cemmap

centre for microdata methods and practice

Testing for threshold effects in regression models

Sokbae Lee
Myung Hwan Seo
Youngki Shin

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP36/10

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

TESTING FOR THRESHOLD EFFECTS IN REGRESSION MODELS

SOKBAE LEE, MYUNG HWAN SEO, AND YOUNGKI SHIN

ABSTRACT. In this article, we develop a general method for testing threshold effects in regression models, using sup-likelihood-ratio (LR)-type statistics. Although the sup-LR-type test statistic has been considered in the literature, our method for establishing the asymptotic null distribution is new and nonstandard. The standard approach in the literature for obtaining the asymptotic null distribution requires that there exist a certain quadratic approximation to the objective function. The article provides an alternative, novel method that can be used to establish the asymptotic null distribution, even when the usual quadratic approximation is intractable. We illustrate the usefulness of our approach in the examples of the maximum score estimation, maximum likelihood estimation, quantile regression, and maximum rank correlation estimation. We establish consistency and local power properties of the test. We provide some simulation results and also an empirical application to tipping in racial segregation. This article has supplementary materials online.

KEY WORDS. Davies problem, empirical process, maximum score estimation, maximum rank correlation estimation, U-process, threshold model.

AMS SUBJECT CLASSIFICATION. 62F03, 62F05.

Date: 17 November 2010.
We would like to thank Jesse Rothstein for providing the data used in the article and also thank two referees and an associate editor for their helpful comments. Lee thanks the Economic and Social Research Council for the ESRC Centre for Microdata Methods and Practice (RES-589-28-0001) and the European Research Council for the research grant (ERC-2009-StG-240910-ROMETA). Shin thanks the Social Sciences and Humanities Research Council of Canada (410-2010-0345) for the research grant.

1. Introduction

This article develops general tests for threshold effects in a variety of regression models, including mean, median and quantile regression, binary response, censored or truncated regression, and proportional hazards models as special cases. To illustrate our testing problem, consider a binary regression model as an example. In this model, an observed binary outcome Y is modeled typically as $Y = 1(Y^* \geq 0)$, where $1(A)$ denotes the indicator function, i.e., $1(A) = 1$ if A is true and zero otherwise, and Y^* is a latent continuous variable that determines the binary outcome Y (see e.g. Manski, 1988). Suppose that Y^* has the following form:

$$Y^* = g(\mathbf{W}, \theta_0, \gamma_0) + U, \tag{1.1}$$

$$g(\mathbf{w}, \theta, \gamma) = \mathbf{x}'\beta + \mathbf{z}'\alpha 1\{t > \gamma\}, \tag{1.2}$$

where $\mathbf{W}$ is a vector of regressors that consist of distinct elements of $(\mathbf{X}, \mathbf{Z}, T)$, U is an unobserved random variable, and $\theta_0 := (\beta_0', \alpha_0')'$ and γ_0 are unknown true parameter values and belong to $\Theta := \mathcal{B} \times \mathcal{A}$ and Γ, respectively, which are subsets of finite-dimensional Euclidean spaces. Without loss of generality, assume that the vector $\mathbf{Z}$ is a subset of $\mathbf{X}$ such that $\mathbf{Z} = R'\mathbf{X}$ for some known matrix R and that T might be an element of $\mathbf{X}$. The random variable T is the threshold variable and γ_0 is the unknown threshold parameter. Note that we specify the threshold effect as a change-point due to an unknown threshold in a particular covariate.

Threshold models have a large number of applications in empirical research. In economics and sociology, racial segregation can be modeled as a threshold effect. For example, Card et al. (2008) recently investigated the existence of race-based tipping in neighborhoods using U.S. Census data. In their setup, the hypothesis of interest is whether there exist discontinuities in the dynamics of neighborhood racial composition: once the minority share in a neighborhood exceeds a threshold level ("tipping point"), most of the whites would leave the neighborhood. In a simple model developed by Card et al. (2008), whites' willingness to pay for homes depends on the neighborhood minority share. In their model, the location of the tipping point can vary depending on whites' preferences, thereby implying that the location of the tipping point is unknown. In Section 4, we illustrate our methodology by applying it to the data used by Card et al. (2008).

There are more examples of threshold models. In economics, Durlauf and Johnson (1995) argue that cross-country growth models with multiple equilibria can exhibit threshold effects. In addition, Khan and Senhadji (2001) examine the existence of threshold effects in the relationship between inflation and growth. In empirical finance, Pesaran and Pick (2007) argue that the effect of financial contagion (see, e.g. Forbes and Rigobon, 2002) can be described as a discontinuous threshold effect, hence testing for threshold effects implies testing for the presence of financial contagion. In biostatistics, dose-response models are typically specified with some unknown threshold parameters (see, e.g. Cox, 1987; Schwartz et al., 1995). In epidemiology, logistic regressions with unknown change-points are used to model the relationship between the continuous exposure variable and disease risk (see Pastor and Guallar, 1998; Pastor-Barriuso et al., 2003).

We consider a test of no threshold effect against the presence of threshold effects. That is, the null and alternative hypotheses are that

$$\mathcal{H}_0 : \alpha_0 = 0 \ \text{ for any } \gamma_0 \in \Gamma \ vs. \ \mathcal{H}_1 : \alpha_0 \neq 0 \ \text{ for some } \gamma_0 \in \Gamma.$$

In general, unknown parameters in (1.2) are identifiable under the alternative hypothesis; however, the threshold parameter γ_0 is not identified under the null hypothesis. This feature that the threshold parameter is not identified under the null hypothesis is an example of the so-called "Davies problem" (see Davies, 1977, 1987).

As common in the literature (see, e.g., Andrews and Ploberger, 1995; Hansen, 1996; Andrews, 2001), we develop our tests following Roy's union-intersection principle (Roy, 1953) to deal with the Davies problem. Specifically, in our setup, we suppose that there exist an objective function and a corresponding extreme estimator for the null hypothesis of no threshold model and those for the alternative hypothesis of a threshold model. Then our test statistic is based on the difference between the maximum values of the objective functions under the null and alternative hypotheses. This test statistic can be viewed as a sup-likelihood-ratio (LR)-type statistics.

The main objective of this article is to provide a general testing framework in regression models using the sup-LR-type statistic under weak conditions. Most of the prior literature has focused mainly on applications in time series analysis (see, e.g., Tong, 1990; Chan, 1993; Andrews and Ploberger, 1994; Hansen, 1996; Cho and White, 2007). More recently, threshold models have been considered for nonparametric models (e.g. Delgado and Hidalgo, 2000), for panel data models (e.g. Hansen, 1999), for transformation models (e.g. Pons, 2003; Kosorok and Song, 2007), and for binary response models (e.g. Lee and Seo, 2008), among others.

In this article, we focus on cross-sectional applications and aim to provide a unifying testing framework that includes objective functions that are sufficiently different from standard log-likelihood functions. For example, we consider an objective function for the maximum score estimator (Manski, 1975, 1985), and also consider an objective function based on U-processes such as the maximum rank correlation estimator (Han, 1987). To our best knowledge, we are the first to propose tests for threshold effects that can include maximum score and maximum rank correlation estimators as special cases.

1.1. **Related Literature.**

Although the sup-LR-type test statistic is well known in the literature, our method for establishing the asymptotic null distribution is new and nonstandard. The standard approach in the literature for obtaining the asymptotic null distribution requires that there exist a certain quadratic approximation to the objective function (see, e.g., Andrews, 2001; Liu and Shao, 2003; Zhu and Zhang, 2006; Song et al., 2009). For example, Andrews (2001) assumes that the objective function has a quadratic expansion in identifiable parameters for each value of a nuisance parameter that is unidentified under the null hypothesis. In this article, we provide an alternative, novel method that can be used to establish the asymptotic null distribution, even when the usual quadratic approximation such as one in Andrews (2001) is intractable (see Section 3 for details). For example, no existing method can be applied to the objective function for the maximum score estimator. However, it is worth noting that when quadratic approximations are available, Andrews (2001) covers a more general case such as one where parameter vectors may lie on the boundary of the parameter space under the null hypothesis.

In the literature, there exist articles that establish asymptotic distributions for likelihood ratio types of statistics, without requiring usual asymptotic quadratic approximations. For example, Fan et al. (2000) establish that Wilks results hold as long as likelihood contour sets are fan-shaped. As a result, they show that the likelihood ratio statistics can still be asymptotically chi-squared, even if the maximum likelihood estimator is not asymptotically normal. In addition, Zhang and Li (1993) develop an empirical process approach to deriving the asymptotic null distribution of the sup-LR-type statistics without requiring the usual quadratic approximation to the likelihood function, using the general result in Zhang and Cheng (1989). Their approach is closely related to ours in that it employs the empirical process method; however, their method is different from ours in the sense that they focus on the case when the objective function is a likelihood function and when the class of likelihood functions is assumed to be Hölder continuous in parameters. Again, neither Fan et al. (2000) nor Zhang and Li (1993) can include our maximum score estimator example as a special case.

1.2. **Structure of the Paper.** The remainder of the article is as follows. In Section 2, we provide an informal description of our test statistic and a couple of examples. Section 3 provides an informal overview of our method for obtaining the asymptotic null distribution. Section 4 illustrates the usefulness of our test by applying it to real data used by Card et al. (2008). Formal results are given in Section 5 and they are illustrated in Section 6. A summary of Monte Carlo simulation results are provided briefly in Section D. In Section 8, we provide some concluding remarks. All the proofs, some additional theoretical results, and details of Monte Carlo experiments are contained in the online supplementary materials.

2. The Test Statistic

This section describes our test statistic. To develop a general testing framework without being tied down to a particular statistical model, we suppose that under the null hypothesis, the remaining unknown parameters in (1.2) can be estimated by optimizing a particular objective function and also that under the alternative hypothesis, all unknown parameters including α_0 can be estimated by optimizing a suitable objective function. In other words, we develop our test statistic based on the distance between optimized restricted and unrestricted objective function values.

To be more specific, let $Q_n : \Theta \otimes \Gamma \mapsto \mathbb{R}$ denote an objective function of interest based on a random sample $\{(Y_i, \mathbf{W}_i) : i = 1, \ldots, n\}$. For a given $\gamma \in \Gamma$, let $\hat{\theta}(\gamma)$ denote an estimator of θ_0 that maximizes the objective function $Q_n(\theta, \gamma)$. Define $Q_n(\gamma) := Q_n\left(\hat{\theta}(\gamma), \gamma\right)$ to be a profiled objective function and let

$$\hat{\gamma} = \underset{\gamma}{\operatorname{argmax}}\, Q_n(\gamma), \ \hat{\theta} = \hat{\theta}(\hat{\gamma}), \text{ and } \hat{Q}_n = Q_n(\hat{\gamma}).$$

In addition, let

$$\tilde{\beta} = \underset{\beta:\alpha=0}{\operatorname{argmax}}\, Q_n(\theta, \gamma) \text{ and } \tilde{Q}_n = \max_{\beta:\alpha=0} Q_n(\theta, \gamma).$$

Recall that Q_n does not depend on γ when $\alpha = 0$. That is, $\tilde{Q}_n$ is the maximum value of the objective function under the null hypothesis and $\hat{Q}_n$ is the maximum value without imposing the null hypothesis.

Our test statistic is based on the difference between $\hat{Q}_n$ and $\tilde{Q}_n$, analogous to the likelihood ratio (LR) statistic. Define the quasi-LR (QLR) statistic by

$$QLR_n = r_n^2 \left(\hat{Q}_n - \tilde{Q}_n\right),$$

where r_n is a rate of convergence in probability of $\hat{\theta}(\gamma)$ for a given γ. Let

$$QLR_n(\gamma) = r_n^2 \left[Q_n(\gamma) - \tilde{Q}_n \right]$$

for each γ, and note that

$$QLR_n = \sup_{\gamma \in \Gamma} QLR_n(\gamma).$$

Thus, the statistic QLR_n can be viewed as a sup LR-type statistic. This statistic is relatively easier to implement and analyze than some alternative statistics, e.g. a sup Wald test statistic because it would not be straightforward to studentize the latter and to show the uniform tightness of $\hat{\alpha}(\gamma)$ in some cases, e.g. in the maximum score estimation for binary response models. Also, we expect that the objective-function-based statistic would have better finite sample performance as it is more immune to local maxima problems.

We consider two types of $Q_n(\theta, \gamma)$: the first type is a sample mean statistic and the second type is a U-statistic. For the first case, the objective function has the form

$$Q_n(\theta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} q(Y_i, \mathbf{W}_i; \theta, \gamma), \tag{2.1}$$

where q is a known function up to parameters θ and γ. For example, the maximum score estimator maximizes $Q_n(\theta, \gamma)$ with

$$q(y, \mathbf{w}; \theta, \gamma) = (2y - 1) 1\{g(\mathbf{w}, \theta, \gamma) \geq 0\}.$$

In this example, the rate of convergence is $r_n = n^{1/3}$. For the second case, the objective function has the form

$$Q_n(\theta, \gamma) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \chi(Y_i, \mathbf{W}_i, Y_j, \mathbf{W}_j; \theta, \gamma), \tag{2.2}$$

where χ is again a known function up to parameters, and is symmetric in the sense that $\chi(y_i, \mathbf{w}_i, y_j, \mathbf{w}_j; \theta, \gamma) = \chi(y_j, \mathbf{w}_j, y_i, \mathbf{w}_i; \theta, \gamma)$. For example, the maximum rank correlation estimator maximizes $Q_n(\theta, \gamma)$ with

$$\chi(y_1, \mathbf{w}_1, y_2, \mathbf{w}_2; \theta, \gamma) = 1\{y_1 > y_2\} 1\{g(\mathbf{w}_1, \theta, \gamma) > g(\mathbf{w}_2, \theta, \gamma)\} + 1\{y_1 < y_2\} 1\{g(\mathbf{w}_1, \theta, \gamma) < g(\mathbf{w}_2, \theta, \gamma)\}.$$

In this example, $r_n = n^{1/2}$. In both cases, we assume that q or χ depends on (θ, γ) only through the regression function g.

Additional examples include the maximum likelihood estimator of the probit (or logit) model, the quantile regression estimator (see Koenker, 2005, for the comprehensive treatment of the methodology), and the partial maximum likelihood estimator of the proportional hazard model (see Cox, 1972, 1975) in the first class, and various rank correlation based estimators such as the monotone rank estimator (Cavanagh and Sherman, 1998) and the pairwise rank estimator (Abrevaya, 1999) in the second class.

3. Informal Overview of the Results

This section provides an informal overview of our method for obtaining the asymptotic null distribution. Formal results are given in Section 5. The main idea is to represent our test as a continuous functional of an empirical process for a certain transformation of objective functions of interest without referring to the estimators under the null and alternative hypotheses. Therefore, our method for obtaining the asymptotic null distribution does not require an expansion of the objective functions, and can be used even in cases when the usual quadratic approximation is unavailable or difficult to obtain. In general, the asymptotic null distribution is not pivotal; however, a method for computing asymptotic p-values is illustrated with a couple of examples (subsampling is another option).

In what follows, we use the conventional notation in empirical process theory. Denote by $\mathbf{P}$ the common probability measure, by $\mathbb{P}_n$ the empirical measure of the random sample of size n from $\mathbf{P}$, and by $\mathbb{G}_n$ the empirical process indexed by a class $\mathcal{F}$ of functions q such that $\mathbb{G}_n q = \sqrt{n}(\mathbb{P}_n - \mathbf{P}) q$.

To provide the main idea behind our method, we focus on M-estimation, that is the objective function has the form (2.1). Define

$$m_{\xi,\gamma} = q_{\theta,\gamma} - \tilde{q}_b,$$

where $\xi = (\theta', b')'$, $q_{\theta,\gamma} = q(y, \mathbf{w}; \theta, \gamma)$, and $\tilde{q}_b = q_{(b',0)',\gamma}$. Note that $\tilde{q}_b$ is the same for any γ and thus it is a function of b only. We have introduced the index b to denote arguments for β_0 in the objective function with the restriction $\alpha = 0$ to distinguish this from the index β that denotes arguments for β_0 in the unrestricted objective function.

Also, note that $q_{\theta,\gamma}$ is the same for all γ if $\theta = \theta_0$, using the fact that $\alpha_0 = 0$ under $\mathcal{H}_0$. Thus, under $\mathcal{H}_0$, $q_{\theta_0,\gamma} = \tilde{q}_{\beta_0}$, and when b is restricted to β_0,

$$m_{\xi,\gamma} = q_{\theta,\gamma} - q_{\theta_0,\gamma}.$$

Similarly, when θ is fixed at θ_0,

$$m_{\xi,\gamma} = q_{\theta_0,\gamma} - \tilde{q}_b.$$

It now follows that

$$\begin{aligned} QLR_n &= r_n^2 \left[\sup_{\theta,\gamma} \mathbb{P}_n q_{\theta,\gamma} - \sup_b \mathbb{P}_n \tilde{q}_b \right] \\ &= r_n^2 \left[\sup_{\xi,\gamma: b=\beta_0} \mathbb{P}_n m_{\xi,\gamma} - \sup_{\xi,\gamma:\theta=\theta_0} (-\mathbb{P}_n m_{\xi,\gamma}) \right], \end{aligned} \tag{3.1}$$

which is a continuous transformation of $r_n^2 \mathbb{P}_n m_{\xi,\gamma}$.

