

ROBUST INFERENCE UNDER MOMENT RESTRICTIONS

YUICHI KITAMURA AND TAISUKE OTSU

ABSTRACT. Suppose one wishes to test a parametric hypothesis, or to form a confidence interval for a parameter, in a moment condition model. If the observations may be subject to measurement errors or other data contamination problems, the validity of conventional procedures is called into question. This paper demonstrates that a test based on a minimum Hellinger distance estimator (MHDE) possesses desirable optimal robust properties, which make it suitable for such a situation. First, it is asymptotically minimax optimal, in a general class of tests, in terms of type I error probabilities: its worst case size distortion is asymptotically minimal when the probability law of the data is perturbed within infinitesimal neighborhoods. Second, the local power of the Hellinger distance-based test is optimal in a minimax robust criterion.

1. INTRODUCTION

Consider a probability measure $P_0 \in \mathcal{M}$, where \mathcal{M} is the set of all probability measures on the Borel σ -field $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ of $\mathcal{X} \subseteq \mathbb{R}^d$. Let $g : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^m$ be a vector of functions, which satisfies the moment condition:

$$(1.1) \quad E_{P_0} [g(x, \theta_0)] = \int g(x, \theta_0) dP_0 = 0, \quad \theta_0 \in \Theta,$$

and $\Theta \subseteq \mathbb{R}^p$ is the parameter space. Suppose the model (1.1) is overidentified, that is, $m > p$.

The focus of this paper is the issue of robustness when one wishes to test

$$(1.2) \quad H_0 : \tau(\theta_0) = 0 \quad \text{against} \quad H_1 : \tau(\theta_0) \neq 0$$

for a known function $\tau : \Theta \rightarrow \mathbb{R}$.¹ Define $\Theta_0 = \{\theta \in \Theta | \tau(\theta) = 0\}$, and we assume that $\Theta_0 \subset \text{int}(\Theta)$ in order to avoid the technicalities associated with the inference at the boundary of Θ . This simplifies our presentation. If observations drawn from P_0 or its local alternative are available, existing tests as reviewed, for example, in Newey and McFadden (1994) work well according to the standard asymptotic theory. One may however wish to consider testing (1.2) when observations are potentially subject to measurement error or other data

Date: This Version: May 31, 2010.

Keywords: Asymptotic Minimax Theorem, Hellinger Distance, Size Distortion.

JEL Classification Number: C14.

We acknowledge financial support from the National Science Foundation via grants SES-0851759 (Kitamura) and SES-0720961 (Otsu). We are grateful for helpful discussions with Don Andrews.

¹Although we consider the two-sided alternative hypothesis H_1 , we can derive an analogous result for the one-sided alternatives ($H'_1 : \tau(\theta_0) > 0$ or $H''_1 : \tau(\theta_0) < 0$).

contamination problems. This paper addresses this issue by developing an asymptotic robust testing theory for the moment condition model (1.1). Following the literature of robust statistics (see, for example, Bickel (1981) and Rieder (1994)) we formulate our theory of robustness by introducing infinitesimal perturbations of the measure P_0 (for the analysis of size) or its local alternatives (for the analysis of power). Two main results are discussed in the subsequent sections. Our first result shows that a certain test possesses an asymptotic minimax robust property in terms of size distortion due to perturbation of P_0 . This naturally leads to the next question: Is such extra robustness is gained at a cost of its power? Our second result proves it is not the case, that is, the same test is asymptotically minimax optimal in terms of its type II error probability. The last result also take account of robustness by allowing the probability measures under local alternatives to be subject to perturbation.

We now present a formal setting that is useful in developing our theory of robust testing. Let

$$\mathcal{X}_n = \left\{ x \in \mathcal{X} : \sup_{\theta \in \Theta} |g(x, \theta)| \leq m_n \right\},$$

where $\{m_n\}_{n \in \mathbb{N}}$ is a sequence of positive numbers satisfying $m_n \rightarrow \infty$ as $n \rightarrow \infty$. The set \mathcal{X}_n is used for trimming for technical reasons, and this and the trimming constant m_n are not used in the practical implementation of the test we advocate in this paper. They appear only in the definition of \mathcal{P}_θ and Assumption 2.1 below. In what follows we use the standard notation $E_P(X)$ and $Var_P(X)$ to signify the expectation and the variance of $X \sim P$. Also, if $x_i, i = 1, \dots, n \sim_{iid} P$ for a probability measure P , with a slight abuse of notation we write $E_P[a]$ and $Var_P(a)$ to denote the expectation and the variance of a function a of (x_1, \dots, x_n) . Moreover, we consider IID observations throughout the paper. Define $\dot{g}(x, \theta) = \partial g(x, \theta) / \partial \theta'$.

Definition 1. For each θ in Θ , let \mathcal{P}_θ denote the set of all probability measures that satisfy the following conditions:

[(i)]

- (1) $g(x, \tilde{\theta})$ is continuous over $\tilde{\theta} \in \Theta$ at P -a.e. x ,
- (2) $E_P[g(x, \theta)] = 0$ and $E_P[g(x, \tilde{\theta})] \neq 0$ for all $\tilde{\theta} \neq \theta$,
- (3) $E_P \left[\sup_{\tilde{\theta} \in \Theta} |g(x, \tilde{\theta})|^\eta \right] \leq M$ for some $\eta > 2$,
- (4) $E_P \left[\sup_{\tilde{\theta} \in \mathcal{N}_\theta} |g(x, \tilde{\theta})|^4 \right] < M$,
- (5) $E_P \left[\sup_{\tilde{\theta} \in \mathcal{N}_\theta} |\dot{g}(x, \tilde{\theta})|^2 \right] < M$,
- (6) $\det \left(E_P [\dot{g}(x, \theta)]' E_P [g(x, \theta) g(x, \theta)']^{-1} E_P [\dot{g}(x, \theta)] \right) > c$,
- (7) $\det \left(E_P [g(x, \theta) g(x, \theta)'] \right) > c$,
- (8) $Var_P(g_j(x, \theta)) \in (c, M)$ for $j = 1, \dots, m$,

where $c, M \in (0, \infty)$ are fixed constants (which do not depend on θ or P).

The null hypothesis in (1.2) leads to the following definition:

$$\mathcal{P}^N = \cup_{\theta: \tau(\theta)=0} \mathcal{P}_\theta,$$

for $n \in \mathbb{N}$. A test $\psi_n = \psi_n(x_1, \dots, x_n)$ is defined as a binary function of the sample $\{x_1, \dots, x_n\}$, where $\psi_n = 1$ means rejection and $\psi_n = 0$ means acceptance. The size of the test ψ_n over the set \mathcal{P}^N is given by

$$(1.3) \quad \sup_{Q \in \mathcal{P}^N} E_Q [\psi_n].$$

If a measure P belongs to the \mathcal{P}^N and we observe data drawn from P , then it can be shown that various standard tests mentioned above are able to control the rejection probability $E_P [\psi_n]$ over \mathcal{P}^N asymptotically, under the uniform boundedness conditions in Definition 1. Even if P is in \mathcal{P}^N , if the observations are drawn from a measure Q which is perturbed from P due to, say, measurement errors, then that potentially leads to over-rejection. To account for this, we propose to consider the maximal value of the rejection rate when data are drawn from a measure that is potentially deviated from \mathcal{P}^N . For each $r_0 > 0$, define

$$(1.4) \quad \alpha_{\psi, n} = \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_n],$$

where

$$\mathcal{Q}_n^N = \cup_{P \in \mathcal{P}^N} \left\{ Q \in \mathcal{M} : H(Q, P) \leq \frac{r_0}{\sqrt{n}}, E_Q \left[\sup_{\theta \in \Theta} |g(x, \theta)|^\eta \right] < M \right\},$$

for some $\eta > 2$, and

$$H(Q, P) = \left\{ \int (p^{1/2} - q^{1/2})^2 d\nu \right\}^{1/2} = \left\{ 2 - 2 \int p^{1/2} q^{1/2} d\nu \right\}^{1/2}$$

is the Hellinger distance between two measures Q and P with densities p and q with respect to a dominating measure ν . The Hellinger distance is often used to define robustness neighborhoods in the literature: see, for example, Beran (1977), Bickel (1981) and Rieder (1994). Kitamura, Otsu, and Evdokimov (2009) provide desirable theoretical properties of the Hellinger distance in their robustness analysis. Except for the restriction $H(Q, P) \leq \frac{r_0}{\sqrt{n}}$, we do not impose any parametric structure on Q about how it deviates from P . The maximal rejection rate $\alpha_{\psi, n}$ characterizes the robustness of the size properties of the test ψ_n under local perturbations.

Let $\{P_{\theta+t/\sqrt{n}}\}_{t \in \mathbb{R}^p}$ be a parametric submodel satisfying

- (i): P_θ belongs to \mathcal{P}_θ where $\theta \in \Theta_0$
- (ii): there exists a score function $\eta_\theta \in L_2(P_\theta)$ such that $\int \eta_\theta dP_\theta = 0$ and

$$n \left\| dP_{\theta+t/\sqrt{n}}^{1/2} - dP_\theta^{1/2} - \frac{1}{2\sqrt{n}} t' \eta_\theta dP_\theta^{1/2} \right\|^2 \rightarrow 0,$$

for all $t \in \mathbb{R}^p$.

The local power function along this model is given by $E_{P_{\theta+t/\sqrt{n}}}[\psi_n]$. Instead, consider the minimal power of ψ against a class of perturbed versions of $P_{\theta+t/\sqrt{n}}$:

$$(1.5) \quad \beta_{\psi,n}(P_{\theta+t/\sqrt{n}}) = \inf_{Q \in \mathcal{Q}_n^{LA}(P_{\theta+t/\sqrt{n}})} E_Q[\psi_n],$$

where

$$\mathcal{Q}_n^{LA}(P) = \left\{ Q \in \mathcal{M} : H(Q, P) \leq \frac{r_1}{\sqrt{n}}, E_Q \left[\sup_{\theta \in \Theta} |g(x, \theta)|^\eta \right] < M \right\},$$

for $r_1 > 0$. In contrast to \mathcal{Q}_n^N in the definition of $\alpha_{\psi,n}$, the set $\mathcal{Q}_n^{LA}(P_{\theta+t/\sqrt{n}})$ consists of local deviations from the measure $P_{\theta+t/\sqrt{n}}$ (as opposed to the *set* of measures \mathcal{P}^N). It is of great interest to obtain a minimax test that maximizes the minimal power against the set of perturbed versions of $P_{\theta+t/\sqrt{n}}$, at least asymptotically. In the next section we demonstrate that a test, which is essentially a Wald test based on the minimum Hellinger distance estimator (MHDE) of θ achieves asymptotic optimality for both $\alpha_{\psi,n}$ and $\beta_{\psi,n}(P_{\theta+t/\sqrt{n}})$. The test minimizes the size distortion α asymptotically within a class of tests that includes standard parametric testing procedures. This optimal robustness for different tests and seeks an asymptotically optimal test based on some optimality criteria using $\alpha_{\psi,n}$ and $\beta_{\psi,n}(P_{\theta+t/\sqrt{n}})$.

The current paper is related to the existing research on robust testing, though the literature is mainly concerned with parametric models: see, for example, Huber and Strassen (1973), Rieder (1978), and Beran (1981).

We close this section by introducing our proposed test. Let \hat{P}_n denote the empirical measure of observations $\{x_i\}_{i=1}^n$. As noted in Kitamura, Otsu, and Evdokimov (2009), the MHDE of Beran (1977) applied to the moment condition model (1.1) yields

$$(1.6) \quad \hat{\theta} = \arg \min_{\theta \in \Theta} \inf_{P: \int g(x, \theta) dP = 0, P \ll \hat{P}_n} H(P, \hat{P}_n) = \arg \min_{\theta \in \Theta} \max_{\gamma \in \mathbb{R}^m} -\frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \gamma' g(x_i, \theta)},$$

where the second equality follows from the convex duality theory (Kitamura (2006)). In practice, we use the last expression in (1.6) to implement the MHDE. If the data are generated from a measure $P_0 \in \mathcal{P}_{\theta_0}$, $\theta_0 \in \Theta$, it is known that (see, Newey and Smith (2004))

$$(1.7) \quad \sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, \left(E_{P_0}[\dot{g}(x, \theta_0)]' E_{P_0}[g(x, \theta_0)g(x, \theta_0)']^{-1} E_{P_0}[\dot{g}(x, \theta_0)]\right)^{-1}\right).$$

The asymptotic variance can be estimated by

$$\hat{V} = \left(\hat{G}'\hat{\Omega}^{-1}\hat{G}\right)^{-1},$$

where $\hat{G} = \frac{1}{n} \sum_{i=1}^n \dot{g}(x_i, \hat{\theta})$ and $\hat{\Omega} = \frac{1}{n} \sum_{i=1}^n g(x_i, \hat{\theta})g(x_i, \hat{\theta})'$. Define the Hellinger-based Wald test for H_0 against H_1 as follows:

$$(1.8) \quad \psi_{H,n} = \mathbb{I} \left\{ \left(\frac{\sqrt{n}\tau(\hat{\theta})}{\hat{\sigma}} \right)^2 \geq \text{critical value} \right\},$$

where $\mathbb{I}\{\cdot\}$ is the indicator function, $\hat{\sigma} = \dot{\tau}(\hat{\theta})' \hat{V} \dot{\tau}(\hat{\theta})$, and $\dot{\tau}(\theta) = d\tau(\theta)/d\theta$. Based on the asymptotic distribution in (1.7) under P_0 , the critical value is typically set as the $100(1 - \alpha)\%$ critical value of the χ_1^2 distribution. It is known that there are several tests which show the same (first-order) asymptotic properties under the measures P_0 for the null hypothesis and P_n for the a sequence of local alternatives defined on a parametric submodel of $\mathcal{P} = \cup_{\theta \in \Theta} \mathcal{P}_\theta$. We show in the next section that the Hellinger-based Wald test $\psi_{H,n}$ possesses asymptotic optimal robust properties in both criteria $\alpha_{\psi,n}$ and $\beta_{\psi,t,n}$. Thus it works well both in terms of size and power in a situation where robustness is a concern, because of, say, possible measurement errors.

2. MAIN RESULTS

We impose the following assumptions.

Assumption 2.1. *Suppose the following conditions hold:*

- (i): Θ is compact;
- (ii): $\{m_n\}_{n \in \mathbb{N}}$ satisfies $m_n \rightarrow \infty$, $nm_n^{-\eta} \rightarrow 0$, and $n^{-1/2}m_n^{1+\epsilon} = O(1)$ for some $0 < \epsilon < 2$ as $n \rightarrow \infty$;
- (iii): The second-order derivative of τ is bounded on Θ ;
- (iv): For each $\theta \in \Theta_0$, $\sup_{x \in \mathcal{X}_n, \tilde{\theta} \in \mathcal{N}_\theta} \left| \dot{g}(x, \tilde{\theta}) \right| = o(n^{1/2})$ for a neighborhood \mathcal{N}_θ around θ .

2.1. Size Optimality. This section investigates the size properties of tests $\alpha_{\psi,n}$ in (1.4) for a general class of tests. To define the class of tests, we first introduce some notions. Let $\hat{\theta}_a = T_a(\hat{P}_n)$ be a point estimator based on a mapping $T_a : \mathcal{M} \rightarrow \Theta$. The mapping T for the MHDE $\hat{\theta} = T(\hat{P}_n)$ is defined as $T(Q) = \arg \min_{\theta \in \Theta} \inf_{P \in \mathcal{P}_\theta, P \ll Q} H(P, Q)$. We use the following concepts on the mapping T_a .

