1-2020

Inference in Threshold Models

Yoonseok Lee
Syracuse University, ylee41@maxwell.syr.edu

Yulong Wang
Syracuse University

CENTER FOR POLICY RESEARCH

THE MAXWELL SCHOOL

WORKING PAPER SERIES

Inference in Threshold Models

Yoonseok Lee and Yulong Wang

Paper No. 223
January 2020

ISSN: 1525-3066

Maxwell
Syracuse University
Center for Policy Research
426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020
(315) 443-3114/ email: ctrpol@syr.edu
http://www.maxwell.syr.edu/CPR_Working_Papers.aspx

CENTER FOR POLICY RESEARCH – Spring 2020

Leonard M. Lopoo, Director
Professor of Public Administration and International Affairs (PAIA)

Associate Directors

Margaret Austin
Associate Director, Budget and Administration

John Yinger
Trustee Professor of Economics (ECON) and Public Administration and International Affairs (PAIA)
Associate Director, Center for Policy Research

SENIOR RESEARCH ASSOCIATES

Badi Baltagi, ECON
Robert Bifulco, PAIA
Leonard Burman, PAIA
Carmen Carrión-Flores, ECON
Alfonso Flores-Lagunes, ECON
Sarah Hamersma, PAIA
Madonna Harrington Meyer, SOC
Colleen Heflin, PAIA
William Horrace, ECON
Yilin Hou, PAIA
Hugo Jales, ECON
Jeffrey Kubik, ECON
Yoonseok Lee, ECON
Amy Lutz, SOC
Yingyi Ma, SOC
Katherine Michelmore, PAIA
Jerry Miner, ECON
Shannon Monnat, SOC
Jan Ondrich, ECON
David Popp, PAIA
Stuart Rosenthal, ECON
Michah Rothbart, PAIA
Alexander Rothenberg, ECON
Rebecca Schewe, SOC
Amy Ellen Schwartz, PAIA/ECON
Ying Shi, PAIA
Saba Siddiki, PAIA
Perry Singleton, ECON
Yulong Wang, ECON
Michael Wasylenko, ECON
Peter Wilcoxen, PAIA
Maria Zhu, ECON

GRADUATE ASSOCIATES

Rhea Acuña, PAIA
Mariah Brennan, SOC. SCI.
Jun Cai, ECON
Ziqiao Chen, PAIA
Yoon Jung Choi, PAIA
Dahae Choo, PAIA
Stephanie Coffey, ECON
Giuseppe Germinario, ECON
Myriam Gregoire-Zawilski, PAIA
Emily Gutierrez, PAIA
Jeehee Han, PAIA
Mary Helander, Lerner
Hyoung Kwon, PAIA
Mattie Mackenzie-Liu, PAIA
Maeve Maloney, ECON
Austin McNeill Brown, SOC. SCI.
Qasim Mehdi, PAIA
Claire Pendergrast, SOC
Jonathan Presler, ECON
Krushna Ranaware, SOC
Christopher Rick, PAIA
David Schwegman, PAIA
Saied Toossi, PAIA
Huong Tran, ECON
Joaquin Urrego, ECON
Yao Wang, ECON
Yi Yang, ECON
Xiaoyan Zhang, ECON
Bo Zheng, PAIA
Dongmei Zhu, SOC. SCI.

STAFF

Joanna Bailey, Research Associate
Joseph Boskovski, Manager, Maxwell X Lab
Katrina Fiacchi, Administrative Specialist
Michelle Kincaid, Senior Associate, Maxwell X Lab
Emily Minnoe, Administrative Assistant
Candi Patterson, Computer Consultant
Samantha Trajkovski, Postdoctoral Scholar
Laura Walsh, Administrative Assistant

Abstract

This paper develops new statistical inference methods for the parameters in threshold regression models. In particular, we develop a test for homogeneity of the threshold parameter and a test for linear restrictions on the regression coefficients. The tests are built upon a transformed partial-sum process after re-ordering the observations based on the rank of the threshold variable, which recasts the cross-sectional threshold problem into the time-series structural break analogue. The asymptotic distributions of the test statistics are derived using this novel approach, and the finite sample properties are studied in Monte Carlo simulations. We apply the new tests to the tipping point problem studied by Card, Mas, and Rothstein (2008), and statistically justify that the location of the tipping point varies across tracts..

JEL No.: C12, C24

Keywords: Threshold Regression, Test, Homogeneous Threshold, Linear Restriction, Tipping Point

Authors: Yoonseok Lee, Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020, ylee41@maxwell.syr.edu; Yulong Wang, Department of Economics and Center for Policy Research, 127 eggers Hall, Syracuse University, Syracuse, NY 13244-1020, ywang402@maxwell.syr.edu

Acknowledgement

The authors thank Ulrich Müller, Bo Honoré, Mark Watson, Kirill Evidokimov, Simon Lee, Myung Seo, Zhijie Xiao, and particpants at numerous seminar/conference presentations for very helpful discussions. Lee acknowledges financial support from the CUSE grant; Wang acknowledges financial support from the Appleby-Mosher grant.

1 Introduction

Threshold regression models have been widely used and studied in economics and statistics. See, among many others, Hansen (2000a), Caner and Hansen (2004), Seo and Linton (2007), Lee, Seo, and Shin (2011), Li and Ling (2012), Yu (2012), Lee, Liao, Seo, and Shin (2018), Hidalgo, Lee, and Seo (2019), and Yu and Fan (2019).

This paper proposes a new framework for testing hypotheses about the parameters in threshold regression models. In particular, we treat the rank statistics of the threshold variable as time and recast cross-sectional threshold models into the time-series structural break counterparts. Based on this transformation, we develop a test for homogeneity of the threshold parameter (i.e., a constant threshold) and a test for linear restrictions on the regression coefficients. The latter test can be used to test whether there exists a threshold effect. Both tests are empirically motivated by the tipping point problem.

The tipping point model is proposed by Schelling (1971) to analyze the phenomenon that the neighborhood's white population substantially decreases once the minority share exceeds a certain threshold, called the tipping point. Card, Mas, and Rothstein (2008) empirically study this phenomenon by considering the following threshold regression model:

$$y_i = \beta_{01} + \delta_{01}\mathbf{1}\left[q_i < \gamma_0\right] + x_i^{\intercal}\beta_{02} + u_i \tag{1}$$

for neighborhoods $i = 1, \ldots, n$, where the observed variables y_i, q_i, and x_i denote the white population change in a decade, the initial minority share, and other social characteristics in the ith tract, respectively. The unknown parameters, $(\beta_{01}, \beta_{02}^{\intercal}, \delta_{01})^{\intercal}$ and γ_0, denote the regression coefficients and the threshold, respectively. With the model (1), one is interested in testing whether $\delta_{01} = 0$ or not, that is, testing if the tipping point phenomenon exists. More generally, we construct a new test for linear restrictions on the regression coefficients, the average likelihood-ratio (LR) type test (named as the AG test), which is inspired by Andrews (1993), Andrews and Ploberger (1994), and Elliott, Müller, and Watson (2015). In the problem of testing whether $\delta_{01} = 0$, the AG test strongly rejects the null hypothesis of no threshold effect and reinforces the existing founding in Card, Mas, and Rothstein (2008). See also Lee, Seo, and Shin (2011). Compared with existing methods, this new test has substantially higher powers as we show in Monte Carlo simulations.

When the test rejects the null hypothesis of no threshold, one wants to examine the assumption that the tipping point γ_0 remains constant across neighborhoods. Card, Mas, and Rothstein (2008) first assume γ_0 to be a constant within a city and estimate the model

(1) with tract-level data. After collecting the results from all the cities, they further regress the estimated γ_0 on a measure of white population's attitude to the minority at the city level, and find that the tipping point highly depends on this measure. This finding raises the concern that γ_0 may vary across neighborhoods (tracts), which motivates our constant-threshold test (named as the CT test). Specifically, we develop a test for a constant threshold γ_0 against any types of heterogeneous thresholds γ_i (or nonparametric alternatives), which is new to the literature. This test strongly rejects the null hypothesis of a constant threshold, implying that the model (1) is insufficient to characterize the tipping phenomenon. See Lee and Wang (2019) and Yu and Fan (2019) for other motivating examples.

To develop the new tests, we first reframe the cross-sectional threshold model (1) into its time-series structural break analogue. This is done by re-ordering the data according to q_i and treating the rank statistic of q_i as time. We then construct the partial sum process of the re-ordered $x_i\widehat{u}_i$ along the rank of q_i, where $\widehat{u}_i$ is the fitted residual. We construct tests based on the limiting distribution of this partial sum process, which is close in spirit to the methods developed by the aforementioned works in structural break problems (e.g., Elliott and Müller (2007) and Elliott and Müller (2014)).

It should be noted that, however, this re-ordering does not allow us to directly apply the existing tests in the structural break literature. It is mainly because, once we see the rank based on the quantiles of q_i as time, the re-ordering results in time-varying moments of other induced order statistics that lead to a nonstandard limiting distribution of the partial sum process. In comparison, the corresponding moments are time invariant in the structural break models. To solve this nonstationarity issue, we construct a novel transformation that recovers the simple and tractable limiting observation, which consists of a standard Wiener process and a piecewise linear drift term. Recovering this simple limit allows us to develop tests whose limiting distributions are free from nuisance parameters.

The rest of the paper is organized as follows. Section 2 introduces the re-ordering and transformation idea and studies asymptotics of the partial sum process of the induced order statistics. Using this asymptotic results, Section 3 constructs two new tests and studies their limiting properties. Section 4 examines their finite sample performance by Monte Carlo simulations, and Section 5 revisits the tipping point problem as an illustration. Section 6 concludes with some remarks. All proofs are collected in the Appendix.

We use the following notations. Let $\rightarrow_p$ denote convergence in probability, $\rightarrow_d$ convergence in distribution, and $\Rightarrow$ weak convergence of stochastic processes as $n \rightarrow \infty$. Let $=_d$ denote equivalence in distribution. Let $\lfloor a \rfloor$ denote the biggest integer smaller than a and

$\mathbf{1}[A]$ the indicator function of a generic event A. Let $\|B\|$ denote the Euclidean norm of a vector or matrix B.

2 Preliminaries

2.1 Partial sum process

We consider a threshold regression model given by

$$y_i = x_i^\intercal \beta_0 + x_i^\intercal \delta_0 \mathbf{1}\left[q_i \le \gamma_0\right] + u_i \tag{2}$$

for $i = 1, \ldots, n$, where the variables $(y_i, x_i^\intercal, q_i)^\intercal \in \mathbb{R}^{1+k+1}$ are observed but the threshold parameter $\gamma_0 \in \mathbb{R}$ as well as the regression coefficients $\theta_0 = (\beta_0^\intercal, \delta_0^\intercal)^\intercal \in \mathbb{R}^{2k}$ are unknown. All these parameters can be consistently estimated by the standard profile least squares method as Bai and Perron (1998) and Hansen (2000a). Specifically, we estimate γ_0 by minimizing

$$\sum_{i=1}^{n} \left(y_i - x_i^\intercal \widehat{\beta}(\gamma) + x_i^\intercal \widehat{\delta}(\gamma) \mathbf{1}\left[q_i \le \gamma\right] \right)^2$$

in γ, where $(\widehat{\beta}^\intercal(\gamma), \widehat{\delta}^\intercal(\gamma))^\intercal$ are the least squares estimators of (2) with a fixed γ. Once $\widehat{\gamma}$ is obtained, we let $\widehat{\theta} = (\widehat{\beta}^\intercal, \widehat{\delta}^\intercal)^\intercal = (\widehat{\beta}^\intercal(\widehat{\gamma}), \widehat{\delta}^\intercal(\widehat{\gamma}))^\intercal$.

Similarly as Nyblom (1989) and Elliott and Müller (2007), we develop tests based on the partial sum process of $x_i \widehat{u}_i$, where $\widehat{u}_i$ is the fitted residual. When the index i has some natural ordering, such as time in the structural break model, the definition of the partial sum is straightforward. However, we do not have such natural ordering for the cross-sectional observations in general. In this section, we propose an ordering method based on the rank of the threshold variable q_i, and study the partial sum process with the re-ordered observations.

We suppose q_i is a continuous random variable whose distribution function, $F(\cdot)$, is continuous and monotonically increasing. We define $r_0 = F(\gamma_0) \in (0, 1)$. Then we can rewrite (2) as

$$y_i = x_i^\intercal \beta_i + u_i,$$

where

$$\beta_i = \beta_0 + \delta_0 \mathbf{1}\left[F(q_i) \le r_0\right]. \tag{3}$$

In this setup, the threshold variable q_i affects the parameter stability through $F(q_i)$, where

$F(q_i)$ is a standard uniform random variable. Once we sort $\{F(q_i)\}_{i=1}^n$ ascendingly, we can treat them as an irregularly-spaced "time" from the perspective of structural break. In practice, we replace $F(\cdot)$ with the empirical distribution, $\widehat{F}_n(\cdot)$, and then $\widehat{F}_n(q_i)$ equals to R_i/n, where R_i denotes the rank statistic of q_i. By doing so, we can form an equi-spaced time (i.e., ordering) induced by the rank of q_i.

More precisely, we let $Q(\cdot) = F^{-1}(\cdot)$ denote the quantile function of q_i. We assume the density of q_i, denoted as $f(\cdot)$, to be continuous and positive over the support of q_i, implying that $Q(\cdot)$ is continuous and strictly increasing. By sorting $\{q_i\}_{i=1}^n$ into the order statistics $q_{(1:n)} \leq q_{(2:n)} \leq \ldots \leq q_{(n:n)}$ and re-arranging the data according to their ranks, we denote the re-ordered observations $(y_i, x_i^\intercal)^\intercal$ associated with $q_{(i:n)}$ as $(y_{[i:n]}, x_{[i:n]}^\intercal)^\intercal$, that is, $(y_{[i:n]}, x_{[i:n]}^\intercal)^\intercal = (y_j, x_j^\intercal)^\intercal$ if $q_{(i:n)} = q_j$.[1] Such re-ordered values are called induced order statistics or concomitants in the statistics literature (e.g., Bhattacharya (1974) and Yang (1985)). Similarly, we write the re-ordered β_i as

$$\beta_{[i:n]} = \beta_0 + \delta_0 \mathbf{1}\left[F(q_{(i:n)}) \leq r_0\right].$$

Such a re-ordering naturally covers structural break models, in which $q_{(i:n)} = q_i = i$ is the time. In what follows, we drop ": n" in the subscripts for simplicity. The subscript $[i]$ is reserved for the ith induced order statistics associated with the order statistic $q_{(i:n)}$.

Based on the re-ordering, we now construct the partial sum of $x_i \widehat{u}_i$ along the rank of q_i as

$$\widehat{G}_n(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} x_{[i]} \widehat{u}_{[i]} \tag{4}$$

for $s \in [0,1]$, where $\widehat{u}_i = y_i - x_i^\intercal \widehat{\beta} - x_i^\intercal \widehat{\delta} \mathbf{1}[q_i \leq \widehat{\gamma}]$. In order to derive the weak limit of $\widehat{G}_n(\cdot)$, we impose the following regularity conditions, which are similar to Condition 1 in Hansen (2000a). We define

$$\begin{aligned} D(r) &= \mathbb{E}\left[x_i x_i^\intercal | q_i = Q(r)\right] \\ V(r) &= \mathbb{E}\left[x_i x_i^\intercal u_i^2 | q_i = Q(r)\right] \end{aligned}$$

for $r \in [0,1]$.

[1] Since q_i is continuous, the probability of seeing ties is negligible. In finite samples, we may simply drop duplicate (i.e., tied) observations of q_i.

Condition 1

1. $(x_i^\intercal, u_i, q_i)^\intercal$ *is i.i.d.*

2. $\mathbb{E}[x_i u_i | q_i] = 0$ *almost surely.*

3. q_i *has a continuous density function* f *such that for all* q, $0 < f(q) < C$ *for some* $C < \infty$.

4. $\delta_0 = c_0 n^{-\epsilon}$ *for some* $c_0 \neq 0$ *and* $\epsilon \in (0, 1/2)$; $(c_0^\intercal, \beta_0^\intercal)^\intercal$ *belongs to some compact subset of* $\mathbb{R}^{2k}$.

5. $r_0 \in [\tau, 1-\tau]$ for some $\tau \in (0, 1/2)$.

6. $D(r)$ and $V(r)$ *are well-defined matrix-valued functions that are positive definite and continuously differentiable with bounded derivatives at all* $r \in (0, 1)$.

7. $\mathbb{E}[x_i x_i^\intercal] > \mathbb{E}[x_i x_i^\intercal \mathbf{1}[F(q_i) \leq r]] > 0$ *for all* $r \in (0, 1)$.

8. $\sup_{q \in \mathbb{R}} \mathbb{E}[||x_i u_i||^4 | q_i = q] < \infty$ *and* $\sup_{q \in \mathbb{R}} \mathbb{E}[||x_i||^4 | q_i = q] < \infty$.

Condition 1.1 assumes i.i.d. observations. Under this condition, we can show that $\{x_{[i]} u_{[i]}\}_{i=1}^n$ is a martingale difference array, which is a key condition for our main result. Weak dependence would break such a martingale property after re-ordering and hence dramatically complicates the analysis. We leave this to future research. Condition 1.2 assumes a correctly specified model. Condition 1.3 implies that the quantile function of q_i is continuous and uniquely defined for all i. Condition 1.4 adopts the widely used shrinking change size setup as in Bai and Perron (1998) and Hansen (2000a), under which $\widehat{\theta} = (\widehat{\beta}^\intercal, \widehat{\delta}^\intercal)^\intercal$ is $\sqrt{n}$-consistent and asymptotically normal.[2] In Condition 1.5, the truncation is to avoid the threshold being close to the boundary so that there are infinitely many observations on both sides of the threshold. This is commonly assumed in both the structural break and the threshold model literature.

Condition 1.6 requires the moment function to be smooth so that $D(\cdot)$ and $V(\cdot)$ are well defined. These two functions are usually treated as constant matrices in the structural break literature (e.g., Li and Müller (2009) and Elliott and Müller (2014)). However, they can

[2] The case with $\epsilon = 0$ is also allowed in our approach by using the argument in Chan (1993). In this case, the limiting distribution of $\widehat{\gamma}$ is non-standard and non-pivotal. However, it is still consistent and converges at the rate n, which is sufficient for constructing our tests. We do not consider this case for illustrational simplicity.

be any continuous matrix-valued functions here. The smoothness of $D(\cdot)$ and $V(\cdot)$ can be generalized to piecewise smoothness with a finite number of jumps. It is worth noting that invertibility of $D(\cdot)$ excludes the situation that q_i is a linear combination of x_i or q_i is one of the elements of x_i when including a constant term. Condition 1.7 is a full-rank condition, and Condition 1.8 bounds the conditional moments.

Under Condition 1 and from Hansen (2000a), we can verify that the least squares estimator $\widehat{\gamma}$ is consistent and asymptotically independent of $\widehat{\theta} = (\widehat{\beta}^\intercal, \widehat{\delta}^\intercal)^\intercal$. Furthermore, it holds that (e.g., eq.(11) in Hansen (2000a))

$$\sqrt{n}\begin{pmatrix}\widehat{\beta} - \beta_0 \\ \widehat{\delta} - \delta_0\end{pmatrix} \to_d \begin{pmatrix}\Phi_\beta \\ \Phi_\delta\end{pmatrix} \tag{5}$$

as $n \to \infty$ for some k-dimensional normal random vectors Φ_β and Φ_δ. The following theorem derives the weak limit of $\widehat{G}_n(s)$ in (4).

Theorem 1 *Suppose Condition 1 holds. Then, $\widehat{G}_n(\cdot) \Rightarrow G(\cdot)$ as $n \to \infty$, where*

$$G(s) =_d \int_0^s V(t)^{1/2} dW_k(t) - \left(\int_0^s D(t)\,dt\right)\Phi_\beta - \left(\int_0^{\min\{s,r_0\}} D(t)\,dt\right)\Phi_\delta \tag{6}$$

for $s \in [0,1]$, Φ_β and Φ_δ are given in (5), and $W_k(\cdot)$ is the $k \times 1$ vector standard Wiener process defined on $[0,1]$.

