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 $\sigma$-field $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ of $\mathcal{X} \subseteq \mathbb{R}^d$. Let $g : \mathcal{X} \times \Theta \to \mathbb{R}^m$ be a vector of functions, which satisfies the moment condition:

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

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

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

for a known function $\tau : \Theta \to \mathbb{R}$. Define $\Theta_0 = \{ \theta \in \Theta | \tau(\theta) = 0 \}$, and we assume that $\Theta_0 \subset int(\Theta)$ in order to avoid the technicalities associated with the inference at the boundary of $\Theta$. 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 contamination.

\[ \text{Date: This Version: May 31, 2010.} \]
\[ \text{Keywords: Asymptotic Minimax Theorem, Hellinger Distance, Size Distortion.} \]
\[ \text{JEL Classification Number: C14.} \]

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.

\[ ^1 \text{Although we consider the two-sided alternative hypothesis } H_1, \text{ we can derive an analogous result for the one-sided alternatives } (H_1^+ : \tau(\theta_0) > 0 \text{ or } H_1^- : \tau(\theta_0) < 0). \]
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 is 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

\[ X_n = \left\{ x \in X : \sup_{\theta \in \Theta} |g(x, \theta)| \leq m_n \right\}, \]

where \( \{m_n\}_{n \in \mathbb{N}} \) is a sequence of positive numbers satisfying \( m_n \to \infty \) as \( n \to \infty \). The set \( X_n \) 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 \( \mathcal{P}_\theta \) 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 \( X \sim P \). Moreover, we consider IID observations throughout the paper. Define \( \dot{g}(x, \theta) = \partial g(x, \theta)/\partial \theta' \).

**Definition 1.** For each \( \theta \) in \( \Theta \), let \( \mathcal{P}_\theta \) denote the set of all probability measures that satisfy the following conditions:

\[(i)\]

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

where \( c, M \in (0, \infty) \) are fixed constants (which do not depend on \( \theta \) or \( P \)).
The null hypothesis in (1.2) leads to the following definition:

\[ \mathcal{P}_N = \bigcup_{\theta \in \Theta} \mathcal{P}_\theta, \]

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

(1.3) \[ \sup_{Q \in \mathcal{P}_N} E_Q[\psi_n]. \]

If a measure \( P \) belongs to \( \mathcal{P}_N \) 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) \[ \alpha_{\psi,n} = \sup_{Q \in \mathcal{Q}_n} E_Q[\psi_n], \]

where

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

for some \( \eta > 2 \), and

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

is the Hellinger distance between two measures \( Q \) and \( P \) with densities \( p \) and \( q \) with respect to a dominating measure \( \nu \). 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 \( \psi_n \) under local perturbations.

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

(i): \( P_\theta \) belongs to \( \mathcal{P}_\theta \) 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

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

for all \( t \in \mathbb{R}^p \).
The local power function along this model is given by $E_{P_{\hat{\theta} + t/\sqrt{n}}} [\psi_n]$. Instead, consider the minimal power of $\psi$ against a class of perturbed versions of $P_{\hat{\theta} + t/\sqrt{n}}$:

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

where

$$
Q_n^{LA}(P) = \left\{ Q \in \mathcal{M} : H(Q, P) \leq \frac{r_1}{\sqrt{n}}, E_Q \left[ \sup_{\theta \in \Theta} |g(x, \theta)|^2 \right] \leq M \right\},
$$

for $r_1 > 0$. In contrast to $Q_n^N$ in the definition of $\alpha_{\psi, n}$, the set $Q_n^{LA}(P_{\hat{\theta} + t/\sqrt{n}})$ consists of local deviations from the measure $P_{\hat{\theta} + t/\sqrt{n}}$ (as opposed to the set of measures $P^N$). It is of great interest to obtain a minimax test that maximizes the minimal power against the set of perturbed versions of $P_{\hat{\theta} + t/\sqrt{n}}$, 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 $\theta$ achieves asymptotic optimality for both $\alpha_{\psi, n}$ and $\beta_{\psi, n} (P_{\hat{\theta} + t/\sqrt{n}})$. The test minimizes the size distortion $\alpha$ 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 $\beta_{\psi, n} (P_{\hat{\theta} + t/\sqrt{n}})$.

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 not in Kitamura, Otsu, and Evdokimov (2009), the MHDE of Beran (1977) applied to the moment condition model (1.1) yields

$$
\hat{\theta} = \arg \min_{\theta \in \Theta} \inf_{\gamma : f \in \mathcal{F}, P \in \mathcal{P}_0} H(P, \hat{P}_n) = \arg \min_{\theta \in \Theta} \max_{\gamma \in \mathbb{R}^n} - \frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \gamma' g(x_i, \theta)},
$$

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))

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

The asymptotic variance can be estimated by

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

where $\hat{G} = \frac{1}{n} \sum_{i=1}^n g(x_i, \hat{\theta})$ 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:

$$
\psi_{H,n} = \mathbb{I} \left\{ \left( \frac{\sqrt{n} \tau (\hat{\theta})}{\hat{\sigma}} \right)^2 \geq \text{critical value} \right\},
$$
where \( \mathbb{1}\{ \cdot \} \) is the indicator function, \( \hat{\sigma} = \hat{\tau} \left( \hat{\theta} \right)' \hat{V} \hat{\tau} \left( \hat{\theta} \right) \), and \( \hat{\tau} \left( \theta \right) = d \tau \left( \theta \right)/d\theta \). 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 \( \chi^2 \) 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): \( \Theta \) is compact;

(ii): \( \{m_n\}_{n \in \mathbb{N}} \) satisfies \( m_n \to \infty \), \( nm_n^{-\eta} \to 0 \), and \( n^{-1/2}m_n^{1+\epsilon} = O(1) \) for some \( 0 < \epsilon < 2 \) as \( n \to \infty \);

(iii): The second-order derivative of \( \tau \) is bounded on \( \Theta \);

(iv): For each \( \theta \in \Theta_0 \), \( \sup_{x \in X_\theta, \theta \in N_\theta} \left| \hat{g}(x, \hat{\theta}) \right| = o(n^{1/2}) \) for a neighborhood \( N_\theta \) around \( \theta \).

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 \( \hat{\theta}_a = T_a \left( \hat{P}_n \right) \) be a point estimator based on a mapping \( T_a : \mathcal{M} \to \Theta \). The mapping \( T \) for the MHDE \( \hat{\theta} = T \left( \hat{P}_n \right) \) is defined as \( T(Q) = \arg \min_{\theta \in \Theta} \inf_{\bar{P} \in \mathcal{P}_\theta} \inf_{Q < Q} H(P, Q). \) We use the following concepts on the mapping \( T_a \).

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

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

\[
\sqrt{n} \left( T_a \left( P_{\theta+t/\sqrt{n}, \zeta, \rho} \right) - \bar{\theta} \right) \to t,
\]

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

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

\[
\sqrt{n} \left( T_a \left( \hat{P}_n \right) - T_a \left( P_{\theta+t/\sqrt{n}, \zeta, \rho} \right) \right) \overset{d}{\to} N(0, \hat{V}_a), \quad \text{under } P_{\theta+t/\sqrt{n}, \zeta, \rho},
\]

holds for every submodel sequence \( \left\{ P_{\theta+t/\sqrt{n}, \zeta, \rho} \right\}_{n \in \mathbb{N}} \) (with any \( \theta \in \Theta \), \( t \in \mathbb{R} \), \( \zeta_n = O(n^{-1/2}) \), and \( \bar{P} \in \mathcal{P}_\theta \)) satisfying \( P_{\theta+t/\sqrt{n}, \zeta, \rho} \in \left\{ Q : H(Q, \bar{P}) \leq \frac{\rho}{\sqrt{n}} \right\} \) eventually, where the asymptotic variance \( \hat{V}_a \).
Theorem 4. Let $\chi^2$ be the non-central $\chi^2$ distribution with degree of freedom 1 and non-centrality $4\nu^2_0$. Suppose that Assumption 2.1 holds. Then:

(i): For every test $\psi_n \in S$ in Definition 3,

$$\liminf_{n \to \infty} \sup_{Q \in \mathcal{Q}^N_n} E_Q[\psi_n] \geq \Pr \left\{ W \geq \chi_{1,\alpha}^2 \right\},$$

(ii): for some point estimator $\hat{\theta}_n$, it holds $S_n = \left( \frac{\sqrt{n} \tau(\hat{\theta}_n)}{\sigma_n} \right)^2 + o_p(1)$ under each submodel sequence

$$\left\{ P_{\hat{\theta}_n + t/\sqrt{\pi_n} \cdot \zeta_n \cdot \dot{P}} \right\}_{n \in \mathbb{N}},$$

with any $\hat{\theta} \in \Theta$, $t \in \mathbb{R}$, $\zeta_n = O\left(n^{-1/2}\right)$, and $\dot{P} \in \mathcal{P}_\theta$ satisfying $P_{\hat{\theta}_n + t/\sqrt{\pi_n} \cdot \zeta_n \cdot \dot{P}} \in \left\{ Q : H(Q, \dot{P}) \leq \frac{\nu^2_0}{\sqrt{n}} \right\}$ eventually,

(iii): the mapping $T_n$ to define $\hat{\theta}_n = T_n\left( \dot{P}_n \right)$ satisfies the requirements in Definition 2.

Note that the class $S$ contains tests that are asymptotically equivalent to the Wald test based on some estimator $\hat{\theta}_n$ that satisfies the requirements in Definition 2, with the asymptotic critical value $\chi_{1,\alpha}^2$. Therefore, the class $S$ 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 $S$.

