

**A Maximum Score Test for Binary Response Models**

Journal:	<i>Studies in Nonlinear Dynamics & Econometrics</i>
Manuscript ID:	DGSNDE.2012.0038
Manuscript Type:	Article_Transfer from EdiKit
Classifications (use JEL: http://www.aeaweb.org/jel/jel_class_system.php):	c12, c14
Keywords:	maximum score, binary chice, binary chice
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
2009_SNDE.zip 2009_mayer_code.zip	

SCHOLARONE™
Manuscripts

Only

1. Introduction

This paper develops new hypothesis tests for binary response models estimated by maximum score. Manski's (1975, 1985) maximum score imposes an assumption of median independence that is less restrictive than other estimators of the binary response model.¹ The well-known obstacle for empirical implementation is that the asymptotic distribution cannot be used for hypothesis testing (Kim and Pollard 1990 and Lee and Seo 2008). Two approaches that address the problem are smoothed maximum score (Horowitz 1992) extended to threshold models by Seo and Linton (2007), and subsampling proposed by Delgado, Rodriguez-Poo and Wolf (2001).² Smoothed maximum score has a tractable distribution only under stronger assumptions than Manski (1975, 1985), namely that certain densities are sufficiently differentiable. Subsampling does not impose additional assumptions but entails large computational costs (Delgado *et al.* 2001, p. 248). Lee and Seo (2008, p. 495) also note that subsampling may be problematic because "the asymptotics breaks down when the model is close to a continuous one." Moreover, existing results for smoothed maximum score and subsampling are limited to nested hypothesis testing of regression coefficients.

This paper uses a new method to develop tests that that uses a "discretization" argument to circumvent the intractable asymptotic distribution of the maximum score estimator. The tests do not impose assumptions on the data generating process stronger than Manski (1975, 1985) and Lee and Seo (2008), require less computational time than subsampling, and can be applied to a wide range of nested and non-nested hypotheses. The tests are based on a certain asymptotic equivalence that is obtained by discretizing the continuous covariates with a scheme in which the discretization becoming finer as the sample size increases. The discretization parameters (analogous to bandwidths) are specified so that discretization vanishes asymptotically which is critical for the test to be consistent. The basic idea is to restrict the rate at which the discretization vanishes so that the test statistic does not depend on the intractable asymptotic distribution of the maximum score estimator or, more specifically, the asymptotic distribution of the test statistic does not change if the maximum score estimates are replaced by their probability limits.

The test statistics are based on the difference between the objective functions maximized by the maximum score estimators under the null and

¹Other binary response estimators that are consistent under weak distributional assumptions have been proposed by Cosslett (1983), Han (1987), Ichimura (1993), Klein and Spady (1993), Lewbel (2000) and others. For an extension of maximum score to a disequilibrium model, see Mayer and Dorsey (1998) and Mayer (1999).

² Some empirical studies have adopted bootstrapping which Abrevaya and Huang (2005) have shown is inconsistent for maximum score. Lee and Pun (2006) propose a more general bootstrap that can be used but does not allow threshold versions of the model (Lee and Seo 2008, p. 495).

1
2
3 alternative hypotheses and as such reflect differences in predictive accuracy.
4 Potential applications include tests of regression coefficients, and particular
5 parametric models for the error distributions. Since maximum score is robust to
6 distributional misspecification, it provides a natural check on the probit and logit
7 models commonly used in applied work. Although specification tests have been
8 previously proposed for probit and logit, our test appears to be the first to be based on
9 maximum score. The test can also be used to compare two nonnested median
10 regression models estimated by maximum score, both of which can be misspecified.
11 Tests for models of this type have not been previously proposed.
12

13
14 Section 2 describes the general model, required estimators, testable
15 hypotheses, and provides an informal description of the test. One problem is that
16 the limiting distribution of the test statistic is degenerate under nested and
17 overlapping non-nested hypotheses. To address this, we adopt the approach used
18 for asymptotic degeneracy under the null in the nonparametric testing literature,
19 namely sample-splitting (Yatchew 1992, Whang and Andrews 1993, and
20 Horowitz 2009) and stochastic weighting (see Robinson 1991 and Hidalgo 1999,
21 p.377). Following these studies, we specify general weighted test statistics in
22 which sample-splitting is a special case. Section 3 presents the formal results.
23 The tests are shown to be asymptotically normal under the null, and converge to
24 infinity under the alternative. Theorem 1 covers nested hypotheses, while
25 Theorem 2 covers the non-nested case.
26

27
28 Section 4 investigates the finite-sample size and power properties of the
29 test with Monte Carlo experiments for two sets of hypotheses. The first is whether
30 the coefficient of a covariate is zero. The second is whether the errors are
31 logistically distributed. The experiments compare two different weighting
32 schemes for the test statistic, sample-splitting and stochastic weights generated
33 from a normal distribution. The results suggest that the stochastic weighted
34 version has good size in finite samples, while the sample-splitting version tends to
35 be oversized. Both versions appear to have good power. The discretization
36 parameters are analogous to bandwidths in nonparametric statistics. The
37 asymptotic distributions in Theorems 1 and 2 only require that the parameters are
38 chosen so that the discrete approximation converges sufficiently slowly
39 (Assumptions 4-7 below). As in the case of bandwidth selection, a practical
40 problem is that there is an infinite number of asymptotically equivalent choices
41 with potentially different finite-sample size and power properties. The
42 experiments in Section 4 also provide evidence for several different
43 configurations of the discretization parameters. We find that increasing the rate
44 that the discretization vanishes improves power in some cases. One direction for
45 future research is developing criteria for choosing the discretization parameters.
46 Section 5 concludes.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

2. Framework

2.1 Model and Estimation

The proposed test is quite general. It can be applied to models estimated by Manski's (1975, 1985) maximum score and threshold models estimated by the recent extension of Lee and Seo (2008). The latter has a wide range of applications including random utility and asymmetric market response. To cover all these models, we consider the general binary response model:

$$y_i = I(x_i\beta + x_i\alpha I(D_i > \gamma) + \varepsilon_i \geq 0)$$

for $i=1, \dots, n$ where $\theta \equiv (\beta, \alpha, \gamma)$ is an unknown parameter vector, and x_i is a vector of covariates. The threshold model of Lee and Seo (2008) is the special case of $\alpha \neq 0$ with D_i the observed threshold variable. The variable ε_i is an unobserved error with a conditional distribution $F_{\varepsilon_i|x_i, D}$. It is important to emphasize that our test does not require one to assume $\alpha \neq 0$ or $\alpha = 0$. If one wishes to assume $\alpha = 0$, then the restrictions on D and γ stated in the assumptions below can be ignored.

Let Θ denote the parameter space for θ and Φ the space of distribution functions for $F_{\varepsilon_i|x_i, D}$. We consider two competing models: $M^R = \{\Theta^R, \Phi^R\}$ and $M^U = \{\Theta^U, \Phi^U\}$. In what follows, $\hat{\theta}^R$ and $\hat{\theta}^U$ denote two estimators of θ with probability limits $\theta^R \in \Theta^R$ and $\theta^U \in \Theta^U$. For nested models $M^R \subset M^U$, $\hat{\theta}^R$ is a restricted estimator and $\hat{\theta}^U$ is the unrestricted maximum score estimator that maximizes the sample score function:

$$S_n(\theta^*) = n^{-1} \sum_{i=1}^n (2y_i - 1) I(x_i\beta^* + x_i\alpha^* I(D_i \geq \gamma^*) \geq 0) \quad (1)$$

We assume:

Assumption 1

(a) $\{(y_i, x_i, D_i) : i=1, \dots, n\}$ is an i.i.d. sample.

(b) $(\hat{\beta}^j, \hat{\alpha}^j) - (\beta^j, \alpha^j) = O_p(n^{-a(\alpha, \beta, j)})$ and $\hat{\gamma}^j - \gamma^j = O_p(n^{-a(\gamma, j)})$ $j=R, U$ for some $a(\beta, \alpha, j) > 0$ and $a(\gamma, j) > 0$.

Part (b) allows the restricted and unrestricted estimators to converge at different rates which accommodates certain hypotheses discussed below in which

$(\hat{\beta}^U, \hat{\alpha}^U, \hat{\gamma}^U)$ are maximum score estimators and $(\hat{\beta}^R, \hat{\alpha}^R, \hat{\gamma}^R)$ are maximum likelihood estimators. Under the assumptions of Lee and Seo (2008), the

maximum score estimators $\hat{\beta}^U$ and $\hat{\alpha}^U$ converge at rate $n^{1/3}$, $\hat{\gamma}^U$ converges at rate n under a discontinuity assumption (Lee and Seo 2008, Assumption 9) and at rate $n^{1/3}$ otherwise. Consequently, the discontinuity assumption can be relaxed by simply specifying a convergence rate for $\hat{\gamma}^j$ of $n^{1/3}$.

2.2 Nested Hypotheses: $M^R \subset M^U$

Theorem 1 in the next section shows that the test has correct size and unit power asymptotically for null and alternative hypotheses that imply

$$H_0^* : \theta^U = \theta^R \quad \text{and} \quad H_a^* : E(S_n(\theta^U) - S_n(\theta^R)) \neq 0 \quad (2)$$

H_a^* is equivalent to $\theta^R \neq \theta^U$ when θ is identified and either $\theta^U = \theta$ or $\theta^R = \theta$. (see Manski 1985 and Lee and Seo 2008 for identification assumptions.) Examples of hypotheses that imply (2) include nonlinear restrictions on θ with $\hat{\theta}^R$ as the restricted maximizer of (1). Another testable hypothesis is linearity ($\alpha = 0$) against the threshold alternative ($\alpha \neq 0$).³

Other testable hypotheses are restrictions on the conditional distribution $F_{\varepsilon|x,D}$. There has been renewed interest recently in heteroscedasticity (see, for example, Hoderlein 2009). Consequently, an important example is testing for arbitrary heteroscedasticity, that is, whether ε is statistically independent of (x,D) or only median independent:

$$H_0 : F_{\varepsilon|x,D} = F_\varepsilon \quad \text{against} \quad H_a : \text{med}(\varepsilon|x,D) \neq 0 \quad \forall x,D$$

For the $\hat{\theta}^R$ one could use the estimators of Cosslett (1983), Han (1987) or Lewbel (2000), for example. Another important application is a test of the standard probit specification with Φ^R as a normal distribution, Φ^U the space of distribution functions with ε median independent of (X,D) , and $\hat{\theta}^R$ the maximum likelihood estimator.

³ Clearly γ is not identified when $\alpha = 0$. To test the latter, our test requires:

$$\text{plim}(\hat{\beta}^U, \hat{\alpha}^U) = (\beta, \alpha) \text{ under } H_0: \alpha = 0$$

Since $\alpha = 0$ violates Lee and Seo (2008, Assumption 1), their results do not apply. In Appendix 2, however, we show that this in fact holds.

2.3 Non-nested Hypotheses: $M^R \not\subset M^U$ and $M^U \not\subset M^R$

The test can also be used to compare non-nested models. Consider two competing non-nested models of the conditional median of y given x and D :

$$M^R = \{\theta^* \in \Theta^R, \text{med}(y|x, D) = I(x\beta^* + x\alpha^* I(D > \gamma^*) \geq 0)\} \text{ and}$$

$$M^U = \{\theta^* \in \Theta^U, \text{med}(y|x, D) = I(x\beta^* + x\alpha^* I(D > \gamma^*) \geq 0)\},$$

where $\Theta^R \not\subset \Theta^U$, $\Theta^U \not\subset \Theta^R$, and $\hat{\theta}^R$ and $\hat{\theta}^U$ maximize (1) over Θ^R and Θ^U , respectively; θ^R and θ^U again denote the probability limits. The hypotheses are:

$$H_o : E(S_n(\theta^U) - S_n(\theta^R)) = 0 \quad \text{“equivalent” models}$$

$$H_U : E(S_n(\theta^U) - S_n(\theta^R)) > 0 \quad \text{model } M^U \text{ is better}$$

$$H_R : E(S_n(\theta^U) - S_n(\theta^R)) < 0 \quad \text{model } M^R \text{ is better}$$

Asymptotically, the population score function, $E(S_n(\theta))$, measures how accurately $I(x\beta + x\alpha I(D > \gamma) \geq 0)$ predicts the outcome y . This approach is analogous to Vuong (1989) with the log-likelihood replaced by $E(S_n(\theta))$. As in Vuong (1989), neither model needs to be the true model. In this case, analogous to Vuong (1989), H_U , for example, can be interpreted as the hypothesis that M^U is “closer” to the true model than M^R . To see this, suppose the true model is $\text{med}(y|x, D) = I(x\beta + x\alpha I(D > \gamma) \geq 0)$. Then, under the assumptions of Manski (1985) (or Lee and Seo 2008), we have $E(S_n(\theta)) > E(S_n(\theta^j))$ $j=R, U$. Since

$$E(S_n(\theta^U)) - E(S_n(\theta^R)) = E(S_n(\theta^U)) - E(S_n(\theta)) - [E(S_n(\theta^R)) - E(S_n(\theta))]$$

it therefore follows that under H_U model M^U is closest to the true model in terms of asymptotic predictive ability.