Theorem 1 lays the foundation of our asymptotic analysis. In particular, the limiting observation $G(s)$ is to be used to motivate the key structure of our test statistics. Note that, in the special case that the functions $D(\cdot)$ and $V(\cdot)$ are respectively constant matrices $\bar{D}$ and $\bar{V}$, $G(s)$ reduces to

$$\bar{V}^{1/2} W_k(s) - \bar{D} s \Phi_\beta - \bar{D} \min\{s, r_0\}\Phi_\delta. \tag{7}$$

This is the limiting observation studied by Elliott and Müller (2014) and Elliott, Müller, and Watson (2015) for structural break problems. Comparing (6) with (7), the nuisance functions $D(\cdot)$ and $V(\cdot)$ substantially complicate the limiting expression. In the following subsection, we construct a novel transformation that recasts $G(s)$ into to its simpler form in (7).

2.2 Transformation

To construct the transformation, for any $k \times 1$ vector v satisfying $v^\intercal v = 1$, we first define two continuous and strictly increasing functions:

$$h(r) = \int_\tau^r \frac{1}{v^\intercal D(t)^{-1} V(t) (D(t)^{-1})^\intercal v} dt \quad \text{and} \quad g(r) = \frac{h(r)}{h(1-\tau)} \tag{8}$$

for $r \in [\tau, 1-\tau]$, where the truncation parameter τ is specified in Condition 1.5. It sets to ensure $0 < h(r) < \infty$ since $D(\cdot)$ and $V(\cdot)$ may not be well defined near the boundary of 0 or 1. Since the mapping $g : [\tau, 1-\tau] \mapsto [0,1]$ is strictly increasing with $g(\tau) = 0$ and $g(1-\tau) = 1$, we can treat $g(\cdot)$ as a transformed and rescaled time over the unit interval.

Using (8), we define the transformed process of $G(s)$ in (6) as

$$\mathcal{G}_v(s) = h_\tau^{1/2} \int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t) v^\intercal D(t)^{-1} dG(t), \tag{9}$$

where $h_\tau = h(1-\tau)$, $g^{(1)}(t) = \partial g(t)/\partial t = 1/\{v^\intercal D(t)^{-1} V(t) (D(t)^{-1})^\intercal v h_\tau\}$, and $g^{-1}(\cdot)$ is the inverse function of $g(\cdot)$.[3] In what follows, the calligraphic letter $\mathcal{G}$ and its variants are reserved for the transformed processes.

The intuition for constructing $\mathcal{G}_v$ can be explained as follows. First, we standardize the non-constant variance-covariance matrix $D(\cdot)$ by pre-multiplying its inverse matrix function $D(\cdot)^{-1}$. Second, to standardize $V(\cdot)$, we set $g^{(1)}(t)$ as the weighting function that is proportional to the inverse local Fisher information, $v^\intercal D(t)^{-1} V(t) (D(t)^{-1})^\intercal v$. Finally, because $\int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t) dt = s$ for any $s \in [0,1]$, we can transform the stochastic integral to a standard Wiener process while maintaining the deterministic term as a piecewise linear function. If $V(\cdot)$ and $D(\cdot)$ are respectively the constant matrices $\bar{V}$ and $\bar{D}$, this transformation is essentially the same as pre-multiplying $v^\intercal \bar{V}^{-1/2}$ to (7), which yields

$$W_1(s) - s v^\intercal \bar{V}^{-1/2} \bar{D} \Phi_\beta - \min\{s, r_0\} v^\intercal \bar{V}^{-1/2} \bar{D} \Phi_\delta. \tag{10}$$

The novelty of such transformation lies on the design of $g(\cdot)$ and the integral defined in the last step. In comparison, a transformation of the form $\int_0^s \mathcal{T}(t) dG(t)$ with any weighting function $\mathcal{T}(\cdot)$ can simplify only either the stochastic integral or the deterministic function,

[3] We do not have an explicit form of $g^{-1}(\cdot)$ in this case. However, it is well defined as $g^{-1}(s) = h^{-1}(s h_\tau)$, where the inverse function of $h(\cdot)$, $h^{-1}(\cdot)$, exists and is differentiable since $v^\intercal D(t)^{-1} V(t) (D(t)^{-1})^\intercal v$ is strictly positive for all $t \in [\tau, 1-\tau]$.

but not both of them simultaneously. Hence, our time transformation is different from those studied in the nonstationary time-series literature (e.g., Park and Phillips (1999)).

The following theorem gives the main motivation of this transformation: $\mathcal{G}_v(s)$ in (9), the transformed process of the limiting observation $G(s)$, is distributionally equivalent to the simple form in (10). Recall that we define $r_0 = F(\gamma_0)$.

Theorem 2 *The transformed process $\mathcal{G}_v(s)$ in (9) satisfies*

$$\mathcal{G}_v(s) =_d W_1(s) - sv^\intercal \Phi_\beta^h - \min\{s, g(r_0)\} v^\intercal \Phi_\delta^h \tag{11}$$

for $s \in [0,1]$, where $\Phi_\beta^h = h_\tau^{1/2} \Phi_\beta$ and $\Phi_\delta^h = h_\tau^{1/2} \Phi_\delta$.

Theorem 2 implies that the complicate $G(s)$ process can be transformed into the simple $\mathcal{G}_v(s)$, which consists of a standard Wiener process and (piecewise) linear drift terms. Therefore, once properly eliminating the drift terms by demeaning, we can construct test statistics whose limiting distributions are free from nuisance parameters. This idea is similar to some approaches in the structural break models (e.g., Elliott and Müller (2014) and Elliott, Müller, and Watson (2015)) that develop tests based on the partial-sum limit in (7), which resembles (11).

3 Tests for Threshold Models

In this section, we consider two testing problems: testing for homogeneity of the threshold parameter (i.e., a constant threshold) and testing for linear restrictions on the regression coefficients. Both problems can be analyzed in the unified framework introduced in the previous section.

3.1 Test for homogeneous threshold

For the structural break models and threshold regression models, most of the existing studies focus on testing for whether the coefficient change exists. However, once we reject the null hypothesis of no change, a natural question is then to consider whether one single threshold is sufficient to characterize the model. In this subsection, we develop a test for homogeneity of the threshold parameter, which is novel in the threshold model literature.

To construct the new test, we consider a heterogeneous threshold case given by

$$\beta_i = \beta_0 + \delta_0 \mathbf{1}\left[F(q_i) \leq r_i\right]$$

instead of (3), where r_i denotes a random variable defined on $(0,1)$. More precisely, as in Condition 1.5, we assume that $r_i \in [\tau, 1-\tau]$ for some $\tau \in (0, 1/2)$. The hypotheses are then formulated as

$$\begin{aligned} H_0 &: \mathbb{P}(r_i = r_0) = 1 \text{ for some } r_0 \in [\tau, 1-\tau] \\ H_1 &: \mathbb{P}(r_i = r_0) < 1 \text{ for any } r_0 \in [\tau, 1-\tau]. \end{aligned} \tag{12}$$

Note that the alternative hypothesis is very general. It covers the case with multiple thresholds that are the same for all i (cf. Bai and Perron (1998) in the structural break model) and the case with heterogeneous thresholds that vary across i. Moreover, r_i can be a function of some random variables z_i. Examples include an index form, $r_i = z_i^\intercal r$ for some parameter r, as in Yu and Fan (2019); and even a nonparametric from, $r_i = r(z_i)$ for some unknown function $r(\cdot)$, as in Lee and Wang (2019). This setup also covers the tipping point problem, where z_i includes some demographic characteristics of the ith neighborhood that affect the heterogeneous tipping points through some unknown function $r(\cdot)$.

We construct a test statistic for (12), which only requires estimating the model under the null hypothesis (i.e., the constant threshold regression model with (3)). To this end, we first obtain the sample analogue of $\mathcal{G}_v(s)$, denoted as $\widehat{\mathcal{G}}_{vn}(s)$, and study its asymptotic properties.

We first estimate $D(r)$ and $V(r)$ as[4]

$$\widehat{D}(r) = \frac{\sum_{i=1, i\neq \lfloor rn \rfloor}^{n} K\left(\frac{(i/n)-r}{b_n}\right) x_{[i]} x_{[i]}^\intercal}{\sum_{i=1, i\neq \lfloor rn \rfloor}^{n} K\left(\frac{(i/n)-r}{b_n}\right)} \tag{13}$$

[4]We can instead estimate them as

$$\widehat{D}(r) = \frac{1}{nb_n} \sum_{i\neq \lfloor rn \rfloor} K\left(\frac{(i/n)-r}{b_n}\right) x_{[i]} x_{[i]}^\intercal \quad \text{and} \quad \widehat{V}(r) = \frac{1}{nb_n} \sum_{i\neq \lfloor rn \rfloor} K\left(\frac{(i/n)-r}{b_n}\right) x_{[i]} x_{[i]}^\intercal \widehat{u}_{[i]}^2,$$

because (after multuplying $(nb_n)^{-1}$) the denominator in (13) or (14) converges to the pdf of $U[0,1]$ at r by construction and hence 1 in probability.

$$\widehat{V}(r) = \frac{\sum_{i=1, i \neq \lfloor rn \rfloor}^{n} K\left(\frac{(i/n)-r}{b_n}\right) x_{[i]} x_{[i]}^{\intercal} \widehat{u}_{[i]}^{2}}{\sum_{i=1, i \neq \lfloor rn \rfloor}^{n} K\left(\frac{(i/n)-r}{b_n}\right)} \tag{14}$$

for some kernel function $K(\cdot)$ and some bandwidth b_n, where $\widehat{u}_{[i]}$ denotes the re-ordered regression residual under the null hypothesis: $\widehat{u}_i = y_i - x_i^{\intercal}\widehat{\beta} - x_i^{\intercal}\widehat{\delta}\mathbf{1}[q_i \leq \widehat{\gamma}]$. We use the leave-one-out kernel. Given (13) and (14), functions in (8) are estimated by

$$\widehat{h}(r) = \frac{1}{n} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor rn \rfloor} \frac{1}{v^{\intercal} \widehat{D}(i/n)^{-1} \widehat{V}(i/n) (\widehat{D}(i/n)^{-1})^{\intercal} v} \quad \text{and} \quad \widehat{g}(r) = \frac{\widehat{h}(r)}{\widehat{h}(1-\tau)}. \tag{15}$$

Under the following conditions (e.g., Li and Racine (2007) and Yang (1981)), we can verify that all these kernel estimators are uniformly consistent.

Condition 2

1. *$K(\cdot)$ is Lipschitz continuous, continuously differentiable with bounded derivative, and symmetric around zero, which satisfies $\int K(t)\,dt = 1$, $\int tK(t)\,dt = 0$, $0 < \int t^2 K(t) dt < \infty$, $\lim_{t\to\infty} |t| K(t) = 0$, and $\lim_{t\to\infty} t^2(\partial K(t)/\partial t) = 0$.*

2. *$b_n \to 0$, $nb_n/\log n \to \infty$, and $n^{1/4} b_n \to \infty$ as $n \to \infty$.*

Second, since $\widehat{D}(\widehat{F}_n(q_{(i)})) = \widehat{D}(i/n)$ and $\widehat{V}(\widehat{F}_n(q_{(i)})) = \widehat{V}(i/n)$, we can construct the sample analogue of $\mathcal{G}_v$ in (9) as

$$\widehat{\mathcal{G}}_{vn}(s) = \frac{\widehat{h}_{\tau}^{1/2}}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor \widehat{g}^{-1}(s) n \rfloor} \widehat{g}^{(1)}(i/n) v^{\intercal} \widehat{D}(i/n)^{-1} x_{[i]} \widehat{u}_{[i]}, \tag{16}$$

where $\widehat{h}_{\tau}^{1/2} = \widehat{h}(1-\tau)^{1/2}$, $\widehat{g}^{-1}(\cdot)$ is computed as the numerical inverse of $\widehat{g}(\cdot)$, and $\widehat{g}^{(1)}(i/n) = 1/\{v^{\intercal}\widehat{D}(i/n)^{-1}\widehat{V}(i/n)(\widehat{D}(i/n)^{-1})^{\intercal} v \widehat{h}_{\tau}\}$. The following lemma establishes that $\widehat{\mathcal{G}}_{vn}(\cdot)$ weakly converges to $\mathcal{G}_v(\cdot)$ as in (11).

Lemma 1 *Suppose Conditions 1 and 2 hold. Then under the null hypothesis in (12),*
(i) *$\widehat{D}(r)$, $\widehat{V}(r)$, $\widehat{h}(r)$, and $\widehat{g}(r)$ are uniformly consistent on $[\tau, 1-\tau]$;*
(ii) *$\widehat{\mathcal{G}}_{vn}(\cdot) \Rightarrow \mathcal{G}_v(\cdot)$ as $n \to \infty$, where $\mathcal{G}_v(s)$ is given in (11) for $s \in [0,1]$.*

Lemma 1 implies that the transformed partial sum process $\widehat{\mathcal{G}}_{vn}(s)$ has a well-defined weak limit under the null hypothesis. It also shows that the testing problem essentially reduces to testing for additional changes either before or after $g(r_0)$. In order to use this result for the inference purpose, however, we need to eliminate the nuisance terms Φ_β^h and Φ_δ^h. It can be done by constructing some statistic that is invariant to location shift. In particular, we consider the following re-scaled and demeaned sample process

$$\widehat{\mathcal{G}}_n^*(s) = \begin{cases} \widehat{\mathcal{G}}_{1n}^*(s) & \text{if } s \leq \widehat{g}(\widehat{r}) \\ \widehat{\mathcal{G}}_{2n}^*(s) & \text{otherwise,} \end{cases} \tag{17}$$

where $\widehat{r} = \widehat{F}_n(\widehat{\gamma})$,

$$\begin{aligned} \widehat{\mathcal{G}}_{1n}^*(s) &= \frac{1}{\sqrt{\widehat{g}(\widehat{r})}} \left\{ \widehat{\mathcal{G}}_{vn}(s) - \frac{s}{\widehat{g}(\widehat{r})} \widehat{\mathcal{G}}_{vn}(\widehat{g}(\widehat{r})) \right\}, \\ \widehat{\mathcal{G}}_{2n}^*(s) &= \frac{1}{\sqrt{1-\widehat{g}(\widehat{r})}} \left\{ \widehat{\mathcal{G}}_{vn}(1) - \widehat{\mathcal{G}}_{vn}(s) - \frac{1-s}{1-\widehat{g}(\widehat{r})} \left(\widehat{\mathcal{G}}_{vn}(1) - \widehat{\mathcal{G}}_{vn}(\widehat{g}(\widehat{r})) \right) \right\}. \end{aligned}$$

By the continuous mapping theorem and the consistency of $\widehat{g}(\widehat{r})$ to $g(r_0)$, the Φ_β^h and Φ_δ^h terms are canceled out asymptotically so that the weak limits of $\widehat{\mathcal{G}}_{1n}^*(s)$ and $\widehat{\mathcal{G}}_{2n}^*(s)$ are free of nuisance terms. By construction, they behave as the standard Brownian bridges defined on $[0, 1]$ in the limit.

We now define the constant-threshold test statistic, or the CT test statistic, as

$$CT_n = \frac{1}{\lfloor \widehat{g}(\widehat{r}) n \rfloor} \sum_{i=1}^{\lfloor \widehat{g}(\widehat{r}) n \rfloor} \left\{ \widehat{\mathcal{G}}_n^*(i/n) \right\}^2 + \frac{1}{\lfloor n - \widehat{g}(\widehat{r}) n \rfloor} \sum_{i=\lfloor \widehat{g}(\widehat{r}) n \rfloor + 1}^{n} \left\{ \widehat{\mathcal{G}}_n^*(i/n) \right\}^2 \tag{18}$$

in a similar vein to Nyblom (1989) and Elliott and Müller (2007). Theorem 3 below establishes that CT_n converges to the integral of the squared Brownian bridges under the null hypothesis of a constant threshold but diverges under the alternative hypothesis.

Theorem 3 *Suppose Conditions 1 and 2 hold. Then as $n \to \infty$,*

$$CT_n \to_d \int_0^1 \mathcal{B}_2(t)^\intercal \mathcal{B}_2(t)\, dt \tag{19}$$

under the null hypothesis in (12), where $\mathcal{B}_2(t)$ is the 2×1 vector standard Brownian bridge on $[0, 1]$. In addition, $CT_n \to \infty$ under the alternative hypothesis in (12).

The limiting distribution of CT_n is pivotal under the null hypothesis of a constant threshold. Therefore, we can easily simulate the critical values, which are given in Table 1. The

Table 1: Simulated critical values of the CT test

$\mathbb{P}(\int_0^1 \mathcal{B}_2(t)^\intercal \mathcal{B}_2(t)\, dt > \text{cv})$	0.800	0.850	0.900	0.925	0.950	0.975	0.990
cv	0.467	0.527	0.608	0.666	0.744	0.888	1.066

Note: Entries are based on 50,000 replications and 5000 step approximations to the continuous time process.

test for (12) is then conducted as a one-sided test that rejects the null hypothesis if CT_n is larger than the corresponding critical values.

We conclude this subsection by summarizing the steps of implementing the CT test.

Step 1 Obtain the profile least squares estimators $\widehat{\theta}$ and $\widehat{\gamma}$.

Step 2 For each $r \in \{(\lfloor \tau n \rfloor + 1)/n, (\lfloor \tau n \rfloor + 2)/n, ..., \lfloor (1-\tau)n \rfloor /n\}$, obtain the kernel estimators $\widehat{D}(r)$ and $\widehat{V}(r)$ as in (13) and (14), and the estimators $\widehat{h}(r)$, $\widehat{g}(r)$, and $\widehat{g}^{(1)}(r)$ as in (15). Obtain $\widehat{g}^{-1}(\cdot)$ by numerically inverting $\widehat{g}(\cdot)$.

Step 3 Construct $\widehat{\mathcal{G}}_n^*(s)$ for $s \in \{1/n, 2/n, ..., 1\}$ as (17).

Step 4 Compute the CT_n statistic as in (18) and compare it with the critical values from Table 1.

3.2 Test for linear restrictions

With a minor modification to the previous section, we can develop a test for linear restrictions on the regression coefficients. We focus on inference about δ_0 for illustration, which covers the important question about the existence of the threshold.[5] Specifically, we consider the following hypotheses:

$$H_0 : \lambda^\intercal \delta_0 = 0 \quad \text{against} \quad H_1 : \lambda^\intercal \delta_0 \neq 0 \tag{20}$$

for some non-zero $k \times 1$ vector λ. For example, one can consider $\lambda = (1, 0, \ldots, 0)^\intercal$ for testing whether the first element of $\beta_i = \beta_0 + \delta_0 \mathbf{1}[F(q_i) \leq r_0]$ in (3) has a coefficient change, which is the case of the tipping point problem.

[5] Inference about β_0 can also be studied by combining the transformation idea and the test developed in Elliott and Müller (2014).

When γ_0 can be consistently estimated, inference about β_0 and δ_0 becomes straightforward since their least squares estimators based on $\widehat{\gamma}$ are still $\sqrt{n}$-consistent and asymptotically normal (e.g., Lemma A.12 in Hansen (2000a)). Therefore, we focus on the more challenging case where γ_0 cannot be consistently estimated. In particular, we consider a local alternative $\delta_0 = n^{-1/2}c_0$ (i.e., $\epsilon = 1/2$ in Condition 1.4) for some $c_0 \neq 0$, which is contiguous to the no-threshold case. This local alternative leads to non-degenerate asymptotic powers for the hypothesis testing problem (20), as similarly considered in Hansen (2000b), Elliott and Müller (2007), and Elliott, Müller, and Watson (2015).

Now we let $\widehat{u}_i$ be the residual by regressing y_i on x_i only and $v = \lambda/(\lambda^\intercal\lambda)$. Then, we can construct $\widehat{\mathcal{G}}_{vn}$ in (16) in the same way as described in the previous section. In particular, a similar (and even simpler) argument as Lemma 1 yields that $\widehat{\mathcal{G}}_{vn}(\cdot) \Rightarrow \mathcal{G}_v(\cdot)$ as $n \to \infty$, where

$$\mathcal{G}_v(s) =_d W_1(s) - sv^\intercal\Phi_\beta^h - \min\{s, g(r_0)\}v^\intercal c_0 h_\tau^{1/2}$$

for $s \in [0,1]$. In this case, the nuisance term $v^\intercal\Phi_\beta^h$ can be eliminated by constructing

$$\widehat{\mathcal{G}}_n^*(s) = \widehat{\mathcal{G}}_{vn}(s) - s\widehat{\mathcal{G}}_{vn}(1).$$

Then, by the continuous mapping theorem, we have $\widehat{\mathcal{G}}_n^*(\cdot) \Rightarrow \mathcal{G}_v^*(\cdot)$ as $n \to \infty$, where

$$\mathcal{G}_v^*(s) =_d (W_1(s) - sW_1(1)) - (\min\{s, g(r_0)\} - sg(r_0))\, v^\intercal c_0 h_\tau^{1/2} \tag{21}$$

for $s \in [0,1]$. Under the null hypothesis in (20), $v^\intercal c_0 h_\tau^{1/2} = 0$ and hence the right-hand-side of (21) reduces to the standard Brownian bridge.

