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. We provide 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 also 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: 8 February 2010.

We would like to thank Jesse Rothstein for providing the data used in the paper. 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). Lee also thanks the Cowles Foundation for Research in Economics at Yale University, where he is currently a visiting fellow.
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:

(1.1) \[ Y^* = g(W, \theta_0, \gamma_0) + U, \]

(1.2) \[ g(w, \theta, \gamma) = x' \beta + z' \alpha 1_{\{t > \gamma\}}, \]

where $W$ is a vector of regressors that consist of distinct elements of $(X, Z, T)$, $U$ is an unobserved random variable, and $\theta_0 := (\beta_0', \alpha_0')'$ and $\gamma_0$ are unknown true parameter values and belong to $\Theta := B \times A$ and $\Gamma$, respectively, which are subsets of finite-dimensional Euclidean spaces. Without loss of generality, assume that the vector $Z$ is a subset of $X$ such that $Z = R'X$ for some known matrix $R$ and that $T$ might be an element of $X$. The random variable $T$ is the threshold variable and $\gamma_0$ 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 5, 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

$$H_0 : \alpha_0 = 0 \text{ for any } \gamma_0 \in \Gamma \text{ vs. } H_1 : \alpha_0 \neq 0 \text{ for some } \gamma_0 \in \Gamma.$$ 

In general, unknown parameters in (1.2) are identifiable under the alternative hypothesis; however, the threshold parameter \(\gamma_0\) 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, 1994; 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 statistic.

The main objective of this paper is to provide a unified testing framework in regression models using the sup-LR-type statistic. We make two main contributions to the literature. First, 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). We provide an alternative, novel method that can be used to establish the asymptotic null distribution, even when the usual quadratic approximation is intractable. We illustrate our method by applying it to the objective function for the maximum score estimator (Manski, 1975, 1985), for which no existing method can be applied.

Second, 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 paper, 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 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.

The remainder of the paper 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 reports some simulation results and Section 5 illustrates the usefulness of our test by applying it to real data used by Card et al. (2008). Formal results are given in Section 6. In Section 7, we provide some concluding remarks. All the proofs and some additional theoretical results 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 $\alpha_0$ 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 $\{(Y_i, W_i) : i = 1, \ldots, n\}$. For a given $\gamma \in \Gamma$, let $\hat{\theta}(\gamma)$ denote an estimator of $\theta_0$ that maximizes the objective function $Q_n(\theta, \gamma)$. Define $Q_n(\gamma) := Q_n(\hat{\theta}(\gamma), \gamma)$ to be a profiled objective function and let

$$\hat{\gamma} = \arg\max_{\gamma} Q_n(\gamma), \quad \hat{\theta} = \hat{\theta}(\hat{\gamma}), \quad \hat{Q}_n = Q_n(\hat{\gamma}).$$
In addition, let
\[ \tilde{\beta} = \arg\max_{\beta, \alpha = 0} Q_n(\theta, \gamma) \quad \text{and} \quad \tilde{Q}_n = \max_{\beta, \alpha = 0} Q_n(\theta, \gamma). \]

Recall that \( Q_n \) does not depend on \( \gamma \) 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 \( \gamma \). Let
\[
QLR_n(\gamma) = r_n^2 \left[ Q_n(\gamma) - \tilde{Q}_n \right],
\]
for each \( \gamma \), 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
\[
Q_n(\theta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} q(Y_i, W_i; \theta, \gamma), \tag{2.1}
\]
where \( q \) is i.i.d. random variables when \((\theta, \gamma)\) are fixed. For example, the maximum score estimator maximizes \( Q_n(\theta, \gamma) \) with

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

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

\[
Q_n(\theta, \gamma) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \chi(Y_i, W_i, Y_j, W_j; \theta, \gamma),
\]

where \( \chi \) is symmetric in the sense that \( \chi(y_i, w_i, y_j, w_j; \theta, \gamma) = \chi(y_j, w_j, y_i, w_i; \theta, \gamma) \).

