n e^{\lambda_0' g_i(\theta_0)} g_i(\theta_0) g_i(\theta_0)' - E e^{\lambda_0' g_i(\theta_0)} g_i(\theta_0) g_i(\theta_0)'\right\| > \varepsilon\right) = 0, \tag{2.23}$$
$$\lim_{n\to\infty} n^a P\left(\|\hat{\tau} - \tau_0\| > \varepsilon\right) = 0, \tag{2.24}$$
$$\lim_{n\to\infty} n^a P\left(\|g_n(\hat{\theta}) - g_n(\theta_0)\| > \varepsilon\right) = 0, \tag{2.25}$$
$$\lim_{n\to\infty} n^a P\left(\|g_n(\theta_0) - E g_i(\theta_0)\| > \varepsilon\right) = 0. \tag{2.26}$$

The third result (2.24) holds by (2.20). The fourth result (2.25) holds by Assumption

2(b), Lemma 2.2(b) with $h(X_i) = C_g(X_i)$ and $p = q_g$, and (2.18). The last result (2.26) holds by Assumption 2(b) and Lemma 2.2(a) with $h(X_i) = g_i(\theta_0) - Eg_i(\theta_0)$, $c = 0$ and $p = q_g$.

To show the first result (2.22), we apply the triangle inequality to get

$$\begin{aligned}
&\left\| n^{-1} \sum_{i=1}^{n} \left(e^{\hat{\lambda}' g_i(\hat{\theta})} g_i(\hat{\theta}) g_i(\hat{\theta})' - e^{\lambda_0' g_i(\theta_0)} g_i(\theta_0) g_i(\theta_0)' \right) \right\| \\
\leq\ & \left\| n^{-1} \sum_{i=1}^{n} \left| e^{\hat{\lambda}' g_i(\hat{\theta})} - e^{\lambda_0' g_i(\theta_0)} \right| g_i(\hat{\theta}) g_i(\hat{\theta})' \right\| \\
& + \left\| n^{-1} \sum_{i=1}^{n} e^{\lambda_0' g_i(\theta_0)} \left(g_i(\hat{\theta}) g_i(\hat{\theta})' - g_i(\theta_0) g_i(\theta_0)' \right) \right\|. \qquad (2.27)
\end{aligned}$$

By Assumption 2(c), with probability $1 - o(n^{-a})$, the first term of (2.27) satisfies

$$\begin{aligned}
&\left\| n^{-1} \sum_{i=1}^{n} \left| e^{\hat{\lambda}' g_i(\hat{\theta})} - e^{\lambda_0' g_i(\theta_0)} \right| g_i(\hat{\theta}) g_i(\hat{\theta})' \right\| \\
\leq\ & \left\| n^{-1} \sum_{i=1}^{n} C_\tau(X_i) g_i(\hat{\theta}) g_i(\hat{\theta})' \right\| \cdot \|(\hat{\lambda}', \hat{\theta}')' - (\lambda_0, \theta_0)'\| = o_p(1).
\end{aligned}$$

The last conclusion $o_p(1)$ holds by (2.18), (2.19), and the fact that

$$\|n^{-1} \sum_{i=1}^{n} C_\tau(X_i) g_i(\hat{\theta}) g_i(\hat{\theta})'\| = O_p(1),$$

with probability $1 - o(n^{-a})$, which can be proved as follows. By the triangle inequality, Schwarz matrix inequality and Assumption 2(b),

$$\begin{aligned}
\|g_i(\hat{\theta}) g_i(\hat{\theta})'\| &\leq \|g_i(\hat{\theta}) g_i(\hat{\theta})' - g_i(\theta_0) g_i(\theta_0)'\| + \|g_i(\theta_0) g_i(\theta_0)'\| \\
&\leq \|(g_i(\hat{\theta}) - g_i(\theta_0))(g_i(\hat{\theta}) + g_i(\theta_0))'\| + \|g_i(\theta_0)\|^2 \\
&\leq \|\hat{\theta} - \theta_0\|^2 C_g^2(X_i) + 2\|\hat{\theta} - \theta_0\| \cdot C_g(X_i) \|g_i(\theta_0)\| + \|g_i(\theta_0)\|^2.
\end{aligned}$$

Using this inequality relation and the triangle inequality,

$$\begin{aligned}\left\|n^{-1}\sum_{i=1}^{n}C_\tau(X_i)g_i(\hat\theta)g_i(\hat\theta)'\right\| &\leq n^{-1}\sum_{i=1}^{n}C_\tau(X_i)\|g_i(\theta_0)\|^2\\ &\quad+\|\hat\theta-\theta_0\|^2 n^{-1}\sum_{i=1}^{n}C_\tau(X_i)C_g^2(X_i)\\ &\quad+2\|\hat\theta-\theta_0\|\cdot n^{-1}\sum_{i=1}^{n}C_\tau(X_i)C_g(X_i)\|g_i(\theta_0)\|.\end{aligned}$$

Then the right-hand-side is $O_p(1)$ with probability $1-o(n^{-a})$ by using (2.18) and applying Lemma 2.2(b) with $h(X_i)=C_\tau(X_i)\|g_i(\theta_0)\|^2$, $h(X_i)=C_\tau(X_i)C_g^2(X_i)$, and $h(X_i)=C_\tau(X_i)C_g(X_i)\|g_i(\theta_0)\|$, provided that $E\|h(X_i)\|^p<\infty$ for some $p\geq 2$ and $p>2a$. In order to satisfy this condition, we use Hölder's inequality: for some $\zeta>0$,

$$\begin{aligned}EC_\tau^p(X_i)\|g_i(\theta_0)\|^{2p} &\leq \left(EC_\tau^{p(1+\zeta)}(X_i)\right)^{(1+\zeta)^{-1}}\cdot\left(E\|g_i(\theta_0)\|^{2p(1+\zeta^{-1})}\right)^{\zeta(1+\zeta)^{-1}},\\ EC_\tau^p(X_i)C_g^{2p}(X_i) &\leq \left(EC_\tau^{p(1+\zeta)}(X_i)\right)^{(1+\zeta)^{-1}}\cdot\left(EC_g^{2p(1+\zeta^{-1})}\right)^{\zeta(1+\zeta)^{-1}},\\ EC_\tau^p(X_i)C_g^p(X_i)\|g_i(\theta_0)\|^p &\leq \left(EC_\tau^{p(1+\zeta)}(X_i)\right)^{(1+\zeta)^{-1}}\cdot\left(EC_g^{2p(1+\zeta^{-1})}\right)^{\frac{\zeta(1+\zeta)^{-1}}{2}}\\ &\quad\cdot\left(E\|g_i(\theta_0)\|^{2p(1+\zeta^{-1})}\right)^{\frac{\zeta(1+\zeta)^{-1}}{2}},\end{aligned}$$

where Cauchy-Schwarz inequality is further applied for the last result. Letting $p(1+\zeta)=q_\tau$ gives $q_\tau\geq 2(1+\zeta)$ and $q_\tau>2a(1+\zeta)$. Letting $2p(1+\zeta^{-1})=q_g$ gives $q_g\geq 4(1+\zeta^{-1})$ and $q_g>4a(1+\zeta^{-1})$. These conditions are assumed in the present Lemma.

The second term of (2.27) can be proved similarly. By the triangle inequality, Schwarz matrix inequality, and Assumption 2(b),

$$\begin{aligned}&\left\|n^{-1}\sum_{i=1}^{n}e^{\lambda_0' g_i(\theta_0)}(g_i(\hat\theta)g_i(\hat\theta)'-g_i(\theta_0)g_i(\theta_0)')\right\|\\ \leq\ &\|\hat\theta-\theta_0\|^2 n^{-1}\sum_{i=1}^{n}e^{\lambda_0' g_i(\theta_0)}C_g^2(X_i)+2\|\hat\theta-\theta_0\|\cdot n^{-1}\sum_{i=1}^{n}e^{\lambda_0' g_i(\theta_0)}C_g(X_i)\|g_i(\theta_0)\|.\end{aligned}$$

Then, (2.18) and Lemma 2.3(b) with

$$h(X_i) = e^{\lambda_0' g_i(\theta_0)} C_g^2(X_i) \text{ and } h(X_i) = e^{\lambda_0' g_i(\theta_0)} C_g(X_i) \|g_i(\theta_0)\|$$

give the desired result. This proves (2.22).

The result (2.23) can be proved by applying Lemma 2.2(a) with $c = 0$ and $h(X_i) = e^{\lambda_0' g_i(\theta_0)} g_i(\theta_0) g_i(\theta_0)' - E e^{\lambda_0' g_i(\theta_0)} g_i(\theta_0) g_i(\theta_0)'$. To see if $h(X_i)$ satisfies the condition of Lemma 2.2(a), it suffices to show $E e^{\lambda_0' g_i(\theta_0) \cdot p} \|g_i(\theta_0)\|^{2p} < \infty$ by Minkowski's inequality. But we already proved that the assumed q_g and q_τ satisfy this condition in the present Lemma. Since we have shown (2.22)-(2.26), the result (2.21) is proved and this completes the proof of the Lemma. *Q.E.D.*

Proof of Lemma 2.5

Proof. $\hat{\beta}$ solves $n^{-1} \sum_{i=1}^n \phi(X_i, \hat{\beta}) = 0$ with probability $1 - o(n^{-a})$, because $\hat{\beta}$ is in the interior of $\mathcal{B}$ with probability $1 - o(n^{-a})$. By the mean value expansion of $n^{-1} \sum_{i=1}^n \phi(X_i, \hat{\beta}) = 0$ around β_0,

$$\hat{\beta} - \beta_0 = -\left(n^{-1} \sum_{i=1}^n \frac{\partial \phi(X_i, \tilde{\beta})}{\partial \beta'} \right)^{-1} n^{-1} \sum_{i=1}^n \phi(X_i, \beta_0),$$

with probability $1 - o(n^{-a})$, where $\tilde{\beta}$ lies between $\hat{\beta}$ and β_0 and may differ across rows. The Lemma follows from

$$\lim_{n \to \infty} n^a P\left(\left\| n^{-1} \sum_{i=1}^n \frac{\partial \phi(X_i, \tilde{\beta})}{\partial \beta'} - n^{-1} \sum_{i=1}^n \frac{\partial \phi(X_i, \beta_0)}{\partial \beta'} \right\| > \varepsilon \right) = 0, \tag{2.28}$$

$$\lim_{n \to \infty} n^a P\left(\left\| n^{-1} \sum_{i=1}^n \frac{\partial \phi(X_i, \beta_0)}{\partial \beta'} - E \frac{\partial \phi(X_i, \beta_0)}{\partial \beta'} \right\| > \varepsilon \right) = 0, \tag{2.29}$$

$$\lim_{n \to \infty} n^a P\left(\left\| n^{-1} \sum_{i=1}^n \phi(X_i, \beta_0) \right\| > n^{-c} \right) = 0. \tag{2.30}$$

First, we prove (2.30). We apply Lemma 2.2(a) with $h(X_i) = \phi(X_i, \beta_0)$. To satisfy the condition of Lemma 2.2(a), we need to show $E\|\phi(X_i, \beta_0)\|^p < \infty$ for $p \geq 2$

and $p > 2a/(1-2c)$. By investigating the elements of $\|\phi(X_i,\beta_0)\|$, it suffices to show $Ee^{\lambda_0' g_i(\theta_0)\cdot p}\|g_i(\theta_0)\|^{2p} < \infty$ and $Ee^{\lambda_0' g_i(\theta_0)\cdot p}\|g_i(\theta_0)\|^p\|G_i(\theta_0)\|^p < \infty$. By Hölder's inequality and Cauchy-Schwarz inequality, we have $q_g \geq 4(1+\zeta^{-1})$ and $q_g > 4a(1-2c)^{-1}(1+\zeta^{-1})$, $q_\tau \geq 2(1+\zeta)$ and $q_\tau > 2a(1-2c)^{-1}(1+\zeta)$, and $q_G \geq 4(1+\zeta^{-1})$ and $q_G > 4a(1-2c)^{-1}(1+\zeta^{-1})$. But this is implied by the assumption of the Lemma.

Second, we prove (2.29). We apply Lemma 2.2(a) with $h(X_i) = (\partial/\partial\beta')\phi(X_i,\beta_0) - E(\partial/\partial\beta')\phi(X_i,\beta_0)$ and $c = 0$. By investigating the elements of $\|(\partial/\partial\beta')\phi(X_i,\beta_0)\|$, it suffices to show $Ee^{\lambda_0' g_i(\theta_0)\cdot p}\|g_i(\theta_0)\|^{3p} < \infty$, $Ee^{\lambda_0' g_i(\theta_0)\cdot p}\|g_i(\theta_0)\|^{2p}\|G_i(\theta_0)\|^p < \infty$, and $Ee^{\lambda_0' g_i(\theta_0)\cdot p}\|g_i(\theta_0)\|^p\|G_i(\theta_0)\|^{2p} < \infty$, to satisfy the condition of Lemma 2.2(a) with $c = 0$. By Hölder's inequality and Cauchy-Schwarz inequality, the corresponding conditions for q_g, q_τ and q_G are $q_g \geq 6(1+\zeta^{-1})$ and $q_g > 6a(1+\zeta^{-1})$, $q_\tau \geq 2(1+\zeta)$ and $q_\tau > 2a(1+\zeta)$, and $q_G \geq 6(1+\zeta^{-1})$ and $q_G > 6a(1+\zeta^{-1})$, which are implied by the assumption of the Lemma.

Finally, (2.28) can be proved by multiple applications of Lemma 2.2(b) and Lemma 2.4. Note that the matrix $(\partial/\partial\beta')\phi(X_i,\beta)$ consists of terms of the form $\alpha\cdot e^{k_\tau\lambda' g_i(\theta)}g(\theta)^{k_0}\cdot G(\theta)^{k_1}\cdot G^{(2)}(\theta)^{k_2}$ for $k_\tau = 0,1$, $k_0 = 0,1,2,3$, $k_1 = 0,1,2$, $k_2 = 0,1$, and $0 \leq k_0+k_1+k_2 \leq 3$, where $g(\theta)$, $G(\theta)$, and $G^{(2)}(\theta)$ denote elements of $g_i(\theta)$, $G_i(\theta)$, and $G_i^{(2)}(\theta)$, respectively, and where α denotes products of elements of β that are necessarily bounded for $\beta \in \mathcal{B}_\delta(\beta_0)$. Thus, $(\partial/\partial\beta')\phi(X_i,\tilde{\beta}) - (\partial/\partial\beta')\phi(X_i,\beta_0)$ consists of terms of the form

$$\begin{aligned}
& \tilde{\alpha}\cdot e^{k_\tau\tilde{\lambda}' g_i(\tilde{\theta})}g(\tilde{\theta})^{k_0}\cdot G(\tilde{\theta})^{k_1}\cdot G^{(2)}(\tilde{\theta})^{k_2} - \alpha_0\cdot e^{k_\tau\lambda_0' g_i(\theta_0)}g(\theta_0)^{k_0}\cdot G(\theta_0)^{k_1}\cdot G^{(2)}(\theta_0)^{k_2} \\
= \; & (\tilde{\alpha}-\alpha_0)\cdot e^{k_\tau\tilde{\lambda}' g_i(\tilde{\theta})}\cdot g(\tilde{\theta})^{k_0}\cdot G(\tilde{\theta})^{k_1}\cdot G^{(2)}(\tilde{\theta})^{k_2} \\
& +\alpha_0\cdot\left(e^{k_\tau\tilde{\lambda}' g_i(\tilde{\theta})} - e^{k_\tau\lambda_0' g_i(\theta_0)}\right)\cdot g(\tilde{\theta})^{k_0}\cdot G(\tilde{\theta})^{k_1}\cdot G^{(2)}(\tilde{\theta})^{k_2} \\
& +\alpha_0\cdot e^{k_\tau\lambda_0' g_i(\theta_0)}\cdot\left(g(\tilde{\theta})^{k_0} - g(\theta_0)^{k_0}\right)\cdot G(\tilde{\theta})^{k_1}\cdot G^{(2)}(\tilde{\theta})^{k_2} \\
& +\alpha_0\cdot e^{k_\tau\lambda_0' g_i(\theta_0)}\cdot g(\theta_0)^{k_0}\left(G(\tilde{\theta})^{k_1} - G(\theta_0)^{k_1}\right)\cdot G^{(2)}(\tilde{\theta})^{k_2} \\
& +\alpha_0\cdot e^{k_\tau\lambda_0' g_i(\theta_0)}\cdot g(\theta_0)^{k_0}\cdot G(\theta_0)^{k_1}\left(G^{(2)}(\tilde{\theta})^{k_2} - G^{(2)}(\theta_0)^{k_2}\right). \qquad (2.31)
\end{aligned}$$

We apply Lemma 2.4 to show $P(|\tilde{\alpha}-\alpha_0| > \varepsilon) = o(n^{-a})$. By Assumption 2(b)-(c), Assumption 3(c), Lemma 2.2(b) and Lemma 2.4,

$$\begin{aligned} & e^{k_\tau \tilde{\lambda}' g_i(\tilde{\theta})} \cdot g(\tilde{\theta})^{k_0} \cdot G(\tilde{\theta})^{k_1} \cdot G^{(2)}(\tilde{\theta})^{k_2} \\ \leq \; & e^{k_\tau \lambda_0' g_i(\theta_0)} \cdot |g(\theta_0)|^{k_0} \cdot |G(\theta_0)|^{k_1} \cdot |G^{(2)}(\theta_0)|^{k_2} + o_p(1). \end{aligned}$$

To apply Lemma 2.2(b) to the first term of the right-hand side of the above inequality, it suffices to show $Ee^{\lambda_0' g_i(\theta_0)\cdot p} \cdot |g(\theta_0)|^{3p} < \infty$, $Ee^{\lambda_0' g_i(\theta_0)\cdot p} \cdot |g(\theta_0)|^{2p} \cdot |G(\theta_0)|^{p} < \infty$, and $Ee^{\lambda_0' g_i(\theta_0)\cdot p} \cdot |g(\theta_0)|^{p} \cdot |G(\theta_0)|^{p} \cdot |G^{(2)}(\theta_0)|^{p} < \infty$ for $p \geq 2$ and $p > 2a$. Since we already showed this in the proof of (2.29), the first term of (2.31) is $o_p(1)$. Similarly, we can show the remaining terms of (2.31) are also $o_p(1)$ by the binomial theorem under the assumption of the Lemma, and thus, (2.28) is proved.

We illustrate the above proof for a term of $(\partial/\partial\beta')\phi(X_i, \tilde{\beta}) - (\partial/\partial\beta')\phi(X_i, \beta_0)$. For example, we show

$$\begin{aligned} \lim_{n\to\infty} n^a P \Bigg(\Bigg\| & n^{-1} \sum_{i=1}^n e^{\tilde{\lambda}(\tilde{\theta})' g_i(\tilde{\theta})} g_i(\tilde{\theta}) g_i(\tilde{\theta})' \tilde{\kappa} g_i(\tilde{\theta})' \\ & - n^{-1} \sum_{i=1}^n e^{\lambda_0(\theta_0)' g_i(\theta_0)} g_i(\theta_0) g_i(\theta_0)' \kappa_0 g_i(\theta_0)' \Bigg\| > \varepsilon \Bigg) = 0, \end{aligned}$$

as follows. By Assumption 2, and the triangle and Schwarz matrix inequalities,

$$\begin{aligned} & \left\| n^{-1} \sum_{i=1}^n e^{\tilde{\lambda}(\tilde{\theta})' g_i(\tilde{\theta})} g_i(\tilde{\theta}) g_i(\tilde{\theta})' \tilde{\kappa} g_i(\tilde{\theta})' - n^{-1} \sum_{i=1}^n e^{\lambda_0(\theta_0)' g_i(\theta_0)} g_i(\theta_0) g_i(\theta_0)' \kappa_0 g_i(\theta_0)' \right\| \\ \leq \; & \|\tilde{\kappa}\| \cdot \|(\tilde{\lambda}', \tilde{\theta}') - (\lambda_0', \theta_0')\| n^{-1} \sum_{i=1}^n C_\tau(X_i) \|g_i(\tilde{\theta})\|^3 \\ & + \|\tilde{\kappa}\| \cdot n^{-1} \sum_{i=1}^n e^{\lambda_0' g_i(\theta_0)} \|g_i(\tilde{\theta}) g_i(\tilde{\theta})' - g_i(\theta_0) g_i(\theta_0)'\| \cdot \|g_i(\tilde{\theta})\| \\ & + \|\tilde{\kappa} - \kappa_0\| \cdot n^{-1} \sum_{i=1}^n e^{\lambda_0' g_i(\theta_0)} \|g_i(\tilde{\theta})\|^2 \cdot \|g_i(\tilde{\theta})\| \\ & + \|\kappa_0\| \|\tilde{\theta} - \theta_0\| n^{-1} \sum_{i=1}^n e^{\lambda_0' g_i(\theta_0)} C_g(X_i) \|g_i(\tilde{\theta})\|^2. \end{aligned}$$

The conclusion follows from multiple applications of Lemma 2.2(b), Lemma 2.4, and the following inequality relations:

$$\begin{aligned}\|\tilde{\kappa}\| &\leq \|\tilde{\kappa}-\kappa_0\|+\|\kappa_0\|,\\ \|g_i(\tilde{\theta})\| &\leq C_g(X_i)\|\tilde{\theta}-\theta_0\|+\|g_i(\theta_0)\|,\\ \|g_i(\tilde{\theta})g_i(\tilde{\theta})'-g_i(\theta_0)g_i(\theta_0)'\| &\leq C_g^2(X_i)\|\tilde{\theta}-\theta_0\|^2+2C_g(X_i)\|\tilde{\theta}-\theta_0\|\cdot\|g_i(\theta_0)\|.\end{aligned}$$

Q.E.D.