Note that since $\xi := (\theta', b')$, b is still a free parameter after fixing θ at θ_0 and also θ is still a free parameter after fixing $b = \beta_0$. That is, we treat θ and b separate parameters. The reason why $m_{\xi,\gamma}$ is defined in this way is to write the QLR test as a continuous transformation of an empirical process for $m_{\xi,\gamma}$. Note also that $m_{\xi_0,\gamma} = 0$ for any γ, where $\xi_0 = (\theta_0', \beta_0')'$. Then the convergence of $r_n^2 \mathbb{P}_n m_{\xi,\gamma}$ can be derived using the empirical process theory through the decomposition

$$r_n^2 \mathbb{P}_n m_{\xi,\gamma} = \frac{r_n^2}{\sqrt{n}} \mathbb{G}_n m_{\xi,\gamma} + r_n^2 \mathbf{P} m_{\xi,\gamma}. \tag{3.2}$$

Since the supremum is obtained at $\theta = \hat{\theta}(\gamma)$ for each γ and at $b = \tilde{\beta}$, respectively, with the convergence rate r_n, we examine a rescaled version of the process in (3.2) to obtain the asymptotic null distribution.

3.1. **Example 1: Probit.** We use the probit model as our first illustrative example. Define $\mathbf{W}_\gamma := (\mathbf{X}', \mathbf{Z}' 1\{T > \gamma\})'$. The function $q(y, \mathbf{w}; \theta, \gamma)$ for the probit model has the form

$$q(y, \mathbf{w}; \theta, \gamma) = y \log \Phi(g(\mathbf{w}, \theta, \gamma)) + (1-y) \log \Phi(-g(\mathbf{w}, \theta, \gamma)), \tag{3.3}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution and $g(\mathbf{W}, \theta, \gamma) = \mathbf{W}_\gamma' \theta$. It will be shown formally in Section 6 that the limiting distribution of the test statistic is the supremum of a chi-square process indexed by γ as in (5.10). Let $\phi(\cdot)$ denote the probability density function of the standard normal distribution. Let $e = (2Y-1)\phi(\mathbf{X}'\beta_0)/\Phi((2Y-1)\mathbf{X}'\beta_0)$ and $V(\gamma) = E\left[-e^2 \mathbf{W}_\gamma \mathbf{W}_\gamma'\right]$. Also, let $\mathcal{G}$ denote a Gaussian process with the covariance kernel

$$K(\gamma_1, \gamma_2) = E\left[e^2 \mathbf{W}_{\gamma_1} \mathbf{W}_{\gamma_2}'\right].$$

Then, the asymptotic distribution of the QLR_n test becomes

$$\frac{1}{2}\left[\sup_\gamma \mathcal{G}(\gamma)' V(\gamma)^{-1} \mathcal{G}(\gamma) - \mathcal{G}_1' V_\beta^{-1} \mathcal{G}_1\right], \tag{3.4}$$

where $\mathcal{G}_1$ and V_β denote the first k_β elements of $\mathcal{G}$ and the first $k_\beta \times k_\beta$ block of $V(\gamma)$, respectively. Here, k_β denotes the dimension of β.

Note that we cannot tabulate the critical values due to the nonstandard asymptotic distribution and need a simulation method to conduct the testing procedure. For example, we can adopt the p-value transformation method as in Hansen (1996). The basic idea is to approximate the asymptotic distribution by simulating the Gaussian process, which is the empirical process of the score function in our case. For each $i = 1, \ldots, n$, let

$$\nabla_\theta \hat{q}_i = \mathbf{W}_{\gamma,i}(2Y_i - 1)\frac{\phi(\mathbf{W}'_{\gamma,i}\hat{\theta}(\gamma))}{\Phi[(2Y_i - 1)\mathbf{W}'_{\gamma,i}\hat{\theta}(\gamma)]},$$

$$\nabla_b \tilde{q}_i = \mathbf{X}_i(2Y_i - 1)\frac{\phi(\mathbf{X}'_i\tilde{\beta})}{\Phi[(2Y_i - 1)\mathbf{X}'_i\tilde{\beta}]},$$

where $\mathbf{W}_{\gamma,i} := (\mathbf{X}'_i, \mathbf{Z}'_i 1\{T_i > \gamma\})'$. We now carry out the following steps to simulate the p-value.

(1) to generate $i.i.d.$ $N(0,1)$ random variables $\{v_{ij}\}_{i=1}^n$ for $j = 1, ..., J$ for a sufficiently large J;

(2) to simulate unrestricted and restricted score functions, respectively:

$$G^j_{n,\theta}(\gamma) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \nabla_\theta \widehat{q}_i(\gamma) v_{ij} \quad \text{and} \quad G^j_{n,b} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \nabla_b \widetilde{q}_i v_{ij};$$

(3) to simulate test statistics $\{D^j_n\}_{j=1}^J$ using the simulated score functions above and the sample analogue of the asymptotic distribution in (3.4):

$$D^j_n = \sup_\gamma \frac{1}{2}\left[G^j_{n,\theta}(\gamma)' \widehat{\mathcal{I}}^{-1}_{\theta,\gamma} G^j_{n,\theta}(\gamma) - G^{j\prime}_{n,b}\widehat{\mathcal{I}}^{-1}_b G^j_{n,b}\right]$$

where $\widehat{\mathcal{I}}^{-1}_{\theta,\gamma} = (1/n)\sum_{i=1}^n \nabla_\theta \widehat{q}_i(\gamma)\nabla_\theta \widehat{q}_i(\gamma)'$ and $\widehat{\mathcal{I}}^{-1}_b = (1/n)\sum_{i=1}^n \nabla_b \widetilde{q}_i \nabla_b \widetilde{q}'_i$, respectively;

(4) to set $\widehat{p}^J_n = (1/J)\sum_{j=1}^J 1\left\{D^j_n > \widehat{D}_n\right\}$.

3.2. **Example 2: Quantile Regression.** We now consider the quantile regression model. The function $q(y, \mathbf{w}; \theta, \gamma)$ for the quantile regression model has the form

$$q(y, \mathbf{w}; \theta, \gamma) = -\rho_\tau\left[y - g(\mathbf{w}, \theta, \gamma)\right], \tag{3.5}$$

where $\rho_\tau(u) := u(\tau - 1(u < 0))$ is the 'check' function and $g(\mathbf{W}, \theta, \gamma) = \mathbf{W}'_\gamma\theta$. As in the probit model, it will be shown formally in Section 6 that the limiting null distribution of the QLR statistic is characterized by

$$\frac{1}{2}\left[\sup_{\gamma} \mathcal{G}(\gamma)' V(\gamma)^{-1} \mathcal{G}(\gamma) - \mathcal{G}_1' V_1^{-1} \mathcal{G}_1\right], \tag{3.6}$$

where $\mathcal{G}$ is a mean-zero Gaussian process with covariance kernel

$$K(\gamma_1, \gamma_2) = \tau(1-\tau) E W_{\gamma_1} \mathbf{W}_{\gamma_2}', \tag{3.7}$$

$V(\gamma)$ is a matrix such that $V(\gamma) = E\left[\mathbf{W}_\gamma \mathbf{W}_\gamma' f_{Y|\mathbf{W}}(\mathbf{X}'\beta_0|\mathbf{W})\right]$, and $\mathcal{G}_1$ and V_1 denote the first k_1 elements of $\mathcal{G}$ and the first $k_\beta \times k_\beta$ block of $V(\gamma)$, respectively. As before, k_β denotes the dimension of β. Now the p-values can be simulated in the following way:

(1) to generate $i.i.d.$ Unif $[0,1]$ random variables $\{u_{ij}\}_{i=1}^n$ for $j = 1, ..., J$ for a sufficiently large J;

(2) to simulate the following functions, respectively:

$$G_n^j(\gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{W}_{\gamma,i}[\tau - 1(u_{ij} \le \tau)] \quad \text{and} \quad \tilde{G}_n^j = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{X}_i[\tau - 1(u_{ij} \le \tau)];$$

(3) to simulate test statistics $\{D_n^j\}_{j=1}^J$ by:

$$D_n^j = \sup_{\gamma} \frac{1}{2}\left[G_n^j(\gamma)' \widehat{V}(\gamma)^{-1} G_n^j(\gamma) - \tilde{G}_n^{j\prime} \widetilde{V}^{-1} \tilde{G}_n^j\right]$$

where

$$\begin{aligned} \widehat{V}(\gamma) &= \frac{1}{nh_n} \sum_{i=1}^n \mathbf{W}_{\gamma,i} \mathbf{W}_{\gamma,i}' K\left(\frac{Y_i - \mathbf{X}_i'\tilde{\beta}}{h_n}\right), \text{ and} \\ \widetilde{V} &= \frac{1}{nh_n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' K\left(\frac{Y_i - \mathbf{X}_i'\tilde{\beta}}{h_n}\right); \end{aligned} \tag{3.8}$$

(4) to set $\widehat{p}_n^J = (1/J) \sum_{j=1}^J 1\left\{D_n^j > \widehat{D}_n\right\}$.

In step (3), K is a kernel function and h_n is a bandwidth. Recall that $\tilde{\beta}$ is the estimator of β_0 under the null. When the regression error is independent of the regressors, then we can estimate $\widehat{V}(\gamma)$ and $\widetilde{V}$ by

$$\widehat{V}(\gamma) = \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{W}_{\gamma,i}\mathbf{W}'_{\gamma,i}\right] \times \left[\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{Y_i - \mathbf{X}'_i\tilde{\beta}}{h_n}\right)\right], \text{ and}$$
$$\widetilde{V} = \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}'_i\right] \times \left[\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{Y_i - \mathbf{X}'_i\tilde{\beta}}{h_n}\right)\right]. \tag{3.9}$$

4. Application: Tipping in Segregation

We apply the proposed testing procedure to check whether there exists a tipping point for segregation. Using U.S. Census tract-level data, Card et al. (2008) recently showed that the neighborhood's white population decreases substantially when the minority share in the area exceeds a tipping point (or threshold point).

In this application, we use a subsample of the dataset originally used by Card et al. (2008). Among three different base years, we choose a sample of which base year is 1980. Next we pick four major cities and tested if there is a tipping point. To illustrate our testing procedure, we first consider the probit model. We suppose that data $\{(Y_i, \mathbf{X}_i, T_i) : i = 1, \ldots, n\}$ are generated from

$$Dw_i = \beta_0 + \alpha_0 1\{T_i > \gamma_0\} + \mathbf{X}'_i\delta_0 + \epsilon_i,$$
$$Y_i = 1\{Dw_i > 0\},$$

where Dw_i is the ten-year change in the neighborhood's white population, T_i is the base-year minority share in the neighborhood, and $\mathbf{X}_i$ is a vector of six tract-level control variables. The $\mathbf{X}$ variables include the unemployment rate, the log of mean family income, the fractions of single-unit, vacant, and renter-occupied housing units, and the fraction of workers who use public transport to travel to work. See Card et al. (2008) for details on the dataset and variables. In the original dataset, Dw_i is observed but for the time being, we treat this as a latent variable to illustrate our testing procedure for the probit model. The error term ϵ_i follows the standard normal distribution. Thus, the null and alternative hypotheses in our setting are

$$H_0 : \alpha_0 = 0 \quad \text{and} \quad H_1 : \alpha_0 \neq 0, \tag{4.1}$$

respectively.

The four cities we have chosen are Boston, Chicago, New York, and Philadelphia. The p-values are calculated by the simulation method described in Section 3 with 1,000 simulations. For estimating a tipping point (γ) under the alternative, we use the grid search method. The grid points are constructed from T_i that fell in the interval $[l, 50\%]$, where l is the maximum of 5% and the 5th percentile of $\{T_i\}$.

Table 1. Test for Tipping in Segregation: Probit Model

City	*obs.*	p-value	Tipping Points ($\hat{\gamma}$)	$E_{\mathbf{X}}[\Delta \Pr(y=1\vert\mathbf{X})]$
Boston, MA	700	2.8%	46.80	-0.25
Chicago, IL	1813	0.0%	48.74	-0.34
New York, NY	2430	0.0%	14.01	-0.09
Philadelphia, PA	1300	0.0%	39.64	-0.30

We summarize the result in Table 1. The last column of each table shows the average changes in probability that the white population would increase when the minority share crosses the tipping point. We calculate this average marginal effect as

$$E_{\mathbf{X}}[\Delta \Pr(y=1|\mathbf{X})] \;=\; \frac{1}{n}\sum_i \left\{\Phi(\hat{\beta} + \mathbf{X}_i'\hat{\delta}) - \Phi(\hat{\beta} + \hat{\alpha} + \mathbf{X}_i'\hat{\delta})\right\}$$

where $\Phi(\cdot)$ is the CDF of the normal distribution.

First of all, testing results show that there exist tipping points in all four cities. Second, the tipping points vary from 14.01% in New York to 48.74% in Chicago. This shows that cities are heterogeneous in whites' preferences, among other things, implying that tolerance levels against minority shares are quite different across different cities. Third, the average marginal effects are also different across cities. New York shows that the probability drops less than 10%. Meanwhile, Chicago shows that it drops more than 30%.

We now illustrate our testing procedure for the median regression model using observed Dw_i directly, instead of Y_i. We now suppose that data $\{(Dw_i, \mathbf{X}_i, T_i) : i = 1, \ldots, n\}$ are generated from

$$Dw_i = \beta_0 + \alpha_0 1\{T_i > \gamma_0\} + p(T_i) + \mathbf{X}_i' \delta_0 + \epsilon_i,$$

where $\text{median}(\epsilon_i | \mathbf{X}_i, T_i) = 0$ and $p(T)$ is the 4th-order polynomial of T. Note that Card et al. (2008) considered the mean regression model (that is, $E(\epsilon_i | \mathbf{X}_i, T_i) = 0$) with the 4th-order polynomial in $p(T)$. The null and alternative hypotheses are the same as those in (4.1).

The p-values are calculated by the simulation method described in Section 3 with 2,000 simulations. In the application, we estimated $\widehat{V}(\gamma)$ and $\widetilde{V}$ by (3.8) since we do not know whether regression errors are independent of regressors. For estimating a tipping point (γ) under the alternative, we used the grid search method. The grid points were constructed from T_i that fell in the interval $[l, 60\%]$, where l is the maximum of 5% and the 5th percentile of $\{T_i\}$.

Table 2. Median regression model with the 4th-order polynomial

City	*obs.*	p-value	Tipping Point ($\hat{\gamma}$)	Size of the Jump ($\hat{\alpha}$)
Boston, MA	700	4.7%	51.75	-17.640
Chicago, IL	1813	2.9%	48.45	-13.929
New York, NY	2430	1.5%	23.70	-7.309
Philadelphia, PA	1300	1.2%	39.65	-11.599

We summarize the result in Table 2. Testing results show that there exist tipping points in all four cites at the 5% level. The tipping points vary across these cities and are not much different from those from the probit model, especially in Chicago and Philadelphia.

5. The Asymptotic Null Distribution

This section provides asymptotic theory for obtaining the asymptotic null distribution. Our assumptions are quite general and allow for a nonsmooth objective function Q_n, which may not permit usual quadratic approximations. As in Section 3, we focus on the M-estimation in this section. In the online supplements, we provide asymptotic theory for the case when objective functions are based on U-processes and verify regularity conditions for the maximum rank correlation (MRC) estimator. The consistency and local power of the test are included in the online supplements as well.

5.1. **M-estimation.** This section considers the first case when the objective function has the form in (2.1). Our estimators need not be exact maximizers, which might have measurability issues. Thus, we consider an estimator $\hat{\theta}_\gamma$ for a given $\gamma \in \Gamma$ such that

$$Q_n\left(\hat{\theta}_\gamma, \gamma\right) = \sup_{\theta \in \Theta} Q_n(\theta, \gamma) + o_{p\gamma}\left(r_n^{-2}\right),$$

where $o_{p\gamma}(1)$ indicates the sequence under consideration is $o_p(1)$ uniformly over $\gamma \in \Gamma$. We define $o_\gamma(1)$ and $O_{p\gamma}(1)$ similarly. Also, let $\tilde{\beta}$ satisfy

$$\bar{Q}_n\left(\tilde{\beta}\right) = \sup_{\beta \in \mathcal{B}} \bar{Q}_n(\beta) + o_p\left(r_n^{-2}\right),$$

where $\bar{Q}_n$ denotes the restrictive objective function with $\alpha = 0$.

To derive the asymptotic distribution of the statistic QLR_n, we impose some high-level assumptions, which will be verified later for each example. We first introduce some notation. Let

$$\mathcal{F}_\delta = \left\{q_{\theta,\gamma} - q_{\theta_0,\gamma} : |\theta - \theta_0| < \delta, \gamma \in \Gamma\right\}, \tag{5.1}$$

where $|\cdot|$ is the Euclidean norm for a vector (we use the notation $\|\cdot\|$ to indicate a generic norm for a function space). An envelope function of a class $\mathcal{F}$ is a function F such that $\mathbf{P}F^2 < \infty$, $|f(\mathbf{x})| \leq F(\mathbf{x})$ for any $\mathbf{x}$ and $f \in \mathcal{F}$. An envelope function for $\mathcal{F}_\delta$ is denoted by F_δ.

