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 and 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 most powerful in a minimax robust efficiency 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 $g : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^m$ be a vector of functions, which satisfies the moment condition:

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

We assume that $\theta_0$ is scalar for simplicity, i.e., $\Theta \subset \mathbb{R}$. We also assume that the model (1.1) is overidentified, i.e., $m > 1$. The unknown parameter $\theta_0$ can be estimated by, for example, the generalized method of moments (Hansen (1982)), empirical likelihood (Qin and Lawless (1994)), minimum Hellinger distance (Kitamura, Otsu, and Evdokimov (2009)), or their variants (Newey and Smith (2004)).

This paper focuses on the parameter hypotheses testing problem:

$$H_0 : \theta_0 = 0, \quad H_1 : \theta_0 \neq 0.$$ 

Although we consider the two-sided alternative hypothesis $H_1$, we can derive an analogous result for the one-sided alternatives ($H'_1 : \theta_0 > 0$ or $H''_1 : \theta_0 < 0$). To test the parameter hypothesis $H_0$ against $H_1$, several tests are available, such as the Wald or t-value test based on some point estimator of $\theta_0$, GMM distance test, empirical likelihood ratio test, and Lagrange multiplier-type test.

Date: This Version: February 7, 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 consider the asymptotic behaviors of tests when the data generating measure is perturbed from the measure \( P_0 \) satisfying (1.1) or its local alternative measures. In order to introduce our framework, we need some notation. Let

\[
P_\theta = \{ P \in \mathcal{M} : \int g(x, \theta) \, dP = 0 \},
\]

be the set of measures satisfying the moment condition at some \( \theta \in \Theta \). Under the null hypothesis \( H_0 : \theta_0 = 0 \), the data generating measure resides in the set \( P_0 \). Under the local alternative \( H_{0n} : \theta_0 = c/\sqrt{n} \), the (drifting) data generating measure belongs to the set \( P_{c/\sqrt{n}} \). A test \( \psi_n = \psi_n(x_1, \ldots, x_n) \) is defined as a binary function of the sample, where \( \psi_n = 1 \) means rejection and \( \psi_n = 0 \) means acceptance. The size of the test \( \psi_n \) is written as

\[
\sup_{Q \in P_0} E_Q [\psi_n],
\]

and the local power function is

\[
E_{Q_n} [\psi_n] \quad \text{under} \quad Q_n \in P_{c/\sqrt{n}},
\]

for \( c \in \mathbb{R} \) and \( n \in \mathbb{N} \). In this paper, we investigate the size properties of the test \( \psi_n \) under the blow-up version of the set \( P_0 \) for the null hypothesis based on the Hellinger distance (see Kitamura, Otsu, and Evdokimov (2009) for a motivating discussion to adopt the Hellinger distance), that is

\[
\alpha_{\psi,n} = \sup_{Q \in Q_{0n}} E_Q [\psi_n],
\]

where

\[
Q_{0n} = \bigcup_{P \in P_0} \left\{ Q : H(Q, P) \leq \frac{r_0}{\sqrt{n}} \right\},
\]

for \( r_0 > 0 \), 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 \( \nu \). Note that the set \( Q_{0n} \) allows deviations from the set \( P_0 \) for the null hypothesis in terms of the Hellinger distance but the distance from \( P_0 \) should decrease at the \( \sqrt{n} \)-rate. Except for the restriction \( H(Q, P) \leq \frac{r_0}{\sqrt{n}} \), we do not impose any parametric structure on \( Q \) about how to deviate from \( P \). Thus, \( \alpha_{\psi,n} \) is useful to characterize the robustness of the size properties of the test \( \psi_n \) under local perturbations. Similarly, we can consider the locally perturbed version of the local power function, that is

\[
\beta_{\psi,n}(c) = \sup_{Q \in Q_{1n}} E_Q [\psi_n],
\]

where

\[
Q_{1n} = \bigcup_{P \in P_{c/\sqrt{n}}} \left\{ Q : H(Q, P) \leq \frac{r_1}{\sqrt{n}} \right\},
\]

for \( r_1 > 0 \) and \( c \in \mathbb{R} \). Similar to \( Q_{0n} \), the set \( Q_{1n} \) allows local deviations from the set \( P_{c/\sqrt{n}} \) for the local alternative hypothesis. The function \( \beta_{\psi,n}(c) \) is useful to analyze the robustness of the local power properties of the test \( \psi_n \) against the deviations from \( P_{c/\sqrt{n}} \). This paper studies asymptotic behaviors of \( \alpha_{\psi,n} \) and \( \beta_{\psi,n}(c) \)
for different tests and seeks an asymptotically optimal test based on some optimality criteria using \( \alpha_{\psi,n} \) and \( \beta_{\psi,n}(c) \).