By Girsanov's theorem, the Radon-Nikodym derivative of the distribution of $\mathcal{G}_v^*(s)$ relative to the distribution of the standard Brownian bridge $W_1(s) - sW_1(1)$, evaluated at $\mathcal{G}_v^*(s)$, is given by (e.g., Chapter 7 in Liptser and Shiryaev (2013))

$$lr(\mathcal{G}_v^*; g(r_0), v^\intercal c_0 h_\tau^{1/2}) = \exp\left(v^\intercal c_0 h_\tau^{1/2}\mathcal{G}_v^*(g(r_0)) - \frac{\left(v^\intercal c_0 h_\tau^{1/2}\right)^2}{2} g(r_0)(1 - g(r_0)) \right), \tag{22}$$

which yields the likelihood ratio. With two nuisance terms $r_g = g(r_0)$ and $c_v = v^\intercal c_0 h_\tau^{1/2}$ that appear only under the alternative hypothesis, we follow Andrews and Ploberger (1994) and Elliott, Müller, and Watson (2015) to construct a weighted likelihood-ratio test that maximizes the weighted average power criterion:

$$WLR = \int lr\left(\mathcal{G}_v^*; r_g, c_v\right) d\mu(r_g, c_v)$$

for some weight function $\mu(\cdot,\cdot)$ over the values of (r_g, c_v).

For an easy implementation, we choose $\mu(\cdot,\cdot)$ such that the test statistic has a closed-form expression. This can be done by choosing the uniform weight on r_g and the normal-density weight on c_v. Then, we can show that WLR can be written as an integrated form of $\mathcal{G}_v^*(s)^2/(s(1-s))$ as follows.

Lemma 2 *With the choice of $r_g \sim U[\tau, 1-\tau]$ and $c_v|(r_g = s) \sim \mathcal{N}\left(0, \omega^2/s(1-s)\right)$ for some $\omega^2 > 0$,*

$$\lim_{\omega^2 \to 0} \frac{2}{\omega^2}\left(\sqrt{1+\omega^2}WLR - 1\right) = \frac{1}{1-2\tau}\int_{\tau}^{1-\tau} \frac{\mathcal{G}_v^*(s)^2}{s(1-s)} ds. \tag{23}$$

Note that the limit expression in (23) coincides with the "average LR" statistic with a uniform weight in Andrews and Ploberger (1994), which can be obtained by combining equations (2.5) and (3.3) in their paper. Lemma 2 leads to the average $\widehat{\mathcal{G}}_n^*$ test statistic, namely the AG test, defined as

$$AG_n = \frac{1}{(1-2\tau)\,n} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor (1-\tau)n \rfloor} \frac{(\widehat{\mathcal{G}}_n^*\,(i/n))^2}{(i/n)(1-(i/n))}, \tag{24}$$

whose limiting distribution is the same as (23). This is established in the following theorem.

Theorem 4 *Suppose Conditions 1 and 2 hold with $\epsilon = 1/2$ in Condition 1.4. Then as $n \to \infty$,*

$$AG_n \to_d \frac{1}{1-2\tau}\int_{\tau}^{1-\tau} \frac{\mathcal{G}_v^*(s)^2}{s(1-s)} ds \tag{25}$$

where $\mathcal{G}_v^*(\cdot)$ is defined in (21).

Under the null hypothesis that $v^\intercal c_0 = 0$, $\mathcal{G}_v^*(s)$ reduces to $W_1(s) - sW_1(1)$, and hence the limiting distribution of AG_n is the same as the average LR test established by Andrews and Ploberger (1994). Then using the critical values tabulated in their Table II (pp. 1401-1402), the AG test controls size asymptotically.

As a remark, we heuristically discuss the asymptotic admissibility of the AG test. A formal study requires analyzing the higher-order approximation biases in the nonparametric estimation, which is beyond the scope of this paper. On the one hand, nonparametrically estimating $D(\cdot)$ and $V(\cdot)$ may cost efficiency; on the other hand, the fact that the transformation is one-to-one implies that the AG test also shares the optimality of the average LR test established by Andrews and Ploberger (1994). We investigate such ambiguity in Section 4 by Monte Carlo experiments. The results show that the AG test could be substantially more efficient than the average LR test with adjusted critical values, especially when $D(\cdot)$ and $V(\cdot)$ are highly nonlinear. Such a finding is close in spirit to the efficiency gain of the feasible generalized least squares (GLS) regression relative to the ordinary least squares with robust standard errors in the context of classical linear regression with heteroskedasticity.

4 Monte Carlo Experiments

4.1 The CT test

This section examines the small sample performance of the CT test in (18). We consider the following data generating processes (DGPs):

DGP CT-1 $y_i = x_i^\intercal \beta_0 + x_i^\intercal \delta_0 \mathbf{1}[q_i \leq 0] + u_i$;

DGP CT-2 $y_i = x_i^\intercal \beta_0 + x_i^\intercal \delta_0 \mathbf{1}[q_i \leq \sin(z_i)/2] + u_i$;

DGP CT-3 $y_i = x_i^\intercal \beta_0 + x_i^\intercal \delta_0 (\mathbf{1}[q_i \leq 0] + \mathbf{1}[q_i > 0.1]) + u_i$,

where $x_i = (x_{1i}, x_{2i})^\intercal \in \mathbb{R}^2$ with the first element $x_{1i} = 1$ and z_i is some scalar random variable specified later. We set $\beta_0 = 0\iota_2$ and consider $\delta_0 = \delta\iota_2$ for $\delta \in \{0.25, 0.50, 0.75, 1.00\}$, where $\iota_2 = (1,1)^\intercal$.

These DGPs correspond to each of the following three different threshold specifications: (i) one single threshold at 0; (ii) a functional threshold of $\sin(z_i)/2$ for some scalar random variable z_i; and (iii) two thresholds at 0 and 0.1. The first one corresponds to the null hypothesis of the homogeneous threshold in (12), while the other two are for the alternative hypothesis in (12). We set $v = (1,0)^\intercal$, and use the rule-of-thumb choice of the bandwidth $b_n = (1/12)^{1/2} n^{-1/5}$ and the Gaussian kernel. The truncation parameter τ is 0.1. Other choices of bandwidth, kernel, and τ are also implemented, which lead to negligible changes. The sample sizes are $n = 500$, 1000, and 1500, and the significance level is 5%. The results are based on 1000 simulations.

For comparison, we implement two existing methods. The first one is the $F(2|1)$ test proposed by Bai and Perron (1998), which is designed for testing one against two structural breaks. Note that this test is developed for the time-series case with (piecewise) stationary data only, which corresponds to the case that $V(\cdot)$ and $D(\cdot)$ are both constant matrices. To implement this test, one obtains the sum of squared residuals SSR_1 and SSR_2, which are from the change-point regression models with one and two breaks, respectively. The test statistic is then constructed as $F_n(2|1) = n(SSR_1 - SSR_2)/SSR_1$. We use their choice of the parameter $\varepsilon = 0.05n$, which is the minimum number of observations between the two breaks.

The second one is the model selection approach proposed by Gonzalo and Pitarakis (2002). Specifically, Gonzalo and Pitarakis (2002) introduce the following information criterion

$$IC_n(m) = \log SSR_m + \frac{\varphi_n}{n} k(m+1),$$

where m denotes the number of thresholds, SSR_m is the sum of squared residuals from the regression with m thresholds, and φ_n is some tuning parameter that satisfies $\varphi_n \to \infty$ and $\varphi_n / n \to 0$. The number of thresholds is determined by minimizing $IC_n(m)$ over m. To compare with the aforementioned tests for (12), we count the mis-selection probability when $m = 1$ as the rejection probability. We follow Gonzalo and Pitarakis (2002) to choose the BIC approach by setting $\varphi_n = \log n$ and $3 \log n$, denoted BIC1 and BIC3 respectively in Tables 2 and 3 below. The minimum number of observations between the two thresholds is also chosen as $0.05n$.

Table 2 reports the results under the i.i.d. case with $(q_i, z_i, u_i, x_i) \sim \mathcal{N}(0, I_4)$. Several findings can be summarized as follows. First, since q_i is independent of other variables, re-ordering the data leads to the canonical structural break model, in which time is deterministic. Thus both the CT and the $F(2|1)$ tests should control size under the null hypothesis, as illustrated in the first three columns. Second, the $F(2|1)$ test is very conservative while the CT test has approximately the correct size. The middle three columns show the (size-adjusted) powers under the smooth threshold alternative, where the CT test dominates the $F(2|1)$ test. Third, the last three columns show the powers under the alternative with two thresholds. This is the exact alternative that the $F(2|1)$ test is designed for, while our CT test still achieves comparable powers. Finally, the model selection based on BIC has good selection probabilities, especially when the change size is large. However, its performance is very sensitive to the choice of the tuning parameter as we compare the results for BIC1 and BIC3. In particular, BIC3 uses a larger tuning parameter (i.e., heavier penalty) than BIC1,

Table 2: Rejection probabilities with independent q

		DGP CT-1			DGP CT-2			DGP CT-3		
δ	$n=$	500	1000	1500	500	1000	1500	500	1000	1500
CT test										
0.25		0.06	0.05	0.05	0.05	0.05	0.10	0.08	0.08	0.11
0.50		0.06	0.05	0.05	0.14	0.41	0.66	0.14	0.19	0.30
0.75		0.07	0.05	0.05	0.43	0.82	0.97	0.22	0.39	0.55
1.00		0.08	0.05	0.05	0.68	0.95	1.00	0.31	0.55	0.72
F(2|1) test										
0.25		0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.03	0.04
0.50		0.01	0.01	0.01	0.02	0.19	0.41	0.08	0.22	0.37
0.75		0.01	0.01	0.02	0.18	0.67	0.90	0.24	0.54	0.68
1.00		0.00	0.01	0.01	0.43	0.92	1.00	0.44	0.73	0.88
BIC1										
0.25		0.24	0.04	0.01	0.34	0.08	0.03	0.04	0.02	0.03
0.50		0.05	0.03	0.02	0.11	0.27	0.50	0.14	0.28	0.41
0.75		0.07	0.03	0.03	0.44	0.82	0.96	0.43	0.72	0.89
1.00		0.06	0.04	0.03	0.78	0.99	1.00	0.71	0.93	0.98
BIC3										
0.25		0.97	0.74	0.34	0.99	0.94	0.76	0.04	0.00	0.00
0.50		0.04	0.00	0.00	0.32	0.01	0.00	0.00	0.00	0.00
0.75		0.00	0.00	0.00	0.00	0.01	0.08	0.00	0.06	0.17
1.00		0.00	0.00	0.00	0.00	0.17	0.47	0.09	0.36	0.61

Note: Entries are rejection probabilities under the null hypothesis in (12) of the CT test, the $F(2|1)$ test by Bai and Perron (1998), and the model selection using the BIC by Gonzalo and Pitarakis (2002), based on 1000 simulations. The significance level is 5%. Data are generated from DGPs CT-1 to CT-3 with $(q_i, z_i, u_i, x_i) \sim iid\mathcal{N}(0, I_4)$. The first three columns are based on $\gamma_0(s) = 0$; the middle three are based on $\gamma_0(s) = \sin(s)/2$; the third three are based on two thresholds at 0 and 0.1.

Table 3: Rejection probabilities with dependent q

δ	$n=$	CT test			F(2|1) test			F(2|1)-Boot.		
		500	1000	1500	500	1000	1500	500	1000	1500
0.25		0.07	0.06	0.05	0.09	0.12	0.14	0.21	0.26	0.24
0.50		0.07	0.05	0.06	0.08	0.11	0.14	0.23	0.22	0.25
0.75		0.08	0.07	0.06	0.08	0.10	0.12	0.24	0.25	0.26
1.00		0.09	0.07	0.07	0.09	0.12	0.14	0.20	0.24	0.26
					BIC1			BIC3		
δ	$n=$				500	1000	1500	500	1000	1500
0.25					0.62	0.38	0.26	0.99	0.96	0.89
0.50					0.27	0.26	0.23	0.59	0.06	0.00
0.75					0.29	0.24	0.21	0.02	0.00	0.00
1.00					0.30	0.27	0.25	0.00	0.00	0.00

Note: Entries are rejection probabilities under the null hypothesis in (12) of the CT test, the $F(2|1)$ test by Bai and Perron (1998), the $F(2|1)$ test with bootstrap critical values from 100 bootstrap samples, and the model selection using the BIC by Gonzalo and Pitarakis (2002). The results are based on 1000 simulations. The significance level is 5%. Data are generated from DGP CT-1 with $(q_i, z_i) \sim iid\mathcal{N}(0, I_2)$, $x_i | (q_i, z_i) = (q, s) \sim iid\mathcal{N}\left(0, 1/(1+s^2+q^2)\right)$, and $u_i|x_i = x \sim iid\mathcal{N}(0, 1+x^2)$.

which leads to substantially lower powers as BIC3 always chooses one threshold even if the true number of thresholds is more. This feature is also seen in Table 3.

In Table 3, we introduce some correlation between q_i and (x_i, u_i) and investigate the size properties of these three tests. The powers are not presented since only the CT test controls size. In particular, we generate data under the null hypothesis of a single threshold at 0 and use $(q_i, z_i) \sim iid\mathcal{N}(0, I_2)$, $x_i | (q_i, z_i) = (q, s) \sim iid\mathcal{N}(0, 1/(1+s^2+q^2))$ and $u_i|x_i = x \sim iid\mathcal{N}(0, 1+x^2)$. Several findings can be summarized as follows. First, as expected, the $F(2|1)$ test fails to control size since its asymptotic distribution is contaminated by the rank-varying moments. Second, as a remedy, Hansen (1997) suggests using the original test statistics with bootstrap critical values. However, bootstrap is not expected to perform well in this case since the $F(2|1)$ test statistics is not pivotal. This is verified in the top last three columns, where the results are based on 100 bootstrap samples and the residuals from the null model with one single threshold. Third, the CT test performs well in terms of controlling size if the sample size and the break size are large enough, while the $F(2|1)$ test fails to control size with either the original or the bootstrap critical values. Finally, mis-selection probabilities from the BIC are far from 5% because the strong correlation between

q_i and x_i and the conditional heteroskedasticity are difficult to distinguish from the potential coefficient changes. This issue can be alleviated by choosing a larger tuning parameter as in BIC3, which again leads to severe under-rejections under the alternative.

4.2 The AG test

This section examines the small sample performance of the AG test in (24). We consider the following DGPs:

$$y_i = x_i^{\intercal}\beta_0 + x_i^{\intercal}\delta_0\mathbf{1}\left[q_i \leq 0\right] + u_i \tag{26}$$

where $x_i = (1, x_{2i})^{\intercal}$, $(x_{2i}, q_i) \sim iid\mathcal{N}(0\iota_2, (1, 0.5; 0.5, 1))$ and $u_i|q_i = q \sim iid\mathcal{N}\left(0, \sigma_u^2(q)\right)$ with $\sigma_u(q)$ given by

DGP AG-1 $\sigma_u(q) = 1 + |q|^0$;

DGP AG-2 $\sigma_u(q) = 1 + |q|^1$;

DGP AG-3 $\sigma_u(q) = 1 + |q|^2$;

DGP AG-4 $\sigma_u(q) = 1 + |q|^3$.

In these specifications, the effect of q_i on $V(\cdot)$ gets more substantial as the power of $|q_i|$ gets higher. We set $\beta_0 = (\beta_{01}, \beta_{02})^{\intercal} = 0\iota_2$ and consider $\delta_0 = (\delta_{01}, \delta_{02})^{\intercal} = \delta\iota_2$ for $\delta \in \{0.00, 0.25, 0.50\}$. We choose the same v, τ, b_n, and the Gaussian kernel as in the previous experiment. The sample sizes are $n = 500$, 1000, and 1500, and the significance level is 5%. The results are based on 1000 simulations.

We are interested in testing whether the intercept term has a coefficient change (i.e., $\delta_{01} = 0$ or not), as motivated by the tipping point application. We implement the AG test in (24) with the simulated critical values in Table 1. As a comparison, we also consider the average LR test developed in Andrews and Ploberger (1994) and the sup LR test developed in Andrews (1993), which are respectively given by

$$\frac{1}{1-2\tau}\int_{\tau}^{1-\tau}\mathcal{F}_n(s)\,ds \quad \text{and} \quad \sup_{s\in[\tau,1-\tau]}\mathcal{F}_n(s)\,, \tag{27}$$

where $\mathcal{F}_n(s) = n\left(SSR_0 - SSR(s)\right)/SSR(s)$ denotes the Chow-test statistic given the threshold s with SSR_0 and $SSR(s)$ being the restricted and unrestricted sums of squared residuals, respectively (p. 582 in Hansen (2000a)). In particular, we first re-order the data

Table 4: Rejection probabilities of the AG test and the average F test

		DGP AG-1			DGP AG-2			DGP AG-3			DGP AG-4		
δ	$n=$	500	1000	1500	500	1000	1500	500	1000	1500	500	1000	1500
AG test													
0.00		0.07	0.05	0.05	0.06	0.05	0.04	0.04	0.04	0.03	0.03	0.03	0.02
0.25		0.20	0.33	0.45	0.30	0.51	0.70	0.34	0.61	0.79	0.32	0.62	0.82
0.50		0.57	0.85	0.96	0.81	0.98	1.00	0.88	0.99	1.00	0.89	1.00	1.00
bootstrap average LR test													
0.00		0.07	0.07	0.07	0.07	0.06	0.06	0.07	0.06	0.06	0.06	0.06	0.06
0.25		0.28	0.50	0.66	0.16	0.29	0.46	0.10	0.12	0.14	0.08	0.07	0.07
0.50		0.82	0.99	1.00	0.64	0.95	1.00	0.24	0.57	0.84	0.10	0.12	0.13
bootstrap sup LR test													
0.00		0.07	0.06	0.06	0.08	0.06	0.06	0.07	0.05	0.06	0.06	0.06	0.06
0.25		0.29	0.53	0.71	0.22	0.38	0.56	0.12	0.16	0.22	0.08	0.08	0.07
0.50		0.84	0.99	1.00	0.72	0.97	1.00	0.34	0.65	0.88	0.12	0.15	0.20

Note: Entries are finite sample rejection probabilities of the *AG* test and the average F test with bootstrap critical values. Data are generated from (26). See the main text for description of four DGPs and two tests. The significance level is 5%. Based on 1000 simulations.

according to q_i and treat the rank of q_i as time. Then, we construct the average LR and the sup LR test statistics and apply the fixed-bootstrap algorithm given by Hansen (2000b) to adjust the critical value.

Table 4 presents the small sample rejection probabilities of the *AG* test, the average LR test, and the sup LR test in (27). They all have approximately correct size under the null hypothesis ($\delta_{01} = 0$) and reasonable powers under the alternative ($\delta_{01} = 0.25$ and 0.50), which increase in the sample size and the magnitude of δ_{01}. However, a comparison among different DGPs exhibits a sharp difference in the efficiency among these tests. First, in DGP AG-1, q_i is independent of u_i and has only mild correlation with x_i. This feature implies that DGP AG-1 is very close to the classic structural break model with piecewise stationary data. Therefore, the bootstrap critical values are almost identical to the original ones tabulated in Table II of Andrews and Ploberger (1994), and hence the bootstrap tests are almost efficient. In comparison, the nonparametric estimation in the *AG* test suffers from some efficiency loss.