For example, the maximum rank correlation estimator maximizes \( Q_n(\theta, \gamma) \) with

\[
\chi(y_1, w_1, y_2, w_2; \theta, \gamma) = 1\{y_1 > y_2\} \mathbb{1}\{g(w_1, \theta, \gamma) > g(w_2, \theta, \gamma)\} + 1\{y_1 < y_2\} \mathbb{1}\{g(w_1, \theta, \gamma) < g(w_2, \theta, \gamma)\}.
\]

In this example, \( r_n = n^{1/2} \). In both cases, we assume that \( q \) or \( \chi \) depends on \((\theta, \gamma)\) 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 6.
In what follows, we use the conventional notation in empirical process theory. Denote by $P$ the common probability measure, by $P_n$ the empirical measure of the random sample of size $n$ from $P$, and by $G_n$ the empirical process indexed by a class $\mathcal{F}$ of functions $q$ such that $G_n q = \sqrt{n} (P_n - P) q$.

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 $\xi = (\theta', b')'$, $q_{\theta,\gamma} = q(y, w; \theta, \gamma)$, and $\tilde{q}_b = q(\theta', 0)'$. Note that $\tilde{q}_b$ is the same for any $\gamma$ and thus it is a function of $b$ only. We have introduced the index $b$ to denote arguments for $\beta_0$ in the objective function with the restriction $\alpha = 0$ to distinguish this from the index $\beta$ that denotes arguments for $\beta_0$ in the unrestricted objective function.

Also, note that $q_{\theta,\gamma}$ is the same for all $\gamma$ if $\theta = \theta_0$, using the fact that $\alpha_0 = 0$ under $H_0$. Thus, under $H_0$, $q_{\theta_0,\gamma} = \tilde{q}_{\beta_0}$, and when $b$ is restricted to $\beta_0$,

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

Similarly, when $\theta$ is fixed at $\theta_0$,

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

It now follows that

$$QLR_n = \frac{r_n^2}{n} \left[ \sup_{\theta, \gamma} P_n q_{\theta,\gamma} - \sup_b P_n \tilde{q}_b \right]$$

$$= \frac{r_n^2}{n} \left[ \sup_{\xi, \gamma: b = \beta_0} P_n m_{\xi,\gamma} - \sup_{\xi, \gamma: \theta = \theta_0} (-P_n m_{\xi,\gamma}) \right],$$

(3.1)
which is a continuous transformation of $r_n^2 \mathbb{P}_n m_{\xi, \gamma}$. Note also that $m_{\xi_0, \gamma} = 0$ for any $\gamma$, 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

$$
(3.2) \quad r_n^2 \mathbb{P}_n m_{\xi, \gamma} = r_n^2 \sqrt{n} \mathbb{G}_n m_{\xi, \gamma} + r_n^2 \mathbb{P} m_{\xi, \gamma}.
$$

Since the supremum is obtained at $\theta = \hat{\theta}(\gamma)$ for each $\gamma$ 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.

We have described our key idea behind our method for establishing the asymptotic null distribution. This simple yet novel idea enables us to derive the null distribution of $QLR_n$ in a straightforward way even in a situation where the usual quadratic approximation to $QLR_n$ is intractable. To our best knowledge, the literature has not derived the weak convergence of QLR statistics in this way. Instead of resorting to the standard approaches based on stochastic quadratic expansions, we show that the statistic $QLR_n$ can be rewritten as a continuous functional of an empirical process. Therefore, our method for obtaining the asymptotic null distribution is new and can be used even if the usual quadratic approximation is unavailable or difficult to obtain.

