Testing for Homogeneous Thresholds in Threshold Regression Models*

Yoonseok Lee† Yulong Wang‡

Syracuse University Syracuse University

June 2021

Abstract

This paper develops a test for homogeneity of the threshold parameter in threshold regression models. The test has a natural interpretation from time series perspectives and can be also applied to test for additional change points in the structural break models. The limiting distribution of the test statistic is derived, and the finite sample properties are studied in Monte Carlo simulations. We apply the new test 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.

Keywords: threshold regression; test; homogeneous threshold; tipping point

JEL Classifications: C12, C24

*This paper was previously circulated under the title "Inference in Threshold Models". The authors thank Ulrich Müller, Bo Honoré, Mark Watson, Kirill Evidokimov, Simon Lee, Myung Seo, Zhijie Xiao, and participants 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.

†*Address*: Department of Economics and Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244. *E-mail*: `ylee41@syr.edu`

‡*Address*: Department of Economics and Center for Policy Research, Syracuse University, 127 Eggers Hall, Syracuse, NY 13244. *E-mail*: `ywang402@syr.edu`

1 Introduction

Threshold regression models have been widely used and studied in economics and statistics. Most of the existing studies focus on estimating parameters in a given threshold regression model and testing for the threshold effect. However, once tests support the existence of the coefficient change, especially in the cross-sectional threshold models, it is natural to ask whether all the agents share the same threshold location. This paper answers this question by developing a homogeneity test of the threshold parameter (i.e., a constant threshold).

The test is motivated by the tipping point problem (e.g., Schelling (1971)), which analyzes the phenomenon that the neighborhood's white population substantially decreases once the minority share exceeds a certain threshold. Card, Mas, and Rothstein (2008) empirically study this phenomenon by considering the following threshold regression model:

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

for tracts $i = 1, \ldots, n$, 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), when the tipping point feature exists, one may want to examine if the tipping point is the same across tracts. In fact, Card, Mas, and Rothstein (2008) regress the estimated γ_0 on a measure of the white population's attitude to the minority at the aggregated level (more precisely at the city level) and find that the tipping point highly varies across this measure. This finding raises the concern that γ_0 may also vary across tracts depending on some demographics and motivates our constant-threshold test, the CT test, for the homogeneity of γ_0.

More specifically, we develop a test for a constant threshold γ_0 against nonparametric alternatives (or any types of heterogeneous thresholds) with cross-sectional data. In the event of rejection, therefore, one can resort to more flexible models such as those studied by Lee and Wang (2019), Yu and Fan (2020), and Lee, Liao, Seo, and Shin (2020). Furthermore, one could apply the estimation and testing methods proposed by Miao, Su, and Wang (2020) if panel data are available. In this sense, the new CT test can be used as a diagnostic tool for model specification in the threshold regression setup. In the tipping point application, the CT test strongly rejects the null hypothesis of the constant threshold, implying that the model (1) is insufficient to characterize the tipping point phenomenon. See Section 5 for more details.

Our new test statistic builds on a weighted summation of the regression residuals under the null hypothesis of a constant threshold, where the weights are designed to yield a simple limit experiment as exploited by Nyblom (1989), Elliott and Müller (2007) and Elliott and Müller (2014).

By converting the weighted summation into a partial sum process, we bridge the cross-sectional threshold model and the time series change-point model in this testing problem. Hence, the CT test can also be applied to test for any additional change points in the structural break models if we let q_i be the time and γ_0 the break date.

This paper speaks to both the threshold regression and the time series structural break literature. The threshold model with a constant threshold has been extensively investigated. See, among many others, Hansen (2000), Caner and Hansen (2001, 2004), Seo and Linton (2007), Lee, Seo, and Shin (2011), Li and Ling (2012), Yu (2012), Kourtellos, Stengos, and Tan (2016), Yu and Phillips (2018), Hidalgo, Lee, and Seo (2019), and Miao, Su, and Wang (2020). In addition, Seo and Linton (2007), Lee, Liao, Seo, and Shin (2020), and Yu and Fan (2020) study the model where γ_0 has an index form that involves multiple covariates. This paper contributes to the literature by providing a diagnostic method for constancy of the threshold.

When q_i is the time, our method essentially becomes the structural break model. See, among many others, Bai and Perron (1998), Bai, Lumsdaine, and Stock (1998), Elliott and Müller (2007) and Elliott and Müller (2014). Methods in these papers are typically developed under the increasing domain asymptotics and we also develop our test under this classic framework. Alternatively, Jiang, Wang, and Yu (2018a,b) recently develop methods under the infill asymptotics. Casini and Perron (2020, 2021a,b) introduce the generalized Laplace estimation and inference and study a continuous record asymptotic framework.

The rest of the paper is organized as follows. Section 2 constructs the new test and shows the connection to the change-point problem in the time series setup. Section 3 studies the asymptotic properties of the new test. Section 4 examines its finite sample performance by Monte Carlo simulations. 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 $\to_p$ denote convergence in probability, $\to_d$ convergence in distribution, and $\Rightarrow$ weak convergence of stochastic processes as the sample size $n \to \infty$. Let $=_d$ denote equivalence in distribution. Let $\lfloor a \rfloor$ denote the biggest integer smaller than a, $\mathbf{1}[A]$ the indicator function of a generic event A, and $\|B\|$ the Euclidean norm of a vector or matrix B.

2 Testing for a Homogeneous Threshold

2.1 Setup

We consider the threshold regression model with a potentially heterogeneous threshold parameter, which is given by

$$y_i = x_i^\intercal \beta_0 + x_i^\intercal \delta_0 \mathbf{1}\left[q_i \leq \gamma_{0i}\right] + u_i \tag{2}$$

for $i = 1, \ldots, n$. The variables $(y_i, x_i^\intercal, q_i)^\intercal \in \mathbb{R}^{1+k+1}$ are observed but the threshold parameters $\gamma_{0i} \in \mathbb{R}$ as well as the regression coefficients $\theta_0 = (\beta_0^\intercal, \delta_0^\intercal)^\intercal \in \mathbb{R}^{2k}$ are unknown. The threshold γ_{0i} can be considered as a random variable or a constant. Under the assumption of a homogeneous threshold, say $\gamma_{0i} = \gamma_0$ almost surely, the model becomes the classic threshold regression model and all the parameters can be consistently estimated by the standard profile least squares method (e.g., Bai and Perron (1998) and Hansen (2000)). Specifically, under the homogeneous threshold restriction, we estimate γ_0 by minimizing

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

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 $\widehat{\theta} = (\widehat{\beta}^\intercal, \widehat{\delta}^\intercal)^\intercal = (\widehat{\beta}^\intercal(\widehat{\gamma}), \widehat{\delta}^\intercal(\widehat{\gamma}))^\intercal$ and write $\widehat{u}_i = y_i - x_i^\intercal \widehat{\beta} + x_i^\intercal \widehat{\delta} \mathbf{1} \left[q_i \leq \widehat{\gamma} \right]$ as the residual.

