

OPTIMAL TESTING IN A CLASS OF NONREGULAR MODELS

YUYA SHIMIZU AND TAISUKE OTSU

ABSTRACT. This paper studies optimal hypothesis testing for nonregular econometric models with parameter-dependent support. We consider both one-sided and two-sided hypothesis testing and develop asymptotically uniformly most powerful tests based on a limit experiment. Our two-sided test becomes asymptotically uniformly most powerful without imposing further restrictions such as unbiasedness, and can be inverted to construct a confidence set for the nonregular parameter. Simulation results illustrate desirable finite sample properties of the proposed tests.

1. INTRODUCTION

This paper studies optimal hypothesis testing of a class of nonregular econometric models in which the boundary of the support of the observed data depends on some parameter of interest. Such nonregular models, which typically imply discontinuous likelihood functions and nonstandard convergence rates of estimators, have been often studied in the econometrics literature; see, [Flinn and Heckman \(1982\)](#), [Smith \(1985\)](#), [Christensen and Kiefer \(1991\)](#), [Donald and Paarsch \(1993\)](#), [Hong \(1998\)](#), [Donald and Paarsch \(2002\)](#), [Hirano and Porter \(2003\)](#), and [Chernozhukov and Hong \(2004\)](#), among others. In contrast to most existing papers that focus on point estimation, this paper is concerned with optimal (composite) *hypothesis testing* for nonregular models instead of point estimation.

For testing a simple null hypothesis against a simple alternative one, Neyman-Pearson's fundamental lemma yields an optimal power property of the likelihood ratio test even for the case of parameter-dependent support. However, the optimality result is no longer available for general testing problems with composite hypotheses. On the other hand, for regular statistical models, it is known that standard testing methods (such as the likelihood ratio, Wald, and score tests) can achieve certain asymptotic optimal power properties for testing general composite hypotheses (see, Chapter 15 of [Lehmann and Romano, 2022](#)). An open question is whether we can establish an analogous asymptotic optimality result for testing composite hypotheses on nonregular parameters in the case of parameter-dependent support, and this paper addresses this question in a positive way.

In this paper, we consider one-sided and two-sided hypothesis testing for parametric models with parameter-dependent support and develop an asymptotically uniformly most powerful (AUMP) test based on a limit experiment. Interestingly, our two-sided test attains the AUMP

This version: October 4, 2025.

We are grateful to Jack Porter for his valuable comments on the manuscript. We also thank Xiaohong Chen, Harold Chiang, Greg Cox, Bruce Hansen, Ulrich Müller, Kensuke Sakamoto, Xiaoxia Shi, and Kohei Yata for helpful comments.

property without imposing further restrictions such as unbiasedness, and can be inverted to construct a confidence set for the nonregular parameter. For clarity we first present the main results under a benchmark setup in Section 2, where there is no covariate or nuisance parameter. Then we extend our optimality results to a general setup in Section 3 that involves covariates and nuisance parameters. For the general case, we need some independent auxiliary sample to estimate the nuisance parameters, which is typically obtained by sample splitting. Our simulation results in Section 4 illustrate desirable finite sample properties of the proposed tests.

The most closely related papers to ours are [Hirano and Porter \(2003\)](#) and [Chernozhukov and Hong \(2004\)](#). [Hirano and Porter \(2003\)](#) studied efficient point estimation of parameter-dependent support models by extending the limit of experiments argument, and showed that the Bayes estimator is asymptotically efficient under a minimax criterion but the maximum likelihood estimator is generally inefficient. To the best of our knowledge, this paper is the first one that studies optimal hypothesis testing for nonregular models with parameter-dependent support. In contrast to the Bayes estimator in [Hirano and Porter \(2003\)](#) that involves priors on parameters, our optimal testing methods are developed based on the limiting likelihood ratio process without priors. [Chernozhukov and Hong \(2004\)](#) also investigated nonstandard asymptotic properties of estimation and testing methods for parameter-dependent support models. They established asymptotic optimality of the Bayes estimators in terms of the asymptotic average risk, and showed that the Wald test and Bayes posterior quantiles are valid for inference. However, they did not discuss optimality of the testing methods. In addition to these papers, [Chen et al. \(2018, Appendix C\)](#) examined partially identified models with parameter-dependent support, and [Chen et al. \(2025\)](#) adapted their quasi-Bayesian likelihood ratio statistics to partially identified auction models. In our simulation study, we demonstrate that the proposed test exhibits more reasonable performance than the Wald type test by [Chernozhukov and Hong \(2004\)](#).

2. BENCHMARK CASE

This section presents our main results for a simple nonregular model with a scalar parameter and no covariate. This model, covering the uniform distribution as a canonical example, provides a useful benchmark to highlight our developments. We propose one-sided tests for this model and derive the asymptotic properties of these tests, including the AUMP. Based on the one-sided tests, we construct an optimal two-sided test with an unequal-tailed way.

Let $Y \in \mathcal{Y} \subset \mathbb{R}$ follow the parametric model

$$Y \sim f(y|\theta)\mathbb{I}\{y \geq g(\theta)\}, \quad (1)$$

where $\mathbb{I}\{\cdot\}$ is the indicator function, $\theta \in \Theta \subset \mathbb{R}$ is the true value of a scalar parameter, and conditions on the functions f and g are specified below. For this model, we consider the one-sided testing problem

$$H_0^- : \theta \leq \theta_0 \quad \text{against} \quad H_1^- : \theta > \theta_0 \quad (\text{or} \quad H_0^+ : \theta \geq \theta_0 \quad \text{against} \quad H_1^+ : \theta < \theta_0), \quad (2)$$

for a given $\theta_0 \in \Theta$. Hereafter, we focus on the case of $\nabla_{\theta}g(\theta_0) > 0$. The case of $\nabla_{\theta}g(\theta_0) < 0$ is analyzed in the same manner. Based on the reparametrization $h = n(\theta - \theta_0)$, this testing

problem is written as

$$H_0^- : h \leq 0 \quad \text{against} \quad H_1^- : h > 0 \quad (\text{or } H_0^+ : h \geq 0 \quad \text{against} \quad H_1^+ : h < 0). \quad (3)$$

We wish to test H_0^- or H_0^+ based on an independent and identically distributed (iid) sample $Y^n = (Y_1, \dots, Y_n)$ of Y . Although our results can be easily extended to cover a more general reparametrization $h = n(\theta - \theta_0) + \zeta_n$ with $\zeta_n = o(1)$, we focus on the case of $\zeta_n = 0$ for simplicity.

Our testing procedure is constructed based on the limiting likelihood ratio process. Since the joint density of Y^n is written as

$$dP_\theta^n(y^n) = \mathbb{I}\{y_{(1)} \geq g(\theta)\} \prod_{i=1}^n f(y_i|\theta),$$

where $y_{(1)} = \min\{y_1, \dots, y_n\}$, the likelihood ratio process on a parameter space $\mathcal{H} \subset \mathbb{R}$ is

$$Z_n(h, \bar{h}) := \frac{dP_{\theta_0 + \frac{\bar{h}}{n}}^n}{dP_{\theta_0 + \frac{h}{n}}^n}(Y^n) = \frac{\mathbb{I}\{Y_{(1)} \geq g(\theta_0 + \bar{h}/n)\}}{\mathbb{I}\{Y_{(1)} \geq g(\theta_0 + h/n)\}} \prod_{i=1}^n \frac{f(Y_i|\theta_0 + \bar{h}/n)}{f(Y_i|\theta_0 + h/n)},$$

for $h, \bar{h} \in \mathcal{H}$. To characterize asymptotic properties of this process, we impose the following assumptions.

Assumption 1.

- (i): $\{Y_i\}_{i=1}^n$ is an iid sample of $Y \in \mathcal{Y} \subset \mathbb{R}$ with the Lebesgue density in (1). The parameter space Θ is convex.
- (ii): $f(y|\theta)$ is twice continuously differentiable in θ for all y . In some open neighborhood $\mathcal{N}$ of θ_0 , $f(y|\theta)$ and $\nabla_\theta f(y|\theta)$ are continuous in y for $\theta \in \mathcal{N}$, there exists a constant C such that $0 < f(y|\theta) < C < \infty$ for all y and $\theta \in \mathcal{N}$, and

$$\begin{aligned} \int \sup_{\theta \in \mathcal{N}} \|\nabla_\theta f(y|\theta)\| \mathbb{I}\{y \geq g(\theta)\} dy &< \infty, \\ \int \sup_{\theta, \bar{\theta} \in \mathcal{N}} \frac{\|\nabla_\theta f(y|\bar{\theta})\|^2}{f(y|\bar{\theta})^2} \mathbb{I}\{y \geq g(\theta)\} f(y|\theta) dy &< \infty, \\ \int \sup_{\theta, \bar{\theta} \in \mathcal{N}} \frac{\|\nabla_{\theta\theta} f(y|\bar{\theta})\|}{f(y|\bar{\theta})} \mathbb{I}\{y \geq g(\theta)\} f(y|\theta) dy &< \infty. \end{aligned}$$

$g(\theta)$ is continuously differentiable in θ , $\nabla_\theta g(\theta_0) > 0$, and $\sup_{\theta \in \mathcal{N}} \|\nabla_\theta g(\theta)\| < \infty$.

Assumption 1 (i) is standard, and Assumption 1 (ii) contains smoothness and boundedness conditions on the functions f and g in (1). By adapting Hirano and Porter (2003, Theorem 2) to our setup, Assumption 1 guarantees weak convergence (denoted by “ $\rightsquigarrow$ ”) of the likelihood ratio process

$$\{Z_n(h, \bar{h})\}_{\bar{h} \in I} \rightsquigarrow \{Z(h, \bar{h})\}_{\bar{h} \in I} := \{e^{(\bar{h}-h)/\lambda} D_{h, \bar{h}}\}_{\bar{h} \in I} \quad \text{under } P_{\theta_0 + \frac{h}{n}}, \quad (4)$$

for every finite $I \subset \mathcal{H}$, where $\lambda = \{f(g(\theta_0)|\theta_0) \nabla_\theta g(\theta_0)\}^{-1}$ and

$$D_{h, \bar{h}} := \mathbb{I}\{W_h > \bar{h}\} \quad \text{with } W_h \sim f_W(w|h) = \frac{1}{\lambda} e^{-(w-h)/\lambda} \mathbb{I}\{w > h\}.$$

Note that λ is a known constant in the present setup. In contrast to the standard likelihood ratio process, which is locally asymptotically normal, $Z_n(h, \bar{h})$ converges to a limit of experiments

whose randomness is given by the binary variable $D_{h,\bar{h}}$. This is due to lack of differentiability in quadratic mean of the density $dP_\theta^n(y^n)$. Since $D_{h,\bar{h}}$ is discrete, the limiting likelihood ratio process $Z(h, \bar{h})$ is discontinuous in the sense that $\Pr\{Z(h, \bar{h}) \leq z\}$ is not continuous at $z = 0$ and $e^{(\bar{h}-h)/\lambda}$. Thus, we cannot use the conventional asymptotic theory based on the quadratic expansion of the likelihood ratio process to evaluate asymptotic size and power properties of the likelihood ratio test.