We now use the probit model as an illustrative example. Define $W_{\gamma} := (X', Z'1\{T > \gamma\})'$. The function $q(y, w; \theta, \gamma)$ for the probit model has the form

$$
(3.3) \quad q(y, w; \theta, \gamma) = y \log \Phi(g(w, \theta, \gamma)) + (1 - y) \log \Phi(-g(w, \theta, \gamma)),
$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution and $g(W, \theta, \gamma) = W' \gamma \theta$. It will be shown formally in Section 6.2 that the limiting distribution of the test statistic is the supremum of a chi-square process indexed by $\gamma$ as in (6.12). Let $\phi(\cdot)$ denote the probability density function of the
standard normal distribution. Let \( e = (2Y - 1)\phi (X'\beta_0)/\Phi ((2Y - 1) X'\beta_0) \) and

\[
V(\gamma) = E[-e^2W_{\gamma}W'_{\gamma}],
\]

and let \( G \) denote a Gaussian process with the covariance kernel

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

Then, the asymptotic distribution of the \( QLR_n \) test becomes

\[
\left(3.4\right) \frac{1}{2} \left[ \sup_{\gamma} G(\gamma)'V(\gamma)^{-1}G(\gamma) - G_1'V_{\beta}^{-1}G_1 \right],
\]

where \( G_1 \) and \( V_{\beta} \) denote the first \( k_\beta \) elements of \( G \) and the first \( k_\beta \times k_\beta \) block of \( V(\gamma) \), respectively. Here, \( k_\beta \) denotes the dimension of \( \beta \).

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 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

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

\[
\nabla_{\beta} q_i = X_i(2Y_i - 1)\frac{\phi(X'_{\beta}\tilde{\beta})}{\Phi[(2Y_i - 1)X'_{\beta}\tilde{\beta}]},
\]

where \( W_{\gamma,i} := (X'_i, Z'1\{T_i > \gamma}\})' \). Here is a brief description of the procedure:

1. to generate i.i.d. \( N(0,1) \) random variables \( \{v_{ij}\}_{i=1}^n \) for \( j = 1, \ldots, J \) for a sufficiently large \( J \);
(2) to simulate unrestricted and restricted score functions, respectively:

\[ G_{n,\theta}^j (\gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{\theta} \hat{q}_i (\gamma) v_{ij} \]

and

\[ G_{n,b}^j = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{b} \tilde{q}_i v_{ij}; \]

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

\[ D_n^j = \sup_{\gamma} \frac{1}{2} \left[ G_{n,\theta}^j (\gamma)' \hat{T}_{\theta,\gamma}^{-1} G_{n,\theta}^j (\gamma) - G_{n,b}^j \hat{T}_b^{-1} G_{n,b}^j \right] \]

where \( \hat{T}_{\theta,\gamma}^{-1} = (1/n) \sum_{i=1}^{n} \nabla_{\theta} \hat{q}_i (\gamma) \nabla_{\theta} \hat{q}_i (\gamma)' \) and \( \hat{T}_b^{-1} = (1/n) \sum_{i=1}^{n} \nabla_{b} \tilde{q}_i \nabla_{b} \tilde{q}_i \), respectively;

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

We use this simulated p-value \( \hat{p}_n^J \) to decide whether to accept or reject the null hypothesis.

4. Monte Carlo Simulations

In this section we investigate finite sample properties of the proposed test by Monte Carlo experiments. 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:

\[ Y^* = \beta_0 + \beta_1 X + \alpha Z \psi (T, \gamma) + U \]

\[ Y = 1 \{Y^* > 0\}, \]
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 $\alpha$ is equal to zero. Under the alternatives, $\alpha$ has various non-zero values from 0.2 to 1. The covariates $X$ and $Z$ 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 $\gamma$ are estimated by the Newton-Raphson’s method, and the threshold parameter $\gamma$ 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 $J = 2000$.

**Table 1.** Finite Sample Size of Nominal 5% Test

| Sample Size | Probit | | | Logit | | |
|-------------|--------|--------|--------|--------|--------|
| | 50 | 100 | 200 | 50 | 100 | 200 |
| Threshold | 0.067 | 0.049 | 0.052 | 0.055 | 0.056 | 0.052 |
| Smooth Transition | 0.046 | 0.064 | 0.051 | 0.054 | 0.061 | 0.049 |