The main interest of this paper is to test whether the threshold is constant across entities or not. Let Γ be the space of γ_{0i}, which is assumed to be compact and strictly within the support of q_i. The competing hypotheses are stated as

$$\begin{aligned} H_0 &: \mathbb{P}\left(\gamma_{0i} = \gamma_0\right) = 1 \text{ for some constant } \gamma_0 \in \Gamma \\ H_1 &: \mathbb{P}\left(\gamma_{0i} = \gamma_0\right) < 1 \text{ for any } \gamma_0 \in \Gamma. \end{aligned} \tag{3}$$

Under the null hypothesis, there exists only one homogeneous threshold γ_0 and hence the model reduces to the classic threshold regression model as in (1). The alternative hypothesis in (3) states the negation of the null hypothesis and hence it is nonparametric, which encompasses many different cases. A straightforward example is when the threshold varies across i, covering the case that γ_{0i} is a continuous random variable. Moreover, γ_{0i} can be a non-constant function of some random variables z_i. Examples include an index form, $\gamma_{0i} = z_i^\intercal \gamma$ for some parameter γ, as in Yu and Fan (2020) and Lee, Liao, Seo, and Shin (2020); and even a nonparametric from, $\gamma_{0i} = \gamma(z_i)$ for some unknown function $\gamma(\cdot)$, as in Lee and Wang (2019).

It is worthy to note that the alternative hypothesis in (3) also includes the case with multiple thresholds that are the same for all i (cf. Bai and Perron (1998) in the structural break model). For instance, let

$$\gamma_{0i} = \begin{cases} \gamma_{0,1} & \text{with } \mathbb{P}\left(\gamma_{0i} = \gamma_{0,1}\right) = p_0 \\ \gamma_{0,2} & \text{with } \mathbb{P}\left(\gamma_{0i} = \gamma_{0,2}\right) = 1 - p_0 \end{cases} \tag{4}$$

for some $p_0 \in (0, 1)$. We define two random variables $\lambda_{i,1} = p_0 - \mathbf{1}\left[\gamma_{0i} = \gamma_{0,1}\right]$ and $\lambda_{i,2} = 1 - p_0 - \mathbf{1}\left[\gamma_{0i} = \gamma_{0,2}\right]$. Then,

$$\mathbf{1}\left[q_i \leq \gamma_{0i}\right] = \mathbf{1}\left[q_i \leq \gamma_{0,1}\right] \mathbf{1}\left[\gamma_{0i} = \gamma_{0,1}\right] + \mathbf{1}\left[q_i \leq \gamma_{0,2}\right] \mathbf{1}\left[\gamma_{0i} = \gamma_{0,2}\right]$$

$$= \mathbf{1}\left[q_i \le \gamma_{0,1}\right](p_0 - \lambda_{i,1}) + \mathbf{1}\left[q_i \le \gamma_{0,2}\right](1 - p_0 - \lambda_{i,2})$$

and the threshold regression model in (2) can be rewritten as

$$\begin{aligned} y_i &= x_i^\intercal \beta_0 + x_i^\intercal \delta_0 \mathbf{1}\left[q_i \le \gamma_{0,1}\right](p_0 - \lambda_{i,1}) + x_i^\intercal \delta_0 \mathbf{1}\left[q_i \le \gamma_{0,2}\right](1 - p_0 - \lambda_{i,2}) + u_i \\ &= x_i^\intercal \beta_0 + x_i^\intercal \delta_{0,1}^* \mathbf{1}\left[q_i \le \gamma_{0,1}\right] + x_i^\intercal \delta_{0,2}^* \mathbf{1}\left[q_i \le \gamma_{0,2}\right] + u_i^*, \end{aligned} \tag{5}$$

where $\delta_{0,1}^* = \delta_0 p_0$, $\delta_{0,2}^* = \delta_0(1 - p_0) = \delta_0 - \delta_{0,1}^*$, and

$$u_i^* = u_i - x_i^\intercal \delta_0 \left\{\mathbf{1}\left[q_i \le \gamma_{0,1}\right]\lambda_{i,1} + \mathbf{1}\left[q_i \le \gamma_{0,2}\right]\lambda_{i,2}\right\}.$$

It holds that $\mathbb{E}[u_i^*|x_i, q_i] = 0$ when $\mathbb{E}[\lambda_{i,1}|x_i, q_i] = \mathbb{E}[\lambda_{i,2}|x_i, q_i] = \mathbb{E}[u_i|x_i, q_i] = 0$.

This example illustrates that the threshold regression model with a heterogeneous threshold as in (4) can be rewritten as the threshold regression model with two homogeneous thresholds as in (5). In this regard, the alternative hypothesis in (3) amounts to characterizing the scenario where additional coefficient changes exist beyond the original change at γ_0 and can be applied to either the entire sample or only for a subsample. We hence can construct a test for (3) using the idea of Nyblom (1989) and Elliott and Müller (2007) in the change-point problem, where we test for the existence of additional changes before or after the location γ_0. The true threshold γ_0 is not given in the null hypothesis in (3), so we need to consistently estimate it. The key merit of this approach is that our test does not require to specify or estimate the alternative model, unlike the likelihood-ratio tests (e.g., Andrews (1993); Bai and Perron (1998); Lee, Seo, and Shin (2011)).

2.2 Overview of the Test

Here we summarize our test and heuristically present its statistic properties. The formal derivations are postponed to Section 3. First, under the mild primitive conditions given in Section 3.1, 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 (2000))

$$\sqrt{n}\begin{pmatrix} \widehat{\beta} - \beta_0 \\ \widehat{\delta} - \delta_0 \end{pmatrix} \to_d \begin{pmatrix} \Phi_\beta \\ \Phi_\delta \end{pmatrix} \tag{6}$$

as $n \to \infty$ for some k-dimensional normal random vectors Φ_β and Φ_δ. Denote $Q(\cdot)$ as the quantile function of q_i, and define the process $G_n(r) = n^{-1/2} \sum_{i=1}^n x_i \widehat{u}_i \mathbf{1}[q_i \le Q(r)]$. Also define $r_0 \in [0, 1]$ such that $\gamma_0 = Q(r_0)$. For $r \in [0, 1]$, using the standard empirical process results (e.g., van der

Vaart and Wellner (1996) and Kosorok (2008)), we can obtain that

$$\begin{aligned} G_n(r) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n x_i \widehat{u}_i \mathbf{1}[q_i \le Q(r)] \qquad (7)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n x_i u_i \mathbf{1}[q_i \le Q(r)] - \frac{1}{n}\sum_{i=1}^n x_i x_i^\intercal \mathbf{1}[q_i \le Q(r)] \cdot \sqrt{n}(\widehat{\beta}-\beta_0)\\ &\quad -\frac{1}{n}\sum_{i=1}^n x_i x_i^\intercal \mathbf{1}[q_i \le Q(r)]\mathbf{1}[q_i \le Q(r_0)] \cdot \sqrt{n}(\widehat{\delta}-\delta_0) + o_p(1)\\ &\Rightarrow J(r) - \mathbb{E}[x_i x_i^\intercal \mathbf{1}[q_i \le Q(r)]]\Phi_\beta - \mathbb{E}[x_i x_i^\intercal \mathbf{1}[q_i \le \min\{Q(r), Q(r_0)\}]]\Phi_\delta \end{aligned}$$

as $n \to \infty$, where $J(r)$ is a mean-zero Gaussian process[1] defined on $[0,1]$ and Φ_β and Φ_δ are as in (6). Note that we use the quantile function $Q(\cdot)$ in the definition of $G_n(\cdot)$ for the purpose of normalization, so that the process is defined on $[0,1]$.