Theorem 4. Let $W$ be a random variable that obeys the non-central $\chi^2$ distribution with degree of freedom 1 and non-centrality $4\nu^2_0$. Suppose that Assumption 2.1 holds. Then:

(i): For every test $\psi_n \in S$ in Definition 3,

$$\liminf_{n \to \infty} \sup_{Q \in \mathcal{Q}^N_n} E_Q[\psi_n] \geq \Pr \left\{ W \geq \chi_{1,\alpha}^2 \right\},$$

(ii): for some point estimator $\hat{\theta}_n$, it holds $S_n = \left( \frac{\sqrt{n} \tau(\hat{\theta}_n)}{\sigma_n} \right)^2 + o_p(1)$ under each submodel sequence

$$\left\{ P_{\hat{\theta}_n + t/\sqrt{\pi_n} \cdot \zeta_n \cdot \dot{P}} \right\}_{n \in \mathbb{N}},$$

with any $\hat{\theta} \in \Theta$, $t \in \mathbb{R}$, $\zeta_n = O\left(n^{-1/2}\right)$, and $\dot{P} \in \mathcal{P}_\theta$ satisfying $P_{\hat{\theta}_n + t/\sqrt{\pi_n} \cdot \zeta_n \cdot \dot{P}} \in \left\{ Q : H(Q, \dot{P}) \leq \frac{\nu^2_0}{\sqrt{n}} \right\}$ eventually,

(iii): the mapping $T_n$ to define $\hat{\theta}_n = T_n\left( \dot{P}_n \right)$ satisfies the requirements in Definition 2.

Note that the class $S$ contains tests that are asymptotically equivalent to the Wald test based on some estimator $\hat{\theta}_n$ that satisfies the requirements in Definition 2, with the asymptotic critical value $\chi_{1,\alpha}^2$. Therefore, the class $S$ 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)).
for each $r_0 > 0$.

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

$$\lim_{n \to \infty} \sup_{Q \in Q_n^N} E_Q [\psi_{H,n}] = Pr \{ W \geq \chi^2_{1,\alpha} \},$$

for each $r_0 > 0$.

2.2. Power Optimality.

2.2.1. Case I: One-sided $H_1 \&$ Scalar $\theta$. Consider the one-sided alternative hypothesis

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

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

$$\mathcal{P}^N = \mathcal{P}_0,$$

$$Q_n^N = \cup_{P \in \mathcal{P}^N} \left\{ Q \in \mathcal{M} : H(Q,P) \leq \frac{r_0}{\sqrt{n}}, E_Q \left[ \sup_{\theta \in \Theta} |g(x,\theta)|^2 \right] < M \right\},$$

$$P_{t/\sqrt{n},\eta} = \text{parametric submodel with score } \eta,$$

$$Q_n^{LA}(t,\eta) = \left\{ Q \in \mathcal{M} : H(Q,P_{t/\sqrt{n},\eta}) \leq \frac{r_1}{\sqrt{n}}, E_Q \left[ \sup_{\theta \in \Theta} |g(x,\theta)|^2 \right] < M \right\},$$

$$\sigma^2_{\eta} = \left( E_{P_{0,\eta}} [\hat{g}(x,0)]' E_{P_{0,\eta}} [\hat{g}(x,0)\hat{g}(x,0)']^{-1} E_{P_{0,\eta}} [\hat{g}(x,0)] \right)^{-1}.$$

Note that

$$\sqrt{n} \left( \bar{T}(P_n) - \bar{T}(Q_n) \right) \overset{d}{\to} N \left( 0, \sigma^2_{\eta} \right), \quad \bar{\sigma} \overset{p}{\to} \sigma_{\eta} \quad \text{for all sequence } Q_n \in Q_n^{LA}(0,\eta)$$

$$\sqrt{n} \left( \bar{T}(P_n) - \bar{T}(Q_n) \right) \overset{d}{\to} N \left( t, \sigma^2_{\eta} \right), \quad \bar{\sigma} \overset{p}{\to} \sigma_{\eta} \quad \text{for all sequence } Q_n \in Q_n^{LA}(t,\eta)$$

(Thus, $\sigma^2_{\eta}$ is like $\tilde{B}^{-1}$ in Choi-Hall-Schick. On the other hand, van der Vaar is more explicit about the dependence on $\eta$. Also note that the asymptotic variance for $\theta$ using MLE by parametric submodel is $V_{\eta} = \left( \int \eta^2 dP_{0,\eta} \right)^{-1}$ which is always smaller than $\sigma^2_{\eta}$. 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 $z_\alpha + 2r_0 + 2r_1 - \sigma^{-1}_\eta t < 0$ and each $\eta \in L_2(P_{0,\eta})$ satisfying $\int \eta dP_{0,\eta} = 0$.

(i): If a test $\psi_n$ satisfies

$$\limsup_{n \to \infty} \sup_{Q \in Q_n^N} E_Q [\psi_n] \leq \alpha,$$
Note that Theorem 6. demonstrates. The Hellinger-based Wald test is optimal in terms of its robust power property as well, as the following result (2.5)

\[ H_{2.2.2.} \]

\[ \sigma \text{ satisfying} \]

\[ (2.6) \]

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

and

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

2.2.2. Case II: Two-sided \( H_1 \) & Scalar \( \theta \). Consider the one-sided alternative hypothesis

\[ (2.5) \]

\[ H_0 : \theta_0 = 0 \quad \text{against} \quad H_1 : \theta_0 \neq 0, \]

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

\[ \mathcal{P}^N = \mathcal{P}_0, \]

\[ \mathcal{Q}_n^N = \bigcup_{P \in \mathcal{P}^N} \left\{ Q \in \mathcal{M} : H(Q, P) \leq \frac{r_0}{\sqrt{n}}, \ E_Q \left[ \sup_{\theta \in \Theta} |g(x, \theta)|^2 \right] < M \right\}, \]

\[ P_{t/\sqrt{\pi}, \eta} = \text{parametric submodel with score } \eta, \]

\[ \mathcal{Q}_n^{L,A}(t, \eta) = \left\{ Q \in \mathcal{M} : H(Q, P_{t/\sqrt{\pi}, \eta}) \leq \frac{t_1}{\sqrt{n}}, \ E_Q \left[ \sup_{\theta \in \Theta} |g(x, \theta)|^2 \right] < M \right\}, \]

\[ \sigma_\eta^2 = \left( E_{P_{0,n}} [\hat{g}(x, 0)]' E_{P_{0,n}} [g(x, 0) g(x, 0)]^{-1} E_{P_{0,n}} [\hat{g}(x, 0)] \right)^{-1}. \]

Note that

\[ \sqrt{n} \left( \hat{T}(P_n) - T(Q_n) \right) \overset{d}{\to} N \left( 0, \sigma_\eta^2 \right), \quad \hat{\sigma} \overset{p}{\to} \sigma_\eta \quad \text{for all sequence } Q_n \in \mathcal{Q}_n^{L,A}(0, \eta) \]

\[ \sqrt{n} \left( \hat{T}(P_n) - T(Q_n) \right) \overset{d}{\to} N \left( t, \sigma_\eta^2 \right), \quad \hat{\sigma} \overset{p}{\to} \sigma_\eta \quad \text{for all sequence } Q_n \in \mathcal{Q}_n^{L,A}(t, \eta) \]

We say that a test \( \psi_n \) is asymptotically unbiased at \( \eta \) if

\[ \lim sup_{n \to \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_n] \leq \lim inf_{n \to \infty} \inf_{Q \in \mathcal{Q}_n^{L,A}(t, \eta)} E_Q [\psi_n]. \]

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 \( \eta \in L_2(P_{0,n}) \) satisfying \( \int \eta dP_{0,n} = 0. \)

1. **(i):** If an asymptotically unbiased test \( \psi_n \) at \( \eta \) satisfies

\[ (2.6) \]

\[ \lim sup_{n \to \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_n] \leq \frac{\alpha}{2} + \Phi \left( -z_{\alpha/2} - 4r_0 \right), \]
then
\[ \liminf_{n \to \infty} \sup_{Q \in \mathcal{Q}_n^L(\tau, n)} E_Q [1 - \psi_n] \geq \Phi \left( z_{\alpha/2} + 2r_0 + \sigma^{-1}_n |t| + 2r_1 \right) - \Phi \left( -z_{\alpha/2} - 2r_0 + \sigma^{-1}_n |t| + 2r_1 \right). \]

(ii): The Hellinger-based Wald test \( \psi_{H,n} = \mathbb{I} \left\{ \left| \frac{\sqrt{n} \tau(\theta)}{\sigma} \right| \geq z_{\alpha/2} + 2r_0 \right\} \) is asymptotically unbiased and satisfies
\[
\limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_n^N} E_Q [\psi_{H,n}] \leq \frac{\alpha}{2} + \Phi \left( -z_{\alpha/2} - 4r_0 \right),
\]
and
\[
\lim_{n \to \infty} \sup_{Q \in \mathcal{Q}_n^L(\tau, n)} E_Q [1 - \psi_{H,n}] = \Phi \left( z_{\alpha/2} + 2r_0 + \sigma^{-1}_n |t| + 2r_1 \right) - \Phi \left( -z_{\alpha/2} - 2r_0 + \sigma^{-1}_n |t| + 2r_1 \right).
\]

3. Conclusion

To be written.
APPENDIX A. PROOF OF THEOREMS

Notation. Let

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

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

\[
\bar{V} = \left( E_{\bar{P}} \left[ g(x, \bar{\theta}) \right] E_{\bar{P}} \left[ g(x, \bar{\theta}) g(x, \bar{\theta}) \right]^{-1} E_{\bar{P}} \left[ g(x, \bar{\theta}) \right] \right)^{-1},
\]

\[
\bar{t} = \frac{2 (r_0 - \epsilon)}{\sqrt{\tau' \bar{\tau} V_{\bar{\theta}}' \bar{\theta}}} \bar{V}_{\bar{\theta}}.
\]

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

\[
\frac{dP_{\theta + \bar{t}/\sqrt{n}}}{dP} = \frac{1 + \bar{\zeta}_n g_n (x, \theta + \bar{t}/\sqrt{n})}{\int (1 + \bar{\zeta}_n g_n (x, \theta + \bar{t}/\sqrt{n})) d\bar{P}},
\]

where \( \bar{\zeta}_n = -E_{\bar{P}} \left[ g(x, \theta + \bar{t}/\sqrt{n}) g_n (x, \theta + \bar{t}/\sqrt{n}) \right]^{-1} E_{\bar{P}} \left[ g(x, \theta + \bar{t}/\sqrt{n}) \right] \). By a similar argument to the proof of Theorem 3.1 of Kitamura, Otsu, and Evdokimov (2009), we can show that

\[
\lim_{n \to \infty} nH \left( P_{\theta + \bar{t}/\sqrt{n}}, \bar{P} \right)^2 = \frac{1}{4} \bar{t} \bar{V}^{-1} \bar{t} = (r_0 - \epsilon)^2,
\]

which implies

\[
\frac{P_{\theta + \bar{t}/\sqrt{n}}}{\sqrt{n}} \in Q_n,
\]

for all \( n \) large enough.