Our robust analysis framework for tests of moment condition models can be considered as extensions of the one for tests of parametric models considered by e.g., Huber and Strassen (1973), Rieder (1978), and Beran (1981).

We close this section by defining our proposed test, the Hellinger-based Wald test. Let \( P_n \) denote the empirical measure of observations \( \{x_i\}_{i=1}^n \). The minimum Hellinger distance estimator (MHDE), \( \hat{\theta} \), is defined as

\[
\hat{\theta} = \arg \min_{\theta \in \Theta} \inf_{P \in P_n, P \ll P_n} H(P, P_n) = \arg \min_{\theta \in \Theta} \max_{\gamma \in \mathbb{R}^m} -\frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \gamma' g(x_i, \theta)},
\]

where the second equality follows from the convex duality theory (Kitamura (2006)). In practice, we use the last expression in (1.5) to implement the MHDE. If the data are generated from \( P_0 \in P_{\theta_0} \) (i.e., correct specification), it is known that (see, Newey and Smith (2004))

\[
\sqrt{n} \left( \hat{\theta} - \theta_0 \right) \xrightarrow{d} N(0, V),
\]

where \( V = (G' \Omega^{-1} G)^{-1} \) with \( G = \mathbb{E}_{P_0} [\partial g(x, \theta_0)/\partial \theta] \) and \( \Omega = \mathbb{E}_{P_0} [g(x, \theta_0) g(x, \theta_0)'] \). The asymptotic variance \( V \) may be estimated by

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

where \( \hat{G} = \frac{1}{n} \sum_{i=1}^n \partial g(x_i, \hat{\theta}) / \partial \theta' \) and \( \hat{\Omega} = \frac{1}{n} \sum_{i=1}^n g(x_i, \hat{\theta}) g(x_i, \hat{\theta})' \). Based on the MHDE \( \hat{\theta} \) and variance estimator \( \hat{V} \), the Hellinger-based Wald test for \( H_0: \theta_0 = 0 \) against \( H_1: \theta_0 \neq 0 \) is defined as

\[
\psi_{H,n} = \mathbb{I} \left\{ \left| \frac{\hat{\theta}}{\sqrt{\hat{V}/n}} \right| \geq \text{critical value} \right\},
\]

where \( \mathbb{I} \{ \cdot \} \) is the indicator function. Based on the asymptotic distribution in (1.6) under \( P_0 \), the critical value is typically set as the \( 100 \left( 1 - \frac{\alpha}{2} \right) \% \) critical value of the standard normal 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 local alternatives. In the next section, we will argue that this Hellinger-based Wald test \( \psi_{H,n} \) shows some optimal properties in terms of the perturbed versions of the size and local power properties, \( \alpha_{\psi,n} \) and \( \beta_{\psi,n}(c) \).

2. Main Results

Let \( \mathcal{X}_n = \{ x \in \mathcal{X} : \sup_{\theta \in \Theta} |g(x, \theta)| \leq m_n \} \), where \( \{m_n\}_{n \in \mathbb{N}} \) is a sequence of positive numbers satisfying \( m_n \to \infty \) as \( n \to \infty \). The trimming set \( \mathcal{X}_n \) and trimming constant \( m_n \) appear only in Assumption 2.1 (v) and (vii) below. The Hellinger-based Wald test \( \psi_{H,n} \) does not require the trimming. We impose the following assumptions.
Suppose the following conditions hold:

(i): $\{x_i\}_{i=1}^n$ is iid;
(ii): $\Theta$ is compact;
(iii): $\theta_0 \in \text{int}(\Theta)$ is a unique solution to $E_{P_0}[g(x,\theta)] = 0$;
(iv): $g(x,\theta)$ is continuous over $\Theta$ at each $x \in \mathcal{X}$;
(v): $E_{P_0} \left[ \sup_{\theta \in \Theta} |g(x,\theta)|^\eta \right] < \infty$ for some $\eta > 2$, and there exists a neighborhood $\mathcal{N}$ around $\theta_0$ such that $E_{P_0} \left[ \sup_{x \in \mathcal{X}, \theta \in \mathcal{N}} |g(x,\theta)|^4 \right] < \infty$, $g(x,\theta)$ is continuously differentiable a.s. in $\mathcal{N}$, $\sup_{x \in \mathcal{X}, \theta \in \mathcal{N}} |\partial g(x,\theta)/\partial \theta| = o\left(n^{1/2}\right)$, and $E_{P_0} \left[ \sup_{\theta \in \mathcal{N}} |\partial g(x,\theta)/\partial \theta|^2 \right] < \infty$;
(vi): $G$ has the full column rank and $\Omega$ is positive definite;
(vii): $\{m_n\}_{n \in \mathbb{N}}$ satisfies $m_n \to \infty$, $nm_n^{-\eta} \to 0$, and $n^{-1/2}m_n^{1+\epsilon} = O(1)$ for some $0 < \epsilon \leq 2$ as $n \to \infty$.