If we further assume $x_i = 1$ and q_i is independent of u_i, the limiting expression in (7) can be simplified as

$$W_1(r) - r\Phi_\beta - \min\{r, r_0\}\Phi_\delta, \qquad (8)$$

where $W_1(\cdot)$ denotes the standard Wiener process defined on $[0,1]$. This is essentially the limit experiment exploited by the classic structural break literature, based on which Nyblom (1989) constructs the test statistic for an additional change point. In general, however, the limit of $G_n(\cdot)$ in (7) is more complicated since the process $J(\cdot)$ is not the standard Wiener process and the additional terms are not necessarily linear in r. For this reason, it is not straightforward to construct a test statistic directly based on (7).

We can recover the simple limit as in (8) by modifying $G_n(r)$ into a weighted-sum process. Suppose $r_0 \in [\tau, 1-\tau]$ for some $\tau \in (0, 1/2)$, so that we avoid the threshold $\gamma_0 = Q(r_0)$ being close to the boundary. Define[2]

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

and monotonically increasing functions

$$h(r) = \int_\tau^r \frac{1}{v^\intercal D(t)^{-1} V(t) D(t)^{-1} v} dt \quad \text{and} \quad g(r) = \frac{h(r)}{h(1-\tau)} \qquad (9)$$

[1] See Lemma A.4 in Hansen (2000). Its variance-covariance kernel is given in Lemmas A.2 and A.3 in the Appendix.

[2] The threshold regression literature typically uses $q_i = q$ as the index for presentation. We use the alternative presentation so that $D(\cdot)$ and $V(\cdot)$ are defined on $[0,1]$.

for $r \in [\tau, 1-\tau]$ and for any $k \times 1$ vector v that satisfies $v^\intercal v = 1$.[3] Also let $F(\cdot)$ be the distribution function of q_i and define the $k \times 1$ vector of weight

$$w_i = \sqrt{h(1-\tau)} g^{(1)}(F(q_i)) D\left(F(q_i)\right)^{-1} v,$$

where $g^{(1)}(r) = \partial g(r)/\partial r = \{v^\intercal D(r)^{-1} V(r) D(r)^{-1} v h(1-\tau)\}^{-1}$. The modified process is then constructed as

$$\mathcal{G}_n(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i^\intercal x_i \widehat{u}_i \mathbf{1}\left[Q(\tau) \leq q_i \leq Q(g^{-1}(s))\right] \tag{10}$$

for $s \in [0,1]$, where $g^{-1}(\cdot)$ is the inverse function of $g(\cdot)$. Comparing $\mathcal{G}_n(\cdot)$ with $G_n(\cdot)$, the key difference is in two-fold: the weight vector w_i and the indicator function. The intuition for constructing such $\mathcal{G}_n$ is better presented from a time series structural break perspective, which is given in the next subsection. Under the conditions given in Section 3.1, we can show that under the null hypothesis:

$$\mathcal{G}_n(s) \Rightarrow W_1(s) - s\sqrt{h(1-\tau)} v^\intercal \Phi_\beta - \min\{s, g(r_0)\} \sqrt{h(1-\tau)} v^\intercal \Phi_\delta \tag{11}$$

for $s \in [0,1]$ as $n \to \infty$, where $W_1(s)$ is the standard Wiener process on $[0,1]$. See Lemma A.1 for a formal statement. Except for the normalizing constant $\sqrt{h(1-\tau)}$, $\mathcal{G}_n$ now weakly converges to the simple limit as in (8).

To construct a pivotal test statistic using $\mathcal{G}_n$, we further define

$$\mathcal{G}_n^*(s) = \begin{cases} \mathcal{G}_{1n}^*(s) & \text{if } s \leq g(r_0) \\ \mathcal{G}_{2n}^*(s) & \text{otherwise,} \end{cases} \tag{12}$$

where

$$\begin{aligned} \mathcal{G}_{1n}^*(s) &= \frac{1}{\sqrt{g(r_0)}} \left\{ \mathcal{G}_n(s) - \frac{s}{g(r_0)} \mathcal{G}_n(g(r_0)) \right\}, \\ \mathcal{G}_{2n}^*(s) &= \frac{1}{\sqrt{1-g(r_0)}} \left\{ (\mathcal{G}_n(1) - \mathcal{G}_n(s)) - \frac{1-s}{1-g(r_0)} (\mathcal{G}_n(1) - \mathcal{G}_n(g(r_0))) \right\}. \end{aligned} \tag{13}$$

Then $\mathcal{G}_{1n}^*(\cdot)$ and $\mathcal{G}_{2n}^*(\cdot)$ are respectively properly standardized, and hence both weakly converge to two independent standard Brownian Bridge processes. Based on this observation, we construct the test statistic as

$$\frac{1}{g(r_0)} \int_0^{g(r_0)} \mathcal{G}_{1n}^*(s)^2 \, ds + \frac{1}{1-g(r_0)} \int_{g(r_0)}^1 \mathcal{G}_{2n}^*(s)^2 \, ds. \tag{14}$$

[3] The choice of v can be guided by the empirical context to reflect importance attached to different components of the changing coefficients. In the tipping point application, for instance, we use $v = (1, 0, ..., 0)^\intercal$.

As shown in Theorem 1 in Section 3, its limiting null distribution is free of nuisance parameters because the break is only at $s = g(r_0)$. This limiting distribution is also obtained by Elliott and Müller (2007) in the time series structural break setup. Therefore, the critical values can be readily tabulated and no further simulation or bootstrap is needed to conduct the test. Under the alternative hypothesis, when the constant threshold assumption is violated, however, at least one of $\mathcal{G}_{1n}^*(\cdot)$ and $\mathcal{G}_{2n}^*(\cdot)$ are not properly centered, and the test statistic diverges as $n \to \infty$ because of the non-zero drift. See Section 3 ahead.

2.3 Interpretation from Time Series Perspectives

As discussed above, our test uses the idea by Nyblom (1989) and Elliott and Müller (2007), which was originally developed in the time series context where q_i is the time and the observations are obtained sequentially over time. In this subsection, we reformulate the threshold regression into the change-point model and describe the connection between our test with Nyblom (1989) and Elliott and Müller (2007). Instead of deriving the limiting null distribution using the standard empirical process theory (cf., Lee, Seo, and Shin (2011)), we can construct a partial sum process in our setup and obtain the identical limiting null distribution based on the traditional stochastic process results. By doing so, we bridge the cross-sectional threshold model and the time series change-point model in this testing problem. Furthermore, viewing through the time series lens, we can provide a better intuition about how to construct w_i and $g(\cdot)$.

To this end, we first sort the observations according the order of q_i. By sorting the random sample $\{q_i\}_{i=1}^n$ into the order statistics $q_{(1:n)} \leq q_{(2:n)} \leq ... \leq q_{(n:n)}$ and re-arranging the observations according to the rank of q_i, 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 $q_{(i:n)} = q_j$.[4] These re-ordered statistics are called induced order statistics or concomitants (e.g., Bhattacharya (1974), Sen (1976), and Yang (1985)). It gives a natural ordering among the observations as in the time series structural break models, which is the case when $q_{(i:n)} = q_i = i$ 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 $q_{(i:n)}$.