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

\[
S_n = \left( \hat{\sigma}_a^{-1} \sqrt{n} \tau \circ T_a (P_n) \right)^2 + o_P(1)
\]

\[
= \left( \sigma_a^{-1} \sqrt{n} \tau \circ T_a (P_n) \right)^2 + o_P(1)
\]

\[
= \left( \sigma_a^{-1} \hat{\tau} (\tau) \sqrt{n} (T_a (P_n) - \bar{\theta}) \right)^2 + o_n,
\]

under \( \{ P_{\theta + \bar{t}/\sqrt{n}} \}_{n \in \mathbb{N}} \), where \( \sigma_a^2 = \hat{\tau} (\tau) \hat{V}_{\hat{\tau} (\bar{\theta})} \cdot g_n = o_P \left( n (T_a (P_n) - \bar{\theta}) \right)^4 \), the second equality follows from \( \hat{\sigma}_a \overset{p}{\to} \sigma_a \) under \( \{ P_{\theta + \bar{t}/\sqrt{n}} \}_{n \in \mathbb{N}} \) (by Definition 2), and the third equality follows from a second-order expansion.
around $T_a(P_a) = \tilde{\theta}$ (note: $\tau(\tilde{\theta}) = 0$ and the second-order derivative of $\tau$ is bounded on $\Theta$). Thus, we have

$$\liminf_{n \to \infty} \sup_{Q \in \mathcal{Q}_n} E_Q [\psi_n] \geq \liminf_{n \to \infty} E_{P_{\tilde{\theta} + i/\sqrt{n}}} \left[ \mathbb{I} \left\{ S_n \geq \chi_{1,0}^2 \right\} \right] = \liminf_{n \to \infty} \int \mathbb{I} \left\{ \left( \sigma_a^{-1} \hat{\tau} \right)^{\frac{1}{2}} \sqrt{n} \left( T_a(P_a) - \tilde{\theta} \right) \right\}^2 \geq \chi_{1,0}^2 dP_{\tilde{\theta} + i/\sqrt{n}}$$

$$= \liminf_{n \to \infty} \int \mathbb{I} \left\{ \left( \sigma_a^{-1} \hat{\tau} \right)^{\frac{1}{2}} \left( \sqrt{n} \left( T_a(P_a) - T_a(P_{\tilde{\theta} + i/\sqrt{n}}) \right) + \sqrt{n} \left( T_a(P_{\tilde{\theta} + i/\sqrt{n}}) - \tilde{\theta} \right) \right) \right\}^2 \geq \chi_{1,0}^2 dP_{\tilde{\theta} + i/\sqrt{n}}$$

$$= E_Z \left[ \mathbb{I} \left\{ \left( Z + \sigma_a^{-1} \hat{\tau} \hat{\epsilon} \right)^2 \geq \chi_{1,0}^2 \right\} \right],$$

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

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

$$\limsup_{n \to \infty} E_{Q_n} [\psi_{H,n}] \leq E_Z \left[ \mathbb{I} \left\{ (Z + 2r_0)^2 \geq \chi_{1,0}^2 \right\} \right],$$

where $Z \sim N(0,1)$. Pick any sequence $\{Q_n\}_{n \in \mathbb{N}}$ such that $Q_n \in \mathcal{Q}_n$ 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 $\left\{ \theta_n \right\}_{n \in \mathbb{N}}$ 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

$$\limsup_{n \to \infty} E_{Q_n} [\psi_{H,n}] = \limsup_{n \to \infty} \int \mathbb{I} \left\{ \left( \hat{\sigma}^{-1} \sqrt{n} \tau \circ T \left( \hat{P}_n \right) \right)^2 \geq \chi_{1,0}^2 \right\} dQ_n \leq \limsup_{n \to \infty} \int_{(x_1, \ldots, x_n) \in \mathcal{X}_n^m} dQ_n \leq \limsup_{n \to \infty} \int_{(x_1, \ldots, x_n) \in \mathcal{X}_n^m} \mathbb{I} \left\{ \left( \hat{\sigma}^{-1} \sqrt{n} \tau \circ T \left( \hat{P}_n \right) \right)^2 \geq \chi_{1,0}^2 \right\} dQ_n \leq A_1 + A_2,$$

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

$$A_1 \leq \limsup_{n \to \infty} \sum_{i=1}^{n} \int_{X_n} dQ_n \leq \limsup_{n \to \infty} nm^{-n} E_{Q_n} \left[ \sup_{\theta \in \Theta} |g(x, \theta)|^n \right] = 0,$$

where the first inequality follows from a set inclusion relation, the second inequality follows from the Markov inequality, and the equality follows from $nm^{-n} \to 0$ (Assumption 2.1) and $E_{Q_n} \left[ \sup_{\theta \in \Theta} |g(x, \theta)|^n \right] < \infty$ (by
\( Q_n \in Q_n^N \). For \( A_2 \), we have

\[
A_2 \leq \limsup_{n \to \infty} \left\{ \left( \tilde{\sigma}^{-1} \sqrt{\theta} \circ T \left( \hat{P}_n \right) \right)^2 \geq \chi^2_{1, \alpha} \right\} dQ^n_n
\]

\[
= \limsup_{n \to \infty} \int \left\{ \left( \tilde{\sigma}^{-1} \tilde{\tau} (\theta_n) \left\{ \sqrt{n} \left( T \left( \hat{P}_n \right) - T (Q_n) \right) + \sqrt{n} \left( T (Q_n) - \theta_n \right) + \theta_n \right\} \right)^2 \geq \chi^2_{1, \alpha} \right\} dQ^n_n
\]

\[
\leq E_Z \left[ \mathbb{I} \left\{ (Z + 2\rho_0)^2 \geq \chi^2_{1, \alpha} \right\} \right],
\]

where \( \rho_n = o_p(1) \) under \( Q_n \), the inequality follows from the set inclusion relation (\( \mathcal{A}_n^\alpha \) is a subset of support of \( Q^{\infty}_n \)), 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.,

\[
\limsup_{n \to \infty} \sup_{Q \in Q_n^N} \mathbb{E}_Q [\psi_n] \leq \alpha,
\]

\[
\liminf_{n \to \infty} \sup_{Q \in Q_n^N (P_i / \sqrt{\nu_n})} \mathbb{E}_Q [1 - \psi_n] \leq \Phi \left( z_\alpha + 2r_0 + 2r_1 - \sigma^{-1}_n t \right) - \epsilon,
\]

for some \( \epsilon > 0 \).

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

\[
\frac{dP_{i / \sqrt{\nu_n}}^*}{dP_{0, \eta}} = \frac{1 + \zeta_{t, n} g_n \left( x, \frac{t}{\sqrt{n}} \right)}{\int \left( 1 + \zeta_{t, n} g_n \left( x, \frac{t}{\sqrt{n}} \right) \right) dP_{0, \eta}},
\]

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

\[
P_{t / \sqrt{\nu_n}}^* \in P_{t / \sqrt{\eta}},
\]

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

\[
E_{P_i / \sqrt{\nu_n}} \left[ g \left( x, \frac{t}{\sqrt{n}} \right) \right] = \int g \left( x, \frac{t}{\sqrt{n}} \right) dP_{0, \eta} + \int g \left( x, \frac{t}{\sqrt{n}} \right) g_n \left( x, \frac{t}{\sqrt{n}} \right) dP_{0, \eta} \zeta_{t, n},
\]

\[
= \int \left( 1 + \zeta_{t, n} g_n \left( x, \frac{t}{\sqrt{n}} \right) \right) dP_{0, \eta} \zeta_{t, n}.
\]

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

\[
t_0 = 2\sigma_n (r_0 - \delta).
\]

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

\[
nH \left( P_{t_0 / \sqrt{\nu_n}}, P_{0, \eta} \right)^2 \to \frac{1}{4} \sigma_n^{-2} t_0^2 = (r_0 - \delta)^2,
\]

which implies

\[
P_{t_0 / \sqrt{\nu_n}} \in Q_n^N,
\]
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

$$
\sum_{i=1}^{n} \log \frac{dP_{t_{0}/\sqrt{n},\eta}}{dP_{0,\eta}}(x_{i}) \\
= \sum_{i=1}^{n} \log \left( 1 + \zeta^\prime_{t_{0},n} g_{n} \left( x_{i}, \frac{t_{0}}{\sqrt{n}} \right) \right) - n \log \int \left( 1 + \zeta^\prime_{t_{0},n} g_{n} \left( x, \frac{t_{0}}{\sqrt{n}} \right) \right) dP_{0,\eta}
$$

$$
= \zeta^\prime_{t_{0},n} \sum_{i=1}^{n} g_{n} \left( x_{i}, \frac{t_{0}}{\sqrt{n}} \right) - \frac{1}{2} \zeta^\prime_{t_{0},n} \sum_{i=1}^{n} g_{n} \left( x_{i}, \frac{t_{0}}{\sqrt{n}} \right) g_{n} \left( x_{i}, \frac{t_{0}}{\sqrt{n}} \right) \zeta_{t_{0},n}
$$

$$
- n \zeta^\prime_{t_{0},n} \int g_{n} \left( x, \frac{t_{0}}{\sqrt{n}} \right) dP_{0,\eta}
$$

$$
+ \frac{n}{2} \zeta^\prime_{t_{0},n} \int g_{n} \left( x, \frac{t_{0}}{\sqrt{n}} \right) dP_{0,\eta} \int g_{n} \left( x, \frac{t_{0}}{\sqrt{n}} \right) dP_{0,\eta} \zeta_{t_{0},n} + o_{p}(1)
$$

(A.8) $$
\Rightarrow t_{0} \sigma_{\eta}^{-1} Z - \frac{1}{2} \sigma_{\eta}^{-2} t_{0}^{2},
$$