Definition 2. Let $P_{\theta, \zeta, \bar{P}}$ be a regular parametric submodel at $\bar{P} \in \mathcal{P}_{\bar{\theta}}, \bar{\theta} \in \Theta$ (see, Kitamura, Otsu, and Evdokimov (2009)) such that $P_{\bar{\theta}, 0, \bar{P}} = \bar{P} \in \mathcal{P}_{\bar{\theta}}$.

- (i): The mapping T_a is called **Fisher consistent** if

$$(2.1) \quad \sqrt{n} \left(T_a \left(P_{\bar{\theta}+t/\sqrt{n}, \zeta_n, \bar{P}} \right) - \bar{\theta} \right) \rightarrow t,$$

holds for every submodel sequence $\left\{ P_{\bar{\theta}+t/\sqrt{n}, \zeta_n, \bar{P}} \right\}_{n \in \mathbb{N}}$ (with any $\bar{\theta} \in \Theta$, $t \in \mathbb{R}$, $\zeta_n = O(n^{-1/2})$, and $\bar{P} \in \mathcal{P}_{\bar{\theta}}$) satisfying $P_{\bar{\theta}+t/\sqrt{n}, \zeta_n, \bar{P}} \in \left\{ Q : H(Q, \bar{P}) \leq \frac{r_0}{\sqrt{n}} \right\}$ eventually.

- (ii): The mapping T_a is called **Gaussian regular** if

$$(2.2) \quad \sqrt{n} \left(T_a \left(\hat{P}_n \right) - T_a \left(P_{\bar{\theta}+t/\sqrt{n}, \zeta_n, \bar{P}} \right) \right) \xrightarrow{d} N(0, \bar{V}_a), \quad \text{under } P_{\bar{\theta}+t/\sqrt{n}, \zeta_n, \bar{P}},$$

holds for every submodel sequence $\left\{ P_{\bar{\theta}+t/\sqrt{n}, \zeta_n, \bar{P}} \right\}_{n \in \mathbb{N}}$ (with any $\bar{\theta} \in \Theta$, $t \in \mathbb{R}$, $\zeta_n = O(n^{-1/2})$, and $\bar{P} \in \mathcal{P}_{\bar{\theta}}$) satisfying $P_{\bar{\theta}+t/\sqrt{n}, \zeta_n, \bar{P}} \in \left\{ Q : H(Q, \bar{P}) \leq \frac{r_0}{\sqrt{n}} \right\}$ eventually, where the asymptotic variance \bar{V}_a

satisfies

$$\bar{V}_a \geq \bar{V} = \left(E_{\bar{P}} [\dot{g}(x, \bar{\theta})]' E_{\bar{P}} [g(x, \bar{\theta}) g(x, \bar{\theta})']^{-1} E_{\bar{P}} [\dot{g}(x, \bar{\theta})] \right)^{-1}.$$

(iii): An estimator \hat{V}_a satisfies $\hat{V}_a \xrightarrow{P} \bar{V}_a$ under each sequence $\left\{ P_{\bar{\theta}+t/\sqrt{n}, \zeta_n, \bar{P}} \right\}_{n \in \mathbb{N}}$.

These requirements are weak and satisfied by standard estimators, such as the generalized method of moments, empirical likelihood, and exponential tilting estimators. Since \bar{V} is the semiparametric efficiency bound for the estimation of $\bar{\theta}$ in the model (1.1), the requirement $\bar{V}_a \geq \bar{V}$ on the asymptotic variance in (2.2) is reasonable.

Let z_α and $\chi_{1,\alpha}^2$ be the $100(1-\alpha)\%$ critical values of the standard normal distribution and the χ_1^2 distribution, respectively. The Wald test based on the estimator $\hat{\theta}_a = T_a(\hat{P}_n)$ is defined as

$$\psi_{a,n} = \mathbb{I} \left\{ \left(\frac{\sqrt{n}\tau(\hat{\theta}_a)}{\hat{\sigma}_a} \right)^2 \geq \chi_{1,\alpha}^2 \right\}.$$

where $\hat{\sigma}_a^2 = \hat{\tau}(\hat{\theta}_a)' \hat{V}_a \hat{\tau}(\hat{\theta}_a)$. For the size optimality result, we consider the following class of tests.

Definition 3. A test ψ_n belongs to the class \mathcal{S} of tests (denoted by $\psi_n \in \mathcal{S}$) if

- (i): $\psi_n = \mathbb{I} \{ S_n \geq \chi_{1,\alpha}^2 \}$,
- (ii): for some point estimator $\hat{\theta}_a$, it holds $S_n = \left(\frac{\sqrt{n}\tau(\hat{\theta}_a)}{\hat{\sigma}_a} \right)^2 + o_p(1)$ under each submodel sequence $\left\{ P_{\bar{\theta}+t/\sqrt{n}, \zeta_n, \bar{P}} \right\}_{n \in \mathbb{N}}$ (with any $\bar{\theta} \in \Theta$, $t \in \mathbb{R}$, $\zeta_n = O(n^{-1/2})$, and $\bar{P} \in \mathcal{P}_{\bar{\theta}}$) satisfying $P_{\bar{\theta}+t/\sqrt{n}, \zeta_n, \bar{P}} \in \left\{ Q : H(Q, \bar{P}) \leq \frac{r_0}{\sqrt{n}} \right\}$ eventually,
- (iii): the mapping T_a to define $\hat{\theta}_a = T_a(\hat{P}_n)$ satisfies the requirements in Definition 2.

Note that the class \mathcal{S} contains tests that are asymptotically equivalent to the Wald test based on some estimator $\hat{\theta}_a$ that satisfies the requirements in Definition 2, with the asymptotic critical value $\chi_{1,\alpha}^2$. Therefore, the class \mathcal{S} includes many existing parameter hypothesis tests under moment restrictions, such as the Wald, likelihood ratio-type, or Lagrange multiplier-type test based on the generalized method of moments, empirical likelihood, or exponential tilting estimator or criterion function (see, e.g., Newey and West (1987), Qin and Lawless (1994), Kitamura and Stutzer (1997), Smith (1997), and Imbens, Spady, and Johnson (1998)).

The following theorem demonstrates a size optimality result for the Hellinger-based Wald test $\psi_{H,n}$ in the class of tests \mathcal{S} .

Theorem 4. Let W be a random variable that obeys the non-central χ^2 distribution with degree of freedom 1 and non-centrality $4r_0^2$. Suppose that Assumption 2.1 holds. Then:

- (i): For every test $\psi_n \in \mathcal{S}$ in Definition 3,

$$\liminf_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q[\psi_n] \geq \Pr \{ W \geq \chi_{1,\alpha}^2 \},$$

for each $r_0 > 0$.

(ii): The Hellinger-based Wald test $\psi_{H,n} = \mathbb{I} \left\{ \left(\frac{\sqrt{n}\tau(\hat{\theta})}{\hat{\sigma}} \right)^2 \geq \chi_{1,\alpha}^2 \right\}$ belongs to the class \mathcal{S} and satisfies

$$\lim_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_{H,n}] = \Pr \{W \geq \chi_{1,\alpha}^2\},$$

for each $r_0 > 0$.

2.2. Power Optimality.

2.2.1. *Case I: One-sided H_1 & Scalar θ .* Consider the one-sided alternative hypothesis

$$(2.3) \quad H_0 : \theta_0 = 0 \quad \text{against} \quad H_1 : \theta_0 > 0,$$

which can be interpreted as testing $H_0 : t = 0$ against $H_1 : t > 0$. In this subsection, let

$$\begin{aligned} \mathcal{P}^N &= \mathcal{P}_0, \\ \mathcal{Q}_n^N &= \cup_{P \in \mathcal{P}^N} \left\{ Q \in \mathcal{M} : H(Q, P) \leq \frac{r_0}{\sqrt{n}}, E_Q \left[\sup_{\theta \in \Theta} |g(x, \theta)|^\varphi \right] < M \right\}, \\ P_{t/\sqrt{n}, \eta} &= \text{parametric submodel with score } \eta, \\ \mathcal{Q}_n^{LA}(t, \eta) &= \left\{ Q \in \mathcal{M} : H(Q, P_{t/\sqrt{n}, \eta}) \leq \frac{r_1}{\sqrt{n}}, E_Q \left[\sup_{\theta \in \Theta} |g(x, \theta)|^\varphi \right] < M \right\}, \\ \sigma_\eta^2 &= \left(E_{P_{0,\eta}} [\dot{g}(x, 0)]' E_{P_{0,\eta}} [g(x, 0)g(x, 0)']^{-1} E_{P_{0,\eta}} [\dot{g}(x, 0)] \right)^{-1}. \end{aligned}$$

Note that

$$\begin{aligned} \sqrt{n} (\bar{T}(P_n) - \bar{T}(Q_n)) &\xrightarrow{d} N(0, \sigma_\eta^2), \quad \hat{\sigma} \xrightarrow{P} \sigma_\eta \quad \text{for all sequence } Q_n \in \mathcal{Q}_n^{LA}(0, \eta) \\ \sqrt{n} (\bar{T}(P_n) - \bar{T}(Q_n)) &\xrightarrow{d} N(t, \sigma_\eta^2), \quad \hat{\sigma} \xrightarrow{P} \sigma_\eta \quad \text{for all sequence } Q_n \in \mathcal{Q}_n^{LA}(t, \eta) \end{aligned}$$

(Thus, σ_η^2 is like “ B^{*-1} ” in Choi-Hall-Schick. On the other hand, van der Vaar is more explicit about the dependence on η . Also note that the asymptotic variance for θ using MLE by parametric submodel is $V_\eta = (\int \eta^2 dP_{0,\eta})^{-1}$ which is always smaller than σ_η^2 . So we need to create a least favorable submodel ($P_{t/\sqrt{n}, \eta}^*$ defined in the proof) passing through $P_{0,\eta}$). The Hellinger-based Wald test is optimal in terms of its robust power property as well, as the following result demonstrates.

Theorem 5. (like Choi-Hall-Schick & van der Vaart) Suppose that Assumption 2.1 holds. Then the following results hold for each $r_0, r_1 > 0$ and $t \in \mathbb{R}$ satisfying $z_\alpha + 2r_0 + 2r_1 - \sigma_\eta^{-1}t < 0$ and each $\eta \in L_2(P_{0,\eta})$ satisfying $\int \eta dP_{0,\eta} = 0$.

(i): If a test ψ_n satisfies

$$(2.4) \quad \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_n] \leq \alpha,$$

then

$$\liminf_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t, \eta)} E_Q [1 - \psi_n] \geq \Phi(z_\alpha + 2r_0 + 2r_1 - \sigma_\eta^{-1}t).$$

(ii): The Hellinger-based Wald test $\psi_{H,n} = \mathbb{I} \left\{ \frac{\sqrt{n}\tau(\hat{\theta})}{\hat{\sigma}} \geq z_\alpha + 2r_0 \right\}$ satisfies

$$\limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_{H,n}] \leq \alpha,$$

and

$$\lim_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t, \eta)} E_Q [1 - \psi_{H,n}] = \Phi(z_\alpha + 2r_0 + 2r_1 - \sigma_\eta^{-1}t).$$

2.2.2. Case II: Two-sided H_1 & Scalar θ . Consider the one-sided alternative hypothesis

$$(2.5) \quad H_0 : \theta_0 = 0 \quad \text{against} \quad H_1 : \theta_0 \neq 0,$$

which can be interpreted as testing $H_0 : t = 0$ against $H_1 : t \neq 0$. In this subsection, let

$$\begin{aligned} \mathcal{P}^N &= \mathcal{P}_0, \\ \mathcal{Q}_n^N &= \cup_{P \in \mathcal{P}^N} \left\{ Q \in \mathcal{M} : H(Q, P) \leq \frac{r_0}{\sqrt{n}}, E_Q \left[\sup_{\theta \in \Theta} |g(x, \theta)|^\varphi \right] < M \right\}, \\ P_{t/\sqrt{n}, \eta} &= \text{parametric submodel with score } \eta, \\ \mathcal{Q}_n^{LA}(t, \eta) &= \left\{ Q \in \mathcal{M} : H(Q, P_{t/\sqrt{n}, \eta}) \leq \frac{r_1}{\sqrt{n}}, E_Q \left[\sup_{\theta \in \Theta} |g(x, \theta)|^\varphi \right] < M \right\}, \\ \sigma_\eta^2 &= \left(E_{P_{0, \eta}} [\dot{g}(x, 0)]' E_{P_{0, \eta}} [g(x, 0) g(x, 0)']^{-1} E_{P_{0, \eta}} [\dot{g}(x, 0)] \right)^{-1}. \end{aligned}$$

Note that

$$\begin{aligned} \sqrt{n} (\bar{T}(P_n) - \bar{T}(Q_n)) &\xrightarrow{d} N(0, \sigma_\eta^2), \quad \hat{\sigma} \xrightarrow{p} \sigma_\eta \quad \text{for all sequence } Q_n \in \mathcal{Q}_n^{LA}(0, \eta) \\ \sqrt{n} (\bar{T}(P_n) - \bar{T}(Q_n)) &\xrightarrow{d} N(t, \sigma_\eta^2), \quad \hat{\sigma} \xrightarrow{p} \sigma_\eta \quad \text{for all sequence } Q_n \in \mathcal{Q}_n^{LA}(t, \eta) \end{aligned}$$

We say that a test ψ_n is asymptotically unbiased at η if

$$\limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_n] \leq \liminf_{n \rightarrow \infty} \inf_{Q \in \mathcal{Q}_n^{LA}(t, \eta)} E_Q [\psi_n].$$

The Hellinger-based Wald test is optimal in terms of its robust power property as well, as the following result demonstrates.

Theorem 6. *Suppose that Assumption 2.1 holds. Then the following results hold for each $r_0, r_1 > 0$ and $t \in \mathbb{R}$ satisfying $\sigma_\eta^{-1} |t| + 2r_1 > 2r_0$ and each $\eta \in L_2(P_{0, \eta})$ satisfying $\int \eta dP_{0, \eta} = 0$.*

(i): *If an asymptotically unbiased test ψ_n at η satisfies*

$$(2.6) \quad \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_n] \leq \frac{\alpha}{2} + \Phi(-z_{\alpha/2} - 4r_0),$$

then

$$\liminf_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t, \eta)} E_Q [1 - \psi_n] \geq \Phi(z_{\alpha/2} + 2r_0 + \sigma_\eta^{-1} |t| + 2r_1) - \Phi(-z_{\alpha/2} - 2r_0 + \sigma_\eta^{-1} |t| + 2r_1).$$

(ii): The Hellinger-based Wald test $\psi_{H,n} = \mathbb{I} \left\{ \left| \frac{\sqrt{n}\tau(\hat{\theta})}{\hat{\sigma}} \right| \geq z_{\alpha/2} + 2r_0 \right\}$ is asymptotically unbiased and satisfies

$$(2.7) \quad \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_{H,n}] \leq \frac{\alpha}{2} + \Phi(-z_{\alpha/2} - 4r_0),$$

and

$$(2.8) \quad \lim_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t, \eta)} E_Q [1 - \psi_{H,n}] = \Phi(z_{\alpha/2} + 2r_0 + \sigma_\eta^{-1} |t| + 2r_1) - \Phi(-z_{\alpha/2} - 2r_0 + \sigma_\eta^{-1} |t| + 2r_1).$$

3. CONCLUSION

To be written.