These assumptions are basically same as the ones in Kitamura, Otsu, and Evdokimov (2009), which establish some optimal robustness properties of the MHDE $\hat{\theta}$ as a point estimator. Thus, the same comments apply here. Also the above assumptions are sufficient to derive the conventional asymptotic distribution in (1.6) under $P_0$.

2.1. Size Optimality. This section investigates the size properties of tests $\alpha_{\psi,n}$ in (1.3) for a general class of tests. To define the class of tests, we first introduce some notions for point estimators of $\theta_0$. Let $\hat{\theta}_a = T_a(P_n)$ be an estimator of $\theta_0$ based on a mapping $T_a: \mathcal{M} \to \Theta$. The mapping $T$ for the MHDE $\hat{\theta} = T(P_n)$ is defined as $T(P) = \arg\min_{\psi \in \Theta} \inf_{P \in \mathcal{P}_a} H(P,Q)$. We use the following concepts on the mapping $T_a$.

Definition 1. Let $P_{\theta,\zeta}$ be a regular parametric submodel (see, Kitamura, Otsu, and Evdokimov (2009)) of $\cup_{\theta \in \Theta} P_\theta$ such that $P_{\theta,0} = P_0$ and $P_{\theta_0 + t/\sqrt{n},\zeta_n} \in \left\{ Q : H(Q,P_0) \leq \frac{n}{\sqrt{n}} \right\}$ holds for $\zeta_n = O\left(n^{-1/2}\right)$ and each $P_0 \in P_0$ eventually.

(i): The mapping $T_a$ is called **Fisher consistent** if for every $\{P_{\theta_n,\zeta_n}\}_{n \in \mathbb{N}}$ and $t \in \mathbb{R}$,

\begin{equation}
\sqrt{n} \left( T_a(P_{\theta_0 + t/\sqrt{n},\zeta_n}) - \theta_0 \right) \to t.
\end{equation}

(ii): The mapping $T_a$ is called **Gaussian regular** if for every $\{P_{\theta_n,\zeta_n}\}_{n \in \mathbb{N}}$ with $(\theta_n,\zeta_n) = (\theta_0,0) + O\left(n^{-1/2}\right)$,

\begin{equation}
\sqrt{n} \left( T_a(P_n) - T_a(P_{\theta_n,\zeta_n}) \right) \overset{d}{\to} N(0,V_a), \quad \text{under } P_{\theta_n,\zeta_n},
\end{equation}

with $V_a \geq V$.

(iii): An estimator $\hat{V}_a$ satisfies $\hat{V}_a \overset{P}{\to} V_a$ under each sequence $\{P_{\theta_n,\zeta_n}\}_{n \in \mathbb{N}}$.

These requirements are weak and satisfied by the mappings for popular estimators, such as the generalized method of moments, empirical likelihood, and exponential tilting estimators. Since $V = (G'\Omega^{-1}G)^{-1}$ is the
semiparametric efficiency bound to estimate $\theta_0$, the requirement $V_a \geq V$ on the asymptotic variance in (2.2) is reasonable.

Let $z_\alpha$ and $\chi^2_{1,\alpha}$ be the $100(1-\alpha)$% critical values of the standard normal distribution and the $\chi^2_1$ distribution, respectively. The Wald test based on the estimator $\hat{\theta}_a = T_a(P_n)$ is defined as

$$
\psi_{a,n} = \mathbb{I}\left\{ \left( \frac{\hat{\theta}_a}{\sqrt{V_a/n}} \right)^2 \geq \chi^2_{1,\alpha} \right\}.
$$

For the size optimality result, we consider the following class of tests.

**Definition 2.** A test $\psi_n$ belongs to the class $\mathcal{S}$ of tests (denoted by $\psi_n \in \mathcal{S}$) if

(i): $\psi_n = \mathbb{I}\{ T_n \geq \chi^2_{1,\alpha} \}$,

(ii): for some $\theta_a$, it holds $T_n = \left( \frac{\hat{\theta}_a}{\sqrt{V_a/n}} \right)^2 + o_p(1)$ under each sequence $\{ P_{\theta_n,\zeta_n} \}_{n \in \mathbb{N}}$,

(iii): the mapping $T_a$ to define $\hat{\theta}_a = T_a(P_n)$ satisfies the requirements in Definition 1.

Note that the class $\mathcal{S}$ contains asymptotically equivalent tests to the Wald test based on some estimator $\hat{\theta}_a$ which satisfies the requirements in Definition 1. 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 size optimality result for the Hellinger-based Wald test $\psi_{H,n}$ in the class of tests $\mathcal{S}$ is presented as follows.