Even though the likelihood ratio process $\{Z_n(h, \bar{h})\}_{\bar{h} \in I}$ exhibits such nonregularity, it should be noted the limiting likelihood ratio process $\{Z(h, \bar{h})\}_{\bar{h} \in I}$ satisfies the monotone likelihood ratio property, which is defined as follows.

Definition 1 (Monotone likelihood ratio). (Shao, 2003, Definition 6.2) Suppose that the distribution of W is in $\mathcal{P} = \{P_h : h \in \mathcal{H}\}$, a parametric family indexed by a real-valued h , and that $\mathcal{P}$ is dominated by a σ -finite measure μ . The family $\mathcal{P}$ is said to have *monotone likelihood ratio* in $S(W)$ (a real-valued statistic) if and only if, for any $h_1 < h_2$, $dP_{h_2}/dP_{h_1}(w)$ is a nondecreasing function of $S(w)$ for values w at which at least one of $dP_{h_1}(w)$ and $dP_{h_2}(w)$ is positive.

A key feature of distributions with monotone likelihood ratio is the existence of the uniformly most powerful (UMP) test. Relying on the limiting likelihood ratio process is essential since the finite sample likelihood ratio does not exhibit a monotone likelihood ratio property in general.

Lemma 1. (Shao, 2003, Theorem 6.2.1) Suppose that a random variable U_h has a distribution in $\mathcal{P} = \{P_h : h \in \mathcal{H} \subset \mathbb{R}\}$ that has monotone likelihood ratio in $S(U_h)$. Consider the problem of testing $H_0 : h \leq h_0$ against $H_1 : h > h_0$, where h_0 is a given constant. Then there exists a UMP test of size α given by

$$\phi(U_h) = \begin{cases} 1 & \text{if } S(U_h) > c \\ \kappa & \text{if } S(U_h) = c, \\ 0 & \text{if } S(U_h) < c \end{cases} \quad (5)$$

where c and κ are determined by $E_{h_0}[\phi(U_h)] = \alpha$.

Here $\phi(U_h) = 1$ and 0 mean rejection and acceptance of H_0 , respectively, and $\phi(U_h) = \kappa$ means rejection with probability κ . Therefore, to derive an AUMP test for $H_0^- : h \leq 0$, we can still invoke the asymptotic representation lemma below to argue that the sample counterpart of the monotone likelihood ratio test with $S(W_h) = W_h$ is AUMP.

Lemma 2. (van der Vaart, 2000, Theorem 15.1) Let the sequence of experiments $\mathcal{E}_n = \{P_{n,h} : h \in \mathcal{H}\}$ converge to a dominated experiment $\mathcal{E} = \{P_h : h \in \mathcal{H}\}$. Suppose that a sequence of power functions π_n of tests in $\mathcal{E}_n$ converges pointwise, i.e., $\pi_n(h) \rightarrow \pi(h)$ for every h and some function π . Then π is a power function in the limit experiment, i.e., there exists a test ϕ in $\mathcal{E}$ with $\pi(h) = E_h[\phi(X)]$ for every h .

To derive an AUMP test for the other one-sided test $H_0^+ : h \geq 0$, we utilize the fact that the limit experiment can be represented as a function of the random variable W_h . Accordingly, we apply the Neyman-Pearson lemma (e.g., Theorem 3.2.1 (ii) of Lehmann and Romano, 2022) to W_h .

Hereafter, we first formalize this argument for one-sided tests on $H_0^- : h \leq 0$ and $H_0^+ : h \geq 0$ (Section 2.1), and then extend the argument to two-sided testing for $H_0 : h = 0$ (Section 2.2).

2.1. One-sided tests. First, we consider testing $H_0^- : h \leq 0$ against $H_1^- : h > 0$ at the significance level $\alpha \in (0, 1)$. By taking a sample counterpart of $\phi(W_h)$ in (5) with the limiting process W_h in (4), our test is constructed as

$$\phi_n^-(Y^n) = \begin{cases} 1 & \text{if } Y_{(1)} > g\left(\theta_0 + \frac{\lambda}{n} \log\left(\frac{1}{\alpha}\right)\right) \\ 0 & \text{if } Y_{(1)} \leq g\left(\theta_0 + \frac{\lambda}{n} \log\left(\frac{1}{\alpha}\right)\right) \end{cases}. \quad (6)$$

This test achieves an asymptotic optimal property in Definition 2 below, introduced by Choi et al. (1996).

Definition 2 (Asymptotically uniformly most powerful test). For testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ (or $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$ or $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$), a sequence of tests $\{\phi_n\}$ is called *asymptotically uniformly most powerful (AUMP)* in $\mathcal{H}$ at asymptotic level α if $\limsup_n E_{\theta_0+h/n}[\phi_n] \leq \alpha$ for every $h \leq 0$ (or $h \geq 0$ or $h = 0$) in $\mathcal{H}$ and for any other sequence of test functions $\{\psi_n\}$ satisfying $\limsup_n E_{\theta_0+h/n}[\psi_n] \leq \alpha$ for every $h \leq 0$ (or $h \geq 0$ or $h = 0$) in $\mathcal{H}$,

$$\liminf_n E_{\theta_0+h/n}[\phi_n] \geq \limsup_n E_{\theta_0+h/n}[\psi_n],$$

for every $h > 0$ (or $h < 0$ or $h \neq 0$) in $\mathcal{H}$.

The following asymptotic optimality result is established using the monotone likelihood ratio property of the limit experiment, as demonstrated in Appendix.

Theorem 1. Suppose that Assumption 1 holds for the true local parameter $h \in \mathcal{H}$. Then the test $\phi_n^-(Y^n)$ is AUMP in $\mathcal{H}$ at level α for testing $H_0^- : \theta \leq \theta_0$ against $H_1^- : \theta > \theta_0$.

Next, we consider another one-sided testing $H_0^+ : h \geq 0$ against $H_1^+ : h < 0$. The basic idea is same as the previous case. We propose the following test:

$$\phi_n^+(Y^n) = \begin{cases} 1 & \text{if } Y_{(1)} < \max\left\{g(\theta_0), g\left(\theta_0 + \frac{\lambda}{n} \log\left(\frac{1}{1-\alpha}\right)\right)\right\} \\ 0 & \text{if } Y_{(1)} \geq \max\left\{g(\theta_0), g\left(\theta_0 + \frac{\lambda}{n} \log\left(\frac{1}{1-\alpha}\right)\right)\right\} \end{cases}, \quad (7)$$

and the asymptotic optimality of this test is obtained as follows by applying the Neyman-Pearson lemma to W_h .

Theorem 2. Suppose that Assumption 1 holds for the true local parameter $h \in \mathcal{H}$. Then the test $\phi_n^+(Y^n)$ is AUMP in $\mathcal{H}$ at level α for testing $H_0^+ : \theta \geq \theta_0$ against $H_1^+ : \theta < \theta_0$.

Since we assume $\nabla_\theta g(\theta_0) > 0$ and $\lambda > 0$, we have $g(\theta_0) < g\left(\theta_0 + \frac{\lambda}{n} \log\left(\frac{1}{1-\alpha}\right)\right)$ eventually. However, this inequality can be violated in finite samples. We can show that any test ψ_n^+ rejecting the null with probability one if $Y_{(1)} < g(\theta_0)$ and rejecting the null with probability at most α if $Y_{(1)} \geq g(\theta_0)$ is AUMP at level α for testing $H_0^+ : \theta \geq \theta_0$ against $H_1^+ : \theta < \theta_0$. We recommend using (7) since it avoids randomization.

2.2. Two-sided test. This subsection considers two-sided testing $H_0 : h = 0$ against $H_1 : h \neq 0$. By combining the optimal one-sided tests derived in the last subsection, we propose the following (unequal-tailed) two-sided test:

$$\phi_n(Y^n) = \begin{cases} 1 & \text{if } Y_{(1)} > g(\theta_0 + \frac{\lambda}{n} \log(\frac{1}{\alpha})) \text{ or } Y_{(1)} < g(\theta_0) \\ 0 & \text{if } g(\theta_0) \leq Y_{(1)} \leq g(\theta_0 + \frac{\lambda}{n} \log(\frac{1}{\alpha})) \end{cases}. \quad (8)$$

Indeed, this test is shown to be AUMP.

Theorem 3. *Suppose that Assumption 1 holds for the true local parameter $h \in \mathcal{H}$. Then the proposed test $\phi_n(Y^n)$ is AUMP in $\mathcal{H}$ at level α for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.*

Importantly, the proposed test achieves the AUMP property without imposing additional restrictions such as unbiasedness. This feature is shared by a finite sample UMP two-sided test for the uniform distribution (e.g., Lehmann and Romano, 2022, Problem 3.2, p.105).

Since the two-sided test $\phi_n(Y^n)$ does not randomize, we can easily construct a $100(1 - \alpha)\%$ confidence set by the test inversion:

$$CS = \left\{ \theta \in \Theta : g(\theta) \leq Y_{(1)} \leq g\left(\theta + \frac{\lambda}{n} \log\left(\frac{1}{\alpha}\right)\right) \right\}.$$

This confidence set also has a pointwise optimal property (asymptotically uniformly most accurate).