Proof of Lemma 2.6

Proof. See the proof of Lemma 5 of Andrews (2002). *Q.E.D.*

Proof of Lemma 2.7

Proof. See the proof of Lemma 6 of Andrews (2002). *Q.E.D.*

Proof of Lemma 2.8

Proof. The result (a) is proved by the proof of Lemma 8 of Hall and Horowitz (1996). We apply Lemma 6 with $c = 0$ and $p = q_g$ using Assumption 2(b).

Proving (b) and (c) involves several steps. We first show

$$\lim_{n\to\infty} n^a P\left(P^*\left(\sup_{(\lambda',\theta')\in\Lambda(\theta)\times\Theta}\left|n^{-1}\sum_{i=1}^{n}\left(e^{\lambda(\theta)'g_i^*(\theta)}-e^{\lambda(\theta)'g_i(\theta)}\right)\right|>\varepsilon\right)>n^{-a}\right)=0. \tag{2.32}$$

The proof of (2.32) is similar to that of Lemma 8 of Hall and Horowitz (1996). We apply our Lemma 2.7(c) with $c = 0$, $h(X_i) = e^{\lambda_j' g_i(\theta_j)} - Ee^{\lambda_j' g_i(\theta_j)}$ for some $(\lambda_j', \theta_j') \in \Lambda(\theta)\times\Theta$ or $h(X_i) = C_\tau(X_i) - EC_\tau(X_i)$, and $p = q_\tau$. Note that q_τ needs to satisfy $q_\tau \geq 2$ and $q_\tau > 4a$. To see this works, observe that $h(X_i^*) = e^{\lambda_j' g_i^*(\theta_j)} - Ee^{\lambda_j' g_i(\theta_j)}$ or $h(X_i^*) = C_\tau(X_i^*) - EC_\tau(X_i)$. Then, $E^*h(X_i^*) = n^{-1}\sum_{i=1}^{n} e^{\lambda_j' g_i(\theta_j)} - Ee^{\lambda_j' g_i(\theta_j)}$ or

$E^*h(X_i^*) = n^{-1}\sum_{i=1}^n C_\tau(X_i) - EC_\tau(X_i)$. Thus,

$$\begin{aligned} h(X_i^*) - E^*h(X_i^*) &= e^{\lambda_j' g_i^*(\theta_j)} - n^{-1}\sum_{i=1}^n e^{\lambda_j' g_i(\theta_j)} \text{ or} \\ h(X_i^*) - E^*h(X_i^*) &= C_\tau(X_i^*) - n^{-1}\sum_{i=1}^n C_\tau(X_i). \end{aligned}$$

Next, we show

$$\lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}\left\|\bar\lambda^*(\theta) - \bar\lambda(\theta)\right\| > \varepsilon\right) > n^{-a}\right) = 0, \tag{2.33}$$

where $\bar\lambda^*(\theta) = \arg\min_{\lambda\in\Lambda(\theta)} n^{-1}\sum_{i=1}^n e^{\lambda' g_i^*(\theta)}$. For a given $\varepsilon > 0$, let

$$\eta = \inf_{\theta\in\Theta} \inf_{\substack{\lambda\in\Lambda(\theta) \\ \|\lambda-\bar\lambda(\theta)\|>\varepsilon}} (n^{-1}\sum_{i=1}^n e^{\lambda' g_i(\theta)} - n^{-1}\sum_{i=1}^n e^{\bar\lambda(\theta)' g_i(\theta)}),$$

which is positive by the strict convexity of $n^{-1}\sum_{i=1}^n e^{\lambda' g_i(\theta)}$ in λ,

$$\bar\lambda(\theta) \equiv \arg\min_{\lambda\in\Lambda(\theta)} n^{-1}\sum_{i=1}^n e^{\lambda' g_i(\theta)},$$

and the fact that Θ is compact. By the definition of η,

$$P^*\left(\sup_{\theta\in\Theta}\left(n^{-1}\sum_{i=1}^n e^{\bar\lambda^*(\theta)' g_i(\theta)} - n^{-1}\sum_{i=1}^n e^{\bar\lambda(\theta)' g_i(\theta)}\right) \le \eta\right) \le P^*\left(\sup_{\theta\in\Theta}\|\bar\lambda^*(\theta) - \bar\lambda(\theta)\| \le \varepsilon\right). \tag{2.34}$$

Since $n^{-1}\sum_{i=1}^n e^{\bar{\lambda}^*(\theta)' g_i^*(\theta)} - n^{-1}\sum_{i=1}^n e^{\bar{\lambda}(\theta)' g_i^*(\theta)} < 0$,

$$\begin{aligned}
& \sup_{\theta\in\Theta}\left(n^{-1}\sum_{i=1}^n e^{\bar{\lambda}^*(\theta)' g_i(\theta)} - n^{-1}\sum_{i=1}^n e^{\bar{\lambda}(\theta)' g_i(\theta)}\right) \\
\leq\ & \sup_{\theta\in\Theta}\left(n^{-1}\sum_{i=1}^n e^{\bar{\lambda}^*(\theta)' g_i(\theta)} - n^{-1}\sum_{i=1}^n e^{\bar{\lambda}^*(\theta)' g_i^*(\theta)}\right) \\
& + \sup_{\theta\in\Theta}\left(n^{-1}\sum_{i=1}^n e^{\bar{\lambda}^*(\theta)' g_i^*(\theta)} - n^{-1}\sum_{i=1}^n e^{\bar{\lambda}(\theta)' g_i^*(\theta)}\right) \\
& + \sup_{\theta\in\Theta}\left(n^{-1}\sum_{i=1}^n e^{\bar{\lambda}(\theta)' g_i^*(\theta)} - n^{-1}\sum_{i=1}^n e^{\bar{\lambda}(\theta)' g_i(\theta)}\right) \\
\leq\ & 2\sup_{(\lambda',\theta')'\in\Lambda(\theta)\times\Theta}\left|n^{-1}\sum_{i=1}^n e^{\lambda' g_i^*(\theta)} - Ee^{\lambda' g_i(\theta)}\right|.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
& P^*\left(\sup_{\theta\in\Theta}\left(n^{-1}\sum_{i=1}^n e^{\bar{\lambda}^*(\theta)' g_i(\theta)} - n^{-1}\sum_{i=1}^n e^{\bar{\lambda}(\theta)' g_i(\theta)}\right) \leq \eta\right) \\
\geq\ & P^*\left(\sup_{(\lambda',\theta')'\in\Lambda(\theta)\times\Theta}\left|n^{-1}\sum_{i=1}^n e^{\lambda' g_i^*(\theta)} - n^{-1}\sum_{i=1}^n e^{\lambda' g_i(\theta)}\right| \leq \eta/2\right),
\end{aligned}$$

and this implies that

$$\begin{aligned}
& P\left(P^*\left(\sup_{\theta\in\Theta}\left(n^{-1}\sum_{i=1}^n e^{\bar{\lambda}^*(\theta)' g_i(\theta)} - n^{-1}\sum_{i=1}^n e^{\bar{\lambda}(\theta)' g_i(\theta)}\right) > \eta\right) > n^{-a}\right) \\
\leq\ & P\left(P^*\left(\sup_{(\lambda',\theta')'\in\Lambda(\theta)\times\Theta}\left|n^{-1}\sum_{i=1}^n e^{\lambda' g_i^*(\theta)} - n^{-1}\sum_{i=1}^n e^{\lambda' g_i(\theta)}\right| > \eta/2\right) > n^{-a}\right) = o(n^{-a}),
\end{aligned}$$

by (2.32). Using this result, (2.34) implies (2.33).

To prove the present Lemma (c), we need to replace $\bar{\lambda}^*(\theta)$ and $\bar{\lambda}(\theta)$ with $\hat{\lambda}^*(\theta)$ and $\hat{\lambda}(\theta)$, respectively. We have shown that $\bar{\lambda}(\theta) = \hat{\lambda}(\theta)$ in the event that $\bar{\lambda}(\theta) \in int(\Lambda(\theta))$ in the proof of Lemma 2.3. By similar argument, $\bar{\lambda}^*(\theta) = \hat{\lambda}^*(\theta)$ in the event that $\bar{\lambda}^*(\theta) \in int(\Lambda(\theta))$. Take a closed neighborhood $\bar{N}_\delta(\lambda_0(\theta))$ of radius δ around $\lambda_0(\theta)$ such that $\bar{N}_\delta(\lambda_0(\theta)) \subset int(\Lambda(\theta))$. Then, whenever $\bar{\lambda}(\theta), \bar{\lambda}^*(\theta) \in int(\Lambda(\theta))$, $\|\bar{\lambda}(\theta) - \lambda_0(\theta)\| \leq \delta$ and $\|\bar{\lambda}^*(\theta) - \lambda_0(\theta)\| \leq \delta$ for all $\theta \in \Theta$. We use this fact to prove the present Lemma (c). By the triangle inequality,

$$\|\bar{\lambda}^*(\theta) - \lambda_0(\theta)\| \leq \|\bar{\lambda}^*(\theta) - \bar{\lambda}(\theta)\| + \|\bar{\lambda}(\theta) - \lambda_0(\theta)\|,$$

and thus we have

$$\begin{aligned}
& P\left(P^*\left(\sup_{\theta\in\Theta}\|\bar{\lambda}^*(\theta) - \lambda_0(\theta)\| > \delta\right) > n^{-a}\right) \\
\leq\ & P\left(P^*\left(\sup_{\theta\in\Theta}\|\bar{\lambda}^*(\theta) - \bar{\lambda}(\theta)\| > \delta/2\right) > n^{-a}\right) \\
& + P\left(P^*\left(\sup_{\theta\in\Theta}\|\bar{\lambda}(\theta) - \lambda_0(\theta)\| > \delta/2\right) > n^{-a}\right) \\
=\ & o(n^{-a}) + o(n^{-a}) = o(n^{-a}), \qquad (2.35)
\end{aligned}$$

by (2.33) and (2.13). Note that $\bar{\lambda}(\theta)$ is calculated from the original sample and thus

$$\begin{aligned}
& P\left(P^*\left(\sup_{\theta\in\Theta}\|\bar{\lambda}(\theta) - \lambda_0(\theta)\| > \delta/2\right) > n^{-a}\right) \\
=\ & P\left(\mathbf{1}\left\{\sup_{\theta\in\Theta}\|\bar{\lambda}(\theta) - \lambda_0(\theta)\| > \delta/2\right\} > n^{-a}\right) \\
=\ & P\left(\mathbf{1}\left\{\sup_{\theta\in\Theta}\|\bar{\lambda}(\theta) - \lambda_0(\theta)\| > \delta/2\right\} > n^{-a},\ \sup_{\theta\in\Theta}\|\bar{\lambda}(\theta) - \lambda_0(\theta)\| > \delta/2\right) \\
& + P\left(\mathbf{1}\left\{\sup_{\theta\in\Theta}\|\bar{\lambda}(\theta) - \lambda_0(\theta)\| > \delta/2\right\} > n^{-a},\ \sup_{\theta\in\Theta}\|\bar{\lambda}(\theta) - \lambda_0(\theta)\| \leq \delta/2\right) \\
\leq\ & P\left(\sup_{\theta\in\Theta}\|\bar{\lambda}(\theta) - \lambda_0(\theta)\| > \delta/2\right) = o(n^{-a}), \qquad (2.36)
\end{aligned}$$

where $\mathbf{1}\{\cdot\}$ is an indicator function. Now, we are ready to prove the present Lemma (c):

$$
\begin{aligned}
& P^*\left(\sup_{\theta\in\Theta}\left\|\hat{\lambda}^*(\theta)-\hat{\lambda}(\theta)\right\|>\varepsilon\right)\\
= \; & P^*\left(\sup_{\theta\in\Theta}\left\|\hat{\lambda}^*(\theta)-\hat{\lambda}(\theta)\right\|>\varepsilon,\ \sup_{\theta\in\Theta}\|\bar{\lambda}^*(\theta)-\lambda_0(\theta)\|>\delta\right)\\
& +P^*\left(\sup_{\theta\in\Theta}\left\|\hat{\lambda}^*(\theta)-\hat{\lambda}(\theta)\right\|>\varepsilon,\ \sup_{\theta\in\Theta}\|\bar{\lambda}^*(\theta)-\lambda_0(\theta)\|\le\delta\right)\\
\le \; & P^*\left(\sup_{\theta\in\Theta}\|\bar{\lambda}^*(\theta)-\lambda_0(\theta)\|>\delta\right)+P^*\left(\sup_{\theta\in\Theta}\|\bar{\lambda}(\theta)-\lambda_0(\theta)\|>\delta\right)\\
& +P^*\left(\sup_{\theta\in\Theta}\left\|\bar{\lambda}^*(\theta)-\bar{\lambda}(\theta)\right\|>\varepsilon\right),
\end{aligned}
$$

and thus,

$$
\begin{aligned}
& P\left(P^*\left(\sup_{\theta\in\Theta}\left\|\hat{\lambda}^*(\theta)-\hat{\lambda}(\theta)\right\|>\varepsilon\right)>n^{-a}\right)\\
\le \; & P\left(P^*\left(\sup_{\theta\in\Theta}\|\bar{\lambda}^*(\theta)-\lambda_0(\theta)\|>\delta\right)>\frac{n^{-a}}{3}\right)\\
& +P\left(P^*\left(\sup_{\theta\in\Theta}\|\bar{\lambda}(\theta)-\lambda_0(\theta)\|>\delta\right)>\frac{n^{-a}}{3}\right)\\
& +P\left(P^*\left(\sup_{\theta\in\Theta}\left\|\bar{\lambda}^*(\theta)-\bar{\lambda}(\theta)\right\|>\varepsilon\right)>\frac{n^{-a}}{3}\right)\\
= \; & o(n^{-a}),
\end{aligned}
$$

by (2.35), (2.36), and (2.33).

Finally, the present Lemma (b) follows from the results below:

$$
\lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i=1}^n\left(e^{\hat{\lambda}^*(\theta)'g_i^*(\theta)}-e^{\hat{\lambda}(\theta)'g_i^*(\theta)}\right)\right|>\varepsilon\right)>n^{-a}\right)=0, \quad (2.37)
$$

$$
\lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i=1}^n\left(e^{\hat{\lambda}(\theta)'g_i^*(\theta)}-e^{\hat{\lambda}(\theta)'g_i(\theta)}\right)\right|>\varepsilon\right)>n^{-a}\right)=0. \quad (2.38)
$$

Since $\hat{\lambda}(\theta)\in int(\Lambda(\theta))$ with probability $1-o(n^{-a})$ as in (2.36), (2.38) follows from

(2.32). To show (2.37), we use the arguments used in the proof of the present Lemma:

$$\begin{aligned}
& P\left(P^*\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i=1}^n\left(e^{\hat\lambda^*(\theta)'g_i^*(\theta)}-e^{\hat\lambda(\theta)'g_i^*(\theta)}\right)\right|>\varepsilon\right)>n^{-a}\right)\\
\leq\ & P\left(P^*\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i=1}^n\left(e^{\bar\lambda^*(\theta)'g_i^*(\theta)}-e^{\bar\lambda(\theta)'g_i^*(\theta)}\right)\right|>\varepsilon\right)>n^{-a}\right)\\
& +P\left(P^*\left(\sup_{\theta\in\Theta}\left\|\bar\lambda^*(\theta)-\lambda_0(\theta)\right\|>\delta\right)>n^{-a}\right)\\
& +P\left(P^*\left(\sup_{\theta\in\Theta}\left\|\bar\lambda(\theta)-\lambda_0(\theta)\right\|>\delta\right)>n^{-a}\right)\\
\leq\ & o(n^{-a})+P\left(P^*\left(\sup_{\theta\in\Theta}\left\|\bar\lambda^*(\theta)-\lambda_0(\theta)\right\|n^{-1}\sum_{i=1}^n C_\tau(X_i^*)>\varepsilon\right)>n^{-a}\right)=o(n^{-a}),
\end{aligned}$$

by applying Lemma 2.7(d) with $h(X_i^*)=C_\tau(X_i^*)$ and $p=q_\tau$ and using (2.33), (2.35), and (2.36). *Q.E.D.*

Proof of Lemma 2.9

Proof. The proof is analogous to that of Lemma 2.4. Though they are similar, the proof involves some additional steps for the bootstrap version of the estimators. First, we show

$$\lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}|\ln\hat L^*(\theta)-\ln\hat L(\theta)|>\varepsilon\right)>n^{-a}\right)=0. \tag{2.39}$$

Since

$$\ln\hat L^*(\theta) \;=\; -\ln\left(n^{-1}\sum_{i=1}^n e^{\hat\lambda^*(\theta)'g_i^*(\theta)}\right)+\hat\lambda^*(\theta)'g_n^*(\theta),$$

(2.39) follows from

$$\begin{aligned}
&\lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}|n^{-1}\sum_{i=1}^n\left(e^{\hat\lambda^*(\theta)'g_i^*(\theta)}-e^{\hat\lambda(\theta)'g_i(\theta)}\right)|>\varepsilon\right)>n^{-a}\right)=0,\\
&\lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}|\hat\lambda^*(\theta)'g_n^*(\theta)-\hat\lambda(\theta)'g_n(\theta)|>\varepsilon\right)>n^{-a}\right)=0.
\end{aligned}$$

Since the assumption of the present Lemma implies that $q_g \geq 2$, $q_g > 4a$, $q_\tau \geq 2$, and $q_\tau > 4a$, the first result holds by Lemma 2.8(b). To show the second result, we apply Schwarz matrix inequality to get

$$\begin{aligned} & \sup_{\theta\in\Theta} |\hat{\lambda}^*(\theta)' g_n^*(\theta) - \hat{\lambda}(\theta)' g_n(\theta)| \\ \leq\ & \sup_{\theta\in\Theta} \|\hat{\lambda}^*(\theta) - \hat{\lambda}(\theta)\| \cdot \sup_{\theta\in\Theta} \|g_n^*(\theta)\| + \sup_{\theta\in\Theta} \|\hat{\lambda}(\theta)\| \cdot \sup_{\theta\in\Theta} \|g_n^*(\theta) - g_n(\theta)\|. \end{aligned}$$

By Lemma 2.8(a), (c), Lemma 2.7(d) with $h(X_i^*) = \sup_{\theta\in\Theta} \|g_i^*(\theta)\|$ and $p = q_g$, the inequality $\|\hat{\lambda}(\theta)\| \leq \|\hat{\lambda}(\theta) - \lambda_0(\theta)\| + \|\lambda_0(\theta)\|$, Lemma 2.3(c), and the fact that $\lambda_0(\theta)$ exists for all $\theta \in \Theta$ and Θ is compact, the second conclusion follows.