2.4. Informal Description of Test

Our strategy is to construct a test statistic, $T(\hat{\theta}^U, \hat{\theta}^R)$, that has the same asymptotic distribution if the maximum score estimates are replaced by their probability limits:

$$T(\hat{\theta}^R, \hat{\theta}^U) = T(\theta^R, \theta^U) + o_p(1) \quad (3)$$

This allows evaluating hypotheses in terms of the probability limits while avoiding the intractable asymptotic distribution of $\hat{\theta}^U$ and $\hat{\theta}^R$. The asymptotic irrelevance property is achieved using a discretization method in which a discrete random variable is generated for each covariate that has an infinite support.

To motivate the approach, first suppose x_i ($i=1, \dots, n$) is a discrete random variable with a support that has $N < \infty$ points, $P(x_i = 0) = 0$, and $\hat{\beta} - \beta = o_p(1)$ for some estimator $\hat{\beta}$ of $\beta > 0$. Then for some finite scalar $\Delta_N > 0$ and $x^* \in \text{support}(x)$, we have $P(|x_i| \leq \Delta_N) = 1$ and $0 < \inf_{x \in \text{support}(x)} |x\beta| = |x^*\beta| \leq |x_i\beta|$. Therefore

$$\begin{aligned} |\hat{\beta} - \beta| \leq \frac{|x^*\beta|}{\Delta_N} &\Rightarrow |x_i\hat{\beta} - x_i\beta| \leq |x^*\beta| \Rightarrow I(x_i\hat{\beta} \geq 0) = I(x_i\beta \geq 0) \quad \forall i \\ &\Rightarrow \frac{1}{\sqrt{n}} \sum_{i=1}^n (2y_i - 1)I(x_i\hat{\beta} \geq 0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (2y_i - 1)I(x_i\beta \geq 0) \end{aligned} \quad (4)$$

Clearly, $\lim_{n \rightarrow \infty} \Pr(|\hat{\beta} - \beta| \leq \frac{|x^*\beta|}{\Delta_N}) = 1$ since $\hat{\beta} - \beta = o_p(1)$. It therefore follows from (4) that $\hat{\beta}$ is asymptotically irrelevant:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (2y_i - 1)I(x_i\hat{\beta} \geq 0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (2y_i - 1)I(x_i\beta \geq 0) + o_p(1) \quad (5)$$

The above argument breaks down, however, if x has a continuous distribution with a support that is not bounded away from zero. In this case, $|x^*\beta| = 0$ with probability one and, thus, $\lim_{n \rightarrow \infty} \Pr(|\hat{\beta} - \beta| \leq \frac{|x^*\beta|}{\Delta_N}) = 1$ does not hold in general.

Consequently, a discretization scheme is needed for continuous x .

Different discretization schemes can be used. In this paper we adopt one based the uniform distribution. The scheme we adopt can be illustrated with a simple example. Suppose x_i is a continuous random variable with

$P(c_0 < x_i < c_1) = 1$ for finite c_0 and c_1 , $\beta > 0$ and $\hat{\beta} - \beta = O_p(n^{-1/3})$. Let $z_{jN} = c_0 + (c_1 - c_0)j/N + u_{jN}$, where u_{jN} $j=0, \dots, N$, are independent and uniformly distributed on $[0, (c_1 - c_0)/N]$. Define the discrete approximation for x_i :

$$x_{iN} = \begin{cases} z_{0N} & \text{if } c_0 < x_i \leq z_{0N} \\ z_{jN} & \text{if } z_{(j-1)N} < x_i \leq z_{jN} \end{cases} \quad j=1, \dots, N \quad (6)$$

Since $|x_{iN} - x_i| \leq 2(c_1 - c_0)/N$, the approximation error vanishes asymptotically in the sense that $\text{plim}_{N \rightarrow \infty} (x_{iN} - x_i) = 0$.

The basis for constructing test statistics satisfying (3) is that $\hat{\beta}$ is asymptotically irrelevant in functions that depend only on the sign of $x_{iN}\beta$ if $x_{iN}\hat{\beta}$ converges more slowly than $\hat{\beta}$. To illustrate we show that (5) holds for $x_{iN}\hat{\beta}$ and $x_{iN}\beta$ under the restriction that $N \rightarrow \infty$ as $n \rightarrow \infty$ and $\lim_{n \rightarrow \infty} n^{1/3}N^{-d} > 0$ for $d > 2$. Let $|z\beta|_N^{\min} = \min(|z_{0N}\beta|, \dots, |z_{NN}\beta|)$ and $c_N^* = \max(|c_0|, |c_1 + (c_1 - c_0)/N|)$. Clearly, $|z\beta|_N^{\min} \leq |x_{iN}\beta|$ and $|x_{iN}| \leq |c_N^*|$. It follows from an expression analogous to (4) that

$$\lim_{n \rightarrow \infty} P(|\hat{\beta} - \beta| \leq \frac{|z\beta|_N^{\min}}{|c_N^*\beta|}) = 1 \Rightarrow$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (2y_i - 1)I(x_{iN}\hat{\beta} \geq 0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (2y_i - 1)I(x_{iN}\beta \geq 0) + o_p(1) \quad (7)$$

Now for any $\lambda > 0$,

$$P(N^d |z\beta|_N^{\min} \geq \lambda) = (1 - P(N^d |z_{0N}\beta| < \lambda)) \dots (1 - P(N^d |z_{NN}\beta| < \lambda))$$

where $P(N^d |z_{jN}\beta| < \lambda) \leq \min[1, 2\lambda N^{1-d} / \beta(c_1 - c_0)]$. Therefore, for large N and $d > 1$ we have:

$$P(N^d |z\beta|_N^{\min} \geq \lambda) \geq \left(1 - \frac{2\lambda N^{2-d}}{N\beta(c_1 - c_0)}\right)^{N+1} \quad (8)$$

Since for any $\lambda > 0$ the right hand side of (8) converges to one as $N \rightarrow \infty$ for $d > 2$, it follows that $N^d |z\beta|_N^{\min} \xrightarrow{p} \infty$ for $d > 2$. Given this and $\hat{\beta} - \beta = O_p(n^{-1/3})$, we have

$$\lim_{n \rightarrow \infty} P(|\hat{\beta} - \beta| \leq \frac{|z\beta|_N^{\min}}{|c_N^*\beta|}) = \lim_{n \rightarrow \infty} P(n^{1/3} |\hat{\beta} - \beta| \leq \frac{n^{1/3} N^d |z\beta|_N^{\min}}{N^d |c_N^*\beta|}) = 1 \quad (9)$$

if $\lim_{n \rightarrow \infty} n^{1/3}N^{-d} > 0$ for $d > 2$ and the result thus follows from (7). Note that the draws from the uniform distribution ensure that $P(|z\beta|_N^{\min} = 0) = 0$ for all finite N . For this reason, it is important to use randomized discretization. For a deterministic scheme such as $z_{jN} = c_0 + (c_1 - c_0)j/N$, for example, either

$P(|z\beta|_N^{\min} = 0) = 0$ or 1. Ruling out the latter would require restrictions on β that would be difficult to justify.

The next section generalizes the above method for (x_i, D_i) , where x is a vector and components of (x_i, D_i) can take values on infinite intervals. The proposed test statistic is based on an approximation of $S_n(\hat{\theta}^U) - S_n(\hat{\theta}^R)$ denoted by $S_{n,N}(\hat{\theta}^U) - S_{n,N}(\hat{\theta}^R)$. The latter is obtained by replacing the covariates x_i and D_i with approximating discrete random variables, $x_{iN(x)}$ and $D_{iN(D)}$, which are defined similar to (6). The approximating discrete random variable for $D_i, D_{N(D)}$, has a support of size $N(D)+1$ which may be different than $N(x)+1$ to accommodate different rates convergence for estimators of (β, α) and γ . In particular, because estimators of γ typically converge faster (Lee and Seo 2008, p.404), a faster rate of convergence is allowed for $D_{N(D)}$.

Since $S_n(\hat{\theta})$ measures number of times $I(x\hat{\beta} + x\hat{\alpha}I(D > \hat{\gamma}) \geq 0)$ correctly predicts y , $S_{n,N}(\hat{\theta}^U) - S_{n,N}(\hat{\theta}^R)$ measures the difference between the null and alternative hypotheses in terms of predictive ability. Lemma 2 in the next section generalizes the above discretization argument and can be used to show that

$$\sqrt{n}(S_{n,N}(\hat{\theta}^U) - S_{n,N}(\hat{\theta}^R)) = \sqrt{n}(S_{n,N}(\theta^U) - S_{n,N}(\theta^R)) + o_p(1)$$

Consequently, the asymptotic distribution of $\sqrt{n}[S_{n,N}(\hat{\theta}^U) - S_{n,N}(\hat{\theta}^R)]$ is the same as $\sqrt{n}[S_{n,N}(\theta^U) - S_{n,N}(\theta^R)]$.

One problem is that since $\theta^R = \theta^U$ under nested and overlapping non-nested null hypotheses, $\sqrt{n}[S_{n,N}(\theta^U) - S_{n,N}(\theta^R)]$ equals zero and, consequently, $\sqrt{n}[S_{n,N}(\hat{\theta}^U) - S_{n,N}(\hat{\theta}^R)]$ has a degenerate limiting distribution. As noted in the introduction, asymptotic degeneracy under the null also occurs for nonparametric specification tests (see Robinson 1991, Yatchew 1992, Whang and Andrews 1993, Hidalgo 1999, and Horowitz 2009). Following these studies, we address the problem by adopting a weighting scheme. Let W_i^R and W_i^U be random variables and define the “weighted” sample score function:

$$n^{-1} \sum_{i=1}^n W_i^j (2y_i - 1) I(x_{iN(x)}\beta^j + x_{iN(x)}\alpha^j I(D_{iN(D)} > \gamma^j) \geq 0) \quad j=U,R$$

For the weights, we assume:

Assumption 2

- i) $P(W_{in}^R = W_{in}^U) < 1$
- ii) $E(W_{in}^R) = E(W_{in}^U) \neq 0$
- iii) $\text{var}(W_{in}^R) \equiv \sigma_R^2$ and $\text{var}(W_{in}^U) \equiv \sigma_U^2$ are finite.
- iv) W_{in}^R and W_{in}^U are statistically independent of $(x_i, D_i, y_i) \forall i$.

Assumption 2 allows both nonstochastic and stochastic weights. A special case is sample splitting: $W_i^R = I(i \leq n/2)$ and $W_i^U = 1 - W_i^R$. Provided that

Assumption 2 holds, the choice of W_i^R and W_i^U does not affect the asymptotic distribution of the test statistics in Theorems 1 and 2 below but can affect finite sample performance. A specification that produced good size and power for sample sizes typically used in practice is given in Section 4.

Finally, we note that specifying $N \rightarrow \infty$ as $n \rightarrow \infty$ is necessary to ensure that the test presented in the next section has unit power asymptotically against all alternatives, $\theta^R \neq \theta^U$. It is clear from (9) that the asymptotic irrelevancy property holds even if the value of N is fixed as $n \rightarrow \infty$. The problem is that for finite N $E(S_{n,N}(\theta^U) - S_{n,N}(\theta^R)) = 0$ is possible for $\theta^R \neq \theta^U$ since the support of $(x_{N(x)}, D_{N(D)})$ would be finite (see, for example, Horowitz 2009b, p.101).