In this setup, we can view the sorted uniform random variable $F(q_{(i)})$ as a "time" on the unit interval. For the empirical distribution $\widehat{F}_n(\cdot)$, $\widehat{F}_n(q_{(i)}) = i/n$ resembles the equi-spaced time on the unit interval from the perspective of structural break. In fact, Lemma A.3 in the Appendix shows that the effect of replacing $F(\cdot)$ by $\widehat{F}_n(\cdot)$ in two key elements $n^{-1/2}\sum_{i=1}^n x_i u_i \mathbf{1}[q_i \leq Q(r)] = n^{-1/2}\sum_{i=1}^n x_i u_i \mathbf{1}[F(q_i) \leq r]$ and $n^{-1}\sum_{i=1}^n x_i x_i^\intercal \mathbf{1}[q_i \leq Q(r)] = n^{-1}\sum_{i=1}^n x_i x_i^\intercal \mathbf{1}[F(q_i) \leq r]$ in (7)

[4] We suppose q_i is continuous, and the probability of seeing ties is thus negligible. In finite samples, we may simply drop duplicate (i.e., tied) observations of q_i.

are asymptotically negligible in the sense that

$$\sup_{r\in[0,1]}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_iu_i\mathbf{1}\left[q_i\leq Q(r)\right]-\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor rn\rfloor}x_{[i]}u_{[i]}\right\| = o_p\left(1\right), \tag{15}$$

$$\sup_{r\in[0,1]}\left\|\frac{1}{n}\sum_{i=1}^{n}x_ix_i^{\intercal}\mathbf{1}\left[q_i\leq Q(r)\right]-\frac{1}{n}\sum_{i=1}^{\lfloor rn\rfloor}x_{[i]}x_{[i]}^{\intercal}\right\| = o_p\left(1\right), \tag{16}$$

where $n^{-1/2}\sum_{i=1}^{\lfloor rn\rfloor}x_{[i]}u_{[i]}=n^{-1/2}\sum_{i=1}^{n}x_{[i]}u_{[i]}\mathbf{1}[\widehat{F}_n\left(q_{(i)}\right)\leq r]=n^{-1/2}\sum_{i=1}^{n}x_iu_i\mathbf{1}[\widehat{F}_n\left(q_i\right)\leq r]$ and similarly for $n^{-1}\sum_{i=1}^{\lfloor rn\rfloor}x_{[i]}x_{[i]}^{\intercal}$. Therefore, it is asymptotically equivalent to rewrite $\mathcal{G}_n\left(\cdot\right)$ in (10) using the partial sum process of the induced-order statistics and using $\widehat{F}_n(\cdot)$ in place of $F(\cdot)$ for implementation.

Then we can approximate $\mathcal{G}_n\left(\cdot\right)$ by

$$\frac{1}{\sqrt{n}}\sum_{i=\lfloor\tau n\rfloor+1}^{\left\lfloor g^{-1}(s)n\right\rfloor}\sqrt{h\left(1-\tau\right)}g^{(1)}(i/n)v^{\intercal}D(i/n)^{-1}x_{[i]}\widehat{u}_{[i]}, \tag{17}$$

and readily obtain its limit using the traditional stochastic process results from the decomposition of (17) as

$$\begin{aligned}
&= \frac{1}{\sqrt{n}}\sum_{i=\lfloor\tau n\rfloor+1}^{\left\lfloor g^{-1}(s)n\right\rfloor}\sqrt{h\left(1-\tau\right)}g^{(1)}(i/n)v^{\intercal}D(i/n)^{-1}x_{[i]}u_{[i]}\\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=\lfloor\tau n\rfloor+1}^{\left\lfloor g^{-1}(s)n\right\rfloor}\sqrt{h\left(1-\tau\right)}g^{(1)}(i/n)v^{\intercal}D(i/n)^{-1}x_{[i]}x_{[i]}^{\intercal}(\widehat{\beta}-\beta_0)\\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=\lfloor\tau n\rfloor+1}^{\min\{\left\lfloor g^{-1}(s)n\right\rfloor,\lfloor r_0n\rfloor\}}\sqrt{h\left(1-\tau\right)}g^{(1)}(i/n)v^{\intercal}D(i/n)^{-1}x_{[i]}x_{[i]}^{\intercal}(\widehat{\delta}-\delta_0)\\
&\Rightarrow \sqrt{h\left(1-\tau\right)}\int_{g^{-1}(0)}^{g^{-1}(s)}g^{(1)}(t)v^{\intercal}D\left(t\right)^{-1}V\left(t\right)^{1/2}dW_k\left(t\right) &(18)\\
&\quad -\sqrt{h\left(1-\tau\right)}\int_{g^{-1}(0)}^{g^{-1}(s)}g^{(1)}(t)dt\cdot v^{\intercal}\Phi_{\beta} &(19)\\
&\quad -\sqrt{h\left(1-\tau\right)}\int_{g^{-1}(0)}^{\min\{g^{-1}(s),r_0\}}g^{(1)}(t)dt\cdot v^{\intercal}\Phi_{\delta}, &(20)
\end{aligned}$$

where $D(i/n)^{-1}$ asymptotically cancels out with $\mathbb{E}[x_{[i]}x_{[i]}^{\intercal}]$ by construction. Then by the facts that $g\left(\tau\right)=0$ and $\int_0^{g^{-1}(s)}g^{(1)}(t)dt=s$, the terms in (19) and (20) become linear in s. To standardize the first term in (18), we set $g^{(1)}(\cdot)$ to be proportional to the inverse of the local Fisher information, $v^{\intercal} D(\cdot)^{-1} V(\cdot) D(\cdot)^{-1} v$. By doing so, the first term becomes the standard Wiener process. It follows that the limit of $\mathcal{G}_n(s)$ is obtained as (11) in the form of (8) as we claimed in the previous subsection, from which we can readily derive the limiting null distribution that is free from nuisance parameters. A formal statement is given in Lemma A.7 in the Appendix.

The merit of the partial sum process expression in (17) is now evident. First, the observations above explain how we develop the specific forms of the weight w_i and the function $g(\cdot)$ in (10). Note that $g : [\tau, 1-\tau] \mapsto [0,1]$ can be understood as the normalized time. In the structural break literature, in comparison, q_i becomes the time, and the functions $D(\cdot)$ and $V(\cdot)$ are respectively constant matrices $\bar{D}$ and $\bar{V}$ under the piece-wise stationarity assumption (e.g., Bai and Perron (1998)). Then $g(\cdot)$ reduces to the identity function, and the weight w_i becomes $v^{\intercal}\bar{D}^{-1}$. Second, we can readily derive the weak limit of $\mathcal{G}_n(\cdot)$ using the traditional stochastic process results, which naturally bridges the cross-sectional threshold model and the time series change-point model in our testing problem. Therefore, based on the discussion about the alternative hypothesis (3) in Section 2.1, the new test can also be applied to test for any additional change points in the structural break models in time series. Third, $\mathcal{G}_n$ is infeasible since the distribution F of q_i is unknown, whereas the feasible sample analog based on the partial sum process in (17) does not involve F. Therefore, the implementation of our test, as well as the derivation of its limiting distribution, become much simpler. For such reasons, we study the asymptotics of our test using the partial sum process expression in (17) in what follows.

3 Asymptotic Properties

3.1 Limiting Null Distribution

We first introduce some primitive conditions. Recall that we define r_0 such that $\gamma_0 = Q(r_0)$ under the null hypothesis in (3).

Condition 1

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

2. $\mathbb{E}[u_i|x_i, q_i] = 0$ *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. $\mathbb{E}[x_i x_i^\intercal] > \mathbb{E}[x_i x_i^\intercal \mathbf{1}[q_i \leq Q(r)]] > 0$ *for any* $r \in [\tau, 1-\tau]$.