Next we claim that given $\varepsilon > 0$, there exists $\eta > 0$ independent of n such that $\|\theta - \hat{\theta}\| > \varepsilon$ implies that $\ln \hat{L}(\hat{\theta}) - \ln \hat{L}(\theta) \geq \eta > 0$ with probability $1 - o(n^{-a})$. To show this claim, define $M = \inf_{\theta \in N_\epsilon(\hat{\theta})^c \cap \Theta} (\ln L(\theta_0) - \ln L(\theta))$, where $N_\epsilon(\hat{\theta}) = \{\theta : \|\theta - \hat{\theta}\| < \varepsilon\}$. Then, $M > 0$ if $\|\hat{\theta} - \theta_0\| \leq \varepsilon$. Now, conditional on the event that $\|\hat{\theta} - \theta_0\| \leq \varepsilon$, $\sup_{\theta\in\Theta} |\ln \hat{L}(\theta) - \ln L(\theta)| \leq M/6$, and $\ln L(\theta_0) - \ln L(\hat{\theta}) \leq M/3$, we have

$$\begin{aligned} \ln \hat{L}(\hat{\theta}) - \ln \hat{L}(\theta) &= \ln L(\theta_0) - \ln L(\theta) + \ln \hat{L}(\hat{\theta}) - \ln \hat{L}(\theta) \\ &\quad - \ln L(\theta_0) + \ln L(\hat{\theta}) + \ln L(\theta) - \ln L(\hat{\theta}) \\ &\geq \ln L(\theta_0) - \ln L(\theta) - |\ln \hat{L}(\hat{\theta}) - \ln L(\hat{\theta}) + \ln L(\theta) - \ln \hat{L}(\theta)| \\ &\quad - (\ln L(\theta_0) - \ln L(\hat{\theta})) \\ &\geq M - 2 \sup_{\theta\in\Theta} |\ln L(\theta) - \ln \hat{L}(\theta)| - (\ln L(\theta_0) - \ln L(\hat{\theta})) \\ &\geq M/3 > 0, \end{aligned}$$

for any $\|\theta - \hat{\theta}\| > \varepsilon$. Since the event occurs with probability $1 - o(n^{-a})$ by Lemma 2.4, this proves the claim.

By the claim, the triangle inequality, the fact that $\hat{\theta}^*$ maximizes $\ln \hat{L}^*(\theta)$, and

(2.39),

$$\begin{aligned}
& P\left(P^*\left(\|\hat{\theta}-\theta_0\| > \varepsilon\right) > n^{-a}\right) \\
\leq\ & P\left(P^*\left(\ln \hat{L}(\hat{\theta}) - \ln \hat{L}(\hat{\theta}^*) \leq \eta\right) > n^{-a}\right) \\
=\ & P\left(P^*\left(\ln \hat{L}(\hat{\theta}) - \ln \hat{L}^*(\hat{\theta}^*) + \ln \hat{L}^*(\hat{\theta}^*) - \ln \hat{L}(\hat{\theta}^*) \leq \eta\right) > n^{-a}\right) \\
\leq\ & P\left(P^*\left(\ln \hat{L}(\hat{\theta}) - \ln \hat{L}^*(\hat{\theta}) + \ln \hat{L}^*(\hat{\theta}^*) - \ln \hat{L}(\hat{\theta}^*) \leq \eta\right) > n^{-a}\right) \\
\leq\ & P\left(P^*\left(\sup_{\theta\in\Theta}\left|\ln \hat{L}^*(\theta) - \ln \hat{L}(\theta)\right| \leq \eta/2\right) > n^{-a}\right) = o(n^{-a}).
\end{aligned}$$

Thus, we have

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|\hat{\theta}^* - \hat{\theta}\| > \varepsilon\right) > n^{-a}\right) = 0. \tag{2.40}$$

Next, we show

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|\hat{\lambda}^*(\hat{\theta}^*) - \hat{\lambda}(\hat{\theta})\| > \varepsilon\right) > n^{-a}\right) = 0. \tag{2.41}$$

By the triangle inequality,

$$\begin{aligned}
& P^*\left(\|\hat{\lambda}^*(\hat{\theta}^*) - \hat{\lambda}(\hat{\theta})\| > \varepsilon\right) \\
\leq\ & P^*\left(\|\hat{\lambda}^*(\hat{\theta}^*) - \hat{\lambda}(\hat{\theta}^*)\| > \frac{\varepsilon}{4}\right) + P^*\left(\|\hat{\lambda}(\hat{\theta}^*) - \lambda_0(\hat{\theta}^*)\| > \frac{\varepsilon}{4}\right) \\
& + P^*\left(\|\lambda_0(\hat{\theta}^*) - \lambda_0(\hat{\theta})\| > \frac{\varepsilon}{4}\right) + P^*\left(\|\lambda_0(\hat{\theta}) - \hat{\lambda}(\hat{\theta})\| > \frac{\varepsilon}{4}\right) \\
\leq\ & P^*\left(\sup_{\theta\in\Theta}\|\hat{\lambda}^*(\theta) - \hat{\lambda}(\theta)\| > \frac{\varepsilon}{4}\right) + P^*\left(\|\lambda_0(\hat{\theta}^*) - \lambda_0(\hat{\theta})\| > \frac{\varepsilon}{4}\right) \\
& + 2P^*\left(\sup_{\theta\in\Theta}\|\lambda_0(\theta) - \hat{\lambda}(\theta)\| > \frac{\varepsilon}{4}\right).
\end{aligned}$$

Since $\|\hat{\theta}^* - \theta_0\| \leq \|\hat{\theta}^* - \hat{\theta}\| + \|\hat{\theta} - \theta_0\|$,

$$\begin{aligned}
& P\left(P^*\left(\|\hat{\theta}^* - \theta_0\| > \varepsilon\right) > n^{-a}\right) \\
\leq\ & P\left(P^*\left(\|\hat{\theta}^* - \hat{\theta}\| > \varepsilon\right) > n^{-a}\right) + P\left(\|\hat{\theta} - \theta_0\| > \varepsilon\right) = o(n^{-a}),
\end{aligned} \tag{2.42}$$

by Lemma 2.4 and (2.40). Therefore, (2.41) follows from Lemma 2.8(c), Lemma 2.3(c), the fact that $\lambda_0(\theta)$ is continuous in a neighborhood of θ_0, and (2.42).

Next, we show

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|\hat{\tau}^* - \hat{\tau}\| > \varepsilon\right) > n^{-a}\right) = 0. \tag{2.43}$$

Since $\hat{\tau}^* = n^{-1}\sum_{i=1}^n e^{\hat{\lambda}^{*\prime} g_i^*(\hat{\theta}^*)}$, where $\hat{\lambda}^* \equiv \hat{\lambda}^*(\hat{\theta}^*)$, (2.43) follows from

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left|n^{-1}\sum_{i=1}^n \left(e^{\hat{\lambda}^{*\prime} g_i^*(\hat{\theta}^*)} - e^{\hat{\lambda}' g_i^*(\hat{\theta})}\right)\right| > \varepsilon\right) > n^{-a}\right) = 0,$$
$$\lim_{n\to\infty} n^a P\left(P^*\left(\left|n^{-1}\sum_{i=1}^n \left(e^{\hat{\lambda}' g_i^*(\hat{\theta})} - e^{\hat{\lambda}' g_i(\hat{\theta})}\right)\right| > \varepsilon\right) > n^{-a}\right) = 0.$$

Since $\hat{\lambda}^*$ is in the interior of $\Lambda(\theta)$ with P^* probability $1 - o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$ by (2.41), we use Assumption 2(c) and apply Lemma 2.7(d) with $h(X_i^*) = C_\tau(X_i^*)$ and $p = q_\tau$ to show the first result. For the second result, we use the triangle inequality to get

$$e^{\hat{\lambda}' g_i^*(\hat{\theta})} - e^{\hat{\lambda}' g_i(\hat{\theta})} \;\leq\; \left|e^{\lambda_0' g_i^*(\theta_0)} - e^{\lambda_0' g_i(\theta_0)}\right| + \left|e^{\hat{\lambda}' g_i^*(\hat{\theta})} - e^{\lambda_0' g_i^*(\theta_0)}\right| + \left|e^{\lambda_0' g_i(\theta_0)} - e^{\hat{\lambda}' g_i(\hat{\theta})}\right|,$$

and apply Assumption 2(c), Lemma 2.7(a) with $h(X_i^*) = e^{\lambda_0' g_i(\theta_0)} - Ee^{\lambda_0' g_i(\theta_0)}$, $c = 0$ and $p = q_\tau$, Lemma 2.7(d) with $h(X_i^*) = C_\tau(X_i^*)$ and $p = q_\tau$, Lemma 2.2(b) with $h(X_i) = C_\tau(X_i)$, and Lemma 2.4.

Finally, we show

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|\hat{\kappa}^* - \hat{\kappa}\| > \varepsilon\right) > n^{-a}\right) = 0. \tag{2.44}$$

Since

$$\hat{\kappa}^* = -\left(n^{-1}\sum_{i=1}^n e^{\hat{\lambda}^{*\prime} g_i^*(\hat{\theta}^*)} g_i^*(\hat{\theta}^*) g_i^*(\hat{\theta}^*)'\right)^{-1} \hat{\tau}^* g_n^*(\hat{\theta}^*),$$

(2.44) follows from

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^n\left(e^{\hat{\lambda}^{*\prime} g_i^*(\hat{\theta}^*)}g_i^*(\hat{\theta}^*)g_i^*(\hat{\theta}^*)' - e^{\hat{\lambda}' g_i^*(\hat{\theta})}g_i^*(\hat{\theta})g_i^*(\hat{\theta})'\right)\right\| > \varepsilon\right) > n^{-a}\right) = 0,$$

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^n\left(e^{\hat{\lambda}' g_i^*(\hat{\theta})}g_i^*(\hat{\theta})g_i^*(\hat{\theta})' - e^{\hat{\lambda}' g_i^*(\hat{\theta})}g_i^*(\hat{\theta})g_i^*(\hat{\theta})'\right)\right\| > \varepsilon\right) > n^{-a}\right) = 0,$$

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|\hat{\tau}^* - \hat{\tau}\| > \varepsilon\right) > n^{-a}\right) = 0,$$
$$\lim_{n\to\infty} n^a P\left(P^*\left(\|g_n^*(\hat{\theta}^*) - g_n^*(\hat{\theta})\| > \varepsilon\right) > n^{-a}\right) = 0,$$
$$\lim_{n\to\infty} n^a P\left(P^*\left(\|g_n^*(\hat{\theta}) - g_n(\hat{\theta})\| > \varepsilon\right) > n^{-a}\right) = 0.$$

Given the results proved in the present Lemma, these results can be proved in a similar fashion with the proof of Lemma 2.4 and thus omitted. In particular, we apply Lemma 2.7 multiple times and we need q_g and q_τ to be such that $q_g \geq 4(1+\zeta^{-1})$, $q_g > 8a(1+\zeta^{-1})$, $q_\tau \geq 2(1+\zeta)$, and $q_\tau > 4a(1+\zeta)$. This is satisfied by the assumption of the Lemma. *Q.E.D.*

Proof of Lemma 2.10

Proof. $\hat{\beta}^*$ solves $n^{-1}\sum_{i=1}^n \phi(X_i^*, \hat{\beta}^*) = 0$ with P^* probability $1 - o(n^{-a})$, except, possibly, if χ is in a set of P probability $o(n^{-a})$, because $\hat{\beta}^*$ is in the interior of $\mathcal{B}$ 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 expansion of $n^{-1}\sum_{i=1}^n \phi(X_i^*, \hat{\beta}^*) = 0$ around $\hat{\beta}$,

$$\hat{\beta}^* - \hat{\beta} = -\left(n^{-1}\sum_{i=1}^n \frac{\partial\phi(X_i^*, \tilde{\beta}^*)}{\partial\beta'}\right)^{-1} n^{-1}\sum_{i=1}^n \phi(X_i^*, \hat{\beta}),$$

with P^* probability $1 - o(n^{-a})$, except, possibly, if χ is in a set of P probability $o(n^{-a})$, where $\tilde{\beta}^*$ lies between $\hat{\beta}^*$ and $\hat{\beta}$ and may differ across rows. The Lemma follows from

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^n\left(\frac{\partial\phi(X_i^*,\tilde\beta^*)}{\partial\beta'}-\frac{\partial\phi(X_i^*,\beta_0)}{\partial\beta'}\right)\right\|>\varepsilon\right)>n^{-a}\right) = 0, \quad (2.45)$$

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^n\left(\frac{\partial\phi(X_i^*,\beta_0)}{\partial\beta'}-E\frac{\partial\phi(X_i,\beta_0)}{\partial\beta'}\right)\right\|>\varepsilon\right)>n^{-a}\right) = 0, \quad (2.46)$$

$$\lim_{n\to\infty} n^a P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^n\phi(X_i^*,\hat\beta)\right\|>n^{-c}\right)>n^{-a}\right)=0. \quad (2.47)$$

First, we show (2.45). The proof is analogous to that of (2.28) in Lemma 2.5. We use Lemmas 2.4, 2.9, 2.7(d), and the inequality relation $\|\tilde\beta^*-\beta_0\|\le\|\hat\beta^*-\hat\beta\|+\|\hat\beta-\beta_0\|$. In order to apply Lemma 2.7(d) instead of Lemma 2.2(b) as in the proof of Lemma 2.5, we impose stronger assumption of q_g, q_G, and q_τ. In particular, $q_g, q_G \ge 6(1+\zeta^{-1})$, $q_g, q_G > 12a(1+\zeta^{-1})$, $q_\tau \ge 2(1+\zeta)$, and $q_\tau \ge 4a(1+\zeta)$. These are implied by the assumption of the Lemma.

The proof of (2.46) is also analogous to that of (2.29) in Lemma 2.5. We apply Lemma 2.7(c) $h(X_i^*) = (\partial/\partial\beta')\phi(X_i^*,\beta_0) - E(\partial/\partial\beta')\phi(X_i,\beta_0)$ and $c=0$. We need $q_g, q_G \ge 6(1+\zeta^{-1})$, $q_g, q_G > 12a(1+\zeta^{-1})$, $q_\tau \ge 2(1+\zeta)$, and $q_\tau \ge 4a(1+\zeta)$, which are implied by the assumption of the Lemma.

Finally, we show (2.47). By the triangle inequality and the mean value expansion,

$$\begin{aligned}\left\|n^{-1}\sum_{i=1}^n\phi(X_i^*,\hat\beta)\right\| &\le \left\|n^{-1}\sum_{i=1}^n\phi(X_i^*,\beta_0)\right\|+\left\|n^{-1}\sum_{i=1}^n\frac{\partial\phi(X_i^*,\bar\beta)}{\partial\beta'}\right\|\cdot\|\hat\beta-\beta_0\|, \text{ and}\\ \left\|n^{-1}\sum_{i=1}^n\frac{\partial\phi(X_i^*,\bar\beta)}{\partial\beta'}\right\| &\le \left\|n^{-1}\sum_{i=1}^n\frac{\partial\phi(X_i^*,\beta_0)}{\partial\beta'}\right\|+\left\|n^{-1}\sum_{i=1}^n\left(\frac{\partial\phi(X_i^*,\bar\beta)}{\partial\beta'}-\frac{\partial\phi(X_i^*,\beta_0)}{\partial\beta'}\right)\right\|,\end{aligned}$$

where $\bar\beta$ lies between $\hat\beta$ and β_0 and may differ across rows. Since $\|\bar\beta-\beta_0\|\le\|\hat\beta-\beta_0\|$, the proof of (2.45) and (2.46) can be applied to show

$$P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^n\frac{\partial\phi(X_i^*,\bar\beta)}{\partial\beta'}\right\|>K\right)>n^{-a}\right)=o(n^{-a}),$$

by Lemma 2.2(b) for some constant $K > 0$ and Lemma 2.4. Now, (2.47) follows from

$$P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^{n}\phi(X_i^*,\beta_0)\right\| > \frac{n^{-c}}{2}\right) > n^{-a}\right) = o(n^{-a}), \tag{2.48}$$

$$P\left(P^*\left(\left\|\hat{\beta}-\beta_0\right\| > \frac{n^{-c}}{2}\right) > n^{-a}\right) = o(n^{-a}). \tag{2.49}$$

To show (2.48), we apply Lemma 2.7(c) with $h(X_i^*) = \phi(X_i^*,\beta_0)$. To satisfy the condition of Lemma 2.7(c), we need to show $E\|\phi(X_i,\beta_0)\|^p < \infty$ for $p \geq 2$ and $p > 4a/(1-2c)$. By investigating the elements of $\|\phi(X_i,\beta_0)\|$, it suffices to show $Ee^{\lambda_0' g_i(\theta_0)\cdot p}\|g_i(\theta_0)\|^{2p} < \infty$ and $Ee^{\lambda_0' g_i(\theta_0)\cdot p}\|g_i(\theta_0)\|^p\|G_i(\theta_0)\|^p < \infty$. By Hölder's inequality and Cauchy-Schwarz inequality, we have $q_g, q_G \geq 4(1+\zeta^{-1})$ and $q_g, q_G > 8a(1-2c)^{-1}(1+\zeta^{-1})$, and $q_\tau \geq 2(1+\zeta)$ and $q_\tau > 4a(1-2c)^{-1}(1+\zeta)$. But this is also implied by the assumption of the Lemma. (2.49) holds by Lemma 2.5 because

$$\begin{aligned}
& P\left(P^*\left(\left\|\hat{\beta}-\beta_0\right\| > \frac{n^{-c}}{2}\right) > n^{-a}\right) \\
= \; & P\left(\mathbf{1}\left\{\left\|\hat{\beta}-\beta_0\right\| > \frac{n^{-c}}{2}\right\} > n^{-a},\; \left\|\hat{\beta}-\beta_0\right\| > \frac{n^{-c}}{2}\right) \\
& + P\left(\mathbf{1}\left\{\left\|\hat{\beta}-\beta_0\right\| > \frac{n^{-c}}{2}\right\} > n^{-a},\; \left\|\hat{\beta}-\beta_0\right\| \leq \frac{n^{-c}}{2}\right) \\
\leq \; & P\left(\left\|\hat{\beta}-\beta_0\right\| > \frac{n^{-c}}{2}\right) = o(n^{-a}).
\end{aligned}$$

Q.E.D.