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

$$
\alpha \geq \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_{n}} E_{Q}[\psi_{n}]
$$

$$
\geq \liminf_{n \to \infty} E_{P_{t_{0}/\sqrt{n},\eta}}[\psi_{n}]
$$

$$
= \liminf_{n \to \infty} E_{P_{0,\eta}} \left[ \psi_{n} \prod_{i=1}^{n} \frac{dP_{t_{0}/\sqrt{n},\eta}}{dP_{0,\eta}}(x_{i}) \right]
$$

$$
= E_{\psi,Z} \left[ \psi \exp \left( t_{0} \sigma_{\eta}^{-1} Z - \frac{1}{2} \sigma_{\eta}^{-2} t_{0}^{2} \right) \right]
$$

$$
= \frac{1}{\sqrt{2\pi}} \int_{z} \psi_{0}(z) \exp \left( -\frac{1}{2} \left( z - \sigma_{\eta}^{-1} t_{0} \right)^{2} \right) dz
$$

$$
= E_{Z} \left[ \psi_{0} \left( Z + \sigma_{\eta}^{-1} t_{0} \right) \right] = E_{Z} \left[ \psi_{0} \left( Z + 2 (r_{0} - \delta) \right) \right],
$$

(A.9)

where $\psi$ is a random variable such that $\psi_{n} \Rightarrow \psi$ under $P_{0}$, $\psi_{0}(z) = E[\psi|Z = z]$, 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),

$$
nH \left( P_{(t_{1}+t)/\sqrt{n},\eta}^{*} P_{t/\sqrt{n},\eta} \right)^{2} \to \frac{1}{4} t_{1}^{\prime} \sigma_{\eta}^{-1} t_{1} = (r_{1} - \delta)^{2},
$$

(A.10)
which implies

\begin{equation}
  P_{(t_1+t)\sqrt{n\eta}} \in Q_n^{LA}(t, \eta),
\end{equation}

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

\begin{equation}
  \sum_{i=1}^{n} \log \frac{dP_{(t_i+t)\sqrt{n\eta}}}{dP_{0,\eta}}(x_i) \Rightarrow (t_1 + t) \sigma^{-1}_\eta Z - \frac{1}{2} \sigma^{-2}_\eta (t_1 + t).
\end{equation}

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

\begin{equation}
  \Phi \left( z_0 + 2r_0 + 2r_1 - \sigma^{-1}_\eta t \right) - \epsilon \\
  \geq \liminf_{n \to \infty} \sup_{Q \in Q_n^{LA}(t, \eta)} E_Q \left[ 1 - \psi_n \right] \\
  \geq \liminf_{n \to \infty} E_{P_{(t_1+t)\sqrt{n\eta}}} \left[ 1 - \psi_n \right] \\
  = \liminf_{n \to \infty} E_{P_{0,\eta}} \left[ (1 - \psi_n) \prod_{i=1}^{n} \frac{dP_{(t_i+t)\sqrt{n\eta}}}{dP_{0,\eta}}(x_i) \right] \\
  = E_{\psi,\pi} \left[ (1 - \psi) \exp \left( (t_1 + t) \sigma^{-1}_\eta Z - \frac{1}{2} \sigma^{-2}_\eta (t_1 + t)^2 \right) \right] \\
  = \frac{1}{\sqrt{2\pi}} \int_{z} (1 - \psi_0(z)) \exp \left( -\frac{1}{2} (z - \sigma^{-1}_\eta (t_1 + t))^2 \right) dz \\
  \geq 1 - E_{Z} \left[ \psi_0 \left( Z + \sigma^{-1}_\eta (t_1 + t) \right) \right] = 1 - E_{Z} \left[ \psi_0 \left( Z - 2(r_1 - \delta) + \sigma^{-1}_\eta t \right) \right],
\end{equation}

where $\psi$ is a random variable such that $\psi_n \Rightarrow \psi$ under $P_0$, $\psi_0(z) = E[\psi|Z = z]$, 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

\begin{equation}
  E_{Z} \left[ \psi_0 \left( Z + 2(r_0 + \delta) \right) \right] \leq \alpha.
\end{equation}

The requirement (A.13) is written as

\begin{equation}
  E_{Z} \left[ \psi_0 \left( Z - 2(r_1 - \delta) + \sigma^{-1}_\eta t \right) \right] \geq \Phi \left( \sigma^{-1}_\eta t - z_0 - 2r_0 - 2r_1 \right) + \epsilon.
\end{equation}

However, $\Phi \left( \sigma^{-1}_\eta t - z_0 - 2r_0 - 2r_1 \right)$ is the power of the Neyman-Pearson test for testing $H_0 : \mu = 2r_0$ against $H_1 : \mu = \sigma^{-1/2}_\eta t - 2r_1$ under $N(\mu, 1)$. By taking $\delta$ 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 $Q_n \in Q_n^N$ 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),

\[
\limsup_{n \to \infty} E_{Q_n} \left[ \psi_{H,n} \right] = \limsup_{n \to \infty} E_{Q_n} \left[ I \left\{ \hat{\sigma}^{-1} \sqrt{nT} \left( \hat{P}_n \right) \geq z_\alpha + 2r_0 \right\} \right] + o(1)
\]

\[
\leq \limsup_{n \to \infty} E_{Q_n} \left[ I \left\{ \hat{\sigma}^{-1} \sqrt{nT} \left( \hat{P}_n \right) \geq z_\alpha + 2r_0 \right\} + \hat{\sigma}^{-1} \sqrt{n} \sup_{Q \in Q_n^N} \left| T \left( Q \right) \right| \right] + o(1)
\]

\[
= E_Z \left[ I \left\{ Z \geq z_\alpha \right\} \right] = \alpha.
\]

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

\[
\limsup_{n \to \infty} \sup_{Q \in Q_n^{L^A(t, \eta)}} E_{Q_n} \left[ 1 - \psi_{H,n} \right] \leq \Phi \left( z_\alpha + 2r_0 + 2r_1 - \sigma_{\eta}^{-1} t \right).
\]

Observe that the Type II error satisfies

\[
\limsup_{n \to \infty} \sup_{Q \in Q_n^{L^A(t, \eta)}} E_{Q_n} \left[ 1 - \psi_{H,n} \right] \leq \Phi \left( z_\alpha + 2r_0 \right) + o(1)
\]

\[
= \limsup_{n \to \infty} \sup_{Q \in Q_n^{L^A(t, \eta)}} E_{Q_n} \left[ I \left\{ \hat{\sigma}^{-1} \sqrt{nT} \left( \hat{P}_n \right) \leq z_\alpha + 2r_0 \right\} \right] + o(1)
\]

\[
= E_{Q_n} \left[ I \left\{ Z + \sigma_{\eta}^{-1} t - \sigma_{\eta}^{-1} \limsup_{n \to \infty} \sup_{Q \in Q_n^{L^A(t, \eta)}} \sqrt{n} \left| T \left( Q \right) \right| \leq z_\alpha + 2r_0 \right\} \right]
\]

\[
= \Phi \left( z_\alpha + 2r_0 + 2r_1 - \sigma_{\eta}^{-1} t \right),
\]

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.,

\[
(A.16) \quad \limsup_{n \to \infty} \sup_{Q \in Q_n^N} E_{Q_n} \left[ \psi_n \right] \leq \frac{\alpha}{2} + \Phi \left( -z_{\alpha/2} - 4r_0 \right),
\]

\[
(A.17) \quad \liminf_{n \to \infty} \sup_{Q \in Q_n^{L^A(t, \eta)}} E_{Q_n} \left[ 1 - \psi_n \right] \leq \Phi \left( z_{\alpha/2} + 2r_0 + \sigma_{\eta}^{-1} |t| + 2r_1 \right) - \Phi \left( -z_{\alpha/2} - 2r_0 + \sigma_{\eta}^{-1} |t| + 2r_1 \right) - \epsilon,
\]

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

$$
\frac{dP_{t/\sqrt{n},\eta}^*}{dP_{0,\eta}} = \frac{1 + \zeta_{t,\eta} g_n \left( x, \frac{t}{\sqrt{n}} \right)}{\int \left( 1 + \zeta_{t,\eta} g_n \left( x, \frac{t}{\sqrt{n}} \right) \right) dP_{0,\eta}},
$$

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

$$
P_{t/\sqrt{n},\eta}^* \in \mathcal{P}_{t/\sqrt{n}},
$$

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

$$
E_{P_{t/\sqrt{n},\eta}^*} \left[ g \left( x, \frac{t}{\sqrt{n}} \right) \right] = \frac{\int g \left( x, \frac{t}{\sqrt{n}} \right) dP_{0,\eta} + \int g \left( x, \frac{t}{\sqrt{n}} \right) g_n \left( x, \frac{t}{\sqrt{n}} \right) dP_{0,\eta} \zeta_{t,\eta}}{\int \left( 1 + \zeta_{t,\eta} g_n \left( x, \frac{t}{\sqrt{n}} \right) \right) dP_{0,\eta}} = 0.
$$

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),

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

which implies

$$
P_{t_0/\sqrt{n},\eta}^* \in \mathcal{Q}_n.
$$

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

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

under $x \sim P_{0,\eta}$, where $Z \sim N(0, 1)$, the second equality follows from a second-order expansion around $\zeta_{t_0,\eta} = 0$ 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

\[
\frac{\alpha}{2} + \Phi(-z_{\alpha/2} - 4r_0) \\
\geq \lim \sup_{n \to \infty} \sup_{Q \in Q_n^N} E_Q[\psi_n] \\
\geq \lim \inf_{n \to \infty} E_{P^*_{t_0/\pi,\eta}}[\psi_n] \\
= \lim \inf_{n \to \infty} E_{P_{0,\eta}} \left[ \psi_n \prod_{i=1}^n dP^*_{t_0/\pi,\eta}(x_i) \right] \\
= E_{\psi,Z} \left[ \psi \exp \left( t_0 \sigma^{-1}_\eta Z - \frac{1}{2} \sigma^{-2}_\eta t_0 \right) \right] \\
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_0(z) \exp \left( -\frac{1}{2} (z - \sigma^{-1}_\eta t_0)^2 \right) dz \\
(A.21)
= E_Z \left[ \psi_0(Z + \sigma^{-1}_\eta t_0) \right] = E_Z \left[ \psi_0(Z + 2(r_0 - \delta)) \right],
\]

where $\psi$ is a random variable such that $\psi_n \Rightarrow \psi$ under $P_0$, $\psi_0(z) = E[\psi|Z = z]$, 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),

\[
(A.22) \quad nH \left( P^*_{(t_1+t)/\sqrt{\pi,\eta}} P_{t/\sqrt{\pi,\eta}} \right)^2 \to \frac{1}{4} t_1' \sigma^{-1}_\eta t_1 = (r_1 - \delta)^2,
\]

which implies

\[
(A.23) \quad P_{(t_1+t)/\sqrt{\pi,\eta}} \in Q_{n,LA}^L(t,\eta),
\]