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 a random sample, which simplifies our analysis. Under this condition, we can show that the induced order statistic $\{x_{[i]}u_{[i]}\}_{i=1}^n$ is a martingale difference array (e.g., Lemma 2 in Sen (1976); Lemma 3.2 in Bhattacharya (1984)), under which we can readily obtain the weak limit of the partial sum process. A martingale difference array is typically assumed in the time series case, where q_i is time and hence the observations are naturally sorted by q_i. A general form of cross-sectional dependence would break such a martingale property of the induced order statistic and hence substantially complicates the analysis. We leave this generalization for future research. Note that, however, we can allow for some dependence structure as long as the resulting induced order statistic $\{x_{[i]}u_{[i]}\}_{i=1}^n$ remains a martingale difference array.

Condition 1.2 assumes a correctly specified model without endogeneity (cf. Caner and Hansen (2004); Kourtellos, Stengos, and Tan (2016); Yu and Phillips (2018)).[5] 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 (2000), under which $\widehat{\theta} = (\widehat{\beta}^\intercal, \widehat{\delta}^\intercal)^\intercal$ is $\sqrt{n}$-consistent and asymptotically normal under the null hypothesis of constant threshold in (3). A more precise notation should be δ_{0n} in our shrinking size setup, but we still use δ_0 for notational simplicity. Condition 1.5 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 be any continuous matrix-valued functions here. Condition 1.7 is a full-rank condition, and Condition 1.8 bounds the conditional moments.

Under Condition 1, we first derive the weak limit of a partial sum process based on the induced order statistics.

Lemma 1 *Suppose Condition 1 holds. For* $\widehat{G}_n(r) = n^{-1/2}\sum_{i=1}^{\lfloor rn \rfloor} x_{[i]}\widehat{u}_{[i]}$, *we have* $\widehat{G}_n(\cdot) \Rightarrow G(\cdot)$ *as* $n \to \infty$ *under the null hypothesis in (3), where*

$$G(r) =_d \int_0^r V(t)^{1/2} dW_k(t) - \left(\int_0^r D(t)\,dt\right)\Phi_\beta - \left(\int_0^{\min\{r,r_0\}} D(t)\,dt\right)\Phi_\delta \tag{21}$$

[5] Our method can be easily adapted for endogeneity by considering the partial sum of $z_i\hat{u}_i$ and modifying the transformed process (25) ahead accordingly, where z_i denotes the instrument variable and $\hat{u}_i$ the regression residual.

for $r \in [0,1]$*,* Φ_β *and* Φ_δ *are given in (6), and* $W_k(\cdot)$ *is the* $k \times 1$ *vector standard Wiener process defined on* $[0,1]$*.*

In view of (21), we cannot directly use $\widehat{G}_n(r)$ to construct our test statistic because the nonlinear functions $V(\cdot)$ and $D(\cdot)$ are nuisance objects that complicate the asymptotic analysis. As we motivate in the previous section, the modified process $\mathcal{G}_n(\cdot)$ is designed to recover the simple form of the limit as in (8). We proceed to obtain its feasible sample analog, denoted as $\widehat{\mathcal{G}}_n(s)$, and study its asymptotic properties. To this end, we first estimate $D(r)$ and $V(r)$ for any $r \in (0,1)$ as (e.g., Yang (1985))

$$\widehat{D}(r) = \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{(i/n)-r}{b_n}\right) x_{[i]}x_{[i]}^{\intercal} \tag{22}$$

$$\widehat{V}(r) = \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{(i/n)-r}{b_n}\right) x_{[i]}x_{[i]}^{\intercal}\widehat{u}_{[i]}^{2} \tag{23}$$

for some kernel function $K(\cdot)$ and some bandwidth b_n, where $\widehat{u}_{[i]}$ denotes the re-ordered regression residual $\widehat{u}_i = y_i - x_i^{\intercal}\widehat{\beta} - x_i^{\intercal}\widehat{\delta}\mathbf{1}[q_i \le \widehat{\gamma}]$ under the null hypothesis. Given (22) and (23), the functions in (9) are estimated by

$$\widehat{h}(r) = \frac{1}{n}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor rn\rfloor} \frac{1}{v^{\intercal}\widehat{D}(i/n)^{-1}\widehat{V}(i/n)\widehat{D}(i/n)^{-1}v} \quad \text{and} \quad \widehat{g}(r) = \frac{\widehat{h}(r)}{\widehat{h}(1-\tau)}. \tag{24}$$

Under the following conditions, we can verify that all these kernel estimators are uniformly consistent. Note that these conditions are standard in the kernel regression literature (e.g., Li and Racine (2007)), where the last rate restriction in Condition 2.2 is from Yang (1981, Corollary 1).

Condition 2

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

The sample analog of $\mathcal{G}_n(s)$ in (10) is then given as

$$\widehat{\mathcal{G}}_n(s) = \frac{1}{\sqrt{n}}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor \widehat{g}^{-1}(s)n\rfloor} \sqrt{\widehat{h}(1-\tau)}\,\widehat{g}^{(1)}(i/n)v^{\intercal}\widehat{D}(i/n)^{-1}x_{[i]}\widehat{u}_{[i]}, \tag{25}$$

where $\widehat{g}^{(1)}(i/n) = \{v^\intercal \widehat{D}(i/n)^{-1} \widehat{V}(i/n) \widehat{D}(i/n)^{-1} v\widehat{h}(1-\tau)\}^{-1}$, and $\widehat{g}^{-1}(\cdot)$ is computed as the numerical inverse of $\widehat{g}(\cdot)$. The following lemma establishes that $\widehat{\mathcal{G}}_n(\cdot)$ weakly converges to the simple limit expression as in (8).

Lemma 2 *Suppose Conditions 1 and 2 hold. Then, for any v satisfying $v^\intercal v = 1$, under the null hypothesis in (3),*
(i) *$\widehat{D}(r)$, $\widehat{V}(r)$, $\widehat{h}(r)$, and $\widehat{g}(r)$ are uniformly consistent on $r \in [\tau, 1-\tau]$;*
(ii) *$\widehat{\mathcal{G}}_n(\cdot) \Rightarrow \mathcal{G}(\cdot)$ as $n \to \infty$, where*

$$\mathcal{G}(s) =_d W_1(s) - sv^\intercal \Phi_\beta^h - \min\{s, g(r_0)\} v^\intercal \Phi_\delta^h \tag{26}$$

for $s \in [0,1]$ with $\Phi_\beta^h = \sqrt{h(1-\tau)}\Phi_\beta$ and $\Phi_\delta^h = \sqrt{h(1-\tau)}\Phi_\delta$.

Lemma 2 implies that $\widehat{\mathcal{G}}_n(s)$ has a well-defined weak limit under the null hypothesis. Similarly, the sample analog of $\mathcal{G}_n^*(s)$ in (12) is given by

$$\widehat{\mathcal{G}}_n^*(s) = \begin{cases} \widehat{\mathcal{G}}_{1n}^*(s) & \text{if } s \leq \widehat{g}(\widehat{r}) \\ \widehat{\mathcal{G}}_{2n}^*(s) & \text{otherwise,} \end{cases} \tag{27}$$

where $\widehat{r} = \widehat{F}_n(\widehat{\gamma}) = n^{-1}\sum_{i=1}^n \mathbf{1}[q_i \leq \widehat{\gamma}]$,

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

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, each of them behaves as the independent standard Brownian bridge defined on $[0,1]$ in the limit.