Table 1 and Figures 1–2 summarize the result of the simulation study. Overall, the test performs well as expected from the theory. First, Table 1 reports the finite sample size of the test when the nominal level is 5%. Under the null distribution of $\alpha = 0$, the rejection rates of the test are close to the nominal level in most cases. Second, Figures 1–2 show the power of the test when $\alpha$ increases from 0 to 1. The result indicates that, in all cases, the power increases fast as the parameter value of $\alpha$ is farther away from zero. The test shows good performance even with a relatively small sample size, say $n = 100$. 
5. 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 used a subsample of the dataset originally used by Card et al. (2008). Among three different base years, we chose a sample of which base year is 1980. Next we picked eleven major cities and tested if there is a tipping point. To illustrate our testing procedure, specifically, we consider the probit and logit models.
We suppose that data \( \{(Y_i, X_i, T_i) : i = 1, \ldots, n\} \) are generated from

\[
Dw_i = \beta_0 + \alpha_0 1\{T_i > \gamma_0\} + X_i'\delta_0 + \epsilon_i
\]

\[
Y_i = 1\{Dw_i > 0\},
\]

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 \( X_i \) is a vector of six tract-level control variables. The \( X \) 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 in the current application, we treat this as a latent variable to illustrate our testing procedure for probit and logit models. The error term \( \epsilon_i \) follows either the standard normal or the logistic distribution depending on the model specification. Thus, the null and alternative hypotheses in our setting are

\[
H_0 : \alpha_0 = 0 \quad \text{and} \quad H_1 : \alpha_0 \neq 0,
\]

respectively.

The eleven cities we have chosen are Atlanta, Boston, Chicago, Houston, Miami, Nashville, New York, Philadelphia, Pittsburgh, Portland-Vancouver, and Washington DC. The p-values were calculated by the simulation method described in Section 4 with 1,000 simulations. For estimating a tipping point \( (\gamma) \) under the alternative, we used the grid search method. The grid points were constructed from \( T_i \) that fell in the interval \([l, 50\%]\) where \( l \) is the maximum of 5% and the 5th percentile of \( \{T_i\} \).

We summarize the result in Tables 2–3. 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 calculated this average marginal effect
Table 2. Test for Tipping in Segregation: Probit Model

| City                  | obs. | p-value | Tipping Points ($\gamma$) | $E_X[\Delta \Pr(y = 1|X)]$ |
|-----------------------|------|---------|--------------------------|---------------------------|
| Atlanta, GA           | 596  | 0.0%    | 6.55                     | -0.14                     |
| Boston, MA            | 700  | 2.8%    | 46.80                    | -0.25                     |
| Chicago, IL           | 1813 | 0.0%    | 48.74                    | -0.34                     |
| Houston, TX           | 763  | 0.0%    | 42.42                    | -0.25                     |
| Miami, FL             | 341  | 4.5%    | 42.30                    | -0.10                     |
| Nashville, TN         | 247  | 4.7%    | 8.38                     | -0.21                     |
| New York, NY          | 2430 | 0.0%    | 14.01                    | -0.09                     |
| Philadelphia, PA      | 1300 | 0.0%    | 39.64                    | -0.30                     |
| Pittsburgh, PA        | 663  | 0.0%    | 44.74                    | -0.45                     |
| Portland, OR-Vancouver, WA | 409 | 0.0% | 40.03                    | -0.74                     |
| Washington, DC        | 959  | 0.0%    | 41.72                    | -0.20                     |

Table 3. Test for Tipping in Segregation: Logit Model

| City                  | obs. | p-value | Tipping Points ($\gamma$) | $E_X[\Delta \Pr(y = 1|X)]$ |
|-----------------------|------|---------|--------------------------|---------------------------|
| Atlanta, GA           | 596  | 0.0%    | 6.55                     | -0.15                     |
| Boston, MA            | 700  | 3.9%    | 46.80                    | -0.26                     |
| Chicago, IL           | 1813 | 0.0%    | 48.74                    | -0.33                     |
| Houston, TX           | 763  | 0.0%    | 42.42                    | -0.23                     |
| Miami, FL             | 341  | 6.7%    | 42.30                    | -0.10                     |
| Nashville, TN         | 247  | 5.4%    | 8.38                     | -0.20                     |
| New York, NY          | 2430 | 0.0%    | 14.01                    | -0.09                     |
| Philadelphia, PA      | 1300 | 0.0%    | 39.64                    | -0.30                     |
| Pittsburgh, PA        | 663  | 0.1%    | 44.74                    | -0.44                     |
| Portland, OR-Vancouver, WA | 409 | 0.0% | 40.03                    | -0.75                     |
| Washington, DC        | 959  | 0.0%    | 41.72                    | -0.19                     |