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

\[
(A.24) \quad \Rightarrow (t_1 + t) \sigma^{-1}_\eta Z - \frac{1}{2} \sigma^{-2}_\eta (t_1 + t).
\]
Therefore, the power requirement in (A.17) implies

\[
\Phi (z_{\alpha/2} + 2r_0 + \sigma_n^{-1} |t| + 2r_1) - \Phi (-z_{\alpha/2} - 2r_0 + \sigma_n^{-1} |t| + 2r_1) - \epsilon
\]

\[
\geq \liminf_{n \to \infty} \sup_{Q \in \mathcal{Q}^{(1)}(t,\eta)} E_Q [1 - \psi_n]
\]

\[
\geq \liminf_{n \to \infty} E_{P_{(t,1+\epsilon)/\sqrt{\pi}n}} [1 - \psi_n]
\]

\[
= \liminf_{n \to \infty} E_{P_{0,\eta}} \left[ (1 - \psi_n) \prod_{i=1}^{n} \frac{dP^*_{(t_1,t)/\sqrt{\pi}n}}{dP_{0,\eta}} (x_i) \right]
\]

\[
= E_{\psi,Z} \left[ (1 - \psi) \exp \left( (t_1 + t) \sigma_n^{-1} Z - \frac{1}{2} \sigma_n^{-2} (t_1 + t)^2 \right) \right]
\]

\[
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (1 - \psi(z)) \exp \left( -\frac{1}{2} (z - \sigma_n^{-1} (t_1 + t))^2 \right) dz
\]

\[
(A.25) \quad 1 - E_Z \left[ \psi_0 (Z + \sigma_n^{-1} (t_1 + t)) \right] = 1 - E_Z \left[ \psi_0 (Z - 2 (r_1 - \delta) + \sigma_n^{-1} t) \right],
\]

where \( \psi \) is a random variable such that \( \psi_n \Rightarrow \psi \) under \( P_\theta \), \( \psi_0 (z) = E [\psi | Z = z] \), 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

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

However, \( \Phi (-z_{\alpha/2} - 2r_0 - (\sigma_n^{-1} |t| + 2r_1)) + \Phi (-z_{\alpha/2} - 2r_0 + (\sigma_n^{-1} |t| + 2r_1)) \) is the power of the uniformly most powerful unbiased level \( \frac{\alpha}{2} \) test for \( H_0 : |\mu| \leq 2r_0 \) against \( H_1 : |\mu| > 2r_0 \) under \( Z \sim N (\mu, 1) \) evaluated at \( \mu = \sigma_n^{-1} |t| + 2r_1 \). Therefore, by taking \( \delta \) 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 \( H (Q_n, P_n) \leq \frac{\alpha}{\sqrt{n}} \) 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),
\[
\limsup_{n \to \infty} E_{Q_n} [\psi_{H,n}]
= \limsup_{n \to \infty} E_{Q_n} \left[ I \left\{ \left| \hat{\sigma}^{-1} \sqrt{n} \tilde{T} \left( \tilde{P}_n \right) \right| \geq z_{\alpha/2} + 2r_0 \right\} \right]
\leq \limsup_{n \to \infty} E_{Q_n} \left[ I \left\{ \left| \tilde{T} \left( \tilde{P}_n \right) - \tilde{T} \left( Q_n \right) \right| \geq z_{\alpha/2} + 2r_0 \right\} \right]
+ \limsup_{n \to \infty} E_{Q_n} \left[ I \left\{ \left| \tilde{T} \left( \tilde{P}_n \right) - \tilde{T} \left( Q_n \right) \right| \leq -z_{\alpha/2} - 2r_0 \right\} \right] + o(1)
\leq E_{Z} \left[ I \{ Z + 2r_0 \geq z_{\alpha/2} + 2r_0 \} \right] + E_{Z} \left[ I \{ Z + 2r_0 \leq -z_{\alpha/2} - 2r_0 \} \right]
\tag{A.28}
= \frac{\alpha}{2} + \Phi \left( -z_{\alpha/2} - 4r_0 \right).
\]

**Proof of the asymptotic unbiasedness and (2.8).** Observe that the Type II error satisfies
\[
\limsup_{n \to \infty} \sup_{Q \in Q_{n}^{L,t}(n)} E_{Q} [1 - \psi_{H,n}]
\leq \limsup_{n \to \infty} \sup_{Q \in Q_{n}^{L,t}(n)} E_{Q} \left[ I \left\{ -z_{\alpha/2} - 2r_0 \leq \hat{\sigma}^{-1} \sqrt{n} \tilde{T} \left( \tilde{P}_n \right) \leq z_{\alpha/2} + 2r_0 \right\} \right] + o(1)
= \limsup_{n \to \infty} \sup_{Q \in Q_{n}^{L,t}(n)} E_{Q} \left[ I \left\{ -z_{\alpha/2} - 2r_0 \leq -z_{\alpha/2} - 2r_0 \leq z_{\alpha/2} + 2r_0 \right\} \right]
\leq E_{Z} \left[ I \{ Z + 2r_0 \geq z_{\alpha/2} + 2r_0 \} \right] + E_{Z} \left[ I \{ Z + 2r_0 \leq -z_{\alpha/2} - 2r_0 \} \right]
\tag{A.29}
= \Phi \left( z_{\alpha/2} + 2r_0 + \sigma_{\eta}^{-1} |t| + 2r_1 \right) - \Phi \left( -z_{\alpha/2} - 2r_0 + \sigma_{\eta}^{-1} |t| + 2r_1 \right),
\]
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
\[
\limsup_{n \to \infty} \sup_{Q \in Q_{n}^{L,t}(n)} E_{Q} [\psi_{H,n}]
\leq \frac{\alpha}{2} + \Phi \left( -z_{\alpha/2} - 4r_0 \right)
\leq \Phi \left( -z_{\alpha/2} - 2r_0 - (\sigma_{\eta}^{-1} |t| + 2r_1) \right) + \Phi \left( -z_{\alpha/2} - 2r_0 + (\sigma_{\eta}^{-1} |t| + 2r_1) \right)
\leq \limsup_{n \to \infty} \inf_{Q \in Q_{n}^{L,t}(n)} E_{Q} [\psi_{H,n}],
\]
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.

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

$\mathcal{P}_\theta = \left\{ P \in \mathcal{M} : \int g_n(x, \theta) dP = 0 \right\}$,

$\overline{T}(Q) = \arg \min_{\theta \in \Theta} \min_{P \in \mathcal{P}_\theta, P \ll Q} H(P, Q)$, \quad $\overline{T}_Q = \overline{T}(Q_n)$, \quad $\overline{T}_P = \overline{T}(P_n)$,

$G_n = E_{P_n} [g(x, \theta_n)]$, \quad $\Omega_n = E_{P_n} \left[ g(x, \theta_n) g(x, \theta_n)' \right]$, \quad $V_n = (G_n' \Omega_n^{-1} G_n)^{-1}$,

$\Lambda_n = G_n' \Omega_n^{-1} g_n(x, \theta_n)$, \quad $\psi_{n,Q_n} = -2 \left( \int \Lambda_n' dQ_n \right)^{-1} \int \Lambda_n \left\{ dQ_n^{1/2} - d\bar{P}_n^{1/2} \right\} dQ_n^{1/2}$,

$\bar{P}_{\theta,Q} = \arg \min_{P \in \mathcal{P}_\theta, P \ll Q} H(P, Q)$, \quad $R_n(Q, \theta, \gamma) = -\int \frac{1}{1 + \gamma g_n(x, \theta)} dQ$,

$\theta_{t,n} = \theta_n + t/\sqrt{n}$. 

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. \( \hat{T}(Q) \) exists and is upper semi-continuous at each \( Q \in \mathcal{M} \) under the Hellinger distance,

2. \( |\hat{T}_{Q_n} - \theta_n| = o(1), \)

3. \( \sqrt{n} (\hat{T}_{Q_n} - \theta_n) + \sqrt{n} V_n \int \Lambda_n dQ_n = o(1), \)

4. \( \|d\hat{P}_{\theta_n, Q_n}^{1/2} - d\hat{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} (\hat{T}_{Q_n} - \theta_n)^\prime \Lambda_n dQ_n^{1/2}\| = o \left( |\hat{T}_{Q_n} - \theta_n| \right) + o \left( n^{-1/2} \right), \)

5. \( \|d\hat{P}_{\theta_n, \psi_n, Q_n}^{1/2} - d\hat{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi_n^\prime Q_n \Lambda_n dQ_n^{1/2}\| = o \left( |\psi_n, Q_n| \right) + o \left( n^{-1/2} \right), \)

6. \( |E_{P_n} [g_n (x, \theta_n)]| = o \left( n^{-1/2} \right), \) \( |E_{P_n} [g_n (x, \theta_{t,n})]| = O \left( n^{-1/2} \right), \) \( |E_{P_n} [g_n (x, \theta_{t,n}) g_n (x, \theta_n)^\prime] - \Omega_n| = o(1), \) and \( |E_{P_n} [g_n (x, \theta_{t,n}) - G_n]| = o(1), \)

7. \( \gamma_n (\theta_{t,n}, P_n) = \arg \max_{\gamma \in \mathbb{R}^m} - \int \frac{1}{(1 + \gamma g_n (x, \theta_{t,n}))} dP_n \text{ exists for all } n \text{ large enough, } |\gamma_n (\theta_{t,n}, P_n)| = O \left( n^{-1/2} \right), \) and \( \sup_{x \in X} |\gamma_n (\theta_{t,n}, P_n)^\prime g_n (x, \theta_{t,n})| = o(1), \)