3. GENERAL CASE

In this section, as in Hirano and Porter (2003), we generalize the benchmark model to accommodate discrete covariates and nuisance parameters:

$$Y|X = x \sim f(y|x, \theta, \gamma) \mathbb{I}\{y \geq g(x, \theta)\}, \quad (9)$$

where $\theta \in \Theta \subset \mathbb{R}$ is a scalar parameter of interest, $\gamma \in \mathbb{R}^d$ is a d -dimensional vector of (regular) nuisance parameters, Y is a scalar dependent variable, and X is an m -dimensional vector of discrete covariates with support $\mathcal{X} = \{a_1, \dots, a_L\}$. This section considers the one-sided testing problem

$$\begin{aligned} H_0^- : \theta \leq \theta_0, \gamma \in \mathbb{R}^d & \text{ against } H_1^- : \theta > \theta_0, \gamma \in \mathbb{R}^d \\ \text{(or } H_0^+ : \theta \geq \theta_0, \gamma \in \mathbb{R}^d & \text{ against } H_1^+ : \theta < \theta_0, \gamma \in \mathbb{R}^d) \end{aligned}$$

for a given $\theta_0 \in \mathbb{R}$ with the asymptotic level of significance α . By reparametrization $h = n(\theta - \theta_0)$, this testing problem is written as

$$\begin{aligned} H_0^- : h \leq 0, \gamma \in \mathbb{R}^d & \text{ against } H_1^- : h > 0, \gamma \in \mathbb{R}^d \\ \text{(or } H_0^+ : h \geq 0, \gamma \in \mathbb{R}^d & \text{ against } H_1^+ : h < 0, \gamma \in \mathbb{R}^d) \end{aligned}$$

Let $(Y^n, X^n) = ((Y_1, \dots, Y_n), (X_1, \dots, X_n))$ be an iid sample of $(Y, X) \in \mathbb{R} \times \mathcal{X}$. To extend our benchmark results in the last section, we consider the plug-in likelihood ratio process

$$Z_n(h, \hat{h}_n, \hat{\gamma}_n) := \prod_{i=1}^n \frac{f(Y_i|X_i, \theta_0 + \hat{h}_n/n, \hat{\gamma}_n) \mathbb{I}\{Y_i \geq g(X_i, \theta_0 + \hat{h}_n/n)\}}{f(Y_i|X_i, \theta_0 + h/n, \hat{\gamma}_n) \mathbb{I}\{Y_i \geq g(X_i, \theta_0 + h/n)\}}, \quad (10)$$

for $h \in \mathcal{H}$ with a parameter space $\mathcal{H} \subset \mathbb{R}$, where $\hat{\gamma}_n$ and $\hat{h}_n$ are some estimators of the nuisance parameters γ and the user-specified alternative value $\bar{h}$, respectively, based on auxiliary data independent from the main sample (Y^n, X^n) . In contrast to the benchmark case, where the threshold for testing takes the form $g(\theta_0 + \bar{h}/n)$, the optimal values of $\bar{h}$ for our test depend on certain population quantities that must be estimated (see Theorems 4 and 5 below). Thus we introduce an estimator $\hat{h}_n$ for $\bar{h}$ in this general case. Typically we split the sample (say, (Y^{2n}, X^{2n})) into the main sample (Y^n, X^n) and auxiliary one $((Y_{n+1}, \dots, Y_{2n}), (X_{n+1}, \dots, X_{2n}))$ to obtain $\hat{\gamma}_n$ and $\hat{h}_n$.

In this section, we impose the following assumptions.

Assumption 2.

- (i): $\{Y_i, X_i\}_{i=1}^n$ is an iid sample of $\mathcal{Y} \times \mathcal{X}$, where $\mathcal{Y} \subset \mathbb{R}$ and $\mathcal{X} = \{a_1, \dots, a_L\}$ with the conditional density in (9). The parameter space $\Theta \times \Gamma$ of (θ, γ) is convex.
- (ii): $\{\hat{h}_n\}$ is a random sequence independent from (Y^n, X^n) satisfying $\sqrt{n}(\hat{h}_n - \bar{h}) = O_{P_{\theta_0 + \frac{\bar{h}}{n}, \gamma}}(1)$ for some user specified constant $\bar{h}$. $\{\hat{\gamma}_n\}$ is a random sequence independent from (Y^n, X^n) satisfying $\sqrt{n}(\hat{\gamma}_n - \gamma) = O_{P_{\theta_0 + \frac{\bar{h}}{n}, \gamma}}(1)$.
- (iii): Let $\beta = (\theta, \gamma)'$ and $\beta_0 = (\theta_0, \gamma)'$. $f(y|x, \beta)$ is twice continuously differentiable in β for all y and x , and $g(x, \theta)$ is continuously differentiable in θ for all x . In some open neighborhood $\mathcal{N}$ of β_0 , $f(y|x, \beta)$ and $\nabla_\beta f(y|x, \beta)$ are continuous in y for $\beta \in \mathcal{N}$, there exists a constant C such that $0 < f(y|x, \beta) < C < \infty$ for all y, x , and $\beta \in \mathcal{N}$, and for each $j = 1, \dots, L$,

$$\begin{aligned} & \int \sup_{\beta \in \mathcal{N}} \|\nabla_\beta f(y | a_j, \beta)\| \mathbb{I}\{y \geq g(a_j, \theta)\} dy < \infty, \\ & \int \sup_{\bar{\beta}, \beta \in \mathcal{N}} \frac{\|\nabla_\beta f(y | a_j, \bar{\beta})\|^2}{f(y | a_j, \bar{\beta})^2} \mathbb{I}\{y \geq g(a_j, \theta)\} f(y | a_j, \beta) dy < \infty, \\ & \int \sup_{\bar{\beta}, \beta \in \mathcal{N}} \frac{\|\nabla_{\beta\beta} f(y | a_j, \bar{\beta})\|}{f(y | a_j, \bar{\beta})} \mathbb{I}\{y \geq g(a_j, \theta)\} f(y | a_j, \beta) dy < \infty. \end{aligned}$$

$\nabla_\theta g(a_j, \theta_0) > 0$, $\sup_{\beta \in \mathcal{N}} \|\nabla_\theta g(a_j, \theta)\| < \infty$, and $\Pr\{X = a_j\} > 0$ for each $j = 1, \dots, L$.

As specified in Assumption 2 (i), we focus on the case where covariates are discrete variables as in Hirano and Porter (2003). Assumption 2 (ii) requires that the nuisance parameters γ and the alternative value $\bar{h}$ to construct our test statistics below can be estimated at the $\sqrt{n}$ rate. For example, based on an auxiliary sample independent from (Y^n, X^n) , γ can be estimated by the maximum likelihood and $\bar{h}$ can be estimated by the method of moments. This assumption enables us to focus on a neighborhood to establish the weak convergence in Lemma 3 below. Since $\hat{h}_n$ and $\hat{\gamma}_n$ are independent from the main sample (Y^n, X^n) , we can argue the weak convergence in a straightforward way even after conditioning on the concentration to the neighborhood. Assumption 2 (iii) lists boundedness and smoothness conditions on the functions f and g .

To present the limiting distribution of the plug-in likelihood ratio process $\{Z_n(h, \hat{h}_n, \hat{\gamma}_n)\}$, we introduce further notations. Define

$$\lambda = \{E_X[f(g(X, \theta_0)|X, \beta_0)\nabla_{\theta}g(X, \theta_0)]\}^{-1},$$

$$D_{h, \bar{h}} = \text{Bernoulli}(\exp(-E_X[\mathbb{I}\{\bar{h} - h \geq 0\}f(g(X, \theta_0)|X, \beta_0)\nabla_{\theta}g(X, \theta_0)(\bar{h} - h)])),$$

$G_j = \nabla_{\theta}g(a_j, \theta_0)$, and $\lambda_j = \{\Pr(X = a_j)f(g(a_j, \theta_0)|a_j, \beta_0)\}^{-1}$. Also let $(W_{h,1}, \dots, W_{h,L})$ be mutually independent random variables that follow

$$W_{h,j} \sim f_{W_j}(w | h, \gamma) = e^{-(w - G_j h)/\lambda_j} \mathbb{I}\{w > G_j h\} / \lambda_j.$$

The weak convergence of the plug-in likelihood ratio process is established as follows.

Lemma 3. *Under Assumption 2, it holds*

$$\{Z_n(h, \hat{h}_n, \hat{\gamma}_n)\}_{\bar{h} \in I} \rightsquigarrow \{Z(h, \bar{h})\}_{\bar{h} \in I} := \{e^{(\bar{h} - h)/\lambda} D_{h, \bar{h}}\}_{\bar{h} \in I} \quad \text{under } P_{\theta_0 + \frac{h}{n}, \gamma}, \quad (11)$$

for every finite $I \subset \mathcal{H}$, where $\lambda = (\sum_{j=1}^L G_j / \lambda_j)^{-1}$ and $D_{h, \bar{h}} = \prod_{j=1}^L \mathbb{I}\{W_{h,j} > G_j \bar{h}\}$. Moreover, we can show the convergences for the components as

$$\prod_{i=1}^n \mathbb{I}\{Y_i \geq g(X_i, \theta_0 + \hat{h}_n/n)\} \rightsquigarrow D_{h, \bar{h}} \quad \text{under } P_{\theta_0 + \frac{h}{n}, \gamma}, \quad (12)$$

and

$$\prod_{i=1}^n \frac{f(Y_i|X_i, \theta_0 + \hat{h}_n/n, \hat{\gamma}_n)}{f(Y_i|X_i, \theta_0 + h/n, \hat{\gamma}_n)} \xrightarrow{p} e^{(\bar{h} - h)/\lambda} \quad \text{under } P_{\theta_0 + \frac{h}{n}, \gamma}, \quad (13)$$

for any $\bar{h} \in \mathcal{H}$.

This lemma is different from the weak convergence in (4) for the benchmark case in the following aspects. First, the process $\{Z_n(h, \hat{h}_n, \hat{\gamma}_n)\}$ contains the estimated parameters $\hat{h}_n$ and $\hat{\gamma}_n$, which also covers the case of deterministic parameter sequences. Second, due to presence of the discrete covariates $X \in \{a_1, \dots, a_L\}$, the limiting process $Z(h, \bar{h})$ involves an L -dimensional random vector $(W_{h,1}, \dots, W_{h,L})$. Third, the limiting process $Z(h, \bar{h})$ depends on the nuisance parameters γ .

Hereafter, we separately consider testing $H_0^- : h \leq 0, \gamma \in \mathbb{R}^d$ against $H_1^- : h > 0, \gamma \in \mathbb{R}^d$ and $H_0^+ : h \geq 0, \gamma \in \mathbb{R}^d$ against $H_1^+ : h < 0, \gamma \in \mathbb{R}^d$. Let $\hat{\lambda}_{nj}$ and $\hat{\lambda}_n$ be consistent estimators of λ_j and λ , respectively. For example, based on an auxiliary sample $\{X_{n+1}, \dots, X_{2n}\}$, λ_j can be estimated by $\hat{\lambda}_{nj} = \{n^{-1} \sum_{i=n+1}^{2n} \mathbb{I}\{X_i = a_j\} f(g(a_j, \theta_0)|a_j, \theta_0, \hat{\gamma}_n)\}^{-1}$, and then λ can be estimated by $\hat{\lambda}_n = (\sum_{j=1}^L G_j / \hat{\lambda}_{nj})^{-1}$.