**Theorem 3.** Let $W$ be a random variable that obeys the non-central $\chi^2$ distribution with degree of freedom 1 and non-centrality $4\sigma_0^2$. Suppose that Assumption 2.1 holds.

(i): For every test $\psi_n \in \mathcal{S}$ in Definition 2,

$$
\liminf_{n \to \infty} \sup_{Q \in \mathcal{Q}_{0n}} E_Q[\psi_n] \geq \Pr\{ W \geq \chi^2_{1,\alpha} \},
$$

for each $r_0 > 0$.

(ii): The Hellinger-based Wald test $\psi_{H,n} = \mathbb{I}\left\{ \left| \frac{\hat{\theta}_a}{\sqrt{V/a}} \right| \geq z_{\alpha/2} \right\}$ belongs to the class $\mathcal{S}$ and satisfies

$$
\lim_{n \to \infty} \sup_{Q \in \mathcal{Q}_{0n}} E_Q[\psi_{H,n}] = \Pr\{ W \geq \chi^2_{1,\alpha} \},
$$

for each $r_0 > 0$.

**Remark 4.** Part (i) of this theorem derives the minimax bound for the size distortion over the perturbed set $\mathcal{Q}_{0n}$. For any test that belongs to the class $\mathcal{S}$, the worst size $\sup_{Q \in \mathcal{Q}_{0n}} E_Q[\psi_{a,n}]$ is asymptotically bounded from
below by the probability $\Pr \{ W \geq \chi^2_{1, \alpha} \}$. Note that $\Pr \{ W \geq \chi^2_{1, \alpha} \} > 1 - \alpha$. Part (ii) says that the minimax size distortion bound is attained by the Hellinger-based Wald test $\psi_{H,n}$.

This optimal robust size property of the Hellinger-based Wald test directly implies the optimal robust coverage property of the Hellinger-based confidence interval $\left[ \hat{\theta} \pm z_{\alpha/2} \sqrt{\hat{V}/n} \right]$ compared to the one $\left[ \hat{\theta}_a \pm z_{\alpha/2} \sqrt{\hat{V}_a/n} \right]$ based on the alternative estimator $\hat{\theta}_a$ or the one obtained by inverting some likelihood ratio-type statistic based on the generalized method of moments or empirical likelihood criterion function. Kitamura, Otsu, and Evdokimov (2009) derived optimal robust properties of the MHDE $\hat{\theta}$ for point estimation. The present result stretches the optimal robustness of the Hellinger-based method to parameter hypothesis testing and interval estimation.

Although the above theorem is presented for the case where the parameter $\theta_0$ is scalar, it is easy to extend this theorem to the case where the parameter of interest $\tau(\theta_0)$ is scalar but $\theta_0$ is a vector. In this case, we need to assume that the function $\tau$ is continuously differentiable at $\theta_0$.

We can also show that the Hellinger distance test

$$
\psi^{D}_{H,n} = 1 \left\{ \frac{2n}{\hat{\theta}} \left( \ell(0) - \ell(\hat{\theta}) \right) \geq \chi^2_{1, \alpha} \right\}
$$

with $\ell(\theta) = \max_{\gamma \in \mathbb{R}^m} -\frac{1}{2} \sum^n_{i=1} \frac{1}{1 + \gamma^T g(x_i, \theta)}$ satisfies the same optimal property as $\psi_{H,n}$.

### 2.2. Power Optimality.

The optimal robust power property of the Hellinger-based Wald test is presented as follows.

**Theorem 5.** Let $W_1$ be a random variable that obeys the non-central $\chi^2$ distribution with degree of freedom 1 and non-centrality $(|c| V^{-1/2} - 2r_1)^2$. Suppose that Assumption 2.1 holds.

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

$$
\limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_0} \mathbb{E}_Q[\psi_n] \leq \alpha,
$$

for each $r_0 > 0$, then

$$
\limsup_{n \to \infty} \inf_{Q \in \mathcal{Q}_1} \mathbb{E}_Q[\psi_n] \leq \Pr \{ W_1 \geq (z_{\alpha/2} + 2r_0)^2 \},
$$

for each $c \in \mathbb{R}$, $r_0 > 0$, and $r_1 \in (0, |c| V^{-1/2})$.

(ii): The Hellinger-based Wald test $\psi_{H,n} = 1 \left\{ \left| \frac{\hat{\theta}}{\sqrt{\hat{V}/n}} \right| \geq z_{\alpha/2} + 2r_0 \right\}$ satisfies

$$
\limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_0} \mathbb{E}_Q[\psi_{H,n}] \leq \alpha,
$$

and

$$
\liminf_{n \to \infty} \mathbb{E}_Q[\psi_{H,n}] = \Pr \{ W_1 \geq (z_{\alpha/2} + 2r_0)^2 \},
$$

for each $c \in \mathbb{R}$ and $r_1 \in (0, |c| V^{-1/2})$.  