Weak convergence of the statistic QLR_n draws on the size of the class $\mathcal{F}_\delta$ measured by *entropy with or without bracketing*. Let $N(\varepsilon, \mathcal{F}, \|\cdot\|)$ and $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ denote covering and bracketing numbers, respectively. The logarithm of the covering number is called entropy (without bracketing) and that of the bracketing number is called entropy with bracketing. We mostly use the $L_r(Q)$-norm, $\|f\|_{Q,r} = \left(\int |f|^r \, dQ\right)^{1/r}$, where Q is a probability measure. When the entropy without bracketing is concerned, it is common that the required condition is in terms of *uniform entropy,* $\sup_Q \log N(\varepsilon, \mathcal{F}, L_r(Q))$, where the supremum is taken over all the possible probability measures on the sample space, with $0 < QF^r < \infty$. While the measurability is an issue in the formal discussion of uniform entropy conditions, it hardly matters in applications. We assume measurability throughout the article. See e.g. van der Vaart and Wellner (1996) for more general discussions on the empirical process method.

We now present a set of assumptions, whose details will be discussed later on.

Assumption 5.1 (Uniform Consistency). $\hat{\theta}(\gamma) = \theta_0 + o_{p\gamma}(1)$ *and* $\tilde{\beta} = \beta_0 + o_p(1)$.

A set of sufficient conditions for the uniform consistency in Assumption 5.1 is that (i) uniform convergence of the objective function Q_n; (ii) separability of the true value. Formally, we present it as Lemma 5.2 in Section 5.2.

Assumption 5.2 (Uniform Rates of Convergence in Probability). $r_n\left(\tilde{\beta} - \beta_0\right) = O_p(1)$ *and* $r_n\left(\hat{\theta}(\gamma) - \theta_0\right) = O_{p\gamma}(1)$.

Most often, the rate r_n in Assumption 5.2 is already known for linear models and r_n must be the same for $\hat{\theta}(\gamma)$ for each γ since $g(\mathbf{w}, \theta, \gamma)$ is a linear function in θ. Thus, Assumption 5.2 has mainly to do with verifying the uniformity. However, the entropy conditions below in Assumption 5.4 are almost sufficient to ensure it, as will be shown in Lemma 5.3 in Section 5.2.

In what follows, fix $0 < K < \infty$ and assume the following.

Assumption 5.3 (Lindeberg Condition and L_2-Continuity). *For any* $\eta > 0$,

$$\frac{r_n^4}{n}\mathbf{P}F_{K/r_n}^2 = O(1),$$

$$\frac{r_n^4}{n}\mathbf{P}F_{K/r_n}^2 1\left\{\frac{r_n^2}{\sqrt{n}}F_{K/r_n} > \eta\sqrt{n}\right\} = o(1).$$

In addition, for any decreasing sequence $\eta_n \to 0$,

$$\sup_{\substack{|h_1-h_2|<\eta_n \\ |\gamma_1-\gamma_2|<\eta_n}} \frac{r_n^4}{n}\mathbf{P}\left(q_{\theta_0+h_1/r_n,\gamma_1} - q_{\theta_0+h_2/r_n,\gamma_2}\right)^2 = o\left(1\right). \tag{5.2}$$

Assumption 5.3 is a minimal set of conditions on the moments of the envelope function F and on the smoothness of the limit objective function. These are straightforward to verify.

Assumption 5.4 (Entropy Conditions)**.** *For some $\delta_0 > 0$, assume that*

$$\int_0^1 \sup_{\delta<\delta_0} \sup_Q \sqrt{\log N\left(\varepsilon \left\|F_\delta\right\|_{Q,2}, \mathcal{F}_\delta, L_2\left(Q\right)\right)}d\varepsilon < \infty \tag{5.3}$$

or

$$\int_0^1 \sup_{\delta<\delta_0} \sqrt{\log N_{[]}\left(\varepsilon \left\|F_\delta\right\|_{P,2}, \mathcal{F}_\delta, L_2\left(P\right)\right)}d\varepsilon < \infty. \tag{5.4}$$

It is not always trivial to verify these entropy conditions. However, there are well-known classes of functions that satisfy either of the conditions. For example, Vapnik-Červonenkis (VC) classes of functions have the covering numbers that are bounded by a polynomial in ε^{-1}, thus satisfying (5.3) as long as the VC indexes are bounded in n. The bracketing numbers for classes of smooth functions, monotone functions, convex functions, or Lipschitz functions are known, see e.g. Section 2.7 of van der Vaart and Wellner (1996). In particular, the bracketing number of the collection of Lipschitz functions are bounded by the covering number of the index set, thus, being at most the polynomial in $(1/\varepsilon)^p$, where p is the dimension of the parameter space.

Partition h into $(h_\theta', h_b')'$ according to the dimensions of θ and b, respectively.

Assumption 5.5 (Finite-Dimensional Weak Convergence)**.** *Let $h_{1n} = \xi_0 + h_1 r_n^{-1}$ and $h_{2n} = \xi_0 + h_2 r_n^{-1}$. Then, for any $K > 0$, any $\gamma_1, \gamma_2 \in \Gamma$, and any h_1 and h_2 whose Euclidean norms are less than K,*

$$\frac{r_n^4}{n}\mathbf{P}\left(m_{h_{1n},\gamma_1} - m_{h_{2n},\gamma_2}\right)^2 \to E\left(G_1\left(h_1,\gamma_1\right) - G_1\left(h_2,\gamma_2\right)\right)^2,$$

where G_1 is a zero-mean Gaussian process. Furthermore, let $h_n = \xi_0 + hr_n^{-1}$. Then

$$r_n^2 \mathbf{P} m_{h_n,\gamma} \longrightarrow G_2(h,\gamma)$$

uniformly in h and γ over any compact set, for some non-stochastic G_2. Finally, G_1 and G_2 satisfy that

$$\frac{EG_1(h,\gamma)^2}{|h_\theta|^r} \to 0 \quad \text{and} \quad \frac{G_2(h,\gamma)}{|h_\theta|^{r+1/2}} \to -\infty \tag{5.5}$$

as $|h_\theta| \to \infty$, for any γ, h_b, and some $r > 0$.

The limit process over which the supremum will be taken is characterized by the terms given in Assumption 5.5. Considering the definition of $m_{\xi,\gamma}$, the Gaussian process $G_1(h,\gamma)$ is likely to be degenerate in h as shown in later examples. Condition (5.5) in Assumption 5.5 guarantees that the restricted suprema (as in the definition of QLR_n in (3.1)) of $G_1 + G_2$ are $O_p(1)$. When $G_2(h,\gamma)$ is quadratic in h and $G_1(h,\gamma)$ is linear in h for a given γ, then one can choose $r = 1$ in (5.5).

We now present our main theorem.

Theorem 5.1. *Under Assumptions 5.1-5.5,*

$$QLR_n \Rightarrow \sup_\gamma \left[\sup_{h:h_b=0} G(h,\gamma) - \sup_{h:h_\theta=0} \left(-G(h,\gamma)\right) \right], \tag{5.6}$$

where $G = G_1 + G_2$.

While the asymptotic null distribution of QLR_n is well-defined under the restriction in Assumption 5.5, the asymptotic critical values cannot be tabulated due to the unknown covariance kernel of G_1. Therefore, we need to simulate critical values or asymptotic p-values. Alternatively, we need to use resampling methods such as the bootstrap or subsampling. Subsampling works more generally including all the examples we examined in this article. When we can solve out the maximizers explicitly for the expression inside the bracket in

(5.6), simulating the critical values in the spirit of Hansen (1996) can also be applied. Two examples in Section 3 belong to this case.

5.2. **Low-Level Sufficient Conditions for Assumptions.** This section provides low-level sufficient conditions for Assumptions 5.1-5.5. First, we present the following lemma that can be used to verify Assumption 5.1.

Lemma 5.2. *Let $\mathcal{F}$ be a class of functions $\{q_{\theta,\gamma} : (\theta, \gamma) \in \Theta \times \Gamma\}$ with envelope F such that $\mathbf{P}F < \infty$. Suppose either of the following two conditions is satisfied: (i) $N_{[]}(\varepsilon, \mathcal{F}, L_1(\mathbf{P})) < \infty$ for every $\varepsilon > 0$; (ii) For $\mathcal{F}_M$ defined as the class of functions $f1\{F \leq M\}$ for $f \in \mathcal{F}$, $\log N(\varepsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) = o_p(n)$ for every ε and $M > 0$. Then,*

$$\sup_{\theta,\gamma} |Q_n(\theta, \gamma) - Q(\theta, \gamma)| \xrightarrow{p} 0,$$

where $Q(\theta, \gamma) = \mathbf{P}q_{\theta,\gamma}$. Furthermore, assume that

$$\sup_{\gamma \in \Gamma} \left[\sup_{\theta \notin \Theta_0} \{Q(\theta_0, \gamma) - Q(\theta, \gamma)\} \right] > 0 \tag{5.7}$$

for every open set Θ_0 that contains θ_0. Then, $\hat{\theta}(\gamma) - \theta_0 = o_{p\gamma}(1)$.

While there are different ways to present sufficient conditions for Assumption 5.1, we choose this way as the subsequent discussion also draws on the entropy conditions. The entropy conditions in Lemma 5.2 are automatically satisfied if Assumption 5.4 holds. Thus, separability is the one we need to check. Recall that $Q(\theta_0, \gamma)$ is the same for all γ since γ is not identified under the null. However, once we establish the consistency for a given γ and that $Q(\theta_0, \gamma) > \sup_{\theta \notin \Theta} Q(\theta, \gamma)$, the verification of the separability is not very difficult since γ appears only through an indicator function.

We now consider sufficient conditions for Assumption 5.2. The following lemma generalizes a standard method in van der Vaart and Wellner (1996) for obtaining the convergence rate to the case where a uniform rate is needed due to the presence of a nuisance parameter. See Andrews (2001) for a different approach when the quadratic approximation is plausible.

Lemma 5.3. *Assume that for every θ in a neighborhood of θ_0,*

$$\sup_{\gamma} \mathbf{P}\left(q_{\theta,\gamma} - q_{\theta_0,\gamma}\right) \leq -C\left|\theta - \theta_0\right|^2, \tag{5.8}$$

for some finite constant $C > 0$ and that for every n and sufficiently small δ,

$$E \sup_{\gamma} \sup_{|\theta-\theta_0|<\delta} \left|\mathbb{G}_n\left(q_{\theta,\gamma} - q_{\theta_0,\gamma}\right)\right| = O\left(\phi\left(\delta\right)\right), \tag{5.9}$$

for a function ϕ such that $\phi\left(\delta\right)/\delta^r$ is decreasing for some $r < 2$. If Assumption 5.1 holds, then

$$r_n\left(\hat{\theta}\left(\gamma\right) - \theta_0\right) = O_{p\gamma}\left(1\right),$$

for every r_n such that $r_n^2\phi\left(1/r_n\right) \leq \sqrt{n}$ for every n. If the rate r_n is known, then (5.9) can be stated for $\delta = K/r_n$ and $\phi\left(\delta\right) = \sqrt{n}/r_n^2$.

The first condition (5.8) is not difficult to verify. Often, $\mathbf{P}q_{\theta,\gamma}$ is twice continuously differentiable at θ_0 for all γ. In this case, a sufficient condition is the existence of nonsingular second derivative matrices at $\theta = \theta_0$ whose maximum eigenvalues are uniformly bounded away from zero.

The second condition (5.9) is implied by Assumptions 5.3 and 5.4. It is known that the left-hand side term in the equation (5.9) is bounded by the product of the L_2 norm of the envelope function, $\mathbf{P}^{1/2}\left(F_\delta^2\right)$, and the uniform entropy integral or the bracketing integral, which is defined respectively by

$$\sup_{Q} \int_0^1 \sqrt{1 + \log N\left(\varepsilon\left\|F_\delta\right\|_{Q,2}, \mathcal{F}_\delta, L_2\left(Q\right)\right)}d\varepsilon$$

or

$$\int_0^1 \sqrt{1 + \log N_{[]}\left(\varepsilon\left\|F_\delta\right\|_{\mathbf{P},2}, \mathcal{F}_\delta, L_2\left(\mathbf{P}\right)\right)}d\varepsilon.$$

See e.g. Theorems 2.14.1 and 2.14.2 in van der Vaart and Wellner (1996). These are bounded by Assumption 5.4. Thus, in case when the rate r_n is not known *a priori*, it is typical that $\phi^2(\delta) = \mathbf{P}F_\delta^2$ yields the correct rate, leading to the rate as the solution of $r_n^4 \mathbf{P}F_{1/r_n}^2 \sim n$. This is in fact the first condition in Assumption 5.3.

We now provide sufficient conditions for Assumption 5.4. Many interesting examples feature the estimating function q in the form of Lipschitz of order r transformation in the sense that $q_{\theta,\gamma} = q(y, g(\mathbf{w}, \theta, \gamma))$ and

$$|q(y, g(\mathbf{w}; \theta_1, \gamma_1)) - q(y, g(\mathbf{w}, \theta_2, \gamma_2))| \leq L_r(\mathbf{w}) |g(\mathbf{w}; \theta_1, \gamma_1) - g(\mathbf{w}; \theta_2, \gamma_2)|^r,$$

where L_r is square integrable in $\mathbf{P}$. In this case, verification of the entropy conditions and the conditions on the envelope function is straightforward as in the following lemma.

Lemma 5.4. *Suppose that $\mathcal{F}_\delta$ is a class of functions $q_{\theta,\gamma}$, which are Lipschitz of order $r \in (0,1]$ transformations, where $|\theta - \theta_0| < \delta$ and $\gamma \in \Gamma$. Then, for some $\delta_0 > 0$,*

$$\int_0^1 \sup_{\delta < \delta_0} \sup_Q \sqrt{\log N\left(\varepsilon \|F_\delta\|_{Q,2}, \mathcal{F}_\delta, L_2(Q)\right)} d\varepsilon < \infty.$$

Let $\phi(\delta) = \delta^r$. Then, there exists an envelope function F_δ such that for every $\eta > 0$,

$$\lim_{\delta \to 0} \phi^{-2}(\delta) \ \mathbf{P}F_\delta^2 1\left\{F_\delta > \eta \delta^{-2} \phi^2(\delta)\right\} = 0.$$

The lemma specifies the functional form of $\phi(\delta)$ as δ^r, resulting in the convergence rate $r_n = n^{1/(4-2r)}$, upon verifying conditions on $\mathbf{P}q_{\theta,\gamma}$. There are quite a few examples that are Lipschitz of order 1. They include the quantile regression model and the probit model in Section 3.

If $\mathbf{P}q_{\theta,\gamma}$ is twice continuously differentiable at $\theta = \theta_0$ with a unique maximum at θ_0, Assumptions 5.1 and 5.2 may be implied by other conditions as discussed above. Then, the following corollary is more convenient to apply than the main theorem. It provides conditions under which $G_2(h, \gamma)$ is quadratic in h for a given γ and most applications belong to this case.

Corollary 5.5. *Suppose that the function* $Q(\theta,\gamma)$ *has a well-separated maximum* θ_0 *in the sense of (5.7) and it is twice continuously differentiable at* θ_0 *with a negative second derivative matrix, say* $-V(\gamma)$*, whose maximum eigenvalues are bounded away from zero for all* γ*. Let* V_β *denote the block of* $V(\gamma)$ *that is associated with the second derivative with respect to* β*. Then,* $r_n^2 \mathbf{P} m_{h_n,\gamma} \longrightarrow -\frac{1}{2}h_\theta' V(\gamma) h_\theta + \frac{1}{2} h_b' V_\beta h_b = G_2(h,\gamma)$*, uniformly over any compact set. If Assumptions 5.3 and 5.4 hold with a sequence* r_n*, then* $r_n\left(\hat{\theta}(\gamma) - \theta_0\right) = O_{p\gamma}(1)$ *and* $r_n\left(\tilde{\beta} - \beta_0\right) = O_p(1)$*. If Assumption 5.5 holds as well, then*

$$QLR_n \Rightarrow \sup_\gamma \left[\sup_{h:h_b=0} G(h,\gamma) - \sup_{h:h_\theta=0} \left(-G(h,\gamma)\right)\right].$$

If in addition $G_1(h,\gamma)$ is linear in h for a given γ, then a more explicit form of the asymptotic null distribution is available. By construction, we may write

$$G_1(h,\gamma) = h'\mathcal{G}(\gamma) = (h_\beta - h_b)' \mathcal{G}_1 + h_\alpha \mathcal{G}_2(\gamma),$$

where $\mathcal{G}(\gamma) = \left(\mathcal{G}_1', \mathcal{G}_2(\gamma)'\right)'$ is a Gaussian process with some covariance kernel $K(\gamma_1,\gamma_2)$. Then, simple algebra shows that the limiting distribution of QLR_n has the form

$$\frac{1}{2}\left[\sup_\gamma \mathcal{G}(\gamma)'V(\gamma)^{-1}\mathcal{G}(\gamma) - \mathcal{G}_1' V_\beta^{-1}\mathcal{G}_1\right]. \tag{5.10}$$

Standard linear algebra allows us to write this as

$$\frac{1}{2}\sup_\gamma \mathcal{G}(\gamma)' H_\alpha(\gamma) H_\alpha(\gamma)' \mathcal{G}(\gamma),$$

where $H_\alpha(\gamma)$ is a full-column rank matrix whose rank is the dimension of α, say k_α. Furthermore, if efficient estimators are used for both restricted and unrestricted models, then for each γ, $H_\alpha(\gamma)' \mathcal{G}(\gamma)$ is distributed as standard multivariate normal with dimension k_α. Thus, $2QLR_n$ converges in distribution to the supremum of a chi-square process indexed by γ. This is the case with the homoskedastic linear regression model with ordinary least squares estimators (Hansen, 1996) and also with maximum likelihood estimators for logit and probit models.