APPENDIX A. PROOF OF THEOREMS

Notation. Let

$$g_n(x, \theta) = g(x, \theta) \mathbb{I}\{x \in \mathcal{X}_n\}$$

A.1. Proof of Theorem 4. Proof of (i). Pick any $r_0 > 0$, $\epsilon \in (0, r_0)$, and $(\bar{\theta}, \bar{P})$ such that $\tau(\bar{\theta}) = 0$ and $\bar{P} \in \mathcal{P}_{\bar{\theta}}$ (i.e., $\bar{P} \in \mathcal{P}_0$). Let

$$\begin{aligned} \bar{V} &= \left(E_{\bar{P}} [\dot{g}(x, \bar{\theta})]' E_{\bar{P}} [g(x, \bar{\theta}) g(x, \bar{\theta})']^{-1} E_{\bar{P}} [\dot{g}(x, \bar{\theta})] \right)^{-1}, \\ \bar{t} &= \frac{2(r_0 - \epsilon)}{\sqrt{\dot{\tau}(\bar{\theta})' V \dot{\tau}(\bar{\theta})}} \bar{V} \dot{\tau}(\bar{\theta}). \end{aligned}$$

Consider a sequence of parametric submodels $\{P_{\bar{\theta} + \bar{t}/\sqrt{n}}\}_{n \in \mathbb{N}}$ with the Radon-Nikodym density

$$\frac{dP_{\bar{\theta} + \bar{t}/\sqrt{n}}}{d\bar{P}} = \frac{1 + \bar{\zeta}'_n g_n(x, \bar{\theta} + \bar{t}/\sqrt{n})}{\int (1 + \bar{\zeta}'_n g_n(x, \bar{\theta} + \bar{t}/\sqrt{n})) d\bar{P}},$$

where $\bar{\zeta}_n = -E_{\bar{P}} [g(x, \bar{\theta} + \bar{t}/\sqrt{n}) g_n(x, \bar{\theta} + \bar{t}/\sqrt{n})']^{-1} E_{\bar{P}} [g(x, \bar{\theta} + \bar{t}/\sqrt{n})]$. By a similar argument to the proof of Theorem 3.1 of Kitamura, Otsu, and Evdokimov (2009), we can show that

$$\lim_{n \rightarrow \infty} nH \left(P_{\bar{\theta} + \bar{t}/\sqrt{n}}, \bar{P} \right)^2 = \frac{1}{4} \bar{t}' \bar{V}^{-1} \bar{t} = (r_0 - \epsilon)^2,$$

which implies

$$(A.1) \quad P_{\bar{\theta} + \bar{t}/\sqrt{n}} \in \mathcal{Q}_n^N,$$

for all n large enough.

Since ψ_n belongs to the class \mathcal{S} in Definition 3, we can write it as $\psi_n = \mathbb{I}\{S_n \geq \chi_{1, \alpha}^2\}$ and there exists a Wald test statistic $\left(\frac{\sqrt{n} \tau \circ T_a(P_n)}{\hat{\sigma}_a} \right)^2$ satisfying

$$\begin{aligned} S_n &= \left(\hat{\sigma}_a^{-1} \sqrt{n} \tau \circ T_a(P_n) \right)^2 + o_p(1) \\ &= \left(\sigma_a^{-1} \sqrt{n} \tau \circ T_a(P_n) \right)^2 + o_p(1) \\ (A.2) \quad &= \left(\sigma_a^{-1} \dot{\tau}(\bar{\theta})' \sqrt{n} (T_a(P_n) - \bar{\theta}) \right)^2 + \varrho_n, \end{aligned}$$

under $\{P_{\bar{\theta} + \bar{t}/\sqrt{n}}\}_{n \in \mathbb{N}}$, where $\sigma_a^2 = \dot{\tau}(\bar{\theta})' V_a \dot{\tau}(\bar{\theta})$, $\varrho_n = O_p(n(T_a(P_n) - \bar{\theta})^4)$, the second equality follows from $\hat{\sigma}_a \xrightarrow{P} \sigma_a$ under $\{P_{\bar{\theta} + \bar{t}/\sqrt{n}}\}_{n \in \mathbb{N}}$ (by Definition 2), and the third equality follows from a second-order expansion

around $T_a(P_n) = \bar{\theta}$ (note: $\tau(\bar{\theta}) = 0$ and the second-order derivative of τ is bounded on Θ). Thus, we have

$$\begin{aligned}
& \liminf_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q[\psi_n] \\
& \geq \liminf_{n \rightarrow \infty} E_{P_{\bar{\theta} + \bar{t}/\sqrt{n}}} [\mathbb{I}\{S_n \geq \chi_{1,\alpha}^2\}] \\
& = \liminf_{n \rightarrow \infty} \int \mathbb{I}\left\{\left(\sigma_a^{-1} \dot{\tau}(\bar{\theta})' \sqrt{n} \{T_a(P_n) - \bar{\theta}\}\right)^2 \geq \chi_{1,\alpha}^2\right\} dP_{\bar{\theta} + \bar{t}/\sqrt{n}}^{\otimes n} \\
& = \liminf_{n \rightarrow \infty} \int \mathbb{I}\left\{\left(\sigma_a^{-1} \dot{\tau}(\bar{\theta})' \left\{\sqrt{n} \left(T_a(P_n) - T_a\left(P_{\bar{\theta} + \bar{t}/\sqrt{n}}\right)\right) + \sqrt{n} \left(T_a\left(P_{\bar{\theta} + \bar{t}/\sqrt{n}}\right) - \bar{\theta}\right)\right\}\right)^2 \geq \chi_{1,\alpha}^2\right\} dP_{\bar{\theta} + \bar{t}/\sqrt{n}}^{\otimes n} \\
& = E_Z \left[\mathbb{I}\left\{\left(Z + \sigma_a^{-1} \dot{\tau}(\bar{\theta})' \bar{t}\right)^2 \geq \chi_{1,\alpha}^2\right\} \right] \\
& = E_Z \left[\mathbb{I}\left\{\left(Z + 2(r_0 - \epsilon)\right)^2 \geq \chi_{1,\alpha}^2\right\} \right],
\end{aligned}$$

where $Z \sim N(0, 1)$, the inequality follows from (A.1), the first equality follows from (A.2) and $\varrho_n \xrightarrow{P} 0$ under $\left\{P_{\bar{\theta} + \bar{t}/\sqrt{n}}\right\}_{n \in \mathbb{N}}$, the third equality follows from the requirements on the mapping T_a in Definition 2 under $\left\{P_{\bar{\theta} + \bar{t}/\sqrt{n}}\right\}_{n \in \mathbb{N}}$, and the last equality follows from $\sigma_a^2 \geq \sigma^2$ and the definition of \bar{t} . Since ϵ can be arbitrary small, we obtain the conclusion.

Proof of (ii). For the conclusion, it is sufficient to show that for any sequence $Q_n \in \mathcal{Q}_n^N$ generating the triangular array $\{x_i\}_{1 \leq i \leq n, n}$, it holds

$$(A.3) \quad \limsup_{n \rightarrow \infty} E_{Q_n}[\psi_{H,n}] \leq E_Z \left[\mathbb{I}\left\{\left(Z + 2r_0\right)^2 \geq \chi_{1,\alpha}^2\right\} \right],$$

where $Z \sim N(0, 1)$. Pick any sequence $\{Q_n\}_{n \in \mathbb{N}}$ such that $Q_n \in \mathcal{Q}_n^N$ for each $n \in \mathbb{N}$. For this $\{Q_n\}_{n \in \mathbb{N}}$, we can find sequences $\{P_n\}_{n \in \mathbb{N}}$ and $\{\theta_n\}_{n \in \mathbb{N}}$ such that $H(Q_n, P_n) \leq \frac{r_0}{\sqrt{n}}$ and $\tau(\theta_n) = 0$ (i.e., $P_n \in \mathcal{P}_{\theta_n} \subset \mathcal{P}_0$) for each $n \in \mathbb{N}$. Observe that

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} E_{Q_n}[\psi_{H,n}] \\
& = \limsup_{n \rightarrow \infty} \int \mathbb{I}\left\{\left(\hat{\sigma}^{-1} \sqrt{n} \tau \circ T(\hat{P}_n)\right)^2 \geq \chi_{1,\alpha}^2\right\} dQ_n^{\otimes n} \\
& \leq \limsup_{n \rightarrow \infty} \int_{(x_1, \dots, x_n) \notin \mathcal{X}_n^n} dQ_n^{\otimes n} + \limsup_{n \rightarrow \infty} \int_{(x_1, \dots, x_n) \in \mathcal{X}_n^n} \mathbb{I}\left\{\left(\hat{\sigma}^{-1} \sqrt{n} \tau \circ T(\hat{P}_n)\right)^2 \geq \chi_{1,\alpha}^2\right\} dQ_n^{\otimes n} \\
& = A_1 + A_2,
\end{aligned}$$

where the inequality follows from $\mathbb{I}\{\cdot\} \leq 1$ and $T(\hat{P}_n) = \bar{T}(\hat{P}_n)$ for all $(x_1, \dots, x_n) \in \mathcal{X}_n^n$. For A_1 , we have

$$\begin{aligned}
A_1 & \leq \limsup_{n \rightarrow \infty} \sum_{i=1}^n \int_{x_i \notin \mathcal{X}_n} dQ_n \\
& \leq \limsup_{n \rightarrow \infty} n m_n^{-\eta} E_{Q_n} \left[\sup_{\theta \in \Theta} |g(x, \theta)|^\eta \right] = 0,
\end{aligned}$$

where the first inequality follows from a set inclusion relation, the second inequality follows from the Markov inequality, and the equality follows from $n m_n^{-\eta} \rightarrow 0$ (Assumption 2.1) and $E_{Q_n}[\sup_{\theta \in \Theta} |g(x, \theta)|^\eta] < \infty$ (by

$Q_n \in \mathcal{Q}_n^N$). For A_2 , we have

$$\begin{aligned} A_2 &\leq \limsup_{n \rightarrow \infty} \mathbb{I} \left\{ \left(\hat{\sigma}^{-1} \sqrt{n} \tau \circ T \left(\hat{P}_n \right) \right)^2 \geq \chi_{1,\alpha}^2 \right\} dQ_n^{\otimes n} \\ &= \limsup_{n \rightarrow \infty} \int \mathbb{I} \left\{ \left\{ \hat{\sigma}^{-1} \hat{\tau}(\theta_n) \left\{ \sqrt{n} \left(\bar{T} \left(\hat{P}_n \right) - \bar{T} \left(Q_n \right) \right) + \sqrt{n} \left(\bar{T} \left(Q_n \right) - \theta_n \right) + \varrho_n \right\} \right\}^2 \geq \chi_{1,\alpha}^2 \right\} dQ_n^{\otimes n} \\ &\leq E_Z \left[\mathbb{I} \left\{ \left\{ Z + 2r_0 \right\}^2 \geq \chi_{1,\alpha}^2 \right\} \right], \end{aligned}$$

where $\varrho_n = o_p(1)$ under Q_n , the inequality follows from the set inclusion relation (\mathcal{X}_n^n is a subset of support of $Q_n^{\otimes n}$), the first equality follows from a second-order expansion around $T \left(\hat{P}_n \right) = \theta_n$ and $\tau(\theta_n) = 0$, and the second equality follows from Lemmas 3 and 13 and Theorem 3.1 of Kitamura, Otsu, and Evdokimov (2009).

A.2. Proof of Theorem 5 (Case I). Proof of (i). Suppose the conclusion is false, i.e.,

$$(A.4) \quad \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_n] \leq \alpha,$$

$$(A.5) \quad \liminf_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(P_{t/\sqrt{n},\eta})} E_Q [1 - \psi_n] \leq \Phi(z_\alpha + 2r_0 + 2r_1 - \sigma_\eta^{-1}t) - \epsilon,$$

for some $\epsilon > 0$.

First, we explore the implication of the size requirement in (A.4). Consider a sequence of parametric submodels $P_{t/\sqrt{n},\eta}^*$ having the Radon-Nikodym density

$$\frac{dP_{t/\sqrt{n},\eta}^*}{dP_{0,\eta}} = \frac{1 + \zeta'_{t,n} g_n \left(x, \frac{t}{\sqrt{n}} \right)}{\int \left(1 + \zeta'_{t,n} g_n \left(x, \frac{t}{\sqrt{n}} \right) \right) dP_{0,\eta}},$$

where $\zeta_{t,n} = -E_{P_{0,\eta}} \left[g \left(x, \frac{t}{\sqrt{n}} \right) g_n \left(x, \frac{t}{\sqrt{n}} \right) \right]^{-1} E_{P_{0,\eta}} \left[g \left(x, \frac{t}{\sqrt{n}} \right) \right]$. Note that

$$(A.6) \quad P_{t/\sqrt{n},\eta}^* \in \mathcal{P}_{t/\sqrt{n}},$$

for each $t \in \mathbb{R}$ and $n \in \mathbb{N}$ because

$$E_{P_{t/\sqrt{n},\eta}^*} \left[g \left(x, \frac{t}{\sqrt{n}} \right) \right] = \frac{\int g \left(x, \frac{t}{\sqrt{n}} \right) dP_{0,\eta} + \int g \left(x, \frac{t}{\sqrt{n}} \right) g_n \left(x, \frac{t}{\sqrt{n}} \right) dP_{0,\eta} \zeta_{t,n}}{\int \left(1 + \zeta'_{t,n} g_n \left(x, \frac{t}{\sqrt{n}} \right) \right) dP_{0,\eta}} = 0.$$

Pick any $\delta \in (0, \min\{r_0, r_1\})$ and define

$$t_0 = 2\sigma_\eta(r_0 - \delta).$$

From a similar argument to the proof of Theorem 3.1 of Kitamura, Otsu, and Evdokimov (2009),

$$nH \left(P_{t_0/\sqrt{n},\eta}^*, P_{0,\eta} \right)^2 \rightarrow \frac{1}{4} \sigma_\eta^{-2} t_0^2 = (r_0 - \delta)^2,$$

which implies

$$(A.7) \quad P_{t_0/\sqrt{n},\eta}^* \in \mathcal{Q}_n^N,$$

for all n large enough. The limiting behavior of the log likelihood ratio $\sum_{i=1}^n \log \frac{dP_{t_0/\sqrt{n}, \eta}^*}{dP_{0, \eta}}(x_i)$ is obtained as

$$\begin{aligned}
& \sum_{i=1}^n \log \frac{dP_{t_0/\sqrt{n}, \eta}^*}{dP_{0, \eta}}(x_i) \\
&= \sum_{i=1}^n \log \left(1 + \zeta'_{t_0, n} g_n \left(x_i, \frac{t_0}{\sqrt{n}} \right) \right) - n \log \int \left(1 + \zeta'_{t_0, n} g_n \left(x, \frac{t_0}{\sqrt{n}} \right) \right) dP_{0, \eta} \\
&= \zeta'_{t_0, n} \sum_{i=1}^n g_n \left(x_i, \frac{t_0}{\sqrt{n}} \right) - \frac{1}{2} \zeta'_{t_0, n} \sum_{i=1}^n g_n \left(x_i, \frac{t_0}{\sqrt{n}} \right) g_n \left(x_i, \frac{t_0}{\sqrt{n}} \right)' \zeta_{t_0, n} \\
&\quad - n \zeta'_{t_0, n} \int g_n \left(x, \frac{t_0}{\sqrt{n}} \right) dP_{0, \eta} \\
&\quad + \frac{n}{2} \zeta'_{t_0, n} \int g_n \left(x, \frac{t_0}{\sqrt{n}} \right) dP_{0, \eta} \int g_n \left(x, \frac{t_0}{\sqrt{n}} \right)' dP_{0, \eta} \zeta_{t_0, n} + o_p(1) \\
\text{(A.8)} \quad &\Rightarrow t_0 \sigma_\eta^{-1} Z - \frac{1}{2} \sigma_\eta^{-2} t_0^2,
\end{aligned}$$

under $x \sim P_{0, \eta}$, where $Z \sim N(0, 1)$, the second equality follows from a second-order expansion around $\zeta_{t_0, n} = 0$ with Lemmas 6 and 14, the third equality follows from an expansion around $t_0 = 0$ with Lemmas 6 and 14, and the convergence follows from Lemmas 6, 13, 14, and 16. Therefore, the size requirement in (A.4) implies

$$\begin{aligned}
\alpha &\geq \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q[\psi_n] \\
&\geq \liminf_{n \rightarrow \infty} E_{P_{t_0/\sqrt{n}, \eta}^*}[\psi_n] \\
&= \liminf_{n \rightarrow \infty} E_{P_{0, \eta}} \left[\psi_n \prod_{i=1}^n \frac{dP_{t_0/\sqrt{n}, \eta}^*}{dP_{0, \eta}}(x_i) \right] \\
&= E_{\psi, Z} \left[\psi \exp \left(t_0 \sigma_\eta^{-1} Z - \frac{1}{2} \sigma_\eta^{-2} t_0^2 \right) \right] \\
&= \frac{1}{\sqrt{2\pi}} \int_z \psi_0(z) \exp \left(-\frac{1}{2} (z - \sigma_\eta^{-1} t_0)^2 \right) dz \\
\text{(A.9)} \quad &= E_Z[\psi_0(Z + \sigma_\eta^{-1} t_0)] = E_Z[\psi_0(Z + 2(r_0 - \delta))],
\end{aligned}$$

where ψ is a random variable such that $\psi_n \Rightarrow \psi$ under P_θ , $\psi_0(z) = E[\psi|Z=z]$, the first inequality comes from (A.4), the second inequality follows from (A.7), the first equality follows from a change of measure, the second equality follows from (A.8), the third equality follows from the law of iterated expectation, and the last equality follows from a change of variables.