Consequently, the test might not have power against some alternatives if N is held fixed. Lemma 1 in next section shows that this can be addressed by specifying that $N \rightarrow \infty$ as $n \rightarrow \infty$ so that $E(S_{n,N}(\theta^j)) \rightarrow E(S_n(\theta^j))$ for $j=U,R$. The test is then consistent against all alternatives such that $E(S_n(\theta^U) - S_n(\theta^R)) \neq 0$. The latter condition, in turn, is equivalent to $\theta^R \neq \theta^U$ under the usual identification assumptions adopted in maximum score estimation (Manski 1985, and Lee and Seo 2008).⁴

3. Formal Results

This section formally presents assumptions under which the test is asymptotically normal under the null and diverges under alternative. Theorems 1 and 2 cover nested and non-nested hypotheses, respectively. Both are based on two preliminary lemmas. We assume:

⁴ $\hat{\gamma}^j$ has a faster rate of convergence under the discontinuity assumption of Lee and Seo (2008, Assumption 9), which allows one to choose a larger value for $N(D)$ relative to n . In turn the approximate sample score, $S_{N,n}$, can be specified to converge faster to the true sample score, S_n . This suggests that the test might have more power against some alternatives under the discontinuity assumption.

Assumption 3

a) For $j=R, U$: (i) $(x_i \beta^j, x_i \alpha^j) = (\sum_{h=1}^H x_{hi} \beta_h^j + \tilde{x}_i \tilde{\beta}^j, \sum_{h=1}^H x_{hi} \alpha_h^j + \tilde{x}_i \tilde{\alpha}^j)$, where $\beta_1^j \neq 0$,

$\beta_1^j \neq -\alpha_1^j$, the support of \tilde{x} , $\text{support}(\tilde{x})$, is finite, and x_{i1} has a conditional Lebesgue density that is positive everywhere for almost every $(x_{i2}, \dots, x_{iH}, \tilde{x}_i, D_i)$.

(ii) D_i is a continuous random variable with $\gamma \in \text{support}(D)$.

b) $P(X)=1$ where

$X = \{c_{h0} < x_{hi} < c_{h1}, c_{H+1,0} < D_i < c_{H+1,1}, \tilde{x}_i \in \text{support}(\tilde{x}_i); h=1, \dots, H, \forall i\}$ for some

known (possibly infinite) constants $c_{h0}, c_{h1}, h=1, \dots, H+1$.

Assumption 3a is standard (Manski 1985 and Lee and Seo 2008) and allows H components of x to have infinite supports. Assumption 3b specifies a set $X \supset \text{support}(x_i, D_i)$ and allows covariates to have known finite bounds that can be used to construct smoothing parameters below. Without such *a priori* information, one can specify $c_{h0} = -\infty$ and $c_{h1} = +\infty$.

Assumptions 4 and 5 specify the discrete approximations for the x_{hi} and

D_i .

Assumption 4

Let

$$x_{hiN(x)} = \sum_{j=1}^{N(x)} z_{hjN(x)} I(z_{h,j-1,N(x)} < x_{hi} \leq z_{hjN(x)}) + z_{h0N(x)} I(c_{h0N(x)} < x_{hi} \leq z_{h0N(x)})$$

$$D_{iN(D)} = \sum_{j=1}^{N(D)} z_{H+1,j,N(D)} I(z_{H+1,j-1,N(D)} < D_i \leq z_{H+1,j,N(D)}) + z_{H+1,0,N(D)} I(c_{H+1,0,N(D)} < D_i \leq z_{H+1,0,N(D)})$$

where the following hold for $N(k)$ with $k=x$ and $h=1, \dots, H$, and with $k=D$ and $h=H+1$:

(a) $z_{hjN(k)} = c_{h0N(k)} + (c_{h1N(k)} - c_{h0N(k)})j/N(k) + u_{hjN(k)}$ for $j=0, 1, \dots, N$.

(b) $c_{h0N(k)} < c_{h1N(k)} \quad \forall N(k)$,

(c) if c_{h0} is finite, then $c_{h0N} = c_{h0}$; if $c_{h0} = -\infty$, then $\lim_{N(k) \rightarrow \infty} c_{h0N(k)} = -\infty$ with $c_{h0N(k)}$ finite for $N(k) < \infty$.

(d) if c_{h1} is finite, then $c_{h1N(k)} = c_{h1}$; if $c_{h1} = +\infty$, then $\lim_{N(k) \rightarrow \infty} c_{h1N(k)} = +\infty$ with $c_{h1N(k)}$ finite for $N(k) < \infty$.

(e) $N(k)(c_{h1N(k)} - c_{h0N(k)})^{-1} \rightarrow \infty$ as $N(k) \rightarrow \infty$

- (f) $c_{hjN(k)}(c_{11N(k)} - c_{10N(k)})^{-1} = O(1) \quad \forall h; j=0,1.$
 (g) $\forall N, P(c_{1,0N(x)} < x_{1i} < c_{1,1N(x)}, \dots, c_{H+1,0N(D)} < D_i < c_{H+1,1N(D)}) > 0$
 (h) $u_{hjN(k)} \square \text{uniform}(0, (c_{h1N(k)} - c_{h0N(k)}) / N(k))$

Assumption 5

Let $\mathbf{z}_{hN(k)} = (z_{h0N(k)}, z_{h1N(k)}, \dots, z_{hN(k)N(k)})$. Assume:

- (a) The components of $\mathbf{z}_{hN(k)}$ are mutually independent for $k=x$ and $h=1, \dots, H$ and for $k=D$ and $h=H+1$.
 (b) $\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1, N(D)}$ are mutually independent and independent of $(x_i, D_i, y_i) \quad \forall i$.

For example, the $x_{hiN(x)}$ are discrete random variables that approximate x_{hi} on the interval $(c_{h0N(x)}, c_{h1N(x)})$. The $z_{hiN(x)}$ used in the approximation are constructed by adding uniform draws $u_{hjN(x)}$ from $(0, (c_{h1N(x)} - c_{h0N(x)}) / N(x))$ to a sequence of constants that increases from $c_{h0N(x)}$ to $c_{h1N(x)}$ (Assumptions 4a and 4h). Assumption 5 requires that the $z_{hiN(x)}$ are statistically independent of (y, x, D) . Assumption 5a holds if the $u_{hjN(x)}$ are independent across j and h . The values of $z_{hiN(x)}$ divide $(c_{h0N(x)}, c_{h1N(x)})$ into $N(x)+1$ subintervals. Each x_{ih} is approximated by the value of $x_{hiN(x)}$ which is the upper limit of the subinterval in which x_{ih} lies. The order of the approximation error is $O_p(N(x)^{-1}(c_{h1N(x)} - c_{h0N(x)}))$. Assumption 4e ensures that the approximation error vanishes as $N(x) \rightarrow \infty$.

The weighted approximate sample score functions are

$$S_{n,N}^j(\theta^j) = n^{-1} \sum_{i=1}^n W_{in}^j (2y_i - 1) I(x_{iN(x)} \beta^j + x_{iN(x)} \alpha^j I(D_{iN(D)} > \gamma^j) \geq 0) I_{N(x), N(D)}(x_i, D_i)$$

$j=R, U$ where the W_{in}^j satisfy Assumption 2 and $I_N(x_i, D_i) = I(c_{h0N(x)} < x_{hi} < c_{h1N(x)}, \dots, c_{H+1,0N(D)} < D_i < c_{H+1,1N(D)})$. The latter trims out observations for which components of $x_{iN(x)}$ or $D_{iN(D)}$ that approximate covariates with infinite intervals (c_0, c_1) can be zero in finite samples. The function $S_{n,N}^j(\theta^j)$ approximates the weighted sample score function evaluated at the covariates x_i and D_i , denoted $S_n^j(\theta^j)$. Lemma 1 shows that the approximation error vanishes asymptotically.

Lemma 1

Under Assumptions 1, 2, 3, 4 and 5:

$$\lim_{(N(x), N(D)) \rightarrow \infty} E[S_{n,N}^j(\theta^j) - S_n^j(\theta^j)] = 0.$$

Proof: See Appendix 1.

The next assumption restricts $(N(x), N(D))$ relative to the sample size n . Under it, $\sqrt{n}S_{n,N}^j(\hat{\theta}^j)$ and $\sqrt{n}S_{n,N}^j(\theta^j)$ are asymptotically equivalent.

Assumption 6

$(N(x), N(D)) \rightarrow \infty$ as $n \rightarrow \infty$ such that $\lim n^{a(\beta, \alpha, j)} N(x)^{-d} > 0$ and $\lim n^{a(\gamma, j)} N(D)^{-d} > 0$ for some $d > 2$, $j=R, U$.

Assumption 6 implies that $(\hat{\beta}^j, \hat{\alpha}^j, \hat{\gamma}^j)$ converges to $(\beta^j, \alpha^j, \gamma^j)$ faster than $(x_{iN(x)}, D_{iN(D)})$ converges to (x_i, D_i) by factors of $(N(x), N(D))$; or, in other words, the ratios of the convergence rates go to zero for each point in the support of $(x_{iN(x)}, D_{iN(D)})$. To see this, let $d > 2$ and consider, for example,

$$\begin{aligned} n^{a(\beta, \alpha, j)} N(x)^{-d} &= \frac{n^{a(\beta, \alpha, j)}}{N(x)^2 (c_{h1N(x)} - c_{h0N(x)})^{-1}} N(x)^{2-d} (c_{h1N(x)} - c_{h0N(x)})^{-1} \\ &= \frac{n^{a(\beta, \alpha, j)}}{N(x)^2 (c_{h1N(x)} - c_{h0N(x)})^{-1}} o(1) \end{aligned}$$

where the second equality holds by Assumptions 4c and 4d. Thus, Assumption 6

implies:

$$\frac{n^{a(\beta, \alpha, j)}}{N(x)(c_{h1N(x)} - c_{h0N(x)})^{-1}} / N(x) \rightarrow \infty \quad \forall h$$

where $n^{a(\beta, \alpha, j)}$ is the rate of convergence of $\hat{\beta}^j$ by Assumption 1b, and $N(x)(c_{h1N(x)} - c_{h0N(x)})^{-1}$ is the rate of convergence of x_{hiN} (see equation A1 in Appendix 1).

Lemma 2 Under Assumptions 1 to 6:

$$\sqrt{n}S_{n,N}^j(\hat{\theta}^j) = \sqrt{n}S_{n,N}^j(\theta^j) + o_p(1) \quad j=R,U$$

Proof: Appendix 1.

Theorem 1 Suppose Assumptions 1 to 6 hold. Let

$$T_n = \frac{\sqrt{n}[S_{n,N}^U(\hat{\theta}^U) - S_{n,N}^R(\hat{\theta}^R)]}{\hat{\sigma}_S} \quad \text{where } \hat{\sigma}_S \xrightarrow{p} \sigma_S \equiv \sqrt{\text{var}(\sqrt{n}[S_n^U(\theta^U) - S_n^R(\theta^R)])}.$$

Then as $n \rightarrow \infty$: (i) $T_n \xrightarrow{d} N(0,1)$ if $\theta^U = \theta^R$. (ii) $|T_n| \xrightarrow{d} \infty$ if

$$E[S_n(\theta^U) - S_n(\theta^R)] \neq 0$$

Proof: See Appendix 1.

The test is consistent for the null $\theta^U = \theta^R$ against the alternative $\theta^U \neq \theta^R$ provided that θ^U is observationally distinguishable from θ^R in the sense that $E[S_n(\theta^U) - S_n(\theta^R)] \neq 0$. The test requires a consistent estimator of σ_S that can be constructed using Lemmas 1 and 2.

Theorem 2 below covers non-nested hypotheses. For strictly non-nested models, a new issue arises because we can have equivalent models with unequal indexes. To avoid a bias in the asymptotic distribution, we need the approximate sample score to converge to its expected value more slowly than the estimated index converges. This entails averaging the sample score over m instead of n sample values, where m goes to infinity more slowly than n .

Assumption 7 $n \rightarrow \infty$ and $(N(x), N(D)) \rightarrow \infty$ as $m \rightarrow \infty$ with:

$$(i) m^{1/2} C_{N(x)} N(x)^{-1} \rightarrow 0, \text{ where } C_{N(x)} = \max_h (c_{h1N(x)} - c_{h0N(x)}), \text{ and}$$

$$(ii) m^{1/2} N(D)^{-1} (c_{H+1,1,N(D)} - c_{H+1,0,N(D)}) \rightarrow 0.$$

Assumption 7 implies N goes to infinity faster than $m^{1/2}$ which, when combined with Assumption 6 implies $m=o(n)$.