Remark 6. Note that compared to Part (i) of Theorem 3, Part (i) of this theorem holds for any test (i.e., measurable binary function) satisfying (2.3), which may not belong to the set \( S \) in Definition 2.

Part (i) of this theorem says that if the asymptotic rejection probability of the test \( \psi_n \) over \( Q_{0n} \) is controlled to be less than \( \alpha \), the minimum asymptotic power of the test \( \psi_n \) over \( Q_{1n} \) never exceeds \( \Pr \{ W_1 \geq (z_{\alpha/2} + 2r_0)^2 \} \).

Part (ii) says that the Hellinger-based Wald test with the critical value \( z_{\alpha/2} + 2r_0 \) not only satisfies the requirement on the rejection probability over \( Q_{0n} \) but also attains the maxmin asymptotic power bound \( \Pr \{ W_1 \geq (z_{\alpha/2} + 2r_0)^2 \} \).

3. Conclusion

To be written.
Appendix A. Proof of Theorems

Notation. Let “⇒” mean the weak convergence, and

\[ g_n(x, \theta) = g(x, \theta) \mathbb{1}\{x \in \mathcal{X}_n\} \]

A.1. Proof of Theorem 3. Proof of (i). Pick any \( \epsilon \in (0, r_0), r_0 > 0 \), and \( P_0 \in \mathcal{P}_0 \). Let \( t_a = 2(r - \epsilon)V^{1/2} \).

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

\[ \frac{dP_{t_a/\sqrt{n}}}{dP_0} = \frac{1 + \zeta_{t_a,n}g_n \left( x, \frac{t_a}{\sqrt{n}} \right) }{\int \left( 1 + \zeta_{t_a,n}g_n \left( x, \frac{t_a}{\sqrt{n}} \right) \right) dP_0}, \]

where \( \zeta_{t_a,n} = -E_{P_0} \left[ g \left( x, \frac{t_a}{\sqrt{n}} \right) g_n \left( x, \frac{t_a}{\sqrt{n}} \right) \right] ^{-1} E_{P_0} \left[ g \left( x, \frac{t_a}{\sqrt{n}} \right) \right] \). By a similar argument to the proof of Theorem 3.1 of Kitamura, Otsu, and Evdokimov (2009), we can show that

\[ \lim_{n \to \infty} nH \left( P_{t_a/\sqrt{n}}, P_0 \right) ^2 = \frac{1}{4} V^{-1} t_a^2 = (r_0 - \epsilon)^2, \]

which implies

(A.1) \( P_{t_a/\sqrt{n}} \in \mathcal{Q}_0 \),

for all \( n \) large enough.

Since \( \psi_n \) belongs to the class \( S \) in Definition 2, we can write it as \( \psi_n = \mathbb{1}\{T_n \geq \chi^2_{1,\alpha}\} \) and there exists a Wald test statistic \( \frac{nT_a(P_n)^2}{V_a} \) satisfying

(A.2) \( T_n = \frac{nT_a(P_n)^2}{V_a} + r_n \) with \( r_n = o_p(1) \) under \( \{P_{t_a/\sqrt{n}}\}_{n \in \mathbb{N}} \).

Thus, we have

\[
\liminf_{n \to \infty} \sup_{Q \in \mathcal{Q}_0} E_Q [\psi_n] \\
\geq \liminf_{n \to \infty} E_{P_{t_a/\sqrt{n}}} \left[ \mathbb{1}\{T_n \geq \chi^2_{1,\alpha}\} \right] \\
= \liminf_{n \to \infty} \int \mathbb{1}\left\{ \left( \sqrt{n} \left( T_a(P_n) - T_a(P_{t_a/\sqrt{n}}) \right) + \sqrt{n}T_a \left( P_{t_a/\sqrt{n}} \right) \right)^2 + r_n \geq \chi^2_{1,\alpha} \right\} dP_{t_a/\sqrt{n}} \\
= E_Z \left[ \mathbb{1}\left\{ \left( Z + V^{-1/2} t_a \right)^2 \geq \chi^2_{1,\alpha} \right\} \right] \\
\geq E_Z \left[ \mathbb{1}\left\{ \left( Z + 2(r - \epsilon) \right)^2 \geq \chi^2_{1,\alpha} \right\} \right]
\]

where \( Z \sim \mathcal{N}(0, 1) \), the first inequality follows from (A.1), the first equality follows from (A.2), the second equality follows from the requirements on the mapping \( T_a \) in Definition 1 and \( r_n = o_p(1) \) under \( \{P_{t_a/\sqrt{n}}\}_{n \in \mathbb{N}} \), and the second inequality follows from the definition of \( t_a \) and \( V_a \geq V \). Since \( \epsilon \) can be arbitrary small, we obtain the conclusion.
Proof of (ii). Observe that

\[
\limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_n} E_Q [\psi_{H,n}] = \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_n} \int \left\{ \frac{nT(P_n)^2}{V} \geq \chi^2_{1,\alpha} \right\} dQ^\otimes n
\]

\[
\leq \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_n} \int_{(x_1, \ldots, x_n) \notin \mathcal{X}_n^n} dQ^\otimes n + \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_n} \int_{(x_1, \ldots, x_n) \in \mathcal{X}_n^n} I \left\{ \frac{nT(P_n)^2}{V} \geq \chi^2_{1,\alpha} \right\} dQ^\otimes n
\]

\[= A_1 + A_2,
\]