Proof of Lemma 2.11

Proof of (a). The proof is similar to the proof of (2.28) in Lemma 2.5 and is given by as follows. Since $f(X_i,\beta)$ consists of terms of the form $\alpha \cdot e^{k_\tau \lambda' g_i(\theta)} g(\theta)^{k_0} \cdot$

$G(\theta)^{k_1}\cdots G^{(d)}(\theta)^{k_d}$, $f(X_i,\beta)-f(X_i,\beta_0)$ consists of terms of the form

$$\begin{aligned}
& \alpha\cdot e^{k_\tau\lambda' g_i(\theta)}g(\theta)^{k_0}\cdot G(\theta)^{k_1}\cdots G^{(d+1)}(\theta)^{k_{d+1}} \\
& -\alpha_0\cdot e^{k_\tau\lambda_0' g_i(\theta_0)}g(\theta_0)^{k_0}\cdot G(\theta_0)^{k_1}\cdots G^{(d+1)}(\theta_0)^{k_{d+1}} \\
= & \ (\alpha-\alpha_0)\cdot e^{k_\tau\lambda' g_i(\theta)}\cdot g(\theta)^{k_0}\cdot G(\theta)^{k_1}\cdots G^{(d+1)}(\theta)^{k_{d+1}} \\
& +\alpha_0\cdot\left(e^{k_\tau\lambda' g_i(\theta)}-e^{k_\tau\lambda_0' g_i(\theta_0)}\right)\cdot g(\theta)^{k_0}\cdot G(\theta)^{k_1}\cdots G^{(d+1)}(\theta)^{k_{d+1}} \\
& +\alpha_0\cdot e^{k_\tau\lambda_0' g_i(\theta_0)}\cdot\left(g(\theta)^{k_0}-g(\theta_0)^{k_0}\right)\cdot G(\theta)^{k_1}\cdots G^{(d+1)}(\theta)^{k_{d+1}} \\
& +\alpha_0\cdot e^{k_\tau\lambda_0' g_i(\theta_0)}\cdot g(\theta_0)^{k_0}\left(G(\theta)^{k_1}-G(\theta_0)^{k_1}\right)\cdots G^{(d+1)}(\theta)^{k_{d+1}} \\
& +\alpha_0\cdot e^{k_\tau\lambda_0' g_i(\theta_0)}\cdot g(\theta_0)^{k_0}\cdot G(\theta_0)^{k_1}\cdots\left(G^{(d+1)}(\theta)^{k_{d+1}}-G^{(d+1)}(\theta_0)^{k_{d+1}}\right).
\end{aligned}$$

Since α is a product of elements of β, we can write $(\alpha-\alpha_0)\leq M\|\beta-\beta_0\|$ for some constant $M<\infty$. Let $k_\tau=2$, which is the most restrictive case. By Assumption 2(c),

$$\begin{aligned}
e^{2\lambda' g_i(\theta)}-e^{2\lambda_0' g_i(\theta_0)} &= \left(e^{\lambda' g_i(\theta)}-e^{\lambda_0' g_i(\theta_0)}+e^{\lambda_0' g_i(\theta_0)}\right)^2-e^{2\lambda_0' g_i(\theta_0)} \\
&= \left(e^{\lambda' g_i(\theta)}-e^{\lambda_0' g_i(\theta_0)}\right)^2+2\left(e^{\lambda' g_i(\theta)}-e^{\lambda_0' g_i(\theta_0)}\right)\cdot e^{\lambda_0' g_i(\theta_0)} \\
&\leq C_\tau^2(X_i)\left\|(\lambda',\theta')'-(\lambda_0',\theta_0')'\right\|^2 \\
&\quad +2C_\tau(X_i)\cdot e^{\lambda_0' g_i(\theta_0)}\left\|(\lambda',\theta')'-(\lambda_0',\theta_0')'\right\| \\
&\leq \|\beta-\beta_0\|\cdot\left(C_\tau^2(X_i)\|\beta-\beta_0\|+2C_\tau(X_i)\cdot e^{\lambda_0' g_i(\theta_0)}\right).
\end{aligned}$$

By the binomial theorem,

$$\begin{aligned}
g(\theta)^{k_0}-g(\theta_0)^{k_0} &= \left(g(\theta)-g(\theta_0)+g(\theta_0)\right)^{k_0}-g(\theta_0)^{k_0} \\
&= \sum_{r=1}^{k_0}\binom{k_0}{r}\left(g(\theta)-g(\theta_0)\right)^r g(\theta_0)^{k_0-r} \\
&\leq \sum_{r=1}^{k_0}\binom{k_0}{r}C_g^r(X_i)\cdot\|\theta-\theta_0\|^r\cdot\|g_i(\theta_0)\|^{k_0-r} \\
&\leq \|\beta-\beta_0\|\cdot\sum_{r=1}^{k_0}\binom{k_0}{r}C_g^r(X_i)\cdot\|\theta-\theta_0\|^{r-1}\cdot\|g_i(\theta_0)\|^{k_0-r}.
\end{aligned}$$

Similarly, we can show $G^{(j)}(\theta)^{k_j} - G^{(j)}(\theta_0)^{k_j} \leq \|\beta - \beta_0\| \cdot \tilde{C}^{(j)}(X_i)$ for some $\tilde{C}^{(j)}(X_i)$ and for $j = 1, ..., d$.

Now we can write $f(X_i, \beta) - f(X_i, \beta_0) \leq C(X_i)\|\beta - \beta_0\|$, where

$$\begin{aligned} C(X_i) &= Me^{k_\tau \lambda' g_i(\theta)} \cdots G^{(d+1)}(\theta)^{k_{d+1}} \\ &\quad + \alpha_0 \left(C_\tau^2(X_i)\|\beta - \beta_0\| + 2C_\tau(X_i) \cdot e^{\lambda_0' g_i(\theta_0)} \right) g(\theta)^{k_0} \cdots G^{(d+1)}(\theta)^{k_{d+1}} \\ &\quad + \alpha_0 e^{k_\tau \lambda_0' g_i(\theta_0)} g(\theta_0)^{k_0} \tilde{C}^{(1)}(X_i) + \cdots . \end{aligned}$$

To show $C(X_i)$ is bounded with probability $1 - o(n^{-a})$, as is in the proof of Lemma 2.5, the remaining quantities such as $e^{k_\tau \lambda' g_i(\theta)} \cdot g(\theta)^{k_0} \cdot G(\theta)^{k_1} \cdots G^{(d+1)}(\theta)^{k_{d+1}}$, $C_\tau^{k_\tau}(X_i) C_g^{k_0}(X_i)$, or $C_\tau^{k_\tau}(X_i) C_G^{\sum_{j=1}^{d+1} k_j}(X_i)$, need to be shown to be bounded with probability $1 - o(n^{-a})$ by applying Lemma 2.2(b). It suffices to show the most restrictive cases: $EC_\tau^{2p}(X_i) C_g^{(d+3)p}(X_i) < \infty$ and $EC_\tau^{2p}(X_i) C_G^{(d+3)p}(X_i) < \infty$ for $p \geq 2$ and $p > 2a$. By Hölder's inequality, for $\zeta > 0$,

$$\begin{aligned} EC_\tau^{2p}(X_i) C_g^{(d+3)p}(X_i) &\leq \left(EC_\tau^{2p(1+\zeta)}(X_i) \right)^{(1+\zeta)^{-1}} \cdot \left(EC_g^{(d+3)p(1+\zeta^{-1})}(X_i) \right)^{\zeta(1+\zeta)^{-1}}, \\ EC_\tau^{2p}(X_i) C_G^{(d+3)p}(X_i) &\leq \left(EC_\tau^{2p(1+\zeta)}(X_i) \right)^{(1+\zeta)^{-1}} \cdot \left(EC_G^{(d+3)p(1+\zeta^{-1})}(X_i) \right)^{\zeta(1+\zeta)^{-1}}, \end{aligned}$$

and the assumption on q_g, q_G, and q_τ of the present Lemma satisfy the condition. Thus, the present Lemma (a) is proved.

Proof of (b). The proof is analogous to that of the present Lemma (a), except that we use Lemma 2.7(d) in place of Lemma 2.2(b). Since Lemma 2.7(d) requires $p \geq 2$ and $p > 4a$, we need stronger conditions for q_g, q_G, and q_τ, as are assumed. *Q.E.D.*

Proof of Lemma 2.12

Proof of (a). By the definitions of S_n and S, it suffices to show

$$P\left(\left\| n^{-1} \sum_{i=1}^{n} f(X_i, \beta_0) - Ef(X_i, \beta_0) \right\| > \varepsilon \right) = o(n^{-a}).$$

We apply Lemma 2.2(a) with $c = 0$ and $h(X_i) = f(X_i, \beta_0) - Ef(X_i, \beta_0)$. By investigating the components of $f(X_i, \beta_0)$, the most restrictive condition for Lemma 2.2(a) to hold is $Ee^{\lambda_0' g_i(\theta_0)\cdot 2p} C_g^{(d+3)p}(X_i) < \infty$ and $Ee^{\lambda_0' g_i(\theta_0)\cdot 2p} C_G^{(d+3)p}(X_i) < \infty$, where $p \geq 2$ and $p > 2a$. By Hölder's inequality and Assumptions 2(b)-(c) and 3, we need $q_g, q_G \geq 2(d+3)(1+\zeta^{-1})$, $q_g, q_G > 2a(d+3)(1+\zeta^{-1})$, $q_\tau \geq 4(1+\zeta)$ and $q_\tau > 4a(1+\zeta)$. But these are implied by the assumption of the Lemma.

Proof of (b). By the definitions of S_n^* and S^*, it suffices to show

$$P\left(P^*\left(\left\|n^{-1}\sum_{i=1}^{n} f(X_i^*, \hat{\beta}) - n^{-1}\sum_{i=1}^{n} f(X_i, \hat{\beta})\right\| > \varepsilon\right) > n^{-a}\right) = o(n^{-a}). \tag{2.50}$$

By the triangle inequality,

$$\begin{aligned}\left\|n^{-1}\sum_{i=1}^{n} f(X_i^*, \hat{\beta}) - n^{-1}\sum_{i=1}^{n} f(X_i, \hat{\beta})\right\| \leq\; & \left\|n^{-1}\sum_{i=1}^{n}\left(f(X_i^*, \beta_0) - n^{-1}\sum_{i=1}^{n} f(X_i, \beta_0)\right)\right\| \\ & + n^{-1}\sum_{i=1}^{n}\left\|f(X_i^*, \hat{\beta}) - f(X_i^*, \beta_0)\right\| \\ & + n^{-1}\sum_{i=1}^{n}\left\|f(X_i, \hat{\beta}) - f(X_i, \beta_0)\right\|. \end{aligned} \tag{2.51}$$

For the first term of (2.51), we apply Lemma 2.7(a) with $c = 0$ and $h(X_i) = f(X_i, \beta_0) - Ef(X_i, \beta_0)$, where $q_g, q_G \geq 2(d+3)(1+\zeta^{-1})$, $q_g, q_G > 4a(d+3)(1+\zeta^{-1})$, $q_\tau \geq 4(1+\zeta)$ and $q_\tau > 8a(1+\zeta)$. These are also implied by the assumption of the Lemma. By Lemma 2.11, the second and the last terms of (2.51) are bounded by $\|\hat{\beta} - \beta_0\| n^{-1}\sum_{i=1}^{n}(C^*(X_i^*) + C(X_i))$ and we apply Lemmas 2.4. Note that the assumption of the Lemma satisfies the condition of Lemma 2.4. This proves (2.50) and the result (b) of the present Lemma is proved. *Q.E.D.*

Proof of Lemma 2.13

Proof of (a). First, we prove the result (a) with $\Delta_n = \sqrt{n}(\hat{\theta} - \theta_0)$. Let $\delta_n \equiv \hat{\beta} - \beta_0$ and $\delta_{n,j}$ denote the jth element of δ_n. Write $n^{-1}\sum_{i=1}^{n}\phi(X_i, \beta) \equiv \phi_n(\beta)$ for notational brevity. A Taylor series expansion of $0 = \phi_n(\hat{\beta})$ about $\beta = \beta_0$ through order $d-1$ yields

$$\begin{aligned} 0 &= \phi_n(\beta_0) + \frac{\partial \phi_n(\beta_0)}{\partial \beta'}\delta_n + \frac{1}{2}\sum_{j_1=1}^{L_\beta}\sum_{j_2=1}^{L_\beta}\frac{\partial^2 \phi_n(\beta_0)}{\partial \beta_{j_1}\partial \beta_{j_2}}\delta_{n,j_1}\delta_{n,j_2} + \cdots \\ &\quad + \frac{1}{(d-1)!}\sum_{j_1=1}^{L_\beta}\cdots\sum_{j_{d-1}=1}^{L_\beta}\frac{\partial^{d-1}\phi_n(\beta_0)}{\partial \beta_{j_1}\cdots\partial \beta_{j_{d-1}}}\delta_{n,j_1}\cdots\delta_{n,j_{d-1}} + \xi_n, \end{aligned} \tag{2.52}$$

with probability $1 - o(n^{-a})$, where $L_\beta = 1 + 2L_g + L_\theta$ and

$$\xi_n = \frac{1}{(d-1)!}\sum_{j_1=1}^{L_\beta}\cdots\sum_{j_{d-1}=1}^{L_\beta}\left(\frac{\partial^{d-1}\phi_n(\tilde{\beta})}{\partial \beta_{j_1}\cdots\partial \beta_{j_{d-1}}} - \frac{\partial^{d-1}\phi_n(\beta_0)}{\partial \beta_{j_1}\cdots\partial \beta_{j_{d-1}}}\right)\delta_{n,j_1}\cdots\delta_{n,j_{d-1}},$$

where $\tilde{\beta}$ is between $\hat{\beta}$ and β_0 and may differ across rows. Let e_n be the conformable vector $(\xi_n', \mathbf{0}')'$ such that the dimension of e_n is the same with that of S_n. Then, (2.52) can be rewritten as $0 = \Xi(\delta_n, S_n + e_n)$, where $\Xi(\cdot,\cdot)$ is a polynomial and thus, infinitely differentiable with respect to its arguments. By Lemmas 2.4 and 2.12(a), δ_n and S_n converge to 0 and S with probability $1 - o(n^{-a})$. Since $n^{-1}\sum_{i=1}^{n} f(X_i, \beta)$ includes elements of $(\partial^{d-1}/\partial\beta_{j_1}\cdots\beta_{j_{d-1}})\phi_n(\beta)$, e_n converges to zero with probability $1-o(n^{-a})$ by Lemma 2.11(a). Thus, we have $0 = \Xi(0, S)$. Let $\delta \equiv \beta - \beta_0$. If we differentiate Ξ with respect to its first argument and evaluate it at $\delta = 0$, we get $(\partial/\partial\beta')\phi_n(\beta_0)$, the inverse of which exists and bounded with probability $1 - o(n^{-a})$ by (2.29) and Assumption 3(a). Note that the condition for (2.29) to hold is $q_g, q_G \geq 6(1+\zeta^{-1})$, $q_g, q_G \geq 6a(1+\zeta^{-1})$, $q_\tau \geq 2(1+\zeta)$, and $q_\tau > 2a(1+\zeta)$. These are implied by the assumption of the present Lemma.

By applying the implicit function theorem to $\Xi(\delta_n, S_n + e_n)$, there is a function A_1 such that $A_1(S) = 0$, A_1 is infinitely differentiable in a neighborhood of S, and

$$\delta_n \equiv \hat{\beta} - \beta_0 = A_1(S_n + e_n), \tag{2.53}$$

with probability $1-o(n^{-a})$. By Lemma 2.11(a), $\|e_n\| \leq M\|\hat{\beta}-\beta_0\|^d$ for some $M < \infty$, with probability $1 - o(n^{-a})$. By Lemma 2.5, $\|\hat{\beta} - \beta_0\|^d \leq n^{-dc}$ with probability

$1-o(n^{-a})$. The condition for Lemma 2.5 is implied by the assumption of the present Lemma. By the mean value theorem, $A_1(S_n+e_n) = A_1(S_n) + A_1'(S_n+\tilde{e}_n)\cdot e_n$, where $A_1'(\cdot)$ is the first derivative of $A_1(\cdot)$ and $\tilde{e}_n$ lies between $\mathbf{0}'$ and e_n and may differ across rows. Since $A_1'(S_n+\tilde{e}_n) = A_1'(S) + A_1'(S_n+\tilde{e}_n) - A_1'(S)$, A_1' is continuous, and S_n and $\tilde{e}_n$ converges to S and 0 respectively with probability $1-o(n^{-a})$ by Lemmas 2.11(a) and 2.12(a), it follows that

$$P\left(\|A_1(S_n+e_n) - A_1(S_n)\| > n^{-dc}\right) \leq P\left(\tilde{M}\cdot\|e_n\| > n^{-dc}\right) = o(n^{-a}), \tag{2.54}$$

for some $\tilde{M} < \infty$. Since (2.54) holds, we can apply Lemma 2.6(a) if (i) $dc \geq a + 1/2$ for some $c \in [0, 1/2)$, and (ii) $2a$ is a nonnegative integer. (ii) is assumed. Let $d = 2a + 1 + \bar{d}$, where $\bar{d} \geq 1$ is some integer and let $c = (2a+1)(4a+2+2\bar{d})^{-1}$. Then $d \geq 2a+2$ and $dc = a + 1/2$, so that (i) is satisfied. Note that $(1-2c)^{-1} = (2a+1+\bar{d})\bar{d}^{-1} = d(d-2a-1)^{-1}$ with the defined value of c.[4] This term replaces the term $1-2c$ appears in the condition of Lemma 2.5. By Lemma 2.6(a),

$$\lim_{n\to\infty} \sup_{z\in R^{L_\beta}} n^a \left|P\left(\sqrt{n}(\hat{\beta}-\beta_0) \leq z\right) - P\left(\sqrt{n}A_1(S_n) \leq z\right)\right| = 0,$$

and the present Lemma (a) with $\Delta_n = \sqrt{n}(\hat{\theta}-\theta_0)$ holds because Δ_n is a subvector of $\sqrt{n}(\hat{\beta}-\beta_0)$.

Next, we prove the present Lemma (a) with $\Delta_n = T_n \equiv \sqrt{n}(\hat{\theta}_r - \theta_{0,r})/\sqrt{\hat{\Sigma}_{rr}}$. We use the fact that $\hat{\Sigma}$ is a function of β. Define

$$\hat{\Sigma}(\beta) \equiv \left(\frac{\partial \phi_n(\beta)}{\partial \beta'}\right)^{-1} n^{-1}\sum_{i=1}^{n} \phi(X_i,\beta)\phi(X_i,\beta)' \left(\frac{\partial \phi_n(\beta)}{\partial \beta}\right)^{-1},$$

and $H_n(\beta) \equiv (\beta_r - \beta_{0,r})(\hat{\Sigma}_{rr}(\beta))^{-1/2}$ for some r that corresponds to an element of θ.

[4]There is a tradeoff relationship between the value of $\bar{d}$ and c. If we assume infinitely differentiable $g(X_i,\theta)$, then c could be arbitrarily small positive number and this weakens the condition on q_τ. Thus, the tradeoff between $\bar{d}$ and c can be interpreted as a tradeoff between smoothness of the moment function and the existence of the higher moment of $C_\tau(X_i)$.

Consider a Taylor expansion of $H_n(\hat{\beta})$ around β_0, through order $d-1$:

$$\begin{aligned} H_n(\hat{\beta}) &= 0 + \left(\hat{\Sigma}_{rr}(\beta_0)\right)^{-\frac{1}{2}} \delta_{n,r} + \frac{1}{2}\sum_{j_1=1}^{L_\beta}\sum_{j_2=1}^{L_\beta} \frac{\partial^2 H_n(\beta_0)}{\partial\beta_{j_1}\partial\beta_{j_2}}\delta_{n,j_1}\delta_{n,j_2} + \cdots \\ &\quad + \frac{1}{(d-1)!}\sum_{j_1=1}^{L_\beta}\cdots\sum_{j_{d-1}=1}^{L_\beta} \frac{\partial^{d-1} H_n(\beta_0)}{\partial\beta_{j_1}\cdots\partial\beta_{j_{d-1}}}\delta_{n,j_1}\cdots\delta_{n,j_{d-1}} + \eta_n, \end{aligned}$$

where

$$\eta_n = \frac{1}{(d-1)!}\sum_{j_1=1}^{L_\beta}\cdots\sum_{j_{d-1}=1}^{L_\beta}\left(\frac{\partial^{d-1} H_n(\bar{\beta})}{\partial\beta_{j_1}\cdots\partial\beta_{j_{d-1}}} - \frac{\partial^{d-1} H_n(\beta_0)}{\partial\beta_{j_1}\cdots\partial\beta_{j_{d-1}}}\right)\delta_{n,j_1}\cdots\delta_{n,j_{d-1}},$$

where $\bar{\beta}$ is between $\hat{\beta}$ and β_0 and may differ across rows. Since $H_n(\beta)$ and its derivatives through order $d-1$ with respect to the components of β are continuous functions of terms of $n^{-1}\sum_{i=1}^n f(X_i, \beta)$, we can write $H_n(\hat{\beta}) = A_2(S_n, \delta_n) + \eta_n$, where $A_2(\cdot)$ is infinitely differentiable and $A_2(S, 0) = 0$. By (2.53), $A_2(S_n, \delta_n) = A_2(S_n, A_1(S_n + e_n))$. By the mean value expansion, (2.54), and Lemma 2.11(a),

$$A_2(S_n, A_1(S_n + e_n)) - A_2(S_n, A_1(S_n)) \leq M_1 \cdot \|e_n\| \leq M_2 \cdot \|\hat{\beta} - \beta_0\|^d, \qquad (2.55)$$

for some $M_1, M_2 < \infty$, with probability $1 - o(n^{-a})$. Define $A_3(S_n) \equiv A_2(S_n, A_1(S_n))$, so that $A_3(\cdot)$ is infinitely differentiable and $A_3(S) = 0$. Since $\|\eta_n\| < \bar{M}\|\hat{\beta} - \beta_0\|^d$ for some $\bar{M} < \infty$ with probability $1 - o(n^{-a})$ by Lemma 2.11(a), combining this with (2.55) yields $|H_n(\hat{\beta}) - A_3(S_n)| \leq M_3\|\hat{\beta} - \beta_0\|^d$ for some $M_3 < \infty$ with probability $1 - o(n^{-a})$. By Lemma 2.5, we have

$$P\left(\|H_n(\hat{\beta}) - A_3(S_n)\| > n^{-dc}\right) = o(n^{-a}). \qquad (2.56)$$