Theorem 2 Let $T_m = \frac{\sqrt{m}[S_{m,N}^U(\hat{\theta}^U) - S_{m,N}^R(\hat{\theta}^R)]}{\hat{\sigma}_S}$ where

$\hat{\sigma}_S \xrightarrow{p} \sigma_S \equiv \sqrt{\text{var}(\sqrt{m}[S_m^U(\theta^U) - S_m^R(\theta^R)])}$. Under Assumptions 1 to 7, as $m \rightarrow \infty$:

(i) $T_m \xrightarrow{d} N(0,1)$ if

$$E[S_m^U(\theta^U) - S_m^R(\theta^R) | X_N] = O_p(C_{N(x)}N(x)^{-1}) + O_p((c_{H+1,1,N(D)} - c_{H+1,0,N(D)})N(D)^{-1})$$

(ii) $T_m \xrightarrow{d} +\infty$ ($-\infty$) if $E[S_m^U(\theta^U) - S_m^R(\theta^R)] > 0$ (< 0).

If $(c_{h,0}, c_{h,1})$ is finite for all h , then by Assumptions 4c and 4d “equivalent” models in part (i) of Theorem 2 is the same as

$$E[S_m^U(\theta^U) - S_m^R(\theta^R)] = 0.$$

Otherwise, the condition

$$E[S_m^U(\theta^U) - S_m^R(\theta^R) | X_N] = O_p(C_{N(x)}N(x)^{-1}) + O_p((c_{H+1,1,N(D)} - c_{H+1,0,N(D)})N(D)^{-1})$$

implies $E[S_m^U(\theta^U) - S_m^R(\theta^R)] = 0$ but not conversely. Here the stronger condition is required so that differences between the approximate score functions vanish fast enough as the intervals $(c_{h,0N}, c_{h,1N})$ become infinite.

4. Monte Carlo Experiments

Monte Carlo experiments were conducted to investigate the finite-sample performance of the test for the special case of no threshold effects ($\alpha = 0$). All experiments were generated from a binary response model with one continuous and two discrete covariates:

$$y = I(x\beta + \varepsilon \geq 0) \text{ where } x = (x_1, x_2, x_3), \beta = (1, 1, 1)', x_1 \square N(0,1), x_2 = I(d \geq .5),$$

$$d \square \text{uniform}(0,1), x_3 = (1/4)I(w < .2) + (1/2)I(.2 < w < .8) + (3/4)I(w > .8) \text{ and}$$

$$w \square \text{uniform}(0,1).$$

The three distributions for ε used in the experiments follow Horowitz (1992, p.516):

Dist-L: $\varepsilon \square \text{logistic}(0,1)$

Dist-T3: $\varepsilon \sim \text{t-dist. df}=3$

Dist - H : $\varepsilon = 0.25[1 + 2(x\beta)^2 + (x\beta)^4]v$, where $v \square \text{logistic}(0,1)$

Each experiment consists of 1000 replications for a specified distribution and a vector (N, n, c_{10N}, c_{11N}) where c_{10N} and c_{11N} are the sequences of bounds (see Assumption 4) for the continuous covariate x_1 . We used two different weighting schemes for the approximate sample scores: the conventional split-sample method and normally distributed weights. The latter are independent normally distributed random variables with means of one and variances of 0.25.

Experiments were conducted for two sets of experiments. The first is:

$H_o : \beta_2 = 0$ versus $H_a : \beta_2 \neq 0$

For these hypotheses, $\hat{\beta}^R$ and $\hat{\beta}^U$ are, respectively, the restricted ($\beta_2 = 0$) and unrestricted maximum score estimators both of which converge at rate $n^{-1/3}$. Consequently, Assumption 6 holds for $N = \kappa n^{1/7}$ and Assumption 4 holds for $c_{11N} - c_{10N} = \kappa n^{1/8}$ where κ is a positive constant. We used three values of κ : 8, 80, 133 and two values of n : 300 and 1000. The specifications of n and N are summarized as follows:

$n=300, N=18, 180, 300; n=1000, N=22, 215, 357.$

For this set of experiments, Dist-L and Dist-T were used for the error distribution.

Table 1 reports the estimated sizes based on 5% asymptotic critical values. For the split-sample weights, the estimated size is much greater than the nominal size for all experiments except for the largest value of n and N ($n=1000$ and $N=357$). As is often the case for tests under weak distributional assumptions, this suggests that very large samples are needed for the asymptotic distribution to be a reliable indicator of size. On the other hand, for the normally distributed weights, the estimated and nominal sizes are very close in all experiments. Consequently, at least in terms of size, stochastic weighting schemes might be a good alternative to the split-sample approach. Another possible means of correcting the size is bootstrapping. We leave further exploration of this issue to future work.

Table 2 reports the estimated power for $\beta_2 = -3, -1, 1, 3$ under Dist-L. Going from $n=300$ to $n=1000$ the power increases in all cases and in most changing the value of N (given n) has little effect on power. As one would expect, the departures from the null hypothesis that produce the largest change in the response probability tend to have the greatest power. For example, the

greatest power occurs at $\beta_2 = -3$ (average rejection rate over 99% for $n=1000$) while the lowest occurs at $\beta_2 = 1$ (average rejection rate of 16% for $n=1000$). Using the means of the covariates, the change in $P(y=1)$ is 0.55 for $\beta_2 = -3$, and only 0.11 for $\beta_2 = 1$.

Experiments were also conducted to evaluate the ability of the test to detect distributional misspecification. For this set of experiments, only the stochastic weights were used. The null and alternative hypotheses are:

$$H_0 : \varepsilon \square Dist - L \text{ versus } H_a : \varepsilon \text{ not } \square Dist - L$$

The estimators, $\hat{\beta}^R$ and $\hat{\beta}^U$ are, respectively, maximum likelihood (convergence rate of $n^{-1/2}$) and maximum score. The power of the test was evaluated using Dist-H. Under Dist-L, the error is homoscedastic, and the response probability $P(y=1 | x\beta)$ is a monotonic function of $x\beta$. Whereas under Dist-H the error is heteroscedastic and $P(y=1 | x\beta)$ is nonmonotonic and converges to one half as $|x\beta| \rightarrow \infty$. Dist-L is required for consistency of the maximum likelihood estimator. In contrast, maximum score requires only median independence and, thus, remains consistent under both distributions. Consequently, the null implies $\beta^R = \beta^U$ while the alternative $\beta^R \neq \beta^U$.

Table 3 reports the rejection rates under the null hypothesis for a 5% asymptotic critical value. The power nearly doubles going from $n=300$ (average rejection rate 27%) to $n=1000$ (average rejection rate 50%) with the least power corresponding to the smallest values of N . While four out of six size estimates are within 2% of 5% the test tends to be oversized, especially for small n and large N . For $n=300$, the estimated size ranges from 7.3% ($N=18$) to 9.3% ($N=300$). Increasing n from 300 to 1000 generally improves the size with the estimated size ranging from 5.4% ($N=22$) to 7.6% ($N=357$). One explanation for the over rejection of the null is that by definition the maximum score estimates maximize $S_n(b) - S_n(\hat{\beta}^R)$ over values of b . Since the test statistic, T_n , is a function of the corresponding approximation $S_{n,N}^U(\hat{\beta}^U) - S_{n,N}^R(\hat{\beta}^R)$ it follows that the maximum score estimates approximately maximize the value of T_n . For large N , $S_{n,N}^U(\hat{\beta}^U) - S_{n,N}^R(\hat{\beta}^R)$ more closely approximates $S_n(b) - S_n(\hat{\beta}^R)$ and, thus, as Table 3 indicates, the over rejection is greatest. Consistent with this, large values of N also yield the greatest power.

5. Conclusion

The maximum score estimator of the binary response model is consistent under weaker assumptions than other estimators. An obstacle for empirical applications has been the lack of a tractable asymptotic distribution. To address this problem, we used a new discretization method to develop hypothesis tests for models estimated by maximum score. The tests require fewer assumptions than

smoothed maximum score, less computational time than subsampling and can be applied to a wide range of hypotheses. Potential topics for future research include the ability of the tests to detect local alternatives. For example, equation (A13) in the proof of Theorem 1 reveals that the test has unit power against local alternatives such that $\sqrt{n}[E(S_n(\theta^U) - S_n(\theta^R))]$ is bounded away from zero as $n \rightarrow \infty$. This suggests that the test can detect Pitman alternatives of the form $\theta^U = \theta^R + n^{-1/2}$ under additional assumptions on densities of the regressors. Further investigation would be of interest.

Another direction for future research are modifications to the discretization scheme specified by Assumption 4. For example, an analyst comparing non-nested prediction models might want to focus on a particular set of values for the covariates. This could be addressed by specifying a finer approximation grid for them than other values, which might make the test more powerful in finite samples over this set of values. To illustrate such a scheme, suppose the support for a covariate is $[c_0, c_1] = \bigcup_{i=1}^J [c_{0i}, c_{1i}]$. The fineness of the approximation can be varied over the J subintervals by specifying $N = \sum_{i=1}^J N_i$, where the i -th subinterval is divided into $N_i + 1$ subintervals using say $z_{jN_i}^{(i)} = c_{0i} + (c_{1i} - c_{0i})j / N_i + u_{jN_i}^{(i)}$. Another potential topic is alternative weighting schemes used to address asymptotic degeneracy under nested null hypotheses. Our Monte Carlo experiments suggest that stochastic weights perform better than split-sample weights in terms of finite-sample size.

Appendix 1

Below we let

$$\tau_{iN}(\theta^j) = (2y_i - 1)W_i^j I(x_{iN(x)}\beta^j + x_{iN(x)}\alpha^j I(D_{iN(D)} > \gamma) \geq 0) I_N(x_i, D_i),$$

$$\tau_i(\theta^j) = (2y_i - 1)W_i^j I(x_i\beta^j + x_i\alpha^j I(D_i > \gamma) \geq 0),$$

$$\tau_{iN}(\theta^U, \theta^R) = \tau_{iN}(\theta^U) - \tau_{iN}(\theta^R), \text{ and } \tau_i(\theta^U, \theta^R) = \tau_i(\theta^U) - \tau_i(\theta^R).$$

Consequently,

$$S_n^U(\theta^U) - S_n^R(\theta^R) = n^{-1} \sum_{i=1}^n \tau_i(\theta^U, \theta^R). \text{ The proofs of Lemma 1 and Theorems 1}$$

and 2 involve conditioning on the following sets:

$$X_N = \{c_{h0N(x)} < x_{hi} < c_{h1N(x)}, \dots, c_{H+1,0N(D)} < D_i < c_{H+1,1N(D)}, \tilde{x}_i \in \text{support}(\tilde{x}_i), \\ z_{h0,N(x)}, \dots, z_{hN(x),N(x)}, h = 1, \dots, H; z_{H+1,0,N(D)}, \dots, z_{H+1,N(D),N(D)}; \forall i\} \text{ and}$$

$$X = \{c_{h0} < x_{hi} < c_{h1}, c_{H+1,0} < D_i < c_{H+1,1}, \tilde{x}_i \in \text{support}(\tilde{x}_i); h = 1, \dots, H, \forall i\}$$

Two preliminary lemmas are used:

Lemma A

Suppose (i) $P(|w_N - w| < \varepsilon | \Omega_N) \rightarrow 1$ as $N \rightarrow \infty \forall \varepsilon > 0$, and (ii) a function f is continuous on a Borel set B where $P(w \in B | \Omega_N) = 1$.

Then $P(|f(w_N) - f(w)| < \varepsilon | \Omega_N) \rightarrow 0$ as $N \rightarrow \infty \forall \varepsilon > 0$.

Lemma B

Suppose (i) $P(|w_N - w| < \varepsilon | \Omega_N) \rightarrow 1$ as $N \rightarrow \infty \forall \varepsilon > 0$, and (ii) for a finite M , $P(|w_N| \leq M | \Omega_N) = 1$. Then $E(w_N - w | \Omega_N) \rightarrow 0$ as $N \rightarrow \infty$.

Proofs: Replace the unconditional distributions with the conditional distributions. in Bierens (1994, Theorems 2.1.7 and 2.2.2).

