Higher Order Properties of Bootstrap and Jackknife Bias Corrected Maximum Likelihood Estimators

Jinyong Hahn[¤] Guido Kuersteiner[y] Whitney Newey[z]
Brown University MIT MIT

November 7, 2002

Abstract

The purpose of this paper is to consider the third-order asymptotic properties of bias corrected ML. We show third-order e¢ciency of bias corrected maximum likelihood (ML) with a bias correction based on sample averages of certain functions of likelihood derivatives, or on the bootstrap, or on the jacknife. We give an explanation of these results suggesting that any bias correced ML satisfying certain regularity conditions should be third-order e¢cient, i.e. that the form of the bias correction has no e¤ect on the higher (third) order variance for ML. We also …nd a stronger equivalence property for the bootstrap and jacknife bias corrected estimators, that they have the same stochastic expansion to third-order.

1 Introduction

Asymptotic bias corrections provide useful methods for centering estimators nearer the truth. These methods include analytical corrections such as the standard textbook expansion for functions of sample means and the more complicated formulas required for a general maximum likelihood (ML) estimator. They also include the jacknife and bootstrap methods. The purpose of this paper is to consider the third-order asymptotic properties of bias corrected ML. We show third-order efficiency with a bias correction based on sample averages of certain functions of likelihood derivatives, or on the bootstrap, or on the jacknife. We give an explanation of these results suggesting that any bias correced ML satisfying certain regularity conditions should be third-order efficient, i.e. that the form of the bias correction has no effect on the higher (third) order variance for ML. We also find a stronger equivalence property for the bootstrap and jacknife bias corrected estimators, that they have the same stochastic expansion to third-order.

Pfanzagl and Wefelmeyer (1978) had previously shown that the bias-corrected ML is third-order efficient, when the bias correction is based on integrals over the parametric density. Our results show that the expectations in the bias correction formula can be replaced by sample averages without affecting third-order efficiency, which simplifies computation.

The Jackknife bias estimator goes back to Quenouille (1949). Bootstrap bias estimation was discussed by Parr (1983), Shao (1988a,b), Hall (1992), and Horowitz (1998) in the context of nonlinear transformations of OLS estimators of linear models and nonlinear functions of the mean. Akahira (1983) considered second-order properties of the jacknife and bootstrap. We extend the literature on the bootstrap and jackknife bias corrected estimator in two directions. First, we analyze genuinely nonlinear estimators rather than nonlinear transformations of linear estimators as in Shao. Secondly, the literature on bootstrap bias corrected estimators has been focused on analyzing bias properties without investigating the effects of bias correction on the higher order variance. We are instead working with third rather than second order expansions of the bias corrected estimators. This allows us to analyze the effect bias correction has on the higher order variance of the estimator.

In Section 2 we derive the third order stochastic expansion of the bootstrap and jacknife bias corrected ML, showing that they are identical. In Section 3 we consider third-order efficiency of the estimators. Section 4 concludes.

jackknife bias corrected MLE, and argue that they are higher order equivalent. We argue that such bias corrected estimators should have the same higher order variance as the bias corrected MLE developed by Pfanzagl and Wefelmeyer (1978), which was shown to be third order optimal.

2.1 Higher Order Expansion of MLE

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Consider a standard parametric model where $\{Z_i\}_{i=1}^n$ is an iid sample $Z_i \sim f(z, \theta_0)$, such that $f(z, \theta)$ satisfies sufficient smoothness conditions summarized below in Condition 1. The density $f(z, \theta)$ is a member of a parametric family of distributions P_θ indexed by $\theta \in \Theta$ with $\Theta \subset \mathbb{R}$ a compact set. We consider properties of the MLE $\hat{\theta}$ where

$$\hat{\theta} \equiv \sup_{\theta \in \Theta} n^{-1} \sum_{i=1}^{n} \log f(Z_i, \theta).$$

It is convenient to understand $\hat{\theta} \equiv \hat{\theta}\left(\frac{1}{\sqrt{n}}\right)$, where $\theta(\epsilon)$ denotes the solution

$$\hat{\theta}(\epsilon) = \sup_{\theta \in \Theta} \int \log f(\cdot, \theta)\, dF_\epsilon(z).$$

Here,

$$F_\epsilon \equiv F + \epsilon \Delta \equiv F + \epsilon \sqrt{n}\left(\hat{F} - F\right), \qquad \epsilon \in \left[0, \frac{1}{\sqrt{n}}\right]$$

and F and $\hat{F}$ denote the underlying cumulative distribution function and the empirical distribution function $\hat{F}(z) \equiv n^{-1} \sum_{i=1}^{n} 1\{Z_i \le z\}$.

We obtain bootstrapped estimates $\hat{\theta}^*$ by sampling $Z_1^*, \ldots, Z_n^*$ identically and independently from the empirical distribution $\hat{F}$. We denote the empirical distribution of $Z_1^*, \ldots, Z_n^*$ by $\hat{F}^*(z) = n^{-1} \sum_{i=1}^{n} 1\{Z_i^* \le z\}$. Using previous notation it therefore follows that $\hat{\theta}^* \equiv \hat{\theta}^*\left(\frac{1}{\sqrt{n}}\right)$ is the solution

$$\hat{\theta}^*(\epsilon) = \sup_{\theta \in \Theta} \int \log f(\cdot, \theta)\, d\hat{F}_\epsilon(z),$$

where

$$\hat{F}_\epsilon \equiv \hat{F} + \epsilon \hat{\Delta} \equiv \hat{F} + \epsilon \sqrt{n}\left(\hat{F}^* - \hat{F}\right), \qquad \epsilon \in \left[0, \frac{1}{\sqrt{n}}\right].$$

Here, $\hat{\Delta}$ is the bootstrap empirical process $\hat{\Delta} \equiv \sqrt{n}\left(\hat{F}^* - \hat{F}\right)$. We are imposing the following technical conditions to guarantee the validity of our stochastic expansions.

Condition 2 For each $\theta \in \Theta$ and for $m \leq 7$ let $\partial^m \log f(z,\theta)/\partial\theta^m$ be a P-measurable function of z.

Condition 3 Let $\mathcal{F}$ be the class of functions $\partial^m \log f(z,\theta)/\partial\theta^m$ indexed by $\theta \in \Theta$ for $m = 1,..,7$ with envelope $M(z)$. Then,

$$\int_0^1 \sup_{Q \in \mathcal{P}} \sqrt{\log N\left(\varepsilon\left(\int M^2 dQ\right)^{1/2}, \mathcal{F}, L_2(Q)\right)}\, d\varepsilon < \infty, \tag{1}$$

where $\mathcal{P}$ is the class of probability measures on $\mathbb{R}$ that concentrate on a finite set and N is the cover number defined in van der Vaart and Wellner (1996, p.90).

Condition 1 is a standard condition guaranteeing identification of the model and imposing sufficient smoothness conditions as well as existence of higher moments to allow for a higher order stochastic expansion of the estimator. Condition 2 together with separability of the parameter space guarantees measurability of suprema of our empirical processes. As is well known from the probability literature, measurability conditions could be relaxed somewhat at the expense of more refined convergence arguments. We are abstracting from such refinements for the purpose of this paper.

From Giné and Zinn (1990, Theorem 2.4) and Conditions 1,2 and 3 it follows that, almost surely, $n^{1/2}\left(\hat{F}^* - \hat{F}\right) \to T$ weakly in $l^\infty(\mathcal{F})$ where T is a Brownian Bridge Process. We use the result on the convergence of the empirical processes to obtain an expansion of the estimators $\hat{\theta}$ and $\hat{\theta}^*$.

Let $\ell(\cdot,\theta) \equiv \partial \log f(\cdot,\theta)/\partial\theta$, $\ell^\theta(\cdot,\theta) \equiv \partial^2 \log f(\cdot,\theta)\big/\partial\theta^2$, $\ell^{\theta\theta}(\cdot,\theta) \equiv \partial^3 \log f(\cdot,\theta)\big/\partial\theta^3$, etc. Define $\mathcal{I} \equiv -E\left[\ell^\theta(Z_i,\theta_0)\right]$, $Q_1(\theta) \equiv E\left[\ell^{\theta\theta}(Z_i,\theta)\right]$ and $Q_2(\theta) \equiv E\left[\ell^{\theta\theta\theta}(Z_i,\theta)\right]$. It is convenient to express the resulting expansion in terms of U and V-statistics. We define $U_i(\theta) \equiv \ell(Z_i,\theta)$, $V_i(\theta) \equiv \ell^\theta(Z_i,\theta) - E\left[\ell^\theta(Z_i,\theta)\right]$, $W_i \equiv \ell^{\theta\theta}(Z_i) - E\left[\ell^{\theta\theta}(Z_i)\right]$ and let $U(\theta) \equiv n^{-1/2}\sum_{i=1}^n U_i(\theta)$, $V(\theta) \equiv n^{-1/2}\sum_{i=1}^n V_i(\theta)$, and $W(\theta) \equiv n^{-1/2}\sum_{i=1}^n W_i(\theta)$. We obtain the following formal expansion of the ML estimator. Validity of these expansions was established under additional conditions for example by Gusev (1975, 1976).

Proposition 1 Under Condition 1, there exists some $\bar{\epsilon} \in \left[0, \frac{1}{\sqrt{n}}\right]$ such that with probability tending to one, $\hat{\theta}$ satisfies the expansion

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = \theta^{\epsilon}(0) + \frac{1}{2}\frac{1}{\sqrt{n}}\theta^{\epsilon\epsilon}(0) + \frac{1}{6}\frac{1}{n}\theta^{\epsilon\epsilon\epsilon}(0) \tag{2}$$

$$+\frac{1}{24}\frac{1}{n\sqrt{n}}\theta^{\epsilon\epsilon\epsilon\epsilon}(0) + \frac{1}{120}\frac{1}{n^2}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon}(0) + \frac{1}{720}\frac{1}{n^2\sqrt{n}}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}(\bar{\epsilon}) \tag{3}$$

and

$$\theta^{\epsilon\epsilon\epsilon}(0) = I^{-4}Q_2(\theta_0)U(\theta_0)^3 + 3I^{-5}Q_1(\theta_0)^2U(\theta_0)^3 + 9I^{-4}Q_1(\theta)U(\theta_0)^2V(\theta_0) \qquad (6)$$
$$+3I^{-3}U(\theta_0)^2W(\theta_0) + 6I^{-3}U(\theta_0)V(\theta_0)^2.$$

Moreover, $\theta^{\epsilon}(0) = O_p(1)$, $\theta^{\epsilon\epsilon}(0) = O_p(1)$, $\theta^{\epsilon\epsilon\epsilon}(0) = O_p(1)$ and $\max_{\epsilon\in\left[0,\frac{1}{\sqrt{n}}\right]}\theta^{\epsilon\epsilon\epsilon\epsilon}(\epsilon) = O_p(1)$. Finally, let $\tilde{\theta} = \theta_0 + \frac{1}{\sqrt{n}}\theta^{\epsilon}(0) + \frac{1}{2}\frac{1}{n}\theta^{\epsilon\epsilon}(0) + \frac{1}{6}\frac{1}{n^{3/2}}\theta^{\epsilon\epsilon\epsilon}(0)$ such that

$$E_{\theta_0}\left[\left(\sqrt{n}\left(\tilde{\theta}-\theta_0\right)\right)^2\right] = \frac{1}{I} + \frac{\upsilon(\theta_0)}{n} + \frac{b(\theta_0)^2}{n} + o(n^{-1}).$$

where

$$b(\theta_0) \equiv \frac{1}{2}E_{\theta_0}[\theta^{\epsilon\epsilon}] = \frac{1}{2I^2}E_{\theta_0}\left[\ell^{\theta\theta}\right] + \frac{1}{I^2}E_{\theta_0}\left[\ell\ell^{\theta}\right]$$

and

$$\upsilon(\theta_0) \equiv \frac{1}{4}\mathrm{Var}_{\theta_0}(\theta^{\epsilon\epsilon}) + \frac{1}{3}E_{\theta_0}[\theta^{\epsilon\epsilon\epsilon}\theta^{\epsilon}] + n^{1/2}E_{\theta_0}[\theta^{\epsilon\epsilon}\theta^{\epsilon}]$$

Proof. See Appendix A.4. ■

Based on Theorem (1), we can understand $\frac{b(\theta)}{n}$ as the higher order bias of $\hat{\theta}$. Likewise, we can understand $\frac{1}{I} + \frac{\upsilon}{n}$ as the higher order variance of $\hat{\theta}$.

In order to approximate the bias of the bootstrapped estimate $\hat{\theta}^*$ we need a similar higher order expansion as in the case of the ML estimator. Here, however, the reference point around which we develop our approximation is the empirical distribution $\hat{F}$ rather than the original distribution F. The convergence of $\hat{F}$ to F then guarantees that bootstrapped statistics are close to the original statistics.

We replace I, Q_1 and Q_2 with $\hat{I} = -n^{-1}\sum_{i=1}^{n}\ell^{\theta}(Z_i,\hat{\theta})$, $\hat{Q}_1 = n^{-1}\sum_{i=1}^{n}\ell^{\theta\theta}(Z_i,\hat{\theta})$ and $\hat{Q}_2 = n^{-1}\sum_{i=1}^{n}\ell^{\theta\theta\theta}(Z_i,\hat{\theta})$ and define bootstrapped U and V-statistics as $U_i^*(\theta) \equiv \ell(Z_i^*,\theta)$, $V_i^*(\theta) \equiv \ell^{\theta}(Z_i^*,\theta) - n^{-1}\sum_{i=1}^{n}\ell^{\theta}(Z_i,\theta)$, $W_i^* \equiv \ell^{\theta\theta}(Z_i^*) - n^{-1}\sum_{i=1}^{n}\ell^{\theta\theta}(Z_i,\theta)$ and let $U^*(\theta) = n^{-1/2}\sum_{i=1}^{n}U_i^*(\theta)$, $V^*(\theta) = n^{-1/2}\sum_{i=1}^{n}V_i^*(\theta)$ and $W^*(\theta) = n^{-1/2}\sum_{i=1}^{n}W_i^*(\theta)$ we obtain for the following result for the bootstraped estimate $\hat{\theta}^*$.

Proposition 2 Under Conditions 1,2 and 3 $\exists\epsilon\in\left[0,n^{-1/2}\right]$ such that with probability tending to one $P^{\mathcal{N}}$a.s., $\hat{\theta}^*$ satisfies the expansion

$$\sqrt{n}\left(\hat{\theta}^* - \hat{\theta}\right) = \hat{\theta}^{\epsilon}(0) + \frac{1}{2}\frac{1}{\sqrt{n}}\hat{\theta}^{\epsilon\epsilon}(0) + \frac{1}{6}\frac{1}{n}\hat{\theta}^{\epsilon\epsilon\epsilon}(0)$$

2.2 Bootstrap Bias Correction

Bootstrap Bias estimation and Bias correction was analyzed in the context of linear models by Shao (1988a,b). Let E^* be the expectation operator with respect to $\hat{F}$. The idea behind the Bootstrap bias correction is to estimate $E\left[\hat\theta\right] - \theta_0$, if it exists, by $E^*\left[\hat\theta^*\right] - \hat\theta$. We show that $E^*\left[\hat\theta^*\right] - \hat\theta$ is close to $b(\theta)$. This in turn will allow us to construct the bias corrected estimate $2\hat\theta - E^*\left[\hat\theta^*\right]$.

We first establish that $b^* = E^*\left[\hat\theta^*\right] - \hat\theta$ estimates the higher order bias $b(\theta)$ consistently.

Proposition 3 Assume Conditions 1,2 and 3 hold. Then

$$b^* = \frac{b(\theta_0)}{n} + o_p\left(n^{-1}\right).$$

Proof. See Appendix A.4. ∎

While this result establishes that we can consistently estimate the higher order bias it is not sufficient to guarantee good higher order properties of the bias corrected estimator. For this reason we establish the next result.

Proposition 4 Assume Conditions 1,2 and 3 hold. Then

$$\begin{aligned}\sqrt{n}\left(\hat\theta - \left(E^*\left[\hat\theta^*\right] - \hat\theta\right) - \theta_0\right) &= \frac{1}{I}U(\theta_0) + \frac{1}{\sqrt{n}}\left(\frac{1}{2}\theta^{\epsilon\epsilon}(0) - b(\theta_0)\right) \\ &\quad + \frac{1}{6n}\theta^{\epsilon\epsilon\epsilon}(0) - \frac{1}{2n}\mathrm{B} + o_p\left(\frac{1}{n}\right),\end{aligned}$$

where B is defined in (44) in the Appendix.

Proof. See Appendix A.4. ∎

Because

$$E\left[\frac{\theta^{\epsilon\epsilon}(0)}{2} - b(\theta_0)\right] = 0,$$

we can see that the bootstrap successfully removes bias. In a similar way we can approximate the MSE. It then follows that

$$E\left[\left(\sqrt{n}\left(\tilde\theta - \left(E^*\left(\theta^*\right) - \hat\theta\right) - \theta_0\right)\right)^2\right] \approx \mathrm{Var}\left(\sqrt{n}\left(\tilde\theta - \theta_0\right)\right) - \frac{1}{2n}E\left[\mathrm{B}\theta^{\epsilon}\right].$$

2.3 Jackknife Bias Correction

The following proposition establishes the higher order properties of the Jackknife bias corrected ML estimator.

Proposition 5 Assume Condition 1 holds. Then the jackknife bias corrected ML estimator has a higher order expansion as in

$$\sqrt{n}\,(\theta_J - \theta_0) = \theta^{\epsilon} + \frac{1}{\sqrt{n}}\left(\frac{1}{2}\theta^{\epsilon\epsilon}(0) - b(\theta_0)\right) + \frac{1}{6}\frac{1}{n}\theta^{\epsilon\epsilon\epsilon} - \frac{1}{2}\frac{1}{n}J + o_p\left(\frac{1}{n}\right)$$

where J is defined in (52) in the Appendix.

Proof. See Appendix A.4. ∎

It is shown in the appendix that

$$J = B, \tag{7}$$

which means that the Jackknife and Bootstrap bias corrected versions of the ML estimator are higher order equivalent. They do not only have the same higher order variance but agree more generally in terms of their higher order distribution at least as far as the stochastic approximation allows to make such comparisons.

3 Higher Order Efficiency

In this section we obtain the higher order asymptotic properties of the bias corrected estimator of Pfanzagl and Wefelmeyer (1978). Since that estimator was shown to be higher order efficient we will conclude that our bias corrected estimator is higher order efficient under quadratic risk if the variance of the first three terms in the stochastic expansion is the same as for the Pfanzagl and Wefelmeyer estimator.