as

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

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

First of all, testing results show that there exist tipping points in most of the cities. We only cannot reject the null of no tipping in Miami and Nashville at the 5%
significance level with the logit specification. However, their p-values are very close to 5% and we can reject the null for both cities with the probit specification. Second, the tipping points vary from 6.55% in Atlanta 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. For examples, cities like Atlanta, Miami, and New York show that the probability drops less than 15%. Meanwhile, Chicago, Pittsburgh, and Portland-Vancouver show that it drops more than 30%. Finally, not surprisingly, there is no significant difference between probit and logit models.

6. 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. In the online supplementary materials, we verify regularity conditions for maximum score estimation and also for quantile regression. 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.

6.1. 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

$$Q_n \left( \hat{\theta}_\gamma, \gamma \right) = \sup_{\theta \in \Theta} Q_n \left( \theta, \gamma \right) + o_{p^*} \left( r_n^{-2} \right),$$
where \( o_{p} (1) \) 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

\[
Q_{n} (\tilde{\beta}) = \sup_{\beta \in B} Q_{n} (\beta) + o_{p} (r_{n}^{-2}),
\]

where \( Q_{n} \) 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

\[
(6.1) \quad \mathcal{F}_{\delta} = \{ q_{\theta,\gamma} - q_{\theta_{0},\gamma} : |\theta - \theta_{0}| < \delta, \gamma \in \Gamma \},
\]

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 \( P F^{2} < \infty, |f(x)| \leq F(x) \) for any \( x \) and \( f \in \mathcal{F} \). An envelope function for \( \mathcal{F}_{\delta} \) is denoted by \( F_{\delta} \).

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 \( 0 < Q F^{r} < \infty \). While the measurability is an issue in the formal discussion of uniform entropy conditions, it hardly matters in applications. We assume measurability throughout the paper. See e.g. van der Vaart and Wellner (1996) for more general discussions on the empirical process method.
We now present the assumptions, whose details will be discussed later on.

**Assumption 6.1** (Uniform Consistency). \( \hat{\theta}(\gamma) = \theta_0 + o_p(1) \) and \( \tilde{\beta} = \beta_0 + o_p(1) \).

A set of sufficient conditions for the uniform consistency in Assumption 6.1 is that (i) uniform convergence of the objective function \( Q_n \); (ii) separability of the true value. Formally, we present it as Lemma 6.2 in Section 6.2.

**Assumption 6.2** (Uniform Rates of Convergence in Probability). \( r_n (\tilde{\beta} - \beta_0) = O_p(1) \) and \( r_n (\hat{\theta}(\gamma) - \theta_0) = O_{p_{\gamma}}(1) \).

Most often, the rate \( r_n \) in Assumption 6.2 is already known for linear models and \( r_n \) must be the same for \( \hat{\theta}(\gamma) \) for each \( \gamma \) since \( g(w, \theta, \gamma) \) is a linear function in \( \theta \). Thus, Assumption 6.2 has mainly to do with verifying the uniformity. However, the entropy conditions below in Assumption 6.4 are almost sufficient to ensure it, as will be shown in Lemma 6.3 in Section 6.2.