6. Examples

This section presents a few well-known statistical models as examples to illustrate how to check the regularity conditions given in Section 5.

6.1. **Maximum Score Estimation.**

The estimating function for the maximum score estimation is

$$q(y, \mathbf{w}; \theta, \gamma) = (2y-1)\,1\{\mathbf{x}'\beta + \mathbf{z}'\alpha 1\{t>\gamma\} \geq 0\}. \tag{6.1}$$

To check the entropy condition (5.3) in Assumption 5.4, we show that the class $\mathcal{F}$ of these functions $q_{\theta,\gamma}$, where θ and γ belong to any compact subset in the Euclidean space, is a VC class of functions. Indeed, the set $\{\mathbf{x}'\beta + \mathbf{z}'\alpha 1\{t>\gamma\} \geq 0\}$ can be represented by union and intersection of half-spaces in the Euclidean space. Since half spaces are VC class of sets and the VC feature is preserved under unions and intersections, see Lemma 2.6.17 in van der Vaart and Wellner (1996), the sets constitute a VC class, so do the indicator functions of the sets. Now, $\mathcal{F}_\delta = \{q_{\theta,\gamma} : |\theta-\theta_0| < \delta, \gamma \in \Gamma\}$ is also a VC class of functions of the same index at most as $\mathcal{F}$. Thus, the covering numbers of $\mathcal{F}_{K/r_n}$ is bounded in a polynomial in $(1/\varepsilon)^{-1}$, not depending on n, and thus (5.3) is satisfied.

To find an envelope function for $\mathcal{F}_\delta$, note that $|q_{\theta,\gamma} - q_{\theta_0,\gamma}| \leq 1$ and that it takes nonzero values only when $\mathbf{x}'\beta_0$ and $\mathbf{x}'\beta + \mathbf{z}'\alpha 1\{t>\gamma\}$ take different signs. The latter implies that the distance between the two is greater than $\mathbf{x}'\beta_0$ in absolute values. Thus,

$$|\mathbf{x}'\beta_0| \leq \max_{\theta:|\theta-\theta_0|\leq\delta} |\mathbf{x}'(\beta-\beta_0) + \mathbf{z}'\alpha 1\{t>\gamma\}| \leq 2\delta\,|\mathbf{x}|,$$

which yields an envelope function

$$F_\delta = 1\{|\mathbf{x}'\beta_0| \leq 2\delta\,|\mathbf{x}|\}.$$

It is shown in Theorem 6.1 below under some regularity conditions on $\mathbf{P}$ that this envelope function satisfies the conditions in Assumption 5.3 with the rate obtained in Kim and Pollard (1990) for $\tilde{\beta}$, that is, with $r_n = n^{1/3}$.

The following assumption is imposed, which is somewhat more restrictive than required to simplify the exposition. Let $\mathbf{W}_\gamma = (\mathbf{X}', \mathbf{Z}'1\{T > \gamma\})'$.

Assumption 6.1. (i) *The parameter* θ *has unit length, that is,* Θ *is the surface of the unit sphere in* $\mathbb{R}^k$, *and* $\gamma \in \Gamma$, *which is an open subset of the support of* T.
(ii) *The distribution of* U *conditional on* $\mathbf{W} = \mathbf{w}$, *denoted by* $F_{U|\mathbf{W}}(\cdot|\mathbf{w})$, *is absolutely continuous with respect to Lebesgue measure and the corresponding conditional density is uniformly continuous and positive everywhere with probability one. In addition,* $F_{U|\mathbf{W}}(0|\mathbf{w}) = 0.5$ *for almost every* $\mathbf{w}$ *and it is continuously differentiable with respect to* $\mathbf{w}$.
(iii) $\mathbf{X}$ *has a continuously differentiable density* $p_{\mathbf{X}}(\cdot)$ *and the angular component of* $\mathbf{X}$, *considered as a random element in the unit sphere, has a bounded, continuous density with respect to surface measure on the sphere. Furthermore, the density* $p_{\mathbf{X}}$ *has compact support.*
(iv) $\int 1\{\mathbf{x}'\beta_0 = 0\}\ p_{\mathbf{W}}(\mathbf{w})\, d\varpi > 0$, *where* ϖ *denotes the Lebesque measure on* $\{\mathbf{w} : \mathbf{x}'\beta_0 = 0\}$.
(v) T *is continuously distributed.*

The following theorem shows that conditions in Theorem 5.1 are satisfied.

Theorem 6.1. *Suppose Assumption 6.1 hold and* h, h_1 *and* h_2 *belong to the null space of* ξ_0. *Let* $\ell(\mathbf{w}; h_1, h_2, \gamma_1, \gamma_2)$ *be the sum of the lengths of two intervals* $(I_1 - I_2)$ *and* $(I_2 - I_1)$, *where* I_1 *is the interval between* $\mathbf{w}'_{\gamma_1} h_{1\theta}$ *and* $\mathbf{w}'_{\gamma_2} h_{2\theta}$ *and* I_2 *is the interval between* $\mathbf{x}'h_{1b}$ *and* $\mathbf{x}'h_{2b}$. *Also, let*

$$\begin{aligned} \kappa(\mathbf{w}) &:= E\left[1\left\{g\left(\mathbf{W};\theta_0,\gamma_0\right)+U\geq 0\right\}-1\left\{g\left(\mathbf{W};\theta_0,\gamma_0\right)+U<0\right\}|\mathbf{W}=\mathbf{w}\right] \\ &= 1-2F_{U|\mathbf{W}}[-g(\mathbf{w},\theta_0,\gamma_0)|\mathbf{w}]. \end{aligned} \tag{6.2}$$

Then, the covariance kernel of the limit Gaussian process G_1 *is characterized by*

$$E\left(G_1\left(h_1,\gamma_1\right)-G_1\left(h_2,\gamma_2\right)\right)^2=\int \ell\left(\mathbf{w};h_1,h_2,\gamma_1,\gamma_2\right)\ 1\left\{\mathbf{x}'\beta_0=0\right\}\ p_{\mathbf{W}}\left(\mathbf{w}\right)d\varpi,$$

and

$$G_2(h,\gamma) = \int \left((\mathbf{x}'h_b)^2 - (\mathbf{w}'_\gamma h_\theta)^2 \right) 1\{\mathbf{x}'\beta_0 = 0\} [(\partial/\partial \mathbf{x}')\kappa(\mathbf{w})\beta_0] p_{\mathbf{W}}(\mathbf{w}) d\varpi.$$

Furthermore,

$$QLR_n \Rightarrow \sup_\gamma \left[\sup_{h:h_b=0} G(h,\gamma) - \sup_{h:h_\theta=0} (-G(h,\gamma)) \right],$$

where $q_{\theta,\gamma}$ for QLR_n is defined using (6.1) and $G = G_1 + G_2$.

Theorem 6.1 establishes the asymptotic null distribution for the maximum score estimation. The corresponding distribution is nonstandard and cannot be tabulated; however, statistical inference can be carried out by subsampling as in Delgado et al. (2001). Since without Theorem 6.1, it would be difficult to obtain the validity of subsampling, one of the merits of Theorem 6.1 is to provide the asymptotic validity of subsampling.

6.2. **The Probit Model.** We now verify regularity conditions for the probit model. Note that the function $q(y,\mathbf{w};\theta,\gamma)$ in (3.3) is Lipschitz of order 1 transformation and twice continuously differentiable in θ. Therefore, applying Lemma 5.4 and Corollary 5.5, we only need to check the separability condition (5.7) and Assumption 5.5. We assume the following regularity conditions:

Assumption 6.2. *(i) The parameters θ and γ are in the interior of compact sets Θ and Γ where Γ is contained in an open subset of the support of T.*
(ii) For any γ, the matrix $E\left[\mathbf{W}_\gamma \mathbf{W}'_\gamma\right]$ exists and is nonsingular.
(iii) T is continuously distributed.

We first verify the separability condition. Let γ be given. Since $E\left[\mathbf{W}_\gamma \mathbf{W}'_\gamma\right]$ is nonsingular, it is positive definite. This implies that $\mathbf{W}'_\gamma\theta_0 \neq \mathbf{W}_\gamma\theta$ for any $\theta \neq \theta_0$. Therefore, strict monotonicity of $\Phi(\cdot)$ assures identification for each γ, which establishes the separability condition.

Since $q(\cdot)$ is twice continuously differentiable, it follows from the discussion following Corollary 5.5 that the limiting distribution of the test statistic is the supremum of a chi-square process indexed by γ as in (5.10). Then the desired result in (3.4) follows. Using identical arguments, we can obtain the null asymptotic distribution of the test statistic for the logit model. In general, similar arguments can apply to statistical models for which the test statistic can be constructed based on the maximum likelihood estimator.

6.3. **Quantile regression.** Note that the function $q(y, \mathbf{w}; \theta, \gamma)$ in (3.5) is Lipschitz of order 1 as a function of $g(\mathbf{w}, \theta, \gamma)$. Therefore, the bracketing entropy condition (5.4) and the condition on the envelope function in Assumption 5.3 are satisfied due to Lemma 5.4. Furthermore, $r_n = \sqrt{n}$. We verify the other conditions in Corollary 5.5.

Assume that the density of Y conditional on $\mathbf{W}$ and that of T conditional on the other elements in $\mathbf{W}$ exist and are continuously differentiable with uniformly bounded derivatives. Let $f_{Y|\mathbf{W}}(\cdot|\mathbf{w})$ denote the conditional density of Y given $\mathbf{W} = \mathbf{w}$. Then, $\mathbf{P}q_{\theta,\gamma}$ is twice continuously differentiable in any (θ, γ), implying the last condition in Assumption 5.3 is satisfied. Furthermore, G_1 is linear in h as to be shown below. Thus, the limit distribution is characterized by (5.10) with

$$V(\gamma) = E\left[\mathbf{W}_\gamma \mathbf{W}_\gamma' f_{Y|\mathbf{W}}(\mathbf{X}'\beta_0|\mathbf{W})\right],$$

and a mean-zero Guassian process $\mathcal{G}$ with covariance kernel in (3.7). To see this, write

$$\Delta_\tau(a) := \rho_\tau(y - a) - \rho_\tau(y - a_0) - [1(y < a_0) - \tau](a - a_0).$$

Then as in Pollard (1991), simple algebra yields that

$$|\Delta_\tau(a)| \leq |a - a_0| 1\{|y - a_0| \leq |a - a_0|\}. \tag{6.3}$$

Define

$$\varphi_{h,\gamma}(y, \mathbf{w}) := [1(y < \mathbf{x}'\beta_0) - \tau] \left(\mathbf{x}'(h_\beta - h_b) + \mathbf{z}' h_\alpha 1\{t > \gamma\}\right).$$

By (6.3),

$$
\begin{aligned}
& nE\left|m_{h_n,\gamma}-\varphi_{h_n,\gamma}\right|^2 \\
\leq\; & E\left|\mathbf{W}_\gamma' h_\theta\right| 1\left\{\left|Y-\mathbf{X}'\beta_0\right| \leq \frac{\left|\mathbf{W}_\gamma' h_\theta\right|}{\sqrt{n}}\right\}+E\left|\mathbf{X}' h_b\right| 1\left\{\left|Y-\mathbf{X}'\beta_0\right| \leq \frac{\left|\mathbf{X}' h_b\right|}{\sqrt{n}}\right\} \\
\to\; & 0.
\end{aligned}
$$

Therefore, the covariance kernel is given by that of $\varphi_{h,\gamma}$ by applications of Cauchy-Schwarz inequality, which is given by (3.7).

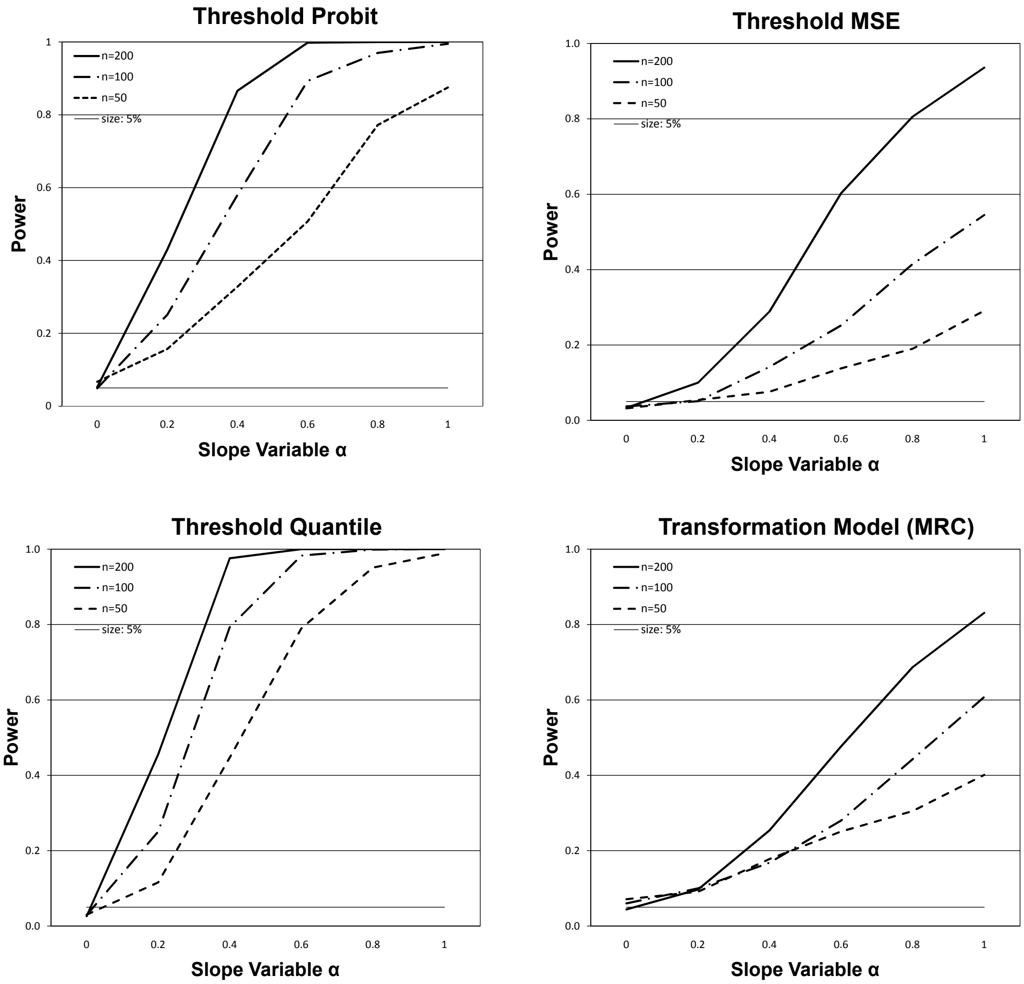
Note that in this example, the asymptotic null distribution is not the supremum of a chi-square process indexed by γ. This is due to the fact that the quantile regression estimator is not an efficient estimator. However, critical values can be simulated by the same method as in the maximum likelihood estimation, which was illustrated in section 3.

7. Monte Carlo Simulations

In this section, we report Monte Carlo simulation results for all four examples considered in the article. Details of simulation designs and testing procedures are provided in the online supplement.

Figure 2 summarizes the results of the simulation study, by showing the power functions for four examples with three different sample sizes: $n = 50, 100$, and 200. The top right and left panels report results from the probit example and those from the maximum score estimation example, respectively. In addition, the bottom right and left panels report results from the quantile regression example and those from the maximum rank correlation estimation example, respectively. First of all, the figure shows the finite sample size of the test when the nominal level is 5%. Under the null hypothesis ($\alpha = 0$), the rejection rates of the test are close to the nominal level in most cases. Secondly, Figure 2 shows the power of the test when α increases from 0 to 1. The result indicates that, in all cases, the power increases fast as the parameter value of α is farther away from zero and also it increases as n gets large.

Figure 1. Power Functions of Threshold Models



8. Conclusions

We have developed a general testing procedure for threshold effects and have proposed a new method for establishing the asymptotic null distribution. Since the new approach does not require to approximate the objective function in a quadratic form, we can construct the test statistic for nonstandard cases like the maximum score estimation. Furthermore, we have proposed the test statistic when the objective function is a U-process. We believe our approach would prove useful in many other occasions where objective function based inferences are made.

Fan et al. (2001) show that a class of the generalized likelihood statistics based on some appropriate nonparametric estimators are asymptotically chi-squared in nonparametric testing problems. However, they do not consider the Davies problem. It is an interesting research topic to see whether one can generalize the methodology of Fan et al. (2001) to cover the case when there may be a nuisance parameter that appears under the alternative, but not under the null. Such examples are partially linear regression models and varying coefficient models with a change-point due to a covariate threshold.