Next, we explore the implication of the local power requirement in (A.5). Define

$$t_1 = -2\sigma_\eta(r_1 - \delta).$$

By a similar argument to (A.6),

$$\text{(A.10)} \quad nH \left(P_{(t_1+t)/\sqrt{n}, \eta}^*, P_{t/\sqrt{n}, \eta} \right)^2 \rightarrow \frac{1}{4} t_1' \sigma_\eta^{-1} t_1 = (r_1 - \delta)^2,$$

which implies

$$(A.11) \quad P_{(t_1+t)/\sqrt{n},\eta} \in \mathcal{Q}_n^{LA}(t, \eta),$$

for all n large enough. By a similar argument to (A.8),

$$(A.12) \quad \begin{aligned} & \sum_{i=1}^n \log \frac{dP_{(t_1+t)/\sqrt{n},\eta}^*(x_i)}{dP_{0,\eta}} \\ \Rightarrow & (t_1+t) \sigma_\eta^{-1} Z - \frac{1}{2} \sigma_\eta^{-2} (t_1+t). \end{aligned}$$

Therefore, the power requirement in (A.5) implies

$$(A.13) \quad \begin{aligned} & \Phi(z_\alpha + 2r_0 + 2r_1 - \sigma_\eta^{-1}t) - \epsilon \\ \geq & \liminf_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t,\eta)} E_Q [1 - \psi_n] \\ \geq & \liminf_{n \rightarrow \infty} E_{P_{(t_1+t)/\sqrt{n},\eta}} [1 - \psi_n] \\ = & \liminf_{n \rightarrow \infty} E_{P_{0,\eta}} \left[(1 - \psi_n) \prod_{i=1}^n \frac{dP_{(t_1+t)/\sqrt{n},\eta}^*(x_i)}{dP_{0,\eta}} \right] \\ = & E_{\psi,Z} \left[(1 - \psi) \exp \left((t_1+t) \sigma_\eta^{-1} Z - \frac{1}{2} \sigma_\eta^{-2} (t_1+t)^2 \right) \right] \\ = & \frac{1}{\sqrt{2\pi}} \int_z (1 - \psi_0(z)) \exp \left(-\frac{1}{2} (z - \sigma_\eta^{-1} (t_1+t))^2 \right) dz \\ = & 1 - E_Z [\psi_0(Z + \sigma_\eta^{-1} (t_1+t))] = 1 - E_Z [\psi_0(Z - 2(r_1 - \delta) + \sigma_\eta^{-1}t)], \end{aligned}$$

where ψ is a random variable such that $\psi_n \Rightarrow \psi$ under P_θ , $\psi_0(z) = E[\psi|Z=z]$, the first inequality comes from (A.5), the second inequality follows from (A.11), the first equality follows from a change of measure, the second equality follows from (A.12), the third equality follows from the law of iterated expectation, and the last equality follows from a change of variables.

Finally, we derive a contradiction. The requirement (A.9) is written as

$$(A.14) \quad E_Z [\psi_0(Z + 2(r_0 + \delta))] \leq \alpha.$$

The requirement (A.13) is written as

$$(A.15) \quad E_Z [\psi_0(Z - 2(r_1 - \delta) + \sigma_\eta^{-1}t)] \geq \Phi(\sigma_\eta^{-1}t - z_\alpha - 2r_0 - 2r_1) + \epsilon.$$

However, $\Phi(\sigma_\eta^{-1}t - z_\alpha - 2r_0 - 2r_1)$ is the power of the Neyman-Pearson test for testing $H_0 : \mu = 2r_0$ against $H_1 : \mu = \sigma_\eta^{-1/2}t - 2r_1$ under $N(\mu, 1)$. By taking δ sufficiently small, we obtain a contradiction.

Proof of (ii). Proof of the first statement. Pick any sequence $\{Q_n\}_{n \in \mathbb{N}}$ such that $Q_n \in \mathcal{Q}_n^N$ for each $n \in \mathbb{N}$. For this $\{Q_n\}_{n \in \mathbb{N}}$, we can find sequences $\{P_n\}_{n \in \mathbb{N}}$ such that $H(Q_n, P_n) \leq \frac{r_0}{\sqrt{n}}$ for each $n \in \mathbb{N}$.

From the same argument to the proof of Part (ii) of Theorem 4 with Lemmas 3 and 13 and Theorem 3.1 of Kitamura, Otsu, and Evdokimov (2009),

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} E_{Q_n} [\psi_{H,n}] \\
&= \limsup_{n \rightarrow \infty} E_{Q_n} \left[\mathbb{I} \left\{ \hat{\sigma}^{-1} \sqrt{n} T \left(\hat{P}_n \right) \geq z_\alpha + 2r_0 \right\} \right] \\
&\leq \limsup_{n \rightarrow \infty} E_{Q_n} \left[\mathbb{I} \left\{ \hat{\sigma}^{-1} \sqrt{n} \bar{T} \left(\hat{P}_n \right) \geq z_\alpha + 2r_0 \right\} \right] + o(1) \\
&\leq \limsup_{n \rightarrow \infty} E_{Q_n} \left[\mathbb{I} \left\{ \hat{\sigma}^{-1} \sqrt{n} \left(\bar{T} \left(\hat{P}_n \right) - \bar{T} \left(Q_n \right) \right) + \hat{\sigma}^{-1} \sup_{Q \in \mathcal{Q}_n^N} \sqrt{n} |\bar{T} \left(Q \right)| \geq z_\alpha + 2r_0 \right\} \right] + o(1) \\
&= E_Z \left[\mathbb{I} \left\{ Z \geq z_\alpha \right\} \right] \\
&= \alpha.
\end{aligned}$$

Proof of the second statement. It is sufficient to show that

$$\limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t,\eta)} E_Q [1 - \psi_{H,n}] \leq \Phi \left(z_\alpha + 2r_0 + 2r_1 - \sigma_\eta^{-1} t \right).$$

Observe that the Type II error satisfies

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t,\eta)} E_Q [1 - \psi_{H,n}] \\
&\leq \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t,\eta)} E_Q \left[\mathbb{I} \left\{ \hat{\sigma}^{-1} \sqrt{n} \bar{T} \left(\hat{P}_n \right) \leq z_\alpha + 2r_0 \right\} \right] + o(1) \\
&= \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t,\eta)} E_Q \left[\mathbb{I} \left\{ \hat{\sigma}^{-1} \sqrt{n} \left(\bar{T} \left(\hat{P}_n \right) - \bar{T} \left(Q \right) \right) + \hat{\sigma}^{-1} \sqrt{n} \bar{T} \left(Q \right) \leq z_\alpha + 2r_0 \right\} \right] \\
&= E_Q \left[\mathbb{I} \left\{ Z + \sigma_\eta^{-1} t - \sigma^{-1} \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t,\eta)} \sqrt{n} |\bar{T} \left(Q \right)| \leq z_\alpha + 2r_0 \right\} \right] \\
&= \Phi \left(z_\alpha + 2r_0 + 2r_1 - \sigma_\eta^{-1} t \right),
\end{aligned}$$

where $Z \sim N(0, 1)$, the first inequality follows from the same argument to the proof of Part (ii) of Theorem 4, the second inequality follows from the set inclusion relation, and the second equality follows from Lemmas 3 and 13, Theorem 3.1 of Kitamura, Otsu, and Evdokimov (2009).

A.3. Proof of Theorem 6 (Case II). Proof of (i). Suppose the conclusion is false, i.e.,

$$\text{(A.16)} \quad \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_n] \leq \frac{\alpha}{2} + \Phi \left(-z_{\alpha/2} - 4r_0 \right),$$

$$\text{(A.17)} \quad \inf_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t,\eta)} E_Q [1 - \psi_n] \leq \Phi \left(z_{\alpha/2} + 2r_0 + \sigma_\eta^{-1} |t| + 2r_1 \right) - \Phi \left(-z_{\alpha/2} - 2r_0 + \sigma_\eta^{-1} |t| + 2r_1 \right) - \epsilon,$$

for some $\epsilon > 0$.

First, we explore the implication of the size requirement in (A.16). Consider a sequence of parametric submodels $P_{t/\sqrt{n},\eta}^*$ having the Radon-Nikodym density

$$\frac{dP_{t/\sqrt{n},\eta}^*}{dP_{0,\eta}} = \frac{1 + \zeta'_{t,n} g_n \left(x, \frac{t}{\sqrt{n}} \right)}{\int \left(1 + \zeta'_{t,n} g_n \left(x, \frac{t}{\sqrt{n}} \right) \right) dP_{0,\eta}},$$

where $\zeta_{t,n} = -E_{P_{0,\eta}} \left[g \left(x, \frac{t}{\sqrt{n}} \right) g_n \left(x, \frac{t}{\sqrt{n}} \right) \right]^{-1} E_{P_{0,\eta}} \left[g \left(x, \frac{t}{\sqrt{n}} \right) \right]$. Note that

$$(A.18) \quad P_{t/\sqrt{n},\eta}^* \in \mathcal{P}_{t/\sqrt{n}},$$

for each $t \in \mathbb{R}$ and $n \in \mathbb{N}$ because

$$E_{P_{t/\sqrt{n},\eta}^*} \left[g \left(x, \frac{t}{\sqrt{n}} \right) \right] = \frac{\int g \left(x, \frac{t}{\sqrt{n}} \right) dP_{0,\eta} + \int g \left(x, \frac{t}{\sqrt{n}} \right) g_n \left(x, \frac{t}{\sqrt{n}} \right)' dP_{0,\eta} \zeta_{t,n}}{\int \left(1 + \zeta'_{t,n} g_n \left(x, \frac{t}{\sqrt{n}} \right) \right) dP_{0,\eta}} = 0.$$

Pick any $\delta \in (0, \min \{r_0, r_1\})$ and define

$$t_0 = 2\sigma_\eta (r_0 - \delta).$$

From a similar argument to the proof of Theorem 3.1 of Kitamura, Otsu, and Evdokimov (2009),

$$nH \left(P_{t_0/\sqrt{n},\eta}^*, P_{0,\eta} \right)^2 \rightarrow \frac{1}{4} \sigma_\eta^{-2} t_0^2 = (r_0 - \delta)^2,$$

which implies

$$(A.19) \quad P_{t_0/\sqrt{n},\eta}^* \in \mathcal{Q}_n^N,$$

for all n large enough. The limiting behavior of the log likelihood ratio $\sum_{i=1}^n \log \frac{dP_{t_0/\sqrt{n},\eta}^*}{dP_{0,\eta}} (x_i)$ is obtained as

$$\begin{aligned} & \sum_{i=1}^n \log \frac{dP_{t_0/\sqrt{n},\eta}^*}{dP_{0,\eta}} (x_i) \\ &= \sum_{i=1}^n \log \left(1 + \zeta'_{t_0,n} g_n \left(x_i, \frac{t_0}{\sqrt{n}} \right) \right) - n \log \int \left(1 + \zeta'_{t_0,n} g_n \left(x, \frac{t_0}{\sqrt{n}} \right) \right) dP_{0,\eta} \\ &= \zeta'_{t_0,n} \sum_{i=1}^n g_n \left(x_i, \frac{t_0}{\sqrt{n}} \right) - \frac{1}{2} \zeta'_{t_0,n} \sum_{i=1}^n g_n \left(x_i, \frac{t_0}{\sqrt{n}} \right) g_n \left(x_i, \frac{t_0}{\sqrt{n}} \right)' \zeta_{t_0,n} \\ & \quad - n \zeta'_{t_0,n} \int g_n \left(x, \frac{t_0}{\sqrt{n}} \right) dP_{0,\eta} \\ & \quad + \frac{n}{2} \zeta'_{t_0,n} \int g_n \left(x, \frac{t_0}{\sqrt{n}} \right) dP_{0,\eta} \int g_n \left(x, \frac{t_0}{\sqrt{n}} \right)' dP_{0,\eta} \zeta_{t_0,n} + o_p(1) \\ (A.20) \quad & \Rightarrow t_0 \sigma_\eta^{-1} Z - \frac{1}{2} \sigma_\eta^{-2} t_0^2, \end{aligned}$$

under $x \sim P_{0,\eta}$, where $Z \sim N(0, 1)$, the second equality follows from a second-order expansion around $\zeta_{t_0,n} = 0$ with Lemmas 6 and 14, the third equality follows from an expansion around $t_0 = 0$ with Lemmas 6 and 14, and

the convergence follows from Lemmas 6, 13, 14, and 16. Therefore, the size requirement in (A.16) implies

$$\begin{aligned}
& \frac{\alpha}{2} + \Phi(-z_{\alpha/2} - 4r_0) \\
& \geq \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q[\psi_n] \\
& \geq \liminf_{n \rightarrow \infty} E_{P_{t_0/\sqrt{n}, \eta}^*}[\psi_n] \\
& = \liminf_{n \rightarrow \infty} E_{P_{0, \eta}} \left[\psi_n \prod_{i=1}^n \frac{dP_{t_0/\sqrt{n}, \eta}^*(x_i)}{dP_{0, \eta}} \right] \\
& = E_{\psi, Z} \left[\psi \exp \left(t_0 \sigma_\eta^{-1} Z - \frac{1}{2} \sigma_\eta^{-2} t_0^2 \right) \right] \\
& = \frac{1}{\sqrt{2\pi}} \int_z \psi_0(z) \exp \left(-\frac{1}{2} (z - \sigma_\eta^{-1} t_0)^2 \right) dz \\
\text{(A.21)} \quad & = E_Z[\psi_0(Z + \sigma_\eta^{-1} t_0)] = E_Z[\psi_0(Z + 2(r_0 - \delta))],
\end{aligned}$$

where ψ is a random variable such that $\psi_n \Rightarrow \psi$ under P_θ , $\psi_0(z) = E[\psi|Z=z]$, the first inequality comes from (A.16), the second inequality follows from (A.19), the first equality follows from a change of measure, the second equality follows from (A.20), the third equality follows from the law of iterated expectation, and the last equality follows from a change of variables.