Since the assumed condition for d and c satisfies the condition of Lemma 2.6, we apply Lemma 2.6 with (2.56) to get our final conclusion:

$$\lim_{n\to\infty}\sup_{z\in R} n^a \left|P\left(T_n \leq z\right) - P\left(\sqrt{n}A_3(S_n) \leq z\right)\right| = 0.$$

Proof of (b). First, we prove the result (b) with $\Delta_n^* = \sqrt{n}(\hat{\theta}^* - \hat{\theta})$. Let $\delta_n^* \equiv \hat{\beta}^* - \hat{\beta}$. Write $n^{-1}\sum_{i=1}^n \phi(X_i^*, \beta) \equiv \phi_n^*(\beta)$ for notational brevity. A Taylor series expansion of the bootstrap first-order condition $0 = \phi_n^*(\hat{\beta}^*)$ about $\beta = \hat{\beta}$ through order $d-1$ yields

$$\begin{aligned} 0 &= \phi_n^*(\hat{\beta}) + \frac{\partial \phi_n^*(\hat{\beta})}{\partial \beta'}\delta_n^* + \frac{1}{2}\sum_{j_1=1}^{L_\beta}\sum_{j_2=1}^{L_\beta}\frac{\partial^2 \phi_n^*(\hat{\beta})}{\partial \beta_{j_1}\partial \beta_{j_2}}\delta_{n,j_1}^*\delta_{n,j_2}^* + \cdots \\ &\quad + \frac{1}{(d-1)!}\sum_{j_1=1}^{L_\beta}\cdots\sum_{j_{d-1}=1}^{L_\beta}\frac{\partial^{d-1}\phi_n^*(\hat{\beta})}{\partial \beta_{j_1}\cdots\partial \beta_{j_{d-1}}}\delta_{n,j_1}^*\cdots\delta_{n,j_{d-1}}^* + \xi_n^*, \end{aligned} \tag{2.57}$$

with P^* probability $1-o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$, where

$$\xi_n^* = \frac{1}{(d-1)!}\sum_{j_1=1}^{L_\beta}\cdots\sum_{j_{d-1}=1}^{L_\beta}\left(\frac{\partial^{d-1}\phi_n^*(\tilde{\beta}^*)}{\partial \beta_{j_1}\cdots\partial \beta_{j_{d-1}}} - \frac{\partial^{d-1}\phi_n^*(\hat{\beta})}{\partial \beta_{j_1}\cdots\partial \beta_{j_{d-1}}}\right)\delta_{n,j_1}^*\cdots\delta_{n,j_{d-1}}^*,$$

where $\tilde{\beta}^*$ is between $\hat{\beta}^*$ and $\hat{\beta}$ and may differ across rows. Let e_n^* be the conformable vector $(\xi_n^{*\prime}, \mathbf{0}')'$ as in the proof of the present Lemma (a). Since all the terms of (2.57) are the same with those of (2.52) by replacing S_n and β_0 with S_n^* and $\hat{\beta}$, respectively, we have $0 = \Xi(\delta_n^*, S_n^* + e_n^*)$, where $\Xi(\cdot,\cdot)$ is the same with that in the proof of the present Lemma (a). By Lemmas 2.9 and 2.12(b), δ_n^* and S_n^* converge to 0 and S^* with P^* probability $1-o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$. To show $P(P^*(\|e_n^*\| > \varepsilon) > n^{-a}) = o(n^{-a})$, we use the triangle inequality and Lemma 2.11(b),

$$\begin{aligned} \frac{\partial^{d-1}\phi_n^*(\tilde{\beta}^*)}{\partial \beta_{j_1}\cdots\partial \beta_{j_{d-1}}} - \frac{\partial^{d-1}\phi_n^*(\hat{\beta})}{\partial \beta_{j_1}\cdots\partial \beta_{j_{d-1}}} &\leq \left|\frac{\partial^{d-1}\phi_n^*(\hat{\beta})}{\partial \beta_{j_1}\cdots\partial \beta_{j_{d-1}}} - \frac{\partial^{d-1}\phi_n^*(\beta_0)}{\partial \beta_{j_1}\cdots\partial \beta_{j_{d-1}}}\right| \\ &\quad + \left|\frac{\partial^{d-1}\phi_n^*(\tilde{\beta}^*)}{\partial \beta_{j_1}\cdots\partial \beta_{j_{d-1}}} - \frac{\partial^{d-1}\phi_n^*(\beta_0)}{\partial \beta_{j_1}\cdots\partial \beta_{j_{d-1}}}\right| \\ &\leq n^{-1}\sum_{i=1}^n C^*(X_i^*)\left(\|\hat{\beta}-\beta_0\| + \|\tilde{\beta}^*-\beta_0\|\right), \end{aligned} \tag{2.58}$$

provided that $\tilde{\beta}^*$ is in $N(\beta_0)$ with P^* probability $1 - o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$, that holds by Lemma 2.9. By Lemmas 2.4 and 2.9, we prove the desired result.

Thus, we have $0 = \Xi(0, S^*)$. Let $\delta^* \equiv \beta - \hat{\beta}$. If we differentiate Ξ with respect to its first argument and evaluate it at $\delta^* = 0$, we get $(\partial/\partial\beta')\phi_n^*(\hat{\beta})$, the inverse of which exists and bounded with P^* probability $1 - o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$ by (2.46) and Assumption 3(a). The conditions for (2.46) to hold are implied by the assumption of the present Lemma.

As in the proof of the present Lemma (a), $A_1(S^*) = 0$, A_1 is infinitely differentiable in a neighborhood of S^*, and

$$\delta_n^* \equiv \hat{\beta}^* - \hat{\beta} = A_1(S_n^* + e_n^*), \tag{2.59}$$

with P^* probability $1 - o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$. Next, we show $P(P^*(\|e_n^*\| > n^{-dc}) > n^{-a}) = o(n^{-a})$. Conditional on the sample χ, by (2.58), for some $M^* < \infty$,

$$\begin{aligned} P^*(\|e_n^*\| > n^{-dc}) \;\; &\leq \;\; P^*\left(M^*\|\hat{\beta} - \beta_0\| \cdot \|\hat{\beta}^* - \hat{\beta}\|^{d-1} > n^{-dc})\right) \\ &\quad\; + P^*\left(M^*\|\hat{\beta}^* - \hat{\beta}\|^d > n^{-dc})\right) \\ &\leq \;\; P^*\left(M^*\|\hat{\beta}^* - \hat{\beta}\|^{d-1} > n^{-(d-1)c})\right) + P^*\left(\|\hat{\beta} - \beta_0\| > n^{-c})\right) \\ &\quad\; + P^*\left(M^*\|\hat{\beta}^* - \hat{\beta}\|^d > n^{-dc})\right). \end{aligned}$$

By Lemmas 2.5 and 2.10, the desired result is proved. Therefore, analogous arguments as in the proof of the present Lemma (a) yield

$$\begin{aligned} P\left(P^*\left(\|A_1(S_n^* + e_n^*) - A_1(S_n^*)\| > n^{-dc}\right) > n^{-a}\right) \;\; &\leq \;\; P\left(P^*\left(\tilde{M}^*\|e_n\| > n^{-dc}\right) > n^{-a}\right) \\ &= \;\; o(n^{-a}), \end{aligned} \tag{2.60}$$

for some $\tilde{M}^* < \infty$. Since the condition of Lemma 2.6(b) is satisfied by the assumption of the present Lemma, Lemma 2.6(b) gives

$$\lim_{n\to\infty} n^a P\left(\sup_{z\in R^{L_\beta}}\left|P^*(\sqrt{n}(\hat{\beta}^*-\hat{\beta})\leq z)-P^*(\sqrt{n}A_1(S_n^*)\leq z)\right|>n^{-a}\right)=0,$$

and the present Lemma (b) with $\Delta_n^*=\sqrt{n}(\hat{\theta}^*-\hat{\theta})$ holds because Δ_n^* is a subvector of $\sqrt{n}(\hat{\beta}^*-\hat{\beta})$.

Next, we prove the present Lemma (b) with $\Delta_n^*=T_n^*\equiv\sqrt{n}(\hat{\theta}_r^*-\hat{\theta}_r)/\sqrt{\hat{\Sigma}_{rr}^*}$. Define

$$\hat{\Sigma}^*(\beta)\equiv\left(\frac{\partial\phi_n^*(\beta)}{\partial\beta'}\right)^{-1} n^{-1}\sum_{i=1}^{n}\phi(X_i^*,\beta)\phi(X_i^*,\beta)'\left(\frac{\partial\phi_n^*(\beta)}{\partial\beta}\right)^{-1},$$

and $H_n^*(\beta)\equiv(\beta_r-\hat{\beta}_r)(\hat{\Sigma}_{rr}^*(\beta))^{-1/2}$ for some r that corresponds to an element of θ. Consider a Taylor expansion of $H_n^*(\hat{\beta}^*)$ around $\hat{\beta}$, through order $d-1$:

$$\begin{aligned}H_n^*(\hat{\beta}^*) &= 0+\left(\hat{\Sigma}_{rr}^*(\hat{\beta})\right)^{-\frac{1}{2}}\delta_{n,r}^*+\frac{1}{2}\sum_{j_1=1}^{L_\beta}\sum_{j_2=1}^{L_\beta}\frac{\partial^2 H_n^*(\hat{\beta})}{\partial\beta_{j_1}\partial\beta_{j_2}}\delta_{n,j_1}^*\delta_{n,j_2}^*+\cdots\\ &\quad+\frac{1}{(d-1)!}\sum_{j_1=1}^{L_\beta}\cdots\sum_{j_{d-1}=1}^{L_\beta}\frac{\partial^{d-1}H_n^*(\hat{\beta})}{\partial\beta_{j_1}\cdots\partial\beta_{j_{d-1}}}\delta_{n,j_1}^*\cdots\delta_{n,j_{d-1}}^*+\eta_n^*,\end{aligned}$$

where

$$\eta_n^*=\frac{1}{(d-1)!}\sum_{j_1=1}^{L_\beta}\cdots\sum_{j_{d-1}=1}^{L_\beta}\left(\frac{\partial^{d-1}H_n^*(\bar{\beta}^*)}{\partial\beta_{j_1}\cdots\partial\beta_{j_{d-1}}}-\frac{\partial^{d-1}H_n^*(\hat{\beta})}{\partial\beta_{j_1}\cdots\partial\beta_{j_{d-1}}}\right)\delta_{n,j_1}^*\cdots\delta_{n,j_{d-1}}^*,$$

where $\bar{\beta}^*$ is between $\hat{\beta}^*$ and $\hat{\beta}$ and may differ across rows. The remainder of the proof proceeds as in the proof of the present Lemma (a). In particular, we use (2.60), $P(P^*(\|\eta_n^*\|>n^{-dc})>n^{-a})=o(n^{-a})$ by a similar argument with (2.58), and Lemma 2.6(b). *Q.E.D.*

Proof of Lemma 2.14

Proof of (a). We show for $m = 2, 3, 4, 5, 6$, because $m = 6$ is the largest number that we need in later Lemmas. Throughout the proof, let $j_\mu = 1$ for $\mu = 1, ..., m$, for notational brevity. By the definition of S_n and S, the first element of S_n and S are $n^{-1}\sum_{i=1}^n f_1(X_i, \beta_0)$ and $Ef_1(X_i, \beta_0)$, respectively, where $f_1(X_i, \beta)$ denote the first element of the vector $f(X_i, \beta)$. Define $s_i(\beta) = f_1(X_i, \beta)$ and $s_n(\beta) = n^{-1}\sum_{i=1}^n s_i(\beta)$ and write $s_i \equiv s_i(\beta_0)$ and $s_n \equiv s_n(\beta_0)$.

First, we show for $m = 2$. Since $n^{\alpha(2)} = 1$, $n^{\alpha(2)}E\Psi_{n,1}^2 = nE(s_n - Es_i)^2 = n(Es_n^2 - (Es_i)^2)$. By Assumption 1,

$$Es_n^2 = E\frac{1}{n^2}\left(\sum_i s_i\right)^2 = \frac{1}{n^2}E\left(\sum_i s_i^2 + \sum_i\sum_{j\neq i} s_i s_j\right) = \frac{1}{n}Es_i^2 + \frac{n-1}{n}(Es_i)^2.$$

Thus, we have $n^{\alpha(2)}E\Psi_{n,1}^2 = Es_i^2 - (Es_i)^2 = \lim_{n\to\infty} n^{\alpha(2)}E\Psi_{n,1}^2$.

Next, we show for $m = 3$. Since $n^{\alpha(3)} = \sqrt{n}$, $n^{\alpha(3)}E\Psi_{n,1}^3 = n^2(Es_n^3 - 3Es_n^2 Es_i + 2(Es_i)^3)$. By Assumption 1,

$$\begin{aligned} Es_n^3 &= E\frac{1}{n^3}\left(\sum_i s_i\right)^3 = \frac{1}{n^3}E\left(\sum_i s_i^3 + \frac{3!}{1!2!}\sum_i\sum_{j\neq i} s_i s_j^2 + \frac{3!}{6}\sum_i\sum_{j\neq i}\sum_{k\neq i,j} s_i s_j s_k\right) \\ &= \frac{1}{n^2}Es_i^3 + \frac{3(n-1)}{n^2}Es_i Es_i^2 + \frac{(n-1)(n-2)}{n^2}(Es_i)^3. \end{aligned}$$

Combining this result with the result for $m = 2$, we have $n^{\alpha(3)}E\Psi_{n,1}^3 = Es_i^3 - 3Es_i Es_i^2 + 2(Es_i)^3 = \lim_{n\to\infty} n^{\alpha(3)}E\Psi_{n,1}^3$.

Next, we show for $m = 4$. Since $n^{\alpha(4)} = 1$,

$$n^{\alpha(4)}E\Psi_{n,1}^4 = n^2(Es_n^4 - 4Es_n^3 Es_i + 6Es_n^2(Es_i)^2 - 3(Es_i)^4).$$

By Assumption 1,

$$\begin{aligned}
Es_n^4 &= E\frac{1}{n^4}\left(\sum_i s_i\right)^4 = \frac{1}{n^4}E\left(\sum_i s_i^4 + \frac{4!}{1!3!}\sum_i\sum_{j\neq i} s_i s_j^3 + \frac{4!}{2!2!}\frac{1}{2}\sum_i\sum_{j\neq i} s_i^2 s_j^2\right.\\
&\quad \left. + \frac{4!}{1!1!2!}\frac{1}{2}\sum_i\sum_{j\neq i}\sum_{k\neq i,j} s_i s_j s_k^2 + \frac{4!}{4!}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k} s_i s_j s_k s_l\right)\\
&= \frac{1}{n^3}Es_i^4 + \frac{4(n-1)}{n^3}Es_iEs_i^3 + \frac{3(n-1)}{n^3}(Es_i^2)^2 + \frac{6(n-1)(n-2)}{n^3}(Es_i)^2Es_i^2\\
&\quad + \frac{(n-1)(n-2)(n-3)}{n^3}(Es_i)^4.
\end{aligned}$$

Combining this result with the results for $m = 2, 3$, we have

$$\begin{aligned}
n^{\alpha(4)}E\Psi_{n,1}^4 &= \frac{1}{n}Es_i^4 - \frac{4}{n}Es_iEs_i^3 + \frac{-6n+12}{n}(Es_i)^2Es_i^2\\
&\quad + \frac{3(n-1)}{n}(Es_i^2)^2 + \frac{3(n-2)}{n}(Es_i)^4\\
&\underset{n\to\infty}{\longrightarrow} 3(Es_i)^4 + 3(Es_i^2)^2 - 6(Es_i)^2Es_i^2.
\end{aligned}$$

Next, we show for $m = 5$. Since $n^{\alpha(5)} = \sqrt{n}$,

$$n^{\alpha(5)}E\Psi_{n,1}^5 = n^3(Es_n^5 - 5Es_n^4Es_i + 10Es_n^3(Es_i)^2 - 10Es_n^2(Es_i)^3 + 4(Es_i)^5).$$

By Assumption 1,

$$
\begin{aligned}
Es_n^5 &= E\frac{1}{n^5}\left(\sum_i s_i\right)^5 = \frac{1}{n^5}E\left(\sum_i s_i^5 + \frac{5!}{1!4!}\sum_i\sum_{j\neq i} s_i s_j^4 + \frac{5!}{2!3!}\sum_i\sum_{j\neq i} s_i^2 s_j^3\right. \\
&\quad +\frac{5!}{1!1!3!}\frac{1}{2}\sum_i\sum_{j\neq i}\sum_{k\neq i,j} s_i s_j s_k^3 + \frac{5!}{1!2!2!}\frac{1}{2}\sum_i\sum_{j\neq i}\sum_{k\neq i,j} s_i s_j^2 s_k^2 \\
&\quad \left.\frac{5!}{1!1!1!2!}\frac{1}{3!}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k} s_i s_j s_k s_l^2 + \frac{5!}{5!}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k}\sum_{m\neq i,j,k,l} s_i s_j s_k s_l s_m\right) \\
&= \frac{1}{n^4}Es_i^5 + \frac{5(n-1)}{n^4}Es_iEs_i^4 + \frac{10(n-1)}{n^4}Es_i^2Es_i^3 + \frac{10(n-1)(n-2)}{n^4}(Es_i)^2Es_i^3 \\
&\quad +\frac{15(n-1)(n-2)}{n^4}Es_i(Es_i^2)^2 + \frac{10(n-1)(n-2)(n-3)}{n^4}(Es_i)^3Es_i^2 \\
&\quad +\frac{(n-1)(n-2)(n-3)(n-4)}{n^4}(Es_i)^5.
\end{aligned}
$$

Combining this result with the results for $m = 2, 3, 4$, we have

$$
\begin{aligned}
n^{\alpha(5)}E\Psi_{n,1}^5 &= \frac{1}{n}Es_i^5 - \frac{5}{n}Es_iEs_i^4 - \frac{30(n-1)}{n}Es_i(Es_i^2)^2 + \frac{10(n-1)}{n}Es_i^2Es_i^3 \\
&\quad +\frac{50n-60}{n}Es_i^2(Es_i)^3 + \frac{-10n+20}{n}(Es_i)^2Es_i^3 + \frac{-20n+24}{n}(Es_i)^5 \\
&\underset{n\to\infty}{\longrightarrow} -30Es_i(Es_i^2)^2 + 10Es_i^2Es_i^3 + 50Es_i^2(Es_i)^3 - 10(Es_i)^2Es_i^3 - 20(Es_i)^5.
\end{aligned}
$$

Finally, we show for $m = 6$. Since $n^{\alpha(6)} = 1$,

$$
n^{\alpha(6)}E\Psi_{n,1}^6 = n^3(Es_n^6 - 6Es_n^5Es_i + 15Es_n^4(Es_i)^2 - 20Es_n^3(Es_i)^3 + 15Es_n^2(Es_i)^4 - 5(Es_i)^6).
$$

By Assumption 1,

$$
\begin{aligned}
Es_n^6 &= \frac{1}{n^6}E\left(\sum_i s_i^6 + \frac{6!}{5!}\sum_i\sum_{j\neq i}s_is_j^5 + \frac{6!}{2!4!}\sum_i\sum_{j\neq i}s_i^2s_j^4 + \frac{6!}{3!3!}\frac{1}{2}\sum_i\sum_{j\neq i}s_i^3s_j^3\right.\\
&+\frac{6!}{4!}\frac{1}{2}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}s_is_js_k^4 + \frac{6!}{2!3!}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}s_is_j^2s_k^3 + \frac{6!}{2!2!2!}\frac{1}{3!}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}s_i^2s_j^2s_k^2\\
&+\frac{6!}{3!}\frac{1}{3!}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k}s_is_js_ks_l^3 + \frac{6!}{2!2!}\frac{1}{2!2!}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k}s_is_js_k^2s_l^2\\
&+\frac{6!}{2!}\frac{1}{4!}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k}\sum_{m\neq i,j,k,l}s_is_js_ks_ls_m^2\\
&\left.+\frac{6!}{6!}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k}\sum_{m\neq i,j,k,l}\sum_{r\neq i,j,k,l,m}s_is_js_ks_ls_ms_r\right)\\
&= \frac{1}{n^5}Es_i^6 + \frac{6(n-1)}{n^5}Es_iEs_i^5 + \frac{15(n-1)}{n^5}Es_i^2Es_i^4 + \frac{10(n-1)}{n^5}(Es_i^3)^2\\
&+\frac{15(n-1)(n-2)}{n^5}(Es_i)^2Es_i^4 + \frac{60(n-1)(n-2)}{n^5}Es_iEs_i^2Es_i^3\\
&+\frac{15(n-1)(n-2)}{n^5}(Es_i^2)^3 + \frac{20(n-1)(n-2)(n-3)}{n^5}(Es_i)^3Es_i^3\\
&+\frac{45(n-1)(n-2)(n-3)}{n^5}(Es_i)^2(Es_i^2)^2\\
&+\frac{15(n-1)(n-2)(n-3)(n-4)}{n^5}(Es_i)^4Es_i^2\\
&+\frac{(n-1)(n-2)(n-3)(n-4)(n-5)}{n^5}(Es_i)^6.
\end{aligned}
$$

Combining this result with the results for $m = 2, 3, 4, 5$, we have

$$\lim_{n\to\infty} n^{\alpha(6)}E\Psi_{n,1}^6 = 15(Es_i^2)^3 - 45(Es_i)^2(Es_i^2)^2 + 45(Es_i)^4Es_i^2 - 15(Es_i)^6.$$

In order for all the quantities to be well defined, the most restrictive case is that Es_i^6 exists. Since s_i is an element of $f(X_i, \beta_0)$, it suffices to show that the condition is satisfied for $s_i = e^{2\lambda_0' g_i(\theta_0)}g^{k_0}G^{k_1}\cdots G^{(d+1)k_{d+1}}$, where $k_0 + \cdots + k_{d+1} = d+3$. By using Assumptions 2-3 and Hölder's inequality, we have $q_g, q_G \geq 6(d+3)(1+\zeta^{-1})$ and $q_\tau \geq 12(1+\zeta)$ for any $\zeta > 0$. For arbitrary a, we use the fact that $\max\{m\} = 2a+2$ to show q_g, q_G, and q_τ should satisfy $q_g, q_G \geq 2(a+1)(d+3)(1+\zeta^{-1})$ and $q_\tau \geq 4(a+1)(1+\zeta)$ for any $\zeta > 0$. This proves the present Lemma (a).