8. \( |E_{Q_n} [g_n (x, \theta_n)]| = O \left( n^{-1/2} \right), \) \( |E_{Q_n} [g_n (x, \theta_n) g_n (x, \theta_n)^\prime] - \Omega_n| = o(1), \)

9. \( \gamma_n (\theta_n, Q_n) = \arg \max_{\gamma \in \mathbb{R}^m} - \int \frac{1}{(1 + \gamma g_n (x, \theta_n))} dQ_n \text{ exists for all } n \text{ large enough, and } |\gamma_n (\theta_n, Q_n)| = O \left( n^{-1/2} \right), \) and \( \sup_{x \in X} |\gamma_n (\theta_n, Q_n)^\prime g_n (x, \theta_n)| = o(1), \)

10. \( |E_{Q_n} [g_n (x, \hat{T}_{Q_n})]| = O \left( n^{-1/2} \right), \) \( |E_{Q_n} [g_n (x, \hat{T}_{Q_n}) g_n (x, \hat{T}_{Q_n})^\prime] - \Omega_n| = o(1), \) and \( |E_{Q_n} [g_n (x, \hat{T}_{Q_n}) - G_n]| = o(1), \)

11. \( \gamma_n (\hat{T}_{Q_n}, Q_n) = \arg \max_{\gamma \in \mathbb{R}^m} - \int \frac{1}{(1 + \gamma g_n (x, \hat{T}_{Q_n}))} dQ_n \text{ exists for all } n \text{ large enough, } |\gamma_n (\hat{T}_{Q_n}, Q_n)| = O \left( n^{-1/2} \right), \) and \( \sup_{x \in X} |\gamma_n (\hat{T}_{Q_n}, Q_n)^\prime g_n (x, \hat{T}_{Q_n})| = o(1), \)

12. \( |\hat{T}_{P_n} - \theta_n| = o_p (1), \)

13. \( \sqrt{n} \left( \hat{T}_{P_n} - \theta_n \right) + \sqrt{n} V_n \int \Lambda_n dP_n = o_p (1) \) and \( V_n^{-1/2} \sqrt{n} \left( \hat{T}_{P_n} - \hat{T}_{Q_n} \right) \xrightarrow{d} N(0, I), \)

14. \( |E_{P_n} [g_n (x, \theta_n)]| = O_p \left( n^{-1/2} \right), \) \( |E_{P_n} [g_n (x, \theta_n) g_n (x, \theta_n)^\prime] - \Omega_n| = o_p (1), \)

15. \( \gamma_n (\theta_n, \hat{P}_n) = \arg \max_{\gamma \in \mathbb{R}^m} - \int \frac{1}{(1 + \gamma g_n (x, \theta_n))} dP_n \text{ exists a.s. for all } n \text{ large enough, } |\gamma_n (\theta_n, \hat{P}_n)| = O_p \left( n^{-1/2} \right), \) and \( \sup_{x \in X} |\gamma_n (\theta_n, \hat{P}_n)^\prime g_n (x, \theta_n)| = o_p (1), \)

16. \( |E_{P_n} [g_n (x, \hat{T}_{P_n})]| = O_p \left( n^{-1/2} \right), \) \( |E_{P_n} [g_n (x, \hat{T}_{P_n}) g_n (x, \hat{T}_{P_n})^\prime] - \Omega_n| = O_p \left( n^{-1/2} \right), \) and \( |E_{P_n} [g_n (x, \hat{T}_{P_n}) - G_n]| = o_p (1), \)

17. \( \gamma_n (\hat{T}_{P_n}, \hat{P}_n) = \arg \max_{\gamma \in \mathbb{R}^m} - \int \frac{1}{(1 + \gamma g_n (x, \hat{T}_{P_n}))} dP_n \text{ exists a.s. for all } n \text{ large enough, } |\gamma_n (\hat{T}_{P_n}, \hat{P}_n)| = O_p \left( n^{-1/2} \right), \) and \( \sup_{x \in X} |\gamma_n (\hat{T}_{P_n}, \hat{P}_n)^\prime g_n (x, \hat{T}_{P_n})| = o_p (1). \)
B.1. **Proof of Lemma 1.** See Kitamura, Otsu, and Evdokimov (2009).

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

(B.1)

\[
\sup_{\theta \in \Theta} |E_{Q_n} [g_n (x, \theta)] - E_{P_n} [g_n (x, \theta)]| \leq \sup_{\theta \in \Theta} |E_{Q_n} [g_n (x, \theta)] - E_{P_n} [g_n (x, \theta)]| + \sup_{\theta \in \Theta} |E_{P_n} [g (x, \theta) \mathbb{1} \{x \notin X_n\}]|.
\]

The first term of (B.1) satisfies

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

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 \( \frac{m_n}{\sqrt{n}} \to 0 \) (Assumption 2.1) and \( P_n \in \mathcal{P}_{\theta_n} \) (particularly \( E_{P_n} \left[ \sup_{\theta \in \Theta} |g (x, \theta)|^\eta \right] \leq M \)).

The second term of (B.1) satisfies

\[
\sup_{\theta \in \Theta} |E_{P_n} [g (x, \theta) \mathbb{1} \{x \notin X_n\}]| \\
\leq \left( \int \sup_{\theta \in \Theta} |g (x, \theta)|^\eta dP_n \right)^{1/\eta} \left( \int \mathbb{1} \{x \notin X_n\} dP_n \right)^{(\eta-1)/\eta} \\
\leq \left( E_{P_n} \left[ \sup_{\theta \in \Theta} |g (x, \theta)|^\eta \right] \right)^{1/\eta} \left( m_n^{-\eta} E_{P_n} \left[ \sup_{\theta \in \Theta} |g (x, \theta)|^\eta \right] \right)^{(\eta-1)/\eta} = o \left( n^{-1/2} \right),
\]

where the first inequality follows from the Hölder inequality, and the second inequality follows from the Markov inequality, and the equality follows from \( m_n^{1-\eta} \to (m_n^{-\eta} n) (m_n/\sqrt{n}) \to 0 \) (Assumption 2.1) and \( P_n \in \mathcal{P}_{\theta_n} \) (particularly \( E_{P_n} \left[ \sup_{\theta \in \Theta} |g (x, \theta)|^\eta \right] \leq M \)).

Combining these results, we obtain the uniform convergence \( \sup_{\theta \in \Theta} |E_{Q_n} [g_n (x, \theta)] - E_{P_n} [g_n (x, \theta)]| \to 0 \). Therefore, from the triangle inequality and \( |E_{Q_n} [g_n (x, \tilde{T}_{Q_n})]| = O \left( n^{-1/2} \right) \) (Lemma 10),

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

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

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
5. From (B.3) and (B.4), this implies
\[
\left\| dQ_n^{1/2} - d\tilde{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} (\tilde{T}_{Q_n} - \theta_n)' \Lambda_n dQ_n^{1/2} \right\|^2 \\
= \left\| dQ_n^{1/2} - d\tilde{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi_n' Q_n \Lambda_n dQ_n^{1/2} \right\|^2 + \left\| \frac{1}{2} (\tilde{T}_{Q_n} - \theta_n - \psi_n, Q_n)' \Lambda_n dQ_n^{1/2} \right\|^2 \\
+ \left\{ \int \left( dQ_n^{1/2} - d\tilde{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi_n' Q_n \Lambda_n dQ_n^{1/2} \right) \Lambda_n' dQ_n^{1/2} \right\} (\tilde{T}_{Q_n} - \theta_n - \psi_n, Q_n) \\
\] (B.3)
where the second equality follows from
\[
\int \left\{ dQ_n^{1/2} - d\tilde{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi_n' Q_n \Lambda_n dQ_n^{1/2} \right\} \Lambda_n' dQ_n^{1/2} \\
= \int \Lambda_n' \left\{ dQ_n^{1/2} - d\tilde{P}_{\theta_n, Q_n}^{1/2} \right\} dQ_n^{1/2} + \frac{1}{2} \psi_n' Q_n \int \Lambda_n \Lambda_n' dQ_n = 0.
\]
The left hand side of (B.3) satisfies
\[
\left\| dQ_n^{1/2} - d\tilde{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} (\tilde{T}_{Q_n} - \theta_n)' \Lambda_n dQ_n^{1/2} \right\| \\
\leq \left\| dQ_n^{1/2} - d\tilde{P}_{\theta_n, Q_n}^{1/2} \right\| + o \left( |\tilde{T}_{Q_n} - \theta_n| \right) + o \left( n^{-1/2} \right) \\
\leq \left\| dQ_n^{1/2} - d\tilde{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi_n' Q_n \Lambda_n dQ_n^{1/2} \right\| + o \left( |\tilde{T}_{Q_n} - \theta_n| \right) + o \left( |\psi_n, Q_n| \right) + o \left( n^{-1/2} \right) \\
\] (B.4)
where the first inequality follows from the triangle inequality and Lemma 4, the second inequality follows from $\tilde{T}_{Q_n} = \arg \min_{\theta \in \Theta} \| dQ_n^{1/2} - d\tilde{P}_{\theta, Q_n}^{1/2} \|$, and the third inequality follows from the triangle inequality and Lemma 5. From (B.3) and (B.4),
\[
\left\| dQ_n^{1/2} - d\tilde{P}_{\theta_n, Q_n}^{1/2} + \frac{1}{2} \psi_n' Q_n \Lambda_n dQ_n^{1/2} \right\|^2 \\
\leq \left\| dQ_n^{1/2} - d\tilde{P}_{\theta_n, Q_n}^{1/2} \right\|^2 + o \left( ||\tilde{T}_{Q_n} - \theta_n|| + o \left( |\psi_n, Q_n| \right) + o \left( n^{-1/2} \right) \right).
\]
This implies
\[
o \left( |\tilde{T}_{Q_n} - \theta_n| \right) + o \left( |\psi_n, Q_n| \right) + o \left( n^{-1/2} \right)
\] (B.5)
for all $n$ large enough, where the second inequality follows from Lemma 8 and $P_n \in P_{\theta_n}$ (particularly $\det \left( E_P [\hat{y}(x, \theta)'] E_P \left[ g(x, \theta) g(x, \theta)' \right]^{-1} E_P [\hat{y}(x, \theta)] \right) > c$).
We now analyze $\psi_{n,Q_n}$. From the definition of $\psi_{n,Q_n}$,

$$
\psi_{n,Q_n} = -2 \left\{ \left( \int \Lambda_n A_n^* dQ_n \right)^{-1} - V_n \right\} \int \Lambda_n \left\{ dQ_n^{1/2} - d\bar{P}_n^{1/2} \right\} dQ_n^{1/2}
$$