We have provided local power results in the online supplements, but have not established the asymptotic admissibility of the test we proposed. Andrews and Ploberger (1995) have established the asymptotic admissibility of the likelihood ratio test. It might be possible to extend their results to more general cases, including our QLR statistics. Alternatively, following Andrews and Ploberger (1994) and Song et al. (2009), we may introduce a class of tests of the following form:

$$Exp - QLR_n = r_n^2 \left[(1+c)^{-p/2} \int \exp\left(\frac{1}{2} \frac{c}{c+1} QLR_n(\gamma) \right) dJ(\gamma) \right],$$

where p is the dimension of b, $J(\cdot)$ is a prespecified weight function over values of γ in Γ, and c is a prespecified scalar constant. It might be possible to establish that $Exp - QLR_n$ has some weighted average power properties in our setup, along the lines of Andrews and Ploberger (1994) and Song et al. (2009). These are interesting topics for future research.

Appendices

The appendices contain all the mathematical proofs, additional theoretical results, and Monte Carlo simulation results. In particular, (i) we provide the proofs of all the theorems; (ii) we provide asymptotic theory for the case when objective functions are based on U-processes and verify regularity conditions for the maximum rank correlation (MRC) estimator; (iii) we discuss the consistency and local power of the test when the null hypothesis is false; and (iv) we report details of Monte Carlo simulation designs and testing procedures.

Appendix A. Proofs of Theorems

Proof of Theorem 5.1. As supremum is a continuous operator, we need to establish the weak convergence of the process $r_n^2 \mathbb{P}_n m_{\xi,\gamma}$. Also note that in view of Lemma 2.5 of Kim and Pollard (1990) the supremum of the limit process G is $O_p(1)$ under Assumption 5. Under the uniform convergence rate given in Assumption 2, it is sufficient to consider the process $r_n^2 \mathbb{P}_n m_{\xi,\gamma}$ only on the r_n^{-1} neighborhood of ξ_0. Furthermore, given the decomposition (3.2) of $r_n^2 \mathbb{P}_n m_{\xi,\gamma}$ and Assumption 5, it remains to obtain the weak convergence of the empirical process indexed by the sequence of classes of functions

$$\mathcal{M}_n = \left\{ \frac{r_n^2}{\sqrt{n}} m_{h_n,\gamma} : |h| \leq K, \gamma \in \Gamma \right\},$$

where h_n is defined in Assumption 5. That is, we derive the weak convergence of

$$\frac{r_n^2}{\sqrt{n}} \mathbb{G}_n m_{h_n,\gamma},$$

for any given $K > 0$. Then, an application of the continuous mapping theorem concludes the proof.

We apply either Theorem 2.11.22 or 2.11.23 of van der Vaart and Wellner (1996) to obtain the weak convergence. While our assumptions are sufficient for both theorems, they are presented in terms of functions q not of functions m. Thus, we need to show that the conditions on q are preserved under the transformation yielding m. First, we verify that the boundedness of both entropy conditions (5.3) and (5.4) is preserved under summation. For the latter, note that the definition of the bracketing numbers implies that for two classes $\mathcal{F}$ and $\mathcal{G}$ of functions,

$$N_{[]}\left(2\varepsilon, \mathcal{F} + \mathcal{G}, L_r(Q)\right) \leq N_{[]}\left(\varepsilon, \mathcal{F}, L_r(Q)\right) N_{[]}\left(\varepsilon, \mathcal{G}, L_r(Q)\right).$$

Therefore, the bounded entropy condition (5.4) for $\mathcal{F}_\delta$ implies the boundedness of the entropy condition for the class

$$\mathcal{M}_\delta = \{m_{\xi,\gamma} : |\xi - \xi_0| < \delta\}.$$

For the former case of uniform entropy, we refer to Theorem 2.10.20 of van der Vaart and Wellner (1996), which shows

$$\log N\left(\varepsilon \left\|2F_\delta\right\|_{Q,2}, \mathcal{F}_\delta + \mathcal{F}_\delta, L_2(Q)\right) \leq 2\log N\left(\varepsilon \left\|F_\delta\right\|_{Q,2}, \mathcal{F}_\delta, L_2(Q)\right).$$

Thus, it is shown that the class $\mathcal{M}_\delta$ also satisfies either the entropy conditions (5.3) or (5.4), which also satisfies that of either Theorem 2.11.22 or 2.11.23 of van der Vaart and Wellner (1996). An envelope function M_δ for $\mathcal{M}_\delta$ is given by $2F_\delta$, which satisfies the conditions of Theorem 2.11.22 or 2.11.23 of van der Vaart and Wellner (1996) under Assumption 3. This completes the proof. ∎

Proof of Lemma 5.2. The proof of this lemma is omitted since the first conclusion is a restatement of Theorems 2.4.1 and 2.4.3 in van der Vaart and Wellner (1996) and the second conclusion follows immediately from Lemma A-1 of Andrews (1993). Specifically speaking, Condition (a) of Lemma A-1 of Andrews (1993) is satisfied by Theorem 2.4.1 and Theorem 2.4.3 in van der Vaart and Wellner (1996). ∎

Proof of Lemma 5.3. To prove the lemma, we modify the peeling device in the proof of Theorem 3.2.5 in van der Vaart and Wellner (1996). For each n, the parameter space can be partitioned into the shells $S_{j,n} = \{\theta : 2^{j-1} < r_n |\theta - \theta_0| \leq 2^j\}$ for integer j's. If $r_n \sup_\gamma \left|\hat{\theta}(\gamma) - \theta_0\right|$ is larger than 2^M for a given M, $\hat{\theta}(\gamma)$ is in one of the shells for some γ. In that case $\sup_{\theta,\gamma} \mathbb{P}_n q_{\theta,\gamma} - \mathbb{P}_n q_{\theta_0} \geq 0$, where the supremum is taken over the shell for θ

and $\gamma \in \Gamma$, due to the definition of $\hat{\theta}$. Therefore, for any $\eta > 0$,

$$\Pr\left(r_n \sup_{\gamma} \left|\hat{\theta}(\gamma) - \theta_0\right| > 2^M\right) \leq \sum_{j \geq M, 2^j \leq \eta r_n} \Pr\left\{\sup_{\theta \in S_{j,n};\gamma} \mathbb{P}_n q_{\theta,\gamma} - \mathbb{P} q_{\theta_0} \geq 0\right\} + \Pr\left(2 \sup_{\gamma} \left|\hat{\theta}(\gamma) - \theta_0\right| \geq \eta\right).$$

The two terms in the right side of the inequality can be shown to be made arbitrarily small by the same argument in the proof of Theorem 3.2.5 in van der Vaart and Wellner (1996). ∎

Proof of Lemma 5.4. Let $\mathcal{G}_\delta$ denote the collection of $g(\mathbf{w};\theta,\gamma)$s such that $|\theta - \theta_0| < \delta$ and $\bar{g}_\delta$ an envelope function of $\mathcal{G}_\delta$. Then, it follows from Theorem 2.10.20 of van der Vaart and Wellner (1996) that the uniform entropy integral of $\mathcal{F}_\delta$ is bounded by

$$\int_0^1 \sup_Q \sqrt{\log N\left(\varepsilon \left\|\bar{g}_\delta\right\|_{Q,2r}, \mathcal{G}_\delta, L_{2r}(Q)\right)} \frac{d\varepsilon}{\varepsilon^{1-r}}, \tag{A.1}$$

where the supremum is taken over all finitely discrete probability measures Q. It is clear that $\mathcal{G}_\delta$ constitutes a VC class of functions since the subgraphs are represented by intersections and unions of half spaces as discussed in Section 6.1. Since the VC-index does not depend on δ and the covering number of a VC class is polynomial in $(1/\varepsilon)$, the uniform entropy integral in (A.1) is bounded uniformly in δ. This in turn implies that the uniform entropy integral condition in (5.3).

An envelope function for $\mathcal{F}_\delta$ is given by $F_\delta = 2L_r \cdot \bar{g}_\delta^r$ since

$$|q(y, g(\mathbf{w};\theta,\gamma)) - q(y, g(\mathbf{w};\theta_0,\gamma_0))|^2 \leq 4L_r^2(\mathbf{w})\, \bar{g}^{2r}(\mathbf{w}).$$

To check the conditions on the envelope function, note that $\bar{g}_\delta(\mathbf{w}) = |\mathbf{w}|\,\delta$ is an envelope function for $\mathcal{G}_\delta$ using the Cauchy-Schwarz inequality. Then, it is straightforward to see that with $\phi(\delta) = \delta^r$

$$\begin{aligned}\phi^{-2}(\delta)\ \mathbf{P}F_\delta^2 1\left\{F_\delta > \eta\delta^{-2}\phi^2(\delta)\right\} &= E\left(4L_r^2(\mathbf{W})|\mathbf{w}|^2 1\{2L_r(\mathbf{W})|\mathbf{w}|\delta^r > \eta\}\right) \\ &\to 0,\end{aligned}$$

as $\delta \to 0$ for any $\eta > 0$. ∎

Proof of Corollary 5.5. From the second order expansion of $\mathbf{P}m_{h_n,\gamma}$, the functional form of G_2 is obvious since the first derivative is zero at $\xi = \xi_0$ from the first order condition. The consistency of $\hat{\theta}(\gamma)$ (and thus $\tilde{\beta}$) follows from Lemma 5.2, and the convergence rates follow from Lemma 5.3 since (5.8) is satisfied due to the presence of the second derivative matrix $V_\theta(\gamma)$ and (5.9) is implied by Assumption 5.4. The last convergence then follows from Theorem 5.1. ∎

Proof of Theorem 6.1. We prove the conditions in Theorem 5.1. Assumption 5.4 is also discussed in the text. The first two conditions for the envelope function F_{K/r_n} in Assumption 3 are verified in Kim and Pollard (1989).

To show the uniform consistency, we need to verify the condition (5.7). Since γ_0 is not identified, it is an arbitrary fixed number, say, zero. Then it can be shown that

$$\begin{aligned}\Delta^*(\theta,\gamma) &= \mathbf{P}\left(q_{\theta_0,\gamma_0} - q_{\theta,\gamma}\right) \\ &= E\Big[\kappa(\mathbf{W})\Big(1\left\{g(\mathbf{W};\theta,\gamma) \geq 0 > g(\mathbf{W};\theta_0,\gamma_0)\right\} \\ &\qquad - 1\left\{g(\mathbf{W};\theta_0,\gamma_0) \geq 0 > g(\mathbf{W};\theta,\gamma)\right\}\Big)\Big]. \end{aligned} \tag{A.2}$$

By the assumption that $F_{U|\mathbf{W}}[0|\mathbf{w}] = 0.5$, note that $\kappa(\mathbf{w}) \geq 0$ when $g(\mathbf{w};\theta_0,\gamma_0) \geq 0$ and that $\kappa(\mathbf{w}) < 0$ when $g(\mathbf{w};\theta_0,\gamma_0) < 0$. Define

$$\mathcal{Q}(\theta,\gamma) = \Big[\mathbf{w} \in \text{supp}(\mathbf{W}) : \{g(\mathbf{w};\theta,\gamma) \geq 0 > g(\mathbf{w},\theta_0,\gamma_0)\} \cup \{g(\mathbf{w},\theta_0,\gamma_0) \geq 0 > g(\mathbf{w},\theta,\gamma)\}\Big].$$

By arguments identical to those used to prove Proposition 2 of Manski (1988), θ_0 is identified if and only if $\inf_\gamma \Pr(\mathbf{W} \in \mathcal{Q}(\theta,\gamma)) > 0$ for any $\theta \neq \theta_0$. Therefore, $\sup_\gamma \Delta^*(\theta,\gamma)$ is non-positive everywhere and is equal to zero only when $(\theta,\gamma) = (\theta_0,\gamma_0)$.

The uniform convergence with $r_n = n^{1/3}$ can be argued from Lemma 5.3 upon proving (5.8), which will be verified when we derive the limit of $r_n^2 \mathbf{P} m_{h_n,\gamma}$.

Now we present the covariance kernel of the limit gaussian process G_1 and verify the last condition in Assumption 3. The following decomposition is useful: for any ξ_1, ξ_2, γ_1, and γ_2,we have

$$
\begin{aligned}
& r_n \mathbf{P} \left(m_{\xi_1,\gamma_1} - m_{\xi_2,\gamma_2} \right)^2 \\
& = r_n \mathbf{P} \left(\left| 1\left\{ g\left(\mathbf{W},\theta_1,\gamma_1\right) \geq 0 \right\} - 1\left\{ g\left(\mathbf{W},\theta_2,\gamma_2\right) \geq 0 \right\} \right| \right) \\
& + r_n \mathbf{P} \left(\left| 1\left\{ \mathbf{X}'\beta_1 \geq 0 \right\} - 1\left\{ \mathbf{X}'\beta_2 \geq 0 \right\} \right| \right) \\
& - r_n 2\mathbf{P} \left(1\left\{ g\left(\mathbf{W},\theta_1,\gamma_1\right) \geq 0 \right\} - 1\left\{ g\left(\mathbf{W},\theta_2,\gamma_2\right) \geq 0 \right\} \right) \left(1\left\{ \mathbf{X}'\beta_1 \geq 0 \right\} - 1\left\{ \mathbf{X}'\beta_2 \geq 0 \right\} \right) \\
& =: A_1 + A_2 + A_3.
\end{aligned}
$$

Some reparameterization and change of variables are useful. In particular to impose the normalization restriction, $|\theta| = |b| = 1$, we characterize the localized parameters as

$$
\begin{aligned}
h_n &= \left(h'_{n\beta}, h'_{n\alpha}, h'_{nb} \right)' \\
&= \left(\sqrt{1 - \left| \dot{h}_\theta / r_n \right|^2} \beta'_0 + \dot{h}'_\beta / r_n, h'_\alpha / r_n, \sqrt{1 - \left| \dot{h}_b / r_n \right|^2} \beta'_0 + \dot{h}'_b / r_n \right)',
\end{aligned}
$$

where $\left|\dot{h}_\beta\right|, \left|\dot{h}_b\right| < K$ and $\dot{h}_\beta$ and $\dot{h}_b$ are orthogonal to β_0. Note here that since $\alpha_0 = 0$ the parameter h_α is not constrained. Let g_n be defined in the same way. Note here that we index by h and g rather than h_1 and h_2 to ease the exposition. Accordingly, decompose $\mathbf{x}$ into

$$
\mathbf{x} = \zeta \beta_0 + \eta, \tag{A.3}
$$

where $\beta_0'\eta = 0$. Then, $\mathbf{x}'h_{n\beta} = \zeta\sqrt{1 - \left|\dot{h}_\theta/r_n\right|^2} + \eta'\dot{h}_\beta/r_n$. Now, take A_1 and note that

$$A_1 = r_n\mathbf{P}1\left\{g\left(\mathbf{W},\theta_1,\gamma_1\right) \geq 0 > g\left(\mathbf{W},\theta_2,\gamma_2\right)\right\} + r_n\mathbf{P}1\left\{g\left(\mathbf{W},\theta_2,\gamma_2\right) \geq 0 > g\left(\mathbf{W},\theta_1,\gamma_1\right)\right\}.$$

Let $\mathbf{Z}_\gamma = \mathbf{Z}\ 1\{T > \gamma\}$ and $p_{\mathbf{W}}\left(\mathbf{x},\mathbf{w}_{-\mathbf{x}}\right) = p_{\mathbf{W}}\left(\mathbf{w}\right)$ for $\mathbf{W} = \left(\mathbf{X}',\mathbf{W}_{-\mathbf{x}}'\right)'$. Also recall that $\mathbf{z} = R'\mathbf{x} = R'\beta_0\zeta + R'\eta$ and let $R_\gamma = R1\{T > \gamma\}$. Take the first term with substitution of (A.3), $\theta_1 = h_{n\theta}$ and $\theta_2 = g_{n\theta}$, that is, consider

$$r_n\mathbf{P}\left\{\zeta\left(\sqrt{1 - \left|\frac{\dot{h}_\theta}{r_n}\right|^2} + \frac{\beta_0'R_{\gamma_1}h_\alpha}{r_n}\right) + \frac{\eta'}{r_n}\left(\dot{h}_\beta + R_{\gamma_1}h_\alpha\right)\right.$$
$$\left.\geq 0 > \zeta\left(\sqrt{1 - \left|\frac{\dot{g}_\theta}{r_n}\right|^2} + \frac{\beta_0'R_{\gamma_2}g_\alpha}{r_n}\right) + \frac{\eta'}{r_n}\left(\dot{g}_\beta + R_{\gamma_2}g_\alpha\right)\right\},$$

which becomes, after rearranging terms and changing the variable $\xi = \dot{\zeta}/r_n$,

$$\iint 1\left\{\frac{\eta'\left(\dot{h}_\beta + R_{\gamma_1}h_\alpha\right)}{-\sqrt{1-\left|\dot{h}_\theta/r_n\right|^2} - \beta_0'R_{\gamma_1}h_\alpha/r_n} \leq \dot{\zeta} < \frac{\eta'\left(\dot{g}_\beta + R_{\gamma_2}g_\alpha\right)}{-\sqrt{1-\left|\dot{g}_\theta/r_n\right|^2} - \beta_0'R_{\gamma_2}g_\alpha/r_n}\right\}$$
$$\times 1\left\{\eta'\beta_0 = 0\right\}p_{\mathbf{W}}\left(\frac{\dot{\zeta}}{r_n}\beta_0 + \eta,\mathbf{w}_{-\mathbf{x}}\right)d\dot{\zeta}d\varpi, \tag{A.4}$$

and converges to

$$\int\left(\eta'\left[\left(\dot{h}_\beta - \dot{g}_\beta\right) + R\left(h_\alpha 1\{t > \gamma_1\} - g_\alpha 1\{t > \gamma_2\}\right)\right]\right)_+ 1\left\{\eta'\beta_0 = 0\right\}p_{\mathbf{W}}\left(\eta,\mathbf{w}_{-\mathbf{x}}\right)d\varpi,$$

where $(\mathbf{x})_+ = \mathbf{x}1\{\mathbf{x} > 0\}$ and ϖ denote the Lebesque measure on $\{\mathbf{w} : \mathbf{x}'\beta_0 = 0\}$. The convergence follows from the dominated convergence theorem.