Proof Lemma 1

Conditional on X_N , under Assumptions 4a and 4h, for all h :

either $z_{h,0,N(x)} \leq x_{hi} \leq z_{h,N(x),N(x)}$ which implies

$$\begin{aligned} |x_{hiN(x)} - x_{hi}| &\leq \max_j |z_{hjN(x)} - z_{hj-1,N(x)}| \leq \max_j |u_{hjN(x)} - u_{hj-1,N(x)}| + (c_{h1N(x)} - c_{h0N(x)})/N(x) \\ &\leq 2(c_{h1N(x)} - c_{h0N(x)})/N(x), \end{aligned}$$

or $x_{hi} < z_{h0N(x)}$ which implies $|x_{hiN(x)} - x_{hi}| \leq (c_{h1N(x)} - c_{h0N(x)})/N(x)$. The same

line of argument applies to $|D_{iN(D)} - D_i|$. Therefore,

$$\begin{aligned} P(\bigcap_{h=1}^H \{|x_{hiN(x)} - x_{hi}| < 2(c_{h1N(x)} - c_{h0N(x)})/N(x)\}), \\ |D_{iN(D)} - D_i| < 2(c_{h1N(D)} - c_{h0N(D)})/N(D) | X_N) = 1 \end{aligned} \tag{A1}$$

By iterated expectations and dominated convergence,

$$\lim_{N \rightarrow \infty} E(\tau_{iN}(\theta^j)) = E(\lim_{N \rightarrow \infty} E(\tau_{iN}(\theta^j) | X_N))$$

Hence, it suffices to show $\lim_{N \rightarrow \infty} E(\tau_{iN}(\theta^j) | X_N) = E(\tau_i(\theta^j))$.

By Assumption 4e, as $(N(x), N(D)) \rightarrow \infty$, we have

$$N(x)(c_{h1N(x)} - c_{h0N(x)})^{-1} \rightarrow \infty \text{ for } h=1, \dots, H, \quad N(D)(c_{H+1,1,N(D)} - c_{H+1,0,N(D)})^{-1} \rightarrow \infty,$$

and, thus, by (A1):

$$P(|(x_{iN(x)}\beta^j, x_{iN(x)}\alpha^j, D_{iN(D)} - \gamma^j) - (x_i\beta^j, x_i\alpha^j, D_i - \gamma^j)| < \varepsilon | X_N) \rightarrow 1 \quad \forall \varepsilon > 0 \quad (\text{A2})$$

Let $G_i(x_{iN(x)}\beta^j, x_{iN(x)}\alpha^j, D_{iN(D)} - \gamma^j) = \tau_{iN}(\theta^j) - \tau_i(\theta^j)$. Clearly

$G_i(x_i\beta, x_i\alpha, D_i - \gamma)$ is continuous at all $(x_i\beta, x_i\alpha, D_i - \gamma)$ such that

$x_i\beta + x_i\alpha I(D_i - \gamma > 0) \neq 0$. By iterated expectations,

$$P(x_i\beta + x_i\alpha I(D_i - \gamma > 0) = 0 | X_N) =$$

$$P(x_i\beta + x_i\alpha = 0 | X_N, D_i - \gamma > 0)P(D_i - \gamma > 0 | X_N) +$$

$$P(x_i\beta = 0 | X_N, D_i - \gamma \leq 0)P(D_i - \gamma \leq 0 | X_N)$$

which equals zero under Assumptions 3a and 5. Consequently, G_i is continuous at all $(x_i\beta, x_i\alpha, D_i - \gamma) \in B$, for a set B with $P(B | X_N) = 1$. Since

$G_i(x_i\beta^j, x_i\alpha^j, D_i - \gamma^j) = 0$, Lemma A and (A2) therefore imply

$$P(|G_i(x_{iN(x)}\beta^j, x_{iN(x)}\alpha^j, D_{iN(D)} - \gamma^j)| < \varepsilon | X_N) \rightarrow 1 \text{ as } (N(x), N(D)) \rightarrow \infty \quad \forall \varepsilon > 0.$$

Since $G_i(x_{iN(x)}\beta^j, x_{iN(x)}\alpha^j, D_{iN(D)} - \gamma^j)$ is bounded, Lemma B thus implies

$$E(G_i(x_{iN(x)}\beta^j, x_{iN(x)}\alpha^j, D_{iN(D)} - \gamma^j) | X_N) \rightarrow 0 \text{ as } (N(x), N(D)) \rightarrow \infty \quad (\text{A3})$$

The z_{h1N}, \dots, z_{hNN} are statistically independent of $\tau_i(\theta^j)$ (Assumption 5b). Thus,

$$E(\tau_i(\theta^j) | X_N) =$$

$$E(\tau_i(\theta^j) | c_{1,0N(x)} < x_{1i} < c_{1,1N(x)}, \dots, c_{H,0N(x)} < x_{Hi} < c_{H,1N(x)}, c_{H+1,0N(D)} < D_i < c_{H+1,1N(D)}) \quad (\text{A4})$$

Equation (A4) is a linear combination of conditional probabilities of the form:

$$g_1(c_{1,0N(x)}, c_{1,1N(x)}, \dots, c_{H+1,0N(D)}, c_{H+1,1N(D)}) / g_2(c_{1,0N(x)}, c_{1,1N(x)}, \dots, c_{H+1,0N(D)}, c_{H+1,1N(D)})$$

where $g_2 \neq 0$ by Assumption 4(g) and g_1 and g_2 are bounded, monotonic functions. Consequently, the right-hand and left-hand limits of g_1 and g_2 exist. Thus, under Assumptions 4c and 4d, as $(N(x), N(D)) \rightarrow \infty$,

$$g_1(c_{1,0N(x)}, c_{1,1N(x)}, \dots, c_{H+1,0N(D)}, c_{H+1,1N(D)}) / g_2(c_{1,0N(x)}, c_{1,1N(x)}, \dots, c_{H+1,0N(D)}, c_{H+1,1N(D)})$$

$$\rightarrow g_1(c_{1,0}, c_{1,1}, \dots, c_{H+1,0}, c_{H+1,1}) / g_2(c_{1,0}, c_{1,1}, \dots, c_{H+1,0}, c_{H+1,1}).$$

Equation (A4) and Assumption 3b thus imply:

$$\lim_{(N(x), N(D)) \rightarrow \infty} E(\tau_i(\theta^j) | X_N) = E(\tau_i(\theta^j) | X) = E(\tau_i(\theta^j)) \quad (A5)$$

Equation (A3), (A5) and the definition of $G_i(x_{iN(x)}\beta^j, x_{iN(x)}\alpha^j, D_{iN(D)} - \gamma^j)$ imply:

$$\lim_{(N(x), N(D)) \rightarrow \infty} E(\tau_{iN}(\theta^j) | X_N) = E(\tau_i(\theta^j)).$$

Proof of Lemma 2

The proof consists of four steps. Steps 1 and 2 specify random variables $\delta(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1N(D)}, \beta^j, \alpha^j) \geq 0$ and $\delta(\mathbf{z}_{H+1, N(D)}, \gamma^j) \geq 0$ such that $|(\hat{\beta}^j, \hat{\alpha}^j) - (\beta^j, \alpha^j)| < \delta(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1, N(x)}, \beta^j, \alpha^j)$ and $|\hat{\gamma}^j - \gamma^j| < \delta(\mathbf{z}_{H+1, N(D)}, \gamma^j)$ imply

$$I_N(x_i, D_i) I[x_{iN(x)}\hat{\beta}^j + x_{iN(x)}\hat{\alpha}^j I(D_{iN(D)} - \hat{\gamma}^j > 0) \geq 0] = I_N(x_i, D_i) I[x_{iN(x)}\beta^j - x_{iN(x)}\alpha^j I(D_{iN(D)} - \gamma^j > 0) \geq 0].$$

Hence Lemma 2 is holds if

$$\lim_{n \rightarrow \infty} P[n^{a(\beta, \alpha, j)} |(\hat{\beta}^j, \hat{\alpha}^j) - (\beta^j, \alpha^j)| < n^{a(\beta, \alpha, j)} \delta(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1, N(D)}, \beta^j, \alpha^j), n^{a(\gamma, j)} |\hat{\gamma}^j - \gamma^j| < n^{a(\gamma, j)} \delta(\mathbf{z}_{H+1, N(D)}, \gamma^j)] = 1$$

which, in turn, given Assumption 1(b) holds if

$$n^{a(\beta, \alpha, j)} \delta(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1, N(D)}, \beta^j, \alpha^j) \xrightarrow{p} \infty \text{ and } n^{a(\gamma, j)} \delta(\mathbf{z}_{H+1, N(D)}, \gamma^j) \xrightarrow{p} \infty \text{ as } n \rightarrow \infty.$$

The latter are shown in Steps 3 and 4.

Step 1: $\exists \delta(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1N(D)}, \beta^j, \alpha^j) \geq 0$ such that

$|(\hat{\beta}^j, \hat{\alpha}^j) - (\beta^j, \alpha^j)| < \delta(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1N(D)}, \beta^j, \alpha^j)$ implies

$$I_N(x_i, D_i) I(x_{iN(x)} \hat{\beta}^j + x_{iN(x)} \hat{\alpha}^j I(D_{iN(D)} - \gamma^j > 0) \geq 0) = \\ I_N(x_i, D_i) I(x_{iN(x)} \beta^j + x_{iN(x)} \alpha^j I(D_{iN(D)} - \gamma^j > 0) \geq 0).$$

If $I_N(x_i, D_i) = 0$, then obviously the result holds for any $\delta \geq 0$. For

$I_N(x_i, D_i) \neq 0$, let

$$\varepsilon(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1N(D)}, \beta^j, \alpha^j) =$$

$$\min_{z_1, \dots, z_H, z_{H+1}, \tilde{x} \in \{\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1N(D)}, \text{support}(\tilde{x})\}} \left| \sum_{h=1}^H z_h \beta^j + \tilde{x} \tilde{\beta}^j + \left(\sum_{h=1}^H z_h \alpha^j + \tilde{x} \tilde{\alpha}^j \right) I(z_{H+1} > \gamma^j) \right|$$

Given the definitions of $x_{iN(x)}$ and $D_{iN(D)}$, we have:

$$P[|x_{iN(x)} \beta^j + x_{iN(x)} \alpha^j I(D_{iN(D)} - \gamma^j > 0)| \geq \varepsilon(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1N(D)}, \beta^j, \alpha^j)] = 1 \quad (\text{A6})$$

Now

$$|x_{iN(x)}| \leq |x_{1N(x)}| + \dots + |x_{HN(x)}| + |\tilde{x}_i| \leq c_{1N(x)}^* + \dots + c_{HN(x)}^* + \max\{|\tilde{x}|\} \equiv \Delta_{N(x)} \quad (\text{A7})$$

where $c_{hN(x)}^* = \max(|c_{h0N(x)}|, |c_{h1N(x)} + (c_{h1N(x)} - c_{h0N(x)})/N(x)|)$ and $\Delta_{N(x)} > 0$.

Define

$$\delta(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1N(D)}, \beta^j, \alpha^j) \equiv \varepsilon(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1N(D)}, \beta^j, \alpha^j) / \Delta_{N(x)}.$$

Clearly,

$$|(\hat{\beta}^j, \hat{\alpha}^j I[D_{iN(D)} > \gamma^j]) - (\beta^j, \alpha^j I[D_{iN(D)} > \gamma^j])| \leq |(\hat{\beta}^j, \hat{\alpha}^j) - (\beta^j, \alpha^j)|.$$

Therefore, $|(\hat{\beta}^j, \hat{\alpha}^j) - (\beta^j, \alpha^j)| < \delta(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1N(D)}, \beta^j, \alpha^j)$ and (A7) imply

$$|x_{iN(x)} \hat{\beta}^j + x_{iN(x)} \hat{\alpha}^j I(D_{iN(D)} - \gamma^j > 0) - x_{iN(x)} \beta^j - x_{iN(x)} \alpha^j I(D_{iN(D)} - \gamma^j > 0)| <$$

$$\varepsilon(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1N(D)}, \beta^j, \alpha^j)$$

which, given (A6), implies

$$I(x_{iN(x)}\hat{\beta}^j + x_{iN(x)}\hat{\alpha}^j I(D_{iN(D)} - \gamma^j > 0) \geq 0) =$$

$$I(x_{iN(x)}\beta^j + x_{iN(x)}\alpha^j I(D_{iN(D)} - \gamma^j > 0) \geq 0).$$

Step 2: $\exists \delta(\mathbf{z}_{H+1,N(D)}, \gamma^j) \geq 0$ such that $|\hat{\gamma}^j - \gamma^j| < \delta(\mathbf{z}_{H+1,N(D)}, \gamma^j)$ implies $I(D_{iN(D)} - \hat{\gamma}^j > 0) = I(D_{iN(D)} - \gamma^j > 0)$.