From the expansion in Proposition 2 we have

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = \theta^{\epsilon}(0) + \frac{1}{2}\frac{1}{\sqrt{n}}\theta^{\epsilon\epsilon}(0) + \frac{1}{6}\frac{1}{n}\theta^{\epsilon\epsilon\epsilon}(0) + O_p\left(\frac{1}{n\sqrt{n}}\right),$$

such that the highest order asymptotic bias of MLE is equal to where $\tau(t_1,t_2,t_3) \equiv \frac{1}{2t_1^2}t_2 + \frac{1}{t_1^2}t_3$, $t_1 \equiv \int \ell(z,\theta)^2 f(z,\theta)\,dz$, $t_2 \equiv \int \ell^{\theta\theta}(z,\theta) f(z,\theta)\,dz$, $t_3 \equiv \int \ell(z,\theta)\,\ell^{\theta}(z,\theta) f(z,\theta)\,dz$, and $m(z,\theta) \equiv \left(\ell(z,\theta)^2, \ell^{\theta\theta}(z,\theta), \ell(z,\theta)\,\ell^{\theta}(z,\theta)\right)'$. This leads to a bias corrected estimator

$$\hat\theta_c \equiv \hat\theta - \frac{b\left(\hat\theta\right)}{n}.$$

This bias correction procedure was shown to be higher order efficient by Pfanzagl and Wefelymeyer (1978). Our next result shows that as long as we restrict ourselves to quadratic loss any other regular estimator of $b(\theta)$ also leads to a higher order efficient bias corrected MLE.

Theorem 1 Assume Conditions 1,2 and 3 hold. Assume that $\sqrt{n}\left(b\left(\hat\theta\right) - b(\theta_0)\right)$ is asymptotically a nonsingular linear combination of $\sqrt{n}\left(\hat\theta - \theta_0\right)$, i.e., $\sqrt{n}\left(b\left(\hat\theta\right) - b(\theta_0)\right) = \Upsilon n^{-1/2}\sum_{i=1}^{n}\psi(Z_i,\theta_0) + o_p(1)$ for some nonsingular Υ where $\psi(Z_i,\theta_0) \equiv I^{-1}\ell(Z_i,\theta_0)$ denotes the efficient influence function. Suppose that b_n is any other regular estimator of $b(\theta_0)$ such that $\sqrt{n}\left(b_n - b(\theta_0)\right) = n^{-1/2}\sum_{i=1}^{n}\varrho(Z_i,\theta_0) + o_p(1)$ for some $\varrho(Z_i,\theta_0)$ such that $E[\varrho(Z_i,\theta_0)] = 0$. Let $\tilde b\left(\hat\theta\right) = b(\theta_0) + \Upsilon n^{-1}\sum_{i=1}^{n}\psi(Z_i,\theta_0)$ and $\tilde b_n = b(\theta_0) + n^{-1}\sum_{i=1}^{n}\varrho(Z_i,\theta_0)$. Then

$$E\left(\sqrt{n}\left(\hat\theta - \frac{\tilde b(\hat\theta)}{n} - \theta_0\right)\right)^2 = E\left(\sqrt{n}\left(\hat\theta - \frac{\tilde b_n}{n} - \theta_0\right)\right)^2.$$

We now consider a few special cases of this result that are relevant in practice. Intstead of analytical or numerical evaluation of the integral one can replace the integral by sample averages. For

$$\hat b(\theta) \equiv \tau\left(n^{-1}\sum_i m(Z_i,\theta)\right),$$

an alternative bias correction is then

$$\hat\theta_a \equiv \hat\theta - \frac{\hat b\left(\hat\theta\right)}{n}.$$

We can show that $\hat\theta_a$ and $\hat\theta_c$ have the same mean squared error up to order $O\left(n^{-1}\right)$ by analyzing their higher order variance. Let

$$\overline{m}(\theta) \equiv E[m(Z_i,\theta)] = \int m(z,\theta) f(z,\theta_0)\,dz \tag{8}$$

with j th element $\overline{m}_j(\theta)$ and write $\overline{m} = \overline{m}(\theta_0)$ ...

where $A_n = \tau_m \left(M - I^{-1} U(\theta_0) \right) + n^{-1/2} \sum_i \left(m(z_i, \theta_0) - \overline{m} \right)$, $C_n = \tau_m (M + ¤) \left(I^{-1} U(\theta_0) \right)$, and

$$E\left[C_n \theta^{\epsilon}(0)' \right] = E\left[A_n \theta^{\epsilon}(0)' \right] = \tau_m (M + ¤) I^{-1}.$$

Proof. See Appendix A.4. ■

This result has an intuitive explanation. Consider any two bias corrrected estimators

$$\overline{\theta} = \hat{\theta} + \overline{b}/n, \tilde{\theta} = \hat{\theta} + \tilde{b}/n.$$

Suppose that $\hat{\theta}, \overline{b}$, and $\tilde{b}$ are joint asymptotically normal estimators of θ_0, $b(\theta_0)$, and $b(\theta_0)$ respectively, so that $\hat{\theta}$ and $\overline{b} - \tilde{b}$ are joint asymptotically normal estimators of θ_0 and 0 respectively. Asymptotic efficiency of the ML means that $\hat{\theta}$ must be asymptotically uncorrelated with $\overline{b} - \tilde{b}$; otherwise, some linear combination of $\hat{\theta}$ and $\overline{b} - \tilde{b}$ would be an estimator with smaller asymptotic variance than $\hat{\theta}$. Consquently, $\hat{\theta}$ must have the same asymptotic covariance with both $\overline{b}$ and $\tilde{b}$. Then, since the presence of the bias correction affects the third-order variance only through the asymptotic covariance of $\hat{\theta}$ with the bias correction (because the bias correction is $O_p(n^{3/2})$) it follows that the bias corrected ML has the same third-order variance for both $\overline{b}$ and $\tilde{b}$.[1]

This result seems to depend on the efficiency of the ML, so that for other estimators the form of the bias correction may affect the third-order variance. It would be interesting to extend this result to estimators that are efficient within some class, to see whether bias correction would effect the third-order variance of these estimators. This extension is beyond the scope of this paper.

Given the preceding discussion, it is perhaps not surprising that the Bootstrap and Jackknife bias corrected Maximum Likelihood estimators have the same approximate MSE as $\hat{\theta}_c$:

Theorem 3 Assume Conditions 1,2 and 3 hold. Then,

$$\frac{1}{2} E\left[B \theta^{\epsilon}(0) \right] = \frac{1}{2} E\left[J \theta^{\epsilon}(0) \right] = \tau_m (M + ¤) I^{-1}.$$

Proof. See Appendix A.4. ■

Remark 1 Theorems 2 and 3 are irrelevant when the relevant loss function is not approximate MSE. On the other hand, equation (7) indicates that the higher order equivalence of Bootstrap and Jackknife goes beyond the MSE comparison.

bootstrap and jackknife procedures can be used to remove bias terms of stochastic order n^{-1} from a ML estimator without affecting higher-order efficiency. Furthermore, we found that the third-order stochastic expansion of the bootstrap and jackknife bias corrected ML are identical, so that they should have the same higher-order properties. These results show that analytical bias corrections are not needed for achieving full third-order efficiency of the ML.

A Proofs

A.1 Some Preliminary Lemmas

Lemma 1 Assume that W_i are iid with $E[W_i] = 0$ and $E\left[W_i^{2k}\right] < \infty$. Then,

$$E\left[\left(\sum_{i=1}^{n} W_i\right)^{2k}\right] = C(k)n^k + o(n^k)$$

for some constant $C(k)$.

Proof. By adopting an argument in the proof of Lemma 5.1 in Lahiri (1992), we have

$$E\left[\left(\sum_{i=1}^{n} W_i\right)^{2k}\right] = \sum_{j=1}^{2k} \sum_{\alpha} C(\alpha_1, ..., \alpha_j) \sum_{I} E\left[\prod_{s=1}^{j} W_{i_s}^{\alpha_s}\right], \tag{9}$$

where for each fixed $j \in \{1, ..., 2k\}$, $\sum_{\alpha}$ extends over all j-tuples of positive integers $(\alpha_1, ..., \alpha_j)$ such that $\alpha_1 + ... + \alpha_j = 2k$ and $\sum_{I}$ extends over all ordered j-tuples $(i_1, ..., i_j)$ of integers such that $1 \leq i_j \leq n$. Also, $C(\alpha_1, ..., \alpha_j)$ stands for a bounded constant. Note, that if $j > k$ then at least one of the indices $\alpha_j = 1$. By independence and the fact that $EW_i = 0$ it follows that $E\prod_{s=1}^{j} W_{i_s}^{\alpha_s} = 0$ whenever $j > k$. This shows that $E\left(\sum_{i=1}^{n} W_i\right)^{2k} = C(k)n^k + o(n^k)$ for some constant $C(k)$. ■

Lemma 2 Suppose that $\{\xi_i, i = 1, 2, \ldots\}$ is a sequence of zero mean i.i.d. random variables. We also assume that $E\left[|\xi_i|^{16}\right] < \infty$. We then have

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} \xi_i\right| > \eta\right] = O\left[n^{-8}\right]$$

for every $\eta > 0$.

Proof. Using Lemma 1, we obtain

Lemma 3 Suppose that, for each i, $\{\xi_i(\phi), i = 1, 2, \ldots\}$ is a sequence of zero mean i.i.d. random variables indexed by some parameter $\phi \in \Theta$. We also assume that $\sup_{\phi\in\Theta} |\xi_i(\phi)| \le B_i$ for some sequence of random variables B_i that is i.i.d. Finally, we assume that $E\left[|B_i|^{16}\right] < \infty$. We then have

$$\Pr\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i(\phi_n)\right| > n^{\frac{1}{12}-\upsilon}\right] = o\left(n^{-1+16\upsilon}\right)$$

for every υ such that $\upsilon < \frac{1}{16}$. For $\upsilon < \frac{1}{48}$ we have

$$\Pr\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i(\phi_n)\right| > n^{\frac{1}{12}-\upsilon}\right] = o(n^{-1}).$$

Here, ϕ_n is an arbitrary sequence in Θ.

Proof. By Markov's inequality, we have

$$\begin{aligned}
\Pr\left[\sup_{\phi\in\Theta}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i(\phi)\right| > n^{\frac{1}{12}-\upsilon}\right] &= \Pr\left[\sup_{\phi\in\Theta}\left|\sum_{i=1}^{n} \xi_i(\phi_n)\right| > n^{\frac{7}{12}-\upsilon}\right] \\
&\le \frac{E\left[\sup_{\phi\in\Theta}\left(\sum_{i=1}^{n}\xi_i(\phi)\right)^{16}\right]}{n^{\frac{28}{3}-16\upsilon}\eta^{16}} \\
&= \frac{\sup_{\phi\in\Theta} E\left[\left(\sum_{i=1}^{n}\xi_i(\phi)\right)^{16}\right]}{n^{\frac{28}{3}-16\upsilon}\eta^{16}},
\end{aligned}$$

where the last equality is based on dominated convergence. By Lemma 1, we have

$$E\left[\left(\sum_{i=1}^{n}\xi_i(\phi)\right)^{16}\right] \le C n^8,$$

where $C > 0$ is a constant. Therefore, we have

$$\Pr\left[\sup_{\phi\in\Theta}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i(\phi)\right| > n^{\frac{1}{12}-\upsilon}\right] \le \frac{Cn^8}{n^{28/3-16\upsilon}\eta^{16}} = O\left(n^{-4/3+16\upsilon}\right).$$

■

Lemma 4 Let $\hat{Q}(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n}\log f(Z_i, \theta)$. Suppose that Condition 1 holds. We then have for all $\eta > 0$ that

Lemma 5 Under Condition 1, we have

$$\Pr\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}|\theta(\epsilon)-\theta_0|\ge\eta\right]=o\left(n^{-\frac{2}{3}}\right)$$

for every $\eta>0$.

Proof. Let η be given, and let $\varepsilon\equiv G(\theta_0)-\sup_{\{\theta:|\theta-\theta_0|>\eta\}}G(\theta)>0$. Letting $g(z,\theta)\equiv\log f(z,\theta)$, we have

$$\int g(z,\theta)\,dF_\epsilon(z)=\left(1-\epsilon\sqrt{n}\right)G(\theta)+\epsilon\sqrt{n}\,\widehat{G}(\theta)$$

and

$$\left|\int g(z,\theta)\,dF_\epsilon(z)-G(\theta)\right|\le\left(1-\epsilon\sqrt{n}\right)\left|\widehat{G}(\theta)-G(\theta)\right|\le\left|\widehat{G}(\theta)-G(\theta)\right|.$$

Here, the last inequality is based on the fact that $0\le\epsilon\le\frac{1}{\sqrt{n}}$. By Lemma 4, we have

$$\Pr\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\sup_{\theta}\left|\int g(z,\theta)\,dF_\epsilon(z)-G(\theta)\right|\ge\eta\right]=o\left(n^{-\frac{2}{3}}\right)$$

Therefore, for every $0\le\epsilon\le\frac{1}{\sqrt{n}}$ with probability equal to $1-o\left(n^{-\frac{2}{3}}\right)$, we have

$$\begin{aligned}\max_{|\theta-\theta_0|>\eta}\int g(z,\theta)\,dF_\epsilon(z)&\le\max_{|\theta-\theta_0|>\eta}G(\theta)+\frac{1}{3}\varepsilon\\&<G(\theta_0)-\frac{2}{3}\varepsilon\\&<\int g(z,\theta_0)\,dF_\epsilon(z)-\frac{1}{3}\varepsilon.\end{aligned}$$

We also have

$$\max_{\theta}\int g(z,\theta)\,dF_\epsilon(z)\ge\int g(z,\theta_0)\,dF_\epsilon(z)$$

by definition. It follows that

$$\max_{|\theta-\theta_0|>\eta}\int g(z,\theta)\,dF_\epsilon(z)<\max_{\theta}\int g(z,\theta)\,dF_\epsilon(z)-\frac{1}{3}\varepsilon$$

for every $0\le\epsilon\le\frac{1}{\sqrt{n}}$. We therefore obtain that $\Pr\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}|\theta(\epsilon)-\theta_0|\ge\eta\right]=o\left(n^{-\frac{2}{3}}\right)$. ■

Also,

$$\Pr\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\int K(\cdot;\theta(\epsilon))\,d\cdot\right| > Cn^{\frac{1}{12}-\upsilon}\right] = o\left(n^{-1+16\upsilon}\right)$$

for some constant $C > 0$ and for every υ such that $\upsilon < \frac{1}{16}$. If $\upsilon < \frac{1}{48}$ then the above order is $o\left(n^{-1}\right)$.

Proof. Note that we may write

$$\begin{aligned}
&\int K(z;\theta(\epsilon))\,dF_\epsilon(z) - E[K(Z_i;\theta_0)]\\
&= \int K(z;\theta(\epsilon))\,dF_\epsilon(z) - \int K(z;\theta_0)\,dF(z)\\
&= \int K(z;\theta(\epsilon))\,dF_\epsilon(z) - \int K(z;\theta_0)\,dF_\epsilon(z) + \int K(z;\theta_0)\,dF_\epsilon(z) - \int K(z;\theta(\epsilon))\,dF(z)\\
&= \int \frac{\partial K(z;\theta^*)}{\partial\theta}(\theta(\epsilon)-\theta_0)\,dF_\epsilon(z) + \epsilon\sqrt{n}\int K(z;\theta_0)\,d\left(\hat{F}-F\right)(z)
\end{aligned}$$

where θ^* is between θ_0 and $\theta(\epsilon)$. Therefore, we have

$$\begin{aligned}
\left|\int K(z;\theta(\epsilon))\,dF_\epsilon(z) - E[K(Z_i;\theta_0)]\right| \le{}& |\theta(\epsilon)-\theta_0|\cdot\left(E[M(Z_i)] + \frac{1}{n}\sum_{i=1} M(Z_i)\right)\\
&+\left|\frac{1}{n}\sum_{i=1}(M(Z_i) - E[M(Z_i)])\right|
\end{aligned}$$

where $M(\cdot)$ is defined in Condition 1. Using Lemma 5, we can bound

$$\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}}\left|\int K(z;\theta(\epsilon))\,dF_\epsilon(z) - E[K(Z_i;\theta_0)]\right|$$

in absolute value by some $\eta > 0$ with probability $1 - o\left(n^{-\frac{23}{3}}\right)$.

Using Condition 1 and Lemmas 3, we can also show that $\left|\int K(\cdot;\theta(\epsilon))\,d\cdot\right|$ can be bounded by $Cn^{\frac{1}{12}-\upsilon}$ for some constant $C > 0$ and υ such that $\upsilon < \frac{1}{16}$ with probability $1 - o\left(n^{-1+16\upsilon}\right)$. Similarly, if $\upsilon < \frac{1}{48}$, then the statement holds with probability $o(n^{-1})$. ■

Lemma 7 Suppose that Condition 1 holds. Then, we have

$$\Pr\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}|\theta^\epsilon(\epsilon)| > Cn^{\frac{1}{12}-\upsilon}\right] = o\left(n^{-1+16\upsilon}\right)$$

Proof. From (28), we have

$$\theta^{\epsilon}(\epsilon) = -\left[\int \ell^{\theta}(z,\epsilon)\, dF_{\epsilon}(z)\right]^{-1} \left[\int \ell(\cdot,\epsilon)\, d\Delta\right]$$

Using Lemma 6, we can bound the denominator by some $C > 0$, and the numerator by some $Cn^{\frac{1}{12}-\upsilon}$ with probability $1 - o\left(n^{-1+16\upsilon}\right)$, from which the first conclusion follows. As for the second conclusion, we note from (29) that we have

$$0 = E_{\epsilon}\left[\ell^{\theta\theta}(Z_i,\epsilon)\right](\theta^{\epsilon}(\epsilon))^2 + E_{\epsilon}\left[\ell^{\theta}(Z_i,\epsilon)\right]\theta^{\epsilon\epsilon}(\epsilon) + 2\left(\int \ell^{\theta}(z,\epsilon)\, d\Delta(z)\right)\theta^{\epsilon}(\epsilon)$$

The second conclusion follows by using Lemmas 6 along with the first conclusion. The rest of the Lemmas can be established similarly. Note that if $\upsilon < \frac{1}{48}$ then we can apply the specialized result of Lemma 6 in the same way as before. ■

Lemma 8 Suppose that Condition 1 holds. Let $\overline{m}_j(\theta)$ be as defined in 8. Then

$$\sqrt{n}\left(\widehat{I} - I\right) = -\left(V(\theta_0) - Q_1(\theta_0)\, I^{-1} U(\theta_0)\right) + o_p(1),$$

$$\sqrt{n}\left(\widehat{Q}_1\left(\widehat{\theta}\right) - Q_1(\theta_0)\right) = W(\theta_0) + Q_2(\theta_0)\, I^{-1} U(\theta_0) + o_p(1),$$

$$\sqrt{n}\left(\overline{m}_1\left(\widehat{\theta}\right) - \overline{m}_1(\theta_0)\right) = 2E\left[U_i(\theta_0)\, V_i(\theta_0)\right] I^{-1} U(\theta_0) + o_p(1),$$

$$\sqrt{n}\left(\overline{m}_3\left(\widehat{\theta}\right) - \overline{m}_3(\theta_0)\right) = \left(E\left[V_i(\theta_0)^2\right] + \left(E\left[\ell^{\theta}(Z_i,\theta_0)\right]\right)^2 + E\left[U_i(\theta_0)\, W_i(\theta_0)\right]\right) I^{-1} U(\theta_0)$$