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

\[
A_1 \leq \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_n} \sum_{i=1}^{n} \int_{x_i \notin \mathcal{X}_n} dQ
\]

\[
\leq \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_n} nm_n^{-\eta} E_Q \sup_{\theta \in \Theta} |g(x, \theta)|^\eta = 0,
\]

where the first inequality follows from a set inclusion relation, the second inequality follows from the Markov inequality, and the equality follows from Assumption 2.1 (vii) and \( E_Q \sup_{\theta \in \Theta} |g(x, \theta)|^\eta < \infty \) for all \( Q \in \mathcal{Q}_n \).

For \( A_2 \), we have

\[
A_2 \leq \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_n} \int \left\{ \left( \sqrt{n} (P_n - \bar{T}(Q)) + \sqrt{n}\bar{T}(Q) \right)^2 \geq \chi^2_{1,\alpha} \right\} dQ^\otimes n
\]

\[
= E_Z \left[ \left\{ \left( Z + 2r \right)^2 \geq \chi^2_{1,\alpha} \right\} \right]
\]

\[
= \Pr \{ W \geq \chi^2_{1,\alpha} \}
\]

where the first inequality follows from the set inclusion relation (\( \mathcal{X}_n^n \) is a subset of support of \( Q^\otimes n \)), and the first equality follows from Lemmas 7.2 and 7.8 of Kitamura, Otsu, and Evdokimov (2009).

A.2. Proof of Theorem 5. Proof of (i). Without loss of generality we assume \( c > 0 \). Suppose the conclusion is false, i.e., there exists a test \( \psi_n \) such that

\[
\limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_n} E_Q [\psi_n] \leq \alpha,
\]

\[
\limsup_{n \to \infty} \inf_{Q \in \mathcal{Q}_n} E_Q [\psi_n] = \Pr \left\{ W_1 \geq \left( z_{\alpha/2} + 2r_0 \right)^2 \right\} + 2\epsilon,
\]

for some \( \epsilon > 0 \). Then there exists a subsequence of \( \mathbb{N} \) such that (but we still use the same subscript \( n \) to simplify the notation)

\[
\limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_n} E_Q [\psi_n] \leq \alpha,
\]

\[
\liminf_{n \to \infty} \sup_{Q \in \mathcal{Q}_n} E_Q [1 - \psi_n] \leq 1 - \Pr \left\{ W_1 \geq \left( z_{\alpha/2} + 2r_0 \right)^2 \right\} - \epsilon,
\]

for some \( \epsilon > 0 \).
Pick any probability measure $P_0 \in \mathcal{P}_0$. Consider a sequence of parametric submodels \( \{P_{t_i}/\sqrt{n}\}_{n \in \mathbb{N}} \) having the Radon-Nikodym density

$$
\frac{dP_{t_i}/\sqrt{n}}{dP_0} = \frac{1 + \zeta'_{t_i,n}g_n \left( x, \frac{c}{\sqrt{n}} \right)}{\int \left( 1 + \zeta'_{t_i,n}g_n \left( x, \frac{c}{\sqrt{n}} \right) \right) dP_0},
$$

where $\zeta_{t,n} = -E_{P_0} \left[ g \left( x, \frac{c}{\sqrt{n}} \right) g_n \left( x, \frac{c}{\sqrt{n}} \right)^t \right]^{-1} E_{P_0} \left[ g \left( x, \frac{c}{\sqrt{n}} \right) \right]$. Note that

$$
P_{c/\sqrt{n}} \in \mathcal{P}_{c/\sqrt{n}},
$$

for each $n \in \mathbb{N}$ because

$$
E_{P_{c/\sqrt{n}}} \left[ g \left( x, \frac{c}{\sqrt{n}} \right) \right] = \frac{\int g \left( x, \frac{c}{\sqrt{n}} \right) dP_0 + \int g \left( x, \frac{c}{\sqrt{n}} \right) g_n \left( x, \frac{c}{\sqrt{n}} \right)^t \int \left( 1 + \zeta'_{t,n}g_n \left( x, \frac{c}{\sqrt{n}} \right) \right) dP_0}{\int \left( 1 + \zeta'_{t,n}g_n \left( x, \frac{c}{\sqrt{n}} \right) \right) dP_0} = 0.
$$

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

$$
nH \left( P_{(t+c)/\sqrt{n}}, P_{c/\sqrt{n}} \right)^2 \rightarrow \frac{1}{4} V^{-1} t^2,
$$

for each $t \in \mathbb{R}$.