Second, in DGP AG-2, q_i enters the standard deviation of u_i linearly, which introduces nonlinearity to $D(\cdot)$ and $V(\cdot)$. Now the bootstrap critical values start to deviate from the original ones, which results in substantial efficiency loss. In contrast, the transformation method substantially outperforms the bootstrap ones. Finally, the relative performance of

Table 5: Tipping point estimation and testing results (1980-1990)

City	n	$\widehat{\gamma}$	AG p-value	CT p-value
Chicago	688	6.94	0.000	0.000
Los Angeles	1263	17.47	0.000	0.000
New York	315	16.08	0.000	0.000
Washington D.C.	719	15.54	0.000	0.000

Note: Entries are sample sizes (n), the constant tipping point estimation ($\widehat{\gamma}$), and the p-values of the AG test (24) and the CT test (18). Data are available from Card, Mas, and Rothstein (2008).

the transformation grows profoundly better in the nonlinearity of $\sigma_u(\cdot)$. In particular, the *AG* test dominates the bootstrap ones by approximately 80% more powers when the sample size is 1500 in DGP AG-4, which is quite remarkable.

5 Application: Tipping Point and Social Segregation

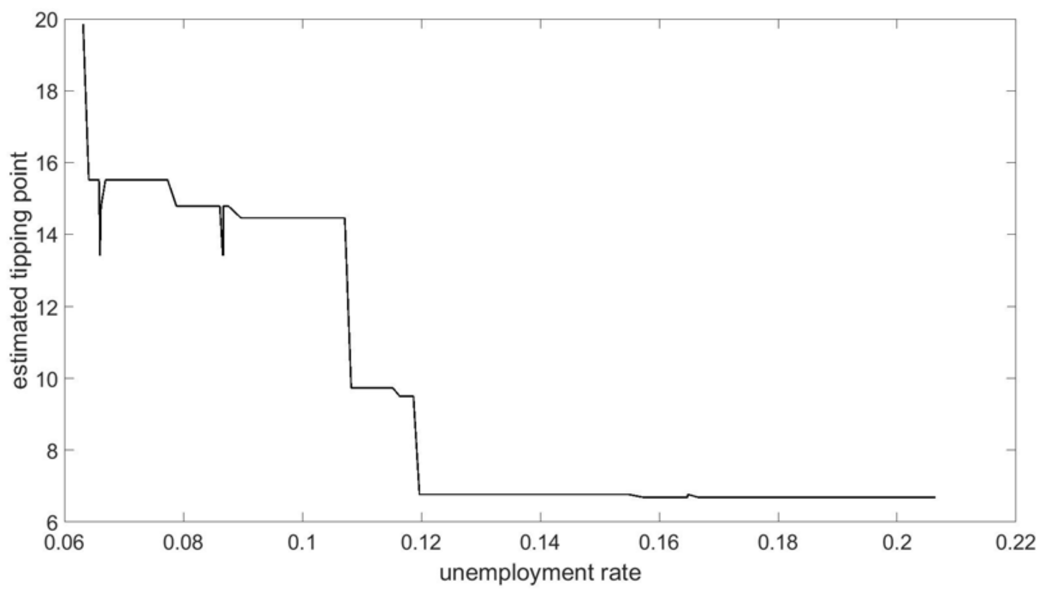
Our motivating example is on social segregation and the tipping point phenomenon. Card, Mas, and Rothstein (2008) empirically examine the theory proposed by Schelling (1971) that the white population substantially decreases once the minority share in a tract exceeds a certain threshold, called the tipping point. In particular, they consider the following threshold regression model:

$$y_i = \beta_{01} + \delta_{01}\mathbf{1}\left[q_i > \gamma_0\right] + x_i^\intercal \beta_{02} + u_i,$$

where for tract i in a certain city, q_i denotes the minority share in percentage at the beginning of a certain decade, y_i the normalized white population change in percentage within this decade, and x_i includes six tract-level control variables: unemployment rate, the logarithm 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. The data are collected from a variety of cities in three periods: 1970-1980, 1980-1990, and 1990-2000. They apply the least squares method to estimate the tipping point γ_0. For most cities and all three periods, they find that white population flows exhibit the tipping-like behavior, with the estimated tipping points ranging approximately from 5% to 20% across cities.

We revisit this problem by first testing for $\delta_{01} = 0$ using the *AG* test in (24). We choose

Figure 1: Estimated tipping point in Chicago, 1980-1990



Note: The figure depicts the point estimate of the tipping point as a function of the tract-level unemployment rate, using the method proposed by Lee and Wang (2019) and the data in Chicago in 1980-1990. Data are available from Card, Mas, and Rothstein (2008).

the rule-of-thumb bandwidth $b_n = (1/12)^{1/2} n^{-1/5}$ and the truncation parameter $\tau = 0.1$ as in the Monte Carlo experiments. We also follow Card, Mas, and Rothstein (2008) to use the tracts in which the initial minority share is between 5% and 60%. As an illustration, Table 5 shows the p-values of the AG test with the data in Chicago, Los Angeles, New York City, and Washington D.C. in the decade 1980-1990. These small p-values reinforce the existing founding that the tipping point feature is statistically significant. See also Lee, Seo, and Shin (2011) for another test based on a sup-likelihood-ratio type statistic, which gives the same conclusion.

Next, we examine the hypothesis that the tipping point remains constant across different tracts. Intuitively, such a null hypothesis can be easily rejected since some social characteristics endogenously determine the tipping points. In particular, Card, Mas, and Rothstein (2008) construct an index that measures white people's attitude against the minority and find that the level of the tipping point strongly depends on this index. To formalize such finding, we consider the model:

$$y_i = \beta_{01} + \delta_{01}\mathbf{1}\left[q_i > \gamma_0(z_i)\right] + x_i^{\intercal}\beta_{02} + u_i,$$

where $\gamma_0(\cdot)$ denotes an unknown tipping point function, and z_i denotes the attitude index. We are interested in testing if the tipping point remains constant across tracts. By treating $\gamma_i = \gamma_0\left(z_i\right)$, the testing problem is then equivalent to (12).

Table 5 shows the results of the CT test in (18) with the same city/decade and tuning parameter choice as above. The small p-values suggest that a single constant threshold is insufficient for fully capturing the social segregation behavior. Data from other cities and decades lead to similar results, which are hence not reported. To have a rough sense of how the tipping point changes, we use the unemployment rate as z_i and nonparametrically estimate the function $\gamma_0\left(\cdot\right)$ using the method proposed by Lee and Wang (2019). Figure 1 shows that the tipping point decreases substantially in the unemployment rate.

6 Conclusion

Threshold models have broad applications in economics. This paper develops a new framework that recasts the cross-sectional threshold problem into the time-series structural break analogue. Using this new framework, we develop two tests empirically motivated by the tipping point problem: a test for homogeneity of the threshold parameter and a test for linear restrictions on the regression coefficients.

Though we focus on these two tests in this paper, we can apply the same approach to develop other types of tests. In particular, our framework allows other inference methods developed in the structural break models to be converted into the threshold model setup, including inference about γ_0 (e.g., Elliott, Müller, and Watson (2015)) and inference about β_0 (e.g., Elliott and Müller (2014)). Moreover, for the CT test, since its alternative hypothesis is unspecified, we can modify it for more general cases as long as we can consistently estimate the null model. For instance, the test can be generalized to test for the null hypothesis of any fixed number of thresholds against additional thresholds.

Appendix: Proofs

We first establish the convergence of the key partial sum processes. Let C denote a generic constant.

Lemma A.1 *Suppose Condition 1 holds. Then, as $n \to \infty$*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} x_{[i]} u_{[i]} \Rightarrow \int_0^s V(t)^{1/2} dW_k(t)$$

for $s \in [0,1]$ and

$$\sup_{s \in [0,1]} \left\| \frac{1}{n} \sum_{i=1}^{\lfloor sn \rfloor} x_{[i]} x_{[i]}^{\intercal} - \int_0^s D(t)\, dt \right\| \to_p 0,$$

where $W_k(\cdot)$ is the $k \times 1$ vector standard Wiener process defined on $[0,1]$.

Proof of Lemma A.1 We prove the first result using Theorem 2 in Bhattacharya (1974). By the Cramér-Wold device, it suffices to show for any $k \times 1$ non-zero vector v,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} v^{\intercal} x_{[i]} u_{[i]} \Rightarrow \int_0^s \left(v^{\intercal} V(t) v\right)^{1/2} dW_1(t). \tag{A.1}$$

Note that $v^{\intercal} x_{[i]} u_{[i]}$ is a scalar random variable and is the induced order statistics of $v^{\intercal} x_i u_i$ associated with q_i. We now check Conditions 1 to 3 in Bhattacharya (1974). Condition 1 requires q_i to be continuous, which is implied by our Condition 1.3. For Condition 2, our Conditions 1.2 and 1.8 imply that $\mathbb{E}[v^{\intercal} x_i u_i | q_i] = 0$ a.s. and

$$\sup_{q \in \mathbb{R}} \mathbb{E}\left[\left(v^{\intercal} x_i u_i\right)^4 | q_i = q\right] \leq C \sup_{q \in \mathbb{R}} \mathbb{E}\left[\|x_i u_i\|^4 | q_i = q\right] < \infty.$$

Condition 3 is directly implied by our Condition 1.6. In particular, the continuous differentiability of $V(\cdot)$ implies that the function $v^{\intercal} V(\cdot) v$ is of bounded variation. Define

$$\phi_V(s) = \int_0^s v^{\intercal} V(t) v dt.$$

By Theorem 2 in Bhattacharya (1974), we have

$$\left(n \phi_V(1)\right)^{-1/2} \sum_{i=1}^{\lfloor sn \rfloor} v^{\intercal} x_{[i]} u_{[i]} \Rightarrow W_1\left(\frac{\phi_V(s)}{\phi_V(1)}\right). \tag{A.2}$$

Then (A.1) follows from the continuous mapping theorem and the fact that

$$\phi_V(1)^{1/2} W_1 \left(\frac{\phi_V(s)}{\phi_V(1)} \right) =_d \int_0^s \phi_V(t)^{1/2} dW_1(s).$$

For the second result, we let $\xi_i = v^\intercal x_i x_i^\intercal v$ and denote $\xi_{[i]}$ as the induced order statistics of ξ_i associated with $q_{(i)}$. Define the processes

$$\phi_{nD}(s) = \int_{-\infty}^{\widehat{F}_n^{-1}(s)} \mathbb{E}[\xi_i | q_i = q] d\widehat{F}_n(q)$$

where $\widehat{F}_n(\cdot)$ is the empirical distribution of q_i, and

$$\phi_D(s) = \int_{-\infty}^{F^{-1}(s)} \mathbb{E}[\xi_i | q_i = q] dF(q).$$

Conditions 1.6 and 1.8 imply that $\sup_{q\in\mathbb{R}} \mathbb{E}[\xi_i | q_i = q] < \infty$ and $\mathbb{E}[\xi_i | q_i = q]$ is of bounded variation. Therefore, $\sup_{s\in[0,1]} |\phi_{nD}(s) - \phi_D(s)| \to 0$ almost surely by integration by parts and application of the Glivenko-Cantelli theorem (e.g., Lemma 2 in Bhattacharya (1974)). By the triangular inequality, it will suffice to show $\sup_{s\in[0,1]} \left| n^{-1} \sum_{i=1}^{\lfloor sn \rfloor} \xi_{[i]} - \phi_{nD}(s) \right| \to_p 0$, which is done in a way analogous to (A.2) (e.g., p. 1038 in Bhattacharya (1974)). The desired result follows by the Cramér-Wold device. ■

Proof of Theorem 1 First note that

$$\begin{aligned}
\widehat{G}_n(s) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} x_{[i]} u_{[i]} \\
&\quad - \frac{1}{n} \sum_{i=1}^{\lfloor sn \rfloor} x_{[i]} x_{[i]}^\intercal \left\{ \sqrt{n} \left(\widehat{\beta} - \beta_0 \right) + \mathbf{1} \left[F(q_{(i)}) \le r_0 \right] \sqrt{n} \left(\widehat{\delta} - \delta_0 \right) \right\} \\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} x_{[i]} x_{[i]}^\intercal \widehat{\delta} \left\{ \mathbf{1} \left[\widehat{F}_n(q_{(i)}) \le \widehat{r} \right] - \mathbf{1} \left[F(q_{(i)}) \le r_0 \right] \right\} \\
&\equiv \widehat{G}_{n1}(s) - \widehat{G}_{n2}(s) - \widehat{G}_{n3}(s),
\end{aligned}$$

where the continuous mapping theorem yields

$$\begin{aligned}
\widehat{G}_{n1}(s) &\Rightarrow \int_0^s V(t)^{1/2} dW_k(t) \\
\widehat{G}_{n2}(s) &\Rightarrow \left(\int_0^s D(t)\, dt \right) \Phi_\beta - \left(\int_0^{\min\{s, r_0\}} D(t)\, dt \right) \Phi_\delta
\end{aligned}$$

from Lemma A.1 and (5). For the last term, we write

$$
\begin{aligned}
\widehat{G}_{n3}(s) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i x_i^{\intercal}\widehat{\delta}\left\{\mathbf{1}\left[q_i \le \gamma_0\right] - \mathbf{1}\left[q_i \le \widehat{\gamma}\right]\right\}\mathbf{1}[q_i \le q_{(\lfloor sn \rfloor)}] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i x_i^{\intercal}\delta_0\left\{\mathbf{1}\left[q_i \le \gamma_0\right] - \mathbf{1}\left[q_i \le \widehat{\gamma}\right]\right\}\mathbf{1}[q_i \le q_{(\lfloor sn \rfloor)}] \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i x_i^{\intercal}\left(\widehat{\delta} - \delta_0\right)\left\{\mathbf{1}\left[q_i \le \gamma_0\right] - \mathbf{1}\left[q_i \le \widehat{\gamma}\right]\right\}\mathbf{1}\left[q_i \le q_{(\lfloor sn \rfloor)}\right] \\
&\equiv \widehat{G}_{n31}(s) + \widehat{G}_{n32}(s).
\end{aligned}
$$

Let $E_{\gamma n}$ be the event that $\widehat{\gamma} \in \mathcal{B}_{Cn^{-1+2\epsilon}}(\gamma_0)$ for some $C > 0$, where $\mathcal{B}_r(x)$ denotes a generic open ball centered at x with radius r. Lemma A.12 in Hansen (2000a) yields that $\mathbb{P}\left(E_{\gamma n}^c\right) \le \varepsilon$ for any $\varepsilon > 0$ if n is sufficiently large. Then for any $\eta > 0$ and any $\varepsilon > 0$, if n is sufficiently large,

$$
\begin{aligned}
\mathbb{P}\left(\sup_{s\in[0,1]}\left\|\widehat{G}_{n31}(s)\right\| > \eta\right) &\le \mathbb{P}\left(\left\{\sup_{s\in[0,1]}\left\|\widehat{G}_{n31}(s)\right\| > \eta\right\}\cap E_{\gamma n}\right) + \mathbb{P}\left(E_{\gamma n}^c\right) \\
&\le \eta^{-1}n^{1/2}\mathbb{E}\left[\left\|x_i x_i^{\intercal}\delta_0\left\{\mathbf{1}\left[q_i \le \gamma_0\right] - \mathbf{1}\left[q_i \le \widehat{\gamma}\right]\right\}\right\|\mathbf{1}\left[E_{\gamma n}\right]\right] + \varepsilon \\
&\le \eta^{-1}n^{1/2-\epsilon}Cn^{-1+2\epsilon} + \varepsilon \\
&\le \varepsilon,
\end{aligned}
$$

where the second inequality is by Markov's inequality; the third inequality is by Conditions 1.4 with $\epsilon \in (0, 1/2)$, 1.7, and 1.8. Using a similar argument, we can also show that $\mathbb{P}\left(\sup_{s\in[0,1]}||\widehat{G}_{n32}(s)|| > \eta\right) \le \varepsilon$. It follows that

$$
\sup_{s\in[0,1]}\left\|\widehat{G}_{n3}(s)\right\| = o_p(1), \tag{A.3}
$$

and hence the desired result is obtained. ■

Proof of Theorem 2 Substituting the definition of $G(\cdot)$ yields that

$$
\begin{aligned}
\mathcal{G}_v(s) &= h_\tau^{1/2}\int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t)v^{\intercal}D(t)^{-1}V(t)^{1/2}\,dW_k(t) \\
&\quad - h_\tau^{1/2}\int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t)dt \times v^{\intercal}\Phi_\beta \\
&\quad - h_\tau^{1/2}\int_{g^{-1}(0)}^{\min\{g^{-1}(s), r_0\}} g^{(1)}(t)dt \times v^{\intercal}\Phi_\delta
\end{aligned}
$$

$$\equiv \quad H_1(s) - H_2(s) - H_3(s). \tag{A.4}$$

First, we show $H_1(s) =_d W_1(s)$. Since both terms are mean-zero Gaussian processes with independent increments, it suffices to show that they have the same variance function, which can be verified as

$$\begin{aligned}
&\int_{g^{-1}(0)}^{g^{-1}(s)} h_\tau \left(g^{(1)}(t)\right)^2 v^\intercal D(t)^{-1} V(t) D(t) v dt \\
&= \int_{g^{-1}(0)}^{g^{-1}(s)} \frac{h_\tau}{(v^\intercal D(t)^{-1} V(t) D(t) v h_\tau)^2} v^\intercal D(t)^{-1} V(t) D(t) v dt \\
&= \int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t) dt \\
&= g\left(g^{-1}(s)\right) - g(g^{-1}(0)) \\
&= s
\end{aligned}$$

for any $s \in [0,1]$. For $H_3(s)$, we have

$$\begin{aligned}
\int_{g^{-1}(0)}^{\min\{g^{-1}(s), r_0\}} g^{(1)}(t) dt &= g(\min\{g^{-1}(s), g^{-1}(g(r_0))\}) - g(g^{-1}(0)) \\
&= \min\{s, g(r_0)\}
\end{aligned}$$

for any $s \in [0,1]$ and hence $H_3(s) =_d \min\{s, g(r_0)\} v^\intercal \Phi_\delta h_\tau^{1/2}$. By the same argument, we have $H_2(s) =_d s v^\intercal \Phi_\beta h_\tau^{1/2}$ as desired. ■

Lemma A.2 *Let*

$$\widehat{V}^0(r) = \frac{\sum_{i=1, i \neq \lfloor rn \rfloor}^n x_{[i]} x_{[i]}^\intercal u_{[i]}^2 K_i(r)}{\sum_{i=1, i \neq \lfloor rn \rfloor}^n K_i(r)},$$

where $K_i(r) = b_n^{-1} K(((i/n) - r)/b_n)$. Under Conditions 1 and 2, $\sup_{r \in [\tau, 1-\tau]} ||\widehat{V}(r) - \widehat{V}^0(r)|| = o_p(1)$.