Proof. Let $\overline{m}_0(\theta) \equiv \int \ell^{\theta}(z,\theta)\, f(z,\theta_0)\, dz$. Note that

$$\begin{aligned}
\widehat{I} - I &= -n^{-1}\sum_{i=1} \ell^{\theta}\left(Z_i,\widehat{\theta}\right) + E\left[\ell^{\theta}(Z_i,\theta_0)\right] \\
&= -n^{-1}\sum_{i=1}\left(\ell^{\theta}(Z_i,\theta_0) - \overline{m}_0(\theta_0)\right) + o_p\left(n^{-1/2}\right) - \left(\overline{m}_0\left(\widehat{\theta}\right) - \overline{m}_0(\theta_0)\right),
\end{aligned}$$

where the last equality is based on the usual stochastic equicontinuity. Also note that $\partial \overline{m}_0(\theta)/\partial\theta = \int \ell^{\theta\theta}(z,\theta)\, f(z,\theta_0)\, dz$ by dominated convergence. We therefore obtain

$$\sqrt{n}\left(\widehat{I} - I\right) = -n^{-1/2}\sum_{i=1}\left(\ell^{\theta}(Z_i,\theta_0) - E\left[\ell^{\theta}(Z_i,\theta_0)\right]\right) - E\left[\ell^{\theta\theta}(Z_i,\theta_0)\right]\sqrt{n}\left(\widehat{\theta} - \theta_0\right) + o_p(1)$$

$$\sqrt{n}\left(\overline{m}_1\left(\hat{\theta}\right) - \overline{m}_1(\theta_0)\right) = 2E\left[\ell(Z_i,\theta_0)\,\ell^{\theta}(Z_i,\theta_0)\right]\sqrt{n}\left(\hat{\theta} - \theta_0\right) + o_p(1)$$

$$= 2E\left[U_i(\theta_0)\,V_i(\theta_0)\right] I^{-1} U(\theta_0) + o_p(1),$$

$$\sqrt{n}\left(\overline{m}_3\left(\hat{\theta}\right) - \overline{m}_3(\theta_0)\right) = \left(E\left[\ell^{\theta}(Z_i,\theta_0)^2\right] + E\left[\ell(Z_i,\theta_0)\,\ell^{\theta\theta}(Z_i,\theta_0)\right]\right)\sqrt{n}\left(\hat{\theta} - \theta_0\right) + o_p(1)$$

$$= \left(E\left[V_i(\theta_0)^2\right] + \left[E\left[\ell^{\theta}(Z_i,\theta_0)\right]\right]^2 + E\left[U_i(\theta_0)\,W_i(\theta_0)\right]\right) I^{-1} U(\theta_0)$$

■

A.2 Lemmas for Bootstrapped Statistics

Proposition 6 Assume that Conditions 1,2 and 3 hold. Let $\mathcal{F}$ be the class of measurable functions defined in Condition 3. Let $\rightsquigarrow$ denote weak convergence. Let $(\Omega, \mathcal{F}, P)$ be a probability space such that $Z_i : \left(\Omega^{\mathbb{N}}, \mathcal{F}^{\mathbb{N}}, P^{\mathbb{N}}\right) \to (\Omega, \mathcal{F}, P)$ are coordinate projections. Then, for $f \in \mathcal{F}$, $\sqrt{n}\left(\hat{F} - F\right)f \rightsquigarrow Tf$ where T is a tight Brownian bridge with variance covariance function $F(t \wedge s) - F(s)F(t)$. Let BL_1 be the set of all function $h : l^{\infty}(\mathcal{F}) \mapsto [0,1]$ such that $|h(z_1) - h(z_2)| \le \|z_1 - z_2\|_{\mathcal{F}}$ for every z_1 and z_2 where $l^{\infty}(\mathcal{F})$ is the set of uniformly bounded real functions on $\mathcal{F}$ and $\|.\|_{\mathcal{F}}$ is the uniform norm for maps from $\mathcal{F}$ to $\mathbb{R}$. Then $\sup_{h \in BL_1}\left|E^*h\left[\sqrt{n}\left(\hat{F}^* - \hat{F}\right)f\right] - Eh[Tf]\right| \to 0$, $P^{\mathbb{N}}$-a.s.

Proof. We first show that for $f \in \mathcal{F}$, $\sqrt{n}\left(\hat{F} - F\right)f \rightsquigarrow Tf$ or in other words that $\mathcal{F}$ is a Donsker class. Define $\mathcal{F}_\delta = \left\{f - g : f, g \in \mathcal{F}, E\left[\|f - g\|^2\right] < \delta\right\}$, $\mathcal{F}_\infty = \{f - g : f, g \in \mathcal{F}\}$ and $\mathcal{F}_\infty^2 = \left\{f^2 : f \in \mathcal{F}_\infty\right\}$. In light of van der Vaart and Wellner (1996, Theorem 2.5.2), it is enough to show that $\mathcal{F}_\delta$ and $\mathcal{F}_\infty^2$ are F measurable classes for every $\delta > 0$ and $E\left[M(z)^2\right] < \infty$. The second requirement is satisfied by Condition 1. Since $\mathcal{F}_\delta \subset \mathcal{F}_\infty$ the first condition holds if for $f \in \mathcal{F}_\infty^2$ and any vector $a \in \mathbb{R}^n$ and any n the function $s(Z_1, ..., Z_n) = \sup_{\theta_1,\theta_2 \in \Theta}\left|\sum_i^n a_i\left[\ell^{(k)}(Z_i,\theta_1) - \ell^{(k)}(Z_i,\theta_2)\right]^2\right|$ is measurable. Let Θ_k be an increasing sequence of countable subsets of Θ whose limit is dense in Θ. Then

$$s_k(Z_1, ..., Z_n) = \sup_{\theta_1,\theta_2 \in \Theta_k}\left|\sum_i a_i\left(\ell^{(k)}(Z_i,\theta_1) - \ell^{(k)}(Z_i,\theta_2)\right)^2\right|$$

is measurable by Condition 2. By continuity of $\ell^{(k)}(Z_i,\theta)$ in θ it follows that

$$\liminf_k s_k(Z_1, ..., Z_n) = s(Z_1, ..., Z_n)$$

variables B_i that is i.i.d. Finally, we assume that $E\left[|B_i|^{16}\right] < \infty$. We then have

$$P^*\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i^*(\phi_n)\right| > n^{\frac{1}{12}-\upsilon}\right] = o_p\left(n^{-1+16\upsilon}\right)$$

for every υ such that $\upsilon \le \frac{1}{16}$. Moreover,

$$P^*\left[\left|\frac{1}{n}\sum_{i=1}^{n} \xi_i^*(\phi_n)\right| > n^{\frac{1}{12}-\upsilon}\right] = o_p\left(n^{-\frac{23}{3}}\right).$$

Here, ϕ_n is an arbitrary sequence in Φ and P^* is the conditional probability measure of Z_i^* given Z_i.

Proof. Note that $\sum_{i=1}^n \xi_i^*(\phi) = \sum_{i=1}^n (N_{ni}-1)\,\tau(Z_i,\phi)$ where $N_{n1},\ldots,N_{nn}$ is multinomially distributed with parameters $(n,1/n,\ldots,1/n) = (k,p_1,\ldots,p_n)$ and independent of Z_i such that $\Pr\left(\cap_{i=1}^n \{N_{ni}=n_i\}\right) = n!/\left(\prod_i n_i!\right)\prod_i n^{-n_i}$ where $\sum_i^n n_i = n$, $n_i \ge 0$. Let $\kappa_{r_1 r_2 \ldots r_n}$ be the mixed higher order cumulant of $N_{n1},\ldots,N_{nn}$ of order $r = r_1+\ldots+r_n$ for $r_i \ge 0$, r_i integer. Mixed higher order cumulants can be obtained from Guldberg's (1935) recurrence relation $\kappa_{r_1 r_2 .. r_i+1 .. r_n} = a_i \partial\left(\kappa_{r_1 r_2 .. r_i .. r_n}\right)/\partial a_i$ where $a_i = p_i/p_1$. Let b be the number of non zero indices r_i. The arguments in Wishart (1949) imply that for $p_i = n^{-1}$ we have $\kappa_{r_1 r_2 \ldots r_n} \le c n^{-b+1}$ for some constant c. For notational convenience we will represent cumulants with zero indices as lower order cumulants of the variables with non-zero indices, i.e. write $\kappa_{\ldots r_i \not{j} ..} = \kappa_{r_1 r_2 \ldots r_n}$ where $r_j = 0$.

Consider

$$\begin{aligned} P^*\left[\sup_{\phi\in\Phi}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i^*(\phi)\right| > n^{\frac{1}{12}-\upsilon}\right] &= P^*\left[\sup_{\phi\in\Phi}\left|\sum_{i=1}^{n} \xi_i^*(\phi_n)\right| > n^{\frac{7}{12}-\upsilon}\right] \\ &\le \frac{E^*\left[\sup_{\phi\in\Phi}\left(\sum_{i=1}^n \xi_i^*(\phi)\right)^{16}\right]}{n^{\frac{28}{3}-16\upsilon}\eta^{16}} \\ &= \frac{\sup_{\phi\in\Phi} E^*\left[\left(\sum_{i=1}^n \xi_i^*(\phi)\right)^{16}\right]}{n^{\frac{28}{3}-16\upsilon}\eta^{16}}, \end{aligned}$$

where the last equality uses the fact that $\sup_{\phi\in\Phi}$ does not involve $N_{n1},\ldots,N_{nn}$. By adopting an argument in the proof of Lemma 5.1 in Lahiri (1992), we have

$$E^*\left(\sum_{i=1}^n \xi_i^*(\phi)\right)^{2k} = \sum_{j=1}^{2k}\sum_{\alpha} C(\alpha_1,\ldots,\alpha_j)\sum_{I}\prod_{t=1}^{j}\tau(Z_{i_t},\phi)^{\alpha_t} E^*\prod_{s=1}^{j}(N_{ni_s}-1)^{\alpha_s}, \qquad (10)$$

where for each fixed $j \in \{1,\ldots,2k\}$, $\sum_\alpha$ extends over all j-tuples of positive integers $(\alpha_1,\ldots,\alpha_j)$ such that $\alpha_1+\ldots+\alpha_j = 2k$ and $\sum_I$ extends over all ordered j-tuples $(i_1,\ldots,i_j)$ of integers such that $1 \le$

where $\sum_{r^{(1)}+\ldots+r^{(q)}=\alpha}$ indicates the sum over all ordered sets of nonnegative integral vectors $r^{(p)}$, $\left|r^{(p)}\right|>0$, whose sum is α. Since the order of 10 depends both on the number of nonzero terms in $\sum_l$ and the size of $\mu(\alpha_1,...,\alpha_j)$ for each j, we analyze the term

$$S(n,j)=\sum_l \prod_{t=1} \tau(Z_{i_t},\phi)^{\alpha_s} E^*\prod_{s=1}(N_{ni_s}-1)^{\alpha_s}$$

for each j. Note that $\left|\prod_{t=1}^{j} \tau(Z_{i_t},\phi)^{\alpha_s}\right|$ is bounded almost surely and therefore does not affect the analysis. Also, $\sum_l$ is a sum over n^j terms and thus is $O(n^j)$ if all these terms are nonzero. The crucial factor in determining the overall order is therefore $E^*\prod_{s=1}^{j}(N_{ni_s}-1)^{\alpha_s}$. We start with $j=1$. Then $\alpha_1=2k$, $q=1...2k$ and $r^{(p)}$ are scalars. Consequently, $\kappa_{r_1^{(p)}}=c_1$ where c_1 is some constant and $S(n,1)\le c_2\sum_{i=1}^{n}|\tau(Z_{i_t},\phi)|^{2k}$ for some other constant c_2. If $j\le k$ then for $q=1...2q$, $r^{(p)}$ are vectors with possibly only one element different from zero. Again, $S(n,j)\le c_2\sum_l\prod_{t=1}^{j}|\tau(Z_{i_t},\phi)|^{\alpha_s}$ for $j\le k$. If $j\ge k$ then α contains at least $2(j-k)$ elements $\alpha_i=1$. Now assume that for some p, $r_i^{(p)}=1$ and $r_j^{(p)}=0$ for $i\ne j$. Then $\kappa_{r_i^{(p)}}=E(N_{ni_s}-1)=0$ and thus $\prod_{p=1}^{q}\kappa_{r_1^{(p)}r_2^{(p)}....r_j^{(p)}}=0$. On the other hand if $r_i^{(p)}=1$ and $r_j^{(p)}\ne 0$ for at least one $j\ne i$ then $\kappa_{r_1^{(p)}r_2^{(p)}....r_n^{(p)}}\le c_1n^{-1}$. Since there must exists p' corresponding to the other $\alpha_{i'}=1$ such that either $r_{i'}^{(p')}=1$ and $r_j^{(p')}=0$ for $i'\ne j$ or $r_{i'}^{(p')}=1$ and $r_j^{(p')}\ne 0$ for at least one $j\ne i'$, it follows that $\prod_{p=1}^{q}\kappa_{r_1^{(p)}r_2^{(p)}....r_j^{(p)}}=c_3n^{-2(j-k)}$, at most. It now follows that $S(n,j)\le c_2n^{-2(j-k)}\sum_l\prod_{t=1}^{j}|\tau(Z_{i_t},\phi)|^{\alpha_s}$ for all $j>k$. Then,

$$E\,|S(n,j)|\le c_2\sum_l E\left(\prod_{t=1}|\tau(Z_{i_t},\phi)|^{\alpha_s}\right)\le c_2n^jE\,|\tau(Z_{i_t},\phi)|^{2k}$$

for $j\le k$ and

$$E\,|S(n,j)|\le c_2n^{-2(j-k)}\sum_l E\left(\prod_{t=1}|\tau(Z_{i_t},\phi)|^{\alpha_s}\right)\le c_2n^{2k-j}E\,|\tau(Z_{i_t},\phi)|^{2k}\le c_2n^kE\,|\tau(Z_{i_t},\phi)|^{2k}$$

for $j>k$. Together these results imply that

$$E\left|E^*\left(\sum_{i=1}^{n}\xi_i^*(\phi)\right)^{2k}\right|\le C(k)n^kE\,|\tau(Z_{i_t},\phi)|^{2k}$$

where $C(k)$ is a constant that depends on k. By the Markov inequality it follows that $E^*\left(\sum_{i=1}^{n}\xi_i^*(\phi)\right)^{2k}=O_p(n^k)$. We conclude that

Lemma 10 Under Condition 1, we have

$$P^*\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\hat\theta^*(\epsilon)-\hat\theta\right|\ge\eta\right]=o_p\left(n^{-\frac{2}{3}}\right).$$

Proof. For any $\eta>0$, there exists some $\delta>0$ such that $|\theta-\theta_0|>\eta/2$ implies $|G(\theta)-G(\theta_0)|>\delta$. Let $\hat G^*(\theta)\equiv\int g(z,\theta)d\hat F^*(z)$ and $\hat G^*_\epsilon(\theta)\equiv\int g(z,\theta)\,d\hat F_\epsilon(z)$. Then,

$$P^*\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\hat\theta^*(\epsilon)-\hat\theta\right|\ge\eta\right]\le P^*\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|G(\hat\theta^*(\epsilon))-G\left(\hat\theta\right)\right|>\delta\right].$$

Because

$$\begin{aligned}G\left(\hat\theta^*(\epsilon)\right)-G\left(\hat\theta\right)=&\left(G\left(\hat\theta^*(\epsilon)\right)-\hat G\left(\hat\theta^*(\epsilon)\right)\right)+\left(\hat G\left(\hat\theta^*(\epsilon)\right)-\hat G^*_\epsilon\left(\hat\theta^*(\epsilon)\right)\right)\\&+\left(\hat G^*_\epsilon\left(\hat\theta^*(\epsilon)\right)-\hat G^*_\epsilon\left(\hat\theta\right)\right)+\left(\hat G^*_\epsilon\left(\hat\theta\right)-G\left(\hat\theta\right)\right)\end{aligned}$$

and

$$\left|\hat G^*_\epsilon(\theta)-\hat G(\theta)\right|\le\left|\hat G^*(\theta)-\hat G(\theta)\right|,$$

we obtain

$$\begin{aligned}&\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|G\left(\hat\theta^*(\epsilon)\right)-G\left(\hat\theta\right)\right|\\
\le\;&\sup_{\theta\in\Theta}\left|\hat G^*(\theta)-\hat G(\theta)\right|+\sup_{\theta\in\Theta}\left|\hat G(\theta)-G(\theta)\right|\\
&+\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\hat G^*_\epsilon\left(\hat\theta^*(\epsilon)\right)-\hat G^*_\epsilon\left(\hat\theta\right)\right|+\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\hat G^*_\epsilon\left(\hat\theta\right)-G\left(\hat\theta\right)\right|\\
\le\;&\sup_{\theta\in\Theta}\left|\hat G^*(\theta)-\hat G(\theta)\right|+\sup_{\theta\in\Theta}\left|\hat G(\theta)-G(\theta)\right|\\
&+\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\hat G^*_\epsilon\left(\hat\theta^*(\epsilon)\right)-\hat G^*_\epsilon\left(\hat\theta\right)\right|+\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\hat G^*_\epsilon\left(\hat\theta\right)-\hat G\left(\hat\theta\right)\right|+\left|\hat G\left(\hat\theta\right)-G\left(\hat\theta\right)\right|\\
\le\;&\sup_{\theta\in\Theta}\left|\hat G^*(\theta)-\hat G(\theta)\right|+\sup_{\theta\in\Theta}\left|\hat G(\theta)-G(\theta)\right|\\
&+\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\hat G^*_\epsilon\left(\hat\theta^*(\epsilon)\right)-\hat G^*_\epsilon\left(\hat\theta\right)\right|+\left|\hat G^*\left(\hat\theta\right)-\hat G\left(\hat\theta\right)\right|+\left|\hat G\left(\hat\theta\right)-G\left(\hat\theta\right)\right|\\
\le\;&2\sup_{\theta\in\Theta}\left|\hat G^*(\theta)-\hat G(\theta)\right|+2\sup_{\theta\in\Theta}\left|\hat G(\theta)-G(\theta)\right|+\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\hat G^*_\epsilon\left(\hat\theta^*(\epsilon)\right)-\hat G^*_\epsilon\left(\hat\theta\right)\right|\qquad(11)\end{aligned}$$

By Lemma 9, we have

where $1\{\cdot\}$ denotes an indicator function. For every $\sigma > 0$, we have

$$\begin{aligned}
&\Pr\left[P^{*}\left(\sup_{\theta\in\Theta}\left|\widehat{G}(\theta)-G(\theta)\right|>\frac{\delta}{6}\right)>\sigma n^{-\frac{2}{3}}\right] \\
=\;&\Pr\left[1\left\{\sup_{\theta\in\Theta}\left|\widehat{G}(\theta)-G(\theta)\right|>\frac{\delta}{6}\right\}>0\right] \\
=\;&\Pr\left[\sup_{\theta\in\Theta}\left|\widehat{G}(\theta)-G(\theta)\right|>\frac{\delta}{6}\right]=o(1)
\end{aligned} \tag{13}$$

where the last equality is implied by Lemma 4. It therefore follows that

$$P^{*}\left[\sup_{\theta\in\Theta}\left|\widehat{G}(\theta)-G(\theta)\right|>\frac{\delta}{6}\right]=o_p\left(n^{-\frac{2}{3}}\right). \tag{14}$$