Pick any $\delta > 0$. From (A.1) and (A.6), it holds that

$$
P_{(c-t_1)/\sqrt{n}} \in \mathcal{Q}_n,
$$

for all $n$ large enough, where $t_1 = 2V^{1/2}r_1 - \delta$.

From Assumption 2.1, the limiting behavior of the log likelihood ratio $\sum_{i=1}^n \log \frac{dP_{(c-t_1)/\sqrt{n}}}{dP_0} (x_i)$ is obtained as

$$
\begin{align*}
&\sum_{i=1}^n \log \frac{dP_{(c-t_1)/\sqrt{n}}}{dP_0} (x_i) \\
&= \sum_{i=1}^n \log \left( 1 + \zeta'_{c-t_1,n}g_n \left( x_i, \frac{c-t_1}{\sqrt{n}} \right) \right) - n \log \int \left( 1 + \zeta'_{c-t_1,n}g_n \left( x, \frac{c-t_1}{\sqrt{n}} \right) \right) dP_0 \\
&= \zeta'_{c-t_1,n} \sum_{i=1}^n g_n \left( x_i, \frac{c-t_1}{\sqrt{n}} \right) - \frac{1}{2} \zeta'_{c-t_1,n} \sum_{i=1}^n g_n \left( x_i, \frac{c-t_1}{\sqrt{n}} \right)^t g_n \left( x_i, \frac{c-t_1}{\sqrt{n}} \right)^t \zeta_{c-t_1,n} \\
&\quad - n \zeta'_{c-t_1,n} \int g_n \left( x, \frac{c-t_1}{\sqrt{n}} \right) dP_0 \\
&\quad + \frac{n}{2} \zeta'_{c-t_1,n} \int g_n \left( x, \frac{c-t_1}{\sqrt{n}} \right) dP_0 \int g_n \left( x, \frac{c-t_1}{\sqrt{n}} \right)^t dP_0 \zeta_{c-t_1,n} + o_p(1) \\
&= -G'\Omega^{-1} \left\{ \frac{c-t_1}{\sqrt{n}} \sum_{i=1}^n g_n \left( x_i, 0 \right) + \frac{\left( c-t_1 \right)^2}{n} \sum_{i=1}^n G \left( x_i, 0 \right) \right\} + \frac{\left( c-t_1 \right)^2}{2} G'\Omega^{-1} G + o_p(1)
\end{align*}
$$

(A.7) \quad \Rightarrow \tilde{t}_1 Z - \frac{t_1^2}{2}
under \( x \sim P_0 \), where \( Z \sim N(0, I) \) and \( \tilde{t}_1 = cV^{-1/2} - 2r_1 + V^{-1/2}\delta \), the second equality follows from a second-order expansion around \( \zeta_{c-t_1, n} = 0 \), the third equality follows from an expansion around \( c - t_1 = 0 \), and the convergence follows from a central limit theorem and law of large numbers combined with Assumption 2.1.

Combining these results, we obtain

\[
1 - \Pr \left\{ W_1 \geq \left( z_{\alpha/2} + 2r_0 \right)^2 \right\} - \epsilon \geq \liminf_{n \to \infty} \sup_{Q \in \mathcal{Q}_1} E_Q [1 - \psi_n]
\]

\[
\geq \liminf_{n \to \infty} E_{P_{(c-t_1)/\sqrt{n}}} [1 - \psi_n]
\]

\[
= \liminf_{n \to \infty} E_{P_0} \left[ (1 - \psi_n) \prod_{i=1}^n \frac{dP_{(c-t_1)/\sqrt{n}}}{dP_0} (x_i) \right]
\]

\[
= E_{\psi, Z} \left[ (1 - \psi) \exp \left( \tilde{t}_1 Z - \frac{1}{2} \tilde{t}_1^2 \right) \right]
\]

\[
= \frac{1}{\sqrt{2\pi}} \int_z (1 - \psi_0(z)) \exp \left( -\frac{1}{2} (z - \tilde{t}_1)^2 \right) dz
\]

\[
= 1 - E_Z \left[ \psi_0 \left( Z + \tilde{t}_1 \right) \right]
\]

(A.8)

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 follows from (A.4), the second inequality follows from (A.1), the first equality follows from the change of measure, the second equality follows from (A.7), the third equality follows from the law of iterated expectation, and the last equality follows from a change of variables.