Next, we explore the implication of the local power requirement in (A.17). Define

$$t_1 = -2\sigma_\eta(r_1 - \delta).$$

By a similar argument to (A.18),

$$\text{(A.22)} \quad nH \left(P_{(t_1+t)/\sqrt{n}, \eta}^*, P_{t/\sqrt{n}, \eta} \right)^2 \rightarrow \frac{1}{4} t_1' \sigma_\eta^{-1} t_1 = (r_1 - \delta)^2,$$

which implies

$$\text{(A.23)} \quad P_{(t_1+t)/\sqrt{n}, \eta} \in \mathcal{Q}_n^{LA}(t, \eta),$$

for all n large enough. By a similar argument to (A.20),

$$\begin{aligned}
& \sum_{i=1}^n \log \frac{dP_{(t_1+t)/\sqrt{n}, \eta}^*(x_i)}{dP_{0, \eta}} \\
\text{(A.24)} \quad & \Rightarrow (t_1 + t) \sigma_\eta^{-1} Z - \frac{1}{2} \sigma_\eta^{-2} (t_1 + t).
\end{aligned}$$

Therefore, the power requirement in (A.17) implies

$$\begin{aligned}
& \Phi(z_{\alpha/2} + 2r_0 + \sigma_\eta^{-1}|t| + 2r_1) - \Phi(-z_{\alpha/2} - 2r_0 + \sigma_\eta^{-1}|t| + 2r_1) - \epsilon \\
\geq & \liminf_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^A(t, \eta)} E_Q[1 - \psi_n] \\
\geq & \liminf_{n \rightarrow \infty} E_{P_{(t_1+t)/\sqrt{n}, \eta}}[1 - \psi_n] \\
= & \liminf_{n \rightarrow \infty} E_{P_{0, \eta}} \left[(1 - \psi_n) \prod_{i=1}^n \frac{dP_{(t_1+t)/\sqrt{n}, \eta}^*}{dP_{0, \eta}}(x_i) \right] \\
= & E_{\psi, Z} \left[(1 - \psi) \exp \left((t_1 + t) \sigma_\eta^{-1} Z - \frac{1}{2} \sigma_\eta^{-2} (t_1 + t)^2 \right) \right] \\
= & \frac{1}{\sqrt{2\pi}} \int_z (1 - \psi_0(z)) \exp \left(-\frac{1}{2} (z - \sigma_\eta^{-1} (t_1 + t))^2 \right) dz \\
(A.25) \quad = & 1 - E_Z [\psi_0(Z + \sigma_\eta^{-1} (t_1 + t))] = 1 - E_Z [\psi_0(Z - 2(r_1 - \delta) + \sigma_\eta^{-1} t)],
\end{aligned}$$

where ψ is a random variable such that $\psi_n \Rightarrow \psi$ under P_θ , $\psi_0(z) = E[\psi|Z = z]$, the first inequality comes from (A.17), the second inequality follows from (A.23), the first equality follows from a change of measure, the second equality follows from (A.24), the third equality follows from the law of iterated expectation, and the last equality follows from a change of variables.

Finally, we derive a contradiction. The requirement (A.21) is written as

$$(A.26) \quad E_Z [\psi_0(Z + 2(r_0 - \delta))] \leq \frac{\alpha}{2} + \Phi(-z_{\alpha/2} - 4r_0).$$

The requirement (A.25) is written as

$$(A.27) \quad E_Z [\psi_0(Z - 2(r_1 - \delta) + \sigma_\eta^{-1} t)] \geq \Phi(-z_{\alpha/2} - 2r_0 - (\sigma_\eta^{-1} |t| + 2r_1)) + \Phi(-z_{\alpha/2} - 2r_0 + (\sigma_\eta^{-1} |t| + 2r_1)) + \epsilon.$$

However, $\Phi(-z_{\alpha/2} - 2r_0 - (\sigma_\eta^{-1} |t| + 2r_1)) + \Phi(-z_{\alpha/2} - 2r_0 + (\sigma_\eta^{-1} |t| + 2r_1))$ is the power of the uniformly most powerful unbiased level $\frac{\alpha}{2} + \Phi(-z_{\alpha/2} - 4r_0)$ test for $H_0 : |\mu| \leq 2r_0$ against $H_1 : |\mu| > 2r_0$ under $Z \sim N(\mu, 1)$ evaluated at $\mu = \sigma_\eta^{-1} |t| + 2r_1$. Therefore, by taking δ sufficiently small, we obtain a contradiction.

Proof of (ii). Proof of (2.7). Pick any sequence $\{Q_n\}_{n \in \mathbb{N}}$ such that $Q_n \in \mathcal{Q}_n^N$ for each $n \in \mathbb{N}$. For this $\{Q_n\}_{n \in \mathbb{N}}$, we can find sequences $\{P_n\}_{n \in \mathbb{N}}$ such that $H(Q_n, P_n) \leq \frac{r_0}{\sqrt{n}}$ for each $n \in \mathbb{N}$. From the same argument to the proof of Part (ii) of Theorem 4 with Lemmas 3 and 13 and Theorem 3.1 of Kitamura, Otsu,

and Evdokimov (2009),

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} E_{Q_n} [\psi_{H,n}] \\
&= \limsup_{n \rightarrow \infty} E_{Q_n} \left[\mathbb{I} \left\{ \left| \hat{\sigma}^{-1} \sqrt{n} T \left(\hat{P}_n \right) \right| \geq z_{\alpha/2} + 2r_0 \right\} \right] \\
&\leq \limsup_{n \rightarrow \infty} E_{Q_n} \left[\mathbb{I} \left\{ \hat{\sigma}^{-1} \sqrt{n} \left(\bar{T} \left(\hat{P}_n \right) - \bar{T} \left(Q_n \right) \right) + \hat{\sigma}^{-1} \sqrt{n} \bar{T} \left(Q \right) \geq z_{\alpha/2} + 2r_0 \right\} \right] \\
&\quad + \limsup_{n \rightarrow \infty} E_{Q_n} \left[\mathbb{I} \left\{ \hat{\sigma}^{-1} \sqrt{n} \left(\bar{T} \left(\hat{P}_n \right) - \bar{T} \left(Q_n \right) \right) + \hat{\sigma}^{-1} \sqrt{n} \bar{T} \left(Q \right) \leq -z_{\alpha/2} - 2r_0 \right\} \right] + o(1) \\
&\leq E_Z \left[\mathbb{I} \left\{ Z + 2r_0 \geq z_{\alpha/2} + 2r_0 \right\} \right] + E_Z \left[\mathbb{I} \left\{ Z + 2r_0 \leq -z_{\alpha/2} - 2r_0 \right\} \right] \\
\text{(A.28)} \quad &= \frac{\alpha}{2} + \Phi \left(-z_{\alpha/2} - 4r_0 \right).
\end{aligned}$$

Proof of the asymptotic unbiasedness and (2.8). Observe that the Type II error satisfies

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t,\eta)} E_Q [1 - \psi_{H,n}] \\
&\leq \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t,\eta)} E_Q \left[\mathbb{I} \left\{ -z_{\alpha/2} - 2r_0 \leq \hat{\sigma}^{-1} \sqrt{n} \bar{T} \left(\hat{P}_n \right) \leq z_{\alpha/2} + 2r_0 \right\} \right] + o(1) \\
&= \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^{LA}(t,\eta)} E_Q \left[\mathbb{I} \left\{ -z_{\alpha/2} - 2r_0 \leq \hat{\sigma}^{-1} \sqrt{n} \left(\bar{T} \left(\hat{P}_n \right) - \bar{T} \left(Q \right) \right) + \hat{\sigma}^{-1} \sqrt{n} \bar{T} \left(Q \right) \leq z_{\alpha/2} + 2r_0 \right\} \right] \\
&\leq E_Z \left[\mathbb{I} \left\{ -z_{\alpha/2} - 2r_0 + \sigma_\eta^{-1} |t| + 2r_1 \leq Z \leq z_{\alpha/2} + 2r_0 + \sigma_\eta^{-1} |t| + 2r_1 \right\} \right] \\
\text{(A.29)} \quad &\Phi \left(z_{\alpha/2} + 2r_0 + \sigma_\eta^{-1} |t| + 2r_1 \right) - \Phi \left(-z_{\alpha/2} - 2r_0 + \sigma_\eta^{-1} |t| + 2r_1 \right),
\end{aligned}$$

where $Z \sim N(0, 1)$, the first inequality follows from the same argument to the proof of Part (ii) of Theorem 4, the second inequality follows from the set inclusion relation, and the second equality follows from Lemmas 3 and 13, Theorem 3.1 of Kitamura, Otsu, and Evdokimov (2009).

Therefore, from (A.28) and (A.29), the asymptotic unbiasedness of $\psi_{H,n}$ is obtained as

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_{H,n}] \\
&\leq \frac{\alpha}{2} + \Phi \left(-z_{\alpha/2} - 4r_0 \right) \\
&\leq \Phi \left(-z_{\alpha/2} - 2r_0 - \left(\sigma_\eta^{-1} |t| + 2r_1 \right) \right) + \Phi \left(-z_{\alpha/2} - 2r_0 + \left(\sigma_\eta^{-1} |t| + 2r_1 \right) \right) \\
&\leq \limsup_{n \rightarrow \infty} \inf_{Q \in \mathcal{Q}_n^{LA}(t,\eta)} E_Q [\psi_{H,n}],
\end{aligned}$$

where the second inequality follows from $\sigma_\eta^{-1} |t| + 2r_1 > 2r_0$.

APPENDIX B. LEMMAS FROM KITAMURA, OTSU AND EVDOKIMOV (2009)

Notation. Let C be a generic positive constant, $\|\cdot\|$ be the L_2 -metric.

$$\begin{aligned} \mathcal{X}_n &= \left\{ x \in \mathcal{X} : \sup_{\theta \in \Theta} |g(x, \theta)| \leq m_n \right\}, & g_n(x, \theta) &= g(x, \theta) \mathbb{I}\{x \in \mathcal{X}_n\}, \\ \bar{\mathcal{P}}_\theta &= \left\{ P \in \mathcal{M} : \int g_n(x, \theta) dP = 0 \right\}, \\ \bar{T}(Q) &= \arg \min_{\theta \in \Theta} \min_{P \in \bar{\mathcal{P}}_\theta, P \ll Q} H(P, Q), & \bar{T}_{Q_n} &= \bar{T}(Q_n), & \bar{T}_{\hat{P}_n} &= \bar{T}(\hat{P}_n), \\ G_n &= E_{P_n} [\dot{g}(x, \theta_n)], & \Omega_n &= E_{P_n} [g(x, \theta_n) g(x, \theta_n)'], & V_n &= (G_n' \Omega_n^{-1} G_n)^{-1}, \\ \Lambda_n &= G_n' \Omega_n^{-1} g_n(x, \theta_n), & \psi_{n, Q_n} &= -2 \left(\int \Lambda_n \Lambda_n' dQ_n \right)^{-1} \int \Lambda_n \left\{ dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} \right\} dQ_n^{1/2}, \\ \bar{P}_{\theta, Q} &= \arg \min_{P \in \bar{\mathcal{P}}_\theta, P \ll Q} H(P, Q), & R_n(Q, \theta, \gamma) &= - \int \frac{1}{1 + \gamma' g_n(x, \theta)} dQ, \\ \theta_{t, n} &= \theta_n + t/\sqrt{n}. \end{aligned}$$

Lemma. Suppose that Assumption 2.1 holds. Then for each $\{Q_n\}_{n \in \mathbb{N}}$, $\{P_n\}_{n \in \mathbb{N}}$, $\{\theta_n\}_{n \in \mathbb{N}}$, and $r_0 > 0$ satisfying $H(Q_n, P_n) \leq \frac{r_0}{\sqrt{n}}$, $P_n \in \mathcal{P}_{\theta_n}$, and $\tau(\theta_n) = 0$, the following statements hold true.

- 1.: $\bar{T}(Q)$ exists and is upper semi-continuous at each $Q \in \mathcal{M}$ under the Hellinger distance,
- 2.: $|\bar{T}_{Q_n} - \theta_n| = o(1)$,
- 3.: $|\sqrt{n}(\bar{T}_{Q_n} - \theta_n) + \sqrt{n}V_n \int \Lambda_n dQ_n| = o(1)$,
- 4.: $\left\| d\bar{P}_{\bar{T}_{Q_n}, Q_n}^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2}(\bar{T}_{Q_n} - \theta_n)' \Lambda_n dQ_n^{1/2} \right\| = o(|\bar{T}_{Q_n} - \theta_n|) + o(n^{-1/2})$,
- 5.: $\left\| d\bar{P}_{\theta_n + \psi_n, Q_n, Q_n}^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2}\psi_n' \Lambda_n dQ_n^{1/2} \right\| = o(|\psi_n|) + o(n^{-1/2})$,
- 6.: $|E_{P_n}[g_n(x, \theta_n)]| = o(n^{-1/2})$, $|E_{P_n}[g_n(x, \theta_{t,n})]| = O(n^{-1/2})$, $|E_{P_n}[g_n(x, \theta_{t,n})g_n(x, \theta_{t,n})'] - \Omega_n| = o(1)$, and $|E_{P_n}[\dot{g}_n(x, \theta_{t,n})] - G_n| = o(1)$,
- 7.: $\gamma_n(\theta_{t,n}, P_n) = \arg \max_{\gamma \in \mathbb{R}^m} - \int \frac{1}{(1+\gamma'g_n(x, \theta_{t,n}))} dP_n$ exists for all n large enough, $|\gamma_n(\theta_{t,n}, P_n)| = O(n^{-1/2})$, and $\sup_{x \in \mathcal{X}} |\gamma_n(\theta_{t,n}, P_n)' g_n(x, \theta_{t,n})| = o(1)$,
- 8.: $|E_{Q_n}[g_n(x, \theta_n)]| = O(n^{-1/2})$, and $|E_{Q_n}[g_n(x, \theta_n)g_n(x, \theta_n)'] - \Omega_n| = o(1)$,
- 9.: $\gamma_n(\theta_n, Q_n) = \arg \max_{\gamma \in \mathbb{R}^m} - \int \frac{1}{(1+\gamma'g_n(x, \theta_n))} dQ_n$ exists for all n large enough, and $|\gamma_n(\theta_n, Q_n)| = O(n^{-1/2})$, and $\sup_{x \in \mathcal{X}} |\gamma_n(\theta_n, Q_n)' g_n(x, \theta_n)| = o(1)$,
- 10.: $|E_{Q_n}[g_n(x, \bar{T}_{Q_n})]| = O(n^{-1/2})$, $|E_{Q_n}[g_n(x, \bar{T}_{Q_n})g_n(x, \bar{T}_{Q_n})'] - \Omega_n| = o(1)$, and $|E_{Q_n}[\dot{g}_n(x, \bar{T}_{Q_n})] - G_n| = o(1)$,
- 11.: $\gamma_n(\bar{T}_{Q_n}, Q_n) = \arg \max_{\gamma \in \mathbb{R}^m} - \int \frac{1}{(1+\gamma'g_n(x, \bar{T}_{Q_n}))} dQ_n$ exists for all n large enough, $|\gamma_n(\bar{T}_{Q_n}, Q_n)| = O(n^{-1/2})$, and $\sup_{x \in \mathcal{X}} |\gamma_n(\bar{T}_{Q_n}, Q_n)' g_n(x, \bar{T}_{Q_n})| = o(1)$,
- 12.: $|\bar{T}_{\hat{P}_n} - \theta_n| = o_p(1)$,
- 13.: $|\sqrt{n}(\bar{T}_{\hat{P}_n} - \theta_n) + \sqrt{n}V_n \int \Lambda_n d\hat{P}_n| = o_p(1)$ and $V_n^{-1/2} \sqrt{n}(\bar{T}_{\hat{P}_n} - \bar{T}_{Q_n}) \xrightarrow{d} N(0, I)$,
- 14.: $|E_{\hat{P}_n}[g_n(x, \theta_n)]| = O_p(n^{-1/2})$, $|E_{\hat{P}_n}[g_n(x, \theta_n)g_n(x, \theta_n)'] - \Omega_n| = o_p(1)$,
- 15.: $\gamma_n(\theta_n, \hat{P}_n) = \arg \max_{\gamma \in \mathbb{R}^m} - \int \frac{1}{(1+\gamma'g_n(x, \theta_n))} d\hat{P}_n$ exists a.s. for all n large enough, $|\gamma_n(\theta_n, \hat{P}_n)| = O_p(n^{-1/2})$, and $\sup_{x \in \mathcal{X}} |\gamma_n(\theta_n, \hat{P}_n)' g_n(x, \theta_n)| = o_p(1)$,
- 16.: $|E_{\hat{P}_n}[g_n(x, \bar{T}_{\hat{P}_n})]| = O_p(n^{-1/2})$, $|E_{\hat{P}_n}[g_n(x, \bar{T}_{\hat{P}_n})g_n(x, \bar{T}_{\hat{P}_n})'] - \Omega_n| = O_p(n^{-1/2})$, and $|E_{\hat{P}_n}[\dot{g}_n(x, \bar{T}_{\hat{P}_n})] - G_n| = o_p(1)$,
- 17.: $\gamma_n(\bar{T}_{\hat{P}_n}, \hat{P}_n) = \arg \max_{\gamma \in \mathbb{R}^m} - \int \frac{1}{(1+\gamma'g_n(x, \bar{T}_{\hat{P}_n}))} d\hat{P}_n$ exists a.s. for all n large enough, $|\gamma_n(\bar{T}_{\hat{P}_n}, \hat{P}_n)| = O_p(n^{-1/2})$, and $\sup_{x \in \mathcal{X}} |\gamma_n(\bar{T}_{\hat{P}_n}, \hat{P}_n)' g_n(x, \bar{T}_{\hat{P}_n})| = o_p(1)$.