Proof of Lemma A.2 For expositional simplicity, we only present the case with scalar x_i. Note that

$$\widehat{V}(r) - \widehat{V}^0(r) = \frac{(1/n) \sum_{i=1, i \neq \lfloor rn \rfloor}^n \left(x_{[i]}^2 \widehat{u}_{[i]}^2 - x_{[i]}^2 u_{[i]}^2\right) K_i(r)}{(1/n) \sum_{i=1, i \neq \lfloor rn \rfloor}^n K_i(r)},$$

where the denominator converges to 1 in probability as $n \to \infty$ for any $r \in [\tau, 1-\tau]$ from Condition 2.1. For the numerator, as $\widehat{u}_i = u_i - x_i(\widehat{\beta} - \beta_0) - x_i(\widehat{\delta} - \delta_0)\mathbf{1}[q_i \leq \gamma_0] - x_i \widehat{\delta}(\mathbf{1}[q_i \leq \widehat{\gamma}] - \mathbf{1}[q_i \leq \gamma_0])$, we have

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1,i\neq\lfloor rn\rfloor}^{n} x_{[i]}^{2}\left(\widehat{u}_{[i]}+u_{[i]}\right)\left(\widehat{u}_{[i]}-u_{[i]}\right)K_{i}\left(r\right)\right| \qquad \text{(A.5)}\\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left|x_{i}^{3}\left(\widehat{u}_{i}+u_{i}\right)(\widehat{\beta}-\beta_{0})K_{i}\left(r\right)\right|\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\left|x_{i}^{3}\left(\widehat{u}_{i}+u_{i}\right)(\widehat{\delta}-\delta_{0})\mathbf{1}\left[q_{i}\leq\gamma_{0}\right]K_{i}\left(r\right)\right|\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\left|x_{i}^{3}\left(\widehat{u}_{i}+u_{i}\right)\widehat{\delta}\left\{\left(\mathbf{1}\left[q_{i}\leq\widehat{\gamma}\right]-\mathbf{1}\left[q_{i}\leq\gamma_{0}\right]\right)\right\}K_{i}\left(r\right)\right|\\
&\equiv M_{1n}(r)+M_{2n}(r)+M_{3n}(r).
\end{aligned}
$$

Let $E_{\theta n}$ be the event that $\widehat{\theta}=(\widehat{\beta}^{\intercal},\widehat{\delta}^{\intercal})^{\intercal}\in\mathcal{B}_{Cn^{-1/2}}(\theta_0)$ and $E_{\gamma n}$ the event that $\widehat{\gamma}\in\mathcal{B}_{Cn^{-1+2\epsilon}}(\gamma_0)$ for some C. Lemma A.12 in Hansen (2000a) implies $\mathbb{P}(E_{\theta n}^{c})\leq\varepsilon$ and $\mathbb{P}\left(E_{\gamma n}^{c}\right)\leq\varepsilon$ for any $\varepsilon>0$ if C and n are large enough. Then for any $\eta>0$,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|M_{1n}(r)\right|>\eta\right)\\
&\leq\mathbb{P}\left(\left\{\sup_{r\in[\tau,1-\tau]}\left|M_{1n}(r)\right|>\eta\right\}\cap E_{\gamma n}\cap E_{\theta n}\right)+\mathbb{P}(E_{\gamma n}^{c}\cup E_{\theta n}^{c})\\
&\leq\eta^{-1}\max_{1\leq i\leq n}\sup_{r\in[0,1]}K_{i}\left(r\right)\times\mathbb{E}\left[\left|x_{i}^{3}\left(\widehat{u}_{i}+u_{i}\right)(\widehat{\beta}-\beta_{0})\right|\left|E_{\theta n}\right.\right]+2\varepsilon\\
&\leq\eta^{-1}\max_{1\leq i\leq n}\sup_{r\in[0,1]}K_{i}\left(r\right)\times\Big\{2\mathbb{E}\left[\left|x_{i}^{3}u_{i}(\widehat{\beta}-\beta_{0})\right|\left|E_{\theta n}\right.\right]+\mathbb{E}\left[\left|x_{i}^{4}(\widehat{\beta}-\beta_{0})^{2}\right|\left|E_{\theta n}\right.\right]\\
&\quad+\mathbb{E}\left[\left|x_{i}^{4}\mathbf{1}\left[q_{i}\leq\gamma_{0}\right](\widehat{\delta}-\delta_{0})(\widehat{\beta}-\beta_{0})^{2}\right|\left|E_{\theta n}\right.\right]\\
&\quad+\mathbb{E}\left[\left|x_{i}^{4}\widehat{\delta}(\widehat{\beta}-\beta_{0})\left(\mathbf{1}\left[q_{i}\leq\widehat{\gamma}\right]-\mathbf{1}\left[q_{i}\leq\gamma_{0}\right]\right)\right|\left|E_{\gamma n}\cap E_{\theta n}\right.\right]\Big\}+2\varepsilon\\
&\leq C\eta^{-1}n^{-1/2}b_{n}^{-1}\left(2\mathbb{E}\left[\left|x_{i}^{3}u_{i}\right|\right]+\mathbb{E}\left[\left|x_{i}^{4}\right|\right]\right)+2\varepsilon\\
&\leq 3\varepsilon
\end{aligned}
$$

for sufficiently large n, where the second inequality is from Markov's inequality; the third inequality follows from the triangular inequality; the fourth inequality follows from Condition 2.1 and the fact that $\mathbf{1}\left[\cdot\right]\leq 1$; and the last inequality follows from Conditions 1.8 and 2.2. For $M_{2n}(r)$ and $M_{3n}(r)$, the same argument yields that $\sup_{r\in[\tau,1-\tau]}\left|M_{2n}(r)\right|=o_p(1)$ and $\sup_{r\in[\tau,1-\tau]}\left|M_{3n}(r)\right|=o_p(1)$ as well because $\widehat{\delta}=O_p(n^{-\epsilon})=o_p(1)$. Hence, the desired result follows. ■

Lemma A.3 *Suppose Conditions 1 and 2 hold. Then under the null hypothesis in (12),* $\sup_{r\in[\tau,1-\tau]}||\widehat{D}(r)-D(r)||=o_p(1)$, $\sup_{r\in[\tau,1-\tau]}||\widehat{V}(r)-V(r)||=o_p(1)$, $\sup_{r\in[\tau,1-\tau]}|\widehat{h}(r)-h(r)|=o_p(1)$, *and* $\sup_{r\in[\tau,1-\tau]}|\widehat{g}(r)-g(r)|=o_p(1)$.

Proof of Lemma A.3 We first prove the uniform consistency of $\widehat{V}(r)$, and the uniform consistency of $\widehat{D}(r)$ follows in the same way. By Lemma A.2, it suffices to show $\sup_{r\in[\tau,1-\tau]}||\widehat{V}^0(r)-V(r)||=o_p(1)$. For expositional simplicity, we only present the case with scalar x_i. We denote

$$\begin{aligned}\widehat{V}^0(r) &= \frac{n^{-1}\sum_{i=1,i\neq\lfloor rn\rfloor}^n x_{[i]}^2u_{[i]}^2K_i(r)}{n^{-1}\sum_{i=1,i\neq\lfloor rn\rfloor}^n K_i(r)}\equiv\frac{\widehat{T}_n(r)}{\widehat{f}_{n,\nu}(r)},\\ V(r) &= \mathbb{E}\left[x_i^2u_i^2|F(q_i)=r\right]=\frac{\iint x^2u^2f_{x,u,\nu}(x,u,r)dxdu}{f_\nu(r)}\equiv\frac{T(r)}{f_\nu(r)},\end{aligned}$$

where $\nu_i=F(q_i)$ is the standard uniform random variable. Hence, $f_\nu(r)=1$ and $\widehat{f}_{n,\nu}(r)\to_p 1$ as $n\to\infty$ for any $r\in[\tau,1-\tau]$ from the standard kernel density estimation result. It follows that

$$\sup_{r\in[\tau,1-\tau]}\left|\widehat{V}^0(r)-V(r)\right|\leq\sup_{r\in[\tau,1-\tau]}\left|\widehat{T}_n(r)-T(r)\right|+o_p(1),$$

and the desired result follows by showing $\sup_{r\in[\tau,1-\tau]}|\widehat{T}_n(r)-T(r)|=o_p(1)$ using a similar argument as in the proof of Lemma A.11 of Lee and Wang (2019). We now provide more details.

The triangular inequality yields

$$\sup_{r\in[\tau,1-\tau]}\left|\widehat{T}_n(r)-T(r)\right|\leq\sup_{r\in[\tau,1-\tau]}\left|\mathbb{E}[\widehat{T}_n(r)]-T(r)\right|+\sup_{r\in[\tau,1-\tau]}\left|\widehat{T}_n(r)-\mathbb{E}[\widehat{T}_n(r)]\right|,$$

where the first item is $o_p(1)$ as established in eqs. (12)-(13) and Lemma 1 in Yang (1981). For the second term, let κ_n be some large truncation parameter to be chosen later, satisfying $\kappa_n\to\infty$ as $n\to\infty$. Define

$$\widehat{T}_n^\kappa(r)=\frac{1}{n}\sum_{i=1}^n x_{[i]}^2u_{[i]}^2K_i(r)\,\mathbf{1}[x_{[i]}^2u_{[i]}^2\leq\kappa_n].$$

The triangular inequality gives that, for any $\eta>0$,

$$\begin{aligned}\mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\widehat{T}_n(r)-\mathbb{E}[\widehat{T}_n(r)]\right|>\eta\right) &\leq \mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\widehat{T}_n(r)-\widehat{T}_n^\kappa(r)\right|>\eta/3\right)\\ &\quad+\mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\mathbb{E}[\widehat{T}_n(r)]-\mathbb{E}[\widehat{T}_n^\kappa(r)]\right|>\eta/3\right)\end{aligned} \tag{A.6}$$

$$
\begin{aligned}
& +\mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\widehat{T}_n^{\kappa}(r)-\mathbb{E}[\widehat{T}_n^{\kappa}(r)]\right|>\eta/3\right) \\
\equiv\ & P_{n1}+P_{n2}+P_{n3}.
\end{aligned}
$$

For P_{n1}, since $\sup_{r\in[\tau,1-\tau]}|K_i(r)|<b_n^{-1}C_1$ for some $0<C_1<\infty$ from Condition 2.1, we have

$$
\begin{aligned}
\mathbb{E}\left[\sup_{r\in[\tau,1-\tau]}\left|\widehat{T}_n(r)-\mathbb{E}[\widehat{T}_n(r)]\right|\right] &\leq \mathbb{E}\left[\frac{C_1}{nb_n}\sum_{i=1}^{n}x_{[i]}^2u_{[i]}^2\mathbf{1}[x_{[i]}^2u_{[i]}^2>\kappa_n]\right] \qquad (A.7)\\
&\leq b_n^{-1}\kappa_n^{-1}C_1\sup_{q\in\mathbb{R}}\mathbb{E}\left[x_i^4u_i^4|q_i=q\right]\\
&\leq C_1b_n^{-1}\kappa_n^{-1}
\end{aligned}
$$

for some $C_1\in(0,\infty)$, where we use Condition 1.8 and the fact that

$$
\int_{|a|>\kappa_n}|a|\,F_A(da)\leq\kappa_n^{-1}\int_{|a|>\kappa_n}|a|^2\,F_A(da)\leq\kappa_n^{-1}\mathbb{E}[A^2]
$$

for a generic random variable $A\sim F_A$. Therefore, $P_{n1}\leq 3C_1/(\eta b_n\kappa_n)$ by Markov's inequality. Similarly,

$$
\sup_{r\in[\tau,1-\tau]}\left|\mathbb{E}[\widehat{T}_n\,(r)]-\mathbb{E}[\widehat{T}_n^{\kappa}(r)]\right|\leq b_n^{-1}\kappa_n^{-1}C_1\sup_{q\in\mathbb{R}}\mathbb{E}\left[x_i^4u_i^4|q_i=q\right]\leq C_1b_n^{-1}\kappa_n^{-1}
$$

and hence $P_{n2}\leq 3C_1/(\eta b_n\kappa_n)$ as well. For P_{n3}, Lemma A.4 below verifies that $P_{n3}\leq(\eta/3)^{-1}C(\log n/(nb_n))^{1/2}$ for some $0<C<\infty$. Therefore, if we choose κ_n such that $\kappa_n=O((b_n\log n/n)^{-1/2})$, we have both P_{n1} and P_{n2} are also bounded by $(\eta/3)^{-1}C(\log n/\,(nb_n))^{1/2}$. A possible choice of κ_n is $\kappa_n=O(n^{4/5})$ or larger as long as $b_n=O\left(n^{-1/5}\right)$. By combining these results, it follows that

$$
\mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\widehat{T}_n(r)-\mathbb{E}[\widehat{T}_n(r)]\right|>\eta\right)\leq\frac{9C}{\eta}\left(\frac{\log n}{nb_n}\right)^{1/2}\to 0
$$

as $n\to\infty$, where $\log n/\,(nb_n)\to 0$ from Condition 2.2.

The uniform consistency of $\widehat{h}(r)$ readily follows since

$$
\begin{aligned}
\widehat{h}(r)-h(r) &= \frac{1}{n}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor rn\rfloor}\frac{\widehat{D}\,(i/n)^2}{\widehat{V}\,(i/n)}-\int_{\tau}^{r}\frac{D\,(t)^2}{V\,(t)}dt\\
&= \frac{1}{n}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor rn\rfloor}\left\{\frac{\widehat{D}\,(i/n)^2}{\widehat{V}\,(i/n)}-\frac{D\,(i/n)^2}{V\,(i/n)}\right\}+\frac{1}{n}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor rn\rfloor}\frac{D\,(i/n)^2}{V\,(i/n)}-\int_{\tau}^{r}\frac{D\,(t)^2}{V\,(t)}dt,
\end{aligned}
$$

where the first term is uniformly $o_p(1)$ by the uniform consistency of $\widehat{D}(\cdot)$ and $\widehat{V}(\cdot)$; the second term is $o(1)$ from the standard Riemann integral, which is guaranteed by Condition 1.6. The uniform convergence of $\widehat{g}(r)$ then follows from that of $\widehat{h}(r)$ and the continuous mapping theorem. ■

Lemma A.4 *Under the same condition as in Lemma A.3, for any $\eta > 0$, P_{n3} in (A.6) satisfies that $P_{n3} \leq (\eta/3)^{-1}C(\log n/(nb_n))^{1/2}$ for some $0 < C < \infty$.*

Proof of Lemma A.4 Since $[\tau, 1-\tau]$ is compact, we can find m_n intervals centered at $r_1, \ldots, r_{m_n}$ with length C_S/m_n that cover $[\tau, 1-\tau]$ for some $C_S \in (0, \infty)$. We denote these intervals as $\mathcal{I}_j$ for $j = 1, \ldots, m_n$ and choose m_n later. The triangular inequality yields

$$\sup_{r\in[\tau,1-\tau]} \left|\widehat{T}_n^{\kappa}(r) - \mathbb{E}[\widehat{T}_n^{\kappa}(r)]\right| \leq T_{1n}^* + T_{2n}^* + T_{3n}^*,$$

where

$$\begin{aligned}
T_{1n}^* &= \max_{1\leq j\leq m_n} \sup_{r\in\mathcal{I}_j} \left|\widehat{T}_n^{\kappa}(r) - \widehat{T}_n^{\kappa}(r_j)\right| \\
T_{2n}^* &= \max_{1\leq j\leq m_n} \sup_{r\in\mathcal{I}_j} \left|\mathbb{E}[\widehat{T}_n^{\kappa}(r)] - \mathbb{E}[\widehat{T}_n^{\kappa}(r_j)]\right| \\
T_{3n}^* &= \max_{1\leq j\leq m_n} \left|\widehat{T}_n^{\kappa}(r_j) - \mathbb{E}[\widehat{T}_n^{\kappa}(r_j)]\right|.
\end{aligned}$$

We first bound T_{3n}^*. Let

$$Z_{n,i}^{\kappa}(r) = n^{-1}\left\{x_{[i]}^2 u_{[i]}^2 K_i(r)\,\mathbf{1}[x_{[i]}^2 u_{[i]}^2 \leq \kappa_n] - \mathbb{E}\left[x_{[i]}^2 u_{[i]}^2 K_i(r)\,\mathbf{1}[x_{[i]}^2 u_{[i]}^2 \leq \kappa_n]\right]\right\},$$

and then

$$\widehat{T}_n^{\kappa}(r) - \mathbb{E}[\widehat{T}_n^{\kappa}(r)] = \sum_{i=1}^{n} Z_{n,i}^{\kappa}(r).$$

Note that, similarly as (A.7), $\sup_{r\in[\tau,1-\tau]} x_{[i]}^2 u_{[i]}^2 K_i(r)\,\mathbf{1}[x_{[i]}^2 u_{[i]}^2 \leq \kappa_n]$ is bounded by $C_2\kappa_n b_n^{-1}$ for some constant $C_2 \in (0, \infty)$ and hence $|Z_{n,i}^{\kappa}(r)| \leq 2C_2\kappa_n/(nb_n)$ for all $i = 1, \ldots, n$. Define $\psi_n = (nb_n \log n)^{1/2}/\kappa_n$. Then $\psi_n|Z_{n,i}^{\kappa}(r)| \leq 2C_2(\log n/(nb_n))^{1/2} \leq 1/2$ for all i when n is sufficiently large. Using the inequality $\exp(x) \leq 1 + x + x^2$ for $|x| \leq 1/2$, we have $\exp(\psi_n|Z_{n,i}^{\kappa}(r)|) \leq 1 + \psi_n|Z_{n,i}^{\kappa}(r)| + \psi_n^2|Z_{n,i}^{\kappa}(r)|^2$. Hence

$$\mathbb{E}[\exp(\psi_n\left|Z_{n,i}^{\kappa}(r)\right|)] \leq 1 + \psi_n^2\mathbb{E}\left[(Z_{n,i}^{\kappa}(r))^2\right] \leq \exp\left(\psi_n^2\mathbb{E}\left[(Z_{n,i}^{\kappa}(r))^2\right]\right) \tag{A.8}$$

since $\mathbb{E}[Z^{\kappa}_{n,i}(r)] = 0$ and $1 + x \leq \exp(x)$ for $x \geq 0$. Using the fact that $\mathbb{P}(X > c) \leq \mathbb{E}[\exp(Xa)]/\exp(ac)$ for any random variable X and nonrandom constants a and c, we have that

$$
\begin{aligned}
\mathbb{P}\left(\left|\widehat{T}^{\kappa}_{n}(r) - \mathbb{E}[\widehat{T}^{\kappa}_{n}(r)]\right| > \eta_n\right) &= \mathbb{P}\left(\widehat{T}^{\kappa}_{n}(r) - \mathbb{E}[\widehat{T}^{\kappa}_{n}(r)] > \eta_n\right) + \mathbb{P}\left(-\widehat{T}^{\kappa}_{n}(r) + \mathbb{E}[\widehat{T}^{\kappa}_{n}(r)] > \eta_n\right) \\
&\leq \frac{\mathbb{E}\left[\exp\left(\psi_n \sum_{i=1}^{n} Z^{\kappa}_{n,i}(r)\right)\right] + \mathbb{E}\left[\exp\left(-\psi_n \sum_{i=1}^{n} Z^{\kappa}_{n,i}(r)\right)\right]}{\exp(\psi_n \eta_n)} \\
&\leq 2\exp(-\psi_n\eta_n)\exp\left(\psi_n^2 \sum_{i=1}^{n} \mathbb{E}\left[(Z^{\kappa}_{n,i}(r))^2\right]\right) \qquad \text{(by (A.8))} \\
&\leq 2\exp(-\psi_n\eta_n)\exp\left(\psi_n^2 C_3 \kappa_n^2/(nb_n)\right)
\end{aligned}
$$

for some sequence $\eta_n \to 0$ as $n \to \infty$, where the last inequality is from

$$
\sum_{i=1}^{n} \mathbb{E}\left[(Z^{\kappa}_{n,i}(r))^2\right] \leq n^{-2} \sum_{i=1}^{n} \mathbb{E}\left[x^4_{[i]} u^4_{[i]} K_i^2(r)\,\mathbf{1}[x^2_{[i]} u^2_{[i]} \leq \kappa_n]\right] \leq C_3 \kappa_n^2 (nb_n)^{-1}
$$

for some $C_3 \in (0, \infty)$. This bound is independent of r given Condition 1.8, and hence it is also the uniform bound, i.e.,

$$
\sup_{r \in [\tau, 1-\tau]} \mathbb{P}\left(\left|\widehat{T}^{\kappa}_{n}(r) - \mathbb{E}[\widehat{T}^{\kappa}_{n}(r)]\right| > \eta_n\right) \leq 2\exp\left(-\psi_n\eta_n + \psi_n^2 C_3 \kappa_n^2/(nb_n)\right). \tag{A.9}
$$

Now given κ_n, we need to choose $\eta_n \to 0$ as fast as possible, and at the same time we let $\psi_n\eta_n \to \infty$ at a rate that ensures (A.9) is summable and $\psi_n\eta_n > \psi_n^2\kappa_n^2/(nb_n)$. This is done by choosing $\psi_n = (nb_n \log n)^{1/2}/\kappa_n$ and $\eta_n = C^*\psi_n^{-1}\log n = C^*\kappa_n((\log n)/(nb_n))^{1/2}$ for some finite constant C^*. This choice yields

$$
-\psi_n\eta_n + \psi_n^2 C_3 \kappa_n^2/(nb_n) = -C^* \log n + C_3 \log n = -(C^* - C_3)\log n.
$$

Therefore, by substituting this into (A.9), we have

$$
\begin{aligned}
\mathbb{P}(T^*_{3n} > \eta_n) &= \mathbb{P}\left(\max_{1\leq j\leq m_n}\left|\widehat{T}^{\kappa}_{n}(r_j) - \mathbb{E}[\widehat{T}^{\kappa}_{n}(r_j)]\right| > \eta_n\right) \\
&\leq m_n \sup_{s\in[\tau,1-\tau]} \mathbb{P}\left(\left|\widehat{T}^{\kappa}_{n}(r) - \mathbb{E}[\widehat{T}^{\kappa}_{n}(r)]\right| > \eta_n\right) \leq 2\frac{m_n}{n^{C^*-C_4}}.
\end{aligned}
$$

Now, we can choose C^* sufficiently large so that $\sum_{n=1}^{\infty} \mathbb{P}(T^*_{3n} > \eta_n)$ is summable, from which we have

$$
T^*_{3n} = O_{a.s.}(\eta_n) = O_{a.s.}\left((\log n/(nb_n))^{1/2}\right)
$$

by the Borel-Cantelli lemma.