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

**Assumption 6.3** (Lindeberg Condition and \( L_2 \)-Continuity). For any \( \eta > 0 \),

\[
\frac{r_n^4}{n} P F_{K/r_n}^2 = O(1), \\
\frac{r_n^4}{n} P F_{K/r_n}^2 \{ \frac{r_n^2}{\sqrt{n}} F_{K/r_n} > \eta \sqrt{n} \} = o(1).
\]

In addition, for any decreasing sequence \( \eta_n \to 0 \),

\[
(6.2) \quad \sup_{|h_1 - h_2| < \eta_n, |\gamma_1 - \gamma_2| < \eta_n} \frac{r_n^4}{n} P \left( q_{\theta_0+h_1/r_n,\gamma_1} - q_{\theta_0+h_2/r_n,\gamma_2} \right)^2 = o(1).
\]

Assumption 6.3 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 6.4** (Entropy Conditions). For every decreasing sequence $\delta_n \to 0$,

\begin{equation}
\sup_Q \int_0^{\delta_n} \sqrt{\log N \left( \varepsilon \left\| F_{K/r_n} \right\|_{Q,2}, \mathcal{F}_{K/r_n}, L_2(Q) \right) } d\varepsilon \to 0,
\end{equation}

or

\begin{equation}
\int_0^{\delta_n} \sqrt{\log N \left[ \varepsilon \left\| F_{K/r_n} \right\|_{P,2}, \mathcal{F}_{K/r_n}, L_2(P) \right] } d\varepsilon \to 0.
\end{equation}

The entropy conditions imposed in Assumption 6.4 are high-level conditions. We provide sufficient conditions and some discussions in Section 6.2. Partition $h$ into $(h', b')'$ according to the dimensions of $\theta$ and $b$, respectively.

**Assumption 6.5** (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 $K$,

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

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

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

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

\begin{equation}
\frac{E G_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
\end{equation}

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

The limit process over which the supremum will be taken is characterized by the terms given in Assumption 6.5. 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
(6.5) in Assumption 6.5 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 \( \gamma \), then one can choose \( r = 1 \) in (6.5).

We now present our main theorem.

**Theorem 6.1.** Under Assumptions 6.1-6.5,

\[
(6.6) \quad QLR_n \Rightarrow \sup_{\gamma} \left[ \sup_{h: h_b=0} G(h, \gamma) - \sup_{h: h_a=0} (-G(h, \gamma)) \right],
\]

where \( G = G_1 + G_2 \).

While the asymptotic null distribution of \( QLR_n \) is well-defined under the restriction in Assumption 6.5, 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 paper. When we can solve out the maximizers explicitly for the expression inside the bracket in (6.6), simulating the critical values in the spirit of Hansen (2006) can also be applied.

6.2. Low-Level Sufficient Conditions for Assumptions. This section provides low-level sufficient conditions for Assumptions 6.1-6.5. First, we present the following lemma that can be used to verify Assumption 6.1.

**Lemma 6.2.** Let \( \mathcal{F} \) be a class of functions \( \{q_{\theta, \gamma} : (\theta, \gamma) \in \Theta \times \Gamma\} \) with envelope \( F \) such that \( P F < \infty \). Suppose either of the following two conditions is satisfied: (i) \( N_{[\varepsilon]}(\varepsilon, \mathcal{F}, L_1(P)) < \infty \) for every \( \varepsilon > 0 \); (ii) For \( \mathcal{F}_M \) defined as the class of functions \( f1 \{ F \leq M \} \) for \( f \in \mathcal{F} \), \( \log N(\varepsilon, \mathcal{F}_M, L_1(P_n)) = o_p(n) \) for every \( \varepsilon \) 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

$$
(6.7) \quad \sup_{\gamma \in \Gamma} \left[ \sup_{\theta \not\in \Theta_0} \{ Q(\theta_0, \gamma) - Q(\theta, \gamma) \} \right] > 0
$$

for every open set $\Theta_0$ that contains $\theta_0$. Then, $\hat{\theta}(\gamma) - \theta_0 = o_P(1)$.

While there are different ways to present sufficient conditions for Assumption 6.1, we chose this way as the subsequent discussion also draws on the entropy conditions. The entropy conditions in Lemma 6.2 are almost automatically satisfied when other regularity conditions that are imposed in the paper are met. Thus, separability is the one we need to check. Recall that $Q(\theta_0, \gamma)$ is the same for all $\gamma$ since $\gamma$ is not identified under the null. However, once we establish the consistency for a given $\gamma$ and that $Q(\theta_0, \gamma) > \sup_{\theta \not\in \Theta} Q(\theta, \gamma)$, the verification of the separability is not very difficult since $\gamma$ appears only through an indicator function.