B.1. Proof of Lemma 1. See Kitamura, Otsu, and Evdokimov (2009).

B.2. Proof of Lemma 2. From the triangle inequality,

(B.1)

$$\sup_{\theta \in \Theta} |E_{Q_n} [g_n(x, \theta)] - E_{P_n} [g(x, \theta)]| \leq \sup_{\theta \in \Theta} |E_{Q_n} [g_n(x, \theta)] - E_{P_n} [g_n(x, \theta)]| + \sup_{\theta \in \Theta} |E_{P_n} [g(x, \theta) \mathbb{I}\{x \notin \mathcal{X}_n\}]|.$$

The first term of (B.1) satisfies

$$\begin{aligned} & \sup_{\theta \in \Theta} |E_{Q_n} [g_n(x, \theta)] - E_{P_n} [g_n(x, \theta)]| \\ & \leq \sup_{\theta \in \Theta} \left| \int g_n(x, \theta) \{dQ_n^{1/2} - dP_n^{1/2}\}^2 \right| + 2 \sup_{\theta \in \Theta} \left| \int g_n(x, \theta) dP_n^{1/2} \{dQ_n^{1/2} - dP_n^{1/2}\} \right| \\ & \leq m_n \frac{r_0^2}{n} + 2 \sqrt{E_{P_n} \left[\sup_{\theta \in \Theta} |g(x, \theta)|^2 \right]} \frac{r_0}{\sqrt{n}} = O(n^{-1/2}), \end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from $H(Q_n, P_n) \leq \frac{r_0}{\sqrt{n}}$ and the Cauchy-Schwarz inequality, and the equality follows from $\frac{m_n}{\sqrt{n}} \rightarrow 0$ (Assumption 2.1) and $P_n \in \mathcal{P}_{\theta_n}$ (particularly $E_{P_n} [\sup_{\theta \in \Theta} |g(x, \theta)|^\eta] \leq M$). The second term of (B.1) satisfies

$$\begin{aligned} & \sup_{\theta \in \Theta} |E_{P_n} [g(x, \theta) \mathbb{I}\{x \notin \mathcal{X}_n\}]| \\ & \leq \left(\int \sup_{\theta \in \Theta} |g(x, \theta)|^\eta dP_n \right)^{1/\eta} \left(\int \mathbb{I}\{x \notin \mathcal{X}_n\} dP_n \right)^{(\eta-1)/\eta} \\ (B.2) \quad & \leq \left(E_{P_n} \left[\sup_{\theta \in \Theta} |g(x, \theta)|^\eta \right] \right)^{1/\eta} \left(m_n^{-\eta} E_{P_n} \left[\sup_{\theta \in \Theta} |g(x, \theta)|^\eta \right] \right)^{(\eta-1)/\eta} = o(n^{-1/2}), \end{aligned}$$

where the first inequality follows from the Hölder inequality, and the second inequality follows from the Markov inequality, and the equality follows from $m_n^{1-\eta} \sqrt{n} = (m_n^{-\eta} n) (m_n / \sqrt{n}) \rightarrow 0$ (Assumption 2.1) and $P_n \in \mathcal{P}_{\theta_n}$ (particularly $E_{P_n} [\sup_{\theta \in \Theta} |g(x, \theta)|^\eta] \leq M$). Combining these results, we obtain the uniform convergence $\sup_{\theta \in \Theta} |E_{Q_n} [g_n(x, \theta)] - E_{P_n} [g(x, \theta)]| \rightarrow 0$. Therefore, from the triangle inequality and $|E_{Q_n} [g_n(x, \bar{T}_{Q_n})]| = O(n^{-1/2})$ (Lemma 10),

$$|E_{P_n} [g(x, \bar{T}_{Q_n})]| \leq |E_{P_n} [g(x, \bar{T}_{Q_n})] - E_{Q_n} [g_n(x, \bar{T}_{Q_n})]| + |E_{Q_n} [g_n(x, \bar{T}_{Q_n})]| \rightarrow 0.$$

The conclusion follows from $P_n \in \mathcal{P}_{\theta_n}$ (particularly $E_{P_n} [g(x, \theta_n)] = 0$ and $E_{P_n} [g(x, \tilde{\theta})] \neq 0$ for all $\tilde{\theta} \neq \theta_n$).

B.3. Proof of Lemma 3. The proof is based on Rieder (1994, proofs of Theorems 6.3.4 and Theorem 6.4.5).

Observe that

$$\begin{aligned}
& \left\| dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} (\bar{T}_{Q_n} - \theta_n)' \Lambda_n dQ_n^{1/2} \right\|^2 \\
&= \left\| dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi'_{n, Q_n} \Lambda_n dQ_n^{1/2} \right\|^2 + \left\| \frac{1}{2} (\bar{T}_{Q_n} - \theta_n - \psi_{n, Q_n})' \Lambda_n dQ_n^{1/2} \right\|^2 \\
&\quad + \left\{ \int \left(dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi'_{n, Q_n} \Lambda_n dQ_n^{1/2} \right) \Lambda_n' dQ_n^{1/2} \right\} (\bar{T}_{Q_n} - \theta_n - \psi_{n, Q_n}) \\
\text{(B.3)} \quad &= \left\| dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi'_{n, Q_n} \Lambda_n dQ_n^{1/2} \right\|^2 + \left\| \frac{1}{2} (\bar{T}_{Q_n} - \theta_n - \psi_{n, Q_n})' \Lambda_n dQ_n^{1/2} \right\|^2,
\end{aligned}$$

where the second equality follows from

$$\begin{aligned}
& \int \left\{ dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi'_{n, Q_n} \Lambda_n dQ_n^{1/2} \right\} \Lambda_n' dQ_n^{1/2} \\
&= \int \Lambda_n' \left\{ dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} \right\} dQ_n^{1/2} + \frac{1}{2} \psi'_{n, Q_n} \int \Lambda_n \Lambda_n' dQ_n = 0.
\end{aligned}$$

The left hand side of (B.3) satisfies

$$\begin{aligned}
& \left\| dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} (\bar{T}_{Q_n} - \theta_n)' \Lambda_n dQ_n^{1/2} \right\| \\
&\leq \left\| dQ_n^{1/2} - d\bar{P}_{\bar{T}_{Q_n}, Q_n}^{1/2} \right\| + o(|\bar{T}_{Q_n} - \theta_n|) + o(n^{-1/2}) \\
&\leq \left\| dQ_n^{1/2} - d\bar{P}_{\theta_n + \psi_{n, Q_n}, Q_n}^{1/2} \right\| + o(|\bar{T}_{Q_n} - \theta_n|) + o(n^{-1/2}) \\
\text{(B.4)} \quad &\leq \left\| dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi'_{n, Q_n} \Lambda_n dQ_n^{1/2} \right\| + o(|\bar{T}_{Q_n} - \theta_n|) + o(|\psi_{n, Q_n}|) + o(n^{-1/2}),
\end{aligned}$$

where the first inequality follows from the triangle inequality and Lemma 4, the second inequality follows from $\bar{T}_{Q_n} = \arg \min_{\theta \in \Theta} \left\| dQ_n^{1/2} - d\bar{P}_{\theta, Q_n}^{1/2} \right\|$, and the third inequality follows from the triangle inequality and Lemma 5. From (B.3) and (B.4),

$$\begin{aligned}
& \left\| \left\| dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi'_{n, Q_n} \Lambda_n dQ_n^{1/2} \right\|^2 + \left\| \frac{1}{2} (\bar{T}_{Q_n} - \theta_n - \psi_{n, Q_n})' \Lambda_n dQ_n^{1/2} \right\|^2 \right\|^{1/2} \\
&\leq \left\| dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi'_{n, Q_n} \Lambda_n dQ_n^{1/2} \right\| + o(|\bar{T}_{Q_n} - \theta_n|) + o(|\psi_{n, Q_n}|) + o(n^{-1/2}).
\end{aligned}$$

This implies

$$\begin{aligned}
& o(|\bar{T}_{Q_n} - \theta_n|) + o(|\psi_{n, Q_n}|) + o(n^{-1/2}) \\
\text{(B.5)} \quad &\geq \sqrt{\frac{1}{4} (\bar{T}_{Q_n} - \theta_n - \psi_{n, Q_n})' \int \Lambda_n \Lambda_n' dQ_n (\bar{T}_{Q_n} - \theta_n - \psi_{n, Q_n})} \geq C |\bar{T}_{Q_n} - \theta_n - \psi_{n, Q_n}|,
\end{aligned}$$

for all n large enough, where the second inequality follows from Lemma 8 and $P_n \in \mathcal{P}_{\theta_n}$

(particularly $\det \left(E_P [\dot{g}(x, \theta)]' E_P [g(x, \theta) g(x, \theta)']^{-1} E_P [\dot{g}(x, \theta)] \right) > c$).

We now analyze ψ_{n,Q_n} . From the definition of ψ_{n,Q_n} ,

$$(B.6) \quad \begin{aligned} \psi_{n,Q_n} &= -2 \left\{ \left(\int \Lambda_n \Lambda_n' dQ_n \right)^{-1} - V_n \right\} \int \Lambda_n \left\{ dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} \right\} dQ_n^{1/2} \\ &\quad - 2V_n \int \Lambda_n \left\{ dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} \right\} dQ_n^{1/2}. \end{aligned}$$

From this and Lemma 8, the first term of (B.6) is $o(n^{-1/2})$. The second term of (B.6) satisfies

$$\begin{aligned} &-2V_n \int \Lambda_n \left\{ dQ_n^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} \right\} dQ_n^{1/2} \\ &= -2V_n G_n' \Omega_n^{-1} \left(\int g_n(x, \theta_n) g_n(x, \theta_n)' dQ_n \right) \gamma_n(\theta_n, Q_n) \\ &\quad + 2V_n G_n' \Omega_n^{-1} \left(\int \frac{\gamma_n(\theta_n, Q_n)' g_n(x, \theta_n)}{1 + \gamma_n(\theta_n, Q_n)' g_n(x, \theta_n)} g_n(x, \theta_n) g_n(x, \theta_n)' dQ_n \right) \gamma_n(\theta_n, Q_n) \\ &= -V_n G_n' \Omega_n^{-1} \left\{ \int g_n(x, \theta_n) dQ_n + \frac{1}{2} \int \varrho_n(x, \theta_n, Q_n) g_n(x, \theta_n) dQ_n \right\} + o(n^{-1/2}) \\ &= -V_n \int \Lambda_n dQ_n + o(n^{-1/2}), \end{aligned}$$

where the first equality follows from (B.7), the second equality follows from (B.8) and Lemmas 8 and 9, and the third equality follows from Lemmas 8 and 9. Therefore,

$$\sqrt{n} \psi_{n, Q_n} = -\sqrt{n} V_n \int \Lambda_n dQ_n + o(1),$$

which also implies $|\psi_{n, Q_n}| = O(n^{-1/2})$ (by Lemma 8). Combining this with (B.5),

$$\sqrt{n} (\bar{T}_{Q_n} - \theta_n) = \sqrt{n} \psi_{n, Q_n} + o(\sqrt{n} |\bar{T}_{Q_n} - \theta_n|) + o(1).$$

By solving this equation for $\sqrt{n} (\bar{T}_{Q_n} - \theta_n)$, the conclusion is obtained.

B.4. Proof of Lemma 4. From the convex duality of partially finite programming (Borwein and Lewis (1993)), the Radon-Nikodym derivative $d\bar{P}_{\theta, Q}/dQ$ is written as

$$(B.7) \quad \frac{d\bar{P}_{\theta, Q}}{dQ} = \frac{1}{(1 + \gamma_n(\theta, Q)' g_n(x, \theta))^2},$$

for each $n \in \mathbb{N}$, $\theta \in \Theta$, and $Q \in \mathcal{M}$, where $\gamma_n(\theta, Q)$ solves

$$(B.8) \quad 0 = \int \frac{g_n(x, \theta)}{(1 + \gamma_n(\theta, Q)' g_n(x, \theta))^2} dQ = E_Q [g_n(x, \theta) \{1 - 2\gamma_n(\theta, Q)' g_n(x, \theta) + \varrho_n(x, \theta, Q)\}],$$

with

$$\varrho_n(x, \theta, Q) = \frac{3(\gamma_n(\theta, Q)' g_n(x, \theta))^2 + 2(\gamma_n(\theta, Q)' g_n(x, \theta))^3}{(1 + \gamma_n(\theta, Q)' g_n(x, \theta))^2}.$$

Denote $t_n = \bar{T}_{Q_n} - \theta_n$. From the triangle inequality and (B.7),

$$\begin{aligned} & \left\| d\bar{P}_{\bar{T}_{Q_n}, Q_n}^{1/2} - d\bar{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2}t'_n \Lambda_n dQ_n^{1/2} \right\| \\ & \leq \left\| \left\{ \gamma_n(\theta_n, Q_n)' g_n(x, \theta_n) - \gamma_n(\bar{T}_{Q_n}, Q_n)' g_n(x, \bar{T}_{Q_n}) \right\} dQ_n^{1/2} + \frac{1}{2}t'_n \Lambda_n dQ_n^{1/2} \right\| \\ & \quad + \left\| \left\{ \gamma_n(\theta_n, Q_n)' g_n(x, \theta_n) - \gamma_n(\bar{T}_{Q_n}, Q_n)' g_n(x, \bar{T}_{Q_n}) \right\} \right. \\ & \quad \left. \times \left\{ \frac{1}{(1 + \gamma_n(\bar{T}_{Q_n}, Q_n)' g_n(x, \bar{T}_{Q_n})) (1 + \gamma_n(\theta_n, Q_n)' g_n(x, \theta_n))} - 1 \right\} dQ_n^{1/2} \right\| = T_1 + T_2. \end{aligned}$$