Proof is similar to Step 1 with $\delta(\mathbf{z}_{H+1,N(D)}, \gamma^j) = \frac{\varepsilon(\mathbf{z}_{H+1,N(D)}, \gamma^j)}{\Delta_{N(D)}}$,

$$\varepsilon(\mathbf{z}_{H+1,N(D)}, \gamma^j) \equiv \min_{z_{H+1,N(D)} \in \mathbf{z}_{H+1,N(D)}} |z_{H+1,N(D)} - \gamma^j|, \text{ and}$$

$$\Delta_{N(D)} = c_{H+1,N(D)}^* \equiv \max(|c_{H+1,0N(D)}|, |c_{H+1,1N(D)} + (c_{H+1,1N(D)} - c_{H+1,0N(D)})/N(D)|).$$

Step 3: $n^{a(\beta, \alpha, j)} \delta(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1,N(D)}, \beta^j, \alpha^j) \xrightarrow{p} \infty$

Recall

$$\delta(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1,N(D)}, \beta^j, \alpha^j) \equiv \varepsilon(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1,N(D)}, \beta^j, \alpha^j) / \Delta_{N(x)},$$

Letting $\mathbf{z}_{-1N} = (\mathbf{z}_{2N(x)}, \dots, \mathbf{z}_{HN(x)})$, we can write:

$$\begin{aligned} \varepsilon(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1,N(D)}, \beta^j, \alpha^j) &= \min \{ |z_{11N(x)}(\beta_1^j + \alpha_1^j I[z_{H+1,N(D)}^{(1)} > \gamma^j]) + \mathbf{z}_{-1N(x)}(\beta_{-1,z}^j + \alpha_{-1,z}^j I[z_{H+1,N(D)}^{(1)} > \gamma^j]) \\ &\quad + \tilde{x}_{(1)}(\tilde{\beta}^j + \tilde{\alpha}^j I[z_{H+1,N(D)}^{(1)} > \gamma^j])|, \dots, \\ &\quad |z_{1N(x)N(x)}(\beta_1^j + \alpha_1^j I[z_{H+1,N(D)}^{(N(x))} > \gamma^j]) + \mathbf{z}_{-1N(x)}(\beta_{-1,z}^j + \alpha_{-1,z}^j I[z_{H+1,N(D)}^{(N(x))} > \gamma^j]) \\ &\quad + \tilde{x}_{(N(x))}(\tilde{\beta}^j + \tilde{\alpha}^j I[z_{H+1,N(D)}^{(N(x))} > \gamma^j])| \} \end{aligned}$$

where, for example, $z_{H+1,N(D)}^{(1)}$, $z_{-1N(x)}$ and $\tilde{x}_{(1)}$ are such that

$$\begin{aligned} |z_{11N(x)}(\beta_1^j + \alpha_1^j I[z_{H+1,N(D)}^{(1)} > \gamma^j]) + \mathbf{z}_{-1N(x)}(\beta_{-1,z}^j + \alpha_{-1,z}^j I[z_{H+1,N(D)}^{(1)} > \gamma^j]) + \tilde{x}_{(1)}(\tilde{\beta}^j + \tilde{\alpha}^j I[z_{H+1,N(D)}^{(1)} > \gamma^j])| = \\ \min_{(z_{-1N(x)}, z_{H+1,N(D)}, \tilde{x}) \in (z_{-1N(x)}, z_{H+1,N(D)}, \text{support}(\tilde{x}))} |z_{11N(x)}(\beta_1^j + \alpha_1^j I[z_{H+1,N(D)}^{(1)} > \gamma^j]) + \mathbf{z}_{-1N(x)}(\beta_{-1,z}^j + \alpha_{-1,z}^j I[z_{H+1,N(D)}^{(1)} > \gamma^j]) \\ + \tilde{x}(\tilde{\beta}^j + \tilde{\alpha}^j I[z_{H+1,N(D)}^{(1)} > \gamma^j])| \end{aligned}$$

The random variables $z_{11N(x)}, \dots, z_{1N(x)N(x)}$ are independent conditional on $\Omega_N \equiv (\mathbf{z}_{-1N(x)}^{(1)}, \dots, \mathbf{z}_{-1N(x)}^{(N(x))}, \mathbf{z}_{H+1,N(D)}^{(1)}, \dots, \mathbf{z}_{H+1,N(D)}^{(N(x))}, \tilde{x}_{(1)}, \dots, \tilde{x}_{(N(x))})$ under Assumption 5. Thus,

$$P(\varepsilon(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1,N(D)}, \beta^j, \alpha^j) \geq \lambda \mid \Omega_N) = \prod_{i=1}^{N(x)} [1 - P(|z_{iN(x)}(\beta_1^j + \alpha_1^j I[z_{H+1,N(D)}^{(i)} > \gamma^j]) + \mathbf{z}_{-1N(x)}^{(i)}(\beta_{-1,z}^j + \alpha_{-1,z}^j I[z_{H+1,N(D)}^{(i)} > \gamma^j]) + \tilde{x}_{(i)}(\tilde{\beta}^j + \tilde{\alpha}^j I[z_{H+1,N(D)}^{(i)} > \gamma^j])| < \lambda \mid \Omega_N)] \quad (\text{A8})$$

Since under Assumptions 4a, 4g and 5, $z_{iN(x)} \mid \Omega_N$ is uniformly distributed on the interval $(c_{10N(x)} + (c_{11N(x)} - c_{10N(x)})i / N, c_{10N(x)} + (c_{11N(x)} - c_{10N(x)})(i+1) / N(x))$, and $\beta_1^j + \alpha_1^j I[z_{H+1,N(D)}^{(i)} > \gamma^j] \neq 0$ by Assumption 3a, we have:

$$\begin{aligned} & P(|z_{iN(x)}(\beta_1^j + \alpha_1^j I[z_{H+1,N(D)}^{(i)} > \gamma^j]) + \mathbf{z}_{-1N(x)}^{(i)}(\beta_{-1,z}^j + \alpha_{-1,z}^j I[z_{H+1,N(D)}^{(i)} > \gamma^j]) \\ & \quad + \tilde{x}_{(i)}(\tilde{\beta}^j + \tilde{\alpha}^j I[z_{H+1,N(D)}^{(i)} > \gamma^j])| < \lambda \mid \Omega_N) \\ & = \\ & P(-\lambda < z_{iN(x)}(\beta_1^j + \alpha_1^j I[z_{H+1,N(D)}^{(i)} > \gamma^j]) + \mathbf{z}_{-1N(x)}^{(i)}(\beta_{-1,z}^j + \alpha_{-1,z}^j I[z_{H+1,N(D)}^{(i)} > \gamma^j]) \\ & \quad + \tilde{x}_{(i)}(\tilde{\beta}^j + \tilde{\alpha}^j I[z_{H+1,N(D)}^{(i)} > \gamma^j]) < \lambda \mid \Omega_N) \\ & \leq \\ & \min[2\lambda / (\beta_1^j + \alpha_1^j I[z_{H+1,N(D)}^{(i)} > \gamma^j]), (c_{11N(x)} - c_{10N(x)}) / N(x)] \div (c_{11N(x)} - c_{10N(x)}) / N(x) \\ & \leq \min[2\lambda N(x) / (c_{11N(x)} - c_{10N(x)})(\beta_1^j + \alpha_1^j I[z_{H+1,N(D)}^{(i)} > \gamma^j]), 1] \quad (\text{A9}) \end{aligned}$$

(A8) and (A9) imply

$$\begin{aligned} & P(\varepsilon(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1,N(D)}, \beta^j, \alpha^j) \geq \lambda \mid \Omega_N) \geq \\ & \prod_{i=1}^N [1 - \min[2\lambda N(x) / (c_{11N(x)} - c_{10N(x)})(\beta_1^j + \alpha_1^j I[z_{H+1,N(D)}^{(i)} > \gamma^j]), 1]] \geq \\ & I[\alpha_1^j > 0][1 - \min[2\lambda N(x) / (c_{11N(x)} - c_{10N(x)})(\beta_1^j + \alpha_1^j), 1]]^{N(x)} + I[\alpha_1^j \leq 0][1 - \min[2\lambda N(x) / (c_{11N(x)} - c_{10N(x)})\beta_1^j, 1]]^{N(x)} \quad (\text{A10}) \end{aligned}$$

Therefore, for all finite λ' :

$$\begin{aligned}
 &P(\delta(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1, N(D)}, \beta^j, \alpha^j)N(x)^d \geq \lambda' | \Omega_N) \geq \\
 &P(\varepsilon(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1, N(D)}, \beta^j, \alpha^j) \geq \Delta_{N(x)} \lambda' N(x)^{-d} | \Omega_N) = \\
 &I(\alpha_1^j > 0)(1 - \min[\rho_+(N(x))/N(x), 1])^{N(x)} + I(\alpha_1^j \leq 0)(1 - \min[\rho_-(N(x))/N(x), 1])^{N(x)}
 \end{aligned}
 \tag{A11}$$

where $\rho_+(N(x)) = 2\Delta_{N(x)} \lambda' N(x)^{2-d} / (c_{11N(x)} - c_{10N(x)})(\beta_1^j + \alpha_1^j)$ and $\rho_-(N(x)) = 2\Delta_{N(x)} \lambda' N(x)^{2-d} / (c_{11N(x)} - c_{10N(x)})\beta_1^j$. By Assumption 6, $N(x) \rightarrow \infty$ as $n \rightarrow \infty$. Therefore, by Assumptions 4e and 4f, $\rho_-(N(x)) \rightarrow 0$ and $\rho_+(N(x)) \rightarrow 0$ as $n \rightarrow \infty$ for $d > 2$. It follows that the RHS of (A11) converges to 1 for $d > 2$. Since λ' is arbitrary and the RHS of (A11) does not depend on the conditioning variables, this implies $\delta(\mathbf{z}_{1N(x)}, \dots, \mathbf{z}_{H+1, N(D)}, \beta^j, \alpha^j)N(x)^d \xrightarrow{p} \infty$ for $d > 2$. Since $\lim n^{\alpha(\beta, \alpha, j)} / N(x)^d > 0$ for some $d > 2$ under Assumption 6, Step 3 holds.

Step 4: $n^{\alpha(\gamma, j)} \delta(\mathbf{z}_{H+1, N(D)}, \gamma^j) \xrightarrow{p} \infty$ as $n \rightarrow \infty$

Proof is completely analogous to Step 3.

Proof of Theorem 1 Let $\sigma_{SN} \equiv \sqrt{\text{var}(n^{-1/2} \sum_{i=1}^n \tau_{iN}(\theta^U, \theta^R) | X_N)}$. By Lemma 2

we can write:

$$T_n = \frac{\sigma_S \sigma_{SN}}{\hat{\sigma}_S \sigma_S} T_n^* + o_p(1) \tag{A12}$$

where

$$T_n^* = T_n^{**} + \sigma_{SN}^{-1} n^{1/2} E(\tau_{iN}(\theta^U, \theta^R) | X_N) \tag{A13}$$

$$\text{and } T_n^{**} = n^{-1/2} \sigma_{SN}^{-1} \sum_{i=1}^n [\tau_{iN}(\theta^U, \theta^R) - E(\tau_{iN}(\theta^U, \theta^R) | X_N)]$$

Proof (i): $\theta^U = \theta^R$ implies $E(\tau_{iN}(\theta^U, \theta^R) | X_N) = 0$ and, thus, by (A12) and (A13):

$$T_n = \frac{\sigma_S \sigma_{SN}}{\hat{\sigma}_S \sigma_S} T_n^{**} + o_p(1) \quad \text{where } T_n^{**} = n^{-1/2} \sigma_{SN}^{-1} \sum_{i=1}^n \tau_{iN}(\theta^U, \theta^R) \tag{A14}$$

Step 1 $\sigma_{SN} - \sigma_S \rightarrow 0$ and $\frac{\sigma_S \sigma_{SN}}{\hat{\sigma}_S \sigma_S} \rightarrow 1$ as $n \rightarrow \infty$

Theorem 1 assumes $\hat{\sigma}_S - \sigma_S \xrightarrow{p} 0$. Under Assumption 6, $(N(x), N(D)) \rightarrow \infty$ as $n \rightarrow \infty$. Hence, under Assumptions 1, 2, 3, 4 and 5, the same line of argument used to prove Lemma 1 can be used to show that: $\sigma_{SN} - \sigma_S \rightarrow 0$.