For T_{1n}^{*}, if n is sufficiently large,

$$\begin{aligned}\mathbb{E}\left|\widehat{T}_{n}^{\kappa}(r)-\widehat{T}_{n}^{\kappa}(r_{j})\right| &= \mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}x_{[i]}^{2}u_{[i]}^{2}\left(K_{i}(r)-K_{i}(r_{j})\right)\mathbf{1}[x_{[i]}^{2}u_{[i]}^{2}\leq\kappa_{n}]\right|\right] \\ &\leq C_{4}\left(1-2\tau\right)\kappa_{n}/m_{n}\end{aligned}$$

for some constant $C_4 < \infty$ given $r \in \mathcal{I}_j$. This bound does not depend on j and hence $T_{1n}^{*} = O_p(\kappa_n/m_n)$. The same argument yields that

$$\begin{aligned}\left|\mathbb{E}[\widehat{T}_{n}^{\kappa}(r)]-\mathbb{E}[\widehat{T}_{n}^{\kappa}(r_{j})]\right| &\leq \mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}x_{[i]}^{2}u_{[i]}^{2}\left(K_{i}(r)-K_{i}(r_{k})\right)\mathbf{1}[x_{[i]}^{2}u_{[i]}^{2}\leq\kappa_{n}]\right|\right] \\ &\leq C_{4}\left(1-2\tau\right)\kappa_{n}/m_{n},\end{aligned}$$

which does not depend on j, and hence it gives the uniform bound $T_{2n}^{*} = O(\tau_n/m_n)$ as well. Therefore, by choosing $m_n = [((\log n)/(nb_n))^{1/2}/\kappa_n]^{-1}$, we have that T_{1n}^{*} and T_{2n}^{*} are both the order of $((\log n)/(nb_n))^{1/2}$. By combining these results, it follows that $P_{n3} \leq (\eta/3)^{-1}C((\log n)/(nb_n))^{1/2}$ for some $C \in (0,\infty)$ by Markov's inequality. ■

Lemma A.5 *Let*

$$\mathcal{G}_{vn}(s) = \frac{h_{\tau}^{1/2}}{\sqrt{n}}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} g^{(1)}(i/n)v^{\intercal}D(i/n)^{-1}x_{[i]}\widehat{u}_{[i]}. \tag{A.10}$$

Suppose Conditions 1 and 2 hold. Then under the null hypothesis in (12), we have $\mathcal{G}_{vn}(\cdot) \Rightarrow \mathcal{G}_{v}(\cdot)$ as $n \to \infty$.

Proof of Lemma A.5 Recall that $\widehat{u}_i = u_i - x_i^{\intercal}(\widehat{\beta} - \beta_0) - x_i^{\intercal}(\widehat{\delta} - \delta_0)\mathbf{1}[q_i \leq \gamma_0] - x_i^{\intercal}\widehat{\delta}\left(\mathbf{1}[q_i \leq \widehat{\gamma}] - \mathbf{1}[q_i \leq \gamma_0]\right)$. Hence, we have

$$\begin{aligned}\mathcal{G}_{vn}(s) = {} & \frac{h_{\tau}^{1/2}}{\sqrt{n}}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} g^{(1)}(i/n)v^{\intercal}D\left(i/n\right)^{-1}x_{[i]}u_{[i]} \\ & -\frac{h_{\tau}^{1/2}}{n}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} g^{(1)}(i/n)v^{\intercal}D\left(i/n\right)^{-1}x_{[i]}x_{[i]}^{\intercal}\sqrt{n}(\widehat{\beta}-\beta_0) \\ & -\frac{h_{\tau}^{1/2}}{n}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} g^{(1)}(i/n)v^{\intercal}D\left(i/n\right)^{-1}x_{[i]}x_{[i]}^{\intercal}\sqrt{n}(\widehat{\delta}-\delta_0)\mathbf{1}\left[q_{(i)}\leq\gamma_0\right]\end{aligned} \tag{A.11}$$

$$
\begin{aligned}
&-\frac{h_\tau^{1/2}}{\sqrt{n}}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} g^{(1)}(i/n)v^\intercal D\left(i/n\right)^{-1}x_{[i]}x_{[i]}^\intercal\widehat{\delta}\left(\mathbf{1}\left[q_{(i)}\leq\widehat{\gamma}\right]-\mathbf{1}\left[q_{(i)}\leq\gamma_0\right]\right)\\
\equiv\ & A_{1n}\left(s\right)-A_{2n}\left(s\right)-A_{3n}\left(s\right)-A_{4n}(s).
\end{aligned}
$$

First, we derive the limit of $A_{1n}\left(s\right)$ by applying Corollary 29.14 in Davidson (1994).[6] To this end, we let $U_{n,i}=h_\tau^{1/2}n^{-1/2}g^{(1)}(i/n)v^\intercal D\left(i/n\right)^{-1}x_{[i]}u_{[i]}$ and $q^{(n)}=\{q_i\}_{i=1}^n$, and check Conditions 29.6(a) to (f′) in the corollary. Condition (a) is satisfied since $\mathbb{E}\left[U_{n,i}\right]=\mathbb{E}[\mathbb{E}\left[U_{n,i}|q^{(n)}\right]]=0$ given our Conditions 1.1 and 1.2. Condition (b) is implied by our Conditions 1.6 and 1.8 by setting $c_{n,i}=1$ in the corollary as seen by

$$
\sup_{i/n\in[\tau,1-\tau]}\left\|U_{n,i}\right\|_4\leq\frac{h_\tau^{1/2}}{\sqrt{n}}\sup_{r\in[\tau,1-\tau]}\left\|v^\intercal D\left(r\right)^{-1}\right\|_4\sup_{r\in[\tau,1-\tau]}\left|g^{(1)}(r)\right|\times\left(\sup_{q\in\mathbb{R}}\mathbb{E}\left[\left|\left|x_iu_i\right|\right|^4|q_i=q\right]\right)^{1/4}<\infty,
$$

where $\left|\left|\cdot\right|\right|_p$ denotes the L^p-norm. Condition (c) is implied by the fact that $\{U_{n,i}\}_{i=1}^n$ is a martingale difference array (see, e.g., Lemma 3.2 of Bhattacharya (1984)). Thus, the NED condition is satisfied. Condition (d) holds by setting $c_{n,i}=1$ and $K_n(t)=\left\lfloor g^{-1}(t)n\right\rfloor$, and from the fact that $g^{-1}\left(\cdot\right)$ is continuously differentiable. Condition (e) is satisfied by setting $c_{n,i}=1$ since $\{U_{n,i}\}_{i=1}^n$ is independent conditional $q^{(n)}$ almost surely (see, e.g., Lemma 3.1 of Bhattacharya (1984)). To satisfy Condition (f′), our Condition 1.6 and Taylor expansion of $V(\cdot)$ at i/n yield that

$$
\begin{aligned}
\mathbb{E}\left[x_{[i]}x_{[i]}^\intercal u_{[i]}^2\right]&=\mathbb{E}\left[\mathbb{E}\left[x_jx_j^\intercal u_j^2|q_j=q_{(i)}\right]\right]\\
&=\mathbb{E}\left[V(F(q_{(i)}))\right]\\
&=V(i/n)+\mathbb{E}\left[\frac{\partial V(t_i)}{\partial t}\left(F\left(q_{(i)}\right)-i/n\right)\right]\\
&=V(i/n)+O\left(n^{-1/2}\right),
\end{aligned}\tag{A.12}
$$

where t_i is between i/n and $F(q_{(i)})$ in the third equality. The last equality follows from

$$
\begin{aligned}
\sup_{i/n\in[\tau,1-\tau]}\left\|\mathbb{E}\left[\frac{\partial V(t_i)}{\partial t}\left(F\left(q_{(i)}\right)-\widehat{F}_n(q_{(i)})\right)\right]\right\| &\leq \sup_{t\in[\tau,1-\tau]}\left\|\frac{\partial V(t)}{\partial t}\right\|\mathbb{E}\left[\sup_{t\in[\tau,1-\tau]}\left|F\left(t\right)-\widehat{F}_n(t)\right|\right]\\
&= O\left(n^{-1/2}\right),
\end{aligned}
$$

[6] Note that we cannot apply Theorem 2 in Bhattacharya (1974) to derive the limit of $A_{1n}\left(s\right)$ as in the proof of Theorem 1. This is because the pre-ordered version of $\{g^{(1)}(i/n)v^\intercal D\left(i/n\right)^{-1}x_{[i]}u_{[i]}\}_{i=1}^n$ is $\{g^{(1)}(R_i/n)v^\intercal D\left(R_i/n\right)^{-1}x_iu_i\}_{i=1}^n$, which is no longer i.i.d. given the rank statistics $\{R_i\}_{i=1}^n$.

which is from Donsker's theorem and Condition 1.6. Then we obtain that

$$
\begin{aligned}
\mathbb{E}\left[\left(\sum_{i=\lfloor\tau n\rfloor+1}^{K_n(s)} U_{n,i}\right)^2\right] &= \mathbb{E}\left[\sum_{i=\lfloor\tau n\rfloor+1}^{K_n(s)} U_{n,i}^2\right] \\
&= \frac{h_\tau}{n}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor}\left(g^{(1)}(i/n)\right)^2 v^\intercal D\left(i/n\right)^{-1}\mathbb{E}\left[x_{[i]}x_{[i]}^\intercal u_{[i]}^2\right]D\left(i/n\right)^{-1}v \\
&= \frac{h_\tau}{n}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor}\left(g^{(1)}(i/n)\right)^2 v^\intercal D\left(i/n\right)^{-1}V(i/n)D\left(i/n\right)^{-1}v + O(n^{-1/2}) \\
&\to h_\tau\int_\tau^{g^{-1}(s)}\left(g^{(1)}\left(t\right)\right)^2 v^\intercal D\left(t\right)^{-1}V(t)D\left(t\right)^{-1}vdt \\
&= \int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t)dt = s,
\end{aligned}
$$

where the first equality is from the fact that $\{U_{n,i}\}_{i=1}^n$ is a martingale difference array; the third equality is by (A.12); the second expression from the bottom is by Riemann integral as $n\to\infty$; the last expression is by the definition of $g^{(1)}(\cdot)$ and $g^{-1}(0)=\tau$. Therefore, Corollary 29.14 Davidson (1994) implies that $A_{1n}\left(s\right)=\sum_{i=\lfloor\tau n\rfloor+1}^{K_n(s)} U_{n,i}^2 \Rightarrow W_1(s)$ for $s\in[0,1]$.

For $A_{2n}(s)$ and $A_{3n}(s)$, we apply Lemma A.1, Lemma A.12 in Hansen (2000a), and the continuous mapping theorem to obtain that

$$
A_{2n}\left(s\right)\to_p\left(\int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t)v^\intercal D\left(t\right)^{-1}D\left(t\right)dt\right)\Phi_\beta h_\tau^{1/2} = sv^\intercal\Phi_\beta h_\tau^{1/2}
$$

and

$$
\begin{aligned}
A_{3n}\left(s\right) &= \frac{h_\tau^{1/2}}{n}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} g^{(1)}(i/n)v^\intercal D\left(i/n\right)^{-1}x_{[i]}x_{[i]}^\intercal\mathbf{1}\left[i/n\le r_0\right]\sqrt{n}\left(\widehat{\delta}-\delta_0\right) \\
&\to_p\left(\int_{g^{-1}(0)}^{\min(g^{-1}(s),r_0)} g^{(1)}(t)v^\intercal D\left(t\right)^{-1}D\left(t\right)dt\right)\Phi_\delta h_\tau^{1/2} \\
&= \min\{s,g\left(r_0\right)\}v^\intercal\Phi_\delta h_\tau^{1/2}.
\end{aligned}
$$

Finally, for A_{4n}, let $E_{\gamma n}$ denote the event that $\widehat{\gamma}\in\mathcal{B}_{Cn^{-1+2\epsilon}}(\gamma_0)$ for some C. Lemma A.12 in Hansen (2000a) yields that $\mathbb{P}\left(E_{\gamma n}^c\right)\le\varepsilon$ for any $\varepsilon>0$ as $n\to\infty$ Then for any $\eta>0$ and $\varepsilon>0$, if

n is sufficiently large,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{s\in[0,1]}|A_{4n}(s)|>\eta\right)\\
&\leq \mathbb{P}\left(\left\{\sup_{s\in[0,1]}|A_{4n}(s)|>\eta\right\}\cap E_{\gamma n}\right)+\varepsilon\\
&\leq \eta^{-1}\frac{h_\tau^{1/2}}{n}\mathbb{E}\left[\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor(1-\tau)n\rfloor}\left|g^{(1)}(i/n)v^\intercal D\left(i/n\right)^{-1}x_{[i]}x_{[i]}^\intercal\widehat{\delta}\left(\mathbf{1}\left[q_{(i)}\leq\widehat{\gamma}\right]-\mathbf{1}\left[q_{(i)}\leq\gamma_0\right]\right)\right|\mathbf{1}\left[E_{\gamma n}\right]\right]+\varepsilon\\
&\leq \eta^{-1}h_\tau^{1/2}\sup_{s\in[\tau,1-\tau]}\left\|g^{(1)}(s)v^\intercal D\left(s\right)^{-1}\right\|\mathbb{E}\left[\|x_ix_i^\intercal\|\,\widehat{\delta}\left(\mathbf{1}\left[q_i\leq\widehat{\gamma}\right]-\mathbf{1}\left[q_i\leq\gamma_0\right]\right)\mathbf{1}\left[E_{\gamma n}\right]\right]+\varepsilon\\
&\leq \eta^{-1}Cn^{-1+2\epsilon}+\varepsilon\\
&\leq 2\varepsilon,
\end{aligned}
$$

where the second inequality is by Markov's inequality and the fourth inequality is by Conditions 1.6 and 1.8. Thus, $\sup_{s\in[0,1]}||A_{4n}(s)||=o_p(1)$. The desired result follows by combining these results. ■

Proof of Lemma 1 The first result follows from Lemma A.3. For the second result, given Lemma A.5, it suffices to establish

$$
\sup_{s\in[0,1]}\left|\widehat{\mathcal{G}}_{vn}\left(s\right)-\mathcal{G}_{vn}(s)\right|=o_p(1).
$$

We first consider $g^{-1}(s)>\widehat{g}^{-1}(s)$. Given Lemma A.5 and $\widehat{h}_\tau^{1/2}=h_\tau^{1/2}+o_p(1)$ from Lemma A.3, we have, for any $s\in[0,1]$,

$$
\begin{aligned}
\widehat{\mathcal{G}}_{vn}\left(s\right)-\mathcal{G}_{vn}(s) &= \frac{h_\tau^{1/2}}{\sqrt{n}}\sum_{i=\lfloor\tau n\rfloor+1}^{\left\lfloor\widehat{g}^{-1}(s)n\right\rfloor}\widehat{g}^{(1)}(i/n)v^\intercal\widehat{D}\left(i/n\right)^{-1}x_{[i]}\widehat{u}_{[i]}\\
&\quad-\frac{h_\tau^{1/2}}{\sqrt{n}}\sum_{i=\lfloor\tau n\rfloor+1}^{\left\lfloor g^{-1}(s)n\right\rfloor}g^{(1)}(i/n)v^\intercal D\left(i/n\right)^{-1}x_{[i]}\widehat{u}_{[i]}+o_p\left(1\right)\\
&= \frac{h_\tau^{1/2}}{\sqrt{n}}\sum_{i=\lfloor\tau n\rfloor+1}^{\left\lfloor\widehat{g}^{-1}(s)n\right\rfloor}\left\{\widehat{g}^{(1)}(i/n)v^\intercal\widehat{D}\left(i/n\right)^{-1}-g^{(1)}(i/n)v^\intercal D\left(i/n\right)^{-1}\right\}x_{[i]}\widehat{u}_{[i]}\\
&\quad-\frac{h_\tau^{1/2}}{\sqrt{n}}\sum_{i=\left\lfloor\widehat{g}^{-1}(s)n\right\rfloor+1}^{\left\lfloor g^{-1}(s)n\right\rfloor}g^{(1)}(i/n)v^\intercal D\left(i/n\right)^{-1}x_{[i]}\widehat{u}_{[i]}+o_p\left(1\right)
\end{aligned}
$$

$$\equiv \quad B_{1n}(s) - B_{2n}(s) + o_p(1). \tag{A.13}$$

For expositional simplicity, we only present the case with scalar x_i. Then v is simply 1.

For $B_{1n}(s)$, we write

$$\begin{aligned} B_{1n}(s) &= \frac{h_\tau^{1/2}}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor +1}^{\lfloor \widehat{g}^{-1}(s)n \rfloor} \left\{ \widehat{g}^{(1)}(i/n)\widehat{D}(i/n)^{-1} - g^{(1)}(i/n)D(i/n)^{-1} \right\} x_{[i]}u_{[i]} \\ &\quad + \frac{h_\tau^{1/2}}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor +1}^{\lfloor \widehat{g}^{-1}(s)n \rfloor} \left\{ \widehat{g}^{(1)}(i/n)\widehat{D}(i/n)^{-1} - g^{(1)}(i/n)D(i/n)^{-1} \right\} x_{[i]} \left(u_{[i]} - \widehat{u}_{[i]} \right) \\ &\equiv B_{11n}(s) + B_{12n}(s). \end{aligned} \tag{A.14}$$

We can verify $\sup_{s\in[0,1]} |B_{11n}(s)| = o_p(1)$ from the argument in Chapter 2 of van der Vaart and Wellner (1996), which we present in Lemma A.6 below. For $B_{12n}(s)$, define the event $E_{\theta n} = \{\widehat{\theta} \in \mathcal{B}_{Cn^{-1/2}}(\theta_0)\}$ for some C. Lemma A.12 in Hansen (2000a) implies that $\mathbb{P}(E_{\theta n}^c) \leq \varepsilon$ for any $\varepsilon > 0$ as $n \to \infty$. Then for any $\varepsilon > 0$, if n is large enough, we have

$$\begin{aligned} &\sup_{s\in[0,1]} |B_{12n}(s)| \\ &\leq h_\tau^{1/2} \sup_{r\in[\tau,1-\tau]} \left| \widehat{g}^{(1)}(r)\widehat{D}(r)^{-1} - g^{(1)}(r)D(r)^{-1} \right| \sup_{r\in[\tau,1-\tau]} \frac{1}{\sqrt{n}} \left| \sum_{i=\lfloor \tau n \rfloor +1}^{\lfloor rn \rfloor} x_{[i]}(\widehat{u}_{[i]} - u_{[i]}) \right| \\ &\leq o_p(1) \left\{ \sup_{r\in[\tau,1-\tau]} \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor +1}^{\lfloor rn \rfloor} x_{[i]}^2 |\widehat{\beta} - \beta_0| \right. \\ &\qquad + \sup_{r\in[\tau,1-\tau]} \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor +1}^{\lfloor rn \rfloor} x_{[i]}^2 \mathbf{1}\left[q_{(i)} \leq \gamma_0 \right] |\widehat{\delta} - \delta_0| \\ &\qquad \left. + \sup_{r\in[\tau,1-\tau]} \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor +1}^{\lfloor rn \rfloor} x_{[i]}^2 |\widehat{\delta}| \left| \mathbf{1}\left[q_{(i)} \leq \gamma_0 \right] - \mathbf{1}\left[q_{(i)} \leq \widehat{\gamma} \right] \right| \right\} \\ &= o_p(1), \end{aligned}$$

where the second inequality is by Lemma A.3, and the last equality follows from Lemma A.1 and (A.3). Therefore, $B_{1n}(s)$ in (A.13) is uniformly $o_p(1)$.