3.1. One-sided tests. First, we consider one-sided testing $H_0^- : h \leq 0, \gamma \in \mathbb{R}^d$ against $H_1^- : h > 0, \gamma \in \mathbb{R}^d$ at the significance level $\alpha \in (0, 1)$. By extending the one-sided test in (6), our test is defined as

$$\phi_n^-(\hat{h}_n^-, Y^n, X^n) = \begin{cases} 1 & \text{if } Y_i > g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for all } i \\ 0 & \text{if } Y_i \leq g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for some } i \end{cases}, \quad (14)$$

where $\hat{h}_n^- = \hat{\lambda}_n \log(\frac{1}{\alpha})$. The idea to construct this test is essentially the same as the benchmark case in Section 2. The main difference is that λ is unknown and needs to be estimated by

plugging-in the consistent estimator $\hat{\lambda}_n$ that is independent from the data. The next theorem shows that this test achieves an asymptotic optimal property.

Theorem 4. Suppose that Assumption 2 holds for the true local parameter $h \in \mathcal{H}$. Let $\{\hat{h}_n^-\}$ be any random sequence independent from the data such that $\sqrt{n}(\hat{h}_n^- - \bar{h}^-) = O_{P_{\theta_0 + \frac{h}{n}, \gamma}}(1)$ for $\bar{h}^- = \lambda \log\left(\frac{1}{\alpha}\right)$ and $\{\hat{\lambda}_{n1}, \dots, \hat{\lambda}_{nL}\}$ be any random sequence independent from the data such that $\sqrt{n}(\hat{\lambda}_{nj} - \lambda_j) = O_{P_{\theta_0 + \frac{h}{n}, \gamma}}(1)$ for each $j = 1, \dots, L$. Then the test $\phi_n^-(\hat{h}_n^-, Y^n, X^n)$ defined with $\hat{h}_n^-$ and $\hat{\lambda}_n$ constructed by $\{\hat{\lambda}_{n1}, \dots, \hat{\lambda}_{nL}\}$ independent from the data is AUMP in $\mathcal{H}$ at level α for testing H_0^- against H_1^- .

Next, we consider another one-sided testing $H_0^+ : h \geq 0, \gamma \in \mathbb{R}^d$ against $H_1^+ : h < 0, \gamma \in \mathbb{R}^d$. In this case, our test is defined as

$$\phi_n^+(\hat{h}_n^+, Y^n, X^n) = \begin{cases} 1 & \text{if } Y_i < \max\left\{g(X_i, \theta_0), g(X_i, \theta_0 + \hat{h}_n^+/n)\right\} \text{ for some } i \\ 0 & \text{if } Y_i \geq \max\left\{g(X_i, \theta_0), g(X_i, \theta_0 + \hat{h}_n^+/n)\right\} \text{ for all } i \end{cases}, \quad (15)$$

where $\hat{h}_n^+ = \hat{\lambda}_n \log\left(\frac{1}{1-\alpha}\right)$. Similar to Theorem 4, asymptotic optimality of this test is obtained as follows.

Theorem 5. Suppose that Assumption 2 holds for the true local parameter $h \in \mathcal{H}$. Let $\{\hat{h}_n^+\}$ be any random sequence independent from the data such that $\sqrt{n}(\hat{h}_n^+ - \bar{h}^+) = O_{P_{\theta_0 + \frac{h}{n}, \gamma}}(1)$ for $\bar{h}^+ = \lambda \log\left(\frac{1}{1-\alpha}\right)$ and $\{\hat{\lambda}_{n1}, \dots, \hat{\lambda}_{nL}\}$ be any random sequence independent from the data such that $\sqrt{n}(\hat{\lambda}_{nj} - \lambda_j) = O_{P_{\theta_0 + \frac{h}{n}, \gamma}}(1)$ for each $j = 1, \dots, L$. Then the test $\phi_n^+(\hat{h}_n^+, Y^n, X^n)$ defined with $\hat{\lambda}_n$ constructed by $\{\hat{\lambda}_{n1}, \dots, \hat{\lambda}_{nL}\}$ independent from the data is AUMP in $\mathcal{H}$ at level α for testing H_0^+ against H_1^+ .

3.2. Two-sided test. This subsection considers two-sided testing $H_0 : h = 0, \gamma \in \mathbb{R}^d$ against $H_1 : h \neq 0, \gamma \in \mathbb{R}^d$. By combining the optimal one-sided tests derived in the last subsection, we propose the following (unequal-tailed) two-sided test:

$$\phi_n(\hat{h}_n^-, Y^n, X^n) = \begin{cases} 1 & \text{if } Y_i > g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for all } i \text{ or } Y_i < g(X_i, \theta_0) \text{ for some } i \\ 0 & \text{if } Y_i \leq g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for some } i \text{ and } Y_i \geq g(X_i, \theta_0) \text{ for all } i \end{cases}. \quad (16)$$

Theorem 6. Suppose that Assumption 2 holds for the true local parameter $h \in \mathcal{H}$. Let $\{\hat{h}_n^-\}$ and $\{\hat{\lambda}_{n1}, \dots, \hat{\lambda}_{nL}\}$ be any random sequences defined in Theorem 1. Then the test $\phi_n(\hat{h}_n^-, Y^n, X^n)$ defined with $\hat{\lambda}_n$ constructed by $\{\hat{\lambda}_{n1}, \dots, \hat{\lambda}_{nL}\}$ independent from the data is AUMP in $\mathcal{H}$ at level α for testing H_0 against H_1 .

Based on this two-sided test, we can construct an asymptotically optimal $100(1 - \alpha)\%$ confidence set for θ by the test inversion:

$$CS = \left\{ \theta \in \Theta : Y_i \leq g(X_i, \theta + \hat{h}_n^-/n) \text{ for some } i \text{ and } Y_i \geq g(X_i, \theta) \text{ for all } i \right\}.$$

4. SIMULATION

In this section, we investigate the finite-sample performance of the proposed test through a simulation study. We consider one-sided testing problems: $H_0^- : h \leq 0$ against $H_1^- : h > 0$ and $H_0^+ : h \geq 0$ against $H_1^+ : h < 0$. The proposed test, $\phi_n^-(Y^n)$ and $\phi_n^+(Y^n)$, are compared with the Wald test based on the maximum likelihood estimator and the asymptotic distribution derived by Chernozhukov and Hong (2004), hereafter referred to as the CH test.

In the CH test, we reject H_0^- if $n(\hat{\theta} - \theta_0) > q_{1-\alpha}(Z^\theta)$, and reject H_0^+ if $n(\hat{\theta} - \theta_0) < 0$, where $\hat{\theta}$ denotes the maximum likelihood estimator. In our benchmark setup in (1), under $P_{\theta+h/n}$, this estimator satisfies $n(\hat{\theta} - (\theta + h/n)) \rightsquigarrow Z^\theta$, where $q_{1-\alpha}(Z^\theta)$ is the $(1 - \alpha)$ -th quantile of Z^θ , and

$$Z^\theta = \frac{\text{Exp}(1)}{f(g(\theta)|\theta)\nabla_\theta g(\theta)}.$$

As the data generating process, we consider a truncated normal distribution $N(\theta, 1)$ restricted to the range $[\theta - 1.25, \infty)$, i.e., the benchmark model in (1) with $f(y|\theta) = \frac{\phi(y-\theta)}{1-\Phi(-1.25)}$ and $g(\theta) = \theta - 1.25$. We set the significance level at $\alpha = 0.05$ and consider two sample sizes $n \in \{20, 40\}$. The number of Monte Carlo replications is set to 2000.

Figures 1 and 2 present the power curves of the proposed test and CH test, and the power envelope for each one-sided test. The power envelopes, derived in the proofs of Theorems 1 and 2 in Appendix, represent the asymptotically optimal values.¹

Figure 1 shows that the proposed test exhibits reasonable power across a range of values for the true parameter value $h > 0$, performing closer to the power envelope, while the CH test demonstrates lower power. In contrast, Figure 2 shows that the CH test exhibits over-rejection under the null $h = 0$, with size 13.0% for $n = 20$ and 6.2% for $n = 40$. The proposed test controls sizes effectively, with values of 4.4% for $n = 20$ and 4.3% for $n = 40$. These simulation results suggest that the proposed test offers better finite-sample performance.

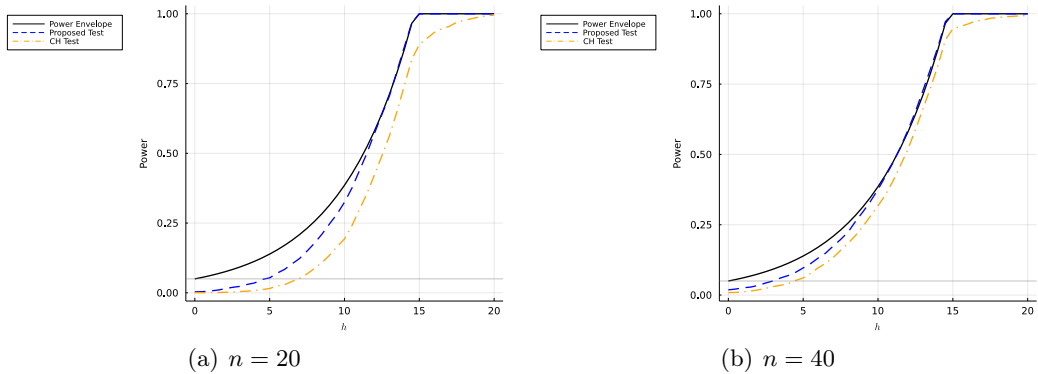


FIGURE 1. Comparison of the proposed test with the CH test (Wald test by Chernozhukov and Hong (2004)) with two sample sizes n for $H_0^- : h \leq 0$ against $H_1^- : h > 0$.

¹The power envelope is given by $E_h[\phi^-(W)]$ for the range $h \geq 0$ and $E_h[\phi^+(W)]$ for the range $h < 0$ using the notation introduced in Appendix.