Proof of (b). Since the bootstrap sample is iid, the proof is analogous to that of the present Lemma (a). In particular, we replace E, X_i, and β_0 with E^*, X_i^*, and $\hat{\beta}$, respectively. Let $s_i^*(\beta) = f_1(X_i^*, \beta)$ and $s_n^*(\beta) = n^{-1}\sum_{i=1}^n f_1(X_i^*, \beta)$. In addition, write $\hat{s}_i^* \equiv s_i^*(\hat{\beta})$, $\hat{s}_i \equiv s_i(\hat{\beta})$, $\hat{s}_n^* \equiv s_n^*(\hat{\beta})$, and $\hat{s}_n \equiv s_n(\hat{\beta})$ for notational brevity.

We describe the proof with $m = 2$, and this illustrates the proof for arbitrary m. Since $n^{\alpha(2)} = 1$,

$$n^{\alpha(2)} E^* \Psi_{n,1}^{*2} = E^* \hat{s}_i^{*2} - (E^* \hat{s}_i^*)^2 = n^{-1}\sum_{i=1}^n \hat{s}_i^2 - \left(n^{-1}\sum_{i=1}^n \hat{s}_i\right)^2.$$

Since $\lim_{n\to\infty} n^{\alpha(2)} E\Psi_{n,1}^2 = Es_i^2 - (Es_i)^2$, combining the following results proves the Lemma for $m = 2$:

$$P\left(\left\|n^{-1}\sum_{i=1}^n \hat{u}_i - n^{-1}\sum_{i=1}^n u_i\right\| > n^{-\gamma}\right) = o(n^{-a}), \tag{2.61}$$

$$P\left(\left\|n^{-1}\sum_{i=1}^n u_i - Eu_i\right\| > n^{-\gamma}\right) = o(n^{-a}), \tag{2.62}$$

where $\hat{u}_i = \hat{s}_i$ or $\hat{u}_i = \hat{s}_i^2$, and $u_i = s_i$ or $u_i = s_i^2$. We use the fact $\|\hat{s}_i^2 - s_i^2\| \leq \|\hat{s}_i - s_i\|(\|\hat{s}_i - s_i\| + 2s_i)$, Lemma 2.11(a), Lemma 2.2, and Lemma 2.5 to show the first result (2.61). The second result is shown by Lemma 2.2(a) with $h(X_i) = s_i^2 - Es_i^2$. By considering the most restrictive form of s_i and combining the conditions for the Lemmas, we need $q_g, q_G \geq 4(d+3)(1+\zeta^{-1})$ and $q_g, q_G > 4a(d+3)(1-2\gamma)^{-1}(1+\zeta^{-1})$, and $q_\tau \geq 8(1+\zeta)$ and $q_\tau > 8a(1-2\gamma)^{-1}(1+\zeta)$.

For arbitrary m, we can show the results (2.61) and (2.62) for $u_i = s_i^m$ by using the binomial expansion as the proof of Lemma 2.11, Lemmas 2.2, 2.5, 2.11(a), and 2.12. Since $\max\{m\} = 2a+2$, q_g, q_G, and q_τ should satisfy $q_g, q_G \geq 4(a+1)(d+3)(1+\zeta^{-1})$ and $q_g, q_G \geq 8a(a+1)(d+3)(1-2\gamma)^{-1}(1+\zeta^{-1})$, and $q_\tau \geq 8(a+1)(1+\zeta)$ and $q_\tau \geq 16a(a+1)(1-2\gamma)^{-1}(1+\zeta)$ for any $\zeta > 0$. This proves the present Lemma (b). *Q.E.D.*

Proof of Lemma 2.15

Proof of (a). By Lemma 2.13(a), it suffices to show that $\sqrt{n}A(S_n)$ possesses Edgeworth expansion with remainder $o(n^{-a})$, where $A(\cdot)$ is an infinitely differentiable real-valued function. The coefficients ν_a are well-defined by Lemma 2.14(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 Assumptions 1-4, and the fact that $A(\cdot)$ is infinitely differentiable and real-valued.

Proof of (b). By Lemma 2.13(b), it suffices to show that $\sqrt{n}A(S_n^*)$ possesses Edgeworth expansion with remainder $o(n^{-a})$. The present Lemma (b) holds 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 2.14(b) with $\gamma = 0$ to ensure that the coefficients $\nu_{n,a}^*$ are well behaved. *Q.E.D.*

2.8 Technical Appendix

The moment function is $g_i = (Y_i\theta - 1, Z_i\theta - 1)'$ and the first derivative of the moment function is $G_i = (Y_i, Z_i)'$. First, we solve for λ_0 as a function of θ. The FOC is given by

$$\begin{aligned} 0 &= E\exp(\lambda' g_i) g_i \\ &= E e^{\lambda_{0,1}(Y_i\theta - 1) + \lambda_{0,2}(Z_i\theta - 1)} \left(Y_i\theta - 1, Z_i\theta - 1\right)'. \end{aligned} \tag{2.63}$$

By using that Y_i and Z_i are independent and Fubini's Theorem, the first row of (2.63) can be rewritten as

$$
\begin{aligned}
0 &= \int\int e^{\lambda_{0,1}(y\theta-1)+\lambda_{0,2}(z\theta-1)}(y\theta-1)dF_y dF_z \\
&= e^{-\lambda_{0,1}-\lambda_{0,2}}\int e^{\lambda_{0,2}\theta z}dF_z\cdot\int e^{\lambda_{0,1}\theta y}(y\theta-1)dF_y \\
&= e^{-\lambda_{0,1}-\lambda_{0,2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{\lambda_{0,2}\theta z-\frac{(z-(1-\delta))^2}{2}}dz\cdot\int_{-\infty}^{\infty}(y\theta-1)\frac{1}{\sqrt{2\pi}}e^{\lambda_{0,1}\theta y-\frac{(y-1)^2}{2}}dy \\
&= C\cdot\int_{-\infty}^{\infty}(y\theta-1)\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-(1+\lambda_{0,1}\theta))^2}{2}}dy \\
&= C\cdot((1+\lambda_{0,1}\theta)\theta-1),
\end{aligned}
$$

where C denotes any nonzero constant. For the fourth equality, we use the fact that the integration of the probability density function (pdf) of a Normal random variable over the real value equals one. For the last equality, we use the fact that the function in the integral is the pdf of a normal random variable with the mean $1+\lambda_{0,1}\theta$. Thus, we have $1=(1+\lambda_{0,1}\theta)\theta$, or $\lambda_{0,1}(\theta)=\theta^{-1}(\theta^{-1}-1)$.

Similarly, the second row of (2.63) can be rewritten as

$$
\begin{aligned}
0 &= \int\int e^{\lambda_{0,1}(y\theta-1)+\lambda_{0,2}(z\theta-1)}(z\theta-1)dF_y dF_z \\
&= e^{-\lambda_{0,1}-\lambda_{0,2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{\lambda_{0,1}\theta y-\frac{(y-1)^2}{2}}dy\cdot\int_{-\infty}^{\infty}(z\theta-1)\frac{1}{\sqrt{2\pi}}e^{\lambda_{0,2}\theta z-\frac{(z-(1-\delta))^2}{2}}dz \\
&= C\cdot\int_{-\infty}^{\infty}(z\theta-1)\frac{1}{\sqrt{2\pi}}e^{-\frac{(z-(1-\delta+\lambda_{0,2}\theta))^2}{2}}dz \\
&= C\cdot((1-\delta+\lambda_{0,2}\theta)\theta-1),
\end{aligned}
$$

where C denotes any nonzero constant. The last equation implies that $\lambda_{0,2}(\theta)=\theta^{-1}(\theta^{-1}-1+\delta)$.

Observe that

$$
\max -\ln Ee^{\lambda_0(\theta)'(g_i-Eg_i)} \Leftrightarrow \min Ee^{\lambda_0(\theta)'(g_i-Eg_i)},
$$

where $Ee^{\lambda_0(\theta)'(g_i - Eg_i)} = Ee^{\lambda_{0,1}\theta(Y_i - EY_i)} \cdot Ee^{\lambda_{0,2}\theta(Z_i - EZ_i)}$, and

$$\begin{aligned} Ee^{\lambda_{0,1}\theta(Y_i - EY_i)} &= \exp\left(\frac{1}{2\theta^2} - \frac{1}{\theta} + \frac{1}{2}\right) \\ Ee^{\lambda_{0,2}\theta(Z_i - EZ_i)} &= \exp\left(\frac{1}{\theta^2} - \frac{2-\delta}{\theta} + \frac{1+(1-\delta)^2}{2}\right). \end{aligned}$$

Since

$$\frac{1}{2\theta^2} - \frac{1}{\theta} + \frac{1}{2} + \frac{1}{\theta^2} - \frac{2-\delta}{\theta} + \frac{1+(1-\delta)^2}{2} = \left(\frac{1}{\theta} - \frac{2-\delta}{2}\right)^2 + \frac{\delta-1}{2},$$

we have $\theta_0 = 2(2-\delta)^{-1}$.

3 BOOTSTRAPPING GMM ESTIMATORS UNDER LOCAL MISSPECIFICATION

3.1 Asymptotic Distribution of GMM

The $L_g \times 1$ dimensional moment function is given by $g(X_i, \theta)$, where θ is the $L_\theta \times 1$ parameter vector. Let a triangular array $\{X_i : i \leq n\}$ be iid over i for fixed n. For notational simplicity, I suppress the additional subscript n on X_i. Assume $L_g \geq L_\theta$, so that the model is overidentified.[1] Write $g_i(\theta) = g(X_i, \theta)$. The moment condition is correctly specified if

$$Eg(X_i, \theta_0) = 0, \tag{3.1}$$

holds for a unique θ_0. A locally misspecified model is defined as follows:

$$Eg(X_i, \theta_0) = \frac{\delta}{\sqrt{n}}, \tag{3.2}$$

where δ is $L_g \times 1$ vector of constants. This type of misspecification can be used to describe situations such that the moment condition is slightly violated with finite n, but becomes correctly specified asymptotically. For example, a set of instruments is not exactly exogeneous when n is finite, but the exogeneity condition is satisfied as n goes to infinity.

When the moment condition is locally misspecified, the GMM estimator is consistent for θ_0, but is not $\sqrt{n}$-consistent. The asymptotic variance of the GMM estimator is not affected.[2] To see this, let $\hat{\theta}$ be the GMM estimator that minimizes the following criterion function:

$$J_n(\theta) = g_n(\theta)' W_n g_n(\theta), \tag{3.3}$$

where $g_n(\theta) = n^{-1} \sum_i g(X_i, \theta)$, and W_n is a weight matrix such that $W_n \to_p W$ and W

[1]The discussion of this note covers a just-identified model ($L_g = L_\theta$). However, a non-standard feature of bootstrapping for GMM arises when the model is over-identified. Thus, $L_g > L_\theta$ is the primary focus of this chapter.

[2]Hall (2005) gives a detailed analysis on the asymptotic behavior of the GMM estimator under local misspecification.

is a positive definite matrix. Let $G_n(\theta) = n^{-1}\sum_i G(X_i,\theta)$, $G(X_i,\theta) = (\partial/\partial\theta')g(X_i,\theta)$, and $G_0 = \lim_{n\to\infty} n^{-1}\sum_i EG(X_i,\theta_0)$. By applying the mean value theorem to the first-order condition multiplied by $\sqrt{n}$,[3] we have

$$\begin{aligned}\sqrt{n}(\hat{\theta}-\theta_0) &= -(G_n(\hat{\theta})'W_nG_n(\bar{\theta}))^{-1}G_n(\hat{\theta})'W_n\sqrt{n}g_n(\theta_0)\\ &= -(G_0'WG_0)^{-1}G_0'W\sqrt{n}(g_n(\theta_0)-Eg(X_i,\theta_0))\\ &\quad -(G_0'WG_0)^{-1}G_0'W\sqrt{n}Eg(X_i,\theta_0)+O_p(n^{-1/2}),\end{aligned}$$

where $\bar{\theta}$ is the mean value between $\hat{\theta}$ and θ_0. By the Lindeberg central limit theorem for a triangular array, we have

$$\sqrt{n}(\hat{\theta}-\theta_0) \to_d N(-(G_0'WG_0)^{-1}G_0'W\delta, \Sigma), \tag{3.4}$$

where

$$\begin{aligned}\Sigma &= (G_0'WG_0)^{-1}G_0'W\Omega WG_0(G_0'WG_0)^{-1},\\ \Omega &= \lim_{n\to\infty} n^{-1}\sum_i Eg(X_i,\theta_0)g(X_i,\theta_0)',\end{aligned}$$

and Ω is positive definite and finite. The asymptotic bias in the limiting distribution arises due to the locally misspecified moment condition. The expression (3.4) becomes more interesting if we rewrite it as

$$\sqrt{n}(\hat{\theta}-\theta_{0(n)}) \to_d N(0,\Sigma), \tag{3.5}$$

where $\theta_{0(n)} = \theta_0 - (G_0'WG_0)^{-1}G_0'WEg(X_i,\theta_0)$. Now the GMM estimator is properly centered. $\theta_{0(n)}$ is called the (pseudo-)true value under local misspecification and

[3]Consistency of $\hat{\theta}$ for θ_0 can be shown easily by the FOC.

indexed by n. Since $Eg(X_i, \theta_0) \to 0$ as $n \to \infty$, $\theta_{0(n)} \to \theta_0$ as $n \to \infty$. Let

$$\begin{aligned} \hat{\Sigma} &= (\hat{G}_n' W_n \hat{G}_n)^{-1} \hat{G}_n' W_n \hat{\Omega}_n W_n \hat{G}_n (\hat{G}_n' W_n \hat{G}_n)^{-1}, \\ \hat{G}_n &= G_n(\hat{\theta}), \\ \hat{\Omega}_n &= n^{-1} \sum_i g(X_i, \hat{\theta}) g(X_i, \hat{\theta})', \end{aligned}$$

then $\hat{\Sigma}$ is a consistent estimator for Σ. Also let θ_r and $\hat{\sigma}_r$ denote the rth and (r,r)th element of θ and $(\hat{\Sigma})^{1/2}$, respectively. A conventional t statistic for testing the null hypothesis $H_0 : \theta_r = \theta_{0(n),r}$ is

$$T_n \equiv \frac{\sqrt{n}(\hat{\theta}_r - \theta_{0(n),r})}{\hat{\sigma}_r}. \tag{3.6}$$

The implication of (3.5) is that the conventional t tests and CI's based on T_n have correct rejection and coverage probabilities asymptotically for the true value $\theta_{0(n)}$. The introduction of $\theta_{0(n)}$ allows flexible interpretation of the estimand under possible misspecification. Under correct specification ($Eg(X_i, \theta_0) = 0$), $\theta_{0(n)} = \theta_0$. Under global misspecification ($Eg(X_i, \theta_0) = \delta$, a vector of constants), $\theta_{0(n)} \neq \theta_0$ as $n \to \infty$. In general, the relationship between $\theta_{0(n)}$ and θ_0 is unknown.

3.2 Conventional Bootstrap Methods

Now consider bootstrap methods for GMM. Hahn (1996) showed first-order validity of the bootstrap for GMM estimators under correct specification. It supports wide use of Efron-type percentile intervals in practice. Furthermore, Hall and Horowitz (1996) (denoted by HH bootstrap, hereinafter) established asymptotic refinements of the bootstrap for t and Wald statistics based on GMM estimators. Therefore, percentile-t CI's constructed by using the HH bootstrap are expected to be more accurate than the asymptotic CI by having smaller error in the coverage probability.

Both Hahn (1996) and Hall and Horowitz (1996) assume correct specification of the moment condition model. Do they work under local misspecification? To answer this question, I allow for local misspecification of the moment condition model and show that the percentile and percentile-t intervals are consistent for $\theta_{0(n)}$ even under local misspecification. However, HH bootstrap fails to achieve asymptotic refinements and a counter example is provided.

First-order Validity

First consider the bootstrap version of the criterion function (3.3) and the first-order condition.[4] Let $G_n^*(\theta) = n^{-1}\sum_i G(X_i^*, \theta)$ and $g_n^*(\theta) = n^{-1}\sum_i g(X_i^*, \theta)$. Let W_n^* be the bootstrap version of the weight matrix such that $W_n^* \to_p W$ conditional on the sample a.s. By the mean value theorem around the GMM estimator $\hat{\theta}$, we have the following expression:

$$\hat{\theta}^* - \hat{\theta} \;\; = \;\; -(G_n^*(\hat{\theta}^*)' W_n^* G_n^*(\bar{\theta}^*))^{-1} G_n^*(\hat{\theta}^*)' W_n^* g_n^*(\hat{\theta}),$$

where $\bar{\theta}^*$ is the mean value between $\hat{\theta}^*$ and $\hat{\theta}$. Using $\hat{\theta} \to_p \theta_0$, we can show $\hat{\theta}^* \to_p \theta_0$ conditional on the sample a.s. Since $E^* g_n^*(\hat{\theta}) = g_n(\hat{\theta})$, the RHS of the above expression is not properly centered and we need further expansion to apply the Lindeberg CLT. We use $G_n^*(\hat{\theta}^*) = G_n(\hat{\theta}) + o_p^*(1)$, $W_n^* = W_n + o_p^*(1)$, and the FOC, $0 = G_n(\hat{\theta})' W_n g_n(\hat{\theta})$. Multiplying both sides by $\sqrt{n}$,

$$\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \;\; = \;\; -(G_0' W G_0)^{-1} G_0' W \sqrt{n}(g_n^*(\hat{\theta}) - g_n(\hat{\theta})) + O_p^*(n^{-1/2}).$$

Finally, we have

$$\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \to_d N(0, \Sigma), \tag{3.7}$$

conditional on the sample a.s., and this establishes the first order validity of the bootstrap under local misspecification. Therefore, the percentile intervals are expected to provide correct coverage probability asymptotically for the (pseudo-)true value $\theta_{0(n)}$ under local misspecification.

[4] Note that this is a *naive* bootstrap for GMM without recentering.

Now consider the HH bootstrap. It bootstraps the t statistic, instead of the GMM estimator itself. Bootstrapping asymptotically pivotal statistic, like the t statistic, yields asymptotic refinements. Before investigating the ability of achieving refinements of the HH bootstrap, I first show the HH bootstrap is still consistent under local misspecification.