By a similar argument and (A.3),

\[
\alpha \geq \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_0} E_Q [\psi_n] \geq \liminf_{n \to \infty} E_{P_{\tilde{t}_0/\sqrt{n}}} [\psi_n] = E_Z \left[ \psi_0 \left( Z + \tilde{t}_0 \right) \right],
\]

(A.9)

where \( \tilde{t}_0 = 2r_0 - V^{-1/2}\delta \) the second inequality follows from (A.6) with setting \( c = 0 \). Combining (A.8) and (A.9), the binary function \( \psi_0(z) \) must satisfy

(i) \( E_Z \left[ \psi_0 \left( Z + \tilde{t}_0 \right) \right] \leq \alpha \) for \( \tilde{t}_0 = 2r_0 - V^{-1/2}\delta \),

(ii) \( E_Z \left[ \psi_0 \left( Z + \tilde{t}_1 \right) \right] \geq \Pr \left\{ W_1 \geq \left( z_{\alpha/2} + 2r_0 \right)^2 \right\} + \epsilon \) for \( \tilde{t}_1 = cV^{-1/2} - 2r_1 + \delta V^{-1/2} \).

Note that \( \Pr \left\{ W_1 \geq \left( z_{\alpha/2} + 2r_0 \right)^2 \right\} \) is power of the Neyman-Pearson test of \( H_0^Z : t = 2r_0 \) versus \( H_1^Z : t = cV^{-1/2} - 2r_1 \) under \( Z \sim N(t, 1) \). Therefore, by taking \( \delta \) arbitrary small, the Neyman-Pearson lemma yields a contradiction.
Proof of (ii). Proof of the first statement. From the same argument to the proof of Part (ii) of Theorem 3,

\[
\limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_{on}} E_Q [\psi_{H,n}]
\leq \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_{on}} E_Q \left[ \left\{ \frac{\sqrt{n}(\bar{T}(P_n) - \bar{T}(Q)) + \sqrt{n}\hat{T}(Q)}{\sqrt{V}} \geq z_{\alpha/2} + 2r_0 \right\} \right]
+ \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_{on}} E_Q \left[ \left\{ \frac{\sqrt{n}(\bar{T}(P_n) - \bar{T}(Q)) + \sqrt{n}\hat{T}(Q)}{\sqrt{V}} \leq -(z_{\alpha/2} + 2r_0) \right\} \right]
= B_1 + B_2.
\]

For \(B_1\), we have

\[
B_1 \leq \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_{on}} E_Q \left[ \left\{ \hat{V}^{-1/2} \sqrt{n} (\bar{T}(P_n) - \bar{T}(Q)) + \hat{V}^{-1/2} \sup_{Q \in \mathcal{Q}_{on}} \left| \frac{\sqrt{n}\hat{T}(Q)}{\sqrt{V/n}} \right| \geq z_{\alpha/2} + 2r_0 \right\} \right]
= E_{\hat{Z}} \left[ \left\{ Z \geq z_{\alpha/2} \right\} \right]
= \frac{\alpha}{2},
\]

where the first equality follows from Lemmas 7.2 and 7.8 of Kitamura, Otsu, and Evdokimov (2009). A similar argument yields \(B_2 \leq \frac{\alpha}{2}\), and we obtain the conclusion.

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

\[
(A.10) \quad \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_{1,n}} E_Q [1 - \psi_{H,n}] \leq 1 - \Pr \left\{ W_1 \geq (z_{\alpha/2} + 2r_0)^2 \right\}.
\]

Observe that

\[
\limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_{1,n}} E_Q [1 - \psi_{H,n}]
\leq \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_{1,n}} E_Q \left[ \left\{ \frac{T(P_n)}{\sqrt{V/n}} \leq (z_{\alpha/2} + 2r_0)^2 \right\} \right]
\leq \limsup_{n \to \infty} \sup_{Q \in \mathcal{Q}_{1,n}} E_Q \left[ \left\{ \hat{V}^{-1/2} \sqrt{n} (\bar{T}(P_n) - \bar{T}(Q)) + \hat{V}^{-1/2} \sup_{Q \in \mathcal{Q}_{1,n}} \left| \frac{\sqrt{n}\hat{T}(Q)}{\sqrt{V/n}} \right| \leq (z_{\alpha/2} + 2r_0)^2 \right\} \right]
= E_{\hat{Z}} \left[ \left\{ (Z + |c| - 2r_1)^2 \leq (z_{\alpha/2} + 2r_0)^2 \right\} \right]
= 1 - \Pr \left\{ W_1 \geq (z_{\alpha/2} + 2r_0)^2 \right\},
\]

where \(Z \sim N(0,1)\), the first inequality follows from the same argument to the proof of Part (ii) of Theorem 3, the second inequality follows from the set inclusion relation, and the first equality follows from Lemmas 7.2 and 7.8 of Kitamura, Otsu, and Evdokimov (2009) and Fisher consistency of \(\bar{T}\).
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