As in (14), we thus define the constant-threshold test statistic, or the CT test statistic, as

$$CT_n = \frac{1}{\lfloor \widehat{g}(\widehat{r})n \rfloor} \sum_{i=1}^{\lfloor \widehat{g}(\widehat{r})n \rfloor} \widehat{\mathcal{G}}_{1n}^*(i/n)^2 + \frac{1}{n - \lfloor \widehat{g}(\widehat{r})n \rfloor} \sum_{i=\lfloor \widehat{g}(\widehat{r})n \rfloor + 1}^{n} \widehat{\mathcal{G}}_{2n}^*(i/n)^2 \tag{28}$$

in a similar vein to Nyblom (1989) and Elliott and Müller (2007). Theorem 1 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.

Table 1: Simulated critical values of the CT test

$\mathbb{P}(\int_0^1 \mathcal{B}_2(t)^\intercal \mathcal{B}_2(t)\, dt >$cv)	0.800	0.850	0.900	0.925	0.950	0.975	0.990
cv	0.467	0.527	0.600	0.666	0.745	0.888	1.067

Note: Entries are based on 50000 replications and 5000 step approximations to the continuous time process.

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

$$CT_n \to_d \int_0^1 \mathcal{B}_2(t)^\intercal \mathcal{B}_2(t)\, dt \tag{29}$$

under the null hypothesis in (3), where $\mathcal{B}_2(t)$ is the 2×1 vector standard Brownian bridge on $[0,1]$. However, $CT_n \to \infty$ in probability under the alternative hypothesis in (3).

The limiting distribution of CT_n is pivotal under the null hypothesis of a constant threshold. It does not depend on the choice of τ and v as long as the latter satisfies $v^\intercal v = 1$. Therefore, we can easily simulate the critical values, which are covered by Elliott and Müller (2007) as the special case with $k = 1$. We reproduce the results in Table 1 for reference. The test for (3) is then conducted as a one-sided test that rejects the null hypothesis if CT_n is larger than the corresponding critical values.

Unlike the conventional quasi-likelihood ratio tests in threshold regression models, the CT test only requires estimating the threshold regression model (2) under the null hypothesis of a constant threshold. It can reject the null hypothesis when the classic threshold regression model is mis-specified and hence can be seen as a specification test. When the CT test rejects the null hypothesis, we can conduct some sequential testing or model selection analysis to search for more flexible specifications as discussed in the introduction.

We summarize the steps to implement the CT test as follows:

Step 1 Under the constant threshold regression model, 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 (22) and (23), and the estimators $\widehat{h}(r)$, $\widehat{g}(r)$, and $\widehat{g}^{(1)}(r)$ as in (24). 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 (27).

Step 4 Compute the CT_n statistic in (28) and conduct a one-sided test using the critical values from Table 1.

3.2 Local Power Analysis

Theorem 1 derives the consistency of the CT test. To examine its local power properties, we now consider a local alternative model given as

$$y_i = x_i^\intercal \beta_0 + x_i^\intercal \delta_0 \mathbf{1}\left[q_i \leq \gamma_0\right] + x_i^\intercal \{n^{-1/2} \alpha\left(q_i\right)\} + u_i, \tag{30}$$

where $\alpha\left(\cdot\right)$ is some non-constant bounded function that characterizes the form of local deviation. Since $\alpha\left(\cdot\right)$ is nonparametric in q_i, this form of local alternative in (30) is very general to cover many empirically relevant cases, including, for example, multiple homogeneous thresholds (e.g., $\alpha\left(q_i\right) = \alpha_0 \mathbf{1}\left[q_i \leq \gamma_1\right]$ for some $\gamma_1 \neq \gamma_0$ and a non-zero finite $k \times 1$ vector α_0) and a single heterogeneous threshold (e.g., $\alpha(q_i) = \alpha_0 \mathbf{1}\left[q_i \leq \gamma_i\right]$ for some random variable γ_i and a non-zero finite $k \times 1$ vector α_0) as we discussed in Section 2.1.

The shrinking magnitude of the local deviation is of the order $n^{-1/2}$, with which the CT test has non-trivial asymptotic power (cf. Elliott, Müller, and Watson (2015)). This local alternative is smaller in order than δ_0 since $\delta_0 = O(n^{-\epsilon})$ for some $\epsilon \in (0, 1/2)$. Therefore, we can still obtain $\widehat{\gamma} - \gamma_0 = O_p\left(n^{-1+2\epsilon}\right)$ and $\widehat{\theta} - \theta_0 = O_p\left(n^{-1/2}\right)$, where $\theta_0 = (\beta_0^\intercal, \delta_0^\intercal)^\intercal$. Furthermore, the kernel estimators are still uniformly consistent on $[\tau, 1-\tau]$. Theorem 2 derives the weak limit of CT_n under the local alternative model in (30).

Theorem 2 *Suppose Conditions 1 and 2 hold. Then under the local alternative in (30),*

$$CT_n \rightarrow_d \int_0^1 \left(\mathcal{B}_2\left(t\right) + \mu(t)\right)^\intercal \left(\mathcal{B}_2\left(t\right) + \mu(t)\right) dt$$

as $n \rightarrow \infty$, *where* $\mu(s) = (\mu_1(s), \mu_2(s))^\intercal$ *with*

$$\begin{aligned} \mu_1(s) &= \sqrt{\frac{h\left(1-\tau\right)}{g(r_0)}} \left\{ \int_\tau^{g^{-1}(s)} \Psi_v(t) dt - \frac{s}{g\left(r_0\right)} \int_\tau^{r_0} \Psi_v(t) dt \right\} \\ \mu_2(s) &= \sqrt{\frac{h\left(1-\tau\right)}{1-g(r_0)}} \left\{ \int_{g^{-1}(s)}^{1-\tau} \Psi_v(t) dt - \frac{1-s}{1-g\left(r_0\right)} \int_{r_0}^{1-\tau} \Psi_v(t) dt \right\} \end{aligned}$$

and $\Psi_v(\cdot) = g^{(1)}(\cdot) v^\intercal \alpha\left(Q(\cdot)\right)$.

The local deviation $n^{-1/2}\alpha\left(\cdot\right)$ introduces a potentially non-zero drift function $\mu\left(\cdot\right)$ to the standard Brownian bridge. As long as $\alpha(\cdot)$ is non-constant either before or after the first break, at least one component of $\mu\left(\cdot\right)$ is non-zero and the limiting distribution is distinguished from the null distribution.

To better understand the drift function $\mu\left(\cdot\right)$, for example, we suppose that the local alternative model assumes an additional (homogeneous) threshold γ_1 in addition to the original threshold

γ_0. Without loss of generality, we let $\alpha(q_i) = \alpha_0 \mathbf{1}[q_i \leq \gamma_1]$ for some $\gamma_1 < \gamma_0$ and a non-zero finite $k \times 1$ vector α_0. Then $\mu_2(s)$ is zero for $s \in (g(r_0), 1]$. We also let $\gamma_1 = Q(r_1)$ for some $r_1 \in [0, r_0]$. In this case, we can show that the weak limit in (26) has an additional drift term, $\sqrt{h(1-\tau)} \min\{s, g(r_1)\} v^\intercal \alpha_0$. This non-zero drift term cannot be removed by the standardization in (13), and thus we have

$$\mu_1(s) = \sqrt{\frac{h(1-\tau)}{g(r_0)}} \left(\min\{s, g(r_1)\} - s\frac{g(r_1)}{g(r_0)} \right) v^\intercal \alpha_0$$

over the region $s \in [0, g(r_0)]$ in the limit experiment. When $v^\intercal \alpha_0 > 0$, $\mu_1(s)$ is positive for all $s \in (0, g(r_0))$ and zero at $s = 0$ and r_0. The optimal choice v might be obtained by maximizing the local power (cf. Andrews (1993) and Andrews and Ploberger (1994)). However, such a choice relies on the unknown knowledge of α_0 and more importantly, the specification that $\alpha(q_i) = \alpha_0 \mathbf{1}[q_i \leq \gamma_1]$. Therefore, the optimality under a general local alternative is very challenging, which is beyond the scope of this paper. We leave this for future research.