(B.6)

$$
-2V_n \int \Lambda_n \left\{ dQ_n^{1/2} - d\bar{P}_n^{1/2} \right\} dQ_n^{1/2}.
$$

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

$$
-2V_n \int \Lambda_n \left\{ dQ_n^{1/2} - d\bar{P}_n^{1/2} \right\} dQ_n^{1/2} = -2V_n G_n^\prime \Omega^{-1} \left\{ \int g_n (x, \theta_n) g_n (x, \theta_n)^\prime dQ_n \right\} \gamma_n (\theta_n, Q_n)
$$

$$
+2V_n G_n^\prime \Omega^{-1} \left\{ \int \frac{\gamma_n (\theta_n, Q_n)^\prime g_n (x, \theta_n)}{1 + \gamma_n (\theta_n, Q_n)^\prime g_n (x, \theta_n)} g_n (x, \theta_n) g_n (x, \theta_n)^\prime dQ_n \right\} \gamma_n (\theta_n, Q_n)
$$

$$
= -V_n G_n^\prime \Omega^{-1} \left\{ \int g_n (x, \theta_n) dQ_n + \frac{1}{2} \int \phi_n (x, \theta_n, Q_n) g_n (x, \theta_n) dQ_n \right\} + o \left( n^{-1/2} \right)
$$

$$
= -V_n \int \Lambda_n dQ_n + o \left( n^{-1/2} \right),
$$

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

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

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

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

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

$$
0 = \int \frac{g_n (x, \theta)}{(1 + \gamma_n (\theta, Q) g_n (x, \theta))^2} dQ = E_Q \left[ g_n (x, \theta) \left\{ 1 - 2 \gamma_n (\theta, Q) g_n (x, \theta) + g_n (x, \theta, Q) \right\} \right],
$$

with

$$
\phi_n (x, \theta, Q) = \frac{3 (\gamma_n (\theta, Q)^\prime g_n (x, \theta))^2 + 2 (\gamma_n (\theta, Q)^\prime g_n (x, \theta))^3}{(1 + \gamma_n (\theta, Q)^\prime g_n (x, \theta))^3}.
$$
Denote \( t_n = \bar{T}_{Q_n} - \theta_n \). From the triangle inequality and (B.7),

\[
\left\| \hat{d} \hat{P}^{1/2}_{T_{Q_n}, Q_n} - d \hat{P}^{1/2}_{\theta_n, Q_n} + \frac{1}{2} t_n' \Lambda_n dQ_n^{1/2} \right\| \\
\leq \left\| \left\{ \gamma_n (\theta_n, Q_n)' g_n(x, \theta_n) - \gamma_n (\bar{T}_{Q_n}, Q_n)' g_n(x, \bar{T}_{Q_n}) \right\} dQ_n^{1/2} + \frac{1}{2} t_n' \Lambda_n dQ_n^{1/2} \right\| \\
+ \left\| \left\{ \gamma_n (\theta_n, Q_n)' g_n(x, \theta_n) - \gamma_n (\bar{T}_{Q_n}, Q_n)' g_n(x, \bar{T}_{Q_n}) \right\} \right\| dQ_n^{1/2} = T_1 + T_2.
\]

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

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

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

\[
T_{11} \leq \left\| \left\{ -\frac{1}{2} E_{Q_n} \left[ g_n(x, \bar{T}_{Q_n})' \right] E_{Q_n} \left[ g_n(x, \bar{T}_{Q_n}) \theta_n(x, \theta_n) \right]^{-1} g_n(x, \bar{T}_{Q_n}) \right\} dQ_n^{1/2} \right\| \\
+ \left\| \left\{ -\frac{1}{2} E_{Q_n} \left[ g_n(x, \bar{T}_{Q_n})' \right] E_{Q_n} \left[ g_n(x, \theta_n) \theta_n(x, \theta_n) \right]^{-1} \left\{ g_n(x, \bar{T}_{Q_n}) - g_n(x, \theta_n) \right\} dQ_n^{1/2} \right\| \\
+ \left\| \frac{1}{2} t_n' \left( \int \hat{g}_n(x, \hat{\theta}) dQ_n - G_n \right)' E_{Q_n} \left[ g_n(x, \theta_n) \theta_n(x, \theta_n) \right]^{-1} g_n(x, \theta_n) dQ_n^{1/2} \right\| \\
+ \left\| \frac{1}{2} t_n' G_n^{-1} - E_{Q_n} \left[ g_n(x, \theta_n) \theta_n(x, \theta_n) \right]^{-1} \right\} g_n(x, \theta_n) dQ_n^{1/2} \right\| \\
= o(n^{-1/2}) + o(t_n),
\]

where \( \hat{\theta} \) is a point on the line joining \( \theta_n \) 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,

$$|E_{P_n} [g_n (x, \theta_{t,n})]| \leq |E_{P_n} [g (x, \theta_{t,n}) 1 \{x \notin X_n\}]| + |E_{P_n} [g (x, \theta_{t,n})]|. \tag{B.9}$$

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

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

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

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

$$\left| E_{P_n} \left[ g_n \left( x, \theta_{t,n} \right) g_n \left( x, \theta_{t,n} \right) \right] - \Omega_n \right|$$

$$\leq \left| E_{P_n} \left[ g_n \left( x, \theta_{t,n} \right) g_n \left( x, \theta_{t,n} \right) \right] - E_{P_n} \left[ g \left( x, \theta_{t,n} \right) g \left( x, \theta_{t,n} \right) \right] \right| + \left| E_{P_n} \left[ g \left( x, \theta_{t,n} \right) g \left( x, \theta_{t,n} \right) \right] - \Omega_n \right|. \tag{B.10}$$

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 $\theta_n$.

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 $a_n m_n \to 0$ and $a_n n^{1/2} \to \infty$. Observe that

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

Since $R_n (P_n, \theta_{t,n}, \gamma)$ is twice continuously differentiable with respect to $\gamma$ and $\Gamma_n$ is compact, $\hat{\gamma} = \arg \max_{\gamma \in \Gamma_n} R_n (P_n, \theta_{t,n}, \gamma)$ exists for each $n \in \mathbb{N}$. A Taylor expansion around $\hat{\gamma} = 0$ yields

$$-1 = R_n (P_n, \theta_{t,n}, 0) \leq R_n (P_n, \theta_{t,n}, \hat{\gamma}) = -1 + \hat{\gamma}' E_{P_n} [g_n (x, \theta_{t,n})] - \hat{\gamma}' E_{P_n} \left[ \frac{g_n (x, \theta_{t,n}) g_n (x, \theta_{t,n})'}{1 + \hat{\gamma}' g_n (x, \theta_{t,n})} \right] \hat{\gamma}$$

$$\leq -1 + \hat{\gamma}' E_{P_n} [g_n (x, \theta_{t,n})] - C |\hat{\gamma}| E_{P_n} [g_n (x, \theta_{t,n}) g_n (x, \theta_{t,n})'] \hat{\gamma}$$

$$\leq -1 + |\gamma| |E_{P_n} [g_n (x, \theta_{t,n})]| - C |\gamma|^2, \tag{B.11}$$

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

Thus, Lemma 6 implies

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

From $a_n n^{1/2} \to \infty$, $\hat{\gamma}$ is an interior point of $\Gamma_n$ and satisfies the first-order condition $\partial R_n \left( Q_n, \theta_n, \hat{\gamma} \right) / \partial \gamma = 0$ for all $n$ large enough. Since $R_n (Q_n, \theta_n, \gamma)$ is concave in $\gamma$ for all $n$ large enough, $\hat{\gamma} = \arg \max_{\gamma \in \mathbb{R}^m} R_n (P_n, \theta_{t,n}, \gamma)$
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 \( \frac{m_n}{\sqrt{n}} \to 0 \) (Assumption 2.1).


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

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

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 \( E_{P_n} [\sup_{\theta \in \Theta} |g (x, \theta)|^q] \leq M \)).

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

\[
E_{Q_n} [g_n (x, \theta_n) g_n (x, \theta_n)'] \leq E_{Q_n} [g_n (x, \theta_n) g_n (x, \theta_n)'] - \Omega_n,
\]

(B.13) \( \leq \left| 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)'] \right| + \left| E_{P_n} [g (x, \theta_n) g (x, \theta_n)'] \mathbb{I} \{ x \notin \mathcal{X}_n \} \right|.
\]

The first term of (B.13) satisfies

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

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 \( \frac{m_n}{\sqrt{n}} \to 0 \) (Assumption 2.1) and \( P_n \in \mathcal{P}_{\theta_n} \) (particularly \( E_{P_n} \left[ \sup_{\theta \in \mathcal{X}} |g (x, \theta)|^q \right] \leq M \)). The second term of (B.13) satisfies

\[
\left| E_{P_n} [g (x, \theta_n) g (x, \theta_n)'] \mathbb{I} \{ x \notin \mathcal{X}_n \}] \right| \\
\leq \left( \int |g (x, \theta_n) g (x, \theta_n)'|^{1+\delta} dP_n \right)^{\frac{1}{1+\delta}} \left( \int \mathbb{I} \{ x \notin \mathcal{X}_n \} dP_n \right)^{\frac{1}{1+\delta}} \\
\leq \left( E_{P_n} \left[ |g (x, \theta_n)|^{2+\delta} \right] \right)^{\frac{1}{1+\delta}} \left( m_n^{-\eta} E_{P_n} \left[ ||g (x, \theta_n)||^q \right] \right)^{\frac{1}{1+\delta}} = o \left( 1 \right),
\]
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 $E_{P_n} \left[ \sup_{\theta \in N_{\theta_n}} |g(x, \theta)|^4 \right] \leq M$).