We now consider sufficient conditions for Assumption 6.2. 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 6.3.** Assume that for every $\theta$ in a neighborhood of $\theta_0$,

$$
(6.8) \quad \sup_{\gamma} \mathbf{P} (q_{\theta, \gamma} - q_{\theta_0, \gamma}) \leq -C |\theta - \theta_0|^2,
$$

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

$$
(6.9) \quad E \sup_{\gamma} \sup_{|\theta - \theta_0| < \delta} |G_n (q_{\theta, \gamma} - q_{\theta_0, \gamma})| = O (\phi(\delta)),
$$
for a function $\phi$ such that $\phi(\delta)/\delta^r$ is decreasing for some $r < 2$. If Assumption 6.1 holds, then

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

for every $r_n$ such that $r_n^2 \phi(1/r_n) \leq \sqrt{n}$ for every $n$. If the rate $r_n$ is known, then (6.9) can be stated for $\delta = K/r_n$ and $\phi(\delta) = \sqrt{n}/r_n^2$.

The first condition (6.8) is not difficult to verify. Often, $P_{q_{\theta,\gamma}}$ is twice continuously differentiable at $\theta_0$ for all $\gamma$. 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 (6.9) is implied by Assumptions 6.3 and 6.4. It is known that the left-hand side term in the equation (6.9) is bounded by the product of the $L_2$ norm of the envelope function, $P^{1/2}(F_2^2)$, and the uniform entropy integral or the bracketing integral, which is defined respectively by

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

or

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

See e.g. Theorems 2.14.1 and 2.14.2 in van der Vaart and Wellner (1996). These are bounded by Assumption 6.4. Thus, in case when the rate $r_n$ is not known $a\ priori$, it is typical that $\phi^2(\delta) = PF^2_\delta$ yields the correct rate, leading to the rate as the solution of $r_n^4 PF^2_1/r_n \sim n$. This is in fact the first condition in Assumption 6.3.

We now provide sufficient conditions for Assumption 6.4. The following condition implies Assumption 6.4.

**Assumption 6.4**. For some $\delta_0 > 0$, assume that

$$\int_0^1 \sup_{\delta < \delta_0} \sup_Q \sqrt{\log N \left( \varepsilon \|F_\delta\|_{Q,2}, F_\delta, L_2(Q) \right)} d\varepsilon < \infty \quad (6.10)$$
(6.11) \[ \int_0^1 \sup_{\delta < \delta_0} \sqrt{\log N_{[1]} \left( \varepsilon \| F_\delta \|_{P,2}, F_\delta, L_2 (P) \right)} d\varepsilon < \infty. \]

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 \( \varepsilon^{-1} \), thus satisfying (6.10) 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. It is also obvious then that these classes of functions satisfy the conditions for the uniform convergence in Lemma 6.2.

Many interesting examples feature the estimating function \( q \) in the form of Lipschitz of order \( r \) transformation in the sense that \( q_{\theta,\gamma} = q(y, g(w, \theta, \gamma)) \) and

\[ |q(y, g(w; \theta_1, \gamma_1)) - q(y, g(w, \theta_2, \gamma_2))| \leq L_r(w) |g(w; \theta_1, \gamma_1) - g(w; \theta_2, \gamma_2)|^r, \]

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

**Lemma 6.4.** Suppose that \( F_\delta \) 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 \( \delta_0 > 0 \),

\[ \int_0^1 \sup_{\delta < \delta_0} \sup_{Q} \sqrt{\log N \left( \varepsilon \| F_\delta \|_{Q,2}, F_\delta, L_2 (Q) \right)} d\varepsilon < \infty. \]
Let \( \phi(\delta) = \delta^r \). Then, there exists an envelope function \( F_\delta \) such that for every \( \eta > 0 \),

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

The lemma specifies the functional form of \( \phi(\delta) \) as \( \delta^r \), 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 and logit models.