Let $g_n^*(\theta) = n^{-1}\sum_i g(X_i^*, \theta)$. The recentered criterion function is

$$\bar{J}_n^*(\theta) = \left(g_n^*(\theta) - g_n(\hat{\theta})\right)' \bar{W}_n^* \left(g_n^*(\theta) - g_n(\hat{\theta})\right), \tag{3.8}$$

where $\bar{W}_n^*$ is constructed using the recentered moment function $g(X_i^*, \theta) - g_n(\hat{\theta})$. For example, a common choice of W_n is

$$W_n(\tilde{\theta}) = \left(\frac{1}{n}\sum_i g(X_i, \tilde{\theta})g(X_i, \tilde{\theta})'\right)^{-1},$$

where $\tilde{\theta}$ is a preliminary estimator. Then, W_n^* and $\bar{W}_n^*$ are

$$\begin{aligned} W_n^* &= \left(\frac{1}{n}\sum_i g(X_i^*, \tilde{\theta}^*)g(X_i^*, \tilde{\theta}^*)'\right)^{-1}, \\ \bar{W}_n^* &= \left(\frac{1}{n}\sum_i (g(X_i^*, \tilde{\theta}^*) - g_n(\hat{\theta}))(g(X_i^*, \tilde{\theta}^*) - g_n(\hat{\theta}))'\right)^{-1}, \end{aligned}$$

where $\tilde{\theta}^*$ is the bootstrap version of the preliminary estimator. Note that W_n^* is used for the naive bootstrap without recentering. If W_n does not depend on the moment function $g(X_i, \theta)$, then $\bar{W}_n^* = W_n^*$. Let $\hat{\theta}_{HH}^*$ be the HH bootstrap version of the GMM estimator that satisfies the following FOC:

$$0 = G_n^*(\hat{\theta}_{HH}^*)'\bar{W}_n^*(g_n^*(\hat{\theta}_{HH}^*) - g_n(\hat{\theta})).$$

By Taylor expanding the RHS around $\hat{\theta}$, we have

$$\hat{\theta}_{HH}^* - \hat{\theta} = -(G_n^*(\hat{\theta}_{HH}^*)'\bar{W}_n^* G_n^*(\bar{\theta}^*))^{-1} G_n^*(\hat{\theta}_{HH}^*)'\bar{W}_n^*(g_n^*(\hat{\theta}) - g_n(\hat{\theta})),$$

where $\bar{\theta}^*$ is the mean value between $\hat{\theta}^*_{HH}$ and $\hat{\theta}$. We can show $\hat{\theta}^*_{HH} \to_p \hat{\theta}$ conditional on the sample a.s. and $\bar{W}^*_n = W + o^*_p(1)$. Multiplying both sides by $\sqrt{n}$,

$$\sqrt{n}(\hat{\theta}^*_{HH} - \hat{\theta}) = -(G_0'WG_0)^{-1}G_0'W\sqrt{n}(g^*_n(\hat{\theta}) - g_n(\hat{\theta})) + O^*_p(n^{-1/2}).$$

Then, the Lindeberg CLT gives

$$\sqrt{n}(\hat{\theta}^*_{HH} - \hat{\theta}) \to_d N(0, \Sigma),$$

conditional on the sample a.s. Let

$$\begin{aligned} \hat{\Sigma}^*_{HH} &= (G^{*\prime}_n \bar{W}^*_n G^*_n)^{-1} G^{*\prime}_n \bar{W}^*_n \bar{\Omega}^*_n \bar{W}^*_n G^*_n (G^{*\prime}_n \bar{W}^*_n G^*_n)^{-1}, \\ \bar{\Omega}^*_n &= n^{-1} \sum_i (g(X^*_i, \hat{\theta}^*_{HH}) - g_n(\hat{\theta}))(g(X^*_i, \hat{\theta}^*_{HH}) - g_n(\hat{\theta}))'. \end{aligned}$$

It is easy to show $\hat{\Sigma}^*_{HH} = \Sigma + o^*_p(1)$. Also let $\sigma^*_{HH,r}$ denote the (r, r)th component of $(\hat{\Sigma}^*_{HH})^{1/2}$. Finally, we have

$$T^*_{HH,n} \equiv \frac{\sqrt{n}(\hat{\theta}^*_{HH,r} - \hat{\theta}_r)}{\hat{\sigma}^*_{HH,r}} \to_d N(0, 1), \tag{3.9}$$

for any $r = 1, ..., L_\theta$, conditional on the sample a.s. This result implies that the percentile-t intervals using the HH bootstrap critical values have correct coverage probabilities asymptotically. Thus, the HH bootstrap is first-order valid even under local misspecification.

Asymptotic Refinements

In this section, I argue that the HH bootstrap does not achieve asymptotic refinements under local misspecification. In order to investigate the higher-order property of the distribution of the t statistic, I use Hall (1997)'s argument based on the Edgeworth expansion. Consider the t statistic defined as (3.6). We may expand the distribution function of T_n as

$$P(T_n \leq z) \;\; = \;\; \Phi(z) + n^{-1/2} q(z, F) + O(n^{-1}),$$

uniformly over z, where $q(z, F)$ is an even function of z for each F. Asymptotic t tests and confidence intervals use the critical value from the standard Normal distribution, ignoring the higher-order terms. Therefore, the Normal approximation is in error by $n^{-1/2}$.

Now consider $T^*_{HH,n}$ defined in (3.9). The expansion of the distribution of $T^*_{HH,n}$ is

$$P^*(T^*_{HH,n} \leq z) \;\; = \;\; \Phi(z) + n^{-1/2} q_{HH}(z, F_n) + O_p(n^{-1}),$$

uniformly over z, where $q_{HH}(z, F_n)$ is an even function of z. Observe that the sample analogue of $T^*_{HH,n}$ is not T_n under local misspecification, because the population analogue of $g_n(\hat{\theta})$, $Eg(X_i, \theta_{0(n)})$ is not equal to zero. To find the sample version of $T^*_{HH,n}$, we consider the recentered version of the criterion function

$$\bar{J}_n(\theta) \;\; = \;\; \left(g_n(\theta) - Eg(X_i, \theta_{0(n)})\right)' \bar{W}_n \left(g_n(\theta) - Eg(X_i, \theta_{0(n)})\right),$$

where $\bar{W}_n$ is constructed using the moment function $g(X_i, \theta) - Eg(X_i, \theta_{0(n)})$. The sample analogue of $\hat{\theta}^*_{HH}$ is $\hat{\theta}_{HH}$ that minimizes $\bar{J}_n(\theta)$. Note that $\hat{\theta}_{HH}$ is an infeasible estimator because $Eg(X_i, \theta_{0(n)})$ is unknown. Note that $\theta_{0(n)} = \theta_0$ and $0 = Eg(X_i, \theta_{0(n)})$ under correct specification, and as a result, $\bar{J}_n(\theta) = J_n(\theta)$ and $\hat{\theta}_{HH} = \hat{\theta}$. However, under local misspecification, the sample analogue of $\hat{\theta}^*_{HH}$ is not $\hat{\theta}$, but the minimizer of the recentered criterion function, $\hat{\theta}_{HH}$. Let

$$\begin{aligned}
\hat{\Sigma}_{HH} &= (G_n' \bar{W}_n G_n)^{-1} G_n' \bar{W}_n \bar{\Omega}_n \bar{W}_n G_n (G_n' \bar{W}_n G_n)^{-1}, \\
\bar{\Omega}_n &= n^{-1} \sum_i (g(X_i, \hat{\theta}_{HH}) - Eg(X_i, \theta_{0(n)}))(g(X_i, \hat{\theta}_{HH}) - Eg(X_i, \theta_{0(n)}))', \\
\hat{\sigma}_{HH,r} &= (r,r)\text{th element of } (\hat{\Sigma}_{HH})^{1/2}.
\end{aligned}$$

Then, the (infeasible) recentered t statistic is

$$T_{HH,n} \equiv \frac{\sqrt{n}(\hat{\theta}_{HH,r} - \theta_{0(n),r})}{\hat{\sigma}_{HH,r}},$$

for $r = 1, ..., L_\theta$. The distribution of $T_{HH,n}$ admits analogous expansion

$$P(T_{HH,n} \leq z) \quad = \quad \Phi(z) + n^{-1/2} q_{HH}(z, F) + O(n^{-1}),$$

uniformly over z, where $q_{HH}(z, F)$ is an even function of z. Typically, $q_{HH}(z, F_n) - q_{HH}(z, F) = O_p(n^{-1/2})$. Thus, the HH bootstrap provides asymptotic refinements for the distribution of $T_{HH,n}$:

$$P^*(T^*_{HH,n} \leq z) - P(T_{HH,n} \leq z) \quad = \quad O_p(n^{-1}).$$

The key question is whether $q_{HH}(z, F)$ is close enough to $q(z, F)$ so that the HH bootstrap also achieves asymptotic refinements for the distribution of T_n. In other words, the HH bootstrap provides refinements if $T_{HH,n} = T_n + O_p(n^{-1})$ so that $P(T_{HH,n} \leq z) = P(T_n \leq z) + O(n^{-1})$ by the delta method because

$$\begin{aligned}
& P^*(T^*_{HH,n} \leq z) - P(T_n \leq z) \\
= \quad & P^*(T^*_{HH,n} \leq z) - P(T_{HH,n} \leq z) + P(T_{HH,n} \leq z) - P(T_n \leq z) \\
= \quad & O_p(n^{-1}),
\end{aligned}$$

Unfortunately, the answer seems to be negative because in general,

$$T_{HH,n} = T_n + \xi_n,$$

where $\xi_n \to_d N(0, V_\xi)$ for some covariance matrix V_ξ. Therefore, I conjecture that $P(T_{HH,n} \leq z)$ and $P(T_n \leq z)$ differs by the size of error $n^{-1/2}$, and we may conclude that the HH bootstrap does not achieve asymptotic refinements for the distribution of T_n under local misspecification. A counter example and simulation result are provided in the next section.

An Example

Suppose that we observe $X_i = (Y_i, Z_i)' \in \mathbb{R}^2$, $i = 1, ...n$, and we are interested in the mean of Z_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 can use this information to efficiently estimate θ by GMM. We form the moment function as

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

However, it may be true that the mean of Y is slightly different from zero with finite n, i.e., $EY_i = \delta/\sqrt{n}$ for some fixed number δ, then the model is locally misspecified.

I use the following weight matrix by using the MOM estimator as a preliminary estimator:

$$W_n = \left(n^{-1}\sum_i^n g(X_i, \bar{Z})g(X_i, \bar{Z})'\right)^{-1}.$$

The two-step GMM estimator $\hat{\theta}$ and the (pseudo-)true value $\theta_{0(n)}$ are given by

$$\begin{aligned} \hat{\theta} &= \bar{Z} - \frac{\widehat{Cov}(Y_i, Z_i)}{\overline{Y^2}}\bar{Y}, \\ \theta_{0(n)} &= EZ_i - \frac{Cov(Y_i, Z_i)}{EY_i^2}EY_i = EZ_i - \frac{Cov(Y_i, Z_i)}{EY_i^2}\frac{\delta}{\sqrt{n}}, \end{aligned}$$

where $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i$, $\overline{Y^2} = n^{-1}\sum_{i=1}^n Y_i^2$, and $\widehat{Cov}(Y_i, Z_i) = n^{-1}\sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z})$. Observe that $\theta_{0(n)}$ is the population analogue of $\hat{\theta}$, by replacing the sample mean with the population mean. Also note that $\theta_{0(n)} \to EZ_i$ as n grows, but $\theta_{0(n)} \neq EZ_i$ for all n unless $\delta = 0$. The variance estimator $\hat{\sigma}_c^2$ is given by

$$\hat{\sigma}_c^2 = n^{-1}\sum_{i=1}^n (Z_i - \hat{\theta})^2 - \frac{(\overline{YZ} - \bar{Y}\hat{\theta})^2}{\overline{Y^2}},$$

where $\overline{YZ} = n^{-1}\sum_{i=1}^{n} Y_i Z_i$. The $100(1-\alpha)\%$ asymptotic confidence interval is

$$CI_c = \left[\hat{\theta} \pm z_{\alpha/2}\hat{\sigma}_c\right],$$

and it has asymptotically correct coverage for $\theta_{0(n)}$.

The HH bootstrap GMM estimator $\hat{\theta}^*_{HH}$ minimizes (3.8) with the weight matrix

$$\bar{W}^*_n = \left(\frac{1}{n}\sum_i (g(X^*_i, \bar{Z}^*) - g_n(\hat{\theta}))(g(X^*_i, \bar{Z}^*) - g_n(\hat{\theta}))'\right)^{-1},$$

where $\bar{Z}^* = n^{-1}\sum_{i=1}^{n} Z^*_i$ and is given by

$$\hat{\theta}^*_{HH} = \bar{Z}^* - \frac{\widehat{Cov}(Y^*_i, Z^*_i) - (\bar{Y}^* - \bar{Y})\frac{\widehat{Cov}(Y_i,Z_i)}{\overline{Y^2}}\bar{Y}}{\widehat{Var}(Y^*_i) + (\bar{Y}^* - \bar{Y})^2}(\bar{Y}^* - \bar{Y}) - \frac{\widehat{Cov}(Y_i, Z_i)}{\overline{Y^2}}\bar{Y},$$

where $\widehat{Cov}(Y^*_i, Z^*_i) = n^{-1}\sum_{i=1}^{n}(Y^*_i - \bar{Y}^*)(Z^*_i - \bar{Z}^*)$ and $\widehat{Var}(Y^*_i) = n^{-1}\sum_{i=1}^{n}(Y^*_i - \bar{Y}^*)^2$. Due to the recentering, $\hat{\theta}^*_{HH}$ is not the bootstrap analogue of $\hat{\theta}$, and in other words, the sample analogue of $\hat{\theta}^*_{HH}$ is $\hat{\theta}_{HH}$ which differs from $\hat{\theta}$. The recentered version of the sample GMM estimator, $\hat{\theta}_{HH}$, is given by

$$\begin{aligned}\hat{\theta}_{HH} &= \bar{Z} - \frac{\widehat{Cov}(Y_i, Z_i) - (\bar{Y} - EY_i)\frac{Cov(Y_i,Z_i)}{EY_i^2}EY_i}{\widehat{Var}(Y_i) + (\bar{Y} - EY_i)^2}(\bar{Y} - EY_i) - \frac{Cov(Y_i, Z_i)}{EY_i^2}EY_i, \\ &= \hat{\theta} + C_{1n}(\bar{Y} - EY_i) + C_{2n}(\widehat{Cov}(Y_i, Z_i) - Cov(Y_i, Z_i)) + C_{3n}\left(\frac{1}{\overline{Y^2}} - \frac{1}{EY_i^2}\right) \\ &\quad + C_{4n}(\bar{Y} - EY_i)^2, \qquad (3.10)\end{aligned}$$

where C_{1n}, C_{2n}, C_{3n}, and C_{4n} are

$$\begin{aligned}
C_{1n} &= \widehat{Cov}(Y_i, Z_i)\left(\frac{1}{\overline{\overline{Y^2}}} - \frac{1}{\widehat{Var}(Y_i) + (\bar{Y} - EY_i)^2}\right) = O_p(1), \\
C_{2n} &= \frac{EY_i}{\overline{\overline{Y^2}}} = O_p(1), \\
C_{3n} &= Cov(Y_i, Z_i) EY_i, \\
C_{4n} &= \frac{EY_i}{EY_i^2} \cdot \frac{Cov(Y_i, Z_i)}{\widehat{Var}(Y_i) + (\bar{Y} - EY_i)^2} = O_p(1).
\end{aligned}$$

The variance estimator $\hat{\sigma}_{HH}^2$ is calculated similarly:

$$\begin{aligned}
\hat{\sigma}_{HH}^2 &= n^{-1} \sum_{i=1}^{n} \left(Z_i - \hat{\theta} - \frac{Cov(Y_i, Z_i)}{EY_i^2} EY_i \right)^2 \\
&\quad - \frac{\left(n^{-1} \sum_{i=1}^{n} (Y_i - EY_i) \left(Z_i - \hat{\theta} - \frac{Cov(Y_i, Z_i)}{EY_i^2} EY_i \right) \right)^2}{n^{-1} \sum_{i=1}^{n} (Y_i - EY_i)^2}.
\end{aligned}$$

By using (3.10) and the fact that $\hat{\sigma}_c^2 - \hat{\sigma}_{HH}^2 = o_p(1)$, we have

$$\begin{aligned}
\frac{\sqrt{n}(\hat{\theta}_{HH} - \theta_{0(n)})}{\hat{\sigma}_{HH}} &= \frac{\sqrt{n}(\hat{\theta} - \theta_{0(n)})}{\hat{\sigma}_c} + \xi_n + \frac{\sqrt{n}(\hat{\theta}_{HH} - \theta_{0(n)})}{\hat{\sigma}_{HH}} \left(1 - \frac{\hat{\sigma}_{HH}}{\hat{\sigma}_c} \right) + O_p(n^{-1/2}) \\
&= \frac{\sqrt{n}(\hat{\theta} - \theta_{0(n)})}{\hat{\sigma}_c} + \xi_n + o_p(1), \qquad (3.11)
\end{aligned}$$

where

$$\xi_n = \frac{C_{1n}}{\hat{\sigma}_c} \sqrt{n}(\bar{Y} - EY_i) + \frac{C_{2n}}{\hat{\sigma}_c} \sqrt{n}(\widehat{Cov}(Y_i, Z_i) - Cov(Y_i, Z_i)) + \frac{C_{3n}}{\hat{\sigma}_c} \sqrt{n} \left(\frac{1}{\overline{\overline{Y^2}}} - \frac{1}{EY_i^2} \right),$$

and $\xi_n \to_d \xi \sim N(0, V_\xi)$, for some positive definite matrix V_ξ. Since both $T_{HH,n}$ and T_n are asymptotically standard normal, we have $V_\xi + Cov(T_n, \xi_n) = 0$. Unless the higher-order cumulants of ξ_n, such as skewness and kurtosis, equal to zero, (3.11) shows that the distribution of $T_{HH,n}$ would be differ from that of T_n in error by $n^{-1/2}$.

Let the true data generating process (DGP) be

$$\text{DGP 1} \quad : \quad \begin{pmatrix} Y_i \\ Z_i \end{pmatrix}_{i\le n} \sim N\left(\begin{pmatrix} \delta/\sqrt{n} \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix} \right),$$

$$\text{DGP 2} \quad : \quad \text{Generate data from DGP 1, then use } \begin{pmatrix} Y_i \\ Z_i^T \end{pmatrix}_{i\le n} \text{ where } Z_i^T = e^{Z_i} - e^{1/2},$$

where $(Y_i, Z_i)'$ is iid given n. Figure 1 and 2 show the sampling distribution of $T_{HH,n}$ and T_n. The number of Monte Carlo repetition is 10,000 and the distributions are nonparametrically estimated. When data is generated by the DGP 1, the distributions of T_n and $T_{HH,n}$ differ greatly when $n = 25$, in particular with respect to the variance and the skewness, as is shown in Figure 1. Figure 2 shows similar results when data is generated by DGP 2, with emphasis on the difference in the mean and the kurtosis of the distributions. When I increase the sample size (to $n = 250$ when DGP 1, and to $n = 2500$ when DGP 2) the distributions of T_n and $T_{HH,n}$ become similar, close to the standard normal distribution. Table 1 compares the first fourth cumulants of the sampling distributions of T_n and $T_{HH,n}$.

Now I compare the coverage probabilities of the asymptotic confidence interval CI_n and the HH bootstrap confidence interval $CI^*_{HH,n}$ under local misspecification.