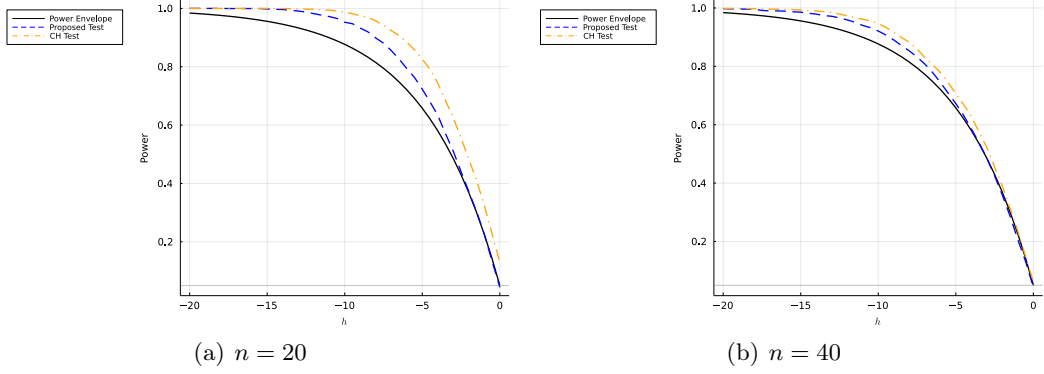


FIGURE 2. Comparison of the proposed test with the CH test (Wald test by Chernozhukov and Hong (2004)) with two sample sizes n for $H_0^+ : h \geq 0$ against $H_1^+ : h < 0$.

APPENDIX A. MATHEMATICAL APPENDIX

Notation: Let $P_h\{\cdot\}$ and $E_h[\cdot]$ be the probability and the expectation under $f_W(w|h)$, respectively.

The following lemma can be shown by adapting the proof of Theorem 2 in Hirano and Porter (2003).

Lemma 4. Suppose that Assumption 1 holds true with any local parameter $h \in \mathcal{H}$. Then

$$\mathbb{I}\left\{Y_{(1)} \geq g\left(\theta_0 + \frac{\bar{h}}{n}\right)\right\} \rightsquigarrow D_{h,\bar{h}} \quad \text{under } P_{\theta_0 + \frac{h}{n}}, \quad (17)$$

and

$$\prod_{i=1}^n \frac{f(Y_i|\theta_0 + \bar{h}/n)}{f(Y_i|\theta_0 + h/n)} \xrightarrow{p} e^{(\bar{h}-h)/\lambda} \quad \text{under } P_{\theta_0 + \frac{h}{n}}, \quad (18)$$

for any $\bar{h} \in \mathcal{H}$.

A.1. Proof of Theorem 1.

Step 1: Derive the UMP test in the limit of experiments. By the definition, $D_{h,h_1} = 0$ implies $D_{h,h_2} = 0$ for any h and $h_1 < h_2$. Thus, if at least one of $dP_{h_1}(w)$ and $dP_{h_2}(w)$ is positive, then we must have $D_{h,h_1} = 1$. Under $D_{h,h_1} = 1$, the likelihood ratio $Z(h_2, h)/Z(h_1, h)$ is well-defined and a nondecreasing function of W_h . Therefore, $\mathcal{P} = \{Z(\bar{h}, h) : \bar{h} \in \mathcal{H}\}$ has monotone likelihood ratio in $W_h \sim f_W(w|h) = \frac{1}{\lambda} e^{-(w-h)/\lambda} \mathbb{I}\{w > h\}$ under $P_{\theta_0 + \frac{h}{n}}$.

Let $\bar{h}^- = \lambda \log(1/\alpha)$. Observe that for any $\bar{h}$, we have

$$P_h\{W_h > \bar{h}\} = \frac{1}{\lambda} \int_{\max\{\bar{h}, h\}}^{\infty} e^{-(w-h)/\lambda} dw = \min\{e^{(h-\bar{h})/\lambda}, 1\},$$

which implies $P_0\{W_0 > \bar{h}^-\} = \alpha$. Combining this with Lemma 1 implies that the test $\phi^-(W_h) := \mathbb{I}\{W_h > \bar{h}^-\}$ is UMP for testing $H_0 : h \leq 0$ against $H_1 : h > 0$ in the limit of experiments. Then the proposed test $\phi_n^-(Y^n)$ in (6) emerges as its sample counterpart.

Step 2: Show size control of $\phi_n^-(Y^n)$. Pick any $h \leq 0$. Since the weak convergence in Lemma 4 implies

$$\lim_n E_{\theta_0 + \frac{h}{n}}[\phi_n^-(Y^n)] = E_h[\phi^-(W_h)],$$

we obviously have

$$\lim_n E_{\theta_0 + \frac{h}{n}}[\phi_n^-(Y^n)] = \alpha e^{h/\lambda} \leq \alpha,$$

i.e., $\phi_n^-(Y^n)$ achieves the asymptotic size control.

Step 3: Show power optimality of $\phi_n^-(Y^n)$. Pick any sequence of tests $\{\psi_n\}$ satisfying $\limsup_n E_{\theta_0}[\psi_n] \leq \alpha$, and pick any $\bar{h} > 0$. Consider a subsequence $\{\psi_m\}$ of $\{\psi_n\}$ such that

$$\begin{aligned} \lim_m E_{\theta_0}[\psi_m] &\leq \limsup_n E_{\theta_0}[\psi_n], \\ \lim_m E_{\theta_0 + \frac{\bar{h}}{m}}[\psi_m] &= \limsup_n E_{\theta_0 + \frac{\bar{h}}{m}}[\psi_n]. \end{aligned}$$

By Lemma 2, if a sequence of power functions of tests $\{E_{\theta_0 + \frac{h}{m}}[\psi_m]\}$ pointwise converges to some function (denoted by $\pi_\psi(h)$) for $h \in \{0, \bar{h}\}$, then $\pi_\psi(h)$ is a power function for testing $H_0 : h = 0$ against $H_1 : h = \bar{h}$ in one sample $W_h \sim f_W(w|h)$ for $h \in \{0, \bar{h}\}$. On the other hand, Neyman-Pearson lemma (e.g., Theorem 3.2.1 (ii) of Lehmann and Romano, 2022) implies that $\phi^-(W_h)$ is the most powerful for this one sample testing problem. Therefore, we have $E_{\bar{h}}[\phi^-(W_h)] \geq \pi_\psi(\bar{h})$. Repeating the argument for every $\bar{h}$, the conclusion is obtained.

A.2. Proof of Theorem 2.

Step 1: Derive the UMP test in the limit of experiments. Consider the testing problem $H_0 : h = 0$ against $H_1 : h = \bar{h}$ for any fixed $\bar{h} < 0$ in the limit of experiments. Define $a(k) = P_0\{f_W(W_0|\bar{h}) > k f_W(W_0|0)\}$, which is nonincreasing. Pick any $\alpha \in (0, 1)$ and let k_0 be such that $a(k_0) \leq \alpha \leq a(k_0 - 0)$. Then the Neyman-Pearson test is defined as

$$\phi^+(W_h) = \begin{cases} 1 & \text{if } f_W(W_h|\bar{h}) > k_0 f_W(W_h|0) \\ \frac{\alpha - a(k_0)}{a(k_0 - 0) - a(k_0)} & \text{if } f_W(W_h|\bar{h}) = k_0 f_W(W_h|0) \\ 0 & \text{if } f_W(W_h|\bar{h}) < k_0 f_W(W_h|0) \end{cases},$$

with $E_0[\phi^+(W_0)] = \alpha$. By Neyman-Pearson lemma, this test is most powerful. We can compute k_0 and $a(k_0)$ from $E_0[\phi^+(W)] = \alpha$ and show that

$$\begin{aligned} \phi^+(W_h) &= \begin{cases} 1 & \text{if } f_W(W_h|\bar{h}) > e^{\bar{h}/\lambda} f_W(W_h|0) \\ \alpha & \text{if } f_W(W_h|\bar{h}) = e^{\bar{h}/\lambda} f_W(W_h|0) \\ 0 & \text{if } f_W(W_h|\bar{h}) < e^{\bar{h}/\lambda} f_W(W_h|0) \end{cases} = \begin{cases} 1 & \text{if } \mathbb{I}\{W_h > \bar{h}\} > \mathbb{I}\{W_h > 0\} \\ \alpha & \text{if } \mathbb{I}\{W_h > \bar{h}\} = \mathbb{I}\{W_h > 0\} \\ 0 & \text{if } \mathbb{I}\{W_h > \bar{h}\} < \mathbb{I}\{W_h > 0\} \end{cases} \\ &= \begin{cases} 1 & \text{if } W_h \leq 0 \\ \alpha & \text{if } W_h > 0 \end{cases} = (1 - D_{h,0}) + \alpha D_{h,0}, \end{aligned}$$

is most powerful, where the second equality follows from the definition of f_W , and the third equality follows from the fact that $\mathbb{I}\{W_h > 0\} = 1$ implies $\mathbb{I}\{W_h > \bar{h}\} = 1$ for any $\bar{h} < 0$.

Note that $\phi^+(\cdot)$ does not depend on the alternative $\bar{h}$. Thus, it is UMP for $H_0 : h = 0$ against $H_1 : h < 0$. Moreover, since $E_h[\phi^+(W_h)] = \alpha$ for any $h > 0$, it is also UMP for $H_0 : h \geq 0$ against $H_1 : h < 0$.

Step 2: Show size control of $\phi_n^+(Y^n)$. Let $\bar{h}^+ = \lambda \log\left(\frac{1}{1-\alpha}\right)$. Observe that

$$\begin{aligned} \lim_n E_{\theta_0 + \frac{h}{n}}[\phi_n^+(Y^n)] &= \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_{(1)} < \max \left\{ g(\theta_0), g\left(\theta_0 + \frac{\bar{h}^+}{n}\right) \right\} \right\} \\ &= \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_{(1)} < g\left(\theta_0 + \frac{\bar{h}^+}{n}\right) \right\} = P_h\{W_h \leq \bar{h}^+\} \\ &= 1 - \min\{(1-\alpha)e^{h/\lambda}, 1\} = \max\{1 - (1-\alpha)e^{h/\lambda}, 0\}, \end{aligned}$$

where the second equality follows by $g(\theta_0) < g\left(\theta_0 + \frac{\bar{h}^+}{n}\right)$ eventually, and the third equality follows from the weak convergence in Lemma 4. Thus, we have $\lim_n E_{\theta_0 + \frac{h}{n}}[\phi_n^+(Y^n)] = \max\{1 - (1-\alpha)e^{h/\lambda}, 0\} \leq \alpha$ for any $h \geq 0$.