Then, after sorting out notation, the limit function of A_1 with $\xi_1 = h_n$ and $\xi_2 = g_n$ can be represented by

$$\iint\left|\mathbf{w}_{\gamma_1}'h_\theta - \mathbf{w}_{\gamma_2}'g_\theta\right|1\left\{\mathbf{x}'\beta_0 = 0\right\}p_{\mathbf{W}}\left(\mathbf{w}\right)d\varpi,$$

where h_θ and g_θ take values on the space orthogonal to θ_0. By the same reasoning, that of A_2 is given by

$$\iint\left|\mathbf{x}'\left[\left(h_b - g_b\right)\right]\right|1\left\{\mathbf{x}'\beta_0 = 0\right\}p_{\mathbf{W}}\left(\mathbf{w}\right)d\varpi,$$

where h_b and g_b orthogonal to β_0.

Note that the limit of A_1 (and A_2) is an integral of the length of the interval between two points indexed by (h_θ, γ_1) and (g_θ, γ_2) (and by h_b and g_b) and that the limit of A_3 is a composite of union and intersection of the two intervals. Then, using the notation in the text, we can write

$$E\left(G_1\left(h, \gamma_1\right) - G_1\left(g, \gamma_2\right)\right)^2 = \int \ell\left(\mathbf{w}; h, g, \gamma_1, \gamma_2\right)\ 1\left\{\mathbf{x}'\beta_0 = 0\right\}\ p_{\mathbf{W}}\left(\mathbf{w}\right) d\varpi,$$

for h and g in the null space of ξ_0. It can also be seen that the condition (5.2) in Assumption 5.3 is satisfied, observing (A.4) is bounded by $\left(\left|\dot{h}_\beta - \dot{g}_\beta\right| + \left|h_\alpha - g_\alpha\right|\right)\left(1 + o\left(1\right)\right)$.

Next, turn to $G_2\left(h, \gamma\right)$ or $r_n^2 \mathbf{P} m_{h_n,\gamma} = r_n^2 \mathbf{P}\left(q_{h_\theta,\gamma} - \tilde{q}_{h_b}\right)$. We analyze $r_n^2 \mathbf{P} q_{h_\theta,\gamma}$, then the limit of $r_n^2 \mathbf{P} \tilde{q}_{h_b}$ can be derived in the same way. Write $r_n^2 \mathbf{P} q_{h_\theta,\gamma} = B_{1n} + B_{2n}$, recalling (A.2), where

$$\begin{aligned} B_{1n} &= r_n^2 E\left(\kappa\left(\mathbf{W}\right) 1\left\{\mathbf{X}'\beta_0 < 0 \leq g\left(\mathbf{W}, h_n, \gamma\right)\right\}\right) \\ B_{2n} &= -r_n^2 E\left(\kappa\left(\mathbf{W}\right) 1\left\{\mathbf{X}'\beta_0 \geq 0 > g\left(\mathbf{W}, h_n, \gamma\right)\right\}\right). \end{aligned}$$

The first term B_{1n} can be written as, using the decomposition of $\mathbf{x}$ in (A.3) and keeping the relevant notation there,

$$\text{(A.5a)} \quad B_{1n} = r_n \iint 1\left\{\frac{\eta'\left(\dot{h}_\beta + R_\gamma h_\alpha\right)}{-\sqrt{1 - \left|\dot{h}_\theta / r_n\right|^2 - \beta_0' R_\gamma h_\alpha / r_n}} \leq \dot{\zeta} < 0\right\}$$

$$\text{(A.5b)} \quad \times 1\left\{\eta'\beta_0 = 0\right\} \kappa\left(\frac{\dot{\zeta}}{r_n}\beta_0 + \eta, \mathbf{w}_{-\mathbf{x}}\right) p_{\mathbf{W}}\left(\frac{\dot{\zeta}}{r_n}\beta_0 + \eta, \mathbf{w}_{-\mathbf{x}}\right) d\dot{\zeta} d\varpi,$$

which, using the fact that $\kappa\left(\mathbf{w}\right) = 0$ for $\mathbf{w}$ such that $\mathbf{x}'\beta_0 = 0$ and an expansion for κ with a mean value $\tilde{\zeta}$,

$$\kappa\left(\frac{\dot{\zeta}}{r_n}\beta_0 + \eta, \mathbf{w}_{-\mathbf{x}}\right) = \kappa\left(\eta, \mathbf{w}_{-\mathbf{x}}\right) + \frac{\partial}{\partial \mathbf{x}'}\kappa\left(\frac{\tilde{\zeta}}{r_n}\beta_0 + \eta, \mathbf{w}_{-\mathbf{x}}\right)\left(\frac{\dot{\zeta}}{r_n}\beta_0\right),$$

converges by the dominated convergence theorem to

$$\begin{aligned} & \int \dot{\zeta} 1 \left\{ -\mathbf{x}' \left(\dot{h}_\beta + R_\gamma h_\alpha \right) \leq \dot{\zeta} < 0 \right\} 1 \left\{ \mathbf{x}'\beta_0 = 0 \right\} \left[\left(\partial / \partial \mathbf{x}' \right) \kappa \left(\mathbf{w} \right) \beta_0 \right] p_{\mathbf{W}} \left(\mathbf{w} \right) d\dot{\zeta} d\varpi \\ = & -\frac{1}{2} \int \left(\mathbf{x}' \left(\dot{h}_\beta + R_\gamma h_\alpha \right) \right)^2 1 \left\{ \mathbf{x}'\beta_0 = 0 \right\} \left[\left(\partial / \partial \mathbf{x}' \right) \kappa \left(\mathbf{w} \right) \beta_0 \right] p_{\mathbf{W}} \left(\mathbf{w} \right) d\varpi. \end{aligned}$$

Similarly, we can see that $B_{2n} - (-B_{1n}) \to 0$. And the limit of $r_n^2 \mathbf{P} \tilde{q}_{h_b}$ is also obvious. Since this result also implies (5.8), we verified all the conditions of Theorem 5.1. ∎

Appendix B. Estimation Based on U-Processes

In this section, we provide asymptotic theory for the case with objective functions based on U-processes. To do so, let $\mathbb{U}_n$ denote the random discrete measure putting mass $2/n(n-1)$ for each of the points $\{(Y_i, \mathbf{W}_i, Y_j, \mathbf{W}_j) : 1 \leq i < j \leq n\}$. Since we assume that χ in (2.2) depends on (θ, γ) only through the regression function $g(\mathbf{W}, \theta, \gamma)$ in this case as well, arguments identical to those in Section 5.1 yields

$$\begin{aligned} QLR_n &= r_n^2 \left[\sup_{\theta,\gamma} Q_n(\theta, \gamma) - \sup_{\beta} Q_n(\beta) \right] \\ &= r_n^2 \sup_{\gamma} \left[\sup_{\xi: b = \beta_0} \mathbb{U}_n \mu_{\xi,\gamma} - \sup_{\xi: \theta = \theta_0} \left(-\mathbb{U}_n \mu_{\xi,\gamma} \right) \right], \end{aligned}$$

where

$$\mu_{\xi,\gamma}(y_i, \mathbf{w}_i, y_j, \mathbf{w}_j) := \chi_{\theta,\gamma}(y_i, \mathbf{w}_i, y_j, \mathbf{w}_j) - \tilde{\chi}_b(y_i, \mathbf{w}_i, y_j, \mathbf{w}_j)$$

and $\tilde{\chi}_b := \chi_{(b',0)',\gamma}$. Therefore, QLR_n is a continuous transformation of $r_n^2 \mathbb{U}_n m_{\xi,\gamma}$. General theory on U-processes provides a method for approximating $r_n^2 \mathbb{U}_n m_{\xi,\gamma}$ by its projection uniformly in ξ and γ (e.g. see Ghosal et al., 2000, Appendix). Therefore, the derivation of the asymptotic null distribution is similar to that of Section 5.1.

In this section, we consider the case with $r_n = n^{1/2}$ since all the estimators, which we are aware of, based on U-Processes are $n^{-1/2}$ consistent. To state an additional regularity condition for this section, consider a class of functions

$$\mathcal{M}_\delta = \{\chi_{\theta,\gamma} - \chi_{\theta_0,\gamma} : |\theta - \theta_0| < \delta, \gamma \in \Gamma\}$$

with an envelope function M_δ.

Assumption B.1 (Envelope Function and Entropy Condition)**.** *(1) Let* $\mathbf{Q}$ *denote the product measure* $\mathbf{P} \otimes \mathbf{P}$*. Then,* $\mathbf{Q}M^2_{K/n^{1/2}} \to 0$ *for any positive* $K < \infty$*. (2) For some* $\delta_0 > 0$*, we have that*

$$\int_0^1 \sup_{\delta<\delta_0} \sup_Q \log N\left(\varepsilon \left\|M_\delta\right\|_{Q,2}, \mathcal{M}_\delta, L_2(Q)\right) d\varepsilon < \infty.$$

The first condition in Assumption B.1 is reasonable given that $\mathcal{M}_\delta$ is defined only for a local neighborhood around θ_0. The entropy condition here is more stringent than that of Assumption 5.4. However, VC classes of functions have the covering numbers that are bounded by a polynomial in ε^{-1}, thus still satisfying condition (2) of Assumption B.1 as long as the VC indexes are bounded in n.

Consider a class of functions $\bar{\mathcal{F}}_\delta$ that is the same as in (5.1) with $q = \Pi\chi$, where

$$\Pi\chi_{\theta,\gamma}(y, \mathbf{w}) = 2\left[E\chi_{\theta,\gamma}(Y, \mathbf{W}, y, \mathbf{w}) - \mathbf{Q}\chi_{\theta,\gamma}\right].$$

An envelope function for $\bar{\mathcal{F}}_\delta$ is denoted by $\bar{F}_\delta$. In addition, define

$$\Pi\mu_{\xi,\gamma}(y, \mathbf{w}) = 2\left[E\mu_{\xi,\gamma}(Y, \mathbf{W}, y, \mathbf{w}) - \mathbf{Q}\mu_{\xi,\gamma}\right].$$

The following theorem establishes the asymptotic null distribution when an estimator is a maximizer of a U-Process.

Theorem B.1. *Let Assumptions 5.1 and 5.2 hold with* $r_n = n^{1/2}$*. Let Assumption 5.3 hold with* $F_\delta = \bar{F}_\delta$ *and* $q = \Pi\chi$*, and Assumption 5.5 hold with* $m = \Pi\mu$*. In addition, let Assumption B.1 hold. Then*

$$QLR_n \Rightarrow \sup_\gamma \left[\sup_{h:h_b=0} G(h,\gamma) - \sup_{h:h_\theta=0} \left(-G(h,\gamma)\right)\right],$$

where $G = G_1 + G_2$ *is redefined suitably with* $m = \Pi\mu$*.*

Proof of Theorem B.1. Define

$$\widehat{U}_n\mu_{\xi,\gamma} := \mathbf{Q}\mu_{\xi,\gamma} + \mathbb{P}_n\Pi\mu_{\xi,\gamma}.$$

Then by Theorem A.1 of Ghosal et al. (2000) and comments following this theorem, there exists a universal constant $C < \infty$ such that

$$\begin{aligned} & E\left(\sup_{\mu_{\xi,\gamma}\in\mathcal{M}_{K/n^{1/2}}} |\mathbb{U}_n\mu_{\xi,\gamma} - \widehat{U}_n\mu_{\xi,\gamma}|\right) \\ \leq\ & Cn^{-1}(\mathbf{Q}M^2_{K/n^{1/2}})^{1/2}\int_0^1 \sup_Q \log N\left(\varepsilon\left\|M_{K/n^{1/2}}\right\|_{Q,2}, \mathcal{M}_{K/n^{1/2}}, L_2(Q)\right) d\varepsilon \\ =\ & o(n^{-1}), \end{aligned}$$

where the last equality follows from Assumption B.1. Then since $\mathbf{P}\Pi\mu_{\xi,\gamma} = 0$ and $r_n = n^{1/2}$, we have that

$$r_n^2\mathbb{U}_n\mu_{\xi,\gamma} = r_n^2\mathbf{Q}\mu_{\xi,\gamma} + \frac{r_n^2}{\sqrt{n}}\mathbb{G}_n\Pi\mu_{\xi,\gamma} + o_p(1),$$

uniformly over $\mathcal{M}_\delta$. All regularity conditions of Theorem 5.1 are assumed directly except for Assumption 5.4. By Lemma A2 of Ghosal et al. (2000), the entropy condition of Assumption B.1 is sufficient for Assumption 5.4. Therefore, Theorem 5.1 proves this theorem. ∎

Since the asymptotic distribution is identical to the M-estimation case with the projected function $\Pi\mu_{\xi,\gamma}$, a corollary similar to Corollary 5.5 can be established. The discussion following the corollary is also valid. Therefore, if $\mathbf{Q}\chi_{\theta,\gamma}$ is twice differentiable at θ_0 with relevant rank conditions satisfied and if G_1 is linear in h, then the asymptotic representation in (5.10) would be obtained.

B.1. **Maximum Rank Correlation Estimation.** The objective function of the maximum rank correlation (MRC) estimator is a second-order U-process with kernel

$$\begin{aligned}&\chi\left(y_1,\mathbf{w}_1,y_2,\mathbf{w}_2;\theta,\gamma\right)\\&:=1\left\{y_1>y_2\right\}1\left\{g\left(\mathbf{w}_1,\theta,\gamma\right)>g\left(\mathbf{w}_2,\theta,\gamma\right)\right\}+1\left\{y_1<y_2\right\}1\left\{g\left(\mathbf{w}_1,\theta,\gamma\right)<g\left(\mathbf{w}_2,\theta,\gamma\right)\right\}.\end{aligned}$$

Recall that the MRC estimator can be applied to a general regression model defined as

$$Y=H\circ F\left(g\left(\mathbf{W},\theta_0,\gamma_0\right),U\right),$$

where H is a non-degenerate monotone function and F is a strictly monotone function for both arguments. To provide its asymptotic null distribution, we need to check the separability condition, Assumptions 5.3, 5.5, and B.1. We assume the following conditions which are slight modification of the standard regularity conditions of the MRC estimator in Sherman (1993) to reflect the threshold effects.

Assumption B.2. *(i) The first element of θ, say θ_1, is normalized to be 1.*
(ii) The first component of $\mathbf{W}$ has an everywhere positive Lebesgue density conditional on the remaining components $\widetilde{\mathbf{W}}=\widetilde{\mathbf{w}}$ for all $\widetilde{\mathbf{w}}$.
(iii) $\mathbf{W}$ is independent of U, and the support of $\mathbf{W}_\gamma$ is not contained in any proper linear subspace of $\mathbf{R}^{\dim(\mathbf{X})+\dim(\mathbf{Z})}$ for any γ.
(iv) T is continuously distributed.
(v) Let $\mathcal{N}$ be a neighborhood of θ_0. Then, the following conditions hold: (a) all mixed second partial derivatives of $\Pi\chi_{\theta,\gamma}$ with respect to θ exist on $\mathcal{N}$ for all $\gamma\in\Gamma$; (b) There exists an integrable function $M\left(y,\mathbf{w}\right)$ such that

$$\sup_\gamma\left|\nabla_{\theta\theta'}\Pi\chi_{\theta,\gamma}\left(Y,\mathbf{W}\right)-\nabla_{\theta\theta'}\Pi\chi_{\theta_0,\gamma}\left(Y,\mathbf{W}\right)\right|\leq M\left(Y,\mathbf{W}\right)\left|\theta-\theta_0\right|;$$

(c) $\sup_\gamma\mathbf{P}\left|\nabla_\theta\Pi\chi_{\theta_0,\gamma}\right|^2<\infty$; (d) $\sup_\gamma\mathbf{P}\sum_{i,j}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\Pi\chi_{\theta_0,\gamma}\right|<\infty$; and (e) $\mathbf{P}\nabla_{\theta\theta'}\Pi\chi_{\theta_0,\gamma}$ is negative definite for all $\gamma\in\Gamma$.