	DGP 1				DGP 2			
	$n = 25$		$n = 250$		$n = 25$		$n = 2500$	
	T_n	$T_{HH,n}$	T_n	$T_{HH,n}$	T_n	$T_{HH,n}$	T_n	$T_{HH,n}$
Mean	-0.11	-0.16	0.04	0.02	0.008	-0.51	-0.14	-0.18
Variance	4.43	1.10	1.40	1.02	1.01	1.89	0.73	1.03
Skewness	-0.18	-0	0.14	-0.01	-0.63	-0.89	-0.22	-0.26
Kurtosis	3.17	3.24	3.09	2.99	5.35	4.02	3.15	3.15

Table 3.1: Moments of the Sampling Distribution of T_n and $T_{HH,n}$

The 90% confidence intervals are constructed as

$$\begin{aligned} CI_n &= \left[\hat{\theta} \pm 1.645 \cdot \frac{\hat{\sigma}_c}{\sqrt{n}}\right], \\ CI^*_{HH,n} &= \left[\hat{\theta} \pm z^*_{|T|,0.90} \cdot \frac{\hat{\sigma}_c}{\sqrt{n}}\right], \end{aligned}$$

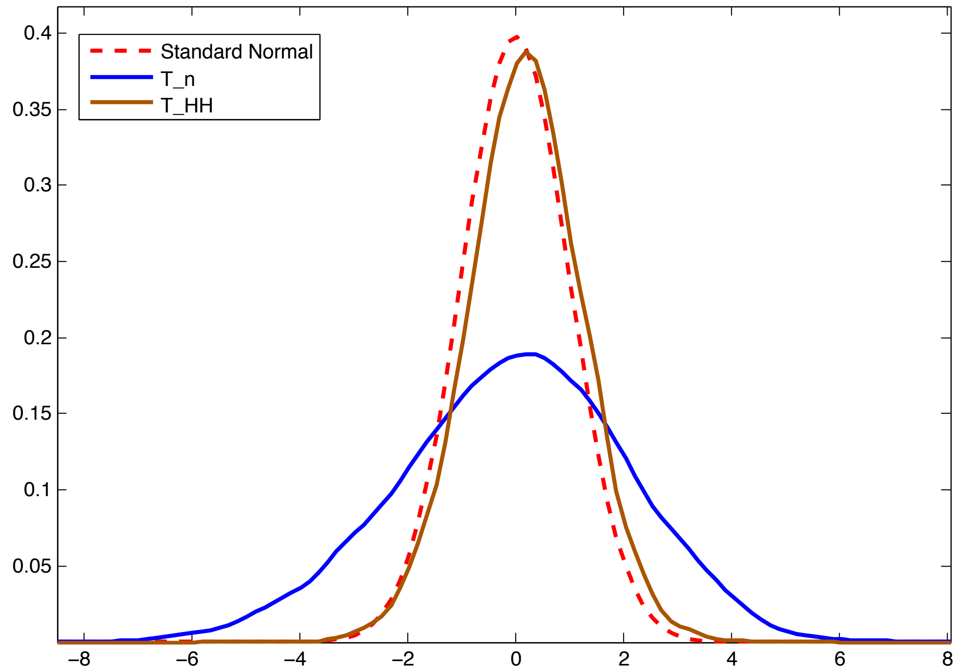
where $z^*_{|T|,0.90}$ is the upper 10th quantile of the distribution of $|T^*_{HH,n}|$. This type of bootstrap confidence interval is called the symmetric percentile-t interval. Hall and Horowitz (1996) and Andrews (2002) show that

$$\begin{aligned} P(\theta_0 \in CI_n) &= 0.90 + O(n^{-1}), \\ P(\theta_0 \in CI^*_{HH,n}) &= 0.90 + O(n^{-2}), \end{aligned}$$

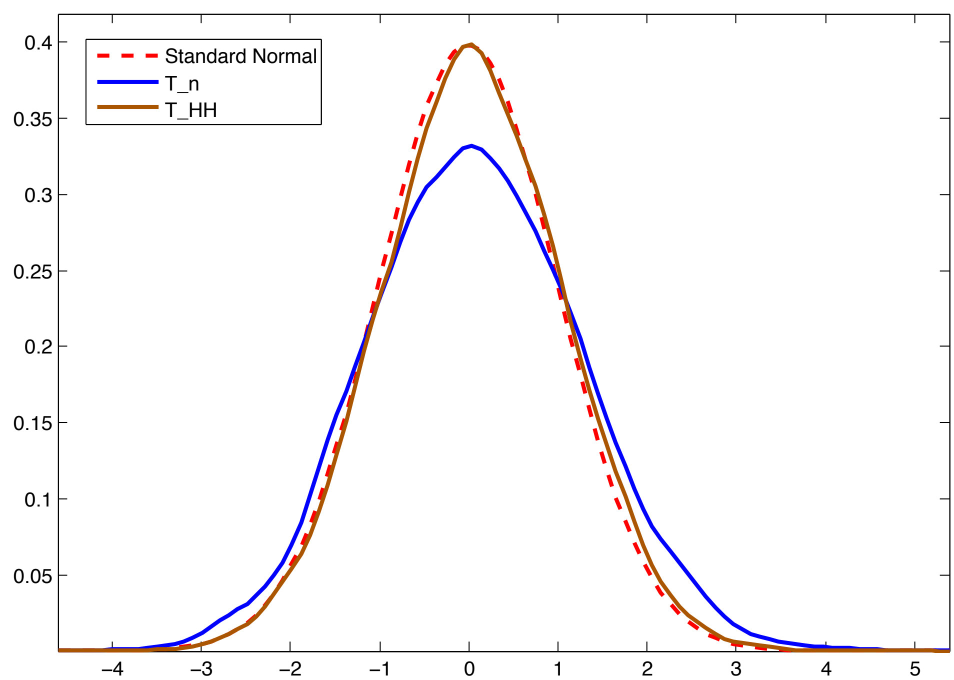
under correct specification. Under local misspecification, we replace θ_0 with $\theta_{0(n)}$. The question is that $CI^*_{HH,n}$ still enjoys refinements over CI_n. Suppose that the HH bootstrap achives asymptotic refinements of the size of n^{-1}. Given the size of the error when $n = 25$, if we increase the sample size to $n = 250$, then the error in $CI^*_{HH,n}$ would decrease by 1/100, while the error in CI_n would decrease by 1/10. If the HH bootstrap fails to achieve asymptotic refinements, then the errors in the coverage probabilities will decrease in a similar rate. Figure 3 shows the actual coverage probabilities of CI_n and $CI^*_{HH,n}$ under local misspecification and correct specification for different sample sizes. When the model is locally misspecified (Figure 3(a)), the actual coverage of the HH bootstrap confidence interval converges to the nominal rate at the same rate as the asymptotic confidence interval. In contrast, the gap between the coverages of the HH bootstrap confidence interval and the asymptotic confidence interval becomes smaller under correct specification. This supports asymptotic refinements of the HH bootstrap under correct specification.

3.3 Conclusion

Bootstrap confidence intervals are often believed to perform better than asymptotic confidence intervals, because of asymptotic refinements of the bootstrap. For GMM, Hall and Horowitz (1996) and Andrews (2002) establish asymptotic refinements of the bootstrap using an ad-hoc procedure, the recentering. The recentered bootstrap works under correct specification, but what if the model is locally misspecified? This paper answers this question by showing that the conventional bootstrap methods for GMM are first-order valid but it does not improve upon first-order asymptotics anymore under local misspecificaiton. A simple example and Monte Carlo experiment result are provided.

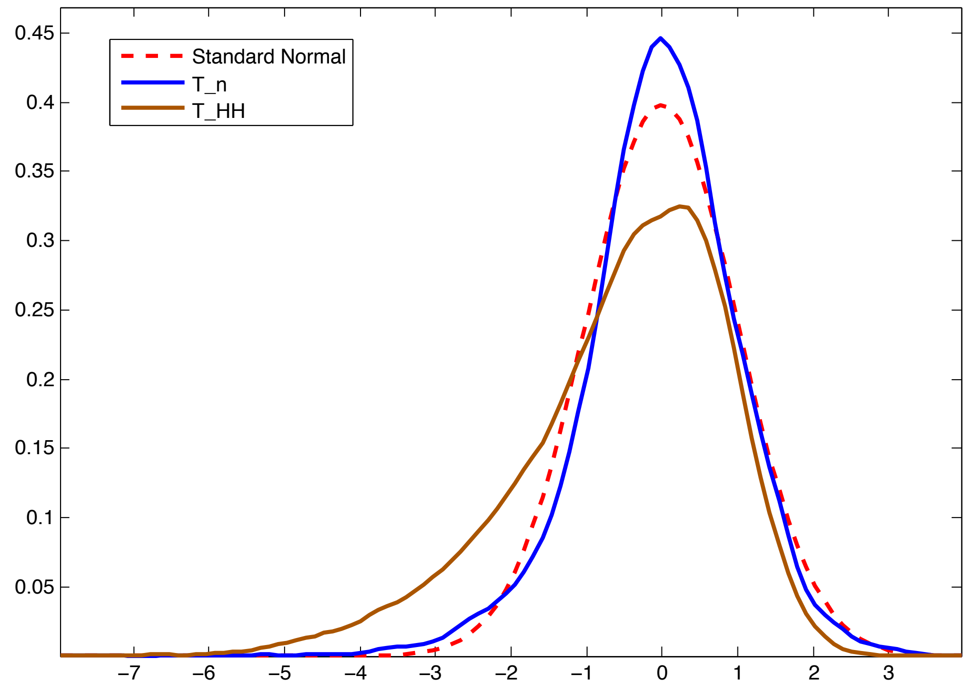


(a) Local Misspecification: $\delta = 5$ and $n = 25$

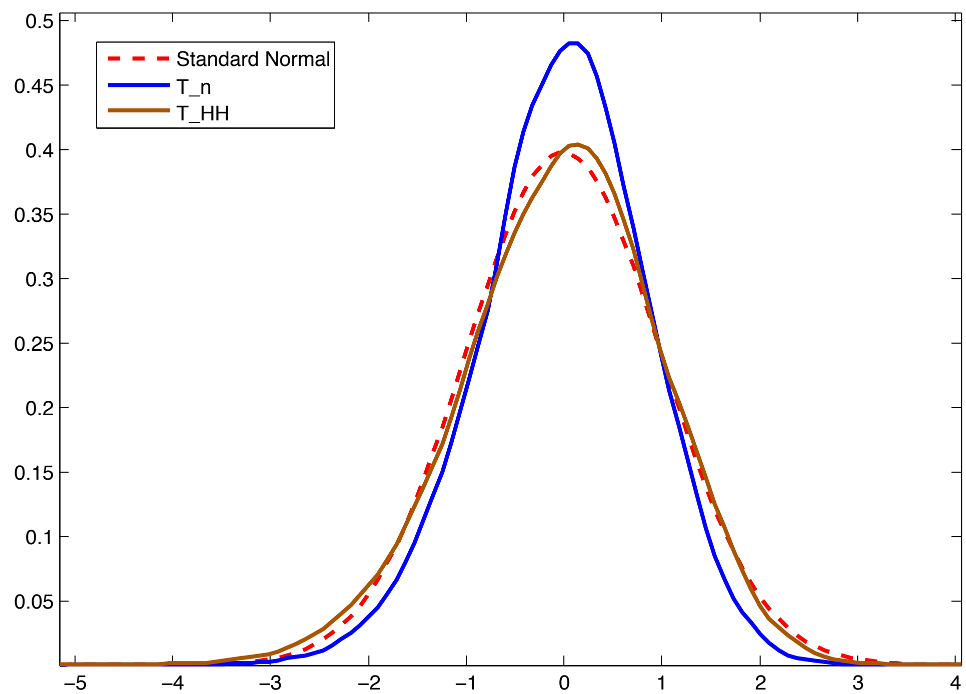


(b) Local Misspecification: $\delta = 5$ and $n = 250$

Figure 3.1: Distribution of T_n and $T_{HH,n}$ when DGP 1

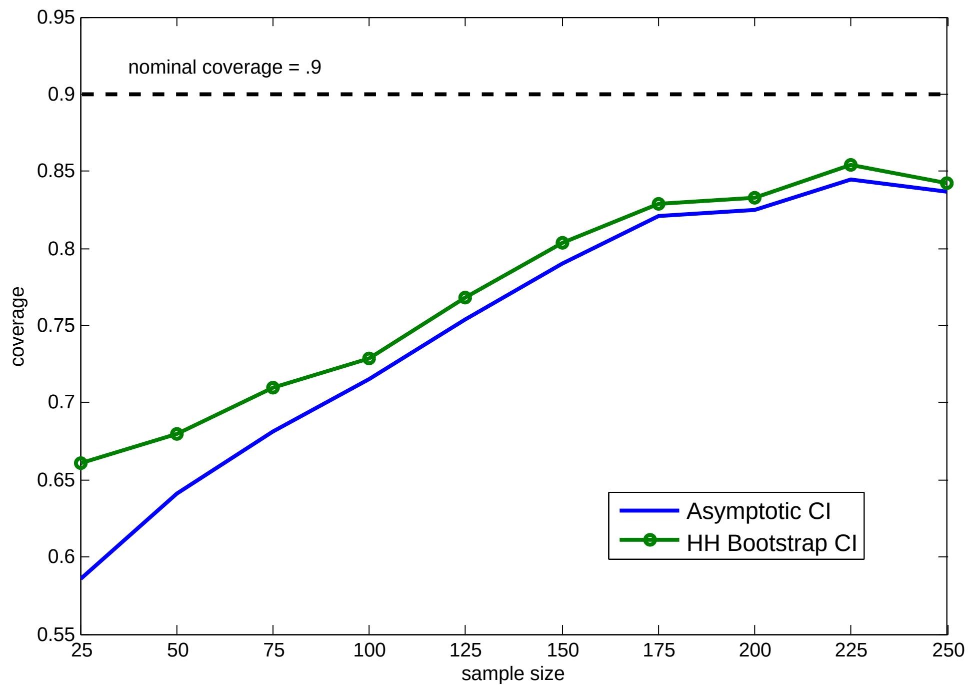


(a) Local Misspecification: $\delta = 5$ and $n = 25$

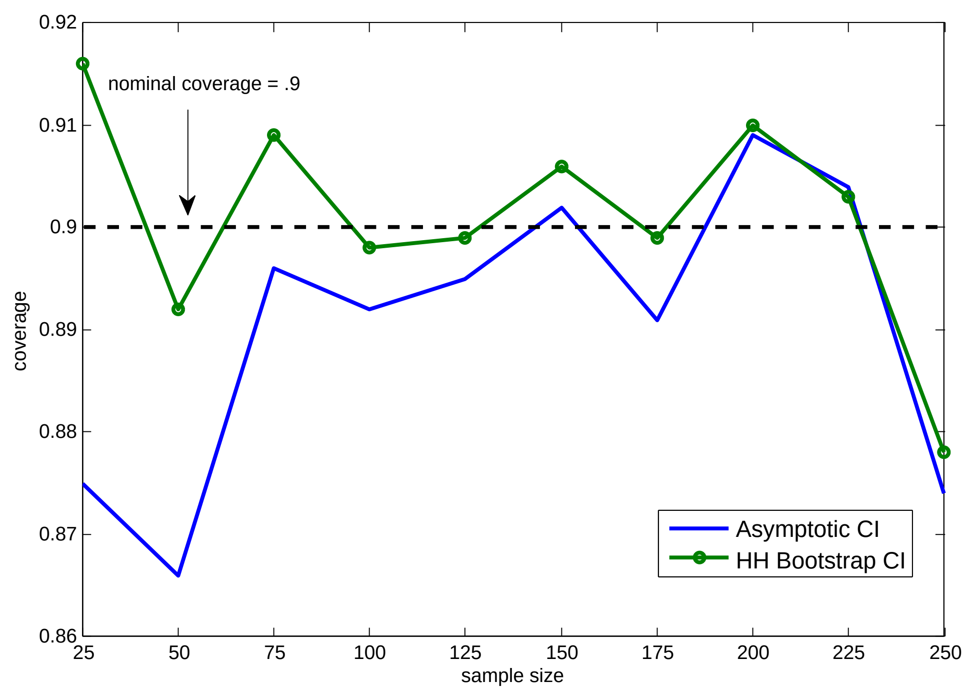


(b) Local Misspecification: $\delta = 5$ and $n = 2500$

Figure 3.2: Distribution of T_n and $T_{HH,n}$ when DGP 2



(a) Local Misspecification: $\delta = 5$



(b) Correct Specification: $\delta = 0$

Figure 3.3: Coverage Probabilities of the Asymptotic and the HH Bootstrap Confidence Intervals:

REFERENCES

Agüero, J.M., and M.S. Marks. 2008. Motherhood and female labor force participation: evidence from infertility shocks. *The American Economic Review* 98(2): 500–504.

Allen, J., A.W. Gregory, and K. Shimotsu. 2011. Empirical likelihood block bootstrapping. *Journal of Econometrics* 161:110–121.

Andrews, D.W.K. 2001. Higher-order improvements of a computationally attractive-step bootstrap for extremum estimators. *Cowles Foundation Discussion Papers*.

———. 2002. Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica* 70(1):119–162.

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

Bhattacharya, R.N. 1987. Some aspects of edgeworth expansions in statistics and probability. *New Perspectives in Theoretical and Applied Statistics* 157–170.

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

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(1):139–164.

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

Canay, I.A. 2010. El inference for partially identified models: Large deviations optimality and bootstrap validity. *Journal of Econometrics* 156(2):408–425.

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(1):109–140.

French, E., and J.B. Jones. 2004. On the distribution and dynamics of health care costs. *Journal of Applied Econometrics* 19(6):705–721.

Gonçalves, S., and H. White. 2004. Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics* 119(1):199–219.

Gowrisankaran, G., and M. Rysman. 2009. Dynamics of consumer demand for new durable goods. Tech. Rep., National Bureau of Economic Research.

Hahn, J. 1996. A note on bootstrapping generalized method of moments estimators. *Econometric Theory* 12(01):187–197.

Hall, A.R. 2000. Covariance matrix estimation and the power of the overidentifying restrictions test. *Econometrica* 68(6):1517–1527.

———. 2005. *Generalized method of moments*. Oxford University Press, USA.

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

Hall, A.R., and D. Pelletier. 2011. Nonnested testing in models estimated via generalized method of moments. *Econometric Theory* 27:443–456.

Hall, P. 1988. On symmetric bootstrap confidence intervals. *Journal of the Royal Statistical Society. Series B (Methodological)* 35–45.

———. 1997. *The bootstrap and edgeworth expansion*. Springer Verlag.

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

Hansen, B.E. 2012. *Econometrics*. Unpublished manuscript.

Hansen, L.P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 1029–1054.

Hansen, L.P., J. Heaton, and A. Yaron. 1996. Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics* 262–280.

Hellerstein, J.K., and G.W. Imbens. 1999. Imposing moment restrictions from auxiliary data by weighting. *Review of Economics and Statistics* 81(1):1–14.

Horowitz, J.L. 2001. The bootstrap. *Handbook of Econometrics* 5:3159–3228.

———. 2003. Bootstrap methods for markov processes. *Econometrica* 71(4):1049–1082.

Imbens, G.W. 1997. One-step estimators for over-identified generalized method of moments models. *The Review of Economic Studies* 64(3):359–383.

———. 2002. Generalized method of moments and empirical likelihood. *Journal of Business and Economic Statistics* 20(4):493–506.

Imbens, G.W., and T. Lancaster. 1994. Combining micro and macro data in microeconometric models. *The Review of Economic Studies* 61(4):655–680.

Imbens, G.W., R.H. Spady, and P. Johnson. 1998. Information theoretic approaches to inference in moment condition models. *Econometrica* 333–357.

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(2):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.*

Kitamura, Y., and M. Stutzer. 1997. An information-theoretic alternative to generalized method of moments estimation. *Econometrica* 861–874.

Maasoumi, E., and P.C.B. Phillips. 1982. On the behavior of inconsistent instrumental variable estimators. *Journal of Econometrics* 19(2):183–201.

Marmer, V., and T. Otsu. 2009. Optimal comparison of misspecified moment restriction models under a chosen measure of fit. *Unpublished working paper*.

Newey, W.K., and D. McFadden. 1994. Large sample estimation and hypothesis testing. *Handbook of Econometrics* 4:2111–2245.

Newey, W.K., and R.J. Smith. 2004. Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica* 72(1):219–255.

Owen, A.B. 1988. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75(2):237–249.

———. 1990. Empirical likelihood ratio confidence regions. *The Annals of Statistics* 18(1):90–120.

———. 2001. *Empirical likelihood*, vol. 92. CRC press.

Parker, J.A., and C. Julliard. 2005. Consumption risk and the cross section of expected returns. *Journal of Political Economy* 113(1):185–222.

Qin, J., and J. Lawless. 1994. Empirical likelihood and general estimating equations. *The Annals of Statistics* 300–325.

Rivers, D., and Q. Vuong. 2002. Model selection tests for nonlinear dynamic models. *The Econometrics Journal* 5(1):1–39.

Sawa, T. 1978. Information criteria for discriminating among alternative regression models. *Econometrica* 1273–1291.

Schennach, S.M. 2007. Point estimation with exponentially tilted empirical likelihood. *The Annals of Statistics* 35(2):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* 307–333.

White, H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 1–25.