If \( \mathbf{P}_{q,\theta,\gamma} \) is twice continuously differentiable at \( \theta = \theta_0 \) with a unique maximum at \( \theta_0 \), Assumptions 6.1 and 6.2 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 \( \gamma \) and most applications belong to this case.

**Corollary 6.5.** Suppose that the function \( \mathbf{Q}(\theta, \gamma) \) has a well-separated maximum \( \theta_0 \) in the sense of (6.7) and it is twice continuously differentiable at \( \theta_0 \) with a negative second derivative matrix, say \( -V(\gamma) \), whose maximum eigenvalues are bounded away from zero for all \( \gamma \). Let \( V_\beta \) denote the block of \( V(\gamma) \) that is associated with the second derivative with respect to \( \beta \). Then, \( r_n^2 \mathbf{P} m_{h_n,\gamma} \to -\frac{1}{2} V_\theta V(\gamma) h_\theta + \frac{1}{2} V_\beta V h_b = G_2(h, \gamma) \), uniformly over any compact set. If Assumptions 6.3 and 6.4 (or 6.4\*) hold with a sequence \( r_n \), then \( r_n \left( \hat{\theta}(\gamma) - \theta_0 \right) = O_p(1) \) and \( r_n \left( \hat{\beta} - \beta_0 \right) = O_p(1) \). If Assumption 6.5 holds as well, then

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

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

\[
G_1(h, \gamma) = h' \mathbf{G}(\gamma) = (h_\beta - h_b)' \mathbf{G}_1 + h_\alpha \mathbf{G}_2(\gamma),
\]
where \( G(\gamma) = (G_1(\gamma), G_2(\gamma))' \) 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

\[
(6.12) \quad \frac{1}{2} \left[ \sup_{\gamma} G(\gamma)' V(\gamma)^{-1} G(\gamma) - G_1(\gamma)' V_\beta^{-1} G_1 \right].
\]

Standard linear algebra allows us to write this as

\[
\frac{1}{2} \sup_{\gamma} G(\gamma)' H(\gamma) H(\gamma)' G(\gamma),
\]

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

We now verify regularity conditions for the probit model. Note that the function \( q(y, w; \theta, \gamma) \) in (3.3) is Lipschitz of order 1 transformation and twice continuously differentiable in \( \theta \). Therefore, applying Lemma 6.4 and Corollary 6.5, we only need to check the separability condition (6.7) and Assumption 6.5. We assume the following regularity conditions:

\textbf{Assumption 6.6.} (i) The parameters \( \theta \) and \( \gamma \) are in the interior of compact sets \( \Theta \) and \( \Gamma \) where \( \Gamma \) is contained in an open subset of the support of \( T \).

(ii) For any \( \gamma \), the matrix \( E \left[ W_\gamma W_\gamma' \right] \) exists and is nonsingular.

(iii) \( T \) is continuously distributed.

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

Since $q (\cdot)$ is twice continuously differentiable, it follows from the discussion following Corollary 6.5 that the limiting distribution of the test statistic is the supremum of a chi-square process indexed by $\gamma$ as in (6.12). 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.

7. 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.

8. Supplemental Materials

The supplement to this article contains all the mathematical proofs and additional theoretical results. In particular, (i) we verify regularity conditions for maximum score estimation and also for quantile regression; (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; and (iii) we discuss the consistency and local power of the test when the null hypothesis is false.


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.
*Journal of Econometrics* 96(1), 113–144.


*Journal of Finance* 57(5), 2223–2261.


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

*Journal of Econometrics* 93(2), 345–368.

*IMF Staff Papers* 48(1), 1–21.


*Journal of Econometrics* 144(2), 492–499.


*Journal of Econometrics* 3(3), 205–228.


Department of Economics, University College London, Gower Street, London, WC1E 6BT, UK.

E-mail address: l.simon@ucl.ac.uk

URL: http://www.homepages.ucl.ac.uk/~uctplso.

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.