Since the functions H and F are monotone, the support condition of $\mathbf{W}$ implies that

$$\Pr\left(Y_1 > Y_2 | \mathbf{W}_1, \mathbf{W}_2\right) > \Pr\left(Y_1 < Y_2 | \mathbf{W}_1, \mathbf{W}_2\right) \quad \Leftrightarrow g\left(\mathbf{W}_1, \theta_0, \gamma\right) > g\left(\mathbf{W}_2, \theta_0, \gamma\right)$$

for any γ. This condition combined with the identification arguments in Han (1987) completes the separability condition. We next turn our attention to Assumption B.1. Consider the class of functions

$$\mathcal{M}_{K/n^{1/2}} = \left\{\chi_{\theta,\gamma} - \chi_{\theta_0,\gamma} : |\theta - \theta_0| < K/n^{1/2}, \gamma \in \Gamma\right\}.$$

From the arguments in Section 5 of Sherman (1993) and in the example of maximum score estimation, we can show that $\mathcal{M}_{K/n^{1/2}}$ is a VC-class of functions and that $\mathcal{M}_{K/n^{1/2}}$ has an envelope function $M_{K/n^{1/2}} = 1\left\{|\mathbf{x}'\beta_0| \le C\left(K/n^{1/2}\right)|\mathbf{x}|\right\}$ for a constant C. Therefore, it satisfies the conditions in Assumption B.1. To verify Assumption 5.3, consider the following class of projected functions

$$\bar{\mathcal{F}}_{K/n^{1/2}} = \left\{\Pi\chi_{\theta,\gamma} - \Pi\chi_{\theta_0,\gamma} : |\theta - \theta_0| < K/n^{1/2}, \gamma \in \Gamma\right\}.$$

Then, Assumptions 5.3 follows from the finite envelope and the smoothness condition of $\Pi\chi_{\theta,\gamma}$. It remains to show Assumption 5.5. However, it follows from the differentiability of $\Pi\chi_{\theta,\gamma}$ that the asymptotic representation in (5.10) applies. In particular, the covariance kernel of $\mathcal{G}(\gamma)$ is given by $\mathbf{P}\left[\left(\nabla_\theta \Pi\chi_{\theta_0,\gamma_1}\right)\left(\nabla_\theta \Pi\chi_{\theta_0,\gamma_2}\right)'\right]$ and $V(\gamma) = \mathbf{P}\nabla_{\theta\theta'}\Pi\chi_{\theta_0,\gamma}$.

Appendix C. Consistency and Local Power

In this section, we present asymptotic properties of our test statistic when the null hypothesis is false. We first consider a fixed alternative $g(\mathbf{w})$ such that

$$g(\mathbf{w}) \neq \mathbf{x}'\beta,$$

for any β. Let $\mathbf{P}_1$ denote the common probability measure under this alternative.

Theorem C.1. *Let $\mathcal{F}$ be a class of functions $q_{\theta,\gamma}$ with envelope F such that $\mathbf{P}_1 F < \infty$. Assume either of the following two conditions: (i) $N_{[]}\left(\varepsilon, \mathcal{F}, L_1\left(\mathbf{P}_1\right)\right) < \infty$ for every $\varepsilon > 0$; (ii) For $\mathcal{F}_M$ defined as the class of functions $f1\left\{F \leq M\right\}$ for $f \in \mathcal{F}$, $\log N\left(\varepsilon, \mathcal{F}_M, L_1\left(\mathbb{P}_n\right)\right) = o_p\left(n\right)$ for every ε and $M > 0$. Let $Q_1\left(\theta,\gamma\right) = \mathbf{P}_1 q_{\theta,\gamma}$ and assume that there exists $\left(\theta,\gamma\right)$ such that $Q_1\left(\theta,\gamma\right) > \sup_{\theta:\alpha=0} Q\left(\theta,\gamma\right)$. Then, the test QLR_n is consistent against the alternative g, that is, the rejection probability of our test goes to one under $\mathbf{P}_1$.*

This theorem states conditions under which our test is consistent. This theorem might not be very constructive to convey some meaningful insight into what alternatives our test can detect as it is difficult to determine the functional form of Q_1 without a specific q and $\mathbf{P}_1$. Roughly speaking, however, it implies that the test can detect an alternative which is better approximated by a piecewise linear structural form than linear one. Clearly, if

$$g\left(\mathbf{w}\right) = \mathbf{x}'\beta_0 + \mathbf{z}'\alpha_0 1\left\{t > \gamma_0\right\}$$

for some $\alpha_0 \neq 0$, the test is consistent under the other conditions of the above theorem. Furthermore, Theorem C.1 suggests that the test be powerful against some other nonlinear alternatives, as we demonstrate via Monte Carlos experiments in Section D.

Next, we investigate the asymptotic power property of our tests under sequences of local alternatives:

$$g_n\left(\mathbf{w}\right) = \mathbf{x}'\beta_0 + \rho_n^{-1} \cdot \mathbf{x}'\alpha\left(t\right),$$

where α is a vector-valued integrable function defined on the support of T and $\rho_n \to \infty$. This alternative is a natural generalization of the threshold model to encompass smooth transition models and varying coefficient models. Let $\mathbf{P}_n$ denote the probability measure for each n under the local alternatives. As above, general statement is less informative than examination of specific examples since local power depends largely on whether or not the limit of $r_n \mathbf{P}_n m_{h_n,\gamma}$ is different from that of $r_n \mathbf{P} m_{h_n,\gamma}$.

When $\rho_n = r_n$, the local asymptotic distribution of QLR_n under $\mathbf{P}_n$ can be obtained under minor modification of previous assumptions. We discuss this. We keep Assumption 1 and 2. Assumptions 3-5 need to be restated in terms of $\mathbf{P}_n$. The uniform entropy condition (5.3) in Assumption 4 remains the same since it does not depend on the true measure (see e.g. section 2.11.1 of van der Vaart and Wellner (1996)). Lemmas 1 and 2 are valid under these modifications on Assumption 3 and 4 and thus Assumptions 1 and 2 can be verified in the same way as under the null. The limit quantities in Assumption 5 would not the same as those under $\mathbf{P}$. Either of the covariance kernel of G_1 or the functional form of G_2 or both change under $\mathbf{P}_n$, yielding the local power. The methods to verify Assumptions 1-5 are similar as under $\mathbf{P}$. When $\mathbf{P}m_{\xi,\gamma}$ is twice differentiable, its first derivative is zero at $\xi = \xi_0$, where $\xi_0 := (\beta_0', 0', \beta_0')'$, and G_2 is quadratic in its second derivative. On the other hand, the first derivative of $\mathbf{P}_n m_{\xi,\gamma}$ is not zero at $\xi = \xi_0$ but $r_n \left(\partial/\partial\xi\right) \mathbf{P}_n m_{\xi_0,\gamma}$ has a non-vanishing limit. This is usually called the "noncentrality parameter", which is the source of the local power and yields the consistency when $\rho_n = o\left(r_n\right)$. All of our examples have nontrivial noncentrality parameters.

We now present two of previous examples to illustrate power properties of our test, focusing on the noncentrality parameter. First, consider the maximum score estimation of the binary response model. We begin with $r_n^2 \mathbf{P}_n m_{h_n,\gamma}$. As shown in section 6.1, $r_n^2 \mathbf{P} m_{h_n,\gamma}$ converges to a quadratic function in h without a linear term as the first term in the expansion vanishes under $\mathbf{P}$. We show that the linear term does not vanish under $\mathbf{P}_n$. In particular, note that $\kappa\left(\mathbf{w}\right)$ in (6.2) need to be replaced by

$$\begin{aligned}\kappa_n\left(\mathbf{w}\right) &= 1 - 2F_{U|\mathbf{W}}\left(-\mathbf{x}'\beta_0 - r_n^{-1}\mathbf{x}'\alpha\left(t\right)|\mathbf{w}\right) \\ &= \kappa\left(\mathbf{w}\right) + 2f_{U|\mathbf{W}}\left(-\mathbf{x}'\beta_0 - \rho_n^{-1}\cdot \mathbf{x}'\tilde{\alpha}\left(t\right)|\mathbf{w}\right)\rho_n^{-1}\mathbf{x}'\alpha\left(t\right),\end{aligned}$$

where $\tilde{\alpha}$ is the mean value. Then, following the steps to derive the limit of $r_n^2 \mathbf{P} m_{h_n,\gamma}$ in the proof of Theorem 6.1 with $\kappa\left(\mathbf{w}\right)$ replaced by $\kappa_n\left(\mathbf{w}\right)$, we can see that the difference $r_n^2 \mathbf{P}_n m_{h_n,\gamma} - r_n^2 \mathbf{P} m_{h_n,\gamma}$, that is, the noncentrality parameter is given by

$$2\int\int \left[\alpha\left(t\right)'\left(xx'\left(h_\beta - h_b\right) + xx'1\left\{t > \gamma\right\}h_\alpha\right)\right]\frac{r_n}{\rho_n}1\left\{\mathbf{x}'\beta_0 = 0\right\}f_{U|\mathbf{W}}\left(0|\mathbf{w}\right)p_{\mathbf{W}}\left(\mathbf{w}\right)d\varpi.$$

If $\alpha(T)$ is nonzero with positive probability, then the noncentrality parameter is nonzero for some γ, regardless of h unless $h_\alpha = 0$. On the other hand, we can easily see from the proof of Theorem 6.1 that the covariance kernel of G_1 does not change. Therefore, our test has local power with $\rho_n = r_n$ and is consistent when $\rho_n = o(r_n)$.

Next consider the MLE of the probit model. Let $\rho_n = r_n = \sqrt{n}$ and examine the score functions for $q(y, \mathbf{w}; \theta, \gamma)$ and $\tilde{q}(y, \mathbf{w}; b)$ under $\mathbf{P}_n$. Their expected values are zero under $\mathbf{P}$ but non-zeros and different from each other under $\mathbf{P}_n$, which yields the noncentrality parameter. In particular, a direct calculation of the expected value with an expansion of the term $\Phi\left(\mathbf{x}'\beta_0 + \mathbf{x}'\alpha(t)/\sqrt{n}\right)$ at $\mathbf{x}'\beta_0$ yields

$$\begin{aligned}
\sqrt{n}\mathbf{P}_n \frac{\partial}{\partial\theta} q(y, \mathbf{w}; \theta_0, \gamma) &= E\left[\frac{\phi(\mathbf{X}'\beta_0)}{\Phi(\mathbf{X}'\beta_0)} \begin{pmatrix} \mathbf{X} \\ \mathbf{X}1\{T > \gamma\} \end{pmatrix} \phi\left(\mathbf{X}'\beta_0 + \mathbf{X}'\frac{\tilde{\alpha}(T)}{\sqrt{n}}\right) \mathbf{X}'\alpha(T)\right] \\
&\quad + E\left[\frac{\phi(\mathbf{X}'\beta_0)}{\Phi(-\mathbf{X}'\beta_0)} \begin{pmatrix} \mathbf{X} \\ \mathbf{X}1\{T > \gamma\} \end{pmatrix} \phi\left(-\mathbf{X}'\beta_0 - \mathbf{X}'\frac{\tilde{\alpha}(T)}{\sqrt{n}}\right) \mathbf{X}'\alpha(T)\right] \\
&\to E\left[\frac{\phi^2(\mathbf{X}'\beta_0)}{\Phi(\mathbf{X}'\beta_0)\Phi(-\mathbf{X}'\beta_0)} \begin{pmatrix} XX'\alpha(T) \\ XX'1\{T > \gamma\}\alpha(T) \end{pmatrix}\right],
\end{aligned}$$

where $\tilde{\alpha}$ lies between α and 0. Similarly,

$$\sqrt{n}\mathbf{P}_n \frac{\partial}{\partial b} \tilde{q}(y, \mathbf{w}; \beta_0) \to E\left[\frac{\phi^2(\mathbf{X}'\beta_0)}{\Phi(\mathbf{X}'\beta_0)\Phi(-\mathbf{X}'\beta_0)} XX'\alpha(T)\right].$$

Then, the noncentrality parameter becomes

$$E\left[\frac{\phi^2(\mathbf{X}'\beta_0)}{\Phi(\mathbf{X}'\beta_0)\Phi(-\mathbf{X}'\beta_0)} XX'\alpha(T)1\{T > \gamma\}\right] \neq 0$$

for some γ as long as $\alpha(T)$ is not zero with positive probability. Therefore, the test has non-trivial local power against local alternatives of the above form as well as of the threshold type. Furthermore, if $\rho_n = o(\sqrt{n})$, then the noncentrality parameter diverges to infinity to yield the consistency of our test.

Appendix D. Monte Carlo Simulations

In this section we investigate finite sample properties of the proposed test by Monte Carlo experiments. we report Monte Carlo simulation results for all four examples considered in the article.

D.1. **Binary Response Models: Probit, Logit, and Maximum Score.**

First, we report Monte Carlo simulation results when the samples are generated from a simple probit or logit model. To see whether the test has power against an alternative that is different from a threshold model, we consider the smooth transition model as well as the threshold model as alternatives. Therefore, we have 4 different models in total, and the baseline model has the following form:

$$\begin{aligned} Y^* &= \beta_0 + \beta_1 \mathbf{X} + \alpha \mathbf{Z} \psi(T, \gamma) + U \\ Y &= 1\{Y^* > 0\}, \end{aligned}$$

where $\psi(T, \gamma) = 1\{T > \gamma\}$ for the threshold model and $\psi(T, \gamma) = 1/(1 + \exp(-(T - \gamma)))$ for the smooth transition model. The true parameter values are set as $\beta_0 = 0.5$, $\beta_1 = 1$, $\gamma = 0.5$ for the threshold model, and $\gamma = 0$ for the smooth transition model. When the null hypothesis is true, the parameter α is equal to zero. Under the alternatives, α has various non-zero values from 0.2 to 1. The covariates $\mathbf{X}$ and $\mathbf{Z}$ are generated independently from $N(0, 1)$ and $N(0, 2)$, respectively. The covariate T follows the uniform distribution on the interval $[0, 1]$ for the threshold model and $N(0, 1)$ for the smooth transition model. The error term U is generated from either $N(0, 1)$ or the logistic distribution.

Parameters other than γ are estimated by the Newton-Raphson's method, and the threshold parameter γ is estimated by the grid search. For the grid, we used the data points of T after trimming at lower and upper 10th percentiles. We considered three different sample sizes, $n = 50, 100$, and 200, and replicated each simulation design 1000 times. For the simulation number of the score functions, we set $J = 2000$.

Figure 2. Power Functions of Threshold Models

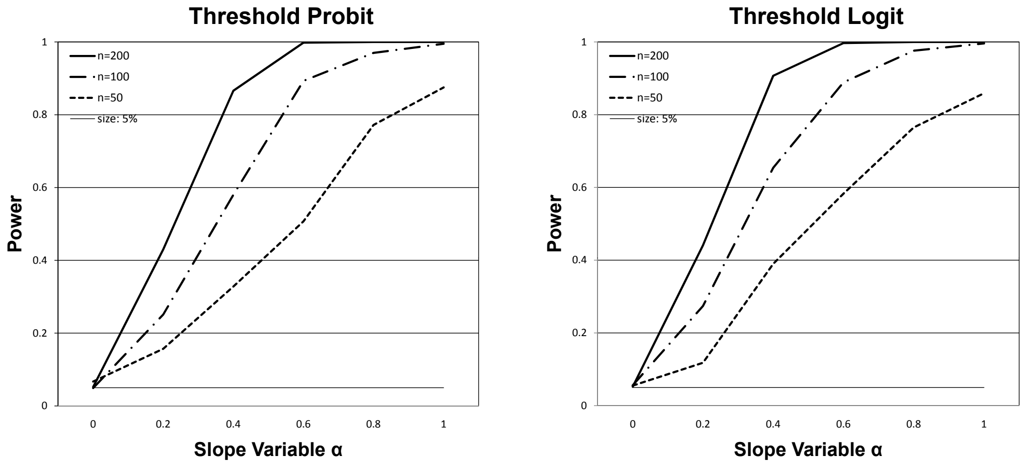
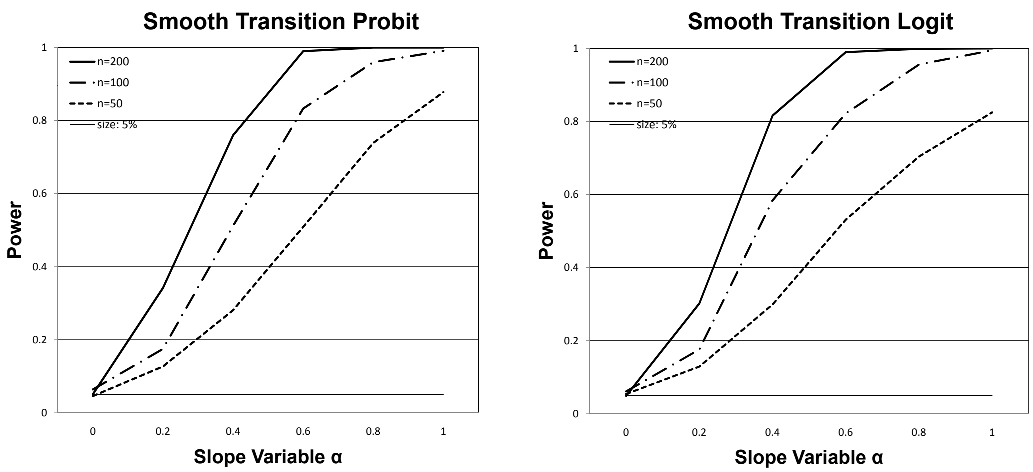


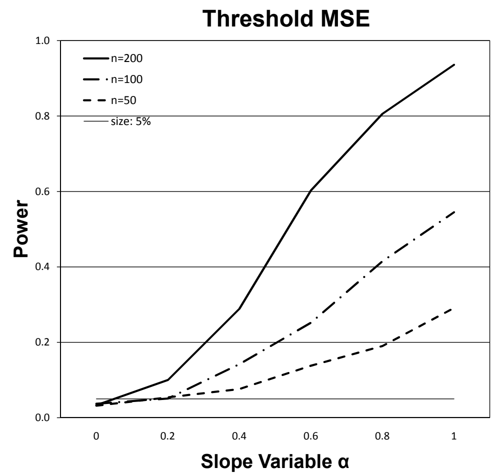
Figure 3. Power Functions of Smooth Transition Models



Figures 2–3 summarize the result of the simulation study. Overall, the test performs well as expected from the theory. First, under the null hypothesis ($\alpha = 0$), the rejection rates of the test are close to the nominal level in most cases. Second, Figures 2–3 show the power of the test when α increases from 0 to 1. The result indicates that, in all cases, the power increases fast as the parameter value of α is farther away from zero. The test shows good performance even with a relatively small sample size, say $n = 100$.