Step 2: As $n \rightarrow \infty$: $T_n^{**} \xrightarrow{d} N(0,1)$.

Let $g_n(\lambda)$ denote the characteristic function of T_n^{**} , and let

$g_n(\lambda | X_N) = E(e^{i\lambda T_n^{**}} | X_N)$. By iterated expectations, $g_n(\lambda) = E[g_n(\lambda | X_N)]$.

Hence, by dominated convergence it suffices to show $g_n(\lambda | X_N)$ converges to $e^{-\lambda^2/2}$ as $n \rightarrow \infty$ for almost every X_N .

The latter follows from arguments similar to those used to prove the Lindeberg-Levy Central Limit Theorem. Under our assumptions, T_n^{**} is a sum of n i.i.d. variables with zero mean *conditional* on X_N . Let

$g(\lambda | X_N) = E(e^{i\lambda \tau_N} | X_N)$. Therefore,

$$g_n(\lambda | X_N) = [g(\lambda / \sigma_{SN} \sqrt{n} | X_N)]^n \quad (\text{A15})$$

A Taylor expansion about zero and properties of the characteristic function yield:

$$g(\lambda / \sigma_{SN} \sqrt{n} | X_N) = 1 - \lambda^2 / 2n + \lambda^3 \sigma_{SN}^{-3} n^{-3/2} 3^{-1} g'''(\xi) \quad (\text{A16})$$

where $g'''(\xi)$ denotes the third derivative and ξ is on the interval connecting 0 and $\lambda / \sigma_{SN} \sqrt{n}$. The third term in (A16) is $o(1)$ since $\sigma_S > 0$ (by Assumption 2), and since τ_N and, thus $g'''(\xi)$ is bounded for all real ξ . It therefore follows from (A15) and (A16) that $g_n(\lambda | X_N)$ converges to $e^{-\lambda^2/2}$ as $n \rightarrow \infty$.

Step 3: As $n \rightarrow \infty$: $T_n \xrightarrow{d} N(0,1)$

Follows from Steps 1 and 2 and (A14).

Proof (ii): Arguments in Step 2 of the proof of Part (i) can be used to show that

$T_n^{**} \xrightarrow{d} N(0,1)$ for $\theta^U \neq \theta^R$. It follows from the proof of Lemma 1 (see the

equation below A5) that $\lim_{N \rightarrow \infty} E(\tau_{iN}(\theta^U, \theta^R) | X_N) = E(\tau_i(\theta^U, \theta^R))$. Part (ii) of Theorem 1 follows from this, and (A13).

Proof of Theorem 2: Let $\sigma_{SN} \equiv \sqrt{\text{var}(m^{-1/2} \sum_{i=1}^m \tau_{iN}(\theta^U, \theta^R) | X_N)}$ and write:

$$T_m = \frac{\sigma_S \sigma_{SN}}{\hat{\sigma}_S \sigma_S} T_m^* + o_p(1)$$

(A17)

where

$$T_m^* = T_m^{**} + \sigma_{SN}^{-1} m^{1/2} E(\tau_{iN}(\theta^U, \theta^R) | X_N)$$

(A18)

$$\text{and } T_m^{**} = m^{-1/2} \sigma_{SN}^{-1} \sum_{i=1}^m [\tau_{iN}(\theta^U, \theta^R) - E(\tau_{iN}(\theta^U, \theta^R) | X_N)]$$

Proof (i): Step 1 $\sigma_{SN} - \sigma_S \rightarrow 0$ and $\frac{\sigma_S \sigma_{SN}}{\hat{\sigma}_S \sigma_S} \rightarrow 1$ as $m \rightarrow \infty$

Theorem 2 assumes $\hat{\sigma}_S - \sigma_S \xrightarrow{p} 0$. Under Assumption 7, $N \rightarrow \infty$ as $m \rightarrow \infty$.

Hence, under Assumptions 4a, 5 and 6, the same line of argument used to prove Lemma 1 can be used to show that: $\sigma_{SN} - \sigma_S \rightarrow 0$.

Step 2: $m^{1/2} E(\tau_{iN}(\theta^U, \theta^R) | X_N) \rightarrow 0$ as $m \rightarrow \infty$.

Under Assumption 7 it suffices to show

$$E(\tau_{iN}(\theta^U, \theta^R) | X_N) = O_p(C_{N(x)} N(x)^{-1}) + O_p((c_{H+1,1,N(D)} - c_{H+1,0,N(D)}) N(D)^{-1})$$

Part (i) of Theorem 2 assumes:

$$E[S_m(\theta^U) - S_m(\theta^R) | X_N] = O_p(C_{N(x)} N(x)^{-1}) + O_p((c_{H+1,1,N(D)} - c_{H+1,0,N(D)}) N(D)^{-1})$$

(A19)

Under Assumption 1a, (A19) implies

$$E(\tau_i(\theta^U, \theta^R) | X_N) = O_p(C_{N(x)} N(x)^{-1}) + O_p((c_{H+1,1,N(D)} - c_{H+1,0,N(D)}) N(D)^{-1}),$$

and thus,

$$E(\tau_{iN}(\theta^U, \theta^R) | X_N) = E(\tau_{iN}(\theta^U, \theta^R) | X_N) - E(\tau_i(\theta^U, \theta^R) | X_N)$$

$$+ O_p(C_{N(x)} N(x)^{-1}) + O_p((c_{H+1,1,N(D)} - c_{H+1,0,N(D)}) N(D)^{-1})$$

$$\begin{aligned}
&= E(W^U)[P(y_i = 1, x_{iN}\beta^U + x_{iN}\alpha^U I(D_{iN} > \gamma^U) < 0 | X_N) \\
&\quad - P(y_i = 1, x_i\beta^U + x_i\alpha^U I(D_i > \gamma^U) < 0 | X_N)] + \dots \\
&\quad + O_p(C_{N(x)}N(x)^{-1}) + O_p((c_{H+1,1,N(D)} - c_{H+1,0,N(D)})N(D)^{-1}) \quad (A20)
\end{aligned}$$

Now

$$\begin{aligned}
&P(y_i = 1, x_{iN}\beta^U + x_{iN}\alpha^U I(D_{iN} > \gamma^U) < 0 | X_N) = \\
&\quad P(y_i = 1, x_{iN}\beta^U + x_{iN}\alpha^U < 0, x_i\beta^U + x_i\alpha^U < 0, D_{iN} > \gamma^U | X_N) \\
&\quad + P(y_i = 1, x_{iN}\beta^U + x_{iN}\alpha^U < 0, x_i\beta^U + x_i\alpha^U > 0, D_{iN} > \gamma^U | X_N) \quad (A21) \\
&\quad + P(y_i = 1, x_{iN}\beta^U < 0, x_i\beta^U < 0, D_{iN} \leq \gamma^U | X_N) \\
&\quad + P(y_i = 1, x_{iN}\beta^U < 0, x_i\beta^U > 0, D_{iN} \leq \gamma^U | X_N)
\end{aligned}$$

$$\begin{aligned}
&P(y_i = 1, x_i\beta^U + x_i\alpha^U I(D_i > \gamma^U) < 0 | X_N) = \\
&\quad P(y_i = 1, x_i\beta^U + x_i\alpha^U < 0, x_{iN}\beta^U + x_{iN}\alpha^U < 0, D_i > \gamma^U | X_N) \\
&\quad + P(y_i = 1, x_i\beta^U + x_i\alpha^U < 0, x_{iN}\beta^U + x_{iN}\alpha^U > 0, D_i > \gamma^U | X_N) \quad (A22) \\
&\quad + P(y_i = 1, x_i\beta^U < 0, x_{iN}\beta^U < 0, D_i \leq \gamma^U | X_N) \\
&\quad + P(y_i = 1, x_i\beta^U < 0, x_{iN}\beta^U > 0, D_i \leq \gamma^U | X_N)
\end{aligned}$$

We first show that the difference between the first term on the RHS of (A21) and the first term on the RHS of (A22) is $O_p((c_{H+1,1,N(D)} - c_{H+1,0,H(D)})N(D)^{-1})$. For any event A, we have

$$\begin{aligned}
&P(A, D_{iN} > \gamma^U | X_N) - P(A, D_i > \gamma^U | X_N) = \\
&\quad P(A, D_{iN} > \gamma^U, D_i \leq \gamma^U | X_N) - P(A, D_i > \gamma^U, D_{iN} \leq \gamma^U | X_N) \quad (A23)
\end{aligned}$$

By (A1), $P(|D_{iN(D)} - D_i| < 2(c_{H+1,1,N(D)} - c_{H+1,0,H(D)})N(D)^{-1} | X_N) = 1$.

Therefore,

$$\begin{aligned}
&\{D_{iN} - \gamma^U \leq 0, D_i - \gamma^U > 0 | X_N\} \Leftrightarrow \\
&\{D_{iN(D)} - \gamma^U \leq 0, D_i - \gamma^U > 0, |D_{iN(D)} - D_i| < 2(c_{H+1,1,N(D)} - c_{H+1,0,N(D)})N(D)^{-1} | X_N\} \\
&\Rightarrow \{0 < D_i - \gamma^U \leq D_i - \gamma^U - (D_{iN(D)} - \gamma^U) \leq 2(c_{H+1,1,N(D)} - c_{H+1,0,N(D)})N(D)^{-1} | X_N\} \\
&\Rightarrow \{0 < D_i - \gamma^U \leq 2(c_{H+1,1,N(D)} - c_{H+1,0,N(D)})N(D)^{-1} | X_N\}
\end{aligned}$$

and, therefore,

$$\begin{aligned} & P(A, D_{iN} > \gamma^U, D_i \leq \gamma^U | X_N) \\ & \leq P(0 < D_i - \gamma^U \leq 2(c_{H+1,1,N(D)} - c_{H+1,0,N(D)})N(D)^{-1} | X_N) \\ & = O_p(c_{H+1,1,N(D)} - c_{H+1,0,N(D)})N(D)^{-1} \end{aligned}$$

where the last equality follows from Assumption 3a. The same line of argument applies to the second term on the RHS of (A23). Consequently, the difference between the first term on the RHS of (A21) and the first term on the RHS of (A22) is $O_p((c_{H+1,1,N(D)} - c_{H+1,0,H(D)})N(D)^{-1})$. The same argument can be applied to the difference between the third term on the RHS of (A21) and the third term on the RHS of (A22).

The second term on the RHS of (A21) has the form:

$$P(A, x_{iN}\beta^U + x_{iN}\alpha^U < 0, x_i\beta^U + x_i\alpha^U > 0 | X_N)$$

By (A1), for a finite $\kappa > 0$:

$$P(|x_{iN}\beta^U + x_{iN}\alpha^U - x_i\beta^U - x_i\alpha^U| \leq \kappa C_{N(x)} N(x)^{-1} | X_N) = 1.$$

Consequently,

$$\{x_{iN}\beta^U + x_{iN}\alpha^U < 0, x_i\beta^U + x_i\alpha^U > 0 | X_N\} \Leftrightarrow$$

$$\begin{aligned} & \{x_{iN}\beta^U + x_{iN}\alpha^U < 0, x_i\beta^U + x_i\alpha^U > 0, \\ & |x_{iN}\beta^U + x_{iN}\alpha^U - x_i\beta^U - x_i\alpha^U| \leq \kappa C_{N(x)} N(x)^{-1} | X_N\} \end{aligned}$$

$$\Rightarrow \{0 < x_i\beta^U + x_i\alpha^U \leq x_i\beta^U + x_i\alpha^U - x_{iN}\beta^U - x_{iN}\alpha^U \leq \kappa C_{N(x)} N(x)^{-1} | X_N\}$$

$$\Rightarrow \{0 < x_i\beta^U + x_i\alpha^U \leq \kappa C_{N(x)} N(x)^{-1} | X_N\}$$

and, therefore,

$$\begin{aligned} & P(A, x_{iN}\beta^U + x_{iN}\alpha^U < 0, x_i\beta^U + x_i\alpha^U > 0 | X_N) \leq \\ & P(0 < x_i\beta^U + x_i\alpha^U \leq \kappa C_{N(x)} N(x)^{-1} | X_N) = O(C_{N(x)} N(x)^{-1}) \end{aligned}$$

where the last equality follows from Assumptions 3a and 5b. Therefore, the second term on the RHS of (A21) is $O(C_{N(x)} N(x)^{-1})$. The same line of argument can be applied to fourth term on the RHS of (A21), and the second and fourth

term of the RHS of (A22). Consequently, the first two terms on the RHS of (A20) are $O_p(C_{N(x)}N(x)^{-1}) + O_p((c_{H+1,1,N(D)} - c_{H+1,0,H(D)})N(D)^{-1})$. Analogous arguments apply to the remaining terms on the RHS of (A20).