**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 $R_n (P_n, \theta_{1,n}, \gamma)$.

**B.10. Proof of Lemma 10.**

**B.10.1. Proof of the first statement.** Define $\tilde{\gamma} = \frac{E_{Q_n} [g_n (x, \tilde{T}_{Q_n})]}{\sqrt{n} \| E_{Q_n} [g_n (x, \tilde{T}_{Q_n})] \|}$. 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

\[
E_{Q_n} \left[ g_n (x, \tilde{T}_{Q_n}) g_n (x, \tilde{T}_{Q_n})' \right] \\
\leq \int \sup_{\theta \in \Theta} |g_n (x, \theta)|^2 \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\}^2 + 2 \int \sup_{\theta \in \Theta} |g_n (x, \theta)|^2 dP_n^{1/2} \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\} + E_{P_n} \left[ \sup_{\theta \in \Theta} |g_n (x, \theta)|^2 \right]
\]

\[(B.15) \quad m_n \sqrt{2 \pi n} + 2 m_n \sqrt{E_{P_n} \left[ \sup_{\theta \in \Theta} |g_n (x, \theta)|^2 \right] + E_{P_n} \left[ \sup_{\theta \in \Theta} |g (x, \theta)|^2 \right]} \leq C E_{P_n} \left[ \sup_{\theta \in \Theta} |g (x, \theta)|^2 \right],
\]

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

\[
R_n (Q_n, \tilde{T}_{Q_n}, \tilde{\gamma}) = -1 + \tilde{\gamma}' E_{P_n} \left[ g_n (x, \tilde{T}_{Q_n}) \right] - \tilde{\gamma}^2 E_{P_n} \left[ \frac{g_n (x, \tilde{T}_{Q_n}) g_n (x, \tilde{T}_{Q_n})'}{(1 + \tilde{\gamma}' g_n (x, \tilde{T}_{Q_n}))^3} \right] \tilde{\gamma}
\]

\[
\geq -1 + n^{-1/2} \left| E_{Q_n} \left[ g_n (x, \tilde{T}_{Q_n}) \right] \right| - C \tilde{\gamma}' E_{Q_n} \left[ g_n (x, \tilde{T}_{Q_n}) g_n (x, \tilde{T}_{Q_n})' \right] \tilde{\gamma}
\]

\[(B.16) \quad \geq -1 + n^{-1/2} \left| E_{Q_n} \left[ g_n (x, \tilde{T}_{Q_n}) \right] \right| - C n^{-1},
\]

for all $n$ large enough, where $\tilde{\gamma}$ is a point on the line joining 0 and $\gamma$, 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)), $\gamma_n (\tilde{T}_{Q_n}, Q_n)$ and $\tilde{T}_{Q_n}$ are written as $\gamma_n (\tilde{T}_{Q_n}, Q_n) = \arg \max_{\gamma \in \mathbb{R}_+} R_n (Q_n, \tilde{T}_{Q_n}, \gamma)$ and $\tilde{T}_{Q_n} = \arg \min_{\theta \in \Theta} R_n (Q_n, \theta, \gamma_n (\theta, Q_n))$. Therefore, from (B.16),

\[
-1 + n^{-1/2} \left| E_{Q_n} \left[ g_n (x, \tilde{T}_{Q_n}) \right] \right| - C n^{-1}
\]

\[(B.17) \quad \leq R_n (Q_n, \tilde{T}_{Q_n}, \tilde{\gamma}) \leq R_n (Q_n, \tilde{T}_{Q_n}, \gamma_n (\tilde{T}_{Q_n}, Q_n)) \leq R_n (Q_n, \theta_n, \gamma_n (\theta_n, Q_n)).
\]

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

\[
R_n (Q_n, \theta_n, \gamma_n (\theta_n, Q_n)) \leq -1 + |\gamma_n (\theta_n, Q_n)| \left| E_{Q_n} [g_n (x, \theta_n)] \right| - C |\gamma_n (\theta_n, Q_n)|^2 = -1 + O \left( n^{-1} \right).
\]
From (B.17) and (B.18), the conclusion follows.


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

\[
|E_{Q_n} [g_n (x, \bar{T}_{Q_n})] - G_n| \leq |E_{Q_n} [g_n (x, \bar{T}_{Q_n})] - E_{P_n} [g_n (x, \bar{T}_{Q_n})]| + |E_{P_n} [1 \{ x \notin X_n \} \hat{g} (x, \bar{T}_{Q_n})]| + |E_{P_n} [\hat{g} (x, \bar{T}_{Q_n})] - G_n|.
\]  

(B.19)

The first term of (B.19) satisfies

\[
|E_{Q_n} [g_n (x, \bar{T}_{Q_n})] - E_{P_n} [g_n (x, \bar{T}_{Q_n})]| 
\leq \left| \int g_n (x, \bar{T}_{Q_n}) \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\}^2 \right| + 2 \left| \int g_n (x, \bar{T}_{Q_n}) dP_n^{1/2} \{ dQ_n^{1/2} - dP_n^{1/2} \} \right|
\]

\[
\leq \sup_{x \in X_n, \theta \in \Omega_n} \left| \hat{g}_n (x, \theta) \right| \sqrt{n} + 2E_{P_n} \left[ \sup_{\theta \in \Omega_n} \left| \hat{g}_n (x, \theta) \right| \right] \frac{\sqrt{n}}{\sqrt{n}} = o(1),
\]

where the first inequality follows from the triangle inequality, the second inequality follows from the Cauchy-Schwarz inequality, and the equality follows from \(P_n \in P_{\theta_n} \) (particularly \(\sup_{x \in X_n, \theta \in \Omega_n} \left| \hat{g} (x, \theta) \right| = o (n^{1/2}) \) and \(E_{P_n} \left[ \sup_{\theta \in \Omega_n} \left| \hat{g} (x, \theta) \right| \right] < M \). 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 \(P_n \in P_{\theta_n} \) (particularly by the continuity of \(\hat{g} (x, \theta)\) at \(\theta_n\)) 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 \(R_n (P_n, \theta_{1,n}, \gamma)\).


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

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

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

\[
|V_nG_n'\Omega_n^{-1}|^{2+\delta} < \infty
\]
Also observe that
\[
E_{Q_n} \left[ |g_n(x, \theta_n)|^{2+\delta} \right] = \int |g_n(x, \theta_n)|^{2+\delta} \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\}^2 + 2 \int |g_n(x, \theta_n)|^{2+\delta} dP_n^{1/2} \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\} + E_{P_n} \left[ |g_n(x, \theta_n)|^{2+\delta} \right]
\]
\[
\leq m_n^{2+\delta} \frac{r_0}{n} + 2m_n^{1+\delta} E_{P_n} \left[ |g(x, \theta_n)|^2 \right] \frac{r_0}{\sqrt{n}} + E_{P_n} \left[ |g(x, \theta_n)|^{2+\delta} \right] < \infty,
\]
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 \( E_{P_n} [\sup_{\theta \in \Theta} |g(x, \theta)|^6] \leq M \)). Therefore, the conclusion is obtained.

**B.14. Proof of Lemma 14.**

**B.14.1. Proof of the first statement.** From the triangle inequality,
\[
\left| E_{P_n} [g_n(x, \theta_n)] \right| \leq \sum_{j=1}^{m} \left| \frac{1}{n} \sum_{i=1}^{n} W_{ijn} \right| \sqrt{\text{Var}_{Q_n} (g_{jn}(x, \theta_n)) + |E_{Q_n} [g_n(x, \theta_n)]|},
\]
where \( W_{ijn} = \frac{g_{jn}(x, \theta_n) - E_{Q_n}[g_{jn}(x, \theta_n)]}{\sqrt{\text{Var}_{Q_n}(g_{jn}(x, \theta_n))}} \). From Lemma 8, the second term is \( O \left( n^{-1/2} \right) \). Thus, if we can show that
\[
\text{(B.20)} \quad \text{Var}_{Q_n} (g_{jn}(x, \theta_n)) < C,
\]
\[
\text{(B.21)} \quad E \left[ |W_{ijn}|^{2+\delta} \right] < \infty \text{ for some } \delta > 0,
\]
for all \( n \) large enough, then Lyapounov’s central limit theorem yields the conclusion.

We first check (B.20). Observe that
\[
\text{Var}_{Q_n} (g_{jn}(x, \theta_n)) = E_{Q_n} \left[ g_{jn}(x, \theta_n)^2 \right] - (E_{Q_n} [g_{jn}(x, \theta_n)])^2 = E_{P_n} \left[ g_j(x, \theta_n)^2 \right] + O \left( n^{-1} \right)
\]
\[
\in (C_1, C_2),
\]
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),
\[
E_{Q_n} \left[ |W_{ijn}|^{2+\delta} \right] \leq CE_{Q_n} \left[ |g_{jn}(x, \theta_n)|^{2+\delta} \right] \leq \int |g_{jn}(x, \theta_n)|^{2+\delta} \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\}^2 + 2 \int |g_{jn}(x, \theta_n)|^{2+\delta} dP_n^{1/2} \left\{ dQ_n^{1/2} - dP_n^{1/2} \right\} + E_{P_n} \left[ |g_{jn}(x, \theta_n)|^{2+\delta} \right]
\]
\[
\leq m_n^{2+\delta} \frac{r_0}{n} + 2m_n^{1+\delta} E_{P_n} \left[ |g(x, \theta_n)|^2 \right] \frac{r_0}{\sqrt{n}} + E_{P_n} \left[ |g(x, \theta_n)|^{2+\delta} \right] \leq C,
\]
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 \( E_{P_n} [\sup_{\theta \in \Theta} |g(x, \theta)|^6] \leq M \)). Therefore, the conclusion is obtained.
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 $E_{P_n} [\|g(x, \theta_n)\|^p] < M$).

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

$$|E_{P_n} [g_n(x, \theta_n)g_n(x, \theta_n)' - \Omega_n]|$$

$$\leq |E_{P_n} [g_n(x, \theta_n)g_n(x, \theta_n)'] - E_{Q_n} [g_n(x, \theta_n)g_n(x, \theta_n)']| + |E_{Q_n} [g_n(x, \theta_n)g_n(x, \theta_n)'] - \Omega_n|.$$

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.17. Proof of Lemma 17. Similar to the proof of Lemma 11.
References


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