Step 3: Show power optimality of $\phi_n^+(Y^n)$. We can derive

$$E_h[\phi^+(W_h)] = 1 - \min\{e^{h/\lambda}, 1\} + \alpha \min\{e^{h/\lambda}, 1\} = \max\{1 - (1-\alpha)e^{h/\lambda}, \alpha\},$$

and

$$\lim_n E_{\theta_0 + \frac{h}{n}}[\phi_n^+(Y^n)] = E_h[\phi^+(W_h)],$$

for each $h < 0$. Therefore, AUMP of $\phi_n^+(Y^n)$ follows from the same argument to show Theorem 1.

A.3. Proof of Theorem 3.

Step 1: Show size control of $\phi_n(Y^n)$. Let $\bar{h}^- = \lambda \log\left(\frac{1}{\alpha}\right)$. For $h = 0$, we have

$$\begin{aligned} \lim_n E_{\theta_0}[\phi_n(Y^n)] &= \lim_n P_{\theta_0} \left\{ Y_{(1)} > g\left(\theta_0 + \frac{\bar{h}^-}{n}\right) \text{ or } Y_{(1)} < g(\theta_0) \right\} \\ &\leq \lim_n P_{\theta_0} \left\{ Y_{(1)} > g\left(\theta_0 + \frac{\bar{h}^-}{n}\right) \right\} + \lim_n P_{\theta_0} \{Y_{(1)} < g(\theta_0)\} \\ &= P_0\{W_0 > \bar{h}^-\} + P_0\{W_0 \leq 0\} = \min\{e^{\bar{h}^-/\lambda}, 1\} + 0 \\ &= \alpha, \end{aligned}$$

where the inequality follows from the union bound, and the second equality follows from the weak convergence in Lemma 4. Thus, $\phi_n(Y^n)$ controls the asymptotic size.

Step 2: Show power optimality of $\phi_n(Y^n)$. For any $h \neq 0$, note that

$$\begin{aligned} &\lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_{(1)} > g\left(\theta_0 + \frac{\bar{h}^-}{n}\right) \text{ or } Y_{(1)} < g(\theta_0) \right\} \\ &= 1 - \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ g(\theta_0) \leq Y_{(1)} \leq g\left(\theta_0 + \frac{\bar{h}^-}{n}\right) \right\} \\ &= 1 - \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_{(1)} \leq g\left(\theta_0 + \frac{\bar{h}^-}{n}\right) \right\} + \lim_n P_{\theta_0 + \frac{h}{n}} \{Y_{(1)} < g(\theta_0)\} \\ &= P_h\{W_h > \bar{h}^-\} + P_h\{W_h \leq 0\}, \end{aligned}$$

where the first and second equalities follow from the set relationship, and the third equality follows from the weak convergence in Lemma 4. For $h > 0$, it holds

$$\lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_{(1)} > g \left(\theta_0 + \frac{\bar{h}^-}{n} \right) \text{ or } Y_{(1)} < g(\theta_0) \right\} = P_h \{W_h > \bar{h}^-\} + 0 = E_h[\phi^-(W)].$$

On the other hand, for $h < 0$, it holds

$$\lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_{(1)} > g \left(\theta_0 + \frac{\bar{h}^-}{n} \right) \text{ or } Y_{(1)} < g(\theta_0) \right\} = \alpha e^{h/\lambda} + (1 - e^{h/\lambda}) = E_h[\phi^+(W_h)].$$

Thus, the AUMP follows from the same argument to show Theorem 1.

A.4. Proof of Lemma 3. Hereafter, let $P_h\{\cdot\}$ and $E_h[\cdot]$ be the probability and the expectation under $f_W(w|h) \equiv \prod_{j=1}^L f_{W_j}(w_j|h, \gamma)$. Define $P_{h,j}\{\cdot\}$ and $E_{h,j}[\cdot]$ are the ones under $f_{W_j}(w_j|h, \gamma)$.

It suffices to show (12) and (13) since (11) follows from the Slutsky lemma (recall that $\prod_{i=1}^n \mathbb{I}\{Y_i \geq g(X_i, \theta_0 + h/n)\} = 1$ under $P_{\theta_0 + \frac{h}{n}, \gamma}$). Pick any $h, \bar{h} \in I$. Let $\delta_n = n^{-1/4}$. Since $P_{\theta_0 + \frac{h}{n}, \gamma} \{|\hat{h}_n - \bar{h}| + \|\hat{\gamma}_n - \gamma\| > \delta_n\} \rightarrow 0$, it suffices to show (12) and (13) under $P_{\theta_0 + \frac{h}{n}, \gamma}$ conditional on $|\hat{h}_n - \bar{h}| + \|\hat{\gamma}_n - \gamma\| \leq \delta_n$. Define

$$\begin{aligned} D_n &= \prod_{i=1}^n \mathbb{I}\{Y_i \geq g(X_i, \theta_0 + \hat{h}_n/n)\}, \\ R_n &= \sum_{i=1}^n \{\log f(Y_i|X_i, \theta_0 + \hat{h}_n/n, \hat{\gamma}_n) - \log f(Y_i|X_i, \theta_0 + h/n, \hat{\gamma}_n)\}. \end{aligned}$$

We first show (13) by proving $R_n \xrightarrow{P} (\bar{h} - h)/\lambda$. Expand R_n as

$$\begin{aligned} R_n &= \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} \log f(Y_i|X_i, \theta_0 + \tilde{h}_n/n, \hat{\gamma}_n)(\hat{h}_n - h) \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \nabla_{\theta} \log f(Y_i|X_i, \theta_0, \gamma) + \frac{\tilde{h}_n}{n} \nabla_{\theta\theta} \log f(Y_i|X_i, \tilde{\theta}_n, \tilde{\gamma}_n) \right. \\ &\quad \left. + \sqrt{n}(\hat{\gamma}_n - \gamma)' \frac{1}{\sqrt{n}} \nabla_{\gamma\theta} \log f(Y_i|X_i, \tilde{\theta}_n, \tilde{\gamma}_n) \right\} (\hat{h}_n - h), \end{aligned}$$

where the first equality follows from an expansion of $\log f(Y_i | X_i, \theta_0 + \hat{h}_n/n, \hat{\gamma}_n)$ around $\theta_0 + \hat{h}_n/n = \theta_0 + h/n$ (with $\tilde{h}_n$ being between $\hat{h}_n$ and h), and the second equality follows from an expansion of $\nabla_{\theta} \log f(Y_i|X_i, \theta_0 + \tilde{h}_n/n, \hat{\gamma}_n)$ around $(\theta_0 + \tilde{h}_n/n, \hat{\gamma}_n) = (\theta_0, \gamma)$ (with $\tilde{\theta}_n$ being between $\theta_0 + \tilde{h}_n/n$ and θ_0 , and $\tilde{\gamma}_n$ being between $\hat{\gamma}_n$ and γ). Since we are conditioning on the event $|\hat{h}_n - \bar{h}| + \|\hat{\gamma}_n - \gamma\| \leq \delta_n$, we have

$$|\tilde{\theta}_n - \theta_0| \leq |\tilde{h}_n|/n \leq (|\hat{h}_n - \bar{h}| + |\bar{h}| + |h|)/n \leq n^{-5/4} + |\bar{h}|n^{-1} + |h|n^{-1},$$

and $\|\tilde{\gamma} - \gamma\| \leq n^{-1/4}$ almost surely. Thus, Assumption 2 (iii) and Markov's inequality imply that $\frac{1}{n^2} \sum_{i=1}^n \nabla_{\theta\theta} \log f(Y_i|X_i, \tilde{\theta}_n, \tilde{\gamma}) = o_{P_{\theta_0 + \frac{h}{n}, \gamma}}(1)$ and $\frac{1}{n^{3/2}} \sum_{i=1}^n \nabla_{\gamma\theta} \log f(Y_i|X_i, \tilde{\theta}_n, \tilde{\gamma}) = o_{P_{\theta_0 + \frac{h}{n}, \gamma}}(1)$. Combining these results, we have

$$R_n = \frac{1}{n} \sum_{i=1}^n \nabla_{\theta} \log f(Y_i|X_i, \theta_0, \gamma)(\hat{h}_n - h) + o_{P_{\theta_0 + \frac{h}{n}, \gamma}}(1) \xrightarrow{P} E_h[\nabla_{\theta} \log f(Y|X, \beta_0)](\bar{h} - h),$$

where the convergence follows from the weak law of large numbers. Also note that

$$\begin{aligned}
E_h[\nabla_\theta \log f(Y|X, \beta_0)] &= E_X \left[\int_{g(X, \theta_0)}^\infty \nabla_\theta f(y|X, \beta_0) dy \right] \\
&= E_X \left[\nabla_\theta \int_{g(X, \theta_0)}^\infty f(y|X, \beta_0) dy + f(g(X, \theta_0)|X, \beta_0) \nabla_\theta g(X, \theta_0) \right] \\
&= E_X[f(g(X, \theta_0)|X, \beta_0) \nabla_\theta g(X, \theta_0)] = \lambda^{-1},
\end{aligned}$$

where the first equality follows from the law of iterated expectations and (9), the second equality follows from Assumption 2 (iii) and the Leibniz integral rule, and the third equality follows from $E_X \left[\int_{g(X, \theta_0)}^\infty f(y|X, \beta_0) dy \right] = 1$. From the continuous mapping theorem, (13) holds true.