4 Monte Carlo Experiments

This section examines the small sample performance of the CT test in (28). We consider the following data generating processes (DGPs):

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

DGP-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-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 $\beta_0 = \iota_2$ 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 zero; (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 (3), while the remaining cases are for the alternative hypothesis in (3). We set $\tau = 0.1$ and $v = (v_1, v_2)^\intercal$ to be proportional to $(1, 1/\mathbb{E}[x_{2i}^2])^\intercal$ with $v^\intercal v = 1$. We use the rule-of-thumb choice of the bandwidth $b_n = (1/12)^{1/2} n^{-1/5}$ and the Gaussian kernel.[6] Other choices of bandwidth, kernel, and τ are also implemented, which lead to negligible changes. The

[6] Results with $v = (0,1)^\intercal$ are very similar and hence omitted.

sample sizes are n = 500, 1000, and 1500, and the significance level is 5%. The results are based on 1000 simulations.

For comparison, we also 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 (3), 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 rejection probabilities under the smooth threshold alternative, where the CT test dominates the $F(2|1)$ test. Third, the next 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. Fourth, 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, 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.

Table 2: Rejection probabilities when q and x are independent

		DGP-1			DGP-2			DGP-3		
δ	$n=$	500	1000	1500	500	1000	1500	500	1000	1500
CT test										
0.25		0.03	0.05	0.05	0.05	0.05	0.10	0.03	0.05	0.09
0.50		0.03	0.05	0.05	0.15	0.41	0.66	0.11	0.20	0.30
0.75		0.03	0.05	0.05	0.38	0.82	0.97	0.16	0.39	0.55
1.00		0.04	0.05	0.05	0.59	0.95	1.00	0.25	0.55	0.72
F(2|1) test										
0.25		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.04
0.50		0.01	0.01	0.01	0.02	0.14	0.42	0.07	0.23	0.36
0.75		0.01	0.01	0.02	0.17	0.66	0.91	0.27	0.54	0.68
1.00		0.00	0.01	0.01	0.46	0.92	0.99	0.42	0.77	0.89
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 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 three DGPs with $(q_i, z_i, u_i, x_{2i}) \sim iid\mathcal{N}(0, I_4)$. The first three columns are based on $\gamma_0(s) = 0$, the middle three are based on $\gamma_0(s) = \sin(s)/2$; and the third three are based on two thresholds at 0 and 0.1.

Table 3: Rejection probabilities when q is correlated with x

		CT test			F(2|1) test		
δ	$n =$	500	1000	1500	500	1000	1500
0.25		0.03	0.03	0.05	0.09	0.12	0.14
0.50		0.03	0.04	0.05	0.08	0.11	0.14
0.75		0.05	0.05	0.04	0.08	0.10	0.12
1.00		0.04	0.05	0.05	0.09	0.12	0.14
		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 (3) of the CT test, the $F(2|1)$ test by Bai and Perron (1998), 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-1 with $(q_i, z_i) \sim iid\mathcal{N}(0, I_2)$, $x_{2i}|\,(q_i, z_i) = (q, s) \sim iid\mathcal{N}\left(0, 1/(1+s^2+q^2)\right)$, and $u_i|x_{2i} = x \sim iid\mathcal{N}(0, 1+x^2)$.

In Table 3, we introduce some correlation between q_i and x_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 from DGP-1, which is under the null hypothesis of a single threshold at zero, and use $(q_i, z_i) \sim iid\mathcal{N}(0, I_2)$, $x_i|\,(q_i, z_i) = (q, s) \sim iid\mathcal{N}\left(0, 1/(1+s^2+q^2)\right)$ 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, the CT test performs well in terms of controlling size if the sample size and the break size are large enough. Third, the 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.

5 Application: Tipping Point and Social Segregation

Our motivating example is 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,

Table 4: Tipping point estimation and testing results (1980-1990)

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

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

called the tipping point. In particular, they consider the following threshold regression model:

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

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 the 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 public transportation commuters. The data are collected from a variety of cities in three periods: 1970-1980, 1980-1990, and 1990-2000. Card, Mas, and Rothstein (2008) 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 point behavior, with the estimated tipping points ranging approximately from 5% to 20% across cities.

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. We want to formally test if the tipping point remains constant across tracts.

Table 4 shows the results of the CT test in (28) using the data in Chicago, Los Angeles, New York City, and Washington D.C. in the decade 1980-1990. We choose the rule-of-thumb bandwidth $b_n = (1/12)^{1/2}n^{-1/5}$ and $\tau = 0.1$ as in the Monte Carlo experiments. We set $v = (1, 0, ..., 0)^{\intercal}$ since only the constant term involves a coefficient change. We also follow Card, Mas, and Rothstein (2008) to use the tracts in which the initial minority share is between 5% and 60%. The small p-values of CT 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.

These test results suggest that we need to use a more flexible form of threshold in the tipping point analysis. For example, we want to consider the fully nonparametric threshold specification as in Lee and Wang (2019), which is given by

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

where $\gamma_0(\cdot)$ is an unknown tipping point function, and z_i denotes the attitude index that Card, Mas, and Rothstein (2008) construct or demographics of tract i. This specification assumes $\gamma_i = \gamma_0\left(z_i\right)$. See Lee and Wang (2019) for details.

6 Conclusion

This paper recasts the cross-sectional threshold problem into the time series structural break problem. Using this new framework, we develop a test for homogeneity of the threshold parameter as empirically motivated by the tipping point problem.

Though we focus on the threshold homogeneity test in this paper, we can apply the novel transformation idea to develop other tests. First, our transformation allows us to convert other inference methods developed in the structural break models into the threshold model setup, including inference about γ_0 (e.g., Elliott, Müller, and Watson (2015)), δ_0 (e.g., Andrews and Ploberger (1994)), and β_0 (e.g., Elliott and Müller (2014)). The inference on δ_0 covers the test for threshold effect. Second, for the CT test, since we only need to estimate the null model and its alternative hypothesis is fully nonparametric, we can modify it for other types of hypotheses as long as we can consistently estimate the null model. For instance, the test can be modified to test for the null hypothesis of any fixed number of thresholds against additional thresholds. Finally, though we do not allow for endogeneity in this paper, the partial sum process and our test can still be constructed even when the model involves endogeneity as long as the parameters can be consistently estimated using instruments.

Appendix: Proofs

Throughout the proofs, we define r_0 and $\widehat{r}$ as $r_0 = F(\gamma_0)$ and $\widehat{r} = \widehat{F}_n(\widehat{\gamma})$, or equivalently $\gamma_0 = Q(r_0)$ and $\widehat{\gamma} = \widehat{Q}_n(\widehat{r})$. We let C denote a generic constant and denote $h_\tau \equiv h(1-\tau)$ and $\widehat{h}_\tau = \widehat{h}(1-\tau)$.

