Australian School of Business Research Paper No. 2014 ECON 02

Asymptotic Refinements of a Misspecification-Robust Bootstrap for GEL Estimators

Seojeong Lee

Asymptotic Refinements of a Misspecification-Robust Bootstrap for GEL Estimators

Seojeong Lee*†

January 21, 2014

Abstract

I propose a nonparametric iid bootstrap procedure for the empirical likelihood, the exponential tilting, and the exponentially tilted empirical likelihood estimators that achieves sharp asymptotic refinements for t tests and confidence intervals based on such estimators. Furthermore, the proposed bootstrap is robust to model misspecification, i.e., it achieves asymptotic refinements regardless of whether the assumed moment condition model is correctly specified or not. This result is new, because asymptotic refinements of the bootstrap based on these estimators have not been established in the literature even under correct model specification. Monte Carlo experiments are conducted in dynamic panel data setting to support the theoretical finding. As an application, bootstrap confidence intervals for the returns to schooling of Hellerstein and Imbens (1999) are calculated. The returns to schooling may be higher.

Keywords: generalized empirical likelihood, bootstrap, asymptotic refinement, model misspecification

JEL Classification: C14, C15, C31, C33

*I am very grateful to Bruce Hansen and Jack Porter for their encouragement and helpful comments. I also thank Guido Imbens, Xiaohong Chen, and Yoon-Jae Whang, as well as seminar participants at UW-Madison, Monash, ANU, Adelaide, UNSW, and U of Sydney for their suggestions and comments. This paper was also presented at the 2013 NASM and SETA 2013.

†School of Economics, Australian School of Business, UNSW, Sydney NSW 2052 Australia, jay.lee@unsw.edu.au, https://sites.google.com/site/misspecified/

1 Introduction

This paper establishes asymptotic refinements of the nonparametric iid bootstrap for t tests and confidence intervals (CI's) based on the empirical likelihood (EL), the exponential tilting (ET), and the exponentially tilted empirical likelihood (ETEL) estimators. This is done without recentering the moment function in implementing the bootstrap, which has been considered as a critical procedure for overidentified moment condition models. Moreover, the proposed bootstrap is robust to misspecification, i.e., the resulting bootstrap CI's achieve asymptotic refinements for the true parameter when the model is correctly specified, and the same rate of refinements is achieved for the pseudo-true parameter when misspecified. This is a new result because in the existing literature, there is no formal proof for asymptotic refinements of the bootstrap for EL, ET, or ETEL estimators even under correct specification. In fact, any bootstrap procedure with recentering for these estimators would be inconsistent if the model is misspecified because recentering imposes the correct model specification in the sample. This paper is motivated by three questions: (i) Why these estimators? (ii) Why bootstrap? (iii) Why care about misspecification?

First of all, EL, ET, and ETEL estimators are used to estimate a finite dimensional parameter characterized by a moment condition model. Traditionally, the generalized method of moments (GMM) estimators of Hansen (1982) have been used to estimate such models. However, it is well known that the two-step GMM may suffer from finite sample bias and inaccurate first-order asymptotic approximation to the finite sample distribution of the estimator when there are many moments, the model is non-linear, or instruments are weak. See Altonji and Segal (1996) and Hansen, Heaton, and Yaron (1996) among others on this matter.

Generalized empirical likelihood (GEL) estimators of Newey and Smith (2004) are alternatives to GMM as they have smaller asymptotic bias. GEL circumvents 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, 1990), Qin and Lawless (1994), and Imbens (1997), the ET estimator of Kitamura and Stutzer (1997) and Imbens, Spady, and Johnson (1998), the continuously updating (CU) estimator of Hansen, Heaton, and Yaron (1996), and the minimum Hellinger distance estimator (MHDE) of Kitamura, Otsu, and Evdokimov (2013). Newey and Smith (2004) show that EL has the most favorable higher-order asymptotic properties than other GEL estimators. Although EL is preferable to other GEL estimators as well as GMM estimators, its nice properties no longer holds under misspecification. In contrast, ET is often considered as robust to misspecification. Schennach (2007) proposes the ETEL estimator that shares the same higher-order property with EL under correct specification while possessing robustness of ET under misspecification. Hence, this paper considers the most widely used, EL, the most robust, ET, and a hybrid of the two, ETEL.[1] An extension of the result to other GEL estimators is possible, but not attempted to make the argument succinct.

Secondly, many efforts have been made to accurately approximate the finite sample distribution of GMM. These include analytic correction of the GMM standard errors by Windmeijer (2005) and the bootstrap by Hahn (1996), Hall and Horowitz (1996), Andrews (2002), Brown and Newey (2002), Inoue and Shintani (2006), Allen, Gregory, and Shimotsu (2011), Lee (2014), among others. The bootstrap tests and CI's based on the GMM estimators achieve asymptotic refinements over the first-order asymptotic tests and CI's, which means their actual test rejection probability and CI coverage probability have smaller errors than the asymptotic tests and CI's. In particular, Lee (2014) applies a similar idea of non-recentering to GMM estimators by using Hall and Inoue (2003)'s misspecification-robust variance estimators to achieve the same sharp rate of refinements with Andrews (2002).

Although GEL estimators are favorable alternatives to GMM, there is little evidence that the finite sample distribution of GEL test statistics is well approximated by the first-order asymptotics. Guggenberger and Hahn (2005) and Guggenberger (2008) find by simulation studies that the first-order asymptotic approximation to the finite sample distribution of EL estimators may be poor. Thus, it is natural to consider bootstrap t tests and CI's based on GEL estimators to improve upon the first-order asymptotic approximation. However, few published papers deal with bootstrapping for GEL. Brown and Newey (2002) and Allen, Gregory, and Shimotsu (2011) employ the EL implied probability in resampling for GMM estimators, but not for GEL estimators. Canay (2010) shows the validity of a bootstrap procedure for the EL ratio statistic in the moment inequality setting. Kundhi and Rilstone

[1]Precisely speaking, ETEL is not a GEL estimator. However, the analysis is quite similar because it is a combination of the two GEL estimators. Therefore, this paper uses the term "GEL" to include ETEL as well as EL and ET to save space and to prevent any confusion.

(2012) argue that analytical corrections by Edgeworth expansion of the distribution of GEL estimators work well compared to the bootstrap, but they assume correct model specification.

Lastly, the validity of inferences and CI's critically depends on the correctly specified model assumption. Although model misspecification can be asymptotically detected by an overidentifying restrictions test, there is always a possibility that one does not reject a misspecified model or reject a correctly specified model in finite sample. Moreover, there is a view that all models are misspecified and will be rejected asymptotically. The consequences of model misspecification are twofold: a potentially biased probability limit of the estimator and a different asymptotic variance. The former is called the pseudo-true value, and it is impossible to correct the bias in general. Nevertheless, there are cases such that the pseudo-true values are still the object of interest: see Hansen and Jagannathan (1997), Hellerstein and Imbens (1999), Bravo (2010), and Almeida and Garcia (2012). GEL pseudo-true values are less arbitrary than GMM ones because the latter depend on a weight matrix, which is an arbitrary choice by a researcher. In contrast, each of the GEL pseudo-true values can be interpreted as a unique minimizer of a well-defined discrepancy measure, e.g. Schennach (2007).

The asymptotic variance of the estimator, however, can be consistently estimated even under misspecification. If a researcher wants to minimize the consequence of model misspecification, a misspecification-robust variance estimator should be used for t tests or confidence intervals. The proposed bootstrap uses the misspecification-robust variance estimator for EL, ET, and ETEL in constructing the t statistic. This makes the proposed bootstrap robust to misspecification without recentering, and enables researchers to make valid inferences and CI's against unknown misspecification.

The remainder of the paper is organized as follows. Section 2 explains the idea of non-recentering by using a misspecification-robust variance estimator for the t statistic. Section 3 defines the estimators and the t statistic. Section 4 describes the nonparametric iid misspecification-robust bootstrap procedure. Section 5 states the assumptions and establishes asymptotic refinements of the misspecification-robust bootstrap. Section 6 presents Monte Carlo experiments. An application to estimate the returns to schooling of Hellerstein and Imbens (1999) is presented in Section 7. Section 8 concludes the paper. Lemmas and proofs are collected in Appendix A.

2 Outline of the Results

This section explains why the proposed procedure achieves asymptotic refinements without recentering. The key idea is to construct an asymptotically pivotal statistic regardless of misspecification. Bootstrapping an asymptotically pivotal statistic is critical to get asymptotic refinements of the bootstrap (e.g. see Beran, 1988; Hall, 1992; Hall and Horowitz, 1996; Horowitz, 2001; Andrews, 2002; and Brown and Newey, 2002). That is, the asymptotic distribution of the test statistic should not depend on unknown population quantities or data generating process (DGP), under the null hypothesis. Thus, we need to construct the t statistic that converges in distribution to the standard normal, both in the sample and in the bootstrap sample. Usually, there is no need to treat the bootstrap sample or statistic specially. For overidentified moment condition models, however, it is important to understand the impact of overidentification when constructing the t statistic in the bootstrap sample.

Suppose that $\chi_n = \{X_i : i \leq n\}$ is an independent and identically distributed (iid) sample. Let F be the corresponding cumulative distribution function (cdf). Let θ be a parameter of interest and $g(X_i, \theta)$ be a moment function. The moment condition model is correctly specified if

$$H_C : Eg(X_i, \theta_0) = 0 \tag{2.1}$$

for a unique θ_0. The hypothesis is denoted by H_C. The hypothesis of interest is

$$H_0 : \theta = \theta_0. \tag{2.2}$$

The usual t statistic T_C is asymptotically standard normal under H_0 *and* H_C.

Now define the bootstrap sample. Let $\chi^*_{n_b} = \{X^*_i : i \leq n_b\}$ be a random draw with replacement from χ_n according to the empirical distribution function (edf) F_n. In this section, I distinguish the number of sample n and the number of bootstrap sample n_b, which helps understand the concept of the conditional asymptotic distribution.[2]

[2] n_b should be distinguished from the number of bootstrap repetition, often denoted by B. For more discussions, see Bickel and Freedman (1981).

The bootstrap versions of H_C and H_0 are

$$H_C^*: \quad E^* g(X_i^*, \hat{\theta}) = 0, \tag{2.3}$$

$$H_0^*: \quad \theta = \hat{\theta}, \tag{2.4}$$

where E^* is the expectation taken over the bootstrap sample and $\hat{\theta}$ is a GEL estimator. Note that $\hat{\theta}$ is considered as the true value in the bootstrap world. The bootstrap version of the usual t statistic T_C^*, however, is not asymptotically pivotal conditional on the sample because H_C^* is not satisfied in the sample if the model is overidentified:

$$E^* g(X_i^*, \hat{\theta}) = n^{-1} \sum_i^n g(X_i, \hat{\theta}) \neq 0. \tag{2.5}$$

Thus, Hall and Horowitz (1996), Andrews (2002), and Brown and Newey (2002) recenter the bootstrap version of the moment function to satisfy H_C^*. The resulting t statistic based on the recentered moment function, $T_{C,R}^*$, tends to the standard normal distribution as n_b grows conditional on the sample almost surely, and asymptotic refinements of the bootstrap are achieved.

This paper takes a different approach. Instead of jointly testing H_C and H_0, I solely focus on H_0, leaving that H_C may not hold. If the model is misspecified, then there is no such θ that satisfies H_C:

$$Eg(X_i, \theta) \neq 0, \forall \theta \in \Theta, \tag{2.6}$$

where Θ is a compact parameter space. This may happen only if the model is overidentified. Since there is no true value, the pseudo-true value θ_0 should be defined. Instead of H_C, θ_0 is defined as a unique minimizer of the population version of the empirical discrepancy used in the estimation. For EL, this discrepancy is the Kullback-Leibler Information Criterion (KLIC). For ET, it maximizes a quantity named entropy. This definition is more flexible since it includes correct specification as a special case when H_C holds at θ_0. Without assuming H_C, we can find regularity conditions for $\sqrt{n}-$consistency and asymptotic normality of $\hat{\theta}$ for the pseudo-true value θ_0. Assume such regularity conditions hold. Then, we have

$$\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, \Sigma_{MR}), \tag{2.7}$$

as the sample size grows where the asymptotic variance matrix Σ_{MR} is different from the standard one. Σ_{MR} can be consistently estimated using the formula given in the next section. Let $\hat{\Sigma}_{MR}$ be a consistent estimator for Σ_{MR}. The misspecification-robust t statistic is given by[3]

$$T_{MR} = \frac{\hat{\theta} - \theta_0}{\sqrt{\hat{\Sigma}_{MR}/n}}, \tag{2.8}$$

and T_{MR} is asymptotically standard normal under H_0, without assuming H_C.

Similarly, the bootstrap version of the t statistic is

$$T^*_{MR} = \frac{\hat{\theta}^* - \hat{\theta}}{\sqrt{\hat{\Sigma}^*_{MR}/n_b}}, \tag{2.9}$$

where $\hat{\theta}^*$ and $\hat{\Sigma}^*_{MR}$ are calculated using the same formula with $\hat{\theta}$ and $\hat{\Sigma}_{MR}$. Conditional on the sample almost surely, T^*_{MR} tends to the standard normal distribution as n_b grows under H^*_0. Since the conditional asymptotic distribution does not depend on H^*_C, we need not recenter the bootstrap moment function to satisfy H^*_C. In other words, the misspecification-robust t statistic T_{MR} is asymptotically pivotal under H_0, while the usual t statistic T_C is asymptotically pivotal under H_0 *and* H_C. This paper develops a theory for bootstrapping T_{MR}, instead of T_C. Note that both can be used to test the null hypothesis $H_0 : \theta = \theta_0$ under correct specification. Under misspecification, however, only T_{MR} can be used to test H_0 because T_C is not asymptotically pivotal. This is useful when the pseudo-true value is an interesting object even if the model is misspecified.

To find the formula for Σ_{MR}, I use a just-identified system of the first-order conditions (FOC's) of EL, ET, and ETEL estimators. This idea is not new, though. Schennach (2007) uses the same idea to find the asymptotic variance matrix of the ETEL estimator robust to misspecification. For GMM estimators, the idea of rewriting the overidentified GMM as a just-identified system appears in Imbens (1997,2002) and Chamberlain and Imbens (2003). Hall and Inoue (2003) find the formula for Σ_{MR} of GMM estimators by expanding the FOC. They show that the formula is different from the one under correct specification, but it coincides with the standard GMM variance matrix if the model is correctly specified.

A natural question is whether we can use GEL implied probabilities to construct

[3] For notational brevity, let θ and Σ_{MR} be scalars in this section.

the cdf estimator $\hat{F}$ and use it instead of the edf 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) argue that using the EL-estimated cdf $\hat{F}_{EL}(z) \equiv \sum_i \mathbf{1}(X_i \leq z)p_i$, where p_i is the EL implied probability, in place of the edf F_n in resampling would improve efficiency of bootstrapping for GMM. Their argument relies on the fact that $\hat{F}_{EL}$ is an efficient estimator of the true cdf F. If the population moment condition is misspecified, however, then the cdf estimator based on the implied probability is inconsistent for F because $E^*g(X_i^*, \hat{\theta}) = 0$ holds even in large sample, while $Eg(X_i, \theta_0) \neq 0$. In contrast, the edf F_n is uniformly consistent for F regardless of whether the population moment condition holds or not by Glivenko-Cantelli Theorem. For this reason, I mainly focus on resampling from F_n rather than $\hat{F}$ in this paper. However, a shrinkage-type cdf estimator combining F_n and $\hat{F}$, similar to Antoine, Bonnal, and Renault (2007), can be used to improve both robustness and efficiency. For example, a shrinkage that has the form

$$\pi_i = \epsilon_n \cdot p_i + (1 - \epsilon_n) \cdot n^{-1}, \;\; \epsilon_n \to 0 \text{ as } n \text{ grows}, \tag{2.10}$$

where p_i is a GEL implied probability, would work with the proposed misspecification-robust bootstrap because

$$E_\pi^* g(X_i, \hat{\theta}) = (1 - \epsilon_n)n^{-1} \sum_i^n g(X_i, \hat{\theta}) \neq 0, \tag{2.11}$$

where the expectation is taken with respect to $\hat{F}_\pi(z) \equiv \sum_i \mathbf{1}(X_i \leq z)\pi_i$. A promising simulation result using this shrinkage estimator in resampling is provided in Section 6.

Note that the definition of misspecification considered in this paper is different from that of White (1982). In his quasi-maximum likelihood (QML) framework, the underlying cdf is misspecified. Since the QML theory deals with just-identified models where the number of parameters is equal to the number of moment restrictions, (2.1) holds even if the underlying cdf is misspecified. Hence, the model is not misspecified in this paper's framework. For bootstrapping QML estimators, see Gonçalves and White (2004).

3 Estimators and Test Statistics

Let $g(X_i,\theta)$ be an $L_g \times 1$ moment function where $\theta \in \Theta \subset \mathbf{R}^{L_\theta}$ is a parameter of interest, where $L_g \geq L_\theta$. 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 an $L_g \times L_\theta$ matrix and $G^{(2)}(X_i,\theta) \equiv (\partial/\partial\theta')vec\{G(X_i,\theta)\}$ is an $L_g L_\theta \times L_\theta$ matrix, where $vec\{\cdot\}$ is the vectorization of a matrix. To simplify notation, write $g_i(\theta) = g(X_i,\theta)$, $G_i^{(j)}(\theta) = G^{(j)}(X_i,\theta)$, $\hat{g}_i = g(X_i,\hat{\theta})$, and $\hat{G}_i^{(j)} = G^{(j)}(X_i,\hat{\theta})$ for $j = 1,...,d+1$, where $\hat{\theta}$ is EL, ET or ETEL estimator. In addition, let $g_{i0} = g_i(\theta_0)$ and $G_{i0} = G_i(\theta_0)$, where θ_0 is the (pseudo-)true value.

3.1 Empirical Likelihood and Exponential Tilting Estimators

To define EL and ET estimators, I follow the notation of Newey and Smith (2004) and Anatolyev (2005). Let $\rho(\nu)$ be a concave function in a scalar ν on the domain that contains zero. For EL, $\rho(\nu) = \log(1-\nu)$ for $\nu \in (-\infty, 1)$. For ET, $\rho(\nu) = 1 - e^{\nu}$ for $\nu \in \mathbf{R}$. In addition, let $\rho_j(\nu) = \partial^j \rho(\nu)/\partial\nu^j$ for $j = 0,1,2,\cdots$.

The EL or the ET estimator, $\hat{\theta}$, and the corresponding Lagrange multiplier, $\hat{\lambda}$, solve a saddle point problem

$$\min_{\theta\in\Theta} \max_{\lambda} n^{-1} \sum_{i=1}^{n} \rho(\lambda' g_i(\theta)). \tag{3.1}$$

The FOC's for $(\hat{\theta}, \hat{\lambda})$ are

$$\underset{L_\theta \times 1}{0} = n^{-1} \sum_{i=1}^{n} \rho_1(\hat{\lambda}'\hat{g}_i)\hat{G}_i'\hat{\lambda}, \quad \underset{L_g \times 1}{0} = n^{-1} \sum_{i=1}^{n} \rho_1(\hat{\lambda}'\hat{g}_i)\hat{g}_i. \tag{3.2}$$

A useful by-product of the estimation is the implied probabilities. The EL and the ET implied probabilities for the observations are, for $i = 1,...,n$,

$$\text{EL:} \qquad p_i = \frac{1}{n(1-\hat{\lambda}'\hat{g}_i)}, \tag{3.3}$$

$$\text{ET:} \qquad p_i = \frac{e^{\hat{\lambda}'\hat{g}_i}}{\sum_{j=1}^{n} e^{\hat{\lambda}'\hat{g}_j}}. \tag{3.4}$$

These probabilities may be used in resampling to increase efficiency under correct specification.

The FOC's hold regardless of model misspecification and form a just-identified moment condition. Let $\psi(X_i, \beta)$ be a $(L_\theta + L_g) \times 1$ vector such that

$$\psi(X_i, \beta) \equiv \begin{bmatrix} \psi_1(X_i, \beta) \\ \psi_2(X_i, \beta) \end{bmatrix} = \begin{bmatrix} \rho_1(\lambda' g_i(\theta)) G_i(\theta)'\lambda \\ \rho_1(\lambda' g_i(\theta)) g_i(\theta) \end{bmatrix}. \tag{3.5}$$

Then, the EL or the ET estimator and the corresponding Lagrange multiplier denoted by an augmented vector, $\hat{\beta} = (\hat{\theta}', \hat{\lambda}')'$, are given by the solution to $n^{-1}\sum_i^n \psi(X_i, \hat{\beta}) = 0$. In the limit, the pseudo-true value $\beta_0 = (\theta_0', \lambda_0')'$ solves the population version of the FOC's:

$$\underset{L_\theta \times 1}{0} = E\rho_1(\lambda_0' g_{i0}) G_{i0}' \lambda_0, \qquad \underset{L_g \times 1}{0} = E\rho_1(\lambda_0' g_{i0}) g_{i0}. \tag{3.6}$$

In this setting, consistency and asymptotic normality of $\hat{\beta} = (\hat{\theta}', \hat{\lambda}')'$ for $\beta_0 = (\theta_0', \lambda_0')'$ can be shown by using standard asymptotic theory of just-identified GMM, e.g. Newey and McFadden (1994).

For EL, Chen, Hong, and Shum (2007) provide regularity conditions for $\sqrt{n}$-consistency and asymptotic normality under misspecification. In particular, they assume that the moment function is uniformly bounded:

$$\textbf{UBC:} \quad \sup_{\theta \in \Theta, x \in \chi} \|g(x, \theta)\| < \infty \quad \text{and} \quad \inf_{\theta \in \Theta, \lambda \in \Lambda(\theta), x \in \chi} (1 - \lambda' g(x, \theta)) > 0, \tag{3.7}$$

where Θ and $\Lambda(\theta)$ are compact sets and χ is the support of X_1. This is a strong condition on the support of the data, e.g., Schennach (2007). Nevertheless, if the data is truncated or the moment function is constructed to satisfy UBC, then the EL estimator would be $\sqrt{n}$-consistent for the pseudo-true value and the bootstrap can be implemented. For ET, UBC is not required. The ET estimator is $\sqrt{n}$-consistent and asymptotically normal under a slightly weaker condition than Assumption 3 of Schennach (2007).[4]

Assuming regularity conditions, we have the following proposition:

Proposition 1. *Suppose regularity conditions hold. In particular, assume that UBC holds for EL. Let $\hat{\beta} = (\hat{\theta}', \hat{\lambda}')'$ be either the EL or the ET estimator and its Lagrange*

[4] We need to replace Schennach(2007)'s Assumption 3(2) with the ET saddle-point problem. In addition, we only require $k_2 = 0, 1, 2$ instead of $k_2 = 0, 1, 2, 3, 4$ in Assumption 3(6).

multiplier, and $\beta_0 = (\theta_0', \lambda_0')'$ *be the corresponding pseudo-true value. Then,*

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

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

The Jacobian matrix for EL or ET is given by

$$\frac{\partial \psi(X_i, \beta)}{\partial \beta'} = \left[\begin{array}{cc} (\partial/\partial\theta')\psi_1(X_i, \beta) & (\partial/\partial\lambda')\psi_1(X_i, \beta) \\ (\partial/\partial\theta')\psi_2(X_i, \beta) & (\partial/\partial\lambda')\psi_2(X_i, \beta) \end{array} \right], \tag{3.8}$$

where

$$\begin{aligned} \frac{\partial \psi_1(X_i, \beta)}{\partial \theta'} &= \rho_1(\lambda' g_i(\theta))(\lambda' \otimes I_{L_\theta})G_i^{(2)}(\theta) + \rho_2(\lambda' g_i(\theta))G_i(\theta)'\lambda\lambda' G_i(\theta), \quad (3.9) \\ \frac{\partial \psi_1(X_i, \beta)}{\partial \lambda'} &= \frac{\partial \psi_2(X_i, \beta)}{\partial \theta} = \rho_1(\lambda' g_i(\theta))G_i(\theta)' + \rho_2(\lambda' g_i(\theta))G_i(\theta)'\lambda g_i(\theta)', \\ \frac{\partial \psi_2(X_i, \beta)}{\partial \lambda'} &= \rho_2(\lambda' g_i(\theta))g_i(\theta)g_i(\theta)'. \end{aligned}$$

Γ and Ψ can be estimated by

$$\hat{\Gamma} = n^{-1}\sum_i^n \frac{\partial \psi(X_i, \hat{\beta})}{\partial \beta'} \quad \text{and} \quad \hat{\Psi} = n^{-1}\sum_i^n \psi(X_i, \hat{\beta})\psi(X_i, \hat{\beta})', \tag{3.10}$$

respectively. The upper left $L_\theta \times L_\theta$ submatrix of $\Gamma^{-1}\Psi(\Gamma')^{-1}$, denoted by Σ_{MR}, is the asymptotic variance matrix of $\sqrt{n}(\hat{\theta} - \theta_0)$. This coincides with the usual asymptotic variance matrix $\Sigma_C = (EG_{i0}'(Eg_{i0}g_{i0}')^{-1}EG_{i0})^{-1}$ under correct specification, but they differ in general under misspecification. Let $\hat{\Sigma}_{MR}$ be the corresponding submatrix of the variance estimator $\hat{\Gamma}^{-1}\hat{\Psi}(\hat{\Gamma}')^{-1}$. Even under correct specification, $\hat{\Sigma}_{MR}$ is different from $\hat{\Sigma}_C$, the conventional variance estimator consistent for Σ_C, because $\hat{\Sigma}_{MR}$ contains additional terms which are assumed to be zero in $\hat{\Sigma}_C$.

3.2 Exponentially Tilted Empirical Likelihood Estimator

Schennach (2007) proposes the ETEL estimator which is designed to be robust to misspecification without UBC, while it maintains the same nice higher-order properties with EL under correct specification. The ETEL estimator and the Lagrange multiplier $(\hat{\theta}, \hat{\lambda})$ solve

$$\arg\min_{\theta \in \Theta} -n^{-1} \sum_{i=1}^{n} \log n\hat{w}_i(\theta), \quad \hat{w}_i(\theta) = \frac{e^{\hat{\lambda}(\theta)' g_i(\theta)}}{\sum_{j=1}^{n} e^{\hat{\lambda}(\theta)' g_j(\theta)}}, \tag{3.11}$$

where $\hat{\lambda} \equiv \hat{\lambda}(\hat{\theta})$ and

$$\hat{\lambda}(\theta) = \arg\max_{\lambda} -n^{-1} \sum_{i=1}^{n} e^{\lambda' g_i(\theta)}. \tag{3.12}$$

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

$$\hat{l}_n(\theta) = \log \left(n^{-1} \sum_{i=1}^{n} e^{\hat{\lambda}(\theta)'(g_i(\theta) - g_n(\theta))} \right), \tag{3.13}$$

where $g_n(\theta) = n^{-1} \sum_{i=1}^{n} g_i(\theta)$. 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 $\beta = (\theta', \lambda', \kappa', \tau)'$, where $\kappa \in \mathbf{R}^{L_g}$ and $\tau \in \mathbf{R}$. By Lemma 9 of Schennach (2007), the ETEL estimator $\hat{\theta}$ is given by the subvector of $\hat{\beta} = (\hat{\theta}', \hat{\lambda}', \hat{\kappa}', \hat{\tau})'$, the solution to

$$n^{-1} \sum_{i}^{n} \psi(X_i, \hat{\beta}) = 0, \tag{3.14}$$

where

$$\psi(X_i, \beta) \equiv \begin{bmatrix} \psi_1(X_i, \beta) \\ \psi_2(X_i, \beta) \\ \psi_3(X_i, \beta) \\ \psi_4(X_i, \beta) \end{bmatrix} = \begin{bmatrix} e^{\lambda' g_i(\theta)} G_i(\theta)' \left(\kappa + \lambda g_i(\theta)'\kappa - \lambda\right) + \tau G_i(\theta)'\lambda \\ (\tau - e^{\lambda' g_i(\theta)}) \cdot g_i(\theta) + e^{\lambda' g_i(\theta)} \cdot g_i(\theta) g_i(\theta)'\kappa \\ e^{\lambda' g_i(\theta)} \cdot g_i(\theta) \\ e^{\lambda' g_i(\theta)} - \tau \end{bmatrix}. \tag{3.15}$$

Note that the estimators of the auxiliary parameters, $\hat{\kappa}$ and $\hat{\tau}$ are given by

$$\hat{\tau} = n^{-1} \sum_{i=1}^{n} e^{\hat{\lambda}' \hat{g}_i} \text{ and } \hat{\kappa} = -\left(n^{-1} \sum_{i=1}^{n} \frac{e^{\hat{\lambda}' \hat{g}_i}}{\hat{\tau}} \hat{g}_i \hat{g}_i' \right)^{-1} \hat{g}_n, \tag{3.16}$$

where $\hat{g}_n = n^{-1}\sum_i \hat{g}_i$. The probability limit of $\hat{\beta}$ is the pseudo-true value $\beta_0 = (\theta_0', \lambda_0', \kappa_0', \tau_0)'$ that solves $E\psi(X_i, \beta_0) = 0$. In particular, a function $\lambda_0(\theta)$ is the solution to $Ee^{\lambda' g_i(\theta)} g_i(\theta) = 0$, where $\lambda_0 \equiv \lambda_0(\theta_0)$ and θ_0 is a unique minimizer of the population objective function:

$$l_0(\theta) = \log\left(Ee^{\lambda_0(\theta)'(g_i(\theta) - Eg_i(\theta))}\right). \tag{3.17}$$

By Theorem 10 of Schennach,

$$\sqrt{n}(\hat{\beta} - \beta_0) \to_d N(0, \Gamma^{-1}\Psi(\Gamma')^{-1}), \tag{3.18}$$

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

Γ and Ψ are estimated by the same formula with (3.10). In order to estimate Γ, we need an exact formula of $(\partial/\partial\beta')\psi(X_i, \beta)$. The partial derivative of $\psi_1(X_i, \beta)$ is given by

$$\frac{\partial\psi_1(X_i,\beta)}{\partial\beta'} = \begin{pmatrix} \underset{L_\theta\times L_\theta}{\frac{\partial\psi_1(X_i,\beta)}{\partial\theta'}} & \underset{L_\theta\times L_g}{\frac{\partial\psi_1(X_i,\beta)}{\partial\lambda'}} & \underset{L_\theta\times L_g}{\frac{\partial\psi_1(X_i,\beta)}{\partial\kappa'}} & \underset{L_\theta\times 1}{\frac{\partial\psi_1(X_i,\beta)}{\partial\tau}} \end{pmatrix}, \tag{3.19}$$

where

$$\begin{aligned} \frac{\partial\psi_1(X_i,\beta)}{\partial\theta'} &= e^{\lambda' g_i(\theta)}\left\{G_i(\theta)'(\kappa\lambda' + \lambda\kappa' + \lambda g_i(\theta)'\kappa\lambda' - \lambda\lambda')G_i(\theta)\right. && (3.20)\\ &\quad \left.+((\kappa' + \kappa' g_i(\theta)\lambda' - \lambda')\otimes I_{L_\theta})G_i^{(2)(\theta)}\right\} + \tau(\lambda'\otimes I_{L_\theta})G_i^{(2)}(\theta), \\ \frac{\partial\psi_1(X_i,\beta)}{\partial\lambda'} &= e^{\lambda' g_i(\theta)}G_i(\theta)'\left\{(\lambda g_i(\theta)'\kappa + \kappa - \lambda)g_i(\theta)' + (g_i(\theta)'\kappa - 1)I_{L_g}\right\} && (3.21)\\ &\quad +\tau G_i(\theta)', \\ \frac{\partial\psi_1(X_i,\beta)}{\partial\kappa'} &= e^{\lambda' g_i(\theta)}G_i(\theta)'(I_{L_g} + \lambda g_i(\theta)'), && (3.22)\\ \frac{\partial\psi_1(X_i,\beta)}{\partial\tau} &= G_i(\theta)'\lambda. && (3.23) \end{aligned}$$

The partial derivative of $\psi_2(X_i, \beta)$ is given by

$$\frac{\partial\psi_2(X_i,\beta)}{\partial\beta'} = \begin{pmatrix} \underset{L_g\times L_\theta}{\frac{\partial\psi_2(X_i,\beta)}{\partial\theta'}} & \underset{L_g\times L_g}{\frac{\partial\psi_2(X_i,\beta)}{\partial\lambda'}} & \underset{L_g\times L_g}{e^{\lambda' g_i(\theta)}g_i(\theta)g_i(\theta)'} & \underset{L_g\times 1}{g_i(\theta)} \end{pmatrix}, \tag{3.24}$$

where

$$\frac{\partial\psi_2(X_i,\beta)}{\partial\theta'} = \frac{\partial\psi_1(X_i,\beta)}{\partial\lambda}, \tag{3.25}$$

$$\frac{\partial\psi_2(X_i,\beta)}{\partial\lambda'} = e^{\lambda' g_i(\theta)} g_i(\theta) g_i(\theta)'(\kappa g_i(\theta)' - I_{L_g}). \tag{3.26}$$

The partial derivative of $\psi_3(X_i,\beta)$ is given by

$$\frac{\partial\psi_3(X_i,\beta)}{\partial\beta'} = \left(\underset{L_g\times L_\theta}{\frac{\partial\psi_1(X_i,\beta)}{\partial\kappa}} \quad \underset{L_g\times L_g}{e^{\lambda' g_i(\theta)} g_i(\theta) g_i(\theta)'} \quad \underset{L_g\times L_g}{\mathbf{0}} \quad \underset{L_g\times 1}{\mathbf{0}} \right), \tag{3.27}$$

and the partial derivative of $\psi_4(X_i,\beta)$ is given by

$$\frac{\partial\psi_4(X_i,\beta)}{\partial\beta'} = \left(\underset{1\times L_\theta}{e^{\lambda' g_i(\theta)}\lambda' G_i(\theta)} \quad \underset{1\times L_g}{e^{\lambda' g_i(\theta)} g_i(\theta)'} \quad \underset{1\times L_g}{\mathbf{0}} \quad \underset{1\times 1}{-1} \right). \tag{3.28}$$

The upper left $L_\theta \times L_\theta$ submatrix of $\Gamma^{-1}\Psi(\Gamma')^{-1}$, denoted by Σ_{MR}, is the asymptotic variance matrix of $\sqrt{n}(\hat{\theta} - \theta_0)$. Let $\hat{\Sigma}_{MR}$ be the corresponding submatrix of the variance estimator $\hat{\Gamma}^{-1}\hat{\Psi}(\hat{\Gamma}')^{-1}$. Again, Σ_{MR} is different from Σ_C in general under misspecification, but they become identical under correct specification.[5]

3.3 Test statistic

Let $\hat{\theta}$ be either the EL, the ET, or the ETEL estimator and let $\hat{\Sigma}_{MR}$ be the corresponding variance 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 $\hat{\Sigma}_{MR,r}$ denote the rth diagonal element of $\hat{\Sigma}_{MR}$. The t statistic for testing the null hypothesis $H_0 : \theta_r = \theta_{0,r}$ is

$$T_{MR} = \frac{\hat{\theta}_r - \theta_{0,r}}{\sqrt{\hat{\Sigma}_{MR,r}/n}}. \tag{3.29}$$

Since the t statistic T_{MR} is studentized with the misspecification-robust variance estimator $\hat{\Sigma}_{MR,r}$, T_{MR} has an asymptotic $N(0,1)$ distribution under H_0, without assuming the correct model, H_C. This is the source of achieving asymptotic refinements without recentering regardless of misspecification. In contrast, the usual t statistic

[5]Under correct specification, the asymptotic variance matrix Σ_C is the same for EL, ET, and ETEL, which is the asymptotic variance matrix of the two-step efficient GMM.

T_C is studentized with $\hat{\Sigma}_C$, a non-robust variance estimator. Hence, it is not asymptotically pivotal if the model is misspecified. Note that the only difference between T_{MR} and T_C is the variance estimator.

Both one-sided and two-sided t tests and CI's are considered. The asymptotic one-sided t test with asymptotic significance level α of $H_0 : \theta_r \leq \theta_{0,r}$ against $H_1 : \theta_r > \theta_{0,r}$ rejects H_0 if $T_{MR} > z_\alpha$, where z_α is the $1-\alpha$ quantile of the standard normal distribution. The upper one-sided CI with asymptotic confidence level $100(1-\alpha)\%$ is $(-\infty, \hat{\theta}_r + z_\alpha\sqrt{\hat{\Sigma}_{MR,r}/n}]$. Note that this asymptotic CI is robust to misspecification because $\hat{\Sigma}_{MR}$ is used. The asymptotic two-sided t test with asymptotic significance level α of $H_0 : \theta_r = \theta_{0,r}$ against $H_1 : \theta_r \neq \theta_{0,r}$ rejects H_0 if $|T_{MR}| > z_{\alpha/2}$. The misspecification-robust two-sided asymptotic CI with asymptotic confidence level $100(1-\alpha)\%$ is $[\hat{\theta}_r \pm z_{\alpha/2}\sqrt{\hat{\Sigma}_{MR,r}/n}]$.

4 The Misspecification-Robust Bootstrap Procedure

The nonparametric iid bootstrap is implemented by resampling $X_1^*, \cdots, X_n^*$ randomly with replacement from the sample $X_1, \cdots, X_n$. Although GEL implied probabilities are useful by-products of the estimation procedure, those probabilities cannot be naively used in resampling, because the cdf estimators based on such implied probabilities would be inconsistent for the true cdf if the model is misspecified. Alternatively, the bootstrap sample can be drawn from a simple shrinkage cdf estimator that combines a GEL implied probability and the empirical probability in the form of (2.10).

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

$$n^{-1}\sum_{i}^{n}\psi(X_i^*, \hat{\beta}^*) = 0, \tag{4.1}$$

where $\psi(X_i, \beta)$ is given by (3.5) for EL or ET, and (3.15) for ETEL. The bootstrap version of the variance matrix estimator is $\hat{\Gamma}^{*-1}\hat{\Psi}^*(\hat{\Gamma}^{*\prime})^{-1}$, which can be calculated using the same formula with (3.10) using the bootstrap sample instead of the original sample. Let $\hat{\Sigma}_{MR}^*$ be the upper left $L_\theta \times L_\theta$ submatrix of the bootstrap covariance estimator $\hat{\Gamma}^{*-1}\hat{\Psi}^*(\hat{\Gamma}^{*\prime})^{-1}$. It should be emphasized that the only difference between the bootstrap and the sample versions of the estimators is that the former are calculated from the bootstrap sample, χ_n^*, in place of the original sample, χ_n, because we need no additional correction such as recentering as in Hall and Horowitz (1996) and Andrews (2002).

The misspecification-robust bootstrap t statistic is

$$T_{MR}^* = \frac{\hat{\theta}_r^* - \hat{\theta}_r}{\sqrt{\hat{\Sigma}_{MR,r}^*/n}}. \tag{4.2}$$

Let $z_{T,\alpha}^*$ and $z_{|T|,\alpha}^*$ denote the $1-\alpha$ quantile of T_{MR}^* and $|T_{MR}^*|$, respectively. Following Andrews (2002), we define $z_{|T|,\alpha}^*$ to be a value that minimizes $|P^*(|T_{MR}^*| \leq z) - (1-\alpha)|$ over $z \in \mathbf{R}$, because the distribution of $|T_{MR}^*|$ is discrete. The definition of $z_{T,\alpha}^*$ is analogous. Each of the following bootstrap tests are of asymptotic significance level α. The one-sided bootstrap t test of $H_0 : \theta_r \leq \theta_{0,r}$ against $H_1 : \theta_r > \theta_{0,r}$ rejects H_0 if $T_{MR} > z_{T,\alpha}^*$. 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 $|T_{MR}| > z_{|T|,\alpha}^*$. The equal-tailed two-sided bootstrap t test of the same hypotheses rejects if $T_{MR} < z_{T,1-\alpha/2}^*$ or $T_{MR} > z_{T,\alpha/2}^*$. Similarly, each of the following bootstrap CI's for $\theta_{0,r}$ are of asymptotic confidence level $100(1-\alpha)\%$. The upper one-sided bootstrap CI is $(-\infty, \hat{\theta}_r + z_{T,\alpha}^* \sqrt{\hat{\Sigma}_{MR,r}/n}]$. The symmetric and the equal-tailed bootstrap percentile-t intervals are $[\hat{\theta}_r \pm z_{|T|,\alpha}^* \sqrt{\hat{\Sigma}_{MR,r}/n}]$ and $[\hat{\theta}_r - z_{T,\alpha/2}^* \sqrt{\hat{\Sigma}_{MR,r}/n}, \hat{\theta}_r - z_{T,1-\alpha/2}^* \sqrt{\hat{\Sigma}_{MR,r}/n}]$, respectively.

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

1. Draw n random observations χ_n^* with replacement from the original sample, χ_n.

2. From the bootstrap sample χ_n^*, calculate $\hat{\theta}^*$ and $\hat{\Sigma}_{MR}^*$ using the same formula with their sample counterpart.

3. Construct and save T_{MR}^*.

4. Repeat steps 1-3 B times and get the distribution of T_{MR}^*, which is discrete.

5. Find $z_{|T|,\alpha}^*$ and $z_{T,\alpha}^*$ from the distribution of $|T_{MR}^*|$ and T_{MR}^*, respectively.

5 Main Result

Let $f(X_i, \beta)$ be a vector containing the unique components of $\psi(X_i, \beta)$ and its derivatives with respect to the components of β through order d, and $\psi(X_i, \beta)\psi(X_i, \beta)'$ and its derivatives with respect to the components of β through order $d-1$.

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

Assumption 2.

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

(b) *$(\hat{\theta}, \hat{\lambda})$ solves (3.1) for EL or ET, or (3.11) for ETEL; (θ_0, λ_0) is the pseudo-true value that uniquely solves the population version of (3.1) for EL or ET, or (3.11) for ETEL.*

(c) *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 $0 < q_g < \infty$.*

(d) *For some function $C_\rho(x)$, $|\rho(\lambda_1' g(x, \theta_1)) - \rho(\lambda_2' g(x, \theta_2))| < C_\rho(x)\|(\theta_1', \lambda_1') - (\theta_2', \lambda_2')\|$ for all x in the support of X_1 and all $(\theta_1', \lambda_1'), (\theta_2', \lambda_2') \in \Theta \times \Lambda(\theta)$; $EC_\rho^{q_1}(X_1) < \infty$ for some $q_1 > 4$. In addition, UBC (3.7) holds for EL.*

Assumption 3.

(a) *Γ is nonsingular and Ψ is positive definite.*

(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 $d \geq 4$.*

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

(d) *There is a function $C_{\partial\rho}(x)$ such that*

$$|\rho_j(\lambda' g(x, \theta)) - \rho_j(\lambda_0' g(x, \theta_0))| \leq C_{\partial\rho}(x)\|(\theta', \lambda') - (\theta_0', \lambda_0')\|$$

for all x in the support of X_1 and all $(\theta', \lambda') \in N(\theta_0) \times \Lambda(\theta)$ for $j = 1, ..., d + 1$; $EC_{\partial\rho}^{q_2}(X_1) < \infty$ for some $q_2 > 16$.

(e) *$f(X_1, \beta_0)$ is once differentiable with respect to X_1 with uniformly continuous first derivative.*

Assumption 4. *For $t \in \mathbf{R}^{dim(f)}$, $\limsup_{\|t\|\to\infty} \left|Ee^{it'f(X_1,\beta_0)}\right| < 1$, where $i = \sqrt{-1}$.*

Assumption 1 is that the sample is iid, which is also assumed in Schennach (2007) and Newey and Smith (2004). Assumption 2(a)-(c) are similar to Assumption 2(a)-(b) of Andrews (2002). Assumption 2(d) is similar to but slightly stronger than Assumption 3(4) of Schennach (2007) for ET or ETEL, and it includes Assumption 3(1) of Chen, Hong, and Shum (2007) for EL to avoid a negative implied probability under misspecification. Assumption 2(c)-(d) are required to have the uniform convergence of the objective function. Assumption 3(a) is a standard regularity condition for a well-defined asymptotic covariance matrix. Assumption 3 except for (d) is similar to Assumption 3 of Andrews (2002). The assumptions on q_g and q_G are slightly stronger than necessary, but yield a simpler result. This is also assumed in Andrews (2002) for the same reason. Assumption 3(d) is similar to but stronger than Assumption 3(6) of Schennach (2007). It ensures that the components of higher-order Taylor expansion of the FOC have well-defined probability limits. Assumption 4 is the standard Cramér condition for Edgeworth expansion.

Throughout the proof, I pay a particular attention to the values of q_1 and q_2 that may restrict DGP's under misspecification for ET and ETEL. For example, since a zero vector is in $\Lambda(\theta)$, Assumption 3(d) implies $Ee^{q_2\lambda_0' g(X_1,\theta_0)} < \infty$, where $\lambda_0 \neq 0$ under misspecification. Lee (2014b) provides a simple example that the model cannot be misspecified too much to have $Ee^{q_2\lambda_0' g(X_1,\theta_0)} < \infty$ for some q_2 for ET and ETEL, and the set of possible misspecification shrinks to zero as q_2 gets larger.

Theorem 1 formally establishes asymptotic refinements of the bootstrap t tests and CI's based on EL, ET, and ETEL estimators. This result is new, because asymptotic refinements of the bootstrap for this class of estimators have not been established in the existing literature even under correct model specifications.

Theorem 1. *(a) Suppose Assumptions 1-4 hold with $q_1 > 4$, $q_2 > 16$, and $d = 4$. Under $H_0 : \theta_r = \theta_{0,r}$, for all $\xi \in [0, 1/2)$,*

$$P(T_{MR} > z^*_{T,\alpha}) = \alpha + o(n^{-(1/2+\xi)}) \text{ and}$$
$$P(T_{MR} < z^*_{T,\alpha/2} \text{ or } T_{MR} > z^*_{T,1-\alpha/2}) = \alpha + o(n^{-(1/2+\xi)}).$$

(b) Suppose Assumptions 1-4 hold with $q_1 > 6$*,* $q_2 > 30$*, and* $d = 5$*. Under* $H_0 : \theta_r = \theta_{0,r}$*, for all* $\xi \in [0, 1/2)$*,*

$$P(|T_{MR}| > z^*_{|T|,\alpha}) = \alpha + o(n^{-(1+\xi)}).$$

(c) Suppose Assumptions 1-4 hold with $q_1 > 8$*,* $q_2 > 48$*, and* $d = 6$*. Under* $H_0 : \theta_r = \theta_{0,r}$*,*

$$P(|T_{MR}| > z^*_{|T|,\alpha}) = \alpha + O(n^{-2}).$$

By the duality of t tests and CI's, asymptotic refinements of the same rate for the bootstrap CI's follow from Theorem 1. The equal-tailed percentile-t CI corresponds to Theorem 1(a). The symmetric percentile-t CI corresponds to Theorem 1(b)-(c). Recall that the corresponding asymptotic t test and CI based on T_{MR} are correct up to $O(n^{-1/2})$, $O(n^{-1})$, and $O(n^{-1})$ for (a), (b), and (c), respectively. The reason that the two-sided t tests and the symmetric CI achieve a higher rate of refinements is due to a symmetry property of Hall (1992).

The proof of Theorem 1 is similar to that of Andrews (2002) that establishes asymptotic refinements of the bootstrap for GMM estimators under correct specification. Since I consider GEL estimators rather than GMM, and allow misspecification rather than assuming correct specification, the detailed proof is slightly different from that of Andrews but the fundamental idea is the same. I use the fact that the FOC of GEL estimators can be written as a just-identified system of moment function regardless of misspecification. Writing an overidentified model as a just-identified system by augmenting additional parameters also appears in Imbens (1997, 2002), Chamberlain and Imbens (2003), and Schennach (2007). Then, consistency and asymptotic normality of the estimator follow by standard arguments using Newey and McFadden (1994). I show that the misspecification-robust t statistic is well approximated by a smooth function of sample averages of the data by taking Taylor expansion of the FOC, and prove asymptotic refinements by using Hall (1988,1992)'s argument on Edgeworth expansion of a smooth function of sample averages.

6 Monte Carlo Results

This section compares the finite sample coverage probabilities of CI's for a scalar parameter of interest, under correct specification and misspecification. To reduce computational burden of calculating GEL estimators B times for each Monte Carlo repetition,[6] the warp-speed Monte Carlo method of Giacomini, Politis, and White (2013) is used. The method also appears in White (2000) and Davidson and MacKinnon (2002, 2007), but the validity of the method is formally established in Giacomini, Politis, and White (2013). The key difference between the warp-speed method and a usual Monte Carlo experiment is that the bootstrap sample is drawn only once for each Monte Carlo repetition rather than B times, and thus computation time is significantly reduced.

I consider the AR(1) dynamic panel model of Blundell and Bond (1998). For $i = 1, ..., n$ and $t = 1, ...T$,

$$y_{it} \quad = \quad \rho_0 y_{i,t-1} + \eta_i + \nu_{it}, \tag{6.1}$$

where η_i is an unobserved individual-specific effect and ν_{it} is an error term. To estimate ρ_0, we use two sets of moment conditions:

$$Ey_{i(t-s)}(\Delta y_{it} - \rho_0 \Delta y_{i(t-1)}) = 0, \qquad t = 3, ...T, \text{ and } s \geq 2, \tag{6.2}$$

$$E\Delta y_{i(t-1)}(y_{it} - \rho_0 y_{i(t-1)}) = 0, \qquad t = 3, ...T. \tag{6.3}$$

The first set (6.2) is derived from taking differences of (6.1), and uses the lagged values of y_{it} as instruments. The second set (6.3) is derived from the initial conditions on DGP and mitigates weak instruments problem from using only the lagged values. Blundell and Bond (1998) suggest to use the system-GMM estimator based on the two sets of moment conditions. The number of moment conditions is $(T+1)(T-2)/2$.

Four DGP's are considered: two correctly specified models and two misspecified models. For each of the DGP's, $T = 4, 6$ and $n = 100, 200$ are considered. To minimize the effect of the initial condition, I generate 100+T time periods and use the last T periods for estimation. In Tables 1-4, "Boot" and "Asymp" mean the bootstrap CI and the asymptotic CI, respectively. The third column shows which estimator the CI is based on. GMM denotes the two-step GMM based on the system moment conditions. The fourth column shows which standard error (or variance estimator) is used: "C" denotes the usual standard error and "MR" denotes the misspecification-robust one. The fifth column shows how the bootstrap is implemented for the bootstrap

[6]For example, if $B = 1,000$ and the number of Monte Carlo repetition is $r = 1,000$, then one simulation round involves $1,000,000$ nonlinear optimizations.

CI's: "L" denotes the misspecification-robust bootstrap proposed in this paper and in Lee (2014), "HH" denotes the recentering method of Hall and Horowitz (1996), and "BNS" denotes the efficient bootstrapping of Brown and Newey (2002) with a shrinkage estimator. The shrinkage is given by (2.10) with $\epsilon_n = n^{-1/4}$. The columns under "CI" show the coverage probabilities. The column under "J test" shows the rejection probability of the overidentification test: the Hall-Horowitz bootstrap J test, the asymptotic J test, the EL likelihood-ratio (LR) test, the ET LR test, and the ETEL LR test results are presented.

In sum, eight bootstrap CI's and eight asymptotic CI's are compared. GMM-C-HH serves as a benchmark, as its properties have been relatively well investigated. GMM-MR-L is suggested by Lee (2014). Both EL-MR-L and EL-MR-BNS are suggested in this paper, while they differ in resampling methods. CI's based on ET and ETEL are defined similarly. Note that CI's using the usual standard error (C) are not robust to misspecification.

The DGP for a correctly specified model is the same as that of Bond and Windmeijer (2005). For $i = 1, ..., n$ and $t = 1, ...T$,

$$\text{DGP C-1:} \qquad \begin{aligned} y_{it} &= \rho_0 y_{i,t-1} + \eta_i + \nu_{it}, \\ \eta_i &\sim N(0,1); \nu_{it} \sim \frac{\chi_1^2 - 1}{\sqrt{2}}, \\ y_{i1} &= \frac{\eta_i}{1-\rho_0} + u_{i1}; u_{i1} \sim N\left(0, \frac{1}{1-\rho_0^2}\right). \end{aligned}$$

Since the bootstrap does not solve weak instruments (Hall and Horowitz, 1996), I let $\rho_0 = 0.4$ so that the performance of the bootstrap is not affected by the problem. The simulation result is given in Table 1. First of all, the bootstrap CI's show significant improvement over the asymptotic CI's across all the cases considered. Second, similar to the result of Bond and Windmeijer (2005), the bootstrap CI's coverage probabilities tend to be too high for $T = 6$. This over-coverage problem becomes less severe as the sample size increases to $n = 200$, especially for those based on EL, ET, and ETEL. Interestingly, resampling from the shrinkage estimator (BNS) seems to mitigate this problem. Third, the asymptotic CI's using the robust standard error (MR) work better than the ones using the usual standard error (C). This result is surprising given that the model is correctly specified. One reason is that both standard errors underestimate the standard deviation of the estimator while the robust standard error is relatively large in this case. For example, when $T = 6$ and $n = 100$, the difference in the coverage probabilities between Asymp-ET-C and Asymp-ET-MR is quite large. The unreported standard deviation of the ET estimator is 0.0819, while the mean of robust and usual standard errors are 0.0592 and 0.0472, respectively. Finally, the overidentification tests based on GEL estimators or the HH bootstrap show significant size distortion, especially when $T = 6$.

Next a heteroskedastic error term across individuals is considered. The DGP is

$$\text{DGP C-2:} \quad \begin{aligned} & y_{it} = \rho_0 y_{i,t-1} + \eta_i + \nu_{it}, \\ & \eta_i \sim N(0,1); \nu_{it} \sim N(0,\sigma_i^2); \sigma_i^2 \sim U[0.2.1.8], \\ & y_{i1} = \frac{\eta_i}{1-\rho_0} + u_{i1}; u_{i1} \sim N\left(0, \frac{\sigma_i^2}{1-\rho_0^2}\right). \end{aligned}$$

The result is given in Table 2. The findings are similar to that of Table 1, except that the over-coverage problem of the bootstrap CI's based on GEL estimators improves quickly as the sample size grows.

To allow misspecification, consider the case that the DGP follows an AR(2) process while the model is still based on the AR(1) specification, (6.1). For $i = 1, ..., n$ and $t = 1, ...T$,

$$\text{DGP M-1:} \quad \begin{aligned} & y_{it} = \rho_1 y_{i,t-1} + \rho_2 y_{i,t-2} + \eta_i + \nu_{it}, \\ & \eta_i \sim N(0,1); \nu_{it} \sim \frac{\chi_1^2 - 1}{\sqrt{2}}, \\ & y_{i1} = \frac{\eta_i}{1-\rho_1-\rho_2} + u_{i1}; u_{i1} \sim N\left(0, \frac{1-\rho_2}{(1+\rho_2)[(1-\rho_2)^2-\rho_1^2]}\right). \end{aligned}$$

Since the EL estimator is not $\sqrt{n}-$consistent under misspecification unless the UBC (3.7) is satisfied, I also consider DGP M-2 which is identical to DGP M-1 except that η_i, u_{i1}^0, and ν_{it} are generated from a truncated standard normal distributed between -3 and 3, where $u_{i1} = \sqrt{\frac{1-\rho_2}{(1+\rho_2)[(1-\rho_2)^2-\rho_1^2]}} u_{i1}^0$.

If the model is misspecified, then there is no true parameter that satisfies the moment conditions simultaneously. It is important to understand what is identified and estimated under misspecification. The moment conditions (6.2) and (6.3) impose

$$\frac{Ey_{i1}\Delta y_{it}}{Ey_{i1}\Delta y_{i(t-1)}} = \cdots = \frac{Ey_{i(t-3)}\Delta y_{it}}{Ey_{i(t-3)}\Delta y_{i(t-1)}} = \frac{Ey_{i(t-2)}\Delta y_{it}}{Ey_{i(t-2)}\Delta y_{i(t-1)}} = \frac{E\Delta y_{i(t-1)} y_{it}}{E\Delta y_{i(t-1)} y_{i(t-1)}}, \quad (6.4)$$

for $t = 3, ..., T$. Under correct specification, the restriction (6.4) holds and a unique parameter is identified. However, each of the ratios identifies different parameters under misspecification, and the probability limits of GMM and GEL estimators are weighted averages of the parameters. For example, when $T = 4$, we have five moment conditions. Four of them identify $\rho_{T4}^a \equiv \rho_1 - \rho_2$ and the other identify $\rho_{T4}^b \equiv \rho_1 + \frac{\rho_2}{\rho_1 - \rho_2}$. When $T = 6$, we have fourteen moment conditions. Eight of them identify ρ_{T4}^a, three identify ρ_{T4}^b, two identify

$$\rho_{T6}^a \equiv \frac{(\rho_1^2 + \rho_2)(\rho_1 - \rho_2) + \rho_1\rho_2}{\rho_1(\rho_1 - \rho_2) + \rho_2}, \tag{6.5}$$

and the other identifies

$$\rho_{T6}^b \equiv \frac{(\rho_1^3 + 2\rho_1\rho_2)(\rho_1 - \rho_2) + \rho_2(\rho_1^2 + \rho_2)}{(\rho_1^2 + \rho_2)(\rho_1 - \rho_2) + \rho_1\rho_2}. \tag{6.6}$$

Thus, the pseudo-true value ρ_0 is defined as

$$T = 4: \quad \rho_0 = w_1\rho_{T4}^a + (1 - w_1)\rho_{T4}^b, \tag{6.7}$$

$$T = 6: \quad \rho_0 = c_1\rho_{T4}^a + c_2\rho_{T4}^b + c_3\rho_{T6}^a + (1 - c_1 - c_2 - c_3)\rho_{T6}^b, \tag{6.8}$$

where w_1 and c_1, c_2, c_3 are some weights between 0 and 1. The pseudo-true values are different for $T = 4$ and $T = 6$. Moreover, GMM and GEL pseudo-true values would be different because their weights are different. Observe that if $\rho_2 = 0$, then the pseudo-true values coincide with ρ_1, the AR(1) coefficient. Thus, the pseudo-true values capture the deviation from the AR(1) model. If $|\rho_2|$ is relatively small, then the pseudo-true value would not be much different from ρ_1, while there is an advantage of using a parsimonious model. If one accepts the possibility of misspecification and decides to proceed with the pseudo-true value, then GEL pseudo-true values have better interpretation than GMM ones because GEL weights are implicitly calculated according to a well-defined distance measure while GMM weights depend on the choice of a weight matrix by a researcher.

Tables 3-4 show the coverage probabilities of CI's under DGP M-1 and M-2, respectively. I set $\rho_1 = 0.6$ and $\rho_2 = 0.2$. The pseudo-true values are calculated using the sample size of $n = 30,000$ for $T = 4$ and $n = 20,000$ for $T = 6$.[7] It is clearly seen

[7]The two-step GMM and GEL pseudo-values are not that different. They are around 0.4 when

that the bootstrap CI's outperform the asymptotic CI's. In particular, the performances of Boot-EL-MR-L, Boot-ET-MR-L, and Boot-ETEL-MR-L CI's are excellent for $T = 4$. When $T = 6$, these CI's exhibit over-coverage but the problem is less severe than Boot-GMM-MR-L. In addition, the bootstrap CI's using the shrinkage in resampling are found to improve on the over-coverage problem. Although DGP M-1 does not satisfy the UBC (3.7), the performance of the CI based on EL does not seem to be affected. One may wonder why the HH bootstrap CI works quite well under misspecification even though the CI is not robust to misspecification. This is spurious and cannot be generalized. In this case, the usual standard error $\sqrt{\hat{\Sigma}_C/n}$ is considerably smaller than the robust standard error $\sqrt{\hat{\Sigma}_{MR}/n}$, while the HH bootstrap critical value is much larger than the asymptotic one, which offsets the smaller standard error. Lee (2014) reports that the performance of the HH bootstrap CI under misspecification is much worse than that of the MR bootstrap CI. In addition, the HH bootstrap J test shows very low power relative to the asymptotic tests. Among the asymptotic CI's, those based on GEL estimators and the robust standard errors show better performances.

Finally, Table 5 compares the width of the bootstrap CI's under different DGP's. Since this paper establishes asymptotic refinements in the size and coverage errors of the MR bootstrap t tests and CI's based on GEL estimators, the width of CI's is not directly related to the main result. Nevertheless, the table clearly demonstrates a reason to consider GEL as an alternative to GMM, especially when misspecification is suspected. Under correct specification (C-1 and C-2), all the bootstrap CI's have similar width. This conclusion changes dramatically under misspecification (M-1 and M-2). The CI's based on GMM are much wider than those based on GEL. For example, when $T = 6$ and $n = 200$ in DGP M-2, the width of the Boot-GMM-MR-L 95% CI is 1.004, while that of Boot-EL-MR-BNS 95% CI is 0.277, almost a fourth. The main reason for this is that the GEL standard errors are smaller than the GMM ones under misspecification. In addition, the bootstrap CI's using the shrinkage in resampling are generally narrower than the nonparametric iid bootstrap CI's.

The findings of Monte Carlo experiments can be summarized as follows. First, the misspecification-robust bootstrap CI's based on GEL estimators are generally more accurate than other bootstrap and asymptotic CI's regardless of misspecification. Not

$T = 4$ and around 0.5 when $T = 6$.

surprisingly, the coverage of non-robust CI's are very poor under misspecification. Second, the GEL-based bootstrap CI's improve on the severe over-coverage of the GMM-based bootstrap CI's, which is also a concern of Bond and Windmeijer (2005). In addition, the GEL-based bootstrap CI's using the shrinkage in resampling (BNS) can mitigate over-coverage of the bootstrap CI's when T is relatively large.[8] Lastly, it is recommended to use the misspecification-robust variance estimator in constructing t statistics and CI's regardless of whether the model is correctly specified or not, because the coverage of the misspecification-robust CI's tends to be more accurate even under correct specification.

7 Application: Returns to Schooling

Hellerstein and Imbens (1999) estimate the Mincer equation by weighted least squares, where the weights are calculated using EL. The equation of interest is

$$\begin{aligned} \log(\text{wage}_i) &= \beta_0 + \beta_1 \cdot \text{education}_i + \beta_2 \cdot \text{experience}_i + \beta_3 \cdot \text{experience}_i^2 \\ &\quad + \beta_4 \cdot \text{IQ}_i + \beta_5 \cdot \text{KWW}_i + \varepsilon_i, \end{aligned} \tag{7.1}$$

where KWW denotes Knowledge of the World of Work, an ability test score. Since the National Longitudinal Survey Young Men's Cohort (NLS) dataset reports both ability test scores and schooling, the equation (7.1) can be estimated by OLS. However, the NLS sample size is relatively small, and it may not correctly represent the whole population. In contrast, the Census data is a very large dataset which is considered as the whole population, but we cannot directly estimate the equation (7.1) using the Census because it does not contain ability measures. Hellerstein and Imbens calculate weights by matching the Census and the NLS moments and use the weights to estimate the equation (7.1) by the least squares. This method can be used to reduce the standard errors or change the estimand toward more representative of the Census.

Let $y_i \equiv \log(\text{wage}_i)$ and $\mathbf{x}_i$ be the regressors on the right-hand-side of (7.1). The Hellerstein-Imbens weighted least squares can be viewed as a special case of the EL

[8] The choice of $n^{-1/4}$ rate in the shrinkage estimator is arbitrary, but no guidance of how to choose the rate is provided. Nevertheless, the shrinkage estimator improves on the over-coverage of the bootstrap CI's and make the CI's narrower. This topic deserves more research.

estimator using the following moment condition:

$$E_s g_i(\beta_0) = 0, \tag{7.2}$$

where $E_s[\cdot]$ is the expectation over a probability density function $f_s(y_i, \mathbf{x}_i)$, which is labeled the *sampled population*. The moment function $g_i(\beta)$ is

$$g_i(\beta) = \begin{pmatrix} \mathbf{x}_i(y_i - \mathbf{x}_i'\beta) \\ m(y_i, \mathbf{x}_i) - E_t m(y_i, \mathbf{x}_i) \end{pmatrix}, \tag{7.3}$$

where β is a parameter vector, $m(y_i, \mathbf{x}_i)$ is a 13×1 vector, and $E_t[\cdot]$ is the expectation over a probability density function $f_t(y_i, \mathbf{x}_i)$, labeled the *target population*. The first set of the moment condition is the FOC of OLS and the second set matches the sample (NLS) moments with the known population (Census) moments. In particular, the thirteen moments consisting of first, second, and cross moments of log(wage), education, experience, and experience squared are matched. If the sampled population is identical to the target population, i.e., the NLS sample is randomly drawn from the Census distribution, the moment condition model is correctly specified and (7.2) holds. Otherwise, the model is misspecified and there is no such β that satisfies (7.2). In this case, the probability limit of the EL estimator solves the FOC of OLS with respect to an artificial population that minimizes a distance between the sampled and the target populations. This pseudo-true value is an interesting estimand because we are ultimately interested in the parameters of the target population, rather than the sampled population.

Table 6 shows the estimation result of OLS, two-step GMM, EL, ET, and ETEL estimators. Without the Census moments, the equation (7.1) is estimated by OLS and the estimate of the returns to schooling is 0.054 with the standard error of 0.010. By using the Census moments, the coefficients estimates and the standard errors change. The two-step GMM estimator is calculated using the OLS estimator as a preliminary estimator, and it serves as a benchmark. EL, ET, and ETEL produce higher point estimates and smaller standard errors than those of OLS. Since the J-test rejects the null hypothesis of correct specification for all of the estimators using the Census moments, it is likely that the target population differs from the sampled population. If this is the case, then the conventional standard errors are no longer valid, and the misspecification-robust standard errors should be used. The misspecification-

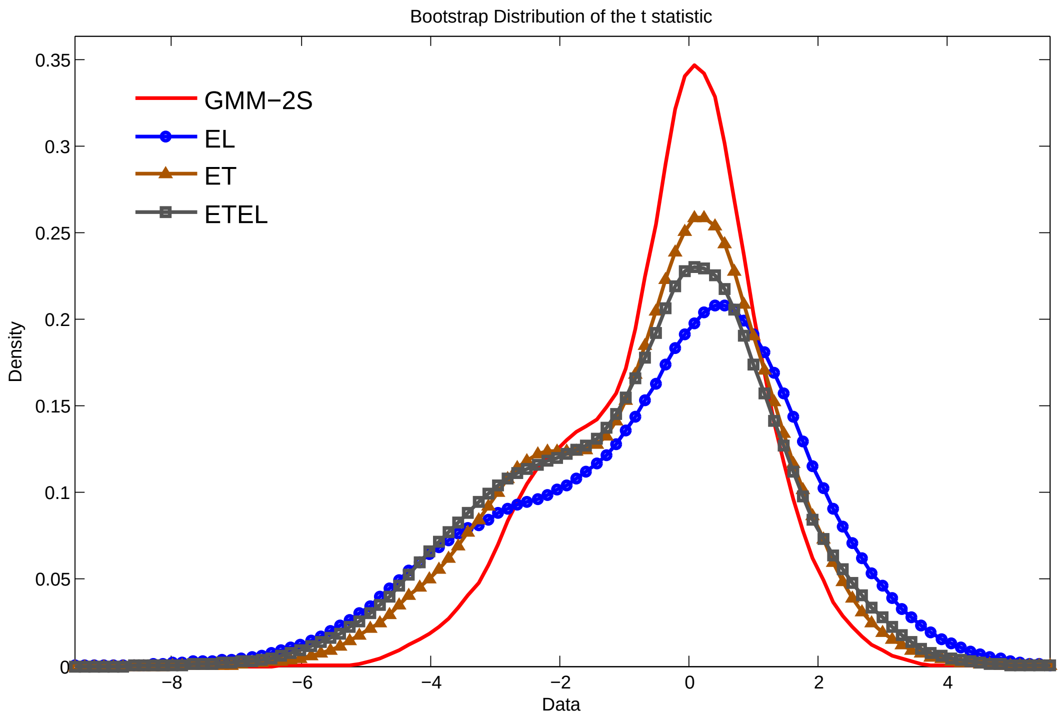

Figure 1: Bootstrap distribution of the t statistics based on 2-step GMM estimator (solid), EL estimator (solid with circle), ET estimator (solid with triangle), and ETEL estimator (solid with rectangle).

robust standard errors, s.e.$_{MR}$, of EL, ET, and ETEL are slightly larger than the usual standard errors assuming correct specification, s.e.$_C$, but still smaller than the standard errors of OLS. In contrast, s.e.$_{MR}$ of GMM is much larger than s.e.$_C$, which is consistent with the simulation result given in Section 6.

Table 7 shows the lower and upper bounds of CI's based on various estimators and their respective width. The width of the GMM based CI's are relatively wide compared to those based on GEL estimators. Among the GEL estimators, the ET estimator has the widest CI, while the EL estimator has the narrowest. Although the bootstrap CI's are generally wider than the asymptotic CI's, using the shrinkage in resampling reduces the width significantly. The upper bounds of the bootstrap CI's range from 8.3% to 11%, which are higher than those of the asymptotic CI's. I also present a nonparametric kernel estimate of the bootstrap distribution of the t statistics based on GMM, EL, ET, and ETEL estimators in Figure 1. The distributions are skewed to the left, which implies the presence of a downward bias. Overall, the estimation of (7.1) using GEL estimators and the resulting bootstrap CI's suggest that the returns to schooling is likely to be higher than originally estimated by Hellerstein and Imbens.

8 Conclusion

GEL estimators are favorable alternatives to GMM. Although asymptotic refinements of the bootstrap for GMM have been established, the same for GEL have not been done yet. In addition, the current literature on bootstrapping does not consider model misspecification that adversely affects the refinement and validity of the bootstrap. This paper formally established asymptotic refinements of the bootstrap for t tests and CI's based on GEL estimators. Moreover, the proposed bootstrap is robust to misspecification, which means the asymptotic refinements of the bootstrap is not affected by unknown model misspecification. Simulation results did support this finding. As an application, the returns to education was estimated by extending the method of Hellerstein and Imbens (1999). The exercise found that the estimates of Hellerstein and Imbens were robust across different GEL estimators, and the returns to education could be even higher.

References

Allen, J., Gregory, A. W., and Shimotsu, K. (2011). Empirical likelihood block bootstrapping. Journal of Econometrics, 161(2), 110-121.

Almeida, C., and Garcia, R. (2012). Assessing misspecified asset pricing models with empirical likelihood estimators. Journal of Econometrics, 170(2), 519-537.

Altonji, J. G., and Segal, L. M. (1996). Small-sample bias in GMM estimation of covariance structures. Journal of Business and Economic Statistics, 14(3), 353-366.

Anatolyev, S. (2005). GMM, GEL, serial correlation, and asymptotic bias. Econometrica, 73(3), 983-1002.

Andrews, D. W. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. Econometrica, 70(1), 119-162.

Antoine, B., Bonnal, H., and Renault, E. (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood. Journal of Econometrics, 138(2), 461-487.

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. Journal of the American Statistical Association, 83(403), 687-697.

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

Blundell, R., and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, 87(1), 115-143.

Bond, S., and Windmeijer, F. (2005). Reliable inference for GMM estimators? Finite sample properties of alternative test procedures in linear panel data models. Econometric Reviews, 24(1), 1-37.

Bravo, F. (2010). Efficient M-estimators with auxiliary information. Journal of Statistical Planning and Inference, 140(11), 3326-3342.

Brown, B. W., and Newey, W. K. (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.

Chamberlain, G., and Imbens, G. W. (2003). Nonparametric applications of Bayesian inference. Journal of Business and Economic Statistics, 21(1), 12-18.

Chen, X., Hong, H., and Shum, M. (2007). Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models. Journal of Econometrics, 141(1), 109-140.

Davidson, R., and MacKinnon, J. G. (2002). Fast double bootstrap tests of nonnested linear regression models. Econometric Reviews, 21(4), 419-429.

Davidson, R., and MacKinnon, J. G. (2007). Improving the reliability of bootstrap tests with the fast double bootstrap. Computational Statistics & Data Analysis, 51(7), 3259-3281.

Giacomini, R., Politis, D. N., and White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. Econometric Theory, 1-23.

Gonçalves, S., and White, H. (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. Journal of Econometrics, 119(1), 199-219.

Guggenberger, P. (2008). Finite sample evidence suggesting a heavy tail problem of the generalized empirical likelihood estimator. Econometric Reviews, 27(4-6), 526-541.

Guggenberger, P., and Hahn, J. (2005). Finite Sample Properties of the Two-Step Empirical Likelihood Estimator. Econometric Reviews, 24(3), 247-263.

Hahn, J. (1996). A note on bootstrapping generalized method of moments estimators. Econometric Theory, 12, 187-197.

Hall, P. (1988). On symmetric bootstrap confidence intervals. Journal of the Royal Statistical Society. Series B (Methodological), 35-45.

Hall, P. (1992). The bootstrap and Edgeworth expansion. Springer.

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

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica, 50(4), 1029-1054.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics, 14(3), 262-280.

Hansen, L. P., and Jagannathan, R. (1997). Assessing specification errors in stochastic discount factor models. The Journal of Finance, 52(2), 557-590.

Hellerstein, J. K., and Imbens, G. W. (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.

Imbens, G. W. (1997). One-step estimators for over-identified generalized method of moments models. The Review of Economic Studies, 64(3), 359-383.

Imbens, G. W. (2002). Generalized method of moments and empirical likelihood. Journal of Business and Economic Statistics, 20(4).

Imbens, G. W., Spady, R. H. and Johnson, P. (1998). Information theoretic approaches to inference in moment condition models. Econometrica, 66(2), 333-357.

Inoue, A., and Shintani, M. (2006). Bootstrapping GMM estimators for time series. Journal of Econometrics, 133(2), 531-555.

Kitamura, Y., and Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation. Econometrica, 65(4), 861-874.

Kitamura, Y., Otsu, T., and Evdokimov, K. (2013). Robustness, infinitesimal neighborhoods, and moment restrictions. Econometrica, 81(3), 1185-1201.

Kundhi, G., and Rilstone, P. (2012). Edgeworth expansions for GEL estimators. Journal of Multivariate Analysis, 106, 118-146.

Lee, S. (2014). Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators. Journal of Econometrics, 178(3), 398-413.

Lee, S. (2014b). On robustness of GEL estimators to model misspecification. Working paper. Australian School of Business, UNSW.

Newey, W. K., and McFadden, D. (1994). Large sample estimation and hypothesis testing. Handbook of Econometrics, 4, 2111-2245.

Newey, W. K., and Smith, R. J. (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.

Owen, A. (1990). Empirical likelihood ratio confidence regions. The Annals of Statistics, 18(1), 90-120.

Qin, J., and Lawless, J. (1994). Empirical likelihood and general estimating equations. The Annals of Statistics, 300-325.

Schennach, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. The Annals of Statistics, 35(2), 634-672.

White, H. (1982). Maximum likelihood estimation of misspecified models. Econometrica, 50(1), 1-25.

White, H. (2000). A reality check for data snooping. Econometrica, 68(5), 1097-1126.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. Journal of Econometrics, 126(1), 25-51.

A Appendix: Lemmas and Proofs

A.1 Proof of Proposition 1

proof. The proof is similar to that of Theorem 10 of Schennach (2007), and thus omitted.

A.2 Lemmas

The lemmas and the proofs are analogous to those of Hall and Horowitz (1996) and Andrews (2002) that show asymptotic refinements of the bootstrap for GMM estimators under correct specification. I also use some proof techniques of Schennach (2007) for GEL estimators. For brevity, Hall and Horowitz (1996) is abbreviated to HH, Andrews (2002) to A2002, and Schennach (2007) to S2007. In the lemmas, a constant a that determines the rate of convergence in probability appears. To show the theorem, we only need $a = 1, 1.5$ and 2, but I assume that $a \geq 0$ throughout the lemmas for generality.

Lemma 1 modifies Lemmas 1, 2, 6, and 7 of A2002 for a nonparametric iid bootstrap under possible misspecification. The modified Lemmas 1, 2, 6, and 7 are denoted by AL1, AL2, AL6, and AL7, respectively. In addition, Lemma 5 of A2002 is denoted by AL5 without modification.

Lemma 1.

(a) *Lemma 1 of A2002 holds by replacing $\widetilde{X}_i$ and N with X_i and n, respectively, under our Assumption 1.*

(b) *Lemma 2 of A2002 for $j = 1$ holds under our Assumptions 1-3.*

(c) *Lemma 6 of A2002 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 A2002 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.*

Proof. The proof is given in Lee (2014). *Q.E.D.*

Lemma 2 shows the uniform convergence of the so-called inner loop and the objective function in θ. Since ET and ETEL solve the same inner loop optimization problem, we let $\rho(\nu) = 1 - e^{\nu}$ for ETEL for the next lemma. Define $\hat{\lambda}(\theta) = \arg\max_{\lambda \in \mathbf{R}^{L_g}} n^{-1} \sum_i \rho(\lambda' g_i(\theta))$ and $\lambda_0(\theta) = \arg\max_{\lambda \in \mathbf{R}^{L_g}} E\rho(\lambda' g_i(\theta))$. Such solutions exist and are continuously differentiable around a neighborhood of $\hat{\theta}$ and θ_0, respectively, by the implicit function theorem (Newey and Smith, 2004, proof of Theorem 2.1).

Lemma 2. *Suppose Assumptions 1-3 hold with $q_1 \geq 2$ and $q_1 > 2a$ for some $a \geq 0$. Then, for all $a \geq 0$ and all $\varepsilon > 0$,*

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

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

Proof. Since the proof is similar to those of Lemma 2 of HH and Theorem 10 of S2007, I provide a sketch of the proof. First, we need to show

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

This is proved by the proof of Lemma 2 of HH with $\rho(\lambda' g_i(\theta))$ in place of their $G(x,\theta)$, except that we use AL1(a) instead of Lemma 1 of HH. In particular, we apply AL1(a) with $c = 0$ and $h(X_i) = C_\rho(X_i) - EC_\rho(X_i)$ or $h(X_i) = \rho(\lambda_j' g_i(\theta_j)) - E\rho(\lambda_j' g_i(\theta_j))$ for some $(\theta_j', \lambda_j') \in \Theta\times\Lambda(\theta)$. Since a zero vector is in $\Lambda(\theta)$, Θ and $\Lambda(\theta)$ are compacts, and $\rho(0) = 0$, Assumption 2(d) implies that $E|\rho(\lambda' g_i(\theta))|^{q_1} < \infty$ for all $(\theta', \lambda') \in \Theta\times\Lambda(\theta)$. Thus, the conditions for AL1(a) is satisfied by letting $p = q_1$ and Assumption 2(d).

Next, we show

$$\lim_{n\to\infty} n^a P\left(\sup_{\theta\in\Theta}\left\|\bar{\lambda}(\theta)-\lambda_0(\theta)\right\| > \varepsilon\right) = 0, \tag{A.2}$$

where $\bar{\lambda}(\theta) = \arg\max_{\lambda\in\Lambda(\theta)} n^{-1}\sum_i \rho(\lambda' g_i(\theta))$. This is proved by using Step 1 of the proof of Theorem 10 of S2007. Then, the present lemma (a) is proved by a similar argument with the proof of Theorem 2.7 of Newey and McFadden (1994) using the concavity of $n^{-1}\sum_i \rho(\lambda' g_i(\theta))$ in λ for any θ.

Finally, the present lemma (b) can be shown as follows. By the triangle inequality, combining the following results proves the desired result.

$$\lim_{n\to\infty} n^a P\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i=1}^{n}\rho(\hat{\lambda}(\theta)'g_i(\theta)) - n^{-1}\sum_{i=1}^{n}\rho(\lambda_0(\theta)'g_i(\theta))\right| > \varepsilon\right) = 0, \tag{A.3}$$

$$\lim_{n\to\infty} n^a P\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i=1}^{n}\rho(\lambda_0(\theta)'g_i(\theta)) - E\rho(\lambda_0(\theta)'g_i(\theta))\right| > \varepsilon\right) = 0. \tag{A.4}$$

By Assumption 2(d), (A.3) follows from the present lemma (a) and AL1(b). Since $\lambda_0(\theta) \in int(\Lambda(\theta))$, (A.4) follows from (A.1). *Q.E.D.*

Let $\beta = (\theta', \lambda')$ and $\mathcal{B} \equiv \Theta\times\Lambda(\theta)$ for EL or ET. For ETEL, we introduce additional notations for the population auxiliary parameters. Define $\tau_0(\theta) \equiv E e^{\lambda_0(\theta)' g_i(\theta)}$ and

$$\kappa_0(\theta) \equiv -(E e^{\lambda_0(\theta)' g_i(\theta)} g_i(\theta) g_i(\theta)')^{-1} \tau_0(\theta) E g_i(\theta).$$

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 ETEL, $\beta \equiv (\theta', \lambda', \kappa', \tau)'$ and we define a compact set $\mathcal{B} \equiv \Theta \times \Lambda(\theta) \times \mathcal{K}(\theta) \times \mathcal{T}(\theta)$.

Let g and $G^{(j)}$ be an element of $g_i(\theta)$ and $G_i^{(j)}(\theta)$, respectively, for $j = 1, ..., d+1$. In addition, let g^k be a multiplication of any k-combination of elements of $g_i(\theta)$. For instance, if $g_i(\theta) = (g_{i,1}(\theta), g_{i,2}(\theta))'$, a 2×1 vector, then $g^2 = (g_{i,1}(\theta))^2$, $g_{i,1}(\theta) g_{i,2}(\theta)$, or $(g_{i,2}(\theta))^2$. $G^{(j)k}$ is defined analogously. To further simply notation, write g_0 and $G_0^{(j)}$ if g_i and $G_i^{(j)}$ are evaluated at θ_0 for $j = 1, 2, ...d+1$.

Lemma 3. *Suppose Assumptions 1-3 hold with $q_1 \geq 2$, $q_1 > 2a$, and $q_2 > \max\left\{2, \frac{2a}{1-2c}\right\}$ for some $c \in [0, 1/2)$ and some $a \geq 0$. Then, for all $c \in [0, 1/2)$ and all $a \geq 0$,*

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

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

Proof. We first show for any $\varepsilon > 0$,

$$\lim_{n \to \infty} n^a P\left(\|\hat{\beta} - \beta_0\| > \varepsilon\right) = 0. \tag{A.5}$$

First, consider EL or ET. Since $\rho(\lambda_0(\theta)' g_i(\theta))$ is continuous in θ and uniquely minimized at $\theta_0 \in int(\Theta)$, standard consistency arguments using Lemma 2(b) show that

$$\lim_{n \to \infty} n^a P\left(\|\hat{\theta} - \theta_0\| > \varepsilon\right) = 0. \tag{A.6}$$

Write $\hat{\lambda} \equiv \hat{\lambda}(\hat{\theta})$ and $\lambda_0 \equiv \lambda_0(\theta_0)$. By Lemma 2(a), (A.6), and the implicit function theorem that $\lambda_0(\theta)$ is continuous in a neighborhood of θ_0, it follows

$$\lim_{n \to \infty} n^a P\left(\|\hat{\lambda} - \lambda_0\| > \varepsilon\right) = 0. \tag{A.7}$$

This proves (A.5) for EL and ET. For ETEL, (A.6) and (A.7) can be shown by Step 2 of the proof of Theorem 10 of S2007 by applying AL1, AL2, and Lemma 2.

Since we have introduced auxiliary parameters (κ, τ) for ETEL, we need to prove consistency of $(\hat{\kappa}, \hat{\tau})$. Since $\hat{\kappa}$ and $\hat{\tau}$ are continuous functions of $\hat{\theta}$ and $\hat{\lambda}$, consistency of $\hat{\theta}$ and

$\hat{\lambda}$ implies that $\hat{\kappa}$ and $\hat{\tau}$ are also consistent. Formally, this can be shown as follows. First, we show

$$\lim_{n\to\infty} n^a P\left(\|\hat{\tau}-\tau_0\| > \varepsilon\right) = 0, \tag{A.8}$$

where $\hat{\tau} = n^{-1}\sum_{i=1}^n e^{\hat{\lambda}'\hat{g}_i}$ and $\tau_0 = Ee^{\lambda_0' g_{i0}}$. This follows from

$$\lim_{n\to\infty} n^a P\left(\left\|n^{-1}\sum_{i=1}^n e^{\hat{\lambda}'\hat{g}_i} - n^{-1}\sum_{i=1}^n e^{\lambda_0' g_{i0}}\right\| > \varepsilon\right) = 0, \tag{A.9}$$

$$\lim_{n\to\infty} n^a P\left(\left\|n^{-1}\sum_{i=1}^n e^{\lambda_0' g_{i0}} - Ee^{\lambda_0' g_{i0}}\right\| > \varepsilon\right) = 0. \tag{A.10}$$

To show (A.9), we apply (A.6), (A.7), and AL1(b) with $h(X_i) = C_{\partial\rho}(X_i)$ and $p = q_2$. The second result (A.10) follows from applying AL1(a) with $c = 0$, $h(X_i) = e^{\lambda_0' g_{i0}} - Ee^{\lambda_0' g_{i0}}$, and $p = q_2$. Next, we show

$$\lim_{n\to\infty} n^a P\left(\|\hat{\kappa}-\kappa_0\| > \varepsilon\right) = 0. \tag{A.11}$$

This can be shown by combining (A.8) and the following results:

$$\lim_{n\to\infty} n^a P\left(\|\hat{g}_n - g_n(\theta_0)\| > \varepsilon\right) = 0, \tag{A.12}$$

$$\lim_{n\to\infty} n^a P\left(\|g_n(\theta_0) - Eg_{i0}\| > \varepsilon\right) = 0, \tag{A.13}$$

$$\lim_{n\to\infty} n^a P\left(\left\|n^{-1}\sum_{i=1}^n e^{\hat{\lambda}'\hat{g}_i}\hat{g}_i\hat{g}_i' - n^{-1}\sum_{i=1}^n e^{\lambda_0' g_{i0}} g_{i0}g_{i0}'\right\| > \varepsilon\right) = 0, \tag{A.14}$$

$$\lim_{n\to\infty} n^a P\left(\left\|n^{-1}\sum_{i=1}^n e^{\lambda_0' g_{i0}} g_{i0}g_{i0}' - Ee^{\lambda_0' g_{i0}} g_{i0}g_{i0}'\right\| > \varepsilon\right) = 0. \tag{A.15}$$

The first result (A.12) holds by Assumption 2(c), AL1(b) with $h(X_i) = C_g(X_i)$ and $p = q_g$, and (A.6). The second result (A.13) holds by Assumption 2(c) and AL1(a) with $h(X_i) = g_i(\theta_0) - Eg_i(\theta_0)$, $c = 0$ and $p = q_g$. The third result (A.14) can be shown by applying the triangle inequality, AL1(b), (A.6) and (A.7), and Schwarz matrix inequality multiple times. In particular, we apply AL1(b) with $h(X_i) = C_{\partial\rho}(X_i)\|g_{i0}\|^2$, $h(X_i) = C_{\partial\rho}(X_i)C_g^2(X_i)$, $h(X_i) = C_{\partial\rho}(X_i)C_g(X_i)\|g_{i0}\|$, $h(X_i) = e^{\lambda_0' g_{i0}}C_g^2(X_i)$, and $h(X_i) = e^{\lambda_0' g_{i0}}C_g(X_i)\|g_{i0}\|$. For $h(X_i) = C_{\partial\rho}(X_i)\|g_i(\theta_0)\|^2$, by Hölder's inequality,

$$EC_{\partial\rho}^p(X_i)\|g_{i0}\|^{2p} \leq \left(EC_{\partial\rho}^{p(1+\epsilon)}(X_i)\right)^{\frac{1}{1+\epsilon}} \cdot \left(E\|g_{i0}\|^{2p(1+\epsilon^{-1})}\right)^{\frac{\epsilon}{1+\epsilon}}, \tag{A.16}$$

for any $0 < \epsilon < \infty$. Since Assumption 2(c) holds for all $q_g < \infty$, we can take small enough ϵ so that $p = q_2 > \max\{2, 2a\}$ implies that (A.16) is finite by Assumption 3(d). Other

$h(X_i)$'s can be shown to satisfy the condition similarly. Note that Assumption 3(d) implies $Ee^{q_2\lambda_0' g_i(\theta_0)} < \infty$ for $q_2 > \max\{2, 2a\}$, because (i) a zero vector is in $\Lambda(\theta)$, (ii) Θ and $\Lambda(\theta)$ are compacts, and (iii) $\rho(0) = 0$. The last result (A.15) can be shown by applying AL1(a) with $c = 0$ and $h(X_i) = e^{\lambda_0' g_{i0}} g_{i0} g_{i0}' - Ee^{\lambda_0' g_{i0}} g_{i0} g_{i0}'$. To see if $h(X_i)$ satisfies the condition of AL1(a), it suffices to show $Ee^{p\lambda_0' g_{i0}} \|g_{i0}\|^{2p} < \infty$ for $p \geq 2$ and $p > 2a$, but this condition is met by letting $p = q_2$ and using Hölder's inequality. Thus, (A.5) is proved for ETEL.

Since we have established consistency of $\hat{\beta}$ for β_0, we now show the present lemma. The proof is similar to that of Lemma 3 of A2002 and Step 3 of the proof of Theorem 10 of S2007. Since $\hat{\beta}$ is in the interior of $\mathcal{B}$ with probability $1 - o(n^{-a})$, $\hat{\beta}$ is the solution to $n^{-1}\sum_{i=1}^n \psi(X_i, \hat{\beta}) = 0$ with probability $1 - o(n^{-a})$. By the mean value expansion of $n^{-1}\sum_{i=1}^n \psi(X_i, \hat{\beta}) = 0$ around β_0,

$$\hat{\beta} - \beta_0 = -\left(n^{-1}\sum_{i=1}^n \frac{\partial \psi(X_i, \tilde{\beta})}{\partial \beta'}\right)^{-1} n^{-1}\sum_{i=1}^n \psi(X_i, \beta_0), \tag{A.17}$$

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

$$\lim_{n\to\infty} n^a P\left(\left\|n^{-1}\sum_{i=1}^n \frac{\partial \psi(X_i, \tilde{\beta})}{\partial \beta'} - n^{-1}\sum_{i=1}^n \frac{\partial \psi(X_i, \beta_0)}{\partial \beta'}\right\| > \varepsilon\right) = 0, \tag{A.18}$$

$$\lim_{n\to\infty} n^a P\left(\left\|n^{-1}\sum_{i=1}^n \frac{\partial \psi(X_i, \beta_0)}{\partial \beta'} - E\frac{\partial \psi(X_i, \beta_0)}{\partial \beta'}\right\| > \varepsilon\right) = 0, \tag{A.19}$$

$$\lim_{n\to\infty} n^a P\left(\left\|n^{-1}\sum_{i=1}^n \psi(X_i, \beta_0)\right\| > n^{-c}\right) = 0. \tag{A.20}$$

First, to show (A.18), observe that the elements of $(\partial/\partial\beta')\psi(X_i, \beta)$ have the form

$$\alpha \cdot \rho_j^{k_\rho}(\lambda' g_i) \cdot g^{k_0} \cdot G^{k_1} \cdot G^{(2)k_2}, \;\; j = 1, 2, \tag{A.21}$$

where α denotes products of components of β, $k_\rho = 1$, $k_0 \leq 2$, $k_1 \leq 2$, and $k_2 \leq 1$ for EL and ET. For ETEL, we replace $\rho_j^{k_\rho}(\lambda' g_{i0})$ with $e^{k_\rho \lambda_0' g_{i0}}$, where $k_\rho = 0, 1$, $k_0 \leq 3$, $k_1 \leq 2$, and $k_2 \leq 1$. For each element, we apply (A.5) and AL1(b) multiple times. For example, $\rho_2(\lambda' g_i(\theta)) g_i(\theta) g_i(\theta)'$ is an element of $(\partial/\partial\beta')\psi(X_i, \beta)$. Then,

$$\begin{aligned} & \|\rho_2(\tilde{\lambda}'\tilde{g}_i)\tilde{g}_i\tilde{g}_i' - \rho_2(\lambda_0' g_{i0}) g_{i0} g_{i0}'\| \qquad \text{(A.22)} \\ \leq \; & \|\tilde{\beta} - \beta_0\| \left(C_{\partial\rho}(X_i) + |\rho_2(\lambda_0' g_{i0})| \cdot C_g(X_i)\left(C_g(X_i)\|\tilde{\beta} - \beta_0\| + 2\|g_{i0}\|\right)\right), \end{aligned}$$

where $\tilde{g}_i = g_i(\tilde{\theta})$. Now by using the fact that $|\rho_2(\lambda_0' g_{i0})| \leq C_{\partial\rho}(X_i)\|\beta_0\|$, Assumptions 2-3, (A.5), and AL1(b), we show

$$P(\|\rho_2(\tilde{\lambda}'\tilde{g}_i)\tilde{g}_i\tilde{g}_i' - \rho_2(\lambda_0' g_{i0})g_{i0}g_{i0}'\| > \varepsilon) = o(n^{-a}). \tag{A.23}$$

Other terms can be shown similarly. The condition of AL1(b) is satisfied by Assumptions 2-3, Hölder's inequality, and letting $p = q_2$. This proves (A.18). The second result (A.19) can be shown analogously by using AL1(a) with $c = 0$ and $h(X_i) = (\partial/\partial\beta')\psi(X_i, \beta_0) - E(\partial/\partial\beta')\psi(X_i, \beta_0)$. The last result (A.20) holds by AL1(a) with $h(X_i) = \psi(X_i, \beta_0)$. By using Hölder's inequality, the conditions of AL1(a) is satisfied if we let $p = q_2 > \max\{2, \frac{2a}{1-2c}\}$, which hold by the assumption of the lemma. *Q.E.D.*

Let P^* be the probability distribution of the bootstrap sample conditional on the original sample. Let E^* denote expectation with respect to P^*. Since we consider the nonparametric iid bootstrap, E^* is taken over the original sample with respect to the edf. For example, $E^* X_i^* = n^{-1}\sum_{i=1}^n X_i$. Write $g_i^*(\theta) \equiv g(X_i^*, \theta)$ and $\hat{g}_i^* \equiv g^*(\hat{\theta}^*)$. Define $\hat{\lambda}^*(\theta) = \arg\max_{\lambda \in \mathbf{R}^{L_g}} n^{-1}\sum_i \rho(\lambda' g_i^*(\theta))$. By the implicit function theorem, this solution exists and is continuously differentiable in a neighborhood of $\hat{\theta}^*$. Write $\hat{\lambda}^* \equiv \hat{\lambda}^*(\hat{\theta}^*)$ for notational brevity. Lemma 4 is the bootstrap version of Lemma 2. Let $\rho(\nu) = 1 - e^{\nu}$ for ETEL in the next lemma.

Lemma 4. *Suppose Assumptions 1-3 hold with $q_1 \geq 2$ and $q_1 > 4a$. Then, for all $a \geq 0$ and all $\varepsilon > 0$,*

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

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

Proof. We first show

$$\lim_{n\to\infty} n^a P\left(P^*\left(\sup_{(\theta,\lambda)\in\Theta\times\Lambda(\theta)}\left|n^{-1}\sum_{i=1}^n \left(\rho(\lambda' g_i^*(\theta)) - \rho(\lambda' g_i(\theta))\right)\right| > \varepsilon\right) > n^{-a}\right) = 0. \tag{A.24}$$

We use the proof of Lemma 8 of HH using AL6(a) with $c = 0$, rather than Lemma 7 of HH. Since $n^{-1}\sum_{i=1}^n \rho(\lambda' g_i(\theta)) = E^*\rho(\lambda' g_i^*(\theta))$, we apply AL6(a) with $h(X_i) = \rho(\lambda_j' g_i(\theta_j)) - E\rho(\lambda_j' g_i(\theta_j))$ for any $(\theta_j, \lambda_j) \in \Theta \times \Lambda(\theta)$ or $h(X_i) = C_\rho(X_i) - EC_\rho(X_i)$. By Minkowski inequality, it suffices to show $E|\rho(\lambda_j' g_i(\theta_j))|^p < \infty$ and $EC_\rho^p(X_i) < \infty$ for $p \geq 2$ and $p > 4a$. This is satisfied by letting $p = q_1$.

Next, we show

$$\lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}\left\|\bar{\lambda}^*(\theta)-\bar{\lambda}(\theta)\right\|>\varepsilon\right)>n^{-a}\right)=0, \tag{A.25}$$

where $\bar{\lambda}^*(\theta) = \arg\max_{\lambda\in\Lambda(\theta)} n^{-1}\sum_{i=1}^n \rho(\lambda' g_i^*(\theta))$. We claim that for a given $\varepsilon > 0$, there exists $\eta > 0$ independent of n such that for any $\theta \in \Theta$ and any $\lambda \in \Lambda(\theta)$, $\|\lambda - \bar{\lambda}(\theta)\| > \varepsilon$ implies that $n^{-1}\sum_i \rho(\bar{\lambda}(\theta)' g_i(\theta)) - n^{-1}\sum_i \rho(\lambda' g_i(\theta)) \geq \eta > 0$ with probability $1 - o(n^{-a})$. This claim can be shown by similar arguments with the proof of Lemma 9 of A2002. For any $\theta \in \Theta$ and any $\lambda \in \Lambda(\theta)$, whenever $\|\lambda - \bar{\lambda}(\theta)\| > \varepsilon$, $\|\lambda - \lambda_0(\theta)\| > \varepsilon/2$ with probability $1-o(n^{-a})$ by the triangle inequality and Lemma 2. Since $E\rho(\lambda' g_i(\theta))$ is uniquely maximized at $\lambda_0(\theta)$ and continuous on $\Lambda(\theta)$, $\|\lambda - \lambda_0(\theta)\| > \varepsilon/2$ implies that there exists $\eta(\theta)$ such that

$$\begin{aligned}
0<\eta(\theta) &\leq E\rho(\lambda_0(\theta)' g_i(\theta)) - E\rho(\lambda' g_i(\theta)) \qquad \text{(A.26)}\\
&\leq n^{-1}\sum_i \rho(\bar{\lambda}(\theta)' g_i(\theta)) - n^{-1}\sum_i \rho(\lambda' g_i(\theta))\\
&\quad + E\rho(\lambda_0(\theta)' g_i(\theta)) - n^{-1}\sum_i \rho(\lambda_0(\theta)' g_i(\theta)) - E\rho(\lambda' g_i(\theta)) + n^{-1}\sum_i \rho(\lambda' g_i(\theta))\\
&\leq n^{-1}\sum_i \rho(\bar{\lambda}(\theta)' g_i(\theta)) - n^{-1}\sum_i \rho(\lambda' g_i(\theta))\\
&\quad + 2\sup_{(\theta,\lambda)\in\Theta\times\Lambda(\theta)} |n^{-1}\sum_i \rho(\lambda' g_i(\theta)) - E\rho(\lambda' g_i(\theta))|.
\end{aligned}$$

Since (A.1) holds for all ε, letting $\varepsilon = \eta(\theta)/3$ in (A.1) and $\eta = \inf_\theta \eta(\theta)$ proves the claim. Then, we have

$$\begin{aligned}
& P(P^*(\sup_{\theta\in\Theta}\|\bar{\lambda}^*(\theta)-\bar{\lambda}(\theta)\|>\varepsilon)>n^{-a}) \qquad \text{(A.27)}\\
\leq\ & P\left(P^*\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_i\left(\rho(\bar{\lambda}(\theta)' g_i(\theta)) - \rho(\bar{\lambda}^*(\theta)' g_i(\theta))\right)\right|>\eta\right)>n^{-a}\right)\\
\leq\ & P\left(P^*\left(\sup_{(\theta,\lambda)\in\Theta\times\Lambda(\theta)}\left|n^{-1}\sum_i\left(\rho(\lambda' g_i^*(\theta)) - \rho(\lambda' g_i(\theta))\right)\right|>\eta/2\right)>n^{-a}\right)=o(n^{-a}).
\end{aligned}$$

The second inequality holds by adding and subtracting $n^{-1}\sum_i \rho(\bar{\lambda}(\theta)' g_i^*(\theta))$, and using the definition of $\bar{\lambda}^*(\theta)$. The last equality follows by (A.24). The present lemma (a) can be obtained by replacing $\bar{\lambda}^*(\theta)$ and $\bar{\lambda}(\theta)$ with $\hat{\lambda}^*(\theta)$ and $\hat{\lambda}(\theta)$, respectively. Since $n^{-1}\sum_i \rho(\lambda' g_i(\theta))$ and $n^{-1}\sum_i \rho(\lambda' g_i^*(\theta))$ are concave in λ for any θ, as long as $\bar{\lambda}(\theta)$ and $\bar{\lambda}^*(\theta)$ are in the interior of $\Lambda(\theta)$, they are maximizers on $\mathbf{R}^{L_g}$ by Theorem 2.7 of Newey and McFadden (1994). But by Assumption 2, $\bar{\lambda}(\theta) \in int(\Lambda(\theta))$ with probability $1 - o(n^{-a})$ and $\bar{\lambda}^*(\theta) \in int(\Lambda(\theta))$

with P^* probability $1 - o(n^{-a})$ except, possibly, if χ is in a set of P probability $o(n^{-a})$. Therefore, the present lemma (a) is proved.

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

$$\lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i=1}^{n}\left(\rho(\hat{\lambda}^*(\theta)'g_i^*(\theta)) - \rho(\hat{\lambda}(\theta)'g_i^*(\theta))\right)\right| > \varepsilon\right) > n^{-a}\right) = 0, \quad (A.28)$$

$$\lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i=1}^{n}\left(\rho(\hat{\lambda}(\theta)'g_i^*(\theta)) - \rho(\hat{\lambda}(\theta)'g_i(\theta))\right)\right| > \varepsilon\right) > n^{-a}\right) = 0. \quad (A.29)$$

(A.28) can be shown as follows. By Assumption 2(d) and standard manipulation,

$$P\left(P^*\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i=1}^{n}\left(\rho(\hat{\lambda}^*(\theta)'g_i^*(\theta)) - \rho(\hat{\lambda}(\theta)'g_i^*(\theta))\right)\right| > \varepsilon\right) > n^{-a}\right) \quad (A.30)$$
$$\leq P\left(P^*\left(n^{-1}\sum_i C_\rho(X_i^*) > \varepsilon\right) > n^{-a}/2\right) + P\left(P^*\left(\sup_{\theta\in\Theta}\|\hat{\lambda}^*(\theta) - \hat{\lambda}(\theta)\| > 1\right) > n^{-a}/2\right).$$

We apply AL6(d) with $h(X_i) = C_\rho(X_i)$ and $p = q_1$ for the first term in the right-hand side (RHS) of the above inequality, and apply the present lemma (a) for the second term to show that the RHS is $o(n^{-a})$. This proves (A.28). Since $\hat{\lambda}(\theta) \in int(\Lambda(\theta))$ with probability $1 - o(n^{-a})$, (A.29) follows from (A.24). *Q.E.D.*

Lemma 5. *Suppose Assumptions 1-3 hold with $q_1 \geq 2$, $q_1 > 4a$, and $q_2 > \max\left\{2, \frac{4a}{1-2c}\right\}$ for some $c \in [0, 1/2)$ and some $a \geq 0$. Then, for all $c \in [0, 1/2)$ and all $a \geq 0$,*

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

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

Proof. The proof is analogous to that of Lemma 3 except that it involves additional steps for the bootstrap versions of the estimators. First, we show

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|\hat{\beta}^* - \hat{\beta}\| > \varepsilon\right) > n^{-a}\right) = 0. \quad (A.31)$$

Consider EL or ET. We claim that for a given $\varepsilon > 0$, there exists $\eta > 0$ independent of n such that $\|\theta - \hat{\theta}\| > \varepsilon$ implies that $0 < \eta \leq n^{-1}\sum_i \rho(\hat{\lambda}(\theta)'g_i(\theta)) - n^{-1}\sum_i \rho(\hat{\lambda}'\hat{g}_i)$ with probability $1 - o(n^{-a})$. This claim can be shown by a similar argument with (A.26) by using the fact that $E\rho(\lambda_0(\theta)'g_i(\theta))$ is uniquely minimized at θ_0 and continuous in θ, AL1(b), Lemma 2(a), (A.4), (A.6), and (A.7). Thus, we have

$$
\begin{aligned}
& P\left(P^*\left(\|\hat{\theta}^* - \hat{\theta}\| > \varepsilon\right) > n^{-a}\right) && \text{(A.32)}\\
\leq\ & P\left(P^*\left(\left|n^{-1}\sum_i \left(\rho(\hat{\lambda}(\hat{\theta}^*)'g_i(\hat{\theta}^*)) - \rho(\hat{\lambda}'\hat{g}_i)\right)\right| > \eta\right) > n^{-a}\right)\\
\leq\ & P\left(P^*\left(\left|n^{-1}\sum_i \left(\rho(\hat{\lambda}(\hat{\theta}^*)'g_i(\hat{\theta}^*)) - \rho(\hat{\lambda}^{*\prime}g_i^*) + \rho(\hat{\lambda}^*(\hat{\theta})'g_i^*(\hat{\theta})) - \rho(\hat{\lambda}'\hat{g}_i)\right)\right| > \eta\right) > n^{-a}\right)\\
\leq\ & P\left(P^*\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_i \left(\rho(\hat{\lambda}^*(\theta)'g_i^*(\theta)) - \rho(\hat{\lambda}(\theta)'g_i(\theta))\right)\right| > \eta/2\right) > n^{-a}\right) = o(n^{-a}),
\end{aligned}
$$

by Lemma 4(b). To show

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|\hat{\lambda}^* - \hat{\lambda}\| > \varepsilon\right) > n^{-a}\right) = 0, \tag{A.33}$$

we use the triangle inequality, (A.6), (A.32), Lemma 2(a), Lemma 4(a), and the implicit function theorem that $\lambda_0(\theta)$ is continuously differentiable around θ_0. This proves (A.31) for EL or ET. For ETEL, an analogous result to Lemma 4(b),

$$\lim_{n\to\infty} n^a P\left(P^*\left(\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i=1}^n \left(\hat{l}_n^*(\theta) - \hat{l}_n(\theta)\right)\right| > \varepsilon\right) > n^{-a}\right) = 0, \tag{A.34}$$

where

$$\hat{l}_n^*(\theta) = \log\left(n^{-1}\sum_{i=1}^n e^{\hat{\lambda}^*(\theta)'(g_i^*(\theta) - g_n^*(\theta))}\right), \tag{A.35}$$

can be shown by Lemma 4(a), AL6, and AL7. Then, replacing $E\rho(\lambda_0(\theta)'g_i(\theta))$ with $l_0(\theta)$ and $n^{-1}\sum_i \rho(\hat{\lambda}(\theta)'g_i(\theta))$ with $\hat{l}_n(\theta)$, and applying a similar argument with (A.26) give (A.32) and (A.33) for ETEL.

For the auxiliary parameters κ and τ, the bootstrap versions of the estimators are

$$
\begin{aligned}
\hat{\kappa}^* &= -\left(n^{-1}\sum_{i=1}^n e^{\hat{\lambda}^{*\prime}\hat{g}_i^*}\hat{g}_i^*\hat{g}_i^{*\prime}\right)^{-1}\hat{\tau}^*\hat{g}_n^*, && \text{(A.36)}\\
\hat{\tau}^* &= n^{-1}\sum_{i=1}^n e^{\hat{\lambda}^{*\prime}\hat{g}_i^*}. && \text{(A.37)}
\end{aligned}
$$

First, the bootstrap version of (A.8) is

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|\hat{\tau}^* - \hat{\tau}\| > \varepsilon\right) > n^{-a}\right) = 0. \tag{A.38}$$

This follows from the triangle inequality, AL6(d) with $h(X_i^*) = C_{\partial\rho}(X_i^*)$, Lemma 4(b), (A.32), (A.33), and the implicit function theorem that $\hat{\lambda}^*(\theta)$ is continuously differentiable around $\hat{\theta}^*$. Second, the bootstrap version of (A.11) is

$$\lim_{n\to\infty} n^a P\left(P^*\left(\|\hat{\kappa}^* - \hat{\kappa}\| > \varepsilon\right) > n^{-a}\right) = 0, \tag{A.39}$$

and this follows from (A.32), (A.33), (A.38), Lemma 4, AL7, and multiple applications of AL6. In particular, the condition of AL6 is satisfied with $p = q_2 > \max\{2, 4a\}$ by using a similar argument with (A.16). Thus, (A.31) is proved for ETEL.

The rest of the proof to show the argument of the lemma (with n^{-c} in place of ε) is analogous to that of Lemma 3 except that we apply AL6 instead of AL1. By Hölder's inequality, the binding condition is $p = q_2 > \max\{2, 4a/(1-2c)\}$ for AL6 but this is satisfied by the assumption of the lemma. *Q.E.D.*

Let $f(X_i, \beta)$ be a vector containing the unique components of $\psi(X_i, \beta)$ and its derivatives with respect to the components of β through order d, and $\psi(X_i, \beta)\psi(X_i, \beta)'$ and its derivatives with respect to the components of β through order $d-1$. We also introduce some additional notation. Let S_n be a 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 a 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 6. *(a) Suppose Assumptions 1-3 hold with $q_2 > \max\{4, 4a\}$ for some $a \geq 0$. Then, for all $\varepsilon > 0$ and all $a \geq 0$,*

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

(b) Suppose Assumptions 1-3 hold with $q_1 \geq 2$, $q_1 > 2a$, and $q_2 > \max\{4, 8a\}$ for some $a \geq 0$. Then, for all $\varepsilon > 0$ and all $a \geq 0$,

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

Proof. The present lemma (a) can be shown as follows. By the definitions of S_n and S, it suffices to show

$$P\left(\left\|n^{-1}\sum_{i=1}^n f(X_i, \beta_0) - Ef(X_i, \beta_0)\right\| > \varepsilon\right) = o(n^{-a}). \tag{A.40}$$

We apply AL1(b) with $c = 0$ and $h(X_i)$ being any unique component of $f(X_i, \beta_0) -$

$Ef(X_i, \beta_0)$. To satisfy the condition of AL1(b), $p \geq 2$ and $p > 2a$, we need to investigate the components of $f(X_i, \beta_0)$. For EL or ET, $f(X_i, \beta)$ consists of terms of the form

$$\alpha \cdot \rho_j^{k_\rho}(\lambda' g_i(\theta)) \cdot g^{k_0} \cdot G^{k_1} \cdots G^{(d+1)k_{d+1}}, \tag{A.41}$$

where α denotes products of components of β and and k_l's are nonnegative integers for $l = 0, 1, ...d+1$. In addition, $j = 1, ..., d+1$, $k_\rho = 1, 2$, $k_0, k_1 \leq d+1$, $k_l \leq d - l + 1$ for $l = 2, ..., d$, $k_{d+1} \leq 1$, and $\sum_{l=0}^{d+1} k_l \leq d+1$. For ETEL, we replace $\rho_j^{k_\rho}(\lambda' g_i(\theta))$ with $e^{k_\rho \lambda' g_i(\theta)}$, where $k_\rho = 0, 1, 2$, $k_0 \leq d+3$, $k_l \leq d - l + 2$ for $l = 1, 2, ..., d+1$, and $\sum_{l=0}^{d+1} k_l \leq d+3$. Since we assume that all the finite moments exist for $g_i(\theta)$, $\forall \theta \in \Theta$ and $G_{i0}^{(j)}$, $j = 1, 2, ..., d+1$, the values of k_l's do not impose additional restriction on the values of q_g and q_G in Assumptions 2-3. What matters is k_ρ, because the value of k_ρ is directly related to q_2 in Assumption 3(d). Since $k_\rho = 2$ is the most restrictive case, it suffices to show $EC_{\partial\rho}^{2p}(X_i)C_g^{(d+3)p}(X_i) < \infty$, $EC_{\partial\rho}^{2p}(X_i)C_G^{(d+3)p}(X_i) < \infty$, $Ee^{2p\lambda_0' g_{i0}}C_g^{(d+3)p}(X_i) < \infty$ and $Ee^{2p\lambda_0' g_{i0}}C_G^{(d+3)p}(X_i) < \infty$ for AL1(b) to be applied. By Hölder's inequality, letting $p = q_2 > \max\{4, 4a\}$ satisfies these conditions.

The present lemma (b) can be shown as follows. By the definitions of S_n^* and S^*, it suffices to show

$$P\left(P^*\left(\left\| n^{-1}\sum_{i=1}^n f(X_i^*, \hat\beta) - n^{-1}\sum_{i=1}^n f(X_i, \hat\beta)\right\| > \varepsilon\right) > n^{-a}\right) = o(n^{-a}). \tag{A.42}$$

By the triangle inequality,

$$\begin{aligned} \left\| n^{-1}\sum_{i=1}^n \left(f(X_i^*, \hat\beta) - f(X_i, \hat\beta)\right)\right\| \quad &\leq \quad \left\| n^{-1}\sum_{i=1}^n \left(f(X_i^*, \beta_0) - f(X_i, \beta_0)\right)\right\| \\ &\quad + \left\| n^{-1}\sum_{i=1}^n \left(f(X_i^*, \hat\beta) - f(X_i^*, \beta_0)\right)\right\| \\ &\quad + \left\| n^{-1}\sum_{i=1}^n \left(f(X_i, \hat\beta) - f(X_i, \beta_0)\right)\right\|. \end{aligned} \tag{A.43}$$

For the first term of the RHS of the inequality (A.43), we apply Lemma AL6(a) with $c = 0$ and $h(X_i) = f(X_i, \beta_0) - Ef(X_i, \beta_0)$. By using a similar argument with the proof of (A.40), the most restrictive condition is met with $p = q_2 > \max\{4, 8a\}$. The second and the last terms are shown by combining Lemma 3 with $c = 0$ and the following results: For all $\beta \in N(\beta_0)$, some neighborhood of β_0, there exist some functions $C(X_i)$ and $C^*(X_i^*)$ such that

$$\begin{aligned} \|f(X_i,\beta)-f(X_i,\beta_0)\| &\le C(X_i)\|\beta-\beta_0\|, \qquad \text{(A.44)}\\ \|f(X_i^*,\beta)-f(X_i^*,\beta_0)\| &\le C^*(X_i^*)\|\beta-\beta_0\|, \qquad \text{(A.45)} \end{aligned}$$

and these functions satisfy for some $K<\infty$,

$$\begin{aligned} &\lim_{n\to\infty} n^a P(\|n^{-1}\sum_{i=1}^n C(X_i)\|>K)=0, \qquad \text{(A.46)}\\ &\lim_{n\to\infty} n^a P(P^*(\|n^{-1}\sum_{i=1}^n C^*(X_i^*)\|>K)>n^{-a})=0. \qquad \text{(A.47)} \end{aligned}$$

After some tedious but straightforward calculation using the binomial theorem, the triangle inequality, and Hölder's inequality, AL1(b) implies that the most restrictive case for the existence of such $C(X_i)$ occurs when $k_\rho=2$, which is satisfied with $p=q_2>\max\{4,4a\}$. Similarly, the condition of AL6(d) with $h(X_i^*)=C^*(X_i^*)$ is satisfied with $p=q_2>\max\{4,8a\}$. *Q.E.D.*

Lemma 7. *Let Δ_n and Δ_n^* denote $\sqrt{n}(\hat\theta-\theta_0)$ and $\sqrt{n}(\hat\theta^*-\hat\theta)$, or T_{MR} and T_{MR}^*. 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 $q_1\ge 2$, $q_1>2a$, and $q_2>\max\left\{4,4a,\frac{2ad}{d-2a-1}\right\}$ and $d\ge 2a+2$ for some $a\ge 0$, where $2a$ is a positive integer. Then,

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

(b) Suppose Assumptions 1-4 hold with $q_1\ge 2$, $q_1>4a$, and $q_2>\max\left\{4,8a,\frac{4ad}{d-2a-1}\right\}$ and $d\ge 2a+2$ for some $a\ge 0$, where $2a$ is a positive integer. Then,

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

Proof. The proof is analogous to that of Lemma 13(a) of A2002 that uses his Lemmas 1 and 3-9. His Lemmas 1, 5, 6, and 7 are used in the proof, and denoted by AL1, AL5, AL6, and AL7, respectively. His Lemma 3 is replaced by our Lemma 3. His Lemmas 4 and 8 are not required because GEL is a one-step estimator without a weight matrix. His Lemma 9 is replaced by our Lemma 5. The main difference is that the conditions on q_1 and q_2 do not appear in the proof of A2002 for GMM. Lemma 6 is used to give conditions for q_1 and q_2.

I provide a sketch of the proof and an explanation where the conditions of the lemma are derived from.

For part (a), the proof proceeds by taking Taylor expansion of the FOC around β_0 through order $d-1$. The remainder term ζ_n from the Taylor expansion satisfies $\|\zeta_n\| \leq M\|\hat{\beta}-\beta_0\|^d \leq n^{-dc}$ for some $M<\infty$ with probability $1-o(n^{-a})$ by Lemma 3. To apply AL5(a), the conditions such that $n^{-dc+1/2}=o(n^{-a})$ or $dc \geq a+1/2$ for some $c \in [0,1/2)$, and that $2a$ is an integer, need to be satisfied. The former is satisfied if $d > 2a+1$ or $d \geq 2a+2$ (both d and $2a$ are integers), and the latter is assumed. Since the condition on q_2 of Lemma 3 is minimized with the smallest c, let $c=(a+1/2)d^{-1}$. By plugging this into the condition of Lemma 3, we have $q_1 \geq 2$, $q_1 > 2a$, and $q_2 > \max\{2, 2ad(d-2a-1)^{-1}\}$. In addition, we use Lemma 6(a) to use the implicit function theorem for the existence of $A(\cdot)$. By collecting the conditions of Lemmas 3 and 6(a), we have the condition for the present lemma.[9]

The proof of part (b) proceeds analogously. By plugging the same c into the condition of Lemma 5, we have $q_2 > \max\{2, 4ad(d-2a-1)^{-1}\}$. The condition of Lemma 6(b) is $q_1 \geq 2$, $q_1 > 2a$, and $q_2 > \max\{4, 8a\}$. The condition of the present lemma collects these conditions. *Q.E.D.*

We define the components of the Edgeworth expansions of the test statistic T_{MR} and its bootstrap analog T^*_{MR}. 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 $n^{\alpha(m)}E\Pi^m_{\mu=1}\Psi_{n,j_\mu}$ and $n^{\alpha(m)}E^*\Pi^m_{\mu=1}\Psi^*_{n,j_\mu}$, 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 8.

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} and T^*_{MR} depend on $\pi_i(\delta,\nu_a)$ and $\pi_i(\delta,\nu^*_{n,a})$, respectively.

Lemma 8. *(a) Suppose Assumptions 1-3 hold with $q_2 > 4(a+1)$ for some $a \geq 0$. Then, for all $a \geq 0$, $\nu_{n,a}$ and $\nu_a \equiv \lim_{n\to\infty}\nu_{n,a}$ exist.*
(b) Suppose Assumptions 1-3 hold with $q_1 \geq 2$, $q_1 > 2a$, and $q_2 > \max\left\{8(a+1), \frac{8a(a+1)}{(1-2\xi)}\right\}$

[9] There is a trade-off between the values of d, smoothness of the moment function, and q_2, the existence of higher moments of $C_{\partial\rho}(X_i)$ or $e^{\lambda_0 g_{i0}}$. Since $\lambda_0 \neq 0$ under misspecification, the value of q_2 may restrict the DGP for the bootstrap to be implemented. This issue is treated separately in Lee (2014b).

for some $a \geq 0$ and some $\xi \in [0, 1/2)$. Then, for all $a \geq 0$ and all $\xi \in [0, 1/2)$,

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

Proof. We first show the present lemma (a). Since $\nu_{n,a}$ contains multiplications of possibly different components of $\Psi_n = \sqrt{n}(S_n - S)$, it suffices to show the result for the least favorable term (with respect to the value of q_2) in Ψ_n. Let s_i be the least favorable term in $f(X_i, \beta_0)$. Since $a = 2$ is the largest number that we need in later lemmas, we show the lemma for $m = 2, 3, 4, 5, 6$. Then, we can show

$$n^{\alpha(2)} E\Pi^2_{\mu=1}\Psi_{n,j_\mu} = Es_i^2 - (Es_i)^2 = \lim_{n\to\infty} n^{\alpha(2)} E\Pi^2_{\mu=1}\Psi_{n,j_\mu}, \tag{A.48}$$

$$n^{\alpha(3)} E\Pi^3_{\mu=1}\Psi_{n,j_\mu} = Es_i^3 - 3Es_iEs_i^2 + 2(Es_i)^3 = \lim_{n\to\infty} n^{\alpha(3)} E\Pi^3_{\mu=1}\Psi_{n,j_\mu}, \tag{A.49}$$

$$\begin{aligned} n^{\alpha(4)} E\Pi^4_{\mu=1}\Psi_{n,j_\mu} &= \frac{1}{n}Es_i^4 - \frac{4}{n}Es_iEs_i^3 + \frac{-6n+12}{n}(Es_i)^2Es_i^2 \\ &\quad + \frac{3(n-1)}{n}(Es_i^2)^2 + \frac{3(n-2)}{n}(Es_i)^4 \\ &\underset{n\to\infty}{\to} 3(Es_i)^4 + 3(Es_i^2)^2 - 6(Es_i)^2Es_i^2 = \lim_{n\to\infty} n^{\alpha(4)} E\Pi^4_{\mu=1}\Psi_{n,j_\mu}. \end{aligned} \tag{A.50}$$

$$\begin{aligned} n^{\alpha(5)} E\Pi^5_{\mu=1}\Psi_{n,j_\mu} &= \frac{1}{n}Es_i^5 - \frac{5}{n}Es_iEs_i^4 - \frac{30(n-1)}{n}Es_i(Es_i^2)^2 \\ &\quad + \frac{10(n-1)}{n}Es_i^2Es_i^3 + \frac{50n-60}{n}Es_i^2(Es_i)^3 \\ &\quad + \frac{-10n+20}{n}(Es_i)^2Es_i^3 + \frac{-20n+24}{n}(Es_i)^5 \\ &\underset{n\to\infty}{\to} -30Es_i(Es_i^2)^2 + 10Es_i^2Es_i^3 + 50Es_i^2(Es_i)^3 \\ &\quad -10(Es_i)^2Es_i^3 - 20(Es_i)^5 = \lim_{n\to\infty} n^{\alpha(5)} E\Pi^5_{\mu=1}\Psi_{n,j_\mu}. \end{aligned} \tag{A.51}$$

$$\begin{aligned} n^{\alpha(6)} E\Pi^6_{\mu=1}\Psi_{n,j_\mu} &= \frac{1}{n^2}Es_i^6 - \frac{6}{n^2}Es_i^5Es_i + \frac{15(n-1)}{n^2}Es_i^4Es_i^2 \\ &\quad + \frac{10(n-1)}{n^2}(Es_i^3)^2 - \frac{15n-30}{n^2}Es_i^4(Es_i)^2 \\ &\quad - \frac{120(n-1)}{n^2}Es_i^3Es_i^2Es_i + \frac{15(n-1)(n-2)}{n^2}(Es_i^2)^3 \\ &\quad + \frac{100n-120}{n^2}Es_i^3(Es_i)^3 - \frac{45(n^2-7n+6)}{n^2}(Es_i^2)^2(Es_i)^2 \\ &\quad + \frac{15(3n^2-26n+24)}{n^2}Es_i^2(Es_i)^4 - \frac{5(3n^2-26n+24)}{n^2}(Es_i)^6 \\ &\underset{n\to\infty}{\to} 15(Es_i^2)^3 - 45(Es_i)^2(Es_i^2)^2 + 45(Es_i)^4Es_i^2 - 15(Es_i)^6 \\ &= \lim_{n\to\infty} n^{\alpha(6)} E\Pi^6_{\mu=1}\Psi_{n,j_\mu}. \end{aligned} \tag{A.52}$$

In order for all the quantities to be well defined, the most restrictive case is the existence of Es_i^6. For EL or ET, $s_i = \alpha_0 \cdot \rho_j^2(\lambda_0' g_{i0}) \cdot g_0^{k_0} \Pi_{l=1}^{d+1} G_0^{(l)k_l}$, $1 \leq j \leq d+1$, where α_0 denotes products of components of β_0. Since $\rho_j(\nu) = (\partial^j)(\partial \nu^j) \log(1-\nu)$, $1 \leq j \leq d+1$ for EL, Es_i^6 exists under Assumptions 2-3. In particular, UBC (3.7) ensures that $E|\rho_j(\lambda_0' g_{i0})|^{k_\rho} < \infty$ for any finite k_ρ and for $j = 1, ..., d+1$. For ET, $\rho_j(\nu) = -e^\nu$ for $1 \leq j \leq d+1$. Thus, $s_i = \alpha_0 \cdot e^{2\lambda_0 g_{i0}} \cdot g_0^{k_0} \Pi_{l=1}^{d+1} G_0^{(l)k_l}$, for $1 \leq j \leq d+1$. This case is not trivial. By Hölder's inequality, we need $q_2 > 12$ for Es_i^6 to exist. Note that the values of k_0 and k_l's do not matter as long as they are finite. Since ETEL has the same term $e^{2\lambda_0 g_{i,0}}$ in s_i with ET, except for different values for k_0 and k_l for $l = 1, ..., d+3$, $q_2 > 12$ is also needed for Es_i^6 to exist. For arbitrary $0 \leq a \leq 2$, we use the fact that $\max\{m\} = 2a+2$ to show $q_2 > 4(a+1)$.

Next we show the present lemma (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 $s_i^*(\beta)$ be the least favorable term in $f(X_i^*, \beta)$ and $s_n^*(\beta) = n^{-1} \sum_{i=1}^n s_i^*(\beta)$. 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 other values of m. Since $n^{\alpha(2)} = 1$,

$$n^{\alpha(2)} E^* \Pi_{\mu=1}^2 \Psi_{n,j_\mu}^* = E^* \hat{s}_i^{*2} - (E^* \hat{s}_i^*)^2 = n^{-1} \sum_{i=1}^n \hat{s}_i^2 - \left(n^{-1} \sum_{i=1}^n \hat{s}_i \right)^2.$$

Since $\lim_{n\to\infty} n^{\alpha(2)} E \Pi_{\mu=1}^2 \Psi_{n,j_\mu} = Es_i^2 - (Es_i)^2$, combining the following results proves the lemma for $m = 2$:

$$P\left(\left\| n^{-1} \sum_{i=1}^n \hat{u}_i - n^{-1} \sum_{i=1}^n u_i \right\| > n^{-\xi} \right) = o(n^{-a}), \tag{A.53}$$

$$P\left(\left\| n^{-1} \sum_{i=1}^n u_i - Eu_i \right\| > n^{-\xi} \right) = o(n^{-a}), \tag{A.54}$$

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 $\|\hat{s}_i^2 - s_i^2\| \leq \|\hat{s}_i - s_i\|(\|\hat{s}_i - s_i\| + 2s_i)$, (A.44), AL1(b), and Lemma 3 to show (A.53). The second result is shown by AL1(a) with $c = \xi$ and $h(X_i) = s_i^2 - Es_i^2$ or $h(X_i) = s_i - Es_i$. By considering the most restrictive form of s_i and combining the conditions of the lemmas, we need $q_2 > \max\left\{8, 8a(1-2\xi)^{-1}\right\}$ by Hölder's inequality. For $m = 3, 4, 5, 6$, we can show similar results with (A.53) and (A.54) for $u_i = s_i^m$ by using the binomial expansion, AL1, Lemma 3, and (A.44). Again, the most restrictive condition arises when we apply AL1(a) with $c = \xi$ and $h(X_i) = s_i^6 - Es_i^6$, and we need $q_2 > \max\left\{24, 24a(1-2\xi)^{-1}\right\}$ by Hölder's inequality. For arbitrary $a \geq 0$, we use the fact that $\max\{m\} = 2a + 2$ to have

$q_2 > \max\left\{8(a+1), 8a(a+1)(1-2\xi)^{-1}\right\}$ and this is assumed in the lemma. *Q.E.D.*

Lemma 9. *(a) Suppose Assumptions 1-4 hold with $q_1 \geq 2$, $q_1 > 2a$, and $q_2 > \max\left\{4(a+1), \frac{2ad}{d-2a-1}\right\}$ and $d \geq 2a+2$ for some $a \geq 0$, where $2a$ is a positive integer. Then,*

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

(b) Suppose Assumptions 1-4 hold with $q_1 \geq 2$, $q_1 > 4a$, and $q_2 > \max\left\{8(a+1), 8a(a+1), \frac{4ad}{d-2a-1}\right\}$ and $d \geq 2a+2$ for some $a \geq 0$, where $2a$ is a positive integer. Then,

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

Proof. The proof is analogous to that of Lemma 16 of A2002. We use our Lemma 7 instead of his Lemma 13. The coefficients ν_a are well defined by Lemma 8(a). Lemma 8(b) with $\xi = 0$ ensures that the coefficients $\nu^*_{n,a}$ are well behaved. *Q.E.D.*

A.3 Proof of Theorem 1

Proof. For part (a), let $a = 1$. To satisfy the conditions of the lemmas, we need $d = 4$, $q_1 > 4$, and $q_2 > 16$. We first show

$$P\left(\sup_{z\in\mathbf{R}} |P(T_{MR} \leq z) - P^*(T^*_{MR} \leq z)| > n^{-(1/2+\xi)}\varepsilon\right) = o(n^{-1}). \tag{A.55}$$

By the triangle inequality,

$$\begin{aligned} & P\left(\sup_{z\in\mathbf{R}} |P(T_{MR} \leq z) - P^*(T^*_{MR} \leq z)| > n^{-(1/2+\xi)}\varepsilon\right) \qquad \text{(A.56)} \\ \leq\ & P\left(\sup_{z\in\mathbf{R}} \left| P(T_{MR} \leq z) - \left(1 + \sum_{i=1}^{2} n^{-i/2}\pi_i(\delta, \nu_1)\right) \Phi(z) \right| > n^{-(1/2+\xi)}\frac{\varepsilon}{4}\right) \\ & + P\left(\sup_{z\in\mathbf{R}} \left| P^*(T^*_{MR} \leq z) - \left(1 + \sum_{i=1}^{2} n^{-i/2}\pi_i(\delta, \nu^*_{n,1})\right) \Phi(z) \right| > n^{-(1/2+\xi)}\frac{\varepsilon}{4}\right) \\ & + P\left(\sup_{z\in\mathbf{R}} n^{-1/2} \left|\pi_1(\delta, \nu_1) - \pi_1(\delta, \nu^*_{n,1})\right| \Phi(z) > n^{-(1/2+\xi)}\frac{\varepsilon}{4}\right) \\ & + P\left(\sup_{z\in\mathbf{R}} n^{-1} \left|\pi_2(\delta, \nu_1) - \pi_2(\delta, \nu^*_{n,1})\right| \Phi(z) > n^{-(1/2+\xi)}\frac{\varepsilon}{4}\right) = o(n^{-1}). \end{aligned}$$

The last equality holds by Lemma 9(a)-(b) and Lemma 8(b). The rest of the proof follows the same argument with (5.32)-(5.34) in the proof of Theorem 2 of Andrews (2001). This establishes the first result of the present theorem (a). The second and the third result can be proved analogously.

For part (b), let $a = 3/2$. Then, we need $d = 5$, $q_1 > 6$, and $q_2 > 30$. We use the evenness of $\pi_i(\delta, \nu_{3/2})\Phi(z)$ and $\pi_i(\delta, \nu^*_{n,3/2})\Phi(z)$ for $i = 1, 3$ to cancel out these terms through $\Phi(z) - \Phi(-z)$. The rest follows analogously.

For part (c), let $a = 2$. Then, we need $d = 6$, $q_1 > 8$, and $q_2 > 48$. The proof is the same with that of Theorem 2(c) of A2002 with his Lemmas 13, 14, and 16 replaced by our Lemmas 7, 8, and 9. The proof relies on the argument of Hall (1988, 1992)'s methods developed for "smooth functions of sample averages," for iid data. *Q.E.D.*

DGP C-1					$n = 100$			$n = 200$		
					CI		J test	CI		J test
					.90	.95	.05	.90	.95	.05
T=4	Boot	GMM	C	HH	.925	.968	.006	.911	.960	.021
		GMM	MR	L	.939	.979	n/a	.924	.969	n/a
		EL	MR	L	.921	.977	n/a	.918	.970	n/a
		EL	MR	BNS	.899	.963		.901	.954	
		ET	MR	L	.922	.976	n/a	.918	.971	n/a
		ET	MR	BNS	.896	.960		.900	.957	
		ETEL	MR	L	.916	.972	n/a	.909	.969	n/a
		ETEL	MR	BNS	.902	.961		.904	.954	
	Asymp	GMM	MR		.781	.843	.037	.830	.888	.039
		GMM	C		.775	.839		.823	.889	
		EL	MR		.742	.812	.119	.812	.871	.081
		EL	C		.730	.807		.795	.867	
		ET	MR		.753	.829	.097	.823	.881	.076
		ET	C		.732	.809		.796	.868	
		ETEL	MR		.745	.817	.165	.813	.874	.108
		ETEL	C		.741	.817		.800	.869	
T=6	Boot	GMM	C	HH	.961	.989	.000	.932	.975	.002
		GMM	MR	L	.981	.994	n/a	.950	.987	n/a
		EL	MR	L	.970	.993	n/a	.934	.974	n/a
		EL	MR	BNS	.933	.973		.919	.960	
		ET	MR	L	.961	.990	n/a	.925	.973	n/a
		ET	MR	BNS	.931	.976		.924	.968	
		ETEL	MR	L	.959	.992	n/a	.926	.973	n/a
		ETEL	MR	BNS	.934	.975		.912	.957	
	Asymp	GMM	MR		.656	.740	.031	.760	.836	.045
		GMM	C		.643	.726		.759	.836	
		EL	MR		.693	.762	.419	.784	.862	.257
		EL	C		.654	.736		.748	.828	
		ET	MR		.732	.800	.325	.808	.878	.210
		ET	C		.656	.740		.761	.841	
		ETEL	MR		.710	.777	.557	.794	.871	.356
		ETEL	C		.664	.743		.754	.839	

Table 1: Coverage Probabilities of 90% and 95% Confidence Intervals for ρ_0 based on GMM, EL, ET, and ETEL under DGP C-1. Number of Monte Carlo repetition $r = 5,000$. The Warp-Speed Monte Carlo method is used.

DGP C-2					$n=100$			$n=200$		
					CI		J test	CI		J test
					.90	.95	.05	.90	.95	.05
T=4	Boot	GMM	C	HH	.907	.957	.033	.898	.943	.044
		GMM	MR	L	.927	.968	n/a	.908	.962	n/a
		EL	MR	L	.908	.957	n/a	.900	.957	n/a
		EL	MR	BNS	.883	.932		.879	.941	
		ET	MR	L	.908	.956	n/a	.896	.953	n/a
		ET	MR	BNS	.895	.939		.889	.942	
		ETEL	MR	L	.907	.953	n/a	.898	.954	n/a
		ETEL	MR	BNS	.892	.941		.883	.941	
	Asymp	GMM	MR		.798	.867	.050	.847	.900	.051
		GMM	C		.795	.860		.846	.901	
		EL	MR		.795	.854	.092	.840	.894	.067
		EL	C		.783	.847		.833	.891	
		ET	MR		.798	.858	.090	.842	.896	.067
		ET	C		.781	.849		.831	.889	
		ETEL	MR		.797	.859	.117	.842	.895	.081
		ETEL	C		.787	.853		.835	.892	
T=6	Boot	GMM	C	HH	.921	.969	.006	.913	.956	.027
		GMM	MR	L	.957	.987	n/a	.940	.977	n/a
		EL	MR	L	.959	.991	n/a	.929	.972	n/a
		EL	MR	BNS	.925	.969		.918	.963	
		ET	MR	L	.945	.984	n/a	.921	.967	n/a
		ET	MR	BNS	.919	.963		.901	.954	
		ETEL	MR	L	.951	.987	n/a	.922	.968	n/a
		ETEL	MR	BNS	.927	.972		.909	.961	
	Asymp	GMM	MR		.709	.783	.053	.804	.876	.056
		GMM	C		.717	.797		.806	.881	
		EL	MR		.747	.817	.284	.838	.900	.135
		EL	C		.731	.809		.823	.891	
		ET	MR		.777	.846	.257	.848	.909	.138
		ET	C		.737	.814		.821	.891	
		ETEL	MR		.756	.829	.391	.846	.904	.196
		ETEL	C		.737	.814		.825	.894	

Table 2: Coverage Probabilities of 90% and 95% Confidence Intervals for ρ_0 based on GMM, EL, ET, and ETEL under DGP C-2. Number of Monte Carlo repetition $r = 5,000$. The Warp-Speed Monte Carlo method is used.

DGP M-1					$n = 100$			$n = 200$		
					CI		J test	CI		J test
					.90	.95	.05	.90	.95	.05
T=4	Boot	GMM	C	HH	.839	.931	.002	.889	.952	.036
		GMM	MR	L	.919	.970	n/a	.949	.982	n/a
		EL	MR	L	.819	.899	n/a	.871	.938	n/a
		EL	MR	BNS	.776	.854		.824	.894	
		ET	MR	L	.819	.899	n/a	.873	.942	n/a
		ET	MR	BNS	.771	.851		.824	.898	
		ETEL	MR	L	.820	.902	n/a	.872	.935	n/a
		ETEL	MR	BNS	.779	.859		.826	.895	
	Asymp	GMM	MR		.511	.564	.172	.642	.697	.277
		GMM	C		.422	.474		.551	.616	
		EL	MR		.585	.648	.251	.697	.760	.301
		EL	C		.558	.625		.636	.701	
		ET	MR		.588	.654	.233	.707	.768	.308
		ET	C		.549	.620		.632	.700	
		ETEL	MR		.596	.660	.312	.713	.776	.362
		ETEL	C		.571	.638		.654	.719	
T=6	Boot	GMM	C	HH	.920	.967	.000	.943	.983	.011
		GMM	MR	L	.971	.993	n/a	.987	.995	n/a
		EL	MR	L	.947	.977	n/a	.918	.969	n/a
		EL	MR	BNS	.849	.926		.852	.917	
		ET	MR	L	.935	.974	n/a	.926	.970	n/a
		ET	MR	BNS	.888	.941		.878	.936	
		ETEL	MR	L	.931	.970	n/a	.914	.959	n/a
		ETEL	MR	BNS	.872	.933		.868	.933	
	Asymp	GMM	MR		.436	.489	.263	.592	.662	.586
		GMM	C		.344	.398		.500	.572	
		EL	MR		.583	.649	.800	.688	.761	.882
		EL	C		.490	.558		.546	.624	
		ET	MR		.634	.697	.734	.739	.814	.857
		ET	C		.482	.560		.562	.646	
		ETEL	MR		.603	.673	.885	.706	.779	.927
		ETEL	C		.482	.552		.555	.631	

Table 3: Coverage Probabilities of 90% and 95% Confidence Intervals for ρ_0 based on GMM, EL, ET, and ETEL under DGP M-1. Number of Monte Carlo repetition $r = 5,000$. The Warp-Speed Monte Carlo method is used.

DGP M-2					$n = 100$			$n = 200$		
					CI		J test	CI		J test
					.90	.95	.05	.90	.95	.05
T=4	Boot	GMM	C	HH	.858	.933	.036	.887	.943	.121
		GMM	MR	L	.917	.969	n/a	.944	.975	n/a
		EL	MR	L	.880	.929	n/a	.911	.956	n/a
		EL	MR	BNS	.861	.909		.898	.954	
		ET	MR	L	.878	.930	n/a	.909	.953	n/a
		ET	MR	BNS	.858	.905		.899	.946	
		ETEL	MR	L	.880	.929	n/a	.906	.957	n/a
		ETEL	MR	BNS	.861	.912		. 906	.956	
	Asymp	GMM	MR		.611	.656	.198	.721	.770	.304
		GMM	C		.525	.582		.653	.707	
		EL	MR		.775	.823	.194	.849	.901	.259
		EL	C		.766	.812		.844	.891	
		ET	MR		.771	.818	.199	.849	.899	.276
		ET	C		.752	.801		.837	.884	
		ETEL	MR		.776	.824	.224	.849	.901	.284
		ETEL	C		.766	.813		.845	.889	
T=6	Boot	GMM	C	HH	.921	.964	.012	.908	.966	.313
		GMM	MR	L	.970	.992	n/a	.982	.995	n/a
		EL	MR	L	.962	.983	n/a	.930	.971	n/a
		EL	MR	BNS	.911	.956		.904	.953	
		ET	MR	L	.954	.979	n/a	.924	.973	n/a
		ET	MR	BNS	.919	.958		.907	.959	
		ETEL	MR	L	.959	.982	n/a	.924	.966	n/a
		ETEL	MR	BNS	.919	.957		.908	.961	
	Asymp	GMM	MR		.549	.617	.277	.692	.766	.619
		GMM	C		.454	.516		.597	.671	
		EL	MR		.778	.839	.555	.856	.913	.727
		EL	C		.717	.791		.794	.865	
		ET	MR		.802	.859	.550	.872	.926	.739
		ET	C		.703	.778		.785	.857	
		ETEL	MR		.792	.851	.649	.863	.922	.784
		ETEL	C		.723	.796		.800	.873	

Table 4: Coverage Probabilities of 90% and 95% Confidence Intervals for ρ_0 based on GMM, EL, ET, and ETEL under DGP M-2. Number of Monte Carlo repetition $r = 5,000$. The Warp-Speed Monte Carlo method is used.

DGP				$T=4$				$T=6$			
				$n=100$		$n=200$		$n=100$		$n=200$	
				.90	.95	.90	.95	.90	.95	.90	.95
C-1	GMM	C	HH	.489	.619	.325	.396	.321	.407	.194	.235
	GMM	MR	L	.546	.703	.349	.435	.389	.495	.212	.266
	EL	MR	L	.595	.807	.397	.513	.439	.582	.212	.269
	EL	MR	BNS	.553	.724	.376	.466	.354	.452	.199	.244
	ET	MR	L	.599	.805	.391	.499	.411	.546	.204	.258
	ET	MR	BNS	.546	.711	.364	.456	.349	.462	.203	.248
	ETEL	MR	L	.589	.781	.379	.494	.410	.551	.204	.266
	ETEL	MR	BNS	.555	.725	.373	.459	.353	.466	.193	.241
C-2	GMM	C	HH	.531	.663	.348	.414	.326	.408	.205	.248
	GMM	MR	L	.584	.730	.362	.456	.395	.499	.232	.280
	EL	MR	L	.600	.776	.382	.475	.427	.576	.219	.269
	EL	MR	BNS	.544	.676	.361	.446	.358	.452	.211	.255
	ET	MR	L	.612	.780	.380	.471	.392	.520	.213	.261
	ET	MR	BNS	.579	.717	.373	.449	.352	.440	.201	.242
	ETEL	MR	L	.596	.757	.379	.468	.403	.537	.214	.260
	ETEL	MR	BNS	.563	.702	.364	.444	.360	.459	.206	.251
M-1	GMM	C	HH	1.157	1.566	.935	1.363	.798	1.101	.491	.723
	GMM	MR	L	2.449	3.533	1.589	2.268	1.544	2.177	.924	1.292
	EL	MR	L	.925	1.340	.707	1.062	.807	1.224	.443	.618
	EL	MR	BNS	.779	1.076	.582	.804	.500	.702	.335	.439
	ET	MR	L	.953	1.382	.742	1.105	.793	1.201	.422	.569
	ET	MR	BNS	.803	1.096	.624	.850	.603	.826	.350	.447
	ETEL	MR	L	.921	1.351	.670	.987	.756	1.092	.420	.559
	ETEL	MR	BNS	.776	1.087	.555	.745	.564	.765	.359	.462
M-2	GMM	C	HH	1.230	1.742	.782	1.111	.705	.974	.371	.512
	GMM	MR	L	2.132	3.093	1.328	1.845	1.342	1.864	.707	1.004
	EL	MR	L	.711	1.041	.415	.539	.585	.847	.247	.306
	EL	MR	BNS	.652	.867	.395	.527	.413	.546	.227	.277
	ET	MR	L	.739	1.119	.436	.562	.579	.827	.249	.317
	ET	MR	BNS	.666	.890	.415	.532	.458	.596	.234	.289
	ETEL	MR	L	.695	1.019	.412	.541	.552	.789	.237	.291
	ETEL	MR	BNS	.640	.874	.413	.538	.420	.542	.225	.279

Table 5: Width of 90% and 95% Bootstrap Confidence Intervals for ρ_0 based on GMM, EL, ET, and ETEL. Number of Monte Carlo repetition $r = 5,000$. The Warp-Speed Monte Carlo method is used.

		OLS	GMM	EL	ET	ETEL
	$\hat{\beta}$	.294	-.561	.016	-.059	-.023
const	s.e.$_C$	(.235)	(.089)	(.097)	(.101)	(.100)
	s.e.$_{MR}$		(.194)	(.109)	(.125)	(.121)
	$\hat{\beta}$	.054	.056	.068	.070	.071
educ	s.e.$_C$	(.010)	(.006)	(.005)	(.006)	(.006)
	s.e.$_{MR}$		(.018)	(.006)	(.009)	(.008)
	$\hat{\beta}$	.068	.140	.076	.081	.082
exper	s.e.$_C$	(.025)	(.006)	(.007)	(.007)	(.007)
	s.e.$_{MR}$		(.022)	(.008)	(.011)	(.010)
	$\hat{\beta}$	-.002	-.004	-.002	-.002	-.002
exper^2	s.e.$_C$	(.001)	(.0002)	(.0002)	(.0002)	(.0002)
	s.e.$_{MR}$		(.0006)	(.0002)	(.0003)	(.0002)
	$\hat{\beta}$	.004	.007	.005	.006	.005
IQ	s.e.$_C$	(.001)	(.001)	(.001)	(.001)	(.001)
	s.e.$_{MR}$		(.002)	(.001)	(.002)	(.002)
	$\hat{\beta}$	.008	-.0003	-.002	-.004	-.005
KWW	s.e.$_C$	(.003)	(.003)	(.003)	(.003)	(.003)
	s.e.$_{MR}$		(.007)	(.003)	(.004)	(.004)
J test	χ^2_{13}		477.3	177.5	285.2	196.2
	p-value		[.000]	[.000]	[.000]	[.000]

Table 6: Estimation of the Mincer equation using Census moments

Estimator	CI	s.e.		LB	Point Est.	UB	Width
OLS	Asymp	n/a		.033	.054	.074	.041
GMM	Asymp	C		.044	.056	.068	.024
	Asymp	MR		.021		.091	.070
	Boot (sym)	MR	L	.003		.108	.105
	Boot (eqt)	MR	L	.019		.115	.096
EL	Asymp	C		.058	.068	.079	.021
	Asymp	MR		.056		.080	.024
	Boot (sym)	MR	L	.041		.096	.055
	Boot (sym)	MR	BNS	.054		.083	.029
	Boot (eqt)	MR	L	.049		.099	.050
	Boot (eqt)	MR	BNS	.055		.085	.030
ET	Asymp	C		.058	.070	.081	.023
	Asymp	MR		.052		.087	.035
	Boot (sym)	MR	L	.035		.104	.069
	Boot (sym)	MR	BNS	.047		.092	.045
	Boots (eqt)	MR	L	.047		.110	.063
	Boots (eqt)	MR	BNS	.047		.093	.046
ETEL	Asymp	C		.060	.071	.083	.023
	Asymp	MR		.056		.086	.030
	Boot (sym)	MR	L	.039		.104	.066
	Boot (sym)	MR	BNS	.052		.090	.038
	Boot (eqt)	MR	L	.051		.108	.057
	Boot (eqt)	MR	BNS	.053		.093	.040

Table 7: 95% Confidence Intervals for the Returns to Schooling. Number of Bootstrap Repetition $B = 5,000$.