For T_2 , Lemmas 8-11 imply $T_2 = o(n^{-1/2})$. For T_1 , the triangle inequality and (B.8) yield

$$\begin{aligned} T_1 & \leq \left\| \left\{ \begin{aligned} & -\frac{1}{2} E_{Q_n} [g_n(x, \bar{T}_{Q_n})]' E_{Q_n} [g_n(x, \bar{T}_{Q_n}) g_n(x, \bar{T}_{Q_n})']^{-1} g_n(x, \bar{T}_{Q_n}) \\ & + \frac{1}{2} E_{Q_n} [g_n(x, \theta_n)]' E_{Q_n} [g_n(x, \theta_n) g_n(x, \theta_n)]^{-1} g_n(x, \theta_n) + \frac{1}{2} t'_n \Lambda_n \end{aligned} \right\} dQ_n^{1/2} \right\| \\ & \quad + \left\| E_{Q_n} [g_n(x, \theta_n, Q_n) g_n(x, \theta_n)]' E_{Q_n} [g_n(x, \theta_n) g_n(x, \theta_n)]^{-1} g_n(x, \theta_n) dQ_n^{1/2} \right\| \\ & \quad + \left\| E_{Q_n} [g_n(x, \bar{T}_{Q_n}, Q_n) g_n(x, \bar{T}_{Q_n})]' E_{Q_n} [g_n(x, \bar{T}_{Q_n}) g_n(x, \bar{T}_{Q_n})']^{-1} g_n(x, \theta_n) dQ_n^{1/2} \right\| \\ & = T_{11} + T_{12} + T_{13}. \end{aligned}$$

Lemmas 8-11 imply that $T_{12} = o(n^{-1/2})$ and $T_{13} = o(n^{-1/2})$. For T_{11} , expansions of $g_n(x, \bar{T}_{Q_n})$ around $\bar{T}_{Q_n} = \theta_n$ yield

$$\begin{aligned} T_{11} & \leq \left\| -\frac{1}{2} E_{Q_n} [g_n(x, \bar{T}_{Q_n})]' \begin{pmatrix} E_{Q_n} [g_n(x, \bar{T}_{Q_n}) g_n(x, \bar{T}_{Q_n})']^{-1} \\ -E_{Q_n} [g_n(x, \theta_n) g_n(x, \theta_n)]^{-1} \end{pmatrix} g_n(x, \bar{T}_{Q_n}) dQ_n^{1/2} \right\| \\ & \quad + \left\| -\frac{1}{2} E_{Q_n} [g_n(x, \bar{T}_{Q_n})]' E_{Q_n} [g_n(x, \theta_n) g_n(x, \theta_n)]^{-1} \{g_n(x, \bar{T}_{Q_n}) - g_n(x, \theta_n)\} dQ_n^{1/2} \right\| \\ & \quad + \left\| -\frac{1}{2} t'_n \left(\int \dot{g}_n(x, \dot{\theta}) dQ_n - G_n \right)' E_{Q_n} [g_n(x, \theta_n) g_n(x, \theta_n)]^{-1} g_n(x, \theta_n) dQ_n^{1/2} \right\| \\ & \quad + \left\| \frac{1}{2} t'_n G'_n \left(\Omega_n^{-1} - E_{Q_n} [g_n(x, \theta_n) g_n(x, \theta_n)]^{-1} \right) g_n(x, \theta_n) dQ_n^{1/2} \right\| \\ & = o(n^{-1/2}) + o(t_n), \end{aligned}$$

where $\dot{\theta}$ is a point on the line joining θ_n and \bar{T}_{Q_n} , and the equality follows from Lemmas 8 and 10.

B.5. Proof of Lemma 5. Similar to the proof of Lemma 4.

B.6. Proof of Lemma 6.

B.6.1. Proof of the first statement. The same argument as (B.2) with $P_n \in \mathcal{P}_{\theta_n}$ (particularly $E_{P_n} [g(x, \theta_n)] = 0$) yields the conclusion.

B.6.2. *Proof of the second statement.* From the triangle inequality,

$$(B.9) \quad |E_{P_n} [g_n(x, \theta_{t,n})]| \leq |E_{P_n} [g(x, \theta_{t,n}) \mathbb{I}\{x \notin \mathcal{X}_n\}]| + |E_{P_n} [g(x, \theta_{t,n})]|.$$

By (B.2), the first term of (B.9) is $o(n^{-1/2})$. The second term of (B.9) satisfies

$$|E_{P_n} [g(x, \theta_{t,n})]| \leq E_{P_n} \left[\sup_{\tilde{\theta} \in \mathcal{N}_{\theta_n}} \left| \dot{g}(x, \tilde{\theta}) \right| \right] \left| \frac{t}{\sqrt{n}} \right| = O(n^{-1/2}),$$

for all n large enough, where the inequality follows from a Taylor expansion around $t = 0$ and $P_n \in \mathcal{P}_{\theta_n}$ (particularly $E_{P_n} [g(x, \theta_n)] = 0$) and the equality follows from $P_n \in \mathcal{P}_{\theta_n}$ (particularly $E_{P_n} \left[\sup_{\tilde{\theta} \in \mathcal{N}_{\theta_n}} \left| \dot{g}(x, \tilde{\theta}) \right|^2 \right] < M$). Combining these results, the conclusion is obtained.

B.6.3. *Proof of the third statement.* From the triangle inequality,

$$\begin{aligned} & |E_{P_n} [g_n(x, \theta_{t,n}) g_n(x, \theta_{t,n})'] - \Omega_n| \\ & \leq |E_{P_n} [g_n(x, \theta_{t,n}) g_n(x, \theta_{t,n})'] - E_{P_n} [g(x, \theta_{t,n}) g(x, \theta_{t,n})']| + |E_{P_n} [g(x, \theta_{t,n}) g(x, \theta_{t,n})'] - \Omega_n|. \end{aligned}$$

The first term is $o(n^{-1/2})$ by the same argument as (B.2) and the second term converges to zero by the continuity of $g(x, \theta)$ at θ_n .

B.6.4. *Proof of the fourth statement.* Similar to the proof of the third statement.

B.7. **Proof of Lemma 7.** Let $\Gamma_n = \{\gamma \in \mathbb{R}^m : |\gamma| \leq a_n\}$ with a positive sequence $\{a_n\}_{n \in \mathbb{N}}$ satisfying $a_n m_n \rightarrow 0$ and $a_n n^{1/2} \rightarrow \infty$. Observe that

$$(B.10) \quad \sup_{\gamma \in \Gamma_n, x \in \mathcal{X}, \theta \in \Theta} |\gamma' g_n(x, \theta)| \leq a_n m_n \rightarrow 0.$$

Since $R_n(P_n, \theta_{t,n}, \gamma)$ is twice continuously differentiable with respect to γ and Γ_n is compact,

$\tilde{\gamma} = \arg \max_{\gamma \in \Gamma_n} R_n(P_n, \theta_{t,n}, \gamma)$ exists for each $n \in \mathbb{N}$. A Taylor expansion around $\tilde{\gamma} = 0$ yields

$$\begin{aligned} -1 & = R_n(P_n, \theta_{t,n}, 0) \leq R_n(P_n, \theta_{t,n}, \tilde{\gamma}) = -1 + \tilde{\gamma}' E_{P_n} [g_n(x, \theta_{t,n})] - \tilde{\gamma}' E_{P_n} \left[\frac{g_n(x, \theta_{t,n}) g_n(x, \theta_{t,n})'}{(1 + \tilde{\gamma}' g_n(x, \theta_{t,n}))^3} \right] \tilde{\gamma} \\ & \leq -1 + \tilde{\gamma}' E_{P_n} [g_n(x, \theta_{t,n})] - C \tilde{\gamma}' E_{P_n} [g_n(x, \theta_{t,n}) g_n(x, \theta_{t,n})'] \tilde{\gamma} \\ (B.11) & \leq -1 + |\tilde{\gamma}| |E_{P_n} [g_n(x, \theta_{t,n})]| - C |\tilde{\gamma}|^2, \end{aligned}$$

for all n large enough, where $\dot{\gamma}$ is a point on the line joining 0 and $\tilde{\gamma}$, the second inequality follows from (B.10), and the last inequality follows from Lemma 6 and $P_n \in \mathcal{P}_{\theta_n}$ (particularly $\det(E_{P_n} [g(x, \theta_n) g(x, \theta_n)']) \geq c$). Thus, Lemma 6 implies

$$(B.12) \quad C |\tilde{\gamma}| \leq |E_{P_n} [g_n(x, \theta_{t,n})]| = O(n^{-1/2}).$$

From $a_n n^{1/2} \rightarrow \infty$, $\tilde{\gamma}$ is an interior point of Γ_n and satisfies the first-order condition $\partial R_n(Q_n, \theta_n, \tilde{\gamma}) / \partial \gamma = 0$ for all n large enough. Since $R_n(Q_n, \theta_n, \gamma)$ is concave in γ for all n large enough, $\tilde{\gamma} = \arg \max_{\gamma \in \mathbb{R}^m} R_n(P_n, \theta_{t,n}, \gamma)$

for all n large enough and the first statement is obtained. Thus, the second statement is obtained from (B.12). The third statement follows from (B.12) and $\frac{m_n}{\sqrt{n}} \rightarrow 0$ (Assumption 2.1).

B.8. Proof of Lemma 8.

B.8.1. *Proof of the first statement.* We have

$$\begin{aligned}
& |E_{Q_n} [g_n(x, \theta_n)]| \\
& \leq \left| \int g_n(x, \theta_n) \{dQ_n - dP_n\} \right| + |E_{P_n} [g_n(x, \theta_n)]| \\
& \leq \left| \int g_n(x, \theta_n) \{dQ_n^{1/2} - dP_n^{1/2}\}^2 \right| + 2 \left| \int g_n(x, \theta_n) dP_n^{1/2} \{dQ_n^{1/2} - dP_n^{1/2}\} \right| + o(n^{-1/2}) \\
& \leq m_n \frac{r_0^2}{n} + 2E_{P_n} [|g(x, \theta_n)|^2] \frac{r_0}{\sqrt{n}} + o(n^{-1/2}) = O(n^{-1/2}),
\end{aligned}$$

where the first and second inequalities follow from the triangle inequality and Lemma 6, the third inequality follows from the Cauchy-Schwarz inequality and $H(Q_n, P_n) \leq \frac{r_0}{\sqrt{n}}$, and the equality follows from Assumption 2.1 and $P_n \in \mathcal{P}_{\theta_n}$ (particularly $E_{P_n} [\sup_{\theta \in \Theta} |g(x, \theta)|^\eta] \leq M$).

B.8.2. *Proof of the second statement.* From the triangle inequality,

$$\begin{aligned}
& |E_{Q_n} [g_n(x, \theta_n) g_n(x, \theta_n)'] - \Omega_n| \\
\text{(B.13)} & \leq |E_{Q_n} [g_n(x, \theta_n) g_n(x, \theta_n)'] - E_{P_n} [g_n(x, \theta_n) g_n(x, \theta_n)']| + |E_{P_n} [g(x, \theta_n) g(x, \theta_n)' \mathbb{I}\{x \notin \mathcal{X}_n\}]|.
\end{aligned}$$

The first term of (B.13) satisfies

$$\begin{aligned}
& |E_{Q_n} [g_n(x, \theta_n) g_n(x, \theta_n)'] - \Omega_n| \\
& \leq \left| \int g_n(x, \theta_n) g_n(x, \theta_n)' \{dQ_n^{1/2} - dP_n^{1/2}\}^2 \right| + 2 \left| \int g_n(x, \theta_n) g_n(x, \theta_n)' dP_n^{1/2} \{dQ_n^{1/2} - dP_n^{1/2}\} \right| \\
& \leq m_n^2 \frac{r_0^2}{n} + 2E_{P_n} [|g(x, \theta_n)|^4] \frac{r_0}{\sqrt{n}} = o(1),
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the Cauchy-Schwarz inequality and $H(Q_n, P_n) \leq \frac{r_0}{\sqrt{n}}$, and the equality follows from $\frac{m_n}{\sqrt{n}} \rightarrow 0$ (Assumption 2.1) and $P_n \in \mathcal{P}_{\theta_n}$ (particularly $E_{P_n} [\sup_{\theta \in \mathcal{N}} |g(x, \theta)|^4] \leq M$). The second term of (B.13) satisfies

$$\begin{aligned}
& |E_{P_n} [g(x, \theta_n) g(x, \theta_n)' \mathbb{I}\{x \notin \mathcal{X}_n\}]| \\
& \leq \left(\int |g(x, \theta_n) g(x, \theta_n)'|^{1+\delta} dP_n \right)^{\frac{1}{1+\delta}} \left(\int \mathbb{I}\{x \notin \mathcal{X}_n\} dP_n \right)^{\frac{\delta}{1+\delta}} \\
& \leq \left(E_{P_n} [|g(x, \theta_n)|^{2+\delta}] \right)^{\frac{1}{1+\delta}} (m_n^{-\eta} E_{P_n} [|g(x, \theta_n)|^\eta])^{\frac{\delta}{1+\delta}} = o(1),
\end{aligned}$$

for sufficiently small $\delta > 0$, where the first inequality follows from the Hölder inequality, the second inequality follows from the Markov inequality, and the equality follows from Assumption 2.1 and $P_n \in \mathcal{P}_{\theta_n}$ (particularly $E_{P_n} \left[\sup_{\tilde{\theta} \in \mathcal{N}_{\theta_n}} \left| g(x, \tilde{\theta}) \right|^4 \right] \leq M$).

B.9. Proof of Lemma 9. Similar to the proof of Lemma 7. Repeat the same argument with $R_n(Q_n, \theta_n, \gamma)$ instead of $R_n(P_n, \theta_{t,n}, \gamma)$.

B.10. Proof of Lemma 10.