For $B_{2n}(s)$ in (A.13), we write

$$\begin{aligned}
B_{2n}(s) &= \frac{h_\tau^{1/2}}{\sqrt{n}} \sum_{i=\lfloor \widehat{g}^{-1}(s)n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} g^{(1)}(i/n) D\left(i/n\right)^{-1} x_{[i]} u_{[i]} \\
&\quad + \frac{h_\tau^{1/2}}{\sqrt{n}} \sum_{i=\lfloor \widehat{g}^{-1}(s)n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} g^{(1)}(i/n) D\left(i/n\right)^{-1} x_{[i]} \left(\widehat{u}_{[i]} - u_{[i]}\right) \\
&\equiv B_{21n}(s) + B_{22n}(s).
\end{aligned} \tag{A.15}$$

For $B_{21n}(s)$, define the event $E_{gn} = \{\sup_{s\in[0,1]} |\widehat{g}^{-1}(s) - g^{-1}(s)| < \eta\}$ for some $\eta > 0$. By Lemma A.3, $\mathbb{P}(E_{gn}^c) \leq \varepsilon$ for any $\varepsilon > 0$ and $\eta > 0$ as $n \to \infty$. On the event E_{gn}, for any given value $\widehat{g}^{-1}(s) = \varrho(s)$, we have that

$$\begin{aligned}
\sup_{s\in[0,1]} |B_{21n}(s)| &\leq \sup_{s\in[0,1]} \sup_{|\varrho(s)-g^{-1}(s)|<\eta} \left| \frac{h_\tau^{1/2}}{\sqrt{n}} \sum_{i=\lfloor \varrho(s)n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} g^{(1)}(i/n) D\left(i/n\right)^{-1} x_{[i]} u_{[i]} \right| \\
&\Rightarrow \sup_{s\in[0,1]} \sup_{|\varrho(s)-g^{-1}(s)|<\eta} \left| h_\tau^{1/2} \int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t) D\left(t\right)^{-1} V\left(t\right)^{1/2} dW_k\left(t\right) \right. \\
&\qquad \left. - h_\tau^{1/2} \int_{g^{-1}(0)}^{\varrho(s)} g^{(1)}(t) D\left(t\right)^{-1} V\left(t\right)^{1/2} dW_k\left(t\right) \right| \\
&=_d \sup_{s\in[0,1]} \sup_{|\varrho(s)-g^{-1}(s)|<\eta} |W_1(s) - W_1(g(\varrho(s)))|
\end{aligned}$$

similarly as $H_1(s)$ in (A.4). Then, we can choose η small enough to obtain that, for any $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}\left(\sup_{s\in[0,1]} |B_{21n}(s)| > \varepsilon \right) &\leq \mathbb{P}\left(\left\{ \sup_{s\in[0,1]} |B_{21n}(s)| > \varepsilon \right\} \cap E_{gn} \right) + \mathbb{P}(E_{gn}^c) \\
&\to \mathbb{P}\left(\sup_{s\in[0,1]} \sup_{|\varrho(s)-g^{-1}(s)|<\eta} |W_1(s) - W_1(g(\varrho(s)))| > \varepsilon \right) + \varepsilon \\
&\leq \varepsilon^{-1} \mathbb{E}\left[\sup_{s\in[0,1]} \sup_{|\varrho(s)-g^{-1}(s)|<\eta} |W_1(s) - W_1(g(\varrho(s)))| \right] + \varepsilon \\
&\leq \varepsilon^{-1} \eta^{1/2} C + \varepsilon \\
&\leq 2\varepsilon,
\end{aligned}$$

where the second inequality is by Markov's inequality; the third inequality follows from the continuity of $g(\cdot)$ and from the fact that $\mathbb{E}\left[\sup_{s\in[0,t]} |W_1(s)|\right] \leq \sqrt{2t/\pi}$; and the last inequality holds with a sufficiently small η. For $B_{22n}(s)$, consider the same events $E_{\theta n}$ and E_{gn} as above. Then, on the these two events, using the same decomposition with the $A_{2n}(s)$, $A_{3n}(s)$, and $A_{4n}(s)$ terms as in (A.11), we have that

$$
\begin{aligned}
&\sup_{s\in[0,1]} |B_{22n}(s)| \\
&\leq h_\tau^{1/2} \sup_{r\in[\tau,1-\tau]} \left| g^{(1)}(r) D(r)^{-1} \right| \sup_{s\in[0,1]} \frac{1}{\sqrt{n}} \sum_{i=\lfloor \widehat{g}^{-1}(s)n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} \left| x_{[i]}(u_{[i]} - \widehat{u}_{[i]}) \right| \\
&\leq C \sup_{s\in[0,1]} \frac{1}{\sqrt{n}} \sum_{i=\lfloor (g^{-1}(s)-\eta)n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} x_{[i]}^2 \left\{ |\widehat{\beta} - \beta_0| + |\widehat{\delta} - \delta_0| \mathbf{1}\left[q_{(i)} \leq \gamma_0\right] + \widehat{\delta} \left| \mathbf{1}\left[q_{(i)} \leq \widehat{\gamma}\right] - \mathbf{1}\left[q_{(i)} \leq \gamma_0\right] \right| \right\} \\
&\leq C \sup_{s\in[0,1]} \frac{1}{n} \sum_{i=\lfloor (g^{-1}(s)-\eta)n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} x_{[i]}^2 \\
&\to_p C \sup_{s\in[0,1]} \int_{g^{-1}(s)-\eta}^{g^{-1}(s)} D(t)\, dt
\end{aligned}
$$

for some constant $0 < C < \infty$, where the second inequality is from Condition 1.6; the third inequality is from the fact that $\mathbf{1}\left[q_{(i)} \leq \gamma\right] \leq 1$ for any γ, result in (A.3), and by conditioning on the events $E_{\theta n}$ and E_{gn}; the last convergence is from Lemma A.1. By choosing a sufficiently small η, therefore, $\sup_{s\in[0,1]} |B_{22n}(s)| = o_p(1)$, which completes the proof. The proof for $g(s) \leq \widehat{g}^{-1}(s)$ is identical and hence omitted. ■

Lemma A.6 *Under the same condition as in Lemma 1,* $\sup_{s\in[0,1]} |B_{11n}(s)| = o_p(1)$*, where* $B_{11n}(\cdot)$ *is defined in (A.14).*

Proof of Lemma A.6 Note that for each n, $\{x_{[i]}u_{[i]}\}_{i=1}^n$ are independent conditional on $q^{(n)} = \{q_1, \ldots, q_n\}$ almost surely (Lemma 3.1 in Bhattacharya (1984)). We aim to use the empirical process argument for independent variables in van der Vaart and Wellner (1996). To this end, we consider the class of functions $J(\cdot) = g^{(1)}(\cdot)v^\intercal D(\cdot)^{-1}$ and the stochastic process

$$
\mathbb{V}_\ell(J) = \sum_{i=\lfloor \tau n \rfloor+1}^{\ell} V_{ni}(J),
$$

where $V_{ni}(J) = h_\tau^{1/2} n^{-1/2} J(i/n) x_{[i]} u_{[i]}$. Define the semi-metric $\rho(J_1, J_2) = \sup_{r\in[\tau,1-\tau]} |J_1(r) - J_2(r)|$. Then the space of continuously differentiable functions defined on $[\tau, 1-\tau]$, denoted $C^1[\tau, 1-\tau]$, is totally bounded. We now apply Theorem 2.11.9 in van der Vaart and Wellner (1996) by checking their conditions. (See also Theorem 3 in Bae, Jun, and Levental (2010) for a martingale difference array argument since $\{x_{[i]}u_{[i]}\}_{i=1}^n$ also form a martingale difference array by Lemma 3.2 in Bhattacharya (1984)).

First, we let their m_n be $\lfloor (1-\tau)n \rfloor$ and their $\mathcal{F}$ be $C^1[\tau, 1-\tau]$. Set their envelope function F as $\bar{C}\,||x||$ for a large enough constant $\bar{C}$. Then, their first condition is satisfied as we write, for any $\varepsilon > 0$,

$$\begin{aligned}
&\sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor (1-\tau)n \rfloor} \mathbb{E}\left[\sup_{J\in\mathcal{F}} |V_{ni}(J)|\,\mathbf{1}\left[\sup_{J\in\mathcal{F}} |V_{ni}(J)| > \varepsilon\right] \middle| q^{(n)}\right] \\
&\leq \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor (1-\tau)n \rfloor} \mathbb{E}\left[\sup_{J\in\mathcal{F}} |V_{ni}(J)|^2 \middle| q^{(n)}\right]^{1/2} \mathbb{P}\left(\sup_{J\in\mathcal{F}} |V_{ni}(J)| > \varepsilon \middle| q^{(n)}\right)^{1/2} \\
&\leq \varepsilon^{-4} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor (1-\tau)n \rfloor} \mathbb{E}\left[\sup_{J\in\mathcal{F}} |V_{ni}(J)|^2 \middle| q^{(n)}\right]^{1/2} \mathbb{E}\left[\sup_{J\in\mathcal{F}} |V_{ni}(J)|^4 \middle| q^{(n)}\right]^{1/2} \\
&\leq \bar{C}^3 n^{-3/2}\varepsilon^{-4} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor (1-\tau)n \rfloor} \mathbb{E}\left[\left\|x_{[i]}u_{[i]}\right\|^2 \middle| q^{(n)}\right]^{1/2} \mathbb{E}\left[\left\|x_{[i]}u_{[i]}\right\|^4 \middle| q^{(n)}\right]^{1/2} \\
&\to 0 \quad \text{a.s.}
\end{aligned}$$

as $n \to \infty$, where the first two inequalities are from Cauchy-Schwarz inequality and the third inequality is by substituting the envelope function $\bar{C}\,||x||$ and from Condition 1.8. Regarding their second condition, we have

$$\begin{aligned}
\sup_{\rho(J,J_1)\leq\varepsilon_n} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor (1-\tau)n \rfloor} \mathbb{E}\left[(V_{ni}(J) - V_{ni}(J_1))^2 \,|q^{(n)}\right] \quad &\leq \quad \bar{C}^2\varepsilon_n n^{-1} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor (1-\tau)n \rfloor} \mathbb{E}\left[\left|x_{[i]}u_{[i]}\right|^2 \middle| q^{(n)}\right] \\
&\to \quad 0 \quad \text{a.s.}
\end{aligned}$$

for every $\varepsilon_n \downarrow 0$. Regarding their third condition, the smoothness of $\mathcal{F}$ is sufficient for Corollary 2.7.2 in van der Vaart and Wellner (1996) by considering their d and α as both 1. This is further sufficient for their uniform bracketing entropy condition. Thus their Theorem 2.11.9 implies that conditional on $q^{(n)}$, the process $\mathbb{V}_n(\cdot)$ is asymptotically tight, that is, for any $\varepsilon > 0$, there exists some η such that if n is large enough,

$$\mathbb{P}\left(\sup_{\rho(J_1,J_2)\leq\eta} |\mathbb{V}_n(J_1) - \mathbb{V}_n(J_2)| > \varepsilon \middle| q^{(n)}\right) \leq \varepsilon \quad \text{a.s.} \tag{A.16}$$

Define $E_{Jn} = \{\rho(\widehat{J}, J) \leq \eta_n\}$ for $\eta_n > 0$, where $\widehat{J}(\cdot) = \widehat{g}^{(1)}(\cdot)v^\intercal \widehat{D}(\cdot)^{-1}$. Then, for any $\varepsilon > 0$, we have

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{s\in[0,1]}|B_{11n}(s)|>\varepsilon\right)\\
&\leq \mathbb{E}\left[\mathbb{P}\left(\left\{\sup_{s\in[0,1]}|B_{11n}(s)|>\varepsilon\right\}\cap E_{Jn}\middle| q^{(n)}\right)\right]+\mathbb{P}\left(E_{Jn}^{c}\right)\\
&\leq \mathbb{E}\left[\mathbb{P}\left(\max_{1\leq\ell\leq n}\sup_{\rho(\widehat{J},J)\leq\eta}\left|\mathbb{V}_{\ell}(J)-\mathbb{V}_{\ell}(\widehat{J})\right|>\varepsilon\middle| q^{(n)}\right)\right]+\varepsilon\\
&\leq \mathbb{E}\left[\frac{\mathbb{P}\left(\sup_{\rho(\widehat{J},J)\leq\eta}|\mathbb{V}_{n}(J)-\mathbb{V}_{n}(\widehat{J})|>\varepsilon\middle| q^{(n)}\right)}{1-\max_{1\leq\ell\leq n}\mathbb{P}\left(\sqrt{\ell/n}\sup_{\rho(\widehat{J},J)\leq\eta}|\mathbb{V}_{\ell}(J)-\mathbb{V}_{\ell}(\widehat{J})|>\varepsilon\middle| q^{(n)}\right)}\right]+\varepsilon\\
&\leq C\varepsilon.
\end{aligned}
$$

The second inequality is from Lemma A.3 that implies $\mathbb{P}(E_{Jn}^{c})\leq\varepsilon$ if n is large enough, and from the law of iterated expectations. The third inequality is from the Ottaviani's inequality (e.g., A.1.1 in van der Vaart and Wellner (1996)) and the fact that $\{x_{[i]}u_{[i]}\}_{i=1}^{n}$ are independent conditional on $q^{(n)}$. The last inequality is from (A.16) and the steps in p. 227 in van der Vaart and Wellner (1996). In particular, for some $1\leq n_0\leq n$,

$$
\begin{aligned}
&\max_{1\leq\ell\leq n}\mathbb{P}\left(\sqrt{\ell/n}\sup_{\rho(\widehat{J},J)\leq\eta}\left|\mathbb{V}_{\ell}(J)-\mathbb{V}_{\ell}(\widehat{J})\right|>\varepsilon\middle| q^{(n)}\right)\\
&\leq \max_{\ell\leq n_0}\mathbb{P}\left(n^{-1/2}\sum_{i=\lfloor\tau n\rfloor+1}^{n_0}C\left|\left|x_{[i]}u_{[i]}\right|\right|>\varepsilon\middle| q^{(n)}\right)\\
&\quad+\max_{n_0\leq\ell}\mathbb{P}\left(\sup_{\rho(\widehat{J},J)\leq\eta}\left|\mathbb{V}_{\ell}(J)-\mathbb{V}_{\ell}(\widehat{J})\right|>\varepsilon\middle| q^{(n)}\right)\\
&\leq C\varepsilon\quad\text{a.s.},
\end{aligned}
$$

where the second inequality follows from Markov's inequality, (A.16), and setting a large enough n_0 satisfying $n_0\to\infty$ and $n_0n^{-1/2}\to 0$. ■

Proof of Theorem 3 We first prove (19) under the null hypothesis. To this end, define

$$
\widetilde{\mathcal{G}}_{n}^{*}(s)=\begin{cases}\widehat{\mathcal{G}}_{1n}^{*}(\cdot) & \text{if } s\leq g(r_0)\\ \widehat{\mathcal{G}}_{2n}^{*}(\cdot) & \text{otherwise,}\end{cases}
$$

which is different from $\widehat{\mathcal{G}}_n^*(\cdot)$ only in a neighborhood of $g(r_0)$. Under the null hypothesis, Lemmas A.1 and A.3 and the continuous mapping theorem yield that

$$\widetilde{\mathcal{G}}_n^*(s) \Rightarrow \begin{cases} \frac{1}{\sqrt{g(r_0)}}\left\{W_1(s) - \frac{s}{g(r_0)}W_1(g(r_0))\right\} & \text{if } s \le g(r_0) \\ \frac{1}{\sqrt{1-g(r_0)}}\left\{W_1(1) - W_1(s) - \frac{1-s}{1-g(r_0)}(W_1(1) - W_1(g(r_0)))\right\} & \text{otherwise} \end{cases}$$

as $n \to \infty$. Therefore, we can establish $\int_0^1 \left|\widehat{\mathcal{G}}_n^*(s) - \widetilde{\mathcal{G}}_n^*(s)\right| ds = o_p(1)$ to obtain the desired result. However, since the empirical cdf is uniformly consistent, Lemma A.3 yields $\widehat{g}(\widehat{r}) - g(r_0) = o_p(1)$. Therefore, it suffices to establish $\int_{g(r_0)-\varepsilon_n}^{g(r_0)+\varepsilon_n} \left|\widehat{\mathcal{G}}_n^*(t) - \widetilde{\mathcal{G}}_n^*(t)\right| dt = o_p(1)$ for some $\varepsilon_n \to 0$ with $n \to \infty$, which is further implied by the fact that both $\sup_{s\in[0,1]}\left|\widehat{\mathcal{G}}_n^*(s)\right|$ and $\sup_{s\in[0,1]}\left|\widetilde{\mathcal{G}}_n^*(s)\right|$ are $O_p(1)$ given Lemma 1. The limiting null distribution of CT_n hence follows as (19) by the continuous mapping theorem.

We now examine the limit of $\widehat{\mathcal{G}}_n(s)$ under the alternative. In this case, $\widehat{\gamma}$ (or $\widehat{r} = \widehat{F}_n(\widehat{\gamma})$) is never consistent since γ_i (or r_i) is a random variable with a non-degenerate variance. Hence, the nonparametric estimators that depend on $\widehat{\gamma}$, $\widehat{V}(\cdot)$, $\widehat{h}(\cdot)$, and $\widehat{g}(\cdot)$, are no longer consistent but still $O_p(1)$. On the other hand, $\widehat{D}(\cdot)$ does not depend on $\widehat{\gamma}$ (or $\widehat{r}$), and hence it is still consistent under the alternative. For $\widehat{\theta} = (\widehat{\beta}^\intercal, \widehat{\delta}^\intercal)^\intercal$, in addition, we can verify that there exists a constant $C_\theta \in [0, \infty)$ such that

$$n^\epsilon(\widehat{\theta} - \theta_0) = C_\theta + o_p(1) \tag{A.17}$$

for any given $\widehat{\gamma}$ (or $\widehat{r}$). In particular, denote $X_i(\gamma) = (x_i^\intercal, x_i^\intercal \mathbf{1}[q_i < \gamma])^\intercal$ and $X_i(\gamma_i) = (x_i^\intercal, x_i^\intercal \mathbf{1}[q_i < \gamma_i])^\intercal$. Given $\widehat{\gamma} = \gamma$ for any γ,

$$\begin{aligned} n^\epsilon\left(\widehat{\theta} - \theta_0\right) &= n^\epsilon\left(\sum_{i=1}^n X_i(\gamma)X_i(\gamma)^\intercal\right)^{-1}\left(\sum_{i=1}^n X_i(\gamma)\left\{y_i - X_i(\gamma)^\intercal\theta_0\right\}\right) \\ &= \left(\frac{1}{n}\sum_{i=1}^n X_i(\gamma)X_i(\gamma)^\intercal\right)^{-1}\left(\frac{n^\epsilon}{n}\sum_{i=1}^n X_i(\gamma)u_i + \frac{n^\epsilon}{n}\sum_{i=1}^n X_i(\gamma)\left(X_i(\gamma_i) - X_i(\gamma)\right)^\intercal\theta_0\right) \\ &\equiv \Theta_{n1}^{-1}\left(\Theta_{n2} + \Theta_{n3}\right). \end{aligned}$$

Similarly as Lemma A.5 of Lee and Wang (2019), we have $\widehat{\Theta}_{n1} \to_p \Theta_1$, which is positive definite by Condition 1.7. For the numerator, since $n^{1/2-\epsilon}\widehat{\Theta}_{n2} = O_p(1)$ by the standard Central Limit Theorem, we have $\widehat{\Theta}_{n2} = o_p(1)$ as $\epsilon \in (0, 1/2)$ in Condition 1.4. Furthermore, since $\delta_0 = c_0 n^{-\epsilon}$ with $c_0 \neq 0$, we have $\widehat{\Theta}_{n3} = O_p(1)$ at most from Conditions 1.4, 5 and 7, though it can be $o_p(1)$ under some special circumstances.