Next, we show that $D_n \rightsquigarrow D_{h, \bar{h}}$ under $P_{\theta_0 + \frac{h}{n}, \gamma}$. Let $D_n^- = \prod_{i=1}^n \mathbb{I}\{Y_i \geq g(X_i, \theta_0 + \bar{h}/n + n^{-5/4})\}$ and $D_n^+ = \prod_{i=1}^n \mathbb{I}\{Y_i \geq g(X_i, \theta_0 + \bar{h}/n - n^{-5/4})\}$. Under $P_{\theta_0 + \frac{h}{n}, \gamma}$ conditional on $|\hat{h}_n - \bar{h}| + \|\hat{\gamma}_n - \gamma\| \leq \delta_n$, we have

$$D_n^- \leq D_n \leq D_n^+ \quad \text{eventually,}$$

where the inequalities hold true eventually by the continuity assumption of $\nabla_\theta g(X_i, \theta_0) > 0$. Let $\bar{\theta}_n = \theta_0 + \bar{h}/n + n^{-5/4}$. Under $P_{\theta_0 + \frac{h}{n}, \gamma}$, we have

$$D_n^- = \prod_{i=1}^n \mathbb{I}\{Y_i \geq g(X_i, \bar{\theta}_n)\} \rightsquigarrow \text{Bernoulli} \left(\lim_{n \rightarrow \infty} E_{\theta_0 + \frac{h}{n}, \gamma} \left[\prod_{i=1}^n \mathbb{I}\{Y_i \geq g(X_i, \bar{\theta}_n)\} \right] \right).$$

We now compute $\lim_{n \rightarrow \infty} E_{\theta_0 + \frac{h}{n}, \gamma} [\prod_{i=1}^n \mathbb{I}\{Y_i \geq g(X_i, \bar{\theta}_n)\}]$. Note that for some $\dot{h}_n$ such that $g(x, \theta_0 + \dot{h}_n/n)$ is between $g(x, \theta_0 + \bar{h}/n + n^{-5/4})$ and $g(x, \theta_0 + h/n)$, and some $\tilde{\theta}_n$ is between $\theta_0 + \bar{h}/n + n^{-5/4}$ and $\theta_0 + h/n$, it holds

$$\begin{aligned}
E_{\theta_0 + \frac{h}{n}, \gamma}[\mathbb{I}\{Y \geq g(X, \bar{\theta}_n)\}] &= 1 - E_X[F_{Y|X, \theta_0 + \frac{h}{n}, \gamma}(g(X, \bar{\theta}_n))] \\
&= 1 - E_X \left[\int_{-\infty}^{g(X, \bar{\theta}_n)} f(y|X, \theta_0 + h/n, \gamma) \mathbb{I}\{y - g(X, \theta_0 + h/n) \geq 0\} dy \right] \\
&= 1 - E_X \left[\mathbb{I}\{g(X, \bar{\theta}_n) - g(X, \theta_0 + h/n) \geq 0\} \int_{g(X, \theta_0 + h/n)}^{g(X, \bar{\theta}_n)} f(y|X, \theta_0 + h/n, \gamma) dy \right] \\
&= 1 - E_X \left[\begin{aligned} &\mathbb{I}\{g(X, \bar{\theta}_n) - g(X, \theta_0 + h/n) \geq 0\} \\ &\times f(g(X, \theta_0 + \dot{h}_n/n)|X, \theta_0 + h/n, \gamma) \{g(X, \bar{\theta}_n) - g(X, \theta_0 + h/n)\} \end{aligned} \right] \\
&= 1 - E_X \left[\begin{aligned} &\mathbb{I}\{\nabla_\theta g(X, \tilde{\theta}_n)(\bar{h} + n^{-1/4} - h) \geq 0\} \\ &\times f(g(X, \theta_0 + \dot{h}_n/n)|X, \theta_0 + h/n, \gamma) \nabla_\theta g(X, \tilde{\theta}_n)(\bar{h} + n^{-1/4} - h)/n \end{aligned} \right]
\end{aligned}$$

where the second inequality follows from the model (9), the fourth equality follows from an expansion of $\int_{g(X, \theta_0 + h/n)}^{g(X, \bar{\theta}_n)} f(y|X, \theta_0 + h/n, \gamma) dy$ around $g(x, \bar{\theta}_n) = g(X, \theta_0 + h/n)$, and the fifth equality follows from an expansion of $g(x, \bar{\theta}_n)$ around $\bar{\theta}_n = \theta_0 + h/n$. Thus, recalling that $\bar{\theta}_n$ and

$\tilde{\theta}_n$ are non-random, we have

$$\begin{aligned}
& E_{\theta_0 + \frac{h}{n}, \gamma} \left[\prod_{i=1}^n \mathbb{I}\{Y_i \geq g(X_i, \tilde{\theta}_n)\} \right] = [P_{\theta_0 + \frac{h}{n}, \gamma}\{Y \geq g(X, \tilde{\theta}_n)\}]^n \\
&= \left[1 - \frac{1}{n} E_X \left[\begin{aligned} & \mathbb{I}\{\nabla_{\theta} g(X, \tilde{\theta}_n)(\bar{h} + n^{-1/4} - h) \geq 0\} \\ & \times f(g(X, \theta_0 + \dot{h}_n/n) | X, \theta_0 + h/n, \gamma) \nabla_{\theta} g(X, \tilde{\theta}_n)(\bar{h} + n^{-1/4} - h) \end{aligned} \right] \right]^n \\
&\rightarrow \exp \left(\lim_{n \rightarrow \infty} -E_X \left[\begin{aligned} & \mathbb{I}\{\nabla_{\theta} g(X, \tilde{\theta}_n)(\bar{h} + n^{-1/4} - h) \geq 0\} \\ & \times f(g(X, \theta_0 + \dot{h}_n/n) | X, \theta_0 + h/n, \gamma) \nabla_{\theta} g(X, \tilde{\theta}_n)(\bar{h} + n^{-1/4} - h) \end{aligned} \right] \right) \\
&= \exp \left(-E_X \left[\lim_{n \rightarrow \infty} \left[\begin{aligned} & \mathbb{I}\{\nabla_{\theta} g(X, \tilde{\theta}_n)(\bar{h} + n^{-1/4} - h) \geq 0\} \\ & \times f(g(X, \theta_0 + \dot{h}_n/n) | X, \theta_0 + h/n, \gamma) \nabla_{\theta} g(X, \tilde{\theta}_n)(\bar{h} + n^{-1/4} - h) \end{aligned} \right] \right] \right) \\
&= \exp \left(-E_X \left[\lim_{n \rightarrow \infty} \left[\begin{aligned} & \mathbb{I}\{\nabla_{\theta} g(X, \tilde{\theta}_n)(\bar{h} + n^{-1/4} - h) \geq 0\} \\ & \times f(g(X, \theta_0) | X, \theta_0, \gamma) \nabla_{\theta} g(X, \theta_0)(\bar{h} - h) \end{aligned} \right] \right] \right),
\end{aligned}$$

where the first equality follows from the iid assumption, the second equality follows from the calculation above, the third equality follows from the dominated convergence theorem with the uniform boundedness of $f(y|x, \theta, \gamma)$ and $\sup_{\beta \in \mathcal{N}} \|\nabla_{\theta} g(x, \theta)\| < \infty$, and the last equality follows from the continuity of $f(y|x, \theta, \gamma)$ in y and θ and the continuity of $g(x, \theta)$ and $\nabla_{\theta} g(x, \theta)$ in θ . Since the assumption $\nabla_{\theta} g(a_j, \theta_0) > 0$ guarantees

$$\begin{aligned}
& \lim_n \mathbb{I}\{\nabla_{\theta} g(a_j, \tilde{\theta}_n)(\bar{h} + n^{-1/4} - h) \geq 0\} \\
&= \lim_n \mathbb{I}\{\bar{h} + n^{-1/4} - h \geq 0\} = \mathbb{I} \left[\bigcap_{n=1}^{\infty} \{\bar{h} - h \geq -n^{-1/4}\} \right] = \mathbb{I}\{\bar{h} - h \geq 0\},
\end{aligned}$$

for each $j = 1, \dots, L$, we obtain

$$\begin{aligned}
D_n^- &\rightsquigarrow \text{Bernoulli}(\exp(-E_X[\mathbb{I}\{\bar{h} - h \geq 0\} f(g(X, \theta_0) | X, \theta_0, \gamma) \nabla_{\theta} g(X, \theta_0)(\bar{h} - h)])) \\
&= D_{h, \bar{h}}.
\end{aligned}$$

Similarly, we obtain $D_n^+ \rightsquigarrow D_{h, \bar{h}}$. Combining these results yields (12). Therefore, the conclusion is obtained.

A.5. Proof of Theorem 4.

Step 1: Derive the UMP test in the limit of experiments. The proof of this part is similar to the one for Theorem 1 after replacing $Z(h, \bar{h})$, $D_{h, \bar{h}}$, and λ in Section 2 with the ones in Section 3 and redefining $f_W(w|h) := \prod_{j=1}^L f_{W_j}(w_j|h, \gamma)$. The only differences are the followings. The likelihood ratio $Z(h_2, h)/Z(h_1, h)$ is a nondecreasing function of $\min_{j \in \{1, \dots, L\}} W_{h,j}/G_j$ since $G_j > 0$ for each j . Observe that

$$P_{h,j}\{W_{h,j} > G_j \bar{h}\} = \frac{1}{\lambda_j} \int_{\max\{G_j \bar{h}, G_j h\}}^{\infty} e^{-(w_j - G_j h)/\lambda_j} dw_j = \exp \left(\min \left\{ \frac{G_j(h - \bar{h})}{\lambda_j}, 0 \right\} \right).$$

Hence,

$$\begin{aligned}
& P_h \left\{ \min_{j \in \{1, \dots, L\}} W_{0,j}/G_j > \bar{h} \right\} \\
&= \prod_{j=1}^L P_{h,j} \{W_{h,j} > G_j \bar{h}\} = \prod_{j=1}^L \exp \left(\min \left\{ \frac{G_j(h - \bar{h})}{\lambda_j}, 0 \right\} \right) \\
&= \exp \left(\min \left\{ \sum_{j=1}^L \frac{G_j(h - \bar{h})}{\lambda_j}, 0 \right\} \right) = \min \left\{ \exp \left(\frac{h - \bar{h}}{\lambda} \right), 1 \right\},
\end{aligned}$$

where the first equality holds since $(W_{h,1}, \dots, W_{h,L})$ are mutually independent, the second equality follows from the above observation, the third equality from $G_j > 0$ and $\lambda_j > 0$ for any j , and the last equality from $\lambda = (\sum_{j=1}^L G_j/\lambda_j)^{-1}$. Thus, $P_0\{\min_{j \in \{1, \dots, L\}} W_{0,j}/G_j > \bar{h}^-\} = \alpha$ since $\bar{h}^- = \lambda \log(1/\alpha)$. Combining this with Lemma 1 implies that the test

$$\phi^-(W_h) := \mathbb{I} \left\{ \min_{j \in \{1, \dots, L\}} W_{0,j}/G_j > \bar{h}^- \right\} = D_{h, \bar{h}^-}$$

is UMP for $H_0 : h \leq 0$ versus $H_1 : h > 0$ in the limit of experiments experiment. Then the proposed test $\phi_n^-(Y^n)$ in (14) emerges as its sample counterpart.