B.10.1. Proof of the first statement. Define $\tilde{\gamma} = \frac{E_{Q_n}[g_n(x, \bar{T}_{Q_n})]}{\sqrt{n} |E_{Q_n}[g_n(x, \bar{T}_{Q_n})]|}$. From $|\tilde{\gamma}| = n^{-1/2}$ and Assumption 2.1,

$$(B.14) \quad \sup_{x \in \mathcal{X}, \theta \in \Theta} |\tilde{\gamma}' g_n(x, \theta)| \leq n^{-1/2} m_n \rightarrow 0.$$

Observe that

$$\begin{aligned} & \left| E_{Q_n} \left[g_n(x, \bar{T}_{Q_n}) g_n(x, \bar{T}_{Q_n})' \right] \right| \\ & \leq \int \sup_{\theta \in \Theta} |g_n(x, \theta)|^2 \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\}^2 + 2 \int \sup_{\theta \in \Theta} |g_n(x, \theta)|^2 dP_n^{1/2} \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\} + E_{P_n} \left[\sup_{\theta \in \Theta} |g_n(x, \theta)|^2 \right] \\ (B.15) \quad & n_n^2 \frac{r_0^2}{n} + 2m_n \sqrt{E_{P_n} \left[\sup_{\theta \in \Theta} |g(x, \theta)|^2 \right]} \frac{r_0}{\sqrt{n}} + E_{P_n} \left[\sup_{\theta \in \Theta} |g(x, \theta)|^2 \right] \leq C E_{P_n} \left[\sup_{\theta \in \Theta} |g(x, \theta)|^2 \right], \end{aligned}$$

for all n large enough, where the first inequality follows from the triangle inequality, the second inequality follows from the Cauchy-Schwarz inequality and $H(Q_n, P_n) \leq \frac{r_0}{\sqrt{n}}$, and the last inequality follows from Assumption 2.1. Thus, an expansion around $\tilde{\gamma} = 0$ yields

$$\begin{aligned} R_n(Q_n, \bar{T}_{Q_n}, \tilde{\gamma}) &= -1 + \tilde{\gamma}' E_{Q_n} [g_n(x, \bar{T}_{Q_n})] - \tilde{\gamma}' E_{Q_n} \left[\frac{g_n(x, \bar{T}_{Q_n}) g_n(x, \bar{T}_{Q_n})'}{(1 + \tilde{\gamma}' g_n(x, \bar{T}_{Q_n}))^3} \right] \tilde{\gamma} \\ &\geq -1 + n^{-1/2} |E_{Q_n} [g_n(x, \bar{T}_{Q_n})]| - C \tilde{\gamma}' E_{Q_n} [g_n(x, \bar{T}_{Q_n}) g_n(x, \bar{T}_{Q_n})'] \tilde{\gamma} \\ (B.16) \quad &\geq -1 + n^{-1/2} |E_{Q_n} [g_n(x, \bar{T}_{Q_n})]| - C n^{-1}, \end{aligned}$$

for all n large enough, where $\tilde{\gamma}$ is a point on the line joining 0 and $\tilde{\gamma}$, the first inequality follows from (B.14), and the second inequality follows from $\tilde{\gamma}' \tilde{\gamma} = n^{-1}$ and (B.15). From the duality of partially finite programming (Borwein and Lewis (1993)), $\gamma_n(\bar{T}_{Q_n}, Q_n)$ and \bar{T}_{Q_n} are written as $\gamma_n(\bar{T}_{Q_n}, Q_n) = \arg \max_{\gamma \in \mathbb{R}^m} R_n(Q_n, \bar{T}_{Q_n}, \gamma)$ and $\bar{T}_{Q_n} = \arg \min_{\theta \in \Theta} R_n(Q_n, \theta, \gamma_n(\theta, Q_n))$. Therefore, from (B.16),

$$\begin{aligned} & -1 + n^{-1/2} |E_{Q_n} [g_n(x, \bar{T}_{Q_n})]| - C n^{-1} \\ (B.17) \quad & \leq R_n(Q_n, \bar{T}_{Q_n}, \tilde{\gamma}) \leq R_n(Q_n, \bar{T}_{Q_n}, \gamma_n(\bar{T}_{Q_n}, Q_n)) \leq R_n(Q_n, \theta_n, \gamma_n(\theta_n, Q_n)). \end{aligned}$$

By a similar argument to (B.11) combined with $|\gamma_n(\theta_n, Q_n)| = O(n^{-1/2})$ and $|E_{Q_n} [g_n(x, \theta_n)]| = O(n^{-1/2})$ (by Lemmas 8 and 9), we have

$$(B.18) \quad R_n(Q_n, \theta_n, \gamma_n(\theta_n, Q_n)) \leq -1 + |\gamma_n(\theta_n, Q_n)| |E_{Q_n} [g_n(x, \theta_n)]| - C |\gamma_n(\theta_n, Q_n)|^2 = -1 + O(n^{-1}).$$

From (B.17) and (B.18), the conclusion follows.

B.10.2. *Proof of the second statement.* Similar to the proof of the second statement of Lemma 8.

B.10.3. *Proof of the third statement.* From the triangle inequality,

$$(B.19) \quad \begin{aligned} |E_{Q_n} [\dot{g}_n(x, \bar{T}_{Q_n})] - G_n| &\leq |E_{Q_n} [\dot{g}_n(x, \bar{T}_{Q_n})] - E_{P_n} [\dot{g}_n(x, \bar{T}_{Q_n})]| \\ &\quad + |E_{P_n} [\mathbb{I}\{x \notin \mathcal{X}_n\} \dot{g}(x, \bar{T}_{Q_n})]| + |E_{P_n} [\dot{g}(x, \bar{T}_{Q_n})] - G_n|. \end{aligned}$$

The first term of (B.19) satisfies

$$\begin{aligned} &|E_{Q_n} [\dot{g}_n(x, \bar{T}_{Q_n})] - E_{P_n} [\dot{g}_n(x, \bar{T}_{Q_n})]| \\ &\leq \left| \int \dot{g}_n(x, \bar{T}_{Q_n}) \{dQ_n^{1/2} - dP_n^{1/2}\}^2 \right| + 2 \left| \int \dot{g}_n(x, \bar{T}_{Q_n}) dP_n^{1/2} \{dQ_n^{1/2} - dP_n^{1/2}\} \right| \\ &\leq \sup_{x \in \mathcal{X}_n, \tilde{\theta} \in \mathcal{N}_{\theta_n}} \left| \dot{g}_n(x, \tilde{\theta}) \right| \frac{r_0^2}{n} + 2E_{P_n} \left[\sup_{\tilde{\theta} \in \mathcal{N}_{\theta_n}} \left| \dot{g}_n(x, \tilde{\theta}) \right|^2 \right] \frac{r_0}{\sqrt{n}} = o(1), \end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the Cauchy-Schwarz inequality, and the equality follows from $P_n \in \mathcal{P}_{\theta_n}$ (particularly $\sup_{x \in \mathcal{X}_n, \tilde{\theta} \in \mathcal{N}_{\theta_n}} \left| \dot{g}(x, \tilde{\theta}) \right| = o(n^{1/2})$ and $E_{P_n} \left[\sup_{\tilde{\theta} \in \mathcal{N}_{\theta_n}} \left| \dot{g}(x, \tilde{\theta}) \right|^2 \right] < M$). The second term of (B.19) is $o(1)$ by the same argument as (B.2). The third term of (B.19) is $o(1)$ by $P_n \in \mathcal{P}_{\theta_n}$ (particularly by the continuity of $\dot{g}(x, \theta)$ at θ_n) and Lemma 2. Therefore, the conclusion is obtained.

B.11. **Proof of Lemma 11.** Similar to the proof of Lemma 7. Repeat the same argument with $R_n(Q_n, \bar{T}_{Q_n}, \gamma)$ instead of $R_n(P_n, \theta_{t,n}, \gamma)$.

B.12. **Proof of Lemma 12.** Similar to the proof of Lemma 2.

B.13. **Proof of Lemma 13.** The proof of the first statement is similar to that of Lemma 3 (replace Q_n with \hat{P}_n and use Lemmas 14-17 instead of Lemmas 8-11). For the second statement, Lemma 3 and the first statement of this lemma imply

$$\sqrt{n} \left(\bar{T}_{\hat{P}_n} - \bar{T}_{Q_n} \right) = -V_n G_n' \Omega_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \{g_n(x_i, \theta_n) - E_{Q_n} [g_n(x, \theta_n)]\} + o_p(1),$$

under Q_n . Thus, it is sufficient to check that we can apply a central limit theorem to the triangular array $\{V_n G_n' \Omega_n^{-1} g_n(x_i, \theta_n)\}_{1 \leq i \leq n, n}$. From $P_n \in \mathcal{P}_{\theta_n}$, we have $|V_n|^{2+\delta} < \infty$ (by $\det \left(E_{P_n} [\dot{g}(x, \theta_n)]' E_{P_n} [g(x, \theta_n) g(x, \theta_n)']^{-1} E_{P_n} [\dot{g}(x, \theta_n)] \right) > c$), $|G_n|^{2+\delta}$ (by $E_{P_n} \left[\sup_{\tilde{\theta} \in \mathcal{N}_{\theta_n}} \left| \dot{g}(x, \tilde{\theta}) \right|^2 \right] < M$), and $|\Omega_n^{-1}|^{2+\delta} < \infty$ (by $\det \left(E_{P_n} [g(x, \theta_n) g(x, \theta_n)'] \right) > c$). Thus,

$$|V_n G_n' \Omega_n^{-1}|^{2+\delta} < \infty$$

Also observe that

$$\begin{aligned}
& E_{Q_n} \left[|g_n(x, \theta_n)|^{2+\delta} \right] \\
&= \int |g_n(x, \theta_n)|^{2+\delta} \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\}^2 + 2 \int |g_n(x, \theta_n)|^{2+\delta} dP_n^{1/2} \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\} + E_{P_n} \left[|g_n(x, \theta_n)|^{2+\delta} \right] \\
&\leq m_n^{2+\delta} \frac{r_0^2}{n} + 2m_n^{1+\delta} E_{P_n} \left[|g(x, \theta_n)|^2 \right] \frac{r_0}{\sqrt{n}} + E_{P_n} \left[|g(x, \theta_n)|^{2+\delta} \right] < \infty,
\end{aligned}$$

for all n large enough, where the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality follows from $P_n \in \mathcal{P}_{\theta_n}$ (particularly $E_{P_n} [\sup_{\theta \in \Theta} |g(x, \theta)|^q] \leq M$). Therefore, the conclusion is obtained.

B.14. Proof of Lemma 14.

B.14.1. *Proof of the first statement.* From the triangle inequality,

$$\left| E_{\hat{P}_n} [g_n(x, \theta_n)] \right| \leq \sum_{j=1}^m \left| \frac{1}{n} \sum_{i=1}^n W_{ijn} \right| \sqrt{\text{Var}_{Q_n}(g_{jn}(x, \theta_n))} + |E_{Q_n} [g_n(x, \theta_n)]|.$$

where $W_{ijn} = \frac{g_{jn}(x_i, \theta_n) - E_{Q_n}[g_{jn}(x, \theta_n)]}{\sqrt{\text{Var}_{Q_n}(g_{jn}(x, \theta_n))}}$. From Lemma 8, the second term is $O(n^{-1/2})$. Thus, if we can show that

$$(B.20) \quad \text{Var}_{Q_n}(g_{jn}(x, \theta_n)) < C,$$

$$(B.21) \quad E \left[|W_{ijn}|^{2+\delta} \right] < \infty \quad \text{for some } \delta > 0,$$

for all n large enough, then Lyapounov's central limit theorem yields the conclusion.

We first check (B.20). Observe that

$$\begin{aligned}
\text{Var}_{Q_n}(g_{jn}(x, \theta_n)) &= E_{Q_n} \left[g_{jn}(x, \theta_n)^2 \right] - (E_{Q_n} [g_{jn}(x, \theta_n)])^2 \\
&= E_{P_n} \left[g_j(x, \theta_n)^2 \right] + O(n^{-1}) \\
(B.22) \quad &\in (C_1, C_2),
\end{aligned}$$

for some $C_1, C_2 > 0$, where the second equality follows from Lemma 8, and the inclusion relation follows from $P_n \in \mathcal{P}_{\theta_n}$ (particularly $\text{Var}_{P_n}(g_j(x, \theta_n)) \in (c, M)$). Thus, (B.20) is satisfied.

We now check (B.21). Using (B.22),

$$\begin{aligned}
& E_{Q_n} \left[|W_{ijn}|^{2+\delta} \right] \leq C E_{Q_n} \left[|g_{jn}(x, \theta_n)|^{2+\delta} \right] \\
&\leq \int |g_{jn}(x, \theta_n)|^{2+\delta} \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\}^2 + 2 \int |g_{jn}(x, \theta_n)|^{2+\delta} dP_n^{1/2} \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\} + E_{P_n} \left[|g_{jn}(x, \theta_n)|^{2+\delta} \right] \\
&\stackrel{(B.22)}{\leq} m_n^{2+\delta} \frac{r_0^2}{n} + 2m_n^{1+\frac{\delta}{2}} \frac{r_0}{\sqrt{n}} \sqrt{E_{P_n} \left[|g_{jn}(x, \theta_n)|^{2+\delta} \right]} + E_{P_n} \left[|g_{jn}(x, \theta_n)|^{2+\delta} \right] \leq C,
\end{aligned}$$

for some $\delta > 0$, where the second inequality follows from the triangle inequality, the third inequality follows from the Cauchy-Schwarz inequality, and the last inequality follows from Assumption 2.1 and $P_n \in \mathcal{P}_{\theta_n}$ (particularly $E_{P_n} [|g(x, \theta_n)|^q] < M$).

B.14.2. *Proof of the second statement.* From the triangle inequality,

$$\begin{aligned} & \left| E_{\hat{P}_n} [g_n(x, \theta_n) g_n(x, \theta_n)' - \Omega_n] \right| \\ & \leq \left| E_{\hat{P}_n} [g_n(x, \theta_n) g_n(x, \theta_n)'] - E_{Q_n} [g_n(x, \theta_n) g_n(x, \theta_n)'] \right| + \left| E_{Q_n} [g_n(x, \theta_n) g_n(x, \theta_n)'] - \Omega_n \right|. \end{aligned}$$

From (B.23), we can apply a law of large numbers for the triangular array $\{g_n(x_i, \theta_n) g_n(x, \theta_n)'\}_{1 \leq i \leq n, n}$ (see Romano (2004)), and the first term is $o_p(1)$. From Lemma 8, the second term is $o(1)$.

B.15. **Proof of Lemma 15.** Similar to the proof of Lemma 7 except using Lemma 14 instead of Lemma 6.

B.16. **Proof of Lemma 16.** Similar to the proof of Lemma 10.

B.17. **Proof of Lemma 17.** Similar to the proof of Lemma 11.

REFERENCES

- BERAN, R. (1977): "Minimum Hellinger distance estimates for parametric models," *Annals of Statistics*, 5, 445–463.
- (1981): "Efficient robust tests in parametric models," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57, 73–86.
- BICKEL, P. J. (1981): "Quelques aspects de la statistique robuste," in *Ecole d'Été de Probabilités de Saint Flour IX 1979*, ed. by P. Hennequin, pp. 1–72. Springer.
- BORWEIN, J. M., AND A. S. LEWIS (1993): "Partially-finite programming in L_1 and the existence of maximum entropy estimates," *SIAM Journal of Optimization*, 3, 248–267.
- HUBER, P. J., AND V. STRASSEN (1973): "Minimax tests and the Neyman-Pearson lemma for capacities," *Annals of Statistics*, 1, 251–263.
- IMBENS, G. W., R. H. SPADY, AND P. JOHNSON (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333–357.
- KITAMURA, Y. (2006): "Empirical Likelihood Methods in Econometrics: Theory and Practice," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by R. Blundell, W. K. Newey, and T. Persson. Cambridge University Press, Cambridge, forthcoming.
- KITAMURA, Y., T. OTSU, AND K. EVDOKIMOV (2009): "Robustness, infinitesimal neighborhoods, and moment restrictions," Manuscript, Department of Economics, Yale University.
- KITAMURA, Y., AND M. STUTZER (1997): "An Information Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65(4), 861–874.
- NEWKEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, vol. IV*, ed. by R. Engle, and D. McFadden, pp. 2111–2245. Elsevier Science B.V.
- NEWKEY, W. K., AND R. J. SMITH (2004): "Higher order properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255.
- NEWKEY, W. K., AND K. D. WEST (1987): "Hypothesis testing with efficient method of moments estimation," *International Economic Review*, 28, 777–787.
- QIN, J., AND J. LAWLESS (1994): "Empirical Likelihood and General Estimating Equations," *Annals of Statistics*, 22, 300–325.
- RIEDER, H. (1978): "A robust asymptotic testing model," *Annals of Statistics*, 6, 1080–1094.
- RIEDER, H. (1994): *Robust Asymptotic Statistics*. Springer-Verlag.
- SMITH, R. J. (1997): "Alternative semi-parametric likelihood approaches to generalized method of moments estimation," *Economic Journal*, 107, 503–519.

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS, YALE UNIVERSITY, NEW HAVEN, CT-06520.

E-mail address: `yuichi.kitamura@yale.edu`

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS, YALE UNIVERSITY, NEW HAVEN, CT-06520.

E-mail address: `taisuke.otsu@yale.edu`