Step 3: As $m \rightarrow \infty$: $T_m^{**} \xrightarrow{d} N(0,1)$

Similar to Step 2 in the proof of Theorem 1.

Step 4: As $m \rightarrow \infty$: $T_m \xrightarrow{d} N(0,1)$ if $E[S_m(\theta^U) - S_m(\theta^R)] | X_N = O_p(C_N N^{-1})$

Follows from Steps 2 and 3, and equations (A17) and (A18).

Proof (ii) By proof of part (ii) of Lemma 1 (see the equation below A4), $\lim_{N \rightarrow \infty} E(\tau_{iN}(\theta^U, \theta^R) | X_N) = E(\tau_i(\theta^U, \theta^R))$. Part (ii) of Theorem 1 follows from this, and equation (A18).

Appendix 2

Clearly γ is not identified when $\alpha = 0$. To test the latter, our test requires:

$$\text{plim}(\hat{\beta}^U, \hat{\alpha}^U) = (\beta, \alpha) \text{ under Ho: } \alpha = 0 \quad (\text{A24})$$

Since $\alpha = 0$ violates the Assumption 1 of Lee and Seo (2008), their results do not apply. In this appendix, we show that (A24) in fact holds. Let $S_n(\beta, \alpha, \gamma)$ denote the sample score function, and $S(\beta, \alpha, \gamma)$ denote $E(S_n(\beta, \alpha, \gamma))$. Suppose $\alpha = 0$ but otherwise Assumptions 1-7 of Lee and Seo (2008) hold. Let Θ denote the parameter space for (β, α, γ) , and define:

$$\Pi = \{(\beta^*, \alpha^*, \gamma^*) \in \Theta : S(\beta^*, \alpha^*, \gamma^*) > S(\beta', \alpha', \gamma'), \forall (\beta', \alpha', \gamma') \neq (\beta^*, \alpha^*, \gamma^*), (\beta', \alpha', \gamma') \in \Theta\}$$

Given $\alpha = 0$, Manski (1985) implies $\Pi = \{(\beta, 0, \gamma^*), \gamma^* \in \Theta\}$. Let $\Pi^c \subset \Theta$

denote the complement of Π , $(\beta', \alpha', \gamma') \in \Pi^c$, $\gamma \in \Theta$,

$\eta \equiv S(\beta, 0, \gamma) - S(\beta', \alpha', \gamma') > 0$, and let A_n denote the event:

$$|S_n(\beta'', \alpha'', \gamma'') - S(\beta'', \alpha'', \gamma'')| < \eta/2 \quad \forall (\beta'', \alpha'', \gamma'') \in \Theta.$$

Now A_n implies:

$$S(\hat{\beta}^U, \hat{\alpha}^U, \hat{\gamma}^U) > S_n(\hat{\beta}^U, \hat{\alpha}^U, \hat{\gamma}^U) - \eta/2 \quad (\text{A25})$$

$$S_n(\beta, 0, \gamma) > S(\beta, 0, \gamma) - \eta/2 \quad (\text{A26})$$

It follows from (A24) and the definition of $(\hat{\beta}^U, \hat{\alpha}^U, \hat{\gamma}^U)$ that

$$S(\hat{\beta}^U, \hat{\alpha}^U, \hat{\gamma}^U) > S_n(\beta, 0, \gamma) - \eta/2 \quad (\text{A27})$$

Adding (A26) and (A27) we get

$$S(\hat{\beta}^U, \hat{\alpha}^U, \hat{\gamma}^U) > S(\beta, 0, \gamma) - \eta = S(\beta', \alpha', \gamma') \quad \forall (\beta', \alpha', \gamma') \in \Pi^c$$

Therefore A_n implies $(\hat{\beta}^U, \hat{\alpha}^U, \hat{\gamma}^U) \in \Pi$. Since $\lim_{n \rightarrow \infty} P(A_n) = 1$ by Newey and McFadden (1994, Theorem 2.1) (see Lee and Seo 2008, p.496), it follows that $\lim_{n \rightarrow \infty} P[(\hat{\beta}^U, \hat{\alpha}^U, \hat{\gamma}^U) \in \Pi] = 1$.

References

- Abrevaya, J. and J. Huang (2005): "On the Bootstrap of the Maximum Score Estimator," *Econometrica* 73-4, 1175-1204.
- Bain, L.E and M. Engelhardt (1992): *Introduction to Probability and Mathematical Statistics*, Pacific Grove, CA: Duxbury Classic Series.
- Bierens, H.J., (1994): *Topics in Advanced Econometrics*, Cambridge, New York: Cambridge University Press.
- Cosslett, S.R. (1983): "Distribution-free Estimators of the Binary Choice Model," *Econometrica* 55, 559-585.
- Delgado, M.A., Rodriguez-Poo, J.M. and M. Wolf (2001): "Subsampling Inference in Cube Root Asymptotics with an Application to Manski's Maximum Score Estimator," *Economic Letters*, 73-2, 241-250.
- Han A.K. (1987): "Nonparametric Analysis of a Generalized Regression Model: The Rank Correlation Estimator," *Journal of Econometrics* 35, 303-316.
- Hidalgo, J. (1999): "Nonparametric Tests for Model Selection with Time Series Data," *Test* 8, 365-398.
- Hoderlein, S. (2009): "Endogenous Semiparametric Binary Choice Models with Heteroscedasticity," Cemmap-Working Paper CWP34/09.
- Horowitz, J.L., (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505-531.
- Horowitz, J.L. (2009): "Specification Testing in Nonparametric Instrumental Variables Estimation," working paper, Northwestern University.
- Horowitz, J.L. (2009b): *Semiparametric and Nonparametric Methods in Econometrics*. Springer: New York.
- Ichimura, H. (1993): "Semiparametric Least squares (SLS) and Weighted SLS Estimation of Single Index Models," *Journal of Econometrics* 58, 71-120.
- Kim, J. and D. Pollard (1990): "Cube Root Asymptotics," *Annals of Statistics* 18-1, 191-219.
- Klein, R.L. and R.H. Spady (1993): "An Efficient Semiparametric Estimator for Discrete Choice Models," *Econometrica* 61, 387-421.

1
2
3 Lee, S.M.S. and M.C. Pun (2006): "On m out of n Bootstrapping for Nonstandard
4 m-Estimation with Nuisance Parameters," *Journal of the American Statistical*
5 *Association* 101, 1185-1197.
6

7
8 Lee, S. and M.H. Seo (2008): "Semiparametric Estimation of a Binary Response
9 Model with a Change-Point due to Covariate Threshold," *Journal of*
10 *Econometrics* 144, 492-499.
11

12
13 Lewbel, A. (2000): "Semiparametric qualitative response model estimation with
14 unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*
15 97-1, pp. 145-177.
16

17
18 Manski, C.F. (1975): "Maximum Score Estimation of the Stochastic Utility
19 Model of Choice," *Journal of Econometrics* 3, 205-228.
20

21
22 Manski, C.F., (1985): "Semiparametric Analysis of Discrete Response:
23 Asymptotic Properties of the Maximum Score Estimator," *Journal of*
24 *Econometrics*, 27, 313-333.
25

26
27 Mayer, W.J. and R.E. Dorsey (1998): "Maximum Score Estimation of
28 Disequilibrium Models and the Role of Anticipatory Price-Setting," *Journal of*
29 *Econometrics* 87, 1-24.
30

31
32 Mayer, W. J. (1999): "An Extension of the Maximum Score Estimator for
33 Disequilibrium Models," *Economics Letters* 64-2, 143-149.
34

35
36 Newey, W.K. and D. McFadden (1994): "Large Sample Estimation and
37 Hypothesis Testing" in *Handbook of Econometrics* Volume IV, North-Holland:
38 Amsterdam.
39

40
41 Robinson, P.M. (1991): "Consistent Nonparametric Entropy-Based Testing,"
42 *Review of Economic Studies* 58, 437-453.
43

44
45 Seo, M. and O. Linton (2007): "A Smoothed Least Squares Estimator for the
46 Threshold Regression," *Journal of Econometrics* 141, 704-735.
47

48
49 Whang, Y. and W.K. Andrews (1993): Tests of Specification for Parametric and
50 Semiparametric Models," *Journal of Econometrics* 57, 277-318.
51

52
53 Vuong, Q.H. (1989): "Likelihood Ratio Tests for Model Selection and Non-
54 nested Hypotheses," *Econometrica* 57-2, 307-333.
55

56
57 Yatchew, A.D. (1992): "Nonparametric Regression Tests based on Least
58 Squares," *Econometric Theory* 8, 435-451.
59
60

Table 1

Estimated Size: Two-Tailed Rejection Percentages for T_n using 5% Asymptotic Critical Values under $H_o : \beta_2 = 0$

Split-Sample weights

n =300	N=18	N=180	N=300
T3	17.7%	17.2%	17.0%
Logistic	16.0%	16.0%	16.0%

n= 1000	N=22	N=215	N=357
T3	14.8%	14.4%	6.2%
Logistic:	16.3%	18.1%	6.2%

Normally distributed weights

n =300	N=18	N=180	N=300
T3	5.2%	7.1%	5.7%
Logistic	6.9%	6.6%	6.3%

n= 1000	N=22	N=215	N=357
T3	7.2%	6.2%	6.4%
Logistic:	6.3%	5.5%	6.7%

1000 replications; n= sample size, N=number of draws to approximate the continuous covariate.

Table 2

Estimated Power: Two-Tailed Rejection Percentages for T_n using 5% Asymptotic Critical Values under $H_a : \beta_2 \neq 0$ for Logistic Errors

n=300	Split-sample weights			Normal weights		
	N=18	N=180	N=300	N=18	N=180	N=300
$\beta_2 = -3$	76.0%	76.0%	77.0%	98.8%	99%	99.1%
$\beta_2 = -1$	22.6%	21.2%	22.0%	23.4%	23.8%	26.1%
$\beta_2 = 1$	15.0%	16.0%	16.1%	8.6%	10.4%	12.6%
$\beta_2 = 3$	21.3%	22.0%	21.1%	22.3%	26%	24.1%

n=1000	Split-sample weights			Normal weights		
	N=22	N=215	N=357	N=22	N=215	N=357
$\beta_2 = -3$	99.0%	99.5%	98.6%	100%	100%	100%
$\beta_2 = -1$	23.0%	23.0%	10.0%	36.7%	44%	45.5%
$\beta_2 = 1$	17.0%	19.0%	10.0%	13.3%	16.2%	19.5%
$\beta_2 = 3$	30.5.5%	33.1%	39.7%	45.8%	55.4%	58.5%

1000 replications; n= sample size, N=number of draws to approximate the continuous covariate.

Table 3

Estimated Power and Size for $H_o : \varepsilon \square Dist - L$ versus

$H_a : \varepsilon$ not $\square Dist - L$

Two-Tailed Rejection Percentages for T_n using 5% Asymptotic
Critical Values

n=300	N=18	N=180	N=300
$H_a : dist - H$	20.6%	30.9%	29.8%
$H_o : dist - L$	7.3%	8.5%	9.3%

n=1000	N=22	N=215	N=357
$H_a : dist - H$	43.2%	52.3%	54.2%
$H_o : dist - L$	6.0%	5.4%	6.7%

1000 replications; n= sample size, N=number of draws to approximate the
continuous covariate; Dist-H:

$\varepsilon = 0.25[1+2(x\beta)^2+(x\beta)^4] v$, where $v \sim \text{logistic}(0, 1)$

Dist-L: $\varepsilon \square \text{logistic}(0,1)$.