A.1 Useful Results

We first prove (11).

Lemma A.1 *Under Condition 1, (11) holds for $s \in [0,1]$ as $n \to \infty$.*

Proof of Lemma A.1 We first decompose (10) into

$$\begin{aligned} & \mathcal{G}_n(s) \\ = & \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i^\intercal x_i u_i \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right] \\ & - \frac{1}{n} \sum_{i=1}^n w_i^\intercal x_i x_i^\intercal \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right] \left\{ \sqrt{n}(\widehat{\beta} - \beta_0) + \mathbf{1}[q_i \le Q(r_0)] \sqrt{n}(\widehat{\delta} - \delta_0) \right\} \\ & - \frac{1}{n} \sum_{i=1}^n w_i^\intercal x_i x_i^\intercal \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right] \{\mathbf{1}[q_i \le \hat{\gamma}] - \mathbf{1}[q_i \le Q(r_0)]\} \sqrt{n}\widehat{\delta}. \end{aligned} \tag{A.1}$$

We can verify (11) from the limits of the first two terms, which can be obtained from, as $n \to \infty$,

$$\begin{aligned} \mathcal{G}_{nA}(s) & \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i^\intercal x_i u_i \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right] \Rightarrow W_1(s), \\ \mathcal{G}_{nB}(s) & \equiv \frac{1}{n} \sum_{i=1}^n w_i^\intercal x_i x_i^\intercal d_0 \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right] \to_p s v^\intercal d_0 h_\tau^{1/2} \end{aligned}$$

uniformly over $s \in [0,1]$ for any bounded $k \times 1$ vector d_0 and by the continuous mapping theorem. For $\mathcal{G}_{nA}(s)$, since it converges to a Gaussian process as in Lemma A.4 of Hansen (2000), it suffices to show that, for any $s \le s'$, the covariance kernel is given as

$$\begin{aligned} & Cov\left[\mathcal{G}_{nA}(s), \mathcal{G}_{nA}(s')\right] \\ = & \mathbb{E}\left[(w_i^\intercal x_i u_i)^2 \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right]\right] \\ = & \int_\tau^{g^{-1}(s)} h_\tau \mathbb{E}\left[\left. \frac{v^\intercal D(F(q_i))^{-1} x_i x_i^\intercal u_i^2 D(F(q_i))^{-1} v}{\left(v^\intercal D(F(q_i))^{-1} V(F(q_i)) D(F(q_i))^{-1} v h_\tau\right)^2} \right| F(q_i) = r \right] dr \end{aligned}$$

$$= \int_{\tau}^{g^{-1}(s)} h_\tau \mathbb{E}\left[\left. \frac{v^\intercal D(r)^{-1} V(r) D(r))^{-1} v}{\left(v^\intercal D(r)^{-1} V(r) D(r)^{-1} v h_\tau\right)^2} \right| q_i = Q(r) \right] dr$$

$$= \int_{\tau}^{g^{-1}(s)} \frac{1}{v^\intercal D(r)^{-1} V(r) D(r)^{-1} v h_\tau} dr = \int_{\tau}^{g^{-1}(s)} g^{(1)}(r)\, dr = s.$$

For $\mathcal{G}_{nB}(s)$, for any $s \in [0,1]$, we have

$$\begin{aligned} \mathbb{E}[\mathcal{G}_{nB}(s)] &= \mathbb{E}\left[w_i^\intercal x_i x_i^\intercal d_0 \mathbf{1}\left[Q(\tau) \leq q_i \leq Q(g^{-1}(s))\right]\right] \\ &= h_\tau^{1/2} \int_{\tau}^{g^{-1}(s)} \mathbb{E}\left[\left. \frac{v^\intercal D(F(q_i))^{-1} x_i x_i^\intercal d_0}{v^\intercal D(F(q_i))^{-1} V(F(q_i)) D(F(q_i))^{-1} v h_\tau} \right| F(q_i) = r \right] dr \\ &= h_\tau^{1/2} v^\intercal d_0 \int_{\tau}^{g^{-1}(s)} \frac{1}{v^\intercal D(r)^{-1} V(r) D(r)^{-1} v h_\tau} dr = s v^\intercal d_0 h_\tau^{1/2}. \end{aligned}$$

Then the pointwise convergence holds under the standard LLN and the uniform convergence follows similarly from the proof of Lemma 1 in Hansen (1996).

It remains to bound the last term in (A.1), which we denote $\Upsilon_n(s)$. Suppose $\hat{\gamma} > \gamma_0$. Then we have

$$|\Upsilon_n(s)| \leq \sup_{r \in [\tau, 1-\tau]} \left\| \sqrt{h(1-\tau)} g^{(1)}(r) D(r)^{-1} v \right\| \left\| \frac{1}{n} \sum_{i=1}^{n} x_i x_i^\intercal \left\{ \mathbf{1}[q_i \leq \widehat{\gamma}] - \mathbf{1}[q_i \leq \gamma_0] \right\} \right\| \left\| \sqrt{n}\widehat{\delta} \right\|$$

for any $s \in [0,1]$, where $\mathbf{1}\left[Q(\tau) \leq q_i \leq Q(g^{-1}(s))\right] \leq 1$. Thus,

$$|\Upsilon_n(s)| \leq C \left\| \frac{1}{n} \sum_{i=1}^{n} x_i x_i^\intercal \left| \mathbf{1}[q_i \leq \widehat{\gamma}] - \mathbf{1}[q_i \leq \gamma_0] \right| \right\| \left\| \sqrt{n}\widehat{\delta} \right\|$$

for some $0 < C < \infty$, because $\sup_{r \in [\tau, 1-\tau]} ||\sqrt{h(1-\tau)} g^{(1)}(r) D(r)^{-1} v|| < \infty$ by Condition 1.6. Note that the bound does not depend on s. By Lemma A.12 in Hansen (2000) and Condition 1.4, we have that $||\sqrt{n}\widehat{\delta}|| \leq ||\sqrt{n}(\widehat{\delta} - \delta_0)|| + ||\sqrt{n}\delta_0|| = O_p(1) + O_p(n^{1/2-\epsilon})$ with $\epsilon \in (0, 1/2)$. Let $E_{\delta n}$ be the event that $||\sqrt{n}\widehat{\delta}|| \leq C_\delta n^{1/2-\epsilon}$ for some $0 < C_\delta < \infty$ and then $\mathbb{P}(E_{\delta n}^c) \leq \varepsilon$ for any $\varepsilon > 0$ if n is sufficiently large. Now let $E_{\gamma n}$ be the event that $\widehat{\gamma} \in (\gamma_0 - C_\gamma n^{-1+2\epsilon}, \gamma_0 + C_\gamma n^{-1+2\epsilon})$ for some $0 < C_\gamma < \infty$. Lemma A.9 in Hansen (2000) yields that $\mathbb{P}\left(E_{\gamma n}^c\right) \leq \varepsilon$ for any $\varepsilon > 0$ if n is sufficiently large. Then for any $\eta > 0$ and any $\varepsilon > 0$, if n is sufficiently large,

$$\begin{aligned} & \mathbb{P}\left(\sup_{s \in [\tau, 1-\tau]} |\Upsilon_n(s)| > \eta \right) && \text{(