Let $r_{[i]}$ be the induced order statistics of $F(\gamma_i)$ associated with $q_{(i)}$. We decompose

$$\begin{aligned}
\widehat{\mathcal{G}}_{vn}(s) &= \frac{\widehat{h}_\tau^{1/2}}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor \widehat{g}^{-1}(s)n \rfloor} \widehat{g}^{(1)}(i/n) v^\intercal \widehat{D}(i/n)^{-1} x_{[i]} \widehat{u}_{[i]} \\
&= \frac{\widehat{h}_\tau^{1/2}}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor \widehat{g}^{-1}(s)n \rfloor} \widehat{g}^{(1)}(i/n) v^\intercal \widehat{D}(i/n)^{-1} x_{[i]} u_{[i]} \\
&\quad - \frac{\widehat{h}_\tau^{1/2}}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor \widehat{g}^{-1}(s)n \rfloor} \widehat{g}^{(1)}(i/n) v^\intercal \widehat{D}(i/n)^{-1} x_{[i]} x_{[i]}^\intercal \{(\widehat{\beta} - \beta_0) + \mathbf{1}[i/n \le \widehat{r}](\widehat{\delta} - \delta_0)\} \\
&\quad - \frac{\widehat{h}_\tau^{1/2}}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor \widehat{g}^{-1}(s)n \rfloor} \widehat{g}^{(1)}(i/n) v^\intercal \widehat{D}(i/n)^{-1} x_{[i]} x_{[i]}^\intercal \delta_0 \left(\mathbf{1}[i/n \le \widehat{r}] - \mathbf{1}\left[i/n \le r_{[i]}\right] \right) \\
&\equiv \widehat{C}_{1n}(s) - \widehat{C}_{2n}(s) - \widehat{C}_{3n}(s),
\end{aligned}$$

and denote their re-scaled and demeaned terms as in (17) as

$$\widehat{\mathcal{G}}_n^*(s) = \widehat{C}_{1n}^*(s) - \widehat{C}_{2n}^*(s) - \widehat{C}_{3n}^*(s).$$

The first $\widehat{C}_{1n}^*(s)$ term is $O_p(1)$ because $\widehat{C}_{1n}(s) = O_p(1)$ given Theorem 1, where the probability limits of $\widehat{h}_\tau$, $\widehat{g}^{(1)}(\cdot)$ are all still bounded and $\widehat{g}(\widehat{r}) \to_p \bar{g} \in [0,1]$ as $n \to \infty$ though $\bar{g}$ is not necessarily the same as $g(r_0)$. For $\widehat{C}_{2n}^*(s)$, since $\widehat{D}(\cdot)$ is still uniformly consistent, a similar argument as Lemma A.5 implies that, for any $s \in [\tau, 1-\tau]$,

$$\frac{1}{n} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor \widehat{g}^{-1}(s)n \rfloor} \widehat{g}^{(1)}(i/n) \widehat{D}(i/n)^{-1} x_{[i]} x_{[i]}^\intercal \to_p s,$$

$$\frac{1}{n} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor \widehat{g}^{-1}(s)n \rfloor} \widehat{g}^{(1)}(i/n) v^\intercal \widehat{D}(i/n)^{-1} x_{[i]} x_{[i]}^\intercal \mathbf{1}[i/n \le r] \to_p \min\{s, \bar{g}\}$$

as $n \to \infty$, which yields

$$\widehat{C}_{2n}(s) = (s + o_p(1))\, n^{1/2}(\widehat{\beta} - \beta_0) + (\min\{s, \bar{g}\} + o_p(1))\, n^{1/2}(\widehat{\delta} - \delta_0) = O_p\left(n^{1/2-\epsilon}\right)$$

since $\widehat{\theta} - \theta_0 = O_p(n^{-\epsilon})$ from (A.17). However, as $\widehat{C}_{2n}(s)$ is linear in s, the re-scaling and demeaning procedure eliminates the leading term and hence we have $\widehat{C}_{2n}^*(s) = o_p\left(n^{1/2-\epsilon}\right)$. This result holds naturally when $\widehat{\theta} - \theta_0 = o_p(n^{-\epsilon})$.

Lastly, the fact that $r_{[i]}$ is a non-degenerate random variable implies

$$\frac{1}{n}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor \widehat{g}^{-1}(s)n\rfloor}\widehat{g}^{(1)}(i/n)v^{\intercal}\widehat{D}\left(i/n\right)^{-1}x_{[i]}x_{[i]}^{\intercal}\left(\mathbf{1}\left[i/n\leq\widehat{r}\right]-\mathbf{1}\left[i/n\leq r_{[i]}\right]\right)=O_{p}(1).$$

Furthermore, since we suppose the support of γ_i is located in the interior of the support of q_i (i.e., Condition 1.5 holds for any values of r_i), $\mathbf{1}\left[q_i<\widehat{\gamma}\right]-\mathbf{1}\left[q_i<\gamma_i\right]$ or equivalently $\mathbf{1}\left[i/n\leq\widehat{r}\right]-\mathbf{1}\left[i/n\leq r_{[i]}\right]$ cannot be zero for all i at the same time unless $\widehat{\gamma}$ locates at the boundary of the support of q_i (or $\widehat{r}$ is either 0 or 1), which is excluded in our case. Hence $\widehat{C}_{3n}\left(s\right)=O_p\left(n^{1/2-\epsilon}\right)$ as $\delta_0=c_0n^{-\epsilon}$ with $c_0\neq 0$ and $\mathbb{E}[x_ix_i^{\intercal}]>\mathbb{E}\left[x_ix_i^{\intercal}\mathbf{1}\left[F\left(q_i\right)\leq r\right]\right]>0$ for any r from Condition 1.7. In this case, even the re-scaling and demeaning procedure cannot eliminate the leading $O_p\left(n^{1/2-\epsilon}\right)$ term unless $r_i=r_0$ for all i.[7] It follows that $\widehat{C}_{3n}^{*}\left(s\right)=O_p\left(n^{1/2-\epsilon}\right)$, which dominates $\widehat{\mathcal{G}}_{vn}^{*}\left(\cdot\right)$. Therefore, since $\epsilon\in(0,1/2)$, $\widehat{\mathcal{G}}_{vn}^{*}\left(\cdot\right)$ diverges and hence $CT_n\to\infty$ with probability approaching to one under the alternative hypothesis. ■

Proof of Lemma 2 Let r_g be uniformly distributed over $[\tau,1-\tau]$ and $c_v|(r_g=s)\sim\mathcal{N}\left(0,\kappa\left(s\right)\right)$ for some function $\kappa\left(s\right)$ to be specified later. Then the weighted likelihood ratio test statistic reads

$$WLR=\frac{1}{1-2\tau}\int_{\tau}^{1-\tau}\int\phi\left(\frac{c_v}{\sqrt{\kappa\left(s\right)}}\right)\exp\left(c_v\mathcal{G}_v^{*}(s)-\frac{c_v^2}{2}s\left(1-s\right)\right)dsdc_v,$$

where $\phi\left(\cdot\right)$ denotes the standard normal density function. Denote $a=\mathcal{G}_v^{*}(s)$ and $b=\kappa\left(s\right)^{-1}+s\left(1-s\right)$. Then, we have

$$\begin{aligned}\phi\left(\frac{c_v}{\sqrt{\kappa\left(s\right)}}\right)\exp\left(c_v\mathcal{G}_v^{*}(s)-\frac{c_v^2}{2}s\left(1-s\right)\right)&=\frac{1}{\sqrt{2\pi\kappa\left(s\right)}}\exp\left(-\frac{c_v^2}{2}b+c_va\right)\\&=\frac{1}{\sqrt{2\pi\kappa\left(s\right)}}\exp\left(-\frac{1}{2}b\left(\frac{a}{b}-c_v\right)^2+\frac{1}{2}\frac{a^2}{b}\right).\end{aligned}$$

Using the fact that a density integrates to 1, we find that

$$\int\phi\left(\frac{c_v}{\sqrt{\kappa\left(s\right)}}\right)\exp\left(c_v\mathcal{G}_v^{*}(s)-\frac{c_v^2}{2}s\left(1-s\right)\right)dc_v=\frac{1}{\sqrt{b\kappa\left(s\right)}}\exp\left(\frac{1}{2}\frac{a^2}{b}\right).$$

[7] For instance, consider the case with two thresholds, r_1 and r_2 with $r_1\neq r_2$. Even when $\widehat{r}$ consistently estimates one threshold, say r_1, and the re-scaling and demeaning procedure in (17) is defined using $\widehat{r}$, one of the right or left sides of r_1 still has a jump at r_2. Because of this nonlinearity, the re-scaling and demeaning procedure cannot completely eliminate the leading $O_p\left(n^{1/2-\epsilon}\right)$ term asymptotically.

Setting $\kappa(s) = \omega^2(s(1-s))^{-1}$ for some constant $\omega^2 > 0$ yields that

$$WLR = \frac{1}{1-2\tau}\int_{\tau}^{1-\tau} \frac{1}{\sqrt{1+\omega^2}} \exp\left(\frac{1}{2}\frac{\omega^2}{1+\omega^2}\frac{\mathcal{G}_v^*(s)^2}{s(1-s)}\right) ds.$$

Then following Andrews and Ploberger (1994), we have

$$\lim_{\omega^2\to 0} \frac{2\left(\sqrt{1+\omega^2}WLR - 1\right)}{\omega^2} = \frac{1}{1-2\tau}\int_{\tau}^{1-\tau} \frac{\mathcal{G}_v^*(s)^2}{s(1-s)} ds$$

as desired. ■

Proof of Theorem 4 We first show $\widehat{\mathcal{G}}_{vn}(s) \Rightarrow \mathcal{G}_v(s)$ for $s \in [0,1]$, where

$$\widehat{\mathcal{G}}_{vn}(s) = \frac{\widehat{h}_\tau^{1/2}}{\sqrt{n}} \sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor \widehat{g}^{-1}(s)n\rfloor} \widehat{g}^{(1)}(i/n) v^\intercal \widehat{D}(i/n)^{-1} x_{[i]}\widehat{u}_{[i]}$$

with $\widehat{u}_i = u_i - x_i(\widehat{\beta} - \beta) + c_0 n^{-1/2}\mathbf{1}[q_i \le \gamma_0]$. To this end, we go through the proof of Lemma 1 under $\delta_0 = c_0 n^{-1/2}$. For simplicity, we present the case for a scalar x_i (so $v = 1$). First, in view of the proof of Lemma A.2, we replace (A.5) by

$$\begin{aligned}
\left|\frac{1}{n}\sum_{i=1, i\neq \lfloor rn\rfloor}^{n} x_{[i]}^2\left(\widehat{u}_{[i]} + u_{[i]}\right)\left(\widehat{u}_{[i]} - u_{[i]}\right) K_i(r)\right| &\le \frac{1}{n}\sum_{i=1}^{n}\left|x_i^3\left(\widehat{u}_i + u_i\right)(\widehat{\beta} - \beta_0) K_i(r)\right| \\
&\quad + \frac{1}{n^{3/2}}\sum_{i=1}^{n}\left|x_i^3\left(\widehat{u}_i + u_i\right) c_0 \mathbf{1}[q_i \le \gamma_0] K_i(r)\right| \\
&\equiv M_{1n}(r) + M_{2n}'(r).
\end{aligned}$$

Then by the same argument as the proof of Lemma A.2, $M_{1n}(r)$ and $M_{2n}'(r)$ are both uniformly $o_p(1)$ over $r \in [\tau, 1-\tau]$.

Second, Lemma A.3 holds identically since it does not rely on the magnitude of δ_0. Third, we establish Lemma A.5. Substituting $\widehat{u}_i$, we obtain

$$\begin{aligned}
\mathcal{G}_{vn}(s) &= \frac{\widehat{h}_\tau^{1/2}}{\sqrt{n}} \sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} g^{(1)}(i/n) D(i/n)^{-1} x_{[i]} u_{[i]} \\
&\quad - \frac{\widehat{h}_\tau^{1/2}}{n} \sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} g^{(1)}(i/n) D(i/n)^{-1} x_{[i]} x_{[i]}^\intercal \sqrt{n}(\widehat{\beta} - \beta_0)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\widehat{h}_{\tau}^{1/2}}{n}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} g^{(1)}(i/n)D\left(i/n\right)^{-1}x_{[i]}x_{[i]}^{\intercal}\sqrt{n}\delta_0\mathbf{1}\left[q_{(i)}\le\gamma_0\right]\\
\equiv\;& A_{1n}\left(s\right)-A_{2n}\left(s\right)+A_{3n}'\left(s\right).
\end{aligned}
$$

By the same argument as the proof of Lemma A.5, the $A_{1n}\left(s\right)$ and $A_{2n}\left(s\right)$ terms have the same limits as before. Regarding $A_{3n}'\left(s\right)$, since $\delta_0=c_0n^{-1/2}$, we apply Lemma A.1 and the continuous mapping theorem to obtain

$$
\begin{aligned}
A_{3n}'\left(s\right)&=\frac{\widehat{h}_{\tau}^{1/2}}{n}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} g^{(1)}(i/n)D\left(i/n\right)^{-1}x_{[i]}x_{[i]}^{\intercal}\mathbf{1}\left[i/n\le r_0\right]\sqrt{n}\delta_0\\
&\to_p h_{\tau}^{1/2}\left(\int_{g^{-1}(0)}^{\min(g^{-1}(s),r_0)} g^{(1)}(t)D\left(t\right)^{-1}D\left(t\right)dt\right)c_0\\
&=\min\{s,g\left(r_0\right)\}c_0h_{\tau}^{1/2}.
\end{aligned}
$$

Finally, in view of the proof of Lemma 1, the B_{11n} term in (A.14) and the B_{21n} term in (A.15) remain unchanged. For B_{12n}, Lemmas A.1 and A.3 and the fact that $\widehat{\beta}-\beta_0=O_p\left(n^{-1/2}\right)$ yield that

$$
\begin{aligned}
&\sup_{s\in[0,1]}\left|B_{12n}(s)\right|\\
&\le h_{\tau}^{1/2}\sup_{r\in[\tau,1-\tau]}\left|\widehat{g}^{(1)}(r)\widehat{D}\left(r\right)^{-1}-g^{(1)}(r)D\left(r\right)^{-1}\right|\sup_{r\in[\tau,1-\tau]}\frac{1}{\sqrt{n}}\left|\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor rn\rfloor}x_{[i]}(\widehat{u}_{[i]}-u_{[i]})\right|\\
&\le o_p(1)\times\left\{\sup_{r\in[\tau,1-\tau]}\frac{1}{\sqrt{n}}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor rn\rfloor}x_{[i]}^2|\widehat{\beta}-\beta_0|+\sup_{r\in[\tau,1-\tau]}\frac{1}{n}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor rn\rfloor}x_{[i]}^2\left|c_0\right|\right\}\\
&=o_p(1).
\end{aligned}
$$

For B_{22n}, consider the events $E_{\beta n}=\{\widehat{\beta}\in\mathcal{B}_{Cn^{-1/2}}(\beta_0)\}$ for some $C>0$ and $E_{gn}=\{\sup_{s\in[0,1]}|\widehat{g}^{-1}(s)-g^{-1}(s)|<\eta\}$ for some $\eta>0$. Then, on the these two events, Condition 1.6 and Lemma A.1 yield that

$$
\begin{aligned}
\sup_{s\in[0,1]}\left|B_{22n}(s)\right|&\le h_{\tau}^{1/2}\sup_{r\in[\tau,1-\tau]}\left|g^{(1)}(r)D\left(r\right)^{-1}\right|\sup_{s\in[0,1]}\frac{1}{\sqrt{n}}\sum_{i=\lfloor \widehat{g}^{-1}(s)n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor}\left|x_{[i]}(u_{[i]}-\widehat{u}_{[i]})\right|\\
&\le C\sup_{s\in[0,1]}\frac{1}{\sqrt{n}}\sum_{i=\lfloor (g^{-1}(s)-\eta)n\rfloor+1}^{\lfloor (g^{-1}(s)+\eta)n\rfloor}\left|x_{[i]}^2\right|\left(|\widehat{\beta}-\beta_0|+n^{-1/2}\left|c_0\right|\right)
\end{aligned}
$$

$$\leq C \sup_{s\in[0,1]} \frac{1}{n} \sum_{i=\lfloor (g^{-1}(s)-\eta)n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} x_{[i]}^2$$
$$\to_p C \sup_{s\in[0,1]} \int_{g^{-1}(s)-\eta}^{g^{-1}(s)} D(t)\, dt$$

for some constant $0 < C < \infty$. Then $\sup_{s\in[0,1]} |B_{22n}(s)| = o_p(1)$ by choosing a sufficiently small η. We thus establish $\widehat{\mathcal{G}}_{vn}(s) \Rightarrow \mathcal{G}_v(s)$ for $s \in [0,1]$ by combining the above four steps. The rest of the proof follows immediately from the continuous mapping theorem. ■

References

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

Andrews, D. W. K., and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

Bae, J., D. Jun, and S. Levental (2010): "The Uniform CLT for Martingagle Difference Arrays under the Uniformly Integrable Entropy," *Bulletin of the Korean Mathematical Society*, 47(1), 39–51.

Bai, J., and P. Perron (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

Bhattacharya, P. K. (1974): "Convergence of Sample Paths of Normalized Sums of Induced Order Statistics," *The Annals of Statistics*, 2(5), 1034–1039.

——— (1984): "18 Induced Order Statistics: Theory and Applications," *Handbook of Statistics*, 4, 383–403.

Caner, M., and B. E. Hansen (2004): "Instrumental Variable Estimation of a Threshold Model," *Econometric Theory*, 20, 813–843.

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

Chan, K. S. (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *Annals of Statistics*, 21, 520–533.

Davidson, J. (1994): *Stochastic Limit Theory*. Oxford University Press, New York.

Elliott, G., and U. K. Müller (2007): "Confidence Sets for the Date of a Single Break in Linear Time Series Regressions," *Journal of Econometrics*, 141, 1196–1218.

——— (2014): "Pre and Post Break Parameter Inference," *Journal of Econometrics*, 180, 141–157.

Elliott, G., U. K. Müller, and M. W. Watson (2015): "Nearly Optimal Tests When a Nuisance Parameter is Present under the Null Hypothesis," *Econometrica*, 83, 771–811.

Gonzalo, J., and J. Pitarakis (2002): "Estimation and Model Selection Based Inference in Single and Multiple Threshold Models," *Journal of Econometrics*, 110, 319–352.

Hansen, B. E. (1997): "Inference in TAR Models," *Studies in Nonlinear Dynamics & Econometrics*, 2(1), 1–14.

——— (2000a): "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 575–603.

——— (2000b): "Testing for Structural Change in Conditional Models," *Journal of Econometrics*, 97, 93–115.

Hidalgo, J., J. Lee, and M. H. Seo (2019): "Robust Inference for Threshold Regression Models," *Journal of Econometrics*, 210, 291–309.

Lee, S., Y. Liao, M. H. Seo, and Y. Shin (2018): "Factor-driven Two-regime Regression," *Working paper.*

Lee, S., M. H. Seo, and Y. Shin (2011): "Testing for Threshold Effects in Regression Models," *Journal of the American Statistical Association*, 106(493), 220–231.

Lee, Y., and Y. Wang (2019): "Nonparametric Sample Splitting," *Working Paper.*

Li, H., and S. Ling (2012): "On the Least Squares Estimation of Multiple-Regime Threshold Autoregressive Models," *Journal of Econometrics*, 1, 240–253.

Li, H., and U. K. Müller (2009): "Valid Inference in Partially Unstable General Method of Moment Models," *Review of Economic Studies*, 76, 343–365.

Li, Q., and J. S. Racine (2007): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Liptser, R., and N. Shiryaev (2013): *Statistics of Random Processes: I. General Theory*, vol. 5. Springer Science & Business Media.

Nyblom, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

Park, J., and P. Phillips (1999): "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric theory*, 15, 269–298.

Schelling, T. C. (1971): "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, 1(2), 143–186.

Seo, M. H., and O. Linton (2007): "A Smooth Least Squares Estimator for Threshold Regression Models," *Journal of Econometrics*, 141(2), 704–735.

van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, New York.

Yang, S. S. (1981): "Linear Functions of Concomitants of Order Statistics with Application to Nonparametric Estimation of a Regression Function," *Journal of the American Statistical Association*, 76(375), 658–662.

——— (1985): "A Smooth Nonparametric Estimator of a Quantile Function," *Journal of the American Statistical Association*, 80(392), 1004–1011.

Yu, P. (2012): "Likelihood Estimation and Inference in Threshold Regression," *Journal of Econometrics*, 167, 274–294.

Yu, P., and X. Fan (2019): "Threshold regression with a threshold Boundary," *Working Paper.*