We now report simulation results for testing the null hypothesis that $\alpha_0 = 0$ for the probit threshold model above with the maximum score objective function. This amounts

FIGURE 4. Power Functions of the Probit Threshold Model with Maximum Score Estimation



to the case when a researcher only relies on the assumption that U has conditional median zero without knowing that U follows the standard normal distribution. The critical values are obtained via subsampling. The subsample sizes (m) and original sample sizes (n) were $(m, n) = (20, 50), (30, 100), (35, 200)$, respectively.

Figure 4 shows the power functions with the 5% level test. Not surprisingly, relative to the left panel of Figure 2, the power does not increase rapidly as α gets large or n increases. Note that this is consistent with the theoretical result that the test with the maximum score estimation has local power at a rate of $n^{-1/3}$.

D.2. **Quantile Regression.** In this section we investigate finite sample properties of the proposed test for the quantile regression model. In particular, we consider the median regression model $(\tau = 0.5)$:

$$Y = \beta_0 + \beta_1 \mathbf{X} + \alpha \mathbf{Z} \psi (T, \gamma) + U,$$

where $\psi (T, \gamma) = 1 \{T > \gamma\}$. The true parameter values are set as $\beta_0 = 0.5$, $\beta_1 = 1$, $\gamma = 0.5$. When the null hypothesis is true, the parameter α is equal to zero. Under the alternatives, α has various non-zero values from 0.2 to 1. The covariates $\mathbf{X}$ and $\mathbf{Z}$ are generated independently from $N (0, 1)$ and $N (0, 2)$, respectively. The covariate T follows the uniform distribution on the interval $[0, 1]$ for the threshold model. The error term U is generated from the standard normal distribution.

Parameters other than γ are estimated by the linear programming method for the standard linear quantile regression model, and the threshold parameter γ is estimated by the grid search. For the grid, we use the data points of T after trimming at lower and upper 10th percentiles. We consider three different sample sizes, $n = 50, 100$, and 200, and replicate each simulation design 1000 times. For the simulation number of the score functions, we set $J = 2000$.

In addition, we estimate $\widehat{V}(\gamma)$ and $\widetilde{V}$ by (3.9) since regression errors are independent of regressors. Finally, we use the standard normal density as the the kernel function K and Silverman's rule of thumb for $h = 1.06 \times \tilde{\sigma} n^{-1/5}$, where $\tilde{\sigma}$ is the sample standard deviation of $\tilde{U} := Y_i - \mathbf{X}_i' \tilde{\beta}$.

Figure 5. Power Functions of Threshold Quantile Regression Models

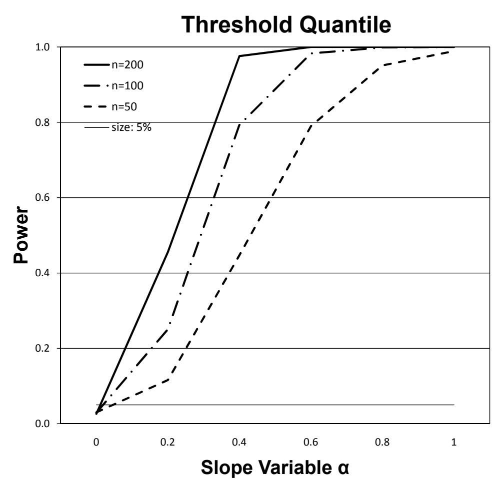


Figure 5 shows the power functions for the 5 % level test. Under the null of $\alpha = 0$, the rejection rates of the test are about 2% lower than the nominal level. The figure shows the power of the test increases fast as α or n gets large.

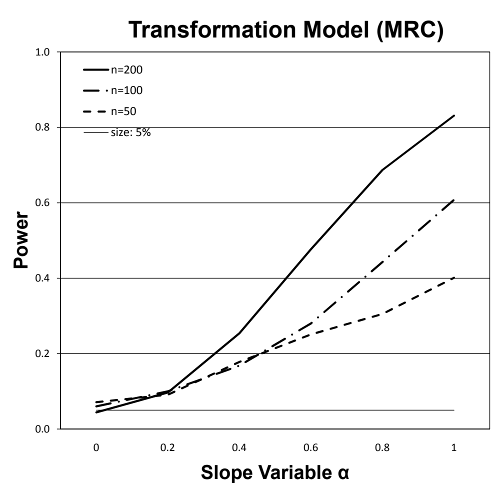
D.3. **Maximum Rank Correlation Estimation.** For the simulation study of the MRC estimator, we use the following data generating procedure:

$$T(Y) = \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \alpha \mathbf{Z} \cdot 1(T > \gamma) + U$$

where covariates, $\mathbf{X}_1, \mathbf{X}_2, \mathbf{Z}$ and T, are generated independently from $N(0,4)$, $N(0,1)$, $N(0,1)$, and the uniform distribution on the interval $[0,1]$, respectively. The error term U is generated from $N(0,1)$. We set the transformation function $T(y) \log y$. The parameter β_1 is normalized as 1, and other parameters are set as $\beta = 1$ and $\gamma = 0.5$. The parameter α is equal to zero under the null hypothesis, and varies from 0.2 to 1 under the alternatives. Note that the constant term is not identified in the unknown transformation model, so we drop it from the model.

We estimate all parameters using the grid search. The girds used for each parameter are as follows: the 51 points equally spaced on the interval $[-1,3]$ are used for estimating β_2, the 51 points on $[-1,2]$ for α, and the 36 points on $[0.1,0.9]$ for γ. We consider three sample sizes, $n = 50, 100$, and 200, and replicate each design 1,000 times. We calculate the simulated p-value with $J = 1,000$.

Figure 6. Power Functions of Threshold Models with Maximum Rank Correlation Estimation



Simulated critical values can be obtained using numerical derivatives as in Section 7 of Sherman (1993). Specifically, we use the smooth objective function in the simulation step by substituting the standard normal cdf for the indicator function with an appropriate bandwidth. Figure 6 shows the power functions for the 5 % level test. Overall, test seems to perform well as in previous examples.

We now explain how to obtain critical values in detail below:

(1) Given the data, estimate the parameter under the null and the alternative, $\widetilde{\beta}$ and $\left(\hat{\theta}, \hat{\gamma}\right)$, respectively. Construct the test statistic QLR_n using the estimates.
(2) Recall some notation here:

$$\begin{aligned} Q_n\left(\theta,\gamma\right) &= \mathbb{U}_n \chi_{\theta,\gamma}(Y_i, Y_j, \mathbf{W}_{\gamma,i}, \mathbf{W}_{\gamma,j}) \\ &= \frac{1}{n\left(n-1\right)} \sum_{i \neq j} 1\left(Y_i > Y_j\right) 1\left(\mathbf{W}_{\gamma,i}' \theta > \mathbf{W}_{\gamma,j}' \theta\right) \end{aligned}$$

and

$$\begin{aligned} \mu_{\xi,\gamma}\left(Y_i, Y_j, \mathbf{W}_i, \mathbf{W}_j\right) &= \chi(Y_i, Y_j, \mathbf{W}_{\gamma,i}, \mathbf{W}_{\gamma,j}) - \widetilde{\chi}_b(Y_i, Y_j, \mathbf{X}_i, \mathbf{X}_j) \\ &= 1\left(Y_i > Y_j\right)\left[1\left(\mathbf{W}_{\gamma,i}' \theta > \mathbf{W}_{\gamma,j}' \theta\right) - 1\left(\mathbf{X}_i' b > \mathbf{X}_j' b\right)\right]. \end{aligned}$$

Replace indicator functions in the objective function with the standard normal cdf. Now, $\mu_{\xi,\gamma}$ is twice differentiable with respect to ξ (slightly abuse notation and use the same μ, χ etc.) The first order derivative of $\mu_{\xi,\gamma}$ is

$$\frac{\partial}{\partial \xi} \mu_{\xi,\gamma} = \begin{bmatrix} \frac{\partial}{\partial \theta} \chi_{\theta,\gamma} \\ -\frac{\partial}{\partial b} \widetilde{\chi}_b \end{bmatrix},$$

where

$$\begin{aligned} \frac{\partial}{\partial \theta} \chi_{\theta,\gamma} &= \Phi\left(\frac{Y_i - Y_j}{a}\right) \phi\left(\frac{\left(\mathbf{W}_{\gamma,i} - \mathbf{W}_{\gamma,j}\right)' \theta}{a}\right) \frac{\mathbf{W}_{\gamma,i} - \mathbf{W}_{\gamma,j}}{a} \\ \frac{\partial}{\partial b} \widetilde{\chi}_b &= \Phi\left(\frac{Y_i - Y_j}{a}\right) \phi\left(\frac{\left(\mathbf{X}_i - \mathbf{X}_j\right)' b}{a}\right) \frac{\mathbf{X}_i - \mathbf{X}_j}{a}. \end{aligned}$$

The bandwidth a is set as $a = 2\hat{\sigma}n^{(-3/5)}$ where $\hat{\sigma}$ is the sample standard deviation of the argument in the function, i.e. $Y_i - Y_j$, $(\mathbf{W}_{\gamma,i} - \mathbf{W}_{\gamma,j})'\theta$ etc. The second derivative is

$$\frac{\partial}{\partial\xi\partial\xi'}\mu_{\xi,\gamma} = \begin{bmatrix} \frac{\partial^2}{\partial\theta\partial\theta'}\chi_{\theta,\gamma} & 0 \\ 0 & -\frac{\partial^2}{\partial b\partial b'}\widetilde{\chi}_b \end{bmatrix},$$

where diagonal elements can be computed easily.

(3) To generate the simulated empirical U-process, say $\widehat{\mathbb{U}}_n$, for each (i, j), we multiply $V_{ij} = V_i + V_j$ to $\mu_{\xi,\gamma}(Y_i, Y_j, \mathbf{W}_i, \mathbf{W}_j)$, where V_i and V_j are generated from $Gamma(0.25, 0.5)$, independently.

(4) Then the simulated test statistic is

$$\sup_{\gamma} \frac{1}{2}\left[r_n\left(\widehat{U}_n - \mathbb{U}_n\right)\frac{\partial}{\partial\xi}\mu_{\tilde{\xi},\gamma}\right]'\left[-\mathbb{U}_n\frac{\partial^2}{\partial\xi\partial\xi'}\mu_{\tilde{\xi},\gamma}\right]^{-1}\left[r_n\left(\widehat{U}_n - \mathbb{U}_n\right)\frac{\partial}{\partial\xi}\mu_{\tilde{\xi},\gamma}\right].$$

(5) Simulate the same statistic J times for a large J, and calculate the simulated p-value as in the main text (that is, the proportion of simulated test statistics that are greater than the original test statistic).

References

Abrevaya, J. (1999). Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics 93*(2), 203–228.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–856.

Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica 69*(3), 683–734.

Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica 62*(6), 1383–1414.

Andrews, D. W. K. and W. Ploberger (1995). Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. *The Annals of Statistics 23*(5), 1609–1629.

Card, D., A. Mas, and J. Rothstein (2008). Tipping and the dynamics of segregation. *Quarterly Journal of Economics 123*(1), 177–218.

Cavanagh, C. and R. Sherman (1998). Rank estimation for monotonic index models. *Journal of Econometrics 84*(2), 351–381.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics 21*, 520–533.

Cho, J. S. and H. White (2007). Testing for regime switching. *Econometrica 75*(6), 1671–1720.

Cox, C. (1987). Threshold dose-response models in toxicology. *Biometrics 43*(3), 511–523.

Cox, D. (1972). Regression models and life tables. *Journal of the Royal Statistical Society Series B 34*, 187–220.

Cox, D. (1975). Partial likelihood. *Biometrika 62*, 269–276.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 64*, 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 74*(1), 33–43.

Delgado, M. A. and J. Hidalgo (2000). Nonparametric inference on structural breaks. *Journal of Econometrics 96*(1), 113–144.

Delgado, M. A., J. M. Rodriguez-Poo, and M. Wolf (2001). Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator. *Economics Letters 73*(2), 241–250.

Durlauf, S. N. and P. A. Johnson (1995). Multiple regimes and cross-country growth behavior. *Journal of Applied Econometrics 10*(4), 365–384.

Fan, J., H.-N. Hung, and W.-H. Wong (2000). Geometric understanding of likelihood ratio statistics. *Journal of the American Statistical Association 95*(451), pp. 836–841.

Fan, J., C. Zhang, and J. Zhang (2001). Generalized likelihood ratio statistics and wilks phenomenon. *The Annals of Statistics 29*(1), pp. 153–193.

Forbes, K. J. and R. Rigobon (2002). No contagion, only interdependence: Measuring stock market co-movements. *Journal of Finance 57*(5), 2223–2261.

Ghosal, S., A. Sen, and A. W. van der Vaart (2000). Testing monotonicity of regression. *Annals of Statistics 28*(4), 1054–1082.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model : The maximum rank correlation estimator. *Journal of Econometrics 35*(2-3), 303–316.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica 64*(2), 413–430.

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics 93*(2), 345–368.

Khan, M. S. and A. S. Senhadji (2001). Threshold effects in the relationship between inflation and growth. *IMF Staff Papers 48*(1), 1–21.

Kim, J. and D. Pollard (1990). Cube root asymptotics. *Annals of Statistics 18*(1), 191–219.

Koenker, R. (2005). *Quantile Regression*, Volume 38 of *Econometric Society monographs.* Cambridge University Press.

Kosorok, M. R. and R. Song (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *Annals of Statistics 35*(3), 957–989.

Lee, S. and M. H. Seo (2008). Semiparametric estimation of a binary response model with a change-point due to a covariate threshold. *Journal of Econometrics 144*(2), 492–499.

Liu, X. and Y. Shao (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *Annals of Statistics 31*(3), 807–832.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*(3), 205–228.

Manski, C. F. (1985). Semiparametric analysis of discrete response. Asymptotic properties of the maximum score estimator. *Journal of Econometrics 27*(3), 313–333.

Manski, C. F. (1988). Identification of binary response models. *Journal of the American Statistical Association 83*(403), 729–738.

Pastor, R. and E. Guallar (1998). Use of tow-segmented logistic regression to estimate change-points in epidemiologic studies. *American Journal of Epidemiology 148*(7), 631–642.

Pastor-Barriuso, R., E. Guallar, and J. Coresh (2003). Transition models for change-point estimation in logistic regression. *Statistics in Medicine 22*, 1141–1162.

Pesaran, M. H. and A. Pick (2007). Econometric issues in the analysis of contagion. *Journal of Economic Dynamics and Control 31*(4), 1245–1277.

Pons, O. (2003). Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *Annals of Statistics 31*(2), 442–463.

Roy, S. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Stat. 24*, 220–238.

Schwartz, P. F., C. Gennings, and V. M. Chinchilli (1995). Threshold models for combination data from reproductive and development experiments. *Journal of the American Statistical Association 90*(431), 862–870.

Sherman, R. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica 61*(1), 123–137.

Song, R., M. Kosorok, and J. Fine (2009). On asymptotically optimal tests under loss of identifiability in semiparametric models. *Annals of Statistics 37*, 2409–2444.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* New York: Oxford University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Process.* Springer, New York.

Zhang, J. and P. Cheng (1989). The asymptotic powers of some pp tests. *Jour. Sys. Sci. and Math. Sci. 9*, pp. 370–382.

Zhang, J. and G. Li (1993). A new approach to asymptotic distributions of maximum likelihood ratio statistics. In K. Matusita, M. L. Puri, and T. Hayakawa (Eds.), *Statistical Science and Data Analysis: Proceedings of the third Pacific Area Statistical Conference*, the Netherlands, pp. 325–336. International Science Publishers.

Zhu, H. and H. Zhang (2006). Asymptotics for estimation and testing procedures under loss of identifiability. *Journal of Multivariate Analysis 97*(1), 19–45.

Department of Economics, Seoul National University, 599 Gwanak-ro, Gwanak-gu, Seoul, 151-742, Republic of Korea, and Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 7 Ridgmount Street, London, WC1E 7AE, UK.

E-mail address: sokbae@gmail.com

URL: http://www.ifs.org.uk/people/profile?id=46.

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

E-mail address: m.seo@lse.ac.uk

URL: http://personal.lse.ac.uk/SEO.

Department of Economics, University of Western Ontario, 1151 Richmond Street N, London, ON N6A 5C2, Canada.

E-mail address: yshin29@uwo.ca

URL: http://publish.uwo.ca/~yshin29.