Step 2: Show size control of $\phi_n^-(\hat{h}_n^-, Y^n, X^n)$. Since the weak convergence in Lemma 3 implies

$$\lim_n E_{\theta_0 + \frac{h}{n}} [\phi_n^-(\hat{h}_n^-, Y^n, X^n)] = E_h[\phi^-(W)],$$

we obtain the asymptotic size control

$$\lim_n E_{\theta_0 + \frac{h}{n}} [\phi_n^-(\hat{h}_n^-, Y^n, X^n)] = \alpha e^{h/\lambda} \leq \alpha,$$

for any $h \leq 0$.

Step 3: Show power optimality of $\phi_n^-(\hat{h}_n^-, Y^n, X^n)$. This step follows from the same argument to show Theorem 1. Therefore, the conclusion is obtained.

A.6. Proof of Theorem 5.

Step 1: Derive the UMP test in the limit of experiments. The proof of this part is similar to the one for Theorem 2 after replacing $D_{h, \bar{h}}$ and λ in Section 2 with the ones in Section 3 and redefining $f_W(w|h) := \prod_{j=1}^L f_{W_j}(w_j|h, \gamma)$. The only differences are the followings. For any fixed

$\bar{h} < 0$,

$$\begin{aligned}
\phi^+(W_h) &= \begin{cases} 1 & \text{if } f_W(W_h|\bar{h}) > e^{\bar{h}/\lambda} f_W(W_h|0) \\ \alpha & \text{if } f_W(W_h|\bar{h}) = e^{\bar{h}/\lambda} f_W(W_h|0) \\ 0 & \text{if } f_W(W_h|\bar{h}) < e^{\bar{h}/\lambda} f_W(W_h|0) \end{cases} \\
&= \begin{cases} 1 & \text{if } \mathbb{I}\{\min_{j \in \{1, \dots, L\}} W_{h,j}/G_j > \bar{h}\} > \mathbb{I}\{\min_{j \in \{1, \dots, L\}} W_{h,j}/G_j > 0\} \\ \alpha & \text{if } \mathbb{I}\{\min_{j \in \{1, \dots, L\}} W_{h,j}/G_j > \bar{h}\} = \mathbb{I}\{\min_{j \in \{1, \dots, L\}} W_{h,j}/G_j > 0\} \\ 0 & \text{if } \mathbb{I}\{\min_{j \in \{1, \dots, L\}} W_{h,j}/G_j > \bar{h}\} < \mathbb{I}\{\min_{j \in \{1, \dots, L\}} W_{h,j}/G_j > 0\} \end{cases} \\
&= \begin{cases} 1 & \text{if } \min_{j \in \{1, \dots, L\}} W_{h,j}/G_j \leq 0 \\ \alpha & \text{if } \min_{j \in \{1, \dots, L\}} W_{h,j}/G_j > 0 \end{cases} = (1 - D_{h,0}) + \alpha D_{h,0},
\end{aligned}$$

is most powerful, where the second equality follows from simple transformations, and the third equality follows from the fact that $\mathbb{I}\{\min_{j \in \{1, \dots, L\}} W_{h,j}/G_j > 0\} = 1$ implies $\mathbb{I}\{\min_{j \in \{1, \dots, L\}} W_{h,j}/G_j > \bar{h}\} = 1$ for any $\bar{h} < 0$.

Step 2: Show size control of $\phi_n^+(\hat{h}_n^+, Y^n, X^n)$. Observe that

$$\begin{aligned}
&\lim_n E_{\theta_0 + \frac{h}{n}} [\phi_n^+(\hat{h}_n^+, Y^n, X^n)] \\
&= \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_i < \max \left\{ g(X_i, \theta_0), g(X_i, \theta_0 + \hat{h}_n^+/n) \right\} \text{ for some } i \right\} \\
&= \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_i < g(X_i, \theta_0 + \hat{h}_n^+/n) \text{ for some } i \right\} = 1 - \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_i \geq g(X_i, \theta_0 + \hat{h}_n^+/n) \text{ for all } i \right\} \\
&= 1 - P_h \left\{ \min_{j \in \{1, \dots, L\}} W_{h,j}/G_j > \bar{h}^+ \right\} = \max \left\{ 1 - \exp \left(\frac{h - \bar{h}^+}{\lambda} \right), 0 \right\},
\end{aligned}$$

where the second equality follows by $g(X_i, \theta_0) < g(X_i, \theta_0 + \hat{h}_n^+/n)$ eventually, and the forth equality follows from the weak convergence in Lemma 3. Since $\bar{h}^+ = \lambda \log(1/(1 - \alpha))$, we obtain the asymptotic size control

$$\lim_n E_{\theta_0 + \frac{h}{n}} [\phi_n^+(\hat{h}_n^+, Y^n, X^n)] = 1 - (1 - \alpha)e^{h/\lambda} \leq \alpha,$$

for any $h \geq 0$.

Step 3: Show power optimality of $\phi_n^+(\hat{h}_n^+, Y^n, X^n)$. This step follows from the same argument to show Theorem 2. Therefore, the conclusion is obtained.

A.7. Proof of Theorem 6. The proof of this part is similar to the one for Theorem 3 after replacing $D_{h,\bar{h}}$ and λ in Section 2 with the ones in Section 3.

Step 1: Show size control of $\phi_n(\hat{h}_n^-, Y^n, X^n)$. For $h = 0$, we have

$$\begin{aligned}
&\limsup_n E_{\theta_0} [\phi_n(\hat{h}_n^-, Y^n, X^n)] \\
&= \limsup_n P_{\theta_0} \left\{ Y_i > g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for all } i \text{ or } Y_i < g(X_i, \theta_0) \text{ for some } i \right\} \\
&\leq \lim_n P_{\theta_0} \left\{ Y_i > g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for all } i \right\} + \lim_n P_{\theta_0} \left\{ Y_i < g(X_i, \theta_0) \text{ for some } i \right\} \\
&= \min \left\{ \exp \left(\frac{h - \bar{h}^-}{\lambda} \right), 1 \right\} + 0 = \alpha,
\end{aligned}$$

where the inequality follows from the union bound, and the second equality follows from the weak convergence in Lemma 3. Thus, $\phi_n(\hat{h}_n^-, Y^n, X^n)$ controls the asymptotic size.

Step 2: Show power optimality of $\phi_n(\hat{h}_n^-, Y^n, X^n)$. For any $h \neq 0$, note that

$$\begin{aligned} & \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_i > g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for all } i \text{ or } Y_i < g(X_i, \theta_0) \text{ for some } i \right\} \\ &= 1 - \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_i \leq g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for some } i \text{ and } Y_i \geq g(X_i, \theta_0) \text{ for all } i \right\} \\ &= 1 - \left[\lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_i \leq g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for some } i \right\} - \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_i < g(X_i, \theta_0) \text{ for some } i \right\} \right] \\ &= \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_i > g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for all } i \right\} + \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_i < g(X_i, \theta_0) \text{ for some } i \right\} \end{aligned}$$

where the second equality follows from $g(X_i, \theta_0) \leq g(X_i, \theta_0 + \hat{h}_n^-/n)$ eventually, which implies the event $\{Y_i \leq g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for some } i\}$ includes $\{Y_i \leq g(X_i, \theta_0) \text{ for some } i\}$ eventually. For $h > 0$, it holds

$$\begin{aligned} & \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_i > g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for all } i \text{ or } Y_i < g(X_i, \theta_0) \text{ for some } i \right\} \\ &= \min \left\{ \exp \left(\frac{h - \bar{h}^-}{\lambda} \right), 1 \right\} + 0 = E_h[\phi^-(W)]. \end{aligned}$$

On the other hand, for $h < 0$, it holds

$$\begin{aligned} & \lim_n P_{\theta_0 + \frac{h}{n}} \left\{ Y_i > g(X_i, \theta_0 + \hat{h}_n^-/n) \text{ for all } i \text{ or } Y_i < g(X_i, \theta_0) \text{ for some } i \right\} \\ &= 1 - (1 - \alpha)e^{h/\lambda} = E_h[\phi^+(W_h)]. \end{aligned}$$

Thus, the AUMP follows from the same argument to show Theorem 1.

REFERENCES

- Chen, X., Christensen, T. M., and Tamer, E. (2018). Monte carlo confidence sets for identified sets. *Econometrica*, 86(6):1965–2018. [Cited on page 2]
- Chen, X., Gentry, M., Li, T., and Lu, J. (2025). Identification and inference in first-price auctions with risk-averse bidders and selective entry. *Review of Economic Studies*, page rda016. [Cited on page 2]
- Chernozhukov, V. and Hong, H. (2004). Likelihood estimation and inference in a class of non-regular econometric models. *Econometrica*, 72(5):1445–1480. [Cited on pages 1, 2, 10, and 11]
- Choi, S., Hall, W. J., and Schick, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *The Annals of Statistics*, 24(2):841–861. [Cited on page 5]
- Christensen, B. J. and Kiefer, N. M. (1991). The exact likelihood function for an empirical job search model. *Econometric Theory*, 7(4):464–486. [Cited on page 1]
- Donald, S. and Paarsch, H. (1993). Maximum likelihood estimation when the support of the distribution depends upon some or all of the unknown parameters. *Typescript. London, Canada: Department of Economics, University of Western Ontario*. [Cited on page 1]

- Donald, S. G. and Paarsch, H. J. (2002). Superconsistent estimation and inference in structural econometric models using extreme order statistics. *Journal of Econometrics*, 109(2):305–340. [Cited on page 1]
- Flinn, C. and Heckman, J. (1982). New methods for analyzing structural models of labor force dynamics. *Journal of Econometrics*, 18(1):115–168. [Cited on page 1]
- Hirano, K. and Porter, J. R. (2003). Asymptotic efficiency in parametric structural models with parameter-dependent support. *Econometrica*, 71(5):1307–1338. [Cited on pages 1, 2, 3, 6, 7, and 11]
- Hong, H. (1998). Maximum likelihood estimation for job search, auction, and frontier production function models. [Cited on page 1]
- Lehmann, E. L. and Romano, J. P. (2022). *Testing statistical hypotheses*, volume 4. Springer. [Cited on pages 1, 4, 6, and 12]
- Shao, J. (2003). *Mathematical statistics*. Springer Science & Business Media. [Cited on page 4]
- Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90. [Cited on page 1]
- van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press. [Cited on page 4]

DEPARTMENT OF ECONOMICS, UNIVERSITY OF WISCONSIN, MADISON, 1180 OBSERVATORY DRIVE, MADISON, WI 53706-1393, USA.

Email address: yuya.shimizu@wisc.edu

DEPARTMENT OF ECONOMICS, LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, LONDON, WC2A 2AE, UK.

Email address: t.otsu@lse.ac.uk