Finally,

$$\begin{aligned}
\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\widehat{G}^{*}_{\epsilon}\left(\widehat{\theta}^{*}(\epsilon)\right)-\widehat{G}^{*}_{\epsilon}\left(\widehat{\theta}\right)\right| &\le \max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\widehat{G}^{*}_{\epsilon}\left(\widehat{\theta}^{*}(\epsilon)\right)-\widehat{G}\left(\widehat{\theta}\right)\right|+\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\widehat{G}^{*}_{\epsilon}\left(\widehat{\theta}\right)-\widehat{G}\left(\widehat{\theta}\right)\right| \\
&= \max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\sup_{\theta}\widehat{G}^{*}_{\epsilon}(\theta)-\sup_{\theta}\widehat{G}(\theta)\right|+\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\widehat{G}^{*}_{\epsilon}\left(\widehat{\theta}\right)-\widehat{G}\left(\widehat{\theta}\right)\right| \\
&\le \max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\sup_{\theta}\widehat{G}^{*}_{\epsilon}(\theta)-\sup_{\theta}\widehat{G}(\theta)\right|+\left|\widehat{G}^{*}\left(\widehat{\theta}\right)-\widehat{G}\left(\widehat{\theta}\right)\right| \\
&\le \max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\sup_{\theta}\left|\widehat{G}^{*}_{\epsilon}(\theta)-\widehat{G}(\theta)\right|+\sup_{\theta}\left|\widehat{G}^{*}(\theta)-\widehat{G}(\theta)\right| \\
&\le \max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\sup_{\theta}\left|\widehat{G}^{*}(\theta)-\widehat{G}(\theta)\right|+\sup_{\theta}\left|\widehat{G}^{*}(\theta)-\widehat{G}(\theta)\right| \\
&\le \sup_{\theta}\left|\widehat{G}^{*}(\theta)-\widehat{G}(\theta)\right|+\sup_{\theta}\left|\widehat{G}^{*}(\theta)-\widehat{G}(\theta)\right| \\
&= 2\sup_{\theta}\left|\widehat{G}^{*}(\theta)-\widehat{G}(\theta)\right|
\end{aligned}$$

Here, the first equality is based on the definitions of $\widehat{\theta}^{*}(\epsilon)$ and $\widehat{\theta}$. Because

$$P^{*}\left(\sup_{\theta\in\Theta}\left|\widehat{G}^{*}(\theta)-\widehat{G}(\theta)\right|>\delta\right)=o_p\left(n^{-\frac{2}{3}}\right)$$

we can conclude that

$$P^{*}\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\widehat{G}^{*}_{\epsilon}\left(\widehat{\theta}^{*}(\epsilon)\right)-\widehat{G}^{*}_{\epsilon}\left(\widehat{\theta}\right)\right|>\frac{\delta}{3}\right]=o_p\left(n^{-\frac{2}{3}}\right). \tag{15}$$

The conclusion follows by combining (11) - (15). ■

Lemma 11 Assume that Condition 1 is satisfied. Let $K(\cdot;\theta(\epsilon))$ be defined as in Lemma 6. Then, for any $n > 0$, we have

Proof. In the same way as in the proof of Lemma 6

$$\int K\left(z;\widehat{\theta}^{*}(\epsilon)\right) d\widehat{F}_{\epsilon}(z) - \int K(z;\theta_0)\, d\widehat{F}(z)$$
$$= \int \frac{\partial K(z;\theta^{*})}{\partial\theta}\left(\widehat{\theta}^{*}(\epsilon) - \theta_0\right) d\widehat{F}_{\epsilon}(z) + \epsilon\sqrt{n}\int K(z;\theta_0)\, d\left(\widehat{F}^{*} - \widehat{F}\right)(z)$$

where θ^{*} is between θ_0 and $\widehat{\theta}^{*}(\epsilon)$. Therefore, we have

$$\left|\int K(z;\theta(\epsilon))\, dF_{\epsilon}(z) - \int K(z;\theta_0)\, d\widehat{F}(z)\right| \le \left|\widehat{\theta}^{*}(\epsilon) - \theta_0\right| \cdot \left(\frac{1}{n}\sum_{i=1} M(Z_i) + \frac{1}{n}\sum_{i=1} M(Z_i^{*})\right)$$
$$+ \left|\frac{1}{n}\sum_{i=1} M(Z_i^{*}) - \frac{1}{n}\sum_{i=1} M(Z_i)\right|$$

where $M(\cdot)$ is defined in Condition 1. Let $\bar{M} = \frac{1}{n}\sum_{i=1}^{n} M(Z_i)$ and $\bar{M}^{*} = \frac{1}{n}\sum_{i=1}^{n} M(Z_i^{*})$. Then, for any η and some c

$$P^{*}\left[\left|\widehat{\theta}^{*}(\epsilon) - \theta_0\right|\bar{M} > \eta\right] \le P^{*}\left[\left|\widehat{\theta}^{*}(\epsilon) - \theta_0\right| > \eta/c\right] + P^{*}\left[\left|\bar{M} - E[M(Z_i)]\right| > c\right] = o_p\left(n^{-\frac{2}{3}}\right)$$

since $P^{*}\left[\left|\bar{M} - E[M(Z_i)]\right| > c\right] = 1$ with probability equal to $P\left(\left|\bar{M} - E[M(Z_i)]\right| > c\right) = o\left(n^{-\frac{2}{3}}\right)$ by Lemma 2 and zero otherwise for some c. Then, $P^{*}\left[\left|\bar{M} - E[M(Z_i)]\right| > c\right] = o_p\left(n^{-\frac{2}{3}}\right)$ by the same argument as in 13 Moreover,

$$P^{*}\left[\left|\widehat{\theta}^{*}(\epsilon) - \theta_0\right|\left|\bar{M}^{*} - \bar{M}\right| > \eta\right] \le P^{*}\left[\left|\widehat{\theta}^{*}(\epsilon) - \theta_0\right| > \eta/c\right] + P^{*}\left[\left|\bar{M}^{*} - \bar{M}\right| > c\right] = o_p\left(n^{-\frac{2}{3}}\right)$$

by Lemmas 9 and 10. It thus follows that for any $\eta > 0$,

$$P^{*}\left(\left|\int K(z;\theta(\epsilon))\, dF_{\epsilon}(z) - \int K(z;\theta_0)\, d\widehat{F}(z)\right| > \eta\right) = o_p\left(n^{-\frac{2}{3}}\right).$$

Finally note that $P^{*}\left(\left|\int K(z;\theta_0)\, d\widehat{F}(z) - EK(z;\theta_0)\right| > \eta\right) = 1$ with probability

$$P\left(\left|\int K(z;\theta_0)\, d\widehat{F}(z) - E[K(z;\theta_0)]\right| > \eta\right) = o(n^{-\frac{2}{3}})$$

by Lemma 2. Thus, by the same argument as in 13

$$P^{*}\left(\left|\int K(z;\theta_0)\, d\widehat{F}(z) - EK(z;\theta_0)\right| > \eta\right) = o_p\left(n^{-\frac{2}{3}}\right).$$

where

$$P^*\left[\sup_{|\theta-\hat\theta|<\delta}\left|\int K(\cdot;\theta)\,d\hat{F}\right| > Cn^{\frac{1}{12}-\upsilon}\right] = o_p\left(n^{-1+16\upsilon}\right)$$

follows directly from Lemma 9 and

$$P^*\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\hat\theta^*(\epsilon)-\hat\theta\right|\ge\delta\right] = o_p\left(n^{-\frac{23}{3}}\right)$$

follows from Lemma 10. ■

Lemma 12 Suppose that Condition 1 holds. Then, we have

$$P^*\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\hat\theta^{\epsilon}(\epsilon)\right| > Cn^{\frac{1}{12}-\upsilon}\right] = o_p\left(\max\left(n^{-\frac{23}{3}}, n^{-1+16\upsilon}\right)\right)$$

$$P^*\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\hat\theta^{\epsilon\epsilon}(\epsilon)\right| > C\left(n^{\frac{1}{12}-\upsilon}\right)^2\right] = o_p\left(\max\left(n^{-\frac{23}{3}}, n^{-1+16\upsilon}\right)\right)$$

$$\vdots$$

$$P^*\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{n}}}\left|\hat\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}(\epsilon)\right| > C\left(n^{\frac{1}{12}-\upsilon}\right)^6\right] = o_p\left(\max\left(n^{-\frac{23}{3}}, n^{-1+16\upsilon}\right)\right)$$

for some constant $C > 0$ and for every υ such that $\upsilon < \frac{1}{16}$.

Proof. Let $\hat{M}_\epsilon = \int \ell^\theta(z,\epsilon)\,d\hat{F}_\epsilon(z)$ such that

$$\hat\theta^{\epsilon}(\epsilon) = -\hat{M}_\epsilon^{-1}\int \ell(\cdot,\epsilon)\,d\hat{F}$$

and for any $\delta > 0$ some $C > 0$ and for every υ such that $\upsilon < \frac{1}{16}$

$$\begin{aligned}
P^*\left[\left|\hat\theta^{\epsilon}(\epsilon)\right| > Cn^{\frac{1}{12}-\upsilon}\right] &\le P^*\left[\sup_\epsilon\left|\int \ell(\cdot,\epsilon)\,d\hat{F}\right| > \delta Cn^{\frac{1}{12}-\upsilon}\right] \\
&\quad + P^*\left[\sup_\epsilon\left|\hat{M}_\epsilon - E\left[\ell^\theta(z,\theta_0)\right]\right| \ge \delta\right] \\
&= o_p\left(\max\left(n^{-\frac{23}{3}}, n^{-1+16\upsilon}\right)\right)
\end{aligned}$$

by Lemma 11. The rest of the Lemma can be established similarly. ■

A.3 Moments of Bootstrapped and Jackknifed Statistics

The following results are stated without proof. They can be derived with straightforward but tedious

Lemma 14 Let $X^*_{k,i} = \tau_k\left(Z^*_i, \hat\theta\right)$ for $k = 1, 2$ be some transformation of Z^*_i, where τ_k possibly depends on the sample $\{Z_i\}_{i=1}^n$ through $\hat\theta$. Then

$$
\begin{aligned}
& E^*\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X^*_{1,i}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X^*_{2,i}\right)\right] \\
= \; & \frac{1}{n}\sum_{i=1}^n X_{1,i}X_{2,i} + \frac{n-1}{n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{1,i}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{2,i}\right)
\end{aligned}
$$

where $X_{k,i} = \tau_k(Z_i, \hat\theta)$.

Lemma 15 Let $X^*_{k,i} = \tau_k\left(Z^*_i, \hat\theta\right)$ for $k = 1, 2$ be some transformation of Z^*_i, where τ_k possibly depends on the sample $\{Z_i\}_{i=1}^n$ through $\hat\theta$. Then

$$
\begin{aligned}
& E^*\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X^*_{1,i}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X^*_{2,i}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X^*_{3,i}\right)\right] \\
= \; & \frac{1}{\sqrt{n}}\frac{1}{n}\sum_{j=1} X_{1,j}X_{2,j}X_{3,j} + \frac{n-1}{n}\left(\frac{1}{n}\sum_{j=1} X_{1,j}X_{2,j}\right)\left(\frac{1}{\sqrt{n}}\sum_{j=1} X_{3,j}\right) \\
& + \frac{n-1}{n}\left(\frac{1}{n}\sum_{j=1} X_{3,j}X_{1,j}\right)\left(\frac{1}{\sqrt{n}}\sum_{j=1} X_{2,j}\right) + \frac{n-1}{n}\left(\frac{1}{n}\sum_{j=1} X_{2,j}X_{3,j}\right)\left(\frac{1}{\sqrt{n}}\sum_{j=1} X_{1,j}\right) \\
& + \frac{n^2-3n+2}{n^2}\left(\frac{1}{\sqrt{n}}\sum_{j=1} X_{1,j}\right)\left(\frac{1}{\sqrt{n}}\sum_{j=1} X_{2,j}\right)\left(\frac{1}{\sqrt{n}}\sum_{j=1} X_{3,j}\right).
\end{aligned}
$$

Lemma 16 Let $U^*_i(\theta) \equiv \ell(Z^*_i, \theta)$, $V^*_i(\theta) \equiv \ell^\theta(Z^*_i, \theta) - \overline{\ell^\theta(\cdot, \theta)} \equiv \ell^\theta(Z^*_i, \theta) - n^{-1}\sum_{i=1}^n \ell^\theta(Z_i, \theta)$, $W^*_i(\theta) \equiv \ell^{\theta\theta}(Z^*_i) - \overline{\ell^{\theta\theta}(\cdot, \theta)} \equiv \ell^{\theta\theta}(Z^*_i) - n^{-1}\sum_{i=1}^n \ell^{\theta\theta}(Z_i, \theta)$ and let $U^*(\theta) = n^{-1/2}\sum_{i=1}^n U^*_i(\theta)$, $V^*(\theta) = n^{-1/2}\sum_{i=1}^n V^*_i(\theta)$ and $W^*(\theta) = n^{-1/2}\sum_{i=1}^n W^*_i(\theta)$. Then (a)

$$
\begin{aligned}
E^*\left[U^*\left(\hat\theta\right)\right] &= 0, \\
E^*\left[V^*\left(\hat\theta\right)\right] &= 0, \\
E^*\left[W^*\left(\hat\theta\right)\right] &= 0,
\end{aligned}
$$

(b)

$$
\begin{aligned}
E^*\left[\left(U^*\left(\hat\theta\right)\right)^2\right] &= \frac{1}{n}\sum_{i=1}^n \ell\left(Z_i, \hat\theta\right)^2 \\
E^*\left[U^*\left(\hat\theta\right)V^*\left(\hat\theta\right)\right] &= \frac{1}{n}\sum_{i=1}^n \ell\left(Z_i, \hat\theta\right)\ell^\theta\left(Z_i, \hat\theta\right)
\end{aligned}
$$

(c)

$$
E^*\left[U^*(\theta)^3\right] = \frac{1}{\sqrt{n}}\frac{1}{n}\sum \ell\left(Z_i, \hat\theta\right)^3
$$

Lemma 17 Let

$$W = \frac{1}{\sqrt{n}} \sum_{i=1} X_i, \qquad W_{(j)} = \frac{1}{\sqrt{n-1}} \sum_{i \neq j} X_i$$

Then, we have

$$nW - \sqrt{n}\sqrt{n-1}\,\frac{1}{n} \sum_{j=1} W_{(j)} = W$$

Lemma 18 Let

$$W = \left(\frac{1}{\sqrt{n}} \sum_{i=1} X_{1,i} \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1} X_{2,i} \right), \qquad W_{(j)} = \left(\frac{1}{\sqrt{n-1}} \sum_{i \neq j} X_{1,i} \right) \left(\frac{1}{\sqrt{n-1}} \sum_{i \neq j} X_{2,i} \right)$$

Then,

$$nW - \sum_{j=1} W_{(j)} = \frac{1}{n-1} \sum_{i \neq j} X_{1,i} X_{2,j}$$

Lemma 19 Let

$$\begin{aligned} W &= \left(\frac{1}{\sqrt{n}} \sum_{i=1} X_{1,i} \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1} X_{2,i} \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1} X_{3,i} \right), \\ W_{(j)} &= \left(\frac{1}{\sqrt{n-1}} \sum_{i \neq j} X_{1,i} \right) \left(\frac{1}{\sqrt{n-1}} \sum_{i \neq j} X_{2,i} \right) \left(\frac{1}{\sqrt{n-1}} \sum_{i \neq j} X_{3,i} \right) \end{aligned}$$

Then,

$$\begin{aligned} & nW - \sqrt{\frac{n}{n-1}} \sum_{j=1} W_{(j)} \\ = \; & \frac{n^2 + n}{(n-1)^2} W \\ & - \frac{n^2}{(n-1)^2} \left(\sqrt{n}\bar{X}_1\right) \left(\frac{1}{n} \sum_{i=1} X_{2,i} X_{3,i} \right) - \frac{n^2}{(n-1)^2} \left(\sqrt{n}\bar{X}_2\right) \left(\frac{1}{n} \sum_{i=1} X_{3,i} X_{1,i} \right) \\ & - \frac{n^2}{(n-1)^2} \left(\sqrt{n}\bar{X}_3\right) \left(\frac{1}{n} \sum_{i=1} X_{1,i} X_{2,i} \right) + \frac{n\sqrt{n}}{(n-1)^2} \left(\frac{1}{n} \sum_{t=1} X_{1,i} X_{2,i} X_{3,i} \right) \end{aligned}$$

Lemma 20 Let

Then,

$$
\begin{aligned}
& nW - \frac{n^2}{n-1}\frac{1}{n}\sum_{i=1} W_{(j)} \\
= \quad & \frac{n\left(n^2+3n-1\right)}{(n-1)^3}W \\
& - \frac{n^3}{(n-1)^3}\left(-\sqrt{n}\bar{X}_1\right)\left(-\sqrt{n}\bar{X}_2\right)\left(\frac{1}{n}\sum_{i=1} X_{3,i}X_{4,i}\right) - \frac{n^3}{(n-1)^3}\left(-\sqrt{n}\bar{X}_1\right)\left(-\sqrt{n}\bar{X}_3\right)\left(\frac{1}{n}\sum_{i=1} X_{2,i}X_{4,i}\right) \\
& - \frac{n^3}{(n-1)^3}\left(-\sqrt{n}\bar{X}_1\right)\left(-\sqrt{n}\bar{X}_4\right)\left(\frac{1}{n}\sum_{i=1} X_{2,i}X_{3,i}\right) - \frac{n^3}{(n-1)^3}\left(-\sqrt{n}\bar{X}_2\right)\left(-\sqrt{n}\bar{X}_3\right)\left(\frac{1}{n}\sum_{i=1} X_{1,i}X_{4,i}\right) \\
& - \frac{n^3}{(n-1)^3}\left(-\sqrt{n}\bar{X}_2\right)\left(-\sqrt{n}\bar{X}_4\right)\left(\frac{1}{n}\sum_{i=1} X_{1,i}X_{3,i}\right) - \frac{n^3}{(n-1)^3}\left(-\sqrt{n}\bar{X}_3\right)\left(-\sqrt{n}\bar{X}_4\right)\left(\frac{1}{n}\sum_{i=1} X_{1,i}X_{2,i}\right) \\
& + \frac{n^3}{\sqrt{n}(n-1)^3}\left(-\sqrt{n}\bar{X}_1\right)\left(\frac{1}{n}\sum_{i=1} X_{2,i}X_{3,i}X_{4,i}\right) + \frac{n^3}{\sqrt{n}(n-1)^3}\left(-\sqrt{n}\bar{X}_2\right)\left(\frac{1}{n}\sum_{i=1} X_{1,i}X_{3,i}X_{4,i}\right) \\
& + \frac{n^3}{\sqrt{n}(n-1)^3}\left(-\sqrt{n}\bar{X}_3\right)\left(\frac{1}{n}\sum_{i=1} X_{1,i}X_{2,i}X_{4,i}\right) + \frac{n^3}{\sqrt{n}(n-1)^3}\left(-\sqrt{n}\bar{X}_4\right)\left(\frac{1}{n}\sum_{i=1} X_{1,i}X_{2,i}X_{3,i}\right) \\
& - \frac{n^2}{(n-1)^3}\left(\frac{1}{n}\sum_{i=1} X_{1,i}X_{2,i}X_{3,i}X_{4,i}\right)
\end{aligned}
$$

Lemma 21 Let

$$
\begin{aligned}
W &= \left(\frac{1}{\sqrt{n}}\sum_{i=1} X_{1,i}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1} X_{2,i}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1} X_{3,i}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1} X_{4,i}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1} X_{5,i}\right), \\
W_{(j)} &= \left(\frac{1}{\sqrt{n-1}}\sum_{i\neq j} X_{1,i}\right)\left(\frac{1}{\sqrt{n-1}}\sum_{i\neq j} X_{2,i}\right)\left(\frac{1}{\sqrt{n-1}}\sum_{i\neq j} X_{3,i}\right) \\
& \quad \times \left(\frac{1}{\sqrt{n-1}}\sum_{i\neq j} X_{4,i}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1} X_{5,i}\right)
\end{aligned}
$$

Then,

$$
\begin{aligned}
& W - \frac{n\sqrt{n}}{(n-1)\sqrt{n-1}} \frac{1}{n} \sum_{j=1} W_{(j)} \\
= \quad & \frac{n^3+6n^2-4n+1}{(n-1)^4} \left(\sqrt{n}\bar{X}_1\right)\left(\sqrt{n}\bar{X}_2\right)\left(\sqrt{n}\bar{X}_3\right)\left(\sqrt{n}\bar{X}_4\right)\left(\sqrt{n}\bar{X}_5\right) \\
& - \frac{n^2}{(n-1)^4} \left(\sqrt{n}\bar{X}_1\right)\left(\sqrt{n}\bar{X}_2\right)\left(\sqrt{n}\bar{X}_5\right)\left(\frac{1}{n}\sum_{j=1} X_{3,j}X_{4,j}\right) \\
& - \frac{n^2}{(n-1)^4} \left(\sqrt{n}\bar{X}_1\right)\left(\sqrt{n}\bar{X}_2\right)\left(\sqrt{n}\bar{X}_3\right)\left(\frac{1}{n}\sum_{j=1} X_{4,j}X_{5,j}\right) \\
& - \frac{n^2}{(n-1)^4} \left(\sqrt{n}\bar{X}_1\right)\left(\sqrt{n}\bar{X}_3\right)\left(\sqrt{n}\bar{X}_5\right)\left(\frac{1}{n}\sum_{j=1} X_{2,j}X_{4,j}\right) \\
& - \frac{n^2}{(n-1)^4} \left(\sqrt{n}\bar{X}_1\right)\left(\sqrt{n}\bar{X}_4\right)\left(\sqrt{n}\bar{X}_5\right)\left(\frac{1}{n}\sum_{j=1} X_{2,j}X_{3,j}\right) \\
& - \frac{n^2}{(n-1)^4} \left(\sqrt{n}\bar{X}_1\right)\left(\sqrt{n}\bar{X}_3\right)\left(\sqrt{n}\bar{X}_4\right)\left(\frac{1}{n}\sum_{j=1} X_{2,j}X_{5,j}\right) \\
& - \frac{n^2}{(n-1)^4} \left(\sqrt{n}\bar{X}_3\right)\left(\sqrt{n}\bar{X}_4\right)\left(\sqrt{n}\bar{X}_5\right)\left(\frac{1}{n}\sum_{j=1} X_{1,j}X_{2,j}\right) \\
& - \frac{n^2}{(n-1)^4} \left(\sqrt{n}\bar{X}_2\right)\left(\sqrt{n}\bar{X}_3\right)\left(\sqrt{n}\bar{X}_5\right)\left(\frac{1}{n}\sum_{j=1} X_{1,j}X_{4,j}\right) \\
& - \frac{n^2}{(n-1)^4} \left(\sqrt{n}\bar{X}_2\right)\left(\sqrt{n}\bar{X}_4\right)\left(\sqrt{n}\bar{X}_5\right)\left(\frac{1}{n}\sum_{j=1} X_{1,j}X_{3,j}\right) \\
& - \frac{n^2}{(n-1)^4} \left(\sqrt{n}\bar{X}_2\right)\left(\sqrt{n}\bar{X}_3\right)\left(\sqrt{n}\bar{X}_4\right)\left(\frac{1}{n}\sum_{j=1} X_{1,j}X_{5,j}\right) \\
& - \frac{n^2}{(n-1)^4} \left(\sqrt{n}\bar{X}_1\right)\left(\sqrt{n}\bar{X}_2\right)\left(\sqrt{n}\bar{X}_4\right)\left(\frac{1}{n}\sum_{j=1} X_{3,j}X_{5,j}\right)
\end{aligned}
$$

(continued)

$$+\frac{n\sqrt{n}}{(n-1)^4}\left(-\sqrt{n}\overline{X}_3\right)\left(-\sqrt{n}\overline{X}_5\right)\left(\frac{1}{n}\sum_{j=1} X_{1,j}X_{2,j}X_{4,j}\right)$$

$$+\frac{n\sqrt{n}}{(n-1)^4}\left(-\sqrt{n}\overline{X}_1\right)\left(-\sqrt{n}\overline{X}_5\right)\left(\frac{1}{n}\sum_{j=1} X_{2,j}X_{3,j}X_{4,j}\right)$$

$$+\frac{n\sqrt{n}}{(n-1)^4}\left(-\sqrt{n}\overline{X}_1\right)\left(-\sqrt{n}\overline{X}_3\right)\left(\frac{1}{n}\sum_{j=1} X_{2,j}X_{4,j}X_{5,j}\right)$$

$$+\frac{n\sqrt{n}}{(n-1)^4}\left(-\sqrt{n}\overline{X}_1\right)\left(-\sqrt{n}\overline{X}_4\right)\left(\frac{1}{n}\sum_{j=1} X_{2,j}X_{3,j}X_{5,j}\right)$$

$$+\frac{n\sqrt{n}}{(n-1)^4}\left(-\sqrt{n}\overline{X}_1\right)\left(-\sqrt{n}\overline{X}_2\right)\left(\frac{1}{n}\sum_{j=1} X_{3,j}X_{4,j}X_{5,j}\right)$$

$$+\frac{n\sqrt{n}}{(n-1)^4}\left(-\sqrt{n}\overline{X}_2\right)\left(-\sqrt{n}\overline{X}_5\right)\left(\frac{1}{n}\sum_{j=1} X_{1,j}X_{3,j}X_{4,j}\right)$$

$$+\frac{n\sqrt{n}}{(n-1)^4}\left(-\sqrt{n}\overline{X}_2\right)\left(-\sqrt{n}\overline{X}_3\right)\left(\frac{1}{n}\sum_{j=1} X_{1,j}X_{4,j}X_{5,j}\right)$$

$$+\frac{n\sqrt{n}}{(n-1)^4}\left(-\sqrt{n}\overline{X}_2\right)\left(-\sqrt{n}\overline{X}_4\right)\left(\frac{1}{n}\sum_{j=1} X_{1,j}X_{3,j}X_{5,j}\right)$$

$$+\frac{n\sqrt{n}}{(n-1)^4}\left(-\sqrt{n}\overline{X}_4\right)\left(-\sqrt{n}\overline{X}_5\right)\left(\frac{1}{n}\sum_{j=1} X_{1,j}X_{2,j}X_{3,j}\right)$$

$$+\frac{n\sqrt{n}}{(n-1)^4}\left(-\sqrt{n}\overline{X}_3\right)\left(-\sqrt{n}\overline{X}_4\right)\left(\frac{1}{n}\sum_{j=1} X_{1,j}X_{2,j}X_{5,j}\right)$$

$$-\frac{n}{(n-1)^4}\left(-\sqrt{n}\overline{X}_1\right)\left(\frac{1}{n}\sum_{j=1} X_{2,j}X_{3,j}X_{4,j}X_{5,j}\right)$$

$$-\frac{n}{(n-1)^4}\left(-\sqrt{n}\overline{X}_2\right)\left(\frac{1}{n}\sum_{j=1} X_{1,j}X_{3,j}X_{4,j}X_{5,j}\right)$$

$$-\frac{n}{(n-1)^4}\left(-\sqrt{n}\overline{X}_3\right)\left(\frac{1}{n}\sum X_{1,j}X_{2,j}X_{4,j}X_{5,j}\right)$$

A.4 Proofs of Main Results

Proof of Proposition 1. Let $\hat{Q}(\theta) = \int \log f(\cdot,\theta)\, d\hat{F}(z)$, $Q_\epsilon(\theta) = \int \log f(\cdot,\theta)\, dF_\epsilon(z)$ and $Q(\theta) = \int \log f(\cdot,\theta)\, dQ(z)$ such that $Q_\epsilon(\theta) - Q(\theta) = \epsilon\sqrt{n}\left(\hat{Q}(\theta) - Q(\theta)\right)$. By Conditions (1), (2) and (3) and van der Vaart and Wellner (1996, Theorem 2.4.3) it follows that $\sup_\theta |Q_\epsilon(\theta) - Q(\theta)| \le \sup_\theta \left|\hat{Q}(\theta) - Q(\theta)\right| \to 0$ in probability. By van der Vaart and Wellner (1996, Corollary 3.2.3), it follows that uniformly in $\epsilon \in \left[-n^{-1/2}, n^{-1/2}\right]$, $\hat{\theta}(\epsilon) \overset{p}{\to} \theta_0$. This implies that for any compact set $K \subset \Theta$ with $\theta_0 \in K$, $P(\hat{\theta}(\epsilon) \in K) \to 1$, as $n \to \infty$. Consider the function $G(\epsilon,\theta) \equiv \int \ell(\cdot,\theta)\, dF_\epsilon(z)$. If $\partial\Theta$ is the boundary of Θ then $P\left(G\left(\epsilon,\hat{\theta}(\epsilon)\right) \neq 0\right) \le P\left(\hat{\theta}(\epsilon) \in \partial\Theta\right) \le 1 - P(\theta(\epsilon) \in K) \to 0$. We now condition on the event $\left\{G\left(\epsilon,\hat{\theta}(\epsilon)\right) = 0\right\}$.

By Taylor's theorem there exists some $\bar{\epsilon} \in [0, 1/\sqrt{n}]$ such that $\hat{\theta}\left(n^{-1/2}\right) = \theta(0) + \sum_{k=1}^{m-1} \frac{1}{k!n^{k/2}} \theta^{(k)}(0) + \frac{1}{m!n^{m/2}} \theta^{(m)}(\bar{\epsilon})$. By Lemmas 5 and 6 it follows that $\max_{0 \le \epsilon \le n^{-1/2}} \theta^{(k)}(\epsilon) = O_p(1)$ such that the remainder term $\frac{1}{m!n^{m/2}} \theta^{(m)}(\bar{\epsilon}) = O_p\left(n^{-m/2}\right)$ for $m \le 6$. To find the derivatives $\theta^{(k)}$, let

$$h(z,\epsilon) \equiv \ell(z,\theta(\epsilon)),$$

and rewrite the first order condition as

$$0 = \int h(z,\epsilon)\, dF_\epsilon(z)$$

Differentiating repeatedly with respect to ϵ, we obtain

$$0 = \int \frac{dh(z,\epsilon)}{d\epsilon} dF_\epsilon(z) + \int h(z,\epsilon)\, d\Delta(z) \tag{16}$$

$$0 = \int \frac{d^2h(z,\epsilon)}{d\epsilon^2} dF_\epsilon(z) + 2\int \frac{dh(z,\epsilon)}{d\epsilon} d\Delta(z) \tag{17}$$

$$0 = \int \frac{d^3h(z,\epsilon)}{d\epsilon^3} dF_\epsilon(z) + 3\int \frac{d^2h(z,\epsilon)}{d\epsilon^2} d\Delta(z) \tag{18}$$

$$0 = \int \frac{d^4h(z,\epsilon)}{d\epsilon^4} dF_\epsilon(z) + 4\int \frac{d^3h(z,\epsilon)}{d\epsilon^3} d\Delta(z) \tag{19}$$

$$0 = \int \frac{d^5h(z,\epsilon)}{d\epsilon^5} dF_\epsilon(z) + 5\int \frac{d^4h(z,\epsilon)}{d\epsilon^4} d\Delta(z) \tag{20}$$

$$0 = \int \frac{d^6h(z,\epsilon)}{d\epsilon^6} dF_\epsilon(z) + 6\int \frac{d^5h(z,\epsilon)}{d\epsilon^5} d\Delta(z) \tag{21}$$

Note that

$$\frac{d^4 h(\epsilon)}{d\epsilon^4} = \ell^{\theta\theta\theta\theta}(\theta^\epsilon)^4 + 6\ell^{\theta\theta\theta}(\theta^\epsilon)^2\theta^{\epsilon\epsilon} + 3\ell^{\theta\theta}(\theta^{\epsilon\epsilon})^2 + 4\ell^{\theta\theta}\theta^\epsilon\theta^{\epsilon\epsilon\epsilon} + \ell^\theta\theta^{\epsilon\epsilon\epsilon\epsilon} \tag{25}$$

$$\begin{aligned} \frac{d^5 h(\epsilon)}{d\epsilon^5} &= \ell^{\theta\theta\theta\theta\theta}(\theta^\epsilon)^5 + 10\ell^{\theta\theta\theta\theta}(\theta^\epsilon)^3\theta^{\epsilon\epsilon} + 15\ell^{\theta\theta\theta}\theta^\epsilon(\theta^{\epsilon\epsilon})^2 \\ &\quad +10\ell^{\theta\theta\theta}(\theta^\epsilon)^2\theta^{\epsilon\epsilon\epsilon} + 10\ell^{\theta\theta}\theta^{\epsilon\epsilon}\theta^{\epsilon\epsilon\epsilon} + 5\ell^{\theta\theta}\theta^\epsilon\theta^{\epsilon\epsilon\epsilon\epsilon} + \ell^\theta\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon} \end{aligned} \tag{26}$$

$$\begin{aligned} \frac{d^6 h(\epsilon)}{d\epsilon^6} &= \ell^{\theta\theta\theta\theta\theta\theta}(\theta^\epsilon)^6 + 15\ell^{\theta\theta\theta\theta\theta}(\theta^\epsilon)^4\theta^{\epsilon\epsilon} + 45\ell^{\theta\theta\theta\theta}(\theta^\epsilon)^2(\theta^{\epsilon\epsilon})^2 \\ &\quad +20\ell^{\theta\theta\theta\theta}(\theta^\epsilon)^3\theta^{\epsilon\epsilon\epsilon} + 15\ell^{\theta\theta\theta}(\theta^{\epsilon\epsilon})^3 + 60\ell^{\theta\theta\theta}\theta^\epsilon\theta^{\epsilon\epsilon}\theta^{\epsilon\epsilon\epsilon} \\ &\quad +15\ell^{\theta\theta\theta}(\theta^\epsilon)^2\theta^{\epsilon\epsilon\epsilon\epsilon} + 10\ell^{\theta\theta}(\theta^{\epsilon\epsilon\epsilon})^2 + 15\ell^{\theta\theta}\theta^{\epsilon\epsilon}\theta^{\epsilon\epsilon\epsilon\epsilon} + 6\ell^{\theta\theta}\theta^\epsilon\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon} \\ &\quad +\ell^\theta\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon} \end{aligned} \tag{27}$$

Here, θ^ϵ denotes the derivative of θ with respect to ϵ. Combining (16) - (19) with (22) - (25), we obtain

$$0 = E_\epsilon\left[\ell^\theta(Z_i,\epsilon)\right]\theta^\epsilon(\epsilon) + \int \ell(z,\epsilon)\,d\text{¢}(z) \tag{28}$$

$$0 = E_\epsilon\left[\ell^{\theta\theta}(Z_i,\epsilon)\right](\theta^\epsilon(\epsilon))^2 + E_\epsilon\left[\ell^\theta(Z_i,\epsilon)\right]\theta^{\epsilon\epsilon}(\epsilon) + 2\left(\int \ell^\theta(z,\epsilon)\,d\text{¢}(z)\right)\theta^\epsilon(\epsilon) \tag{29}$$

$$\begin{aligned} 0 &= E_\epsilon\left[\ell^{\theta\theta\theta}(Z_i,\epsilon)\right](\theta^\epsilon(\epsilon))^3 + 3E_\epsilon\left[\ell^{\theta\theta}(Z_i,\epsilon)\right]\theta^\epsilon(\epsilon)\theta^{\epsilon\epsilon}(\epsilon) + E_\epsilon\left[\ell^\theta(Z_i,\epsilon)\right]\theta^{\epsilon\epsilon\epsilon}(\epsilon) \\ &\quad +3\left(\int \ell^{\theta\theta}(z,\epsilon)\,d\text{¢}(z)\right)(\theta^\epsilon(\epsilon))^2 + 3\left(\int \ell^\theta(z,\epsilon)\,d\text{¢}(z)\right)\theta^{\epsilon\epsilon}(\epsilon) \end{aligned} \tag{30}$$

$$\begin{aligned} 0 &= E_\epsilon\left[\ell^{\theta\theta\theta\theta}(Z_i,\epsilon)\right](\theta^\epsilon(\epsilon))^4 + 6E_\epsilon\left[\ell^{\theta\theta\theta}(Z_i,\epsilon)\right](\theta^\epsilon(\epsilon))^2\theta^{\epsilon\epsilon}(\epsilon) + 3E_\epsilon\left[\ell^{\theta\theta}(Z_i,\epsilon)\right](\theta^{\epsilon\epsilon}(\epsilon))^2 \\ &\quad +4E_\epsilon\left[\ell^{\theta\theta}(Z_i,\epsilon)\right]\theta^\epsilon(\epsilon)\theta^{\epsilon\epsilon\epsilon}(\epsilon) + E_\epsilon\left[\ell^\theta(Z_i,\epsilon)\right]\theta^{\epsilon\epsilon\epsilon\epsilon}(\epsilon) + 4(\theta^\epsilon(\epsilon))^3\left(\int \ell^{\theta\theta\theta}(z,\epsilon)\,d\text{¢}(z)\right) \\ &\quad +12\theta^\epsilon(\epsilon)\theta^{\epsilon\epsilon}(\epsilon)\left(\int \ell^{\theta\theta}(z,\epsilon)\,d\text{¢}(z)\right) + 4\theta^{\epsilon\epsilon\epsilon}(\epsilon)\left(\int \ell^\theta(z,\epsilon)\,d\text{¢}(z)\right) \end{aligned} \tag{31}$$

$$\begin{aligned} 0 &= E_\epsilon\left[\ell^{\theta\theta\theta\theta\theta}(Z_i,\epsilon)\right](\theta^\epsilon(\epsilon))^5 + 10E_\epsilon\left[\ell^{\theta\theta\theta\theta}(Z_i,\epsilon)\right](\theta^\epsilon(\epsilon))^3\theta^{\epsilon\epsilon}(\epsilon) \\ &\quad +15E_\epsilon\left[\ell^{\theta\theta\theta}(Z_i,\epsilon)\right]\theta^\epsilon(\epsilon)(\theta^{\epsilon\epsilon}(\epsilon))^2 \\ &\quad +10E_\epsilon\left[\ell^{\theta\theta\theta}(Z_i,\epsilon)\right](\theta^\epsilon(\epsilon))^2\theta^{\epsilon\epsilon\epsilon}(\epsilon) + 10E_\epsilon\left[\ell^{\theta\theta}(Z_i,\epsilon)\right]\theta^{\epsilon\epsilon}(\epsilon)\theta^{\epsilon\epsilon\epsilon}(\epsilon) \\ &\quad +5E_\epsilon\left[\ell^{\theta\theta}(Z_i,\epsilon)\right]\theta^\epsilon(\epsilon)\theta^{\epsilon\epsilon\epsilon\epsilon}(\epsilon) + E_\epsilon\left[\ell^\theta(Z_i,\epsilon)\right]\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon}(\epsilon) + 5(\theta^\epsilon(\epsilon))^4\left(\int \ell^{\theta\theta\theta\theta}(z,\epsilon)\,d\text{¢}(z)\right) \end{aligned} \tag{32}$$

and

$$
\begin{aligned}
0 = {} & E_\epsilon\left[\ell^{\theta\theta\theta\theta\theta\theta}(Z_i,\epsilon)\right](\theta^\epsilon(\epsilon))^6 + 15E_\epsilon\left[\ell^{\theta\theta\theta\theta\theta}(Z_i,\epsilon)\right](\theta^\epsilon(\epsilon))^4\theta^{\epsilon\epsilon}(\epsilon) \\
& +45E_\epsilon\left[\ell^{\theta\theta\theta\theta}(Z_i,\epsilon)\right](\theta^\epsilon(\epsilon))^2(\theta^{\epsilon\epsilon}(\epsilon))^2 + 20E_\epsilon\left[\ell^{\theta\theta\theta\theta}(Z_i,\epsilon)\right](\theta^\epsilon(\epsilon))^3\theta^{\epsilon\epsilon\epsilon}(\epsilon) \\
& +15E_\epsilon\left[\ell^{\theta\theta\theta}(Z_i,\epsilon)\right](\theta^{\epsilon\epsilon}(\epsilon))^3 + 60E_\epsilon\left[\ell^{\theta\theta\theta}(Z_i,\epsilon)\right]\theta^\epsilon(\epsilon)\theta^{\epsilon\epsilon}(\epsilon)\theta^{\epsilon\epsilon\epsilon}(\epsilon) \\
& +15E_\epsilon\left[\ell^{\theta\theta\theta}(Z_i,\epsilon)\right](\theta^\epsilon(\epsilon))^2\theta^{\epsilon\epsilon\epsilon\epsilon}(\epsilon) + 10E_\epsilon\left[\ell^{\theta\theta}(Z_i,\epsilon)\right](\theta^{\epsilon\epsilon\epsilon}(\epsilon))^2 \\
& +15E_\epsilon\left[\ell^{\theta\theta}(Z_i,\epsilon)\right]\theta^{\epsilon\epsilon}(\epsilon)\theta^{\epsilon\epsilon\epsilon\epsilon}(\epsilon) + 6E_\epsilon\left[\ell^{\theta\theta}(Z_i,\epsilon)\right]\theta^\epsilon(\epsilon)\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon}(\epsilon) \\
& +E_\epsilon\left[\ell^{\theta}(Z_i,\epsilon)\right]\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}(\epsilon) + 6(\theta^\epsilon(\epsilon))^5\left(\int \ell^{\theta\theta\theta\theta\theta}(z,\epsilon)\,d\Delta(z)\right) \\
& +60(\theta^\epsilon(\epsilon))^3\theta^{\epsilon\epsilon}(\epsilon)\left(\int \ell^{\theta\theta\theta\theta}(z,\epsilon)\,d\Delta(z)\right) + 90\theta^\epsilon(\theta^{\epsilon\epsilon}(\epsilon))^2\left(\int \ell^{\theta\theta\theta}(z,\epsilon)\,d\Delta(z)\right) \\
& +60(\theta^\epsilon(\epsilon))^2\theta^{\epsilon\epsilon\epsilon}(\epsilon)\left(\int \ell^{\theta\theta\theta}(z,\epsilon)\,d\Delta(z)\right) + 60\theta^{\epsilon\epsilon}(\epsilon)\theta^{\epsilon\epsilon\epsilon}(\epsilon)\left(\int \ell^{\theta\theta}(z,\epsilon)\,d\Delta(z)\right) \\
& +30\theta^\epsilon(\epsilon)\theta^{\epsilon\epsilon\epsilon\epsilon}(\epsilon)\left(\int \ell^{\theta\theta}(z,\epsilon)\,d\Delta(z)\right) + 6\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon}(\epsilon)\left(\int \ell^{\theta}(z,\epsilon)\,d\Delta(z)\right) \qquad (34)
\end{aligned}
$$

Here, $E_\epsilon[\cdot]$ is defined such that

$$
E_\epsilon[g(Z_i,\epsilon)] \equiv \int g(z,\epsilon)\,dF_\epsilon(z)
$$

Evaluating expressions (28) - (31) at $\epsilon = 0$, we obtain

$$
\theta^\epsilon = \frac{1}{-E[\ell^\theta]}\left(\int \ell\,d\Delta\right) = \frac{1}{\mathcal{I}}\int \ell\,d\Delta, \qquad (35)
$$

$$
\begin{aligned}
\theta^{\epsilon\epsilon} &= \frac{1}{-E[\ell^\theta]}\left(E\left[\ell^{\theta\theta}\right](\theta^\epsilon)^2 + 2\left(\int \ell^\theta\,d\Delta\right)\theta^\epsilon\right) \\
&= \frac{E\left[\ell^{\theta\theta}\right]}{-E[\ell^\theta]}(\theta^\epsilon)^2 + 2\frac{1}{-E[\ell^\theta]}\left(\int \ell^\theta\,d\Delta\right)\theta^\epsilon \\
&= \frac{E\left[\ell^{\theta\theta}\right]}{\mathcal{I}^3}\left(\int \ell\,d\Delta\right)^2 + \frac{2}{\mathcal{I}^2}\left(\int \ell^\theta\,d\Delta\right)\left(\int \ell\,d\Delta\right), \qquad (36)
\end{aligned}
$$

$$
\begin{aligned}
\theta^{\epsilon\epsilon\epsilon} &= \frac{E\left[\ell^{\theta\theta\theta}\right]}{-E[\ell^\theta]}(\theta^\epsilon)^3 + 3\frac{E\left[\ell^{\theta\theta}\right]}{-E[\ell^\theta]}\theta^\epsilon\theta^{\epsilon\epsilon} + 3\frac{1}{-E[\ell^\theta]}\left(\int \ell^{\theta\theta}\,d\Delta\right)(\theta^\epsilon)^2 + 3\frac{1}{-E[\ell^\theta]}\left(\int \ell^\theta\,d\Delta\right)\theta^{\epsilon\epsilon} \\
&= \left(\frac{E\left[\ell^{\theta\theta\theta}\right]}{\mathcal{I}^4} + \frac{3\left(E\left[\ell^{\theta\theta}\right]\right)^2}{\mathcal{I}^5}\right)\left(\int \ell\,d\Delta\right)^3 + \frac{9E\left[\ell^{\theta\theta}\right]}{\mathcal{I}^4}\left(\int \ell\,d\Delta\right)^2\left(\int \ell^\theta\,d\Delta\right) \\
&\quad + \frac{3}{\mathcal{I}^3}\left(\int \ell\,d\Delta\right)^2\left(\int \ell^{\theta\theta}\,d\Delta\right) + \frac{6}{\mathcal{I}^3}\left(\int \ell\,d\Delta\right)\left(\int \ell^\theta\,d\Delta\right)^2 \qquad (37)
\end{aligned}
$$

$$
\begin{aligned}
\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon} = {} & \frac{E\left[\ell^{\theta\theta\theta\theta\theta}\right]}{-E[\ell^{\theta}]}\left(\theta^{\epsilon}\right)^5 + 10\frac{E\left[\ell^{\theta\theta\theta\theta}\right]}{-E[\ell^{\theta}]}\left(\theta^{\epsilon}\right)^3\theta^{\epsilon\epsilon} + 15\frac{E\left[\ell^{\theta\theta\theta}\right]}{-E[\ell^{\theta}]}\theta^{\epsilon}\left(\theta^{\epsilon\epsilon}\right)^2 \\
& + 10\frac{E\left[\ell^{\theta\theta\theta}\right]}{-E[\ell^{\theta}]}\left(\theta^{\epsilon}\right)^2\theta^{\epsilon\epsilon\epsilon} + 10\frac{E\left[\ell^{\theta\theta}\right]}{-E[\ell^{\theta}]}\theta^{\epsilon\epsilon}\theta^{\epsilon\epsilon\epsilon} \\
& + 5\frac{E\left[\ell^{\theta\theta}\right]}{-E[\ell^{\theta}]}\theta^{\epsilon}\theta^{\epsilon\epsilon\epsilon\epsilon} + 5\frac{1}{-E[\ell^{\theta}]}\left(\theta^{\epsilon}\right)^4\left(\int \ell^{\theta\theta\theta\theta}\, d\Delta\right) \\
& + 30\frac{1}{-E[\ell^{\theta}]}\left(\theta^{\epsilon}\right)^2\theta^{\epsilon\epsilon}\left(\int \ell^{\theta\theta\theta}\, d\Delta\right) + 15\frac{1}{-E[\ell^{\theta}]}\left(\theta^{\epsilon\epsilon}\right)^2\left(\int \ell^{\theta\theta}\, d\Delta\right) \\
& + 20\frac{1}{-E[\ell^{\theta}]}\theta^{\epsilon}\theta^{\epsilon\epsilon\epsilon}\left(\int \ell^{\theta\theta}\, d\Delta\right) + 5\frac{1}{-E[\ell^{\theta}]}\theta^{\epsilon\epsilon\epsilon\epsilon}\left(\int \ell^{\theta}\, d\Delta\right). \qquad (39)
\end{aligned}
$$

■

Proof of Proposition 2. Let $\hat{Q}^*(\theta) = \int \log f(\cdot,\theta)\, d\hat{F}^*(z)$, $\hat{Q}_\epsilon(\theta) = \int \log f(\cdot,\theta)\, d\hat{F}_\epsilon(z)$ and $\hat{Q}(\theta) = \int \log f(\cdot,\theta)\, d\hat{F}(z)$ such that $\hat{Q}_\epsilon(\theta) - \hat{Q}(\theta) = \epsilon\sqrt{n}\left(\hat{Q}^*(\theta) - \hat{Q}(\theta)\right)$. By Conditions (1), (2) and (3) and Giné and Zinn (1996, Theorem 2.6) it follows that $\sup_\theta \left|\hat{Q}_\epsilon(\theta) - \hat{Q}(\theta)\right| \le \sup_\theta \left|\hat{Q}^*(\theta) - \hat{Q}(\theta)\right| \to 0$ in probability, P^Na.s. By standard arguments such as Arcones and Giné (1992), it follows that uniformly in $\epsilon \in \left[-n^{-1/2}, n^{-1/2}\right]$, $\hat{\theta}^*(\epsilon) \xrightarrow{P^*} \hat{\theta}$, P^Na.s. This implies that for any compact set $K \subset \Theta$ with $\theta_0 \in K$, $P^*(\hat{\theta}^*(\epsilon) \in K) \to 1, P^N$a.s. as $n \to \infty$. Consider the function $\hat{G}(\epsilon,\theta) \equiv \int \ell(\cdot,\theta)\, d\hat{F}_\epsilon(z)$. If $\partial\Theta$ is the boundary of Θ then $P^*(\hat{G}(\epsilon,\hat{\theta}^*(\epsilon)) \neq 0) \le P^*\left(\hat{\theta}^*(\epsilon) \in \partial\Theta\right) \le 1 - P^*(\theta^*(\epsilon) \in K) \to 0, .P^N$a.s. We now condition on the event $\left\{\hat{G}\left(\epsilon,\hat{\theta}^*(\epsilon)\right) = 0\right\}$. By the same arguments as in the proof of proposition 1 it follows that there exists some $\bar{\epsilon} \in \left[0, n^{-1/2}\right]$ such that $\sqrt{n}\left(\hat{\theta}^* - \hat{\theta}\right) = \hat{\theta}^\epsilon(0) + \sum_{k=1}^{m-1}\frac{1}{k!n^{k/2}}\hat{\theta}^{(k)}(0) + \frac{1}{m!n^{m/2}}\hat{\theta}^{(m)}(\bar{\epsilon})$ P^Na.s., where $\hat{\theta}^\epsilon(0)$ is obtained from evaluating $\int \frac{dh(z,\epsilon)}{d\epsilon}\, d\hat{F}_\epsilon(z) + \int h(z,\epsilon)\, d\hat{\Delta}(z)$ at $\epsilon = 0$. We obtain

$$o_p(n^{-m/2}) = \int \ell^\theta(z,\hat{\theta})\, d\hat{F}(z)\, \hat{\theta}^\epsilon(0) + \int \ell\left(z,\hat{\theta}\right) d\hat{\Delta}(z),$$

where $\int \ell^\theta(z,\hat{\theta})\, d\hat{F}(z) \equiv n^{-1}\sum_{i=1}^n \ell^\theta(Z_i,\hat{\theta})$ and $\hat{\Delta}(z) \equiv \sqrt{n}\left(\hat{F}^*(z) - \hat{F}(z)\right)$. Similar expressions can be found for higher order derivatives of $\hat{\theta}(\epsilon)$. These expressions depend on $n^{-1}\sum_{i=1}^n \ell^{(k)}\left(Z_i,\hat{\theta}\right)$ and $\int \ell^{(k)}\left(z,\hat{\theta}\right) d\hat{\Delta}(z)$ for $k = 0, 1, ..., 6$. By Condition 1 and Lemma 5, it follows that $n^{-1}\sum_{i=1}^n \ell^{(k)}\left(Z_i,\hat{\theta}\right) \xrightarrow{p} E\left[\ell^{(k)}(Z_i,\theta_0)\right]$ by a uniform law of large numbers. By Proposition 6 the class F is Donsker. By the proof of Theorem 2.4 in Gine and Zinn (1990) it follows that the following conditional stochastic equicontinuity property or

$$\int \ell^{(k)}\left(z,\hat\theta\right) d\hat{G}(z) = \int \ell^{(k)}(z,\theta_0)\, d\hat{G}(z) + o_p(1)\ P^{N}\text{a.s.}$$

It now follows from Proposition 6 and Theorem 2.4 of Gine and Zinn (1990) that $\int \ell^{(k)}(z,\theta_0)\, d\hat{G}(z) \rightsquigarrow \int \ell^{(k)}(z,\theta_0)\, dT(z)$ almost surely, where $T(z)$ is a Brownian Bridge process. We finally have to analyze the term $\hat\theta^{(m)}(\epsilon)$ which contains expressions of the form $\int \ell^{(k)}(z,\hat\theta^*(\epsilon))d\hat{P}_\epsilon(z)$ and $\int \ell^{(k)}\left(z,\hat\theta^*(\epsilon)\right) d\hat{G}(z)$. For $\int \ell^{(k)}\left(z,\hat\theta^*(\epsilon)\right) d\hat{G}(z)$ we use the same inequality as in (40) together with Lemma 10 to show that

$$\int \ell^{(k)}\left(z,\hat\theta^*(\epsilon)\right) d\hat{G}(z) = \int \ell^{(k)}(z,\theta_0)\, d\hat{G}(z) + o_p(1)\ P^{N}\text{a.s.}$$

Next consider

$$\begin{aligned}
&\left|\int \ell^{(k)}(z,\hat\theta^*(\epsilon))d\hat{P}_\epsilon(z) - \int \ell^{(k)}(z,\theta_0)dF(z)\right| \\
&\le |\epsilon|\left|\int \ell^{(k)}(z,\hat\theta^*(\epsilon))d\hat{G}(z)\right| + \left|\int \ell^{(k)}(z,\theta_0)\, d\left(F(z) - \hat{P}(z)\right)\right| \\
&\quad + \left|\int \left[\ell^{(k)}(z,\hat\theta^*(\epsilon)) - \ell^{(k)}(z,\theta_0)\right] d\hat{P}(z)\right|
\end{aligned}$$

where $\int \ell^{(k)}(z,\hat\theta^*(\epsilon))d\hat{G}(z) = O_p(1)\ P^{N}$a.s. by Proposition 6 and $\sup|\epsilon| = O(n^{-1/2})$. The second term is $o_p(1)$ by a law of large numbers. Finally,

$$\begin{aligned}
&P^*\left(\sup_\epsilon \left|\int \left[\ell^{(k)}(z,\hat\theta^*(\epsilon)) - \ell^{(k)}(z,\theta_0)\right] d\hat{P}(z)\right| > \eta\right) \\
&\le P^*\left(\sup_{|\theta-\theta_0|<\delta} \left|\int \left[\ell^{(k)}(z,\theta) - \ell^{(k)}(z,\theta_0)\right] d\hat{P}(z)\right| > \eta\right) + P^*\left(\sup_{0\le\epsilon\le 1/\sqrt{n}} \left|\hat\theta^*(\epsilon) - \theta_0\right| \ge \delta\right)
\end{aligned}$$

where the first probability is zero with P^{N}-probability tending to one by stochastic equicontinuity and the second probability goes to zero by Lemma 10. It follows that $\int \ell^{(k)}(z,\hat\theta^*(\epsilon))d\hat{P}_\epsilon(z) \xrightarrow{p} E\ell^{(k)}(z,\theta_0)\ P^{N}$a.s. Together, these results imply that $\sup_\epsilon \left|\hat\theta^{(k)}(\epsilon)\right| = O_p(1)\ P^{N}$a.s. for $k \le 6$. This establishes the validity of the expansion. ∎

Proof of Theorem 3. Introduce the truncation function $h_n(x)$ where

$$h_n(x) = \begin{cases} -n^\alpha & \text{if } x < -n^\alpha \\ x & \text{if } |x| < n^\alpha \end{cases} \tag{41}$$

Using the expansion for $\sqrt{n}\left(\hat\theta^* - \hat\theta\right)$ from Proposition (2) together with Lemma (12) it follows that $P^*\left(\sqrt{n}\left|\hat\theta^* - \hat\theta\right| > n^{\alpha+1/2}\right) = o_p(n^{-20/3})$. This shows that we can replace $E^*\left(\hat\theta^* - \hat\theta\right)$ with a truncated integral $E^* h_n\left(\hat\theta^* - \hat\theta\right)$. Let

$$\hat\theta^*_a \equiv n^{-1/2}\hat\theta'(0) + \frac{1}{2}\frac{1}{n}\hat\theta''(0) + \frac{1}{6}\frac{1}{n^{3/2}}\hat\theta'''(0) + \frac{1}{24}\frac{1}{n^2}\hat\theta''''(0).$$

Because $|h_n(x) - h_n(y)| \le 2n^\alpha \wedge \|x - y\|$, we have

$$\left|h_n\left(\hat\theta^* - \hat\theta\right) - h_n\left(\hat\theta^*_a\right)\right| \le \min\left(2n^\alpha, \frac{1}{96n^{5/2}} \sup_{0\le\epsilon\le 1/\sqrt{n}}\left\|\hat\theta'''''(\epsilon)\right\|\right).$$

Fix $\varepsilon > 0$ and $\frac{7}{96} < \delta < \frac{1}{2}$ arbitrary. Taking expectations with respect to the measure $\hat P$ leads to

$$\begin{aligned}\left|E^*\left[h_n\left(\hat\theta^* - \hat\theta\right)\right] - E^*\left[h_n\left(\hat\theta^*_a\right)\right]\right| \\ \le \varepsilon/n^{2-\delta} + 2n^\alpha \cdot P^*\left[\frac{1}{96n^{5/2}} \sup_{0\le\epsilon\le 1/\sqrt{n}}\left\|\hat\theta'''''(\epsilon)\right\| > \varepsilon/n^{2-\delta}\right].\end{aligned}$$

Use the fact that $P^*\left[\frac{1}{96n^{5/2}}\sup_{0\le\epsilon\le 1/\sqrt{n}}\left\|\hat\theta'''''(\epsilon)\right\| > \varepsilon/n^{2-\delta}\right] = o_p\left(n^{-76/60-(16/5)\delta}\right)$ by setting $-v = 1/60 + \delta/5$ in Lemma 12. Choose $\delta \in (7/96 + (5/16)\alpha, 1/2)$. It follows that

$$\begin{aligned}\left|E^*\left[h_n\left(\hat\theta^* - \hat\theta\right)\right] - E^*\left[h_n\left(\hat\theta^*_a\right)\right]\right| &\le \varepsilon/n^{2-\delta} + 2o_p\left(n^{-76/60-(16/5)\delta+\alpha}\right) \\ &= O_p(n^{\delta-2}) = o_p(n^{-3/2})\end{aligned}$$

Next, we need to show that $E^*\left[h_n\left(\hat\theta^*_a\right)\right] - E^*\left[\hat\theta^*_a\right] = o_p\left(n^{-3/2}\right)$. Note that

$$\begin{aligned}\left|E^*\left[h_n\left(\hat\theta^*_a\right)\right] - E^*\left[\hat\theta^*_a\right]\right| &\le E^*\left[\left|n^\alpha - \hat\theta^*_a\right| 1\left\{\left|\hat\theta^*_a\right| \ge n^\alpha\right\}\right] \\ &\le E^*\left[\left|\hat\theta^*_a\right| 1\left\{\left|\hat\theta^*_a\right| \ge n^\alpha\right\}\right] \\ &\quad + n^\alpha E^*\left[1\left\{\left|\hat\theta^*_a\right| \ge n^\alpha\right\}\right] \\ &\le 2E^*\left[\frac{\left|\hat\theta^*_a\right|^4}{(n^\alpha)^3}\right].\end{aligned}$$

Here, $\left|\hat\theta^*_a\right|^4$ is a forth order polynomial in $a = \hat\theta'(0)$, $b = \frac{1}{2}\hat\theta''(0)$, $c = \frac{1}{6}\hat\theta'''(0)$, and $d = \frac{1}{24}\hat\theta''''(0)$. Expectations of all terms of the from $E^*\left[a^i b^j c^k d^l\right]$ where $i, j, k, l \in \{0, 1, 2, 3, 4\}$ and $i + j + k + l = 4$ are bounded in probability such that $E^*\left[a^i b^j c^k d^l\right] = O_p(1)$ where $E^*\left[\frac{1}{n^2}a^4\right] = O_p(n^{-2})$ is the largest

In order to evaluate $E^*\left[\hat\theta^*_{aa}\right]$ we use Proposition 2 by which $\hat\vartheta'(0) = \hat{\mathrm{I}}^{-1}U^*\left(\hat\theta\right)$, $\hat\vartheta''(0) = \hat{\mathrm{I}}^{-3}\hat{\mathrm{Q}}_1\left(\hat\theta\right)U^*\left(\hat\theta\right)^2 + 2\hat{\mathrm{I}}^{-2}U^*\left(\hat\theta\right)V^*\left(\hat\theta\right)$ and

$$\begin{aligned}\hat\vartheta'''(0) &= \hat{\mathrm{I}}^{-4}\hat{\mathrm{Q}}_2\left(\hat\theta\right)U^*\left(\hat\theta\right)^3 + 3\hat{\mathrm{I}}^{-5}\hat{\mathrm{Q}}_1\left(\hat\theta\right)^2U^*\left(\hat\theta\right)^3 + 9\hat{\mathrm{I}}^{-4}\hat{\mathrm{Q}}_1\left(\hat\theta\right)U^*\left(\hat\theta\right)^2V^*\left(\hat\theta\right) \\ &\quad + 3\hat{\mathrm{I}}^{-3}U^*\left(\hat\theta\right)^2W^*\left(\hat\theta\right) + 6\hat{\mathrm{I}}^{-3}U^*\left(\hat\theta\right)V^*\left(\hat\theta\right)^2.\end{aligned}$$

Note that $\hat{\mathrm{I}}$, $\hat{\mathrm{Q}}_1$ and $\hat{\mathrm{Q}}_2$ are constants with respect to E^*. It thus follows that

$$E^*\left[\hat\vartheta'(0)\right] = \hat{\mathrm{I}}^{-1}E^*\left[U\left(\hat\theta\right)\right] = 0$$

by Lemma 16(a). We consider $E^*\left[U\left(\hat\theta\right)^2\right] = \frac{1}{n}\sum_{i=1}^n \ell\left(Z_i,\hat\theta\right)^2$. By Proposition 6 and van der Waart and Wellner (1996, Theorem 1.5.7) it follows that

$$\limsup_{n\to\infty} P\left(\sup_{|\theta-\theta_0|<\delta}\left|\frac{1}{n}\sum_{i=1}^n \ell(Z_i,\theta)^2 - \frac{1}{n}\sum_{i=1}^n \ell(Z_i,\theta_0)^2\right| > \varepsilon\right) = 0$$

such that by Lemma 5 it follows that

$$E^*\left[U\left(\hat\theta\right)^2\right] = \frac{1}{n}\sum_{i=1}^n \ell(Z_i,\theta_0)^2 + o_p(1).$$

Similar results can be established for the other expressions of Lemma 16. It therefore follows that

$$\begin{aligned}E^*\left[\hat\vartheta''(0)\right] &= \mathrm{I}^{-3}\mathrm{Q}_1(\theta_0)\frac{1}{n}\sum_{i=1}^n \ell(Z_i,\theta_0)^2 + 2\mathrm{I}^{-2}\frac{1}{n}\sum_{i=1}^n \ell(Z_i,\theta_0)\,\ell^\theta(Z_i,\theta_0) + o_p(1) \\ &= \mathrm{I}^{-2}\mathrm{Q}_1(\theta_0) + 2\mathrm{I}^{-2}E\left[\ell\ell^\theta\right] + o_p(1) \\ &= 2b(\theta_0) + o_p(1).\end{aligned}$$

It also follows that $E^*\left[\hat\vartheta'''(0)\right] = O_p\left(n^{-1/2}\right)$ by the same arguments. Therefore

$$E^*\left[\hat\theta^*_{aa}\right] = \frac{b(\theta_0)}{n} + o_p\left(n^{-1}\right),$$

which establishes the result. ■

Proof of Proposition 4. First note that $E^*\left[h_n\left(\hat\theta^* - \hat\theta\right)\right] = E^*\left[\hat\theta^*_{aa}\right] + o_p\left(n^{-3/2}\right)$ by Theorem 3. It follows that

$$\begin{aligned}\sqrt{n}\left(\hat\theta - E^*\left[h_n\left(\hat\theta^*-\hat\theta\right)\right] - \theta_0\right) &= \sqrt{n}\left(\hat\theta - E^*[\theta^*_{aa}] - \theta_0\right) + \sqrt{n}\left(E^*[\theta^*_{aa}] - E^*\left[h_n\left(\hat\theta^*-\hat\theta\right)\right]\right) \\ &= \sqrt{n}\left(\hat\theta - E^*[\theta^*_{aa}] - \theta_0\right) + o_p(n^{-1}).\end{aligned}$$

and

$$E^*\left[\hat\theta^{\epsilon\epsilon\epsilon}(0)\right] = \hat{\mathcal{I}}^{-4}Q_2\left(\hat\theta\right)\left(\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{j=1}\ell\left(Z_i,\hat\theta\right)^3\right) + 3\hat{\mathcal{I}}^{-5}Q_1\left(\hat\theta\right)^2\left(\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{j=1}\ell\left(Z_i,\hat\theta\right)^3\right)$$
$$+9\hat{\mathcal{I}}^{-4}Q_1\left(\hat\theta\right)\left(\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{j=1}\ell\left(Z_i,\hat\theta\right)^2\left(\ell^{\theta}\left(Z_i,\hat\theta\right)-\overline{\ell^{\theta}\left(\cdot,\hat\theta\right)}\right)\right)$$
$$+3\hat{\mathcal{I}}^{-3}\left(\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{j=1}\ell\left(Z_i,\hat\theta\right)^2\left(\ell^{\theta\theta}\left(Z_i,\hat\theta\right)-\overline{\ell^{\theta\theta}\left(\cdot,\hat\theta\right)}\right)\right)$$
$$+6\hat{\mathcal{I}}^{-3}\left(\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{j=1}\ell\left(Z_i,\hat\theta\right)\left(\ell^{\theta}\left(Z_i,\hat\theta\right)-\overline{\ell^{\theta}\left(\cdot,\hat\theta\right)}\right)^2\right)$$
$$= O_p\left(n^{-1/2}\right).$$

Here,

$$\sqrt{n}B_n \equiv -3\mathcal{I}^{-3}\sqrt{n}\left(\hat{\mathcal{I}}-\mathcal{I}\right)Q_1(\theta_0)$$
$$+\mathcal{I}^{-2}\sqrt{n}\left(Q_1\left(\hat\theta\right)-Q(\theta_0)\right)$$
$$+\mathcal{I}^{-3}Q_1(\theta_0)\left(\frac{1}{\sqrt{n}}\sum_{i=1}\left[\ell(Z_i,\theta_0)^2-E\left[\ell(Z_i,\theta_0)^2\right]\right]+\sqrt{n}\left[\hat m_1\left(\hat\theta\right)-\hat m_1(\theta_0)\right]\right)$$
$$-4\mathcal{I}^{-3}\sqrt{n}\left(\hat{\mathcal{I}}-\mathcal{I}\right)E\left[\ell(Z_i,\theta_0)\ell^{\theta}(Z_i,\theta)\right]$$
$$+2\mathcal{I}^{-2}\left(\sqrt{n}\left[\hat m_3\left(\hat\theta\right)-\hat m_3(\theta_0)\right]\right)$$
$$+\sqrt{n}\left(n^{-1}\sum_{i=1}\ell(Z_i,\theta_0)\ell^{\theta}(Z_i,\theta_0)-E\left[\ell(Z_i,\theta_0)\ell^{\theta}(Z_i,\theta_0)\right]\right)$$

It follows that

$$E^*\left[\hat\theta^*_{aa}\right] = \frac{1}{2}\frac{1}{n}\left[\mathcal{I}^{-2}Q_1(\theta_0)+2\mathcal{I}^{-2}E\left[\ell(Z_i,\theta_0)\ell^{\theta}(Z_i,\theta_0)\right]\right] + \frac{1}{2}\frac{1}{n^{3/2}}\left(\sqrt{n}B_n\right)+O_p\left(n^{-2}\right). \tag{42}$$

Using Lemma 8, we obtain

$$\sqrt{n}B_n = B + o_p(1), \tag{43}$$

where

$$B \equiv 3\mathcal{I}^{-4}Q_1(\theta_0)^2U(\theta_0)+\mathcal{I}^{-3}Q_2(\theta_0)U(\theta_0)+6\mathcal{I}^{-4}Q_1(\theta_0)E\left[U_i(\theta_0)V_i(\theta_0)\right]U(\theta_0)$$

Combining (42) and (43), we obtain

$$
\begin{aligned}
&\left(\hat{\theta} - E^{*}\left[h_n\left(\hat{\theta}^{*} - \hat{\theta}\right)\right] - \theta_0\right) \\
&= \left(\hat{\theta} - \theta_0\right) \\
&\quad - \frac{1}{2}\frac{1}{n}\left[I^{-2}Q_1(\theta_0) + 2I^{-2}E\left[\ell(Z_i,\theta_0)\,\ell^{\theta}(Z_i,\theta_0)\right]\right] \\
&\quad - \frac{1}{2}\frac{1}{n^{3/2}}B + o_p\left(\frac{1}{n^{3/2}}\right),
\end{aligned}
$$

from which the conclusion follows. ■

Proof of Proposition 5. Write $\theta^{\epsilon} = \theta^{\epsilon}(0)$, etc, for notational simplicity. Because

$$
\begin{aligned}
\hat{\theta} &= \theta_0 + \theta^{\epsilon} + \frac{1}{2}\frac{1}{n}\theta^{\epsilon\epsilon} + \frac{1}{6}\frac{1}{n\sqrt{n}}\theta^{\epsilon\epsilon\epsilon} \\
&\quad + \frac{1}{24}\frac{1}{n^2}\theta^{\epsilon\epsilon\epsilon\epsilon} + \frac{1}{120}\frac{1}{n^2\sqrt{n}}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon} + \frac{1}{720}\frac{1}{n^3}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}(\tilde{\epsilon}),
\end{aligned}
$$

we should have

$$
\begin{aligned}
\hat{\theta}_{(j)} &= \theta_0 + \theta^{\epsilon}_{(j)} + \frac{1}{2}\frac{1}{n-1}\theta^{\epsilon\epsilon}_{(j)} + \frac{1}{6}\frac{1}{(n-1)\sqrt{n-1}}\theta^{\epsilon\epsilon\epsilon}_{(j)} \\
&\quad + \frac{1}{24}\frac{1}{(n-1)^2}\theta^{\epsilon\epsilon\epsilon\epsilon}_{(j)} + \frac{1}{120}\frac{1}{(n-1)^2\sqrt{n-1}}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon}_{(j)} + \frac{1}{720}\frac{1}{(n-1)^3}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}_{(j)}\left[\tilde{\epsilon}_{(j)}\right].
\end{aligned}
$$

Therefore,

$$\sqrt{n}\left(\tilde{\theta}-\theta_0\right) = \sqrt{n}\left(n\hat{\theta} - \frac{n-1}{n}\sum_{j=1}^{n}\hat{\theta}_{(j)} - \theta_0\right)$$

$$= n\left\{\sqrt{n}\left(\hat{\theta}-\theta_0\right) - \sqrt{\frac{n}{n-1}}\frac{1}{n}\sum_{j=1}^{n}\sqrt{n-1}\left(\hat{\theta}_{(j)}-\theta_0\right)\right\}$$

$$+ \sqrt{\frac{n}{n-1}}\frac{1}{n}\sum_{j=1}^{n}\sqrt{n-1}\left(\hat{\theta}_{(j)}-\theta_0\right)$$

$$= n\left\{\theta' + \frac{1}{2}\frac{1}{\sqrt{n}}\theta'' + \frac{1}{6}\frac{1}{n}\theta'''\right.$$

$$\left. - \sqrt{\frac{n}{n-1}}\frac{1}{n}\sum_{j=1}^{n}\left(\theta'_{(j)} + \frac{1}{2}\frac{1}{\sqrt{n-1}}\theta''_{(j)} + \frac{1}{6}\frac{1}{n-1}\theta'''_{(j)}\right)\right\}$$

$$+ \sqrt{\frac{n}{n-1}}\frac{1}{n}\sum_{j=1}^{n}\left(\theta'_{(j)} + \frac{1}{2}\frac{1}{\sqrt{n-1}}\theta''_{(j)} + \frac{1}{6}\frac{1}{n-1}\theta'''_{(j)}\right)$$

$$+ n\left\{\frac{1}{24}\frac{1}{n\sqrt{n}}\theta'''' + \frac{1}{120}\frac{1}{n^2}\theta'''''\right.$$

$$\left. - \sqrt{\frac{n}{n-1}}\frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{24}\frac{1}{(n-1)\sqrt{n-1}}\theta''''_{(j)} + \frac{1}{120}\frac{1}{(n-1)^2}\theta'''''_{(j)}\right)\right\}$$

$$+ \sqrt{\frac{n}{n-1}}\frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{24}\frac{1}{(n-1)\sqrt{n-1}}\theta''''_{(j)} + \frac{1}{120}\frac{1}{(n-1)^2}\theta'''''_{(j)}\right)$$

$$+ n\left\{\frac{1}{720}\frac{1}{n^2\sqrt{n}}\theta''''''(\mathbf{e}) - \sqrt{\frac{n}{n-1}}\frac{1}{n}\sum_{j=1}^{n}\frac{1}{720}\frac{1}{(n-1)^2\sqrt{n-1}}\theta''''''_{(j)}(\mathbf{e})\right\}$$

$$+ \sqrt{\frac{n}{n-1}}\frac{1}{n}\sum_{j=1}^{n}\frac{1}{720}\frac{1}{(n-1)^2\sqrt{n-1}}\theta''''''_{(j)}(\mathbf{e})$$

or

$$\sqrt{n}\left(\tilde{\theta}-\theta_0\right)$$

$$= \left(n\theta' - \sqrt{n}\sqrt{n-1}\frac{1}{n}\sum_{j=1}^{n}\theta'_{(j)}\right) + \frac{1}{2}\frac{1}{\sqrt{n}}\left(n\theta'' - \sum_{j=1}^{n}\theta''_{(j)}\right) + \frac{1}{6}\frac{1}{n}\left(n\theta''' - \sqrt{\frac{n}{n-1}}\sum_{j=1}^{n}\theta'''_{(j)}\right)$$

$$+ \frac{1}{24}\frac{1}{\sqrt{n}}\left(n\theta'''' - \frac{n^2}{\text{[illegible]}}\frac{1}{\text{[illegible]}}\sum_{j=1}^{n}\theta''''_{(j)}\right) + \frac{1}{120}\frac{1}{\text{[illegible]}}n\left(\theta''''' - \frac{n\sqrt{n}}{\text{[illegible]}}\frac{1}{\text{[illegible]}}\sum_{j=1}^{n}\theta'''''_{(j)}\right)$$

for every υ such that $\upsilon < \frac{1}{48}$. In particular, we have

$$\frac{1}{\sqrt{n}}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}(\mathbf{e}) = o_p(1) \tag{46}$$

By Lemma 7 again, we obtain

$$\begin{aligned}
\Pr\left[\left|\frac{1}{n^{\frac{1}{2}-6\upsilon}}\frac{1}{n}\sum_{j=1}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}_{(j)}\left(\mathbf{e}_{(j)}\right)\right| > C\right] &\leq \sum_{j=1}\Pr\left\{\left|\frac{1}{n^{\frac{1}{2}-6\upsilon}}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}_{(j)}\left(\mathbf{e}_{(j)}\right)\right| > C\right\} \\
&\leq \sum_{j=1}\Pr\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{n}}}\left|\frac{1}{n^{\frac{1}{2}-6\upsilon}}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}_{(j)}(\epsilon)\right| > C\right] \\
&= n\Pr\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{n}}}\left|\frac{1}{n^{\frac{1}{2}-6\upsilon}}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}_{(j)}(\epsilon)\right| > C\right] \\
&= o(1)
\end{aligned}$$

Here, the first equality is based on the fact that Z_i are i.i.d., so that $\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}_{(j)}(\epsilon)$ are identically distributed for $j = 1, \ldots, n$. In particular, we have

$$\frac{1}{(n-1)\sqrt{n}}\sum_{j=1}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}_{(j)}\left(\mathbf{e}_{(j)}\right) = o_p(1) \tag{47}$$

Combining (46) and (47), we obtain

$$\frac{1}{720}\frac{1}{\sqrt{n}}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}(\mathbf{e}) - \frac{1}{720}\frac{1}{(n-1)^2\sqrt{n}}\sum_{i=1}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon\epsilon}_{(j)}\left(\mathbf{e}_{(j)}\right) = o_p\left(\frac{1}{n}\right) \tag{48}$$

Note that $\theta^{\epsilon\epsilon\epsilon\epsilon}$ is a sum of V-statistic of order 4 as considered in Lemma 20. Likewise, $\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon}$ is a sum of V-statistic of order 5 as considered in Lemma 21. Therefore, combining (38) and (39) with Lemmas 20 and 21, we obtain

$$n\theta^{\epsilon\epsilon\epsilon\epsilon} - \frac{n^2}{n-1}\frac{1}{n}\sum_{j=1}\theta^{\epsilon\epsilon\epsilon\epsilon}_{(j)} = O_p(1)$$

$$n\left(\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon} - \frac{n\sqrt{n}}{(n-1)\sqrt{n-1}}\frac{1}{n}\sum_{j=1}\theta^{\epsilon\epsilon\epsilon\epsilon\epsilon}_{(j)}\right) = O_p(1)$$

from which we further obtain

$$\frac{1}{}\frac{1}{\sqrt{n}}\left(n\theta^{\epsilon\epsilon\epsilon\epsilon} - \frac{n^2}{}\frac{1}{}\sum\theta_{(j)}^{\epsilon\epsilon\epsilon\epsilon}\right) = o_p\left(\frac{1}{}\right) \tag{49}$$

Combining (4), (5), (6) with Lemmas 17, 18, 19, we obtain

$$\begin{aligned}
&\left(n\theta' - \sqrt{n}\sqrt{n-1}\,\frac{1}{n}\sum_{j=1}^{n}\theta'_{(j)}\right) + \frac{1}{2}\frac{1}{\sqrt{n}}\left(n\theta'' - \sum_{j=1}^{n}\theta''_{(j)}\right) \\
&\quad + \frac{1}{6}\frac{1}{n}\left(n\theta''' - \sqrt{\frac{n}{n-1}}\sum_{j=1}^{n}\theta'''_{(j)}\right) \\
&= \theta' + \frac{1}{2}\frac{1}{\sqrt{n}}\Big\{ I^{-3}E\left[\ell^{\theta\theta}\right]\left(U(\theta_0)^2 - E\left[U_i(\theta_0)^2\right]\right) \\
&\quad + \frac{2}{I^2}\left(U(\theta_0)V(\theta_0) - E\left[U_i(\theta_0)V_i(\theta_0)\right]\right)\Big\} \\
&\quad + \frac{1}{6}\frac{1}{n}\theta''' - \frac{1}{2}\frac{1}{n}J + o_p\left(\frac{1}{n}\right)
\end{aligned} \tag{51}$$

Combining (45) with (48), (50), (49), and (51), we obtain

$$\begin{aligned}
\sqrt{n}\left(\hat{\theta} - \theta_0\right) &= \theta' + \frac{1}{2}\frac{1}{\sqrt{n}}\Big\{ I^{-3}E\left[\ell^{\theta\theta}\right]\left(U(\theta_0)^2 - E\left[U_i(\theta_0)^2\right]\right) \\
&\quad + \frac{2}{I^2}\left(U(\theta_0)V(\theta_0) - E\left[U_i(\theta_0)V_i(\theta_0)\right]\right)\Big\} \\
&\quad + \frac{1}{6}\frac{1}{n}\theta''' - \frac{1}{2}\frac{1}{n}J + o_p\left(\frac{1}{n}\right)
\end{aligned}$$

where

$$\begin{aligned}
J &= \left(I^{-3}E\left[\ell^{\theta\theta\theta}\right] + 3I^{-4}E\left[\ell^{\theta\theta}\right]^2\right)U(\theta_0) + 6I^{-4}E\left[\ell^{\theta\theta}\right]E\left[U_iV_i\right]U(\theta_0) \\
&\quad + \frac{3}{I^3}E\left[\ell^{\theta\theta}\right]V(\theta_0) + \frac{2}{I^3}E\left[U_iW_i\right]U(\theta_0) + \frac{1}{I^2}W(\theta_0) \\
&\quad + \frac{2}{I^3}E\left[\ell^{\theta}(Z_i,\theta_0)^2\right]U(\theta_0) + \frac{4}{I^3}E\left[U_iV_i\right]V(\theta_0) \\
&\quad + I^{-3}E\left[\ell^{\theta\theta}\right]n^{-1/2}\left(\sum_{i=1}^{n}\left[\ell(Z_i,\theta_0)^2 - E\left[\ell(Z_i,\theta_0)^2\right]\right]\right) \\
&\quad + 2I^{-2}n^{-1/2}\left(\sum_{i=1}^{n}\ell(Z_i,\theta_0)\,\ell^{\theta}(Z_i,\theta_0) - E\left[\ell(Z_i,\theta_0)\,\ell^{\theta}(Z_i,\theta_0)\right]\right)
\end{aligned} \tag{52}$$

∎

Proof of Theorem (1). Because $\hat{\theta}$ is an efficient estimator of θ_0 it follows that $b\left(\hat{\theta}\right)$ is an efficient estimator of $b(\theta_0)$. Denote the limit law of $\sqrt{n}\left(b(\hat{\theta}) - b(\theta_0)\right)$ by L. By the convolution theorem $\sqrt{n}\left(b_n - b(\theta_0)\right) \rightsquigarrow L + W$ where W is independent of L and $\rightsquigarrow$ denotes weak convergence. This can only

Now, note that we have the expansions

$$\sqrt{n}\left(\hat{\theta} - b\left(\hat{\theta}\right)/n - \theta_0\right) = \theta^{\epsilon}(0) + \frac{1}{\sqrt{n}}\left(\frac{1}{2}\theta^{\epsilon\epsilon}(0) - b(\theta_0)\right) + \frac{1}{n}\left(\frac{1}{6}\theta^{\epsilon\epsilon\epsilon}(0) - \text{¨}\, n^{-1/2}\sum_{i=1}\psi(Z_i,\theta_0)\right) + O_p\left(\frac{1}{n\sqrt{n}}\right)$$

and

$$\sqrt{n}\left(\hat{\theta} - \frac{b_n}{n} - \theta_0\right) = \theta^{\epsilon}(0) + \frac{1}{\sqrt{n}}\left(\frac{1}{2}\theta^{\epsilon\epsilon}(0) - b(\theta_0)\right) + \frac{1}{n}\left(\frac{1}{6}\theta^{\epsilon\epsilon\epsilon}(0) - n^{-1/2}\sum_{i=1}\varrho(Z_i,\theta_0)\right) + O_p\left(\frac{1}{n\sqrt{n}}\right)$$

Because $\theta^{\epsilon}(0) = n^{-1/2}\sum_{i=1}^{n}\psi(Z_i,\theta_0)$, equation (53) implies that covariances of the "adjustment terms" of order n^{-1} with $\theta^{\epsilon}(0)$ are equal to each other. ■

Proof of Theorem 2. An expansion of $b\left(\hat{\theta}\right)$ gives

$$\begin{aligned} b\left(\hat{\theta}\right) - b(\theta_0) &= \tau_m\left(\left.\frac{\partial\left[\int m(z,\theta)f(z,\theta)\,dz\right]}{\partial\theta'}\right|_{\theta=\theta_0}\right)\Gamma^{-1}U(\theta_0) + O_p\left(n^{-1}\right) \\ &= \tau_m(M+\Delta)\Gamma^{-1}U(\theta_0) + O_p\left(n^{-1}\right). \end{aligned}$$

A similar expansion for $\hat{b}\left(\hat{\theta}\right)$ gives

$$\begin{aligned} \hat{b}\left(\hat{\theta}\right) - b(\theta_0) &= \hat{b}\left(\hat{\theta}\right) - \hat{b}(\theta_0) + \hat{b}(\theta_0) - b(\theta_0) \\ &= \tau_m\left(\frac{1}{n}\sum_i m\left(z_i,\hat{\theta}\right) - \frac{1}{n}\sum_i m(z_i,\theta_0)\right) \\ &\quad + \tau_m\left(\frac{1}{n}\sum_i m(z_i,\theta_0) - E[m(z_i,\theta_0)]\right) + O_p\left(n^{-1}\right) \\ &= \tau_m M\Gamma^{-1}U(\theta_0) + \tau_m\left(n^{-1}\sum_i\left(m(z_i,\theta_0) - \overline{m}\right)\right) + O_p\left(n^{-1}\right). \end{aligned}$$

Plugging these expansions into that for $\hat{\theta}$ gives

$$\begin{aligned} \sqrt{n}\left(\hat{\theta}_c - \theta_0\right) &= \sqrt{n}\left(\hat{\theta} - \theta_0\right) - \frac{1}{\sqrt{n}}b\left(\hat{\theta}\right) \\ &= \Gamma^{-1}U(\theta_0) + \frac{1}{\sqrt{n}}\left(\frac{1}{2}\theta^{\epsilon\epsilon}(0) - b(\theta_0)\right) \\ &\quad + \frac{1}{n}\left(\frac{1}{6}\theta^{\epsilon\epsilon\epsilon}(0) - \tau_m(M+\Delta)\Gamma^{-1}U(\theta_0)\right) + O_p\left(\frac{1}{n^{3/2}}\right) \end{aligned}$$

Also,

$$
\begin{aligned}
& E\left(M I^{-1}U(\theta_0) + n^{-1/2}\sum_i \left(m(z_i,\theta_0) - \overline{m}\right) \cdot \left[I^{-1}U(\theta_0)\right]'\right) \\
&= M I^{-1} E\left[U(\theta_0)U(\theta_0)'\right] I^{-1} + E\left(n^{-1/2}\sum_i \left(m(z_i,\theta_0) - \overline{m}\right) U(\theta_0)'\right) I^{-1} \\
&= (M + ¤)\, I^{-1}.
\end{aligned}
$$

■

Proof of Theorem 3. The asymptotic bias of the MLE is equal to

$$\frac{b(\theta_0)}{n}$$

where

$$b(\theta) = \frac{1}{2}E[\theta^{\epsilon\epsilon}] = \frac{1}{2I_\theta^2}E_\theta\left[\ell^{\theta\theta}\right] + \frac{1}{I_\theta^2}E_\theta\left[\ell\ell^\theta\right]$$

To show that $\frac{1}{2}E[\mathrm{B}\theta^\epsilon(0)] = \tau_m (M + ¤)\, I^{-1}$, it suffices to prove that $E[\mathrm{B}U(\theta_0)] = 2\tau_m (M + ¤)$. We first note that

$$
\begin{aligned}
E[\mathrm{B}U(\theta_0)] \;=\;& 6I^{-3}Q_1(\theta_0)E\left[\ell(Z_i,\theta_0)\ell^\theta(Z_i,\theta_0)\right] + 2I^{-3}Q_1(\theta_0)^2 \\
& + 4I^{-3}\left[E\left[\ell(Z_i,\theta_0)\ell^\theta(Z_i,\theta_0)\right]\right]^2 + I^{-2}E\left[\ell(Z_i,\theta_0)\ell^{\theta\theta}(Z_i,\theta_0)\right] + I^{-2}Q_2(\theta_0)^2 \\
& + 2I^{-2}\left\{E\left[\ell^\theta(Z_i,\theta_0)^2\right] + E\left[\ell(Z_i,\theta_0)\ell^{\theta\theta}(Z_i,\theta_0)\right] + E\left[\ell(Z_i,\theta_0)^2\ell^\theta(Z_i,\theta_0)\right]\right\},
\end{aligned}
$$

where we have used $E\left[\ell(Z_i,\theta_0)^3\right] = -E\left[\ell^{\theta\theta}(Z_i,\theta_0)\right] - 3E\left[\ell(Z_i,\theta_0)\ell^\theta(Z_i,\theta_0)\right]$. In order to provide an alternative characterization of $2\tau_m(M + ¤)$, we note that

$$
\begin{aligned}
M &= \left(2E\left[\ell(z,\theta)\ell^\theta(Z_i,\theta_0)\right], E\left[\ell^{\theta\theta\theta}(z,\theta)\right], E\left[\ell^\theta(z,\theta)^2 + \ell(z,\theta)\ell^{\theta\theta}(z,\theta)\right]\right)' \\
&= \left(2E\left[\ell(z,\theta)\ell^\theta(Z_i,\theta_0)\right], Q_2(\theta), E\left[\ell^\theta(z,\theta)^2\right] + E\left[\ell(z,\theta)\ell^{\theta\theta}(z,\theta)\right]\right)',
\end{aligned}
$$

$$
\begin{aligned}
¤ &= \left(E\left[\ell(z,\theta)^3\right], E\left[\ell^{\theta\theta}(z,\theta)\ell(z,\theta)\right], E\left[\ell(z,\theta)^2\ell^\theta(z,\theta)\right]\right)' \\
&= \left(-Q_1(\theta) - 3E\left[\ell(Z_i,\theta_0)\ell^\theta(Z_i,\theta_0)\right], E\left[\ell(z,\theta)\ell^{\theta\theta}(z,\theta)\right], E\left[\ell(z,\theta)^2\ell^\theta(z,\theta)\right]\right)'.
\end{aligned}
$$

and

$$\tau_m = \left(-\frac{E\left[\ell^{\theta\theta}(Z_i,\theta_0)\right] + 2E\left[\ell(Z_i,\theta_0)\ell^\theta(Z_i,\theta_0)\right]}{\left[E\left[\ell(z,\theta_0)^2\right]\right]^3}, \frac{1}{\left[2E\left[\ell(z,\theta_0)^2\right]\right]^2}, \frac{1}{\left[E\left[\ell(z,\theta_0)^2\right]\right]^2}\right)'$$

It follows that

$$
\begin{aligned}
2\tau_m(M+\Delta) &= 2\frac{Q_1(\theta)+2E\left[\ell(Z_i,\theta_0)\,\ell^{\theta}(Z_i,\theta_0)\right]}{I^3}\left(Q_1(\theta)+E\left[\ell(Z_i,\theta_0)\,\ell^{\theta}(Z_i,\theta_0)\right]\right) \\
&\quad+\frac{1}{I^2}\left(Q_2(\theta)+E\left[\ell(z,\theta)\,\ell^{\theta\theta}(z,\theta)\right]\right) \\
&\quad+\frac{2}{I^2}\left(E\left[\ell^{\theta}(z,\theta)^2\right]+E\left[\ell(z,\theta)\,\ell^{\theta\theta}(z,\theta)\right]+E\left[\ell(z,\theta)^2\,\ell^{\theta}(z,\theta)\right]\right) \\
&= 2I^{-3}Q_1(\theta)^2+6I^{-3}Q_1(\theta)\,E\left[\ell(Z_i,\theta_0)\,\ell^{\theta}(Z_i,\theta_0)\right] \\
&\quad+4I^{-3}\left(E\left[\ell(Z_i,\theta_0)\,\ell^{\theta}(Z_i,\theta_0)\right]\right)^2+I^{-2}Q_2(\theta)+I^{-2}E\left[\ell(z,\theta)\,\ell^{\theta\theta}(z,\theta)\right] \\
&\quad+2I^{-2}\left(E\left[\ell^{\theta}(z,\theta)^2\right]+E\left[\ell(z,\theta)\,\ell^{\theta\theta}(z,\theta)\right]+E\left[\ell(z,\theta)^2\,\ell^{\theta}(z,\theta)\right]\right) \\
&= E\left[BU(\theta_0)\right]
\end{aligned}
$$

∎

Bibliography

[1] Arcones, M. A. and E. Gine. (1992), "On the bootstrap of M-estimators and other statistical functionals" in Raoul LePage and Lynne Billard eds Exploring the Limits of Bootstrap. Wiley: New York.

[2] Akahira, M. (1983), "Asymptotic Deficiency of the Jackknife Estimator," Australian J. Statistics 25, 123-129.

[3] Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and J. Wellner (1993), Efficient and Adaptive Estimation for Semiparametric Models, Johns Hopkins University Press: Baltimore.

[4] Gine, E. and Zinn, J. (1990), "Bootstrapping General Empirical Measures," Annals of Probability 18, 851-869.

[5] Guldberg, S. (1935), "Recurrence formulae for the semi-invariants of some discontinuous frequency functions of n variables," Skandinavisk Aktuarietidskrift 18, 270-278.

[6] Gusev, S.I. (1975), "Asymptotic Expansions Associated with some Statistical Estimators in the Smooth Case I. Expansions of Random Variables," Theory of Probability and its Applications 20, 470-498.

[7] Gusev, S.I. (1976), "Asymptotic Expansions Associated with some Estimators in the Smooth Case II. Expansions of Moments and Distributions," Theory of Probability and its Applications 21, 14-33.

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

[9] Lahiri, S. N. (1992), "Edgeworth Correction by 'Moving Block' Bootstrap for Stationary and Non-stationary Data" in Raoul LePage and Lynne Billard eds Exploring the Limits of Bootstrap. Wiley: New York.

[10] Lee, A.J. (1990), U-Statistics, Theory and Practice, Marcel Dekker: New York and Basel.

[11] Pfanzagl, J., and W. Wefelmeyer (1978), "A Third-Order Optimum Property of the Maximum

[14] Shiryaev, A.N. (1989), Probability, Springer Verlag: New York.

[15] van der Vaart, A. and J. Wellner. (1996), Weak Convergence and Empirical Processes with Applications to Statistics, Springer Verlag: New York.

[16] Wishart, J. (1947), "Cumulants of Multivariate Multinomial Distributions," Biometrika 36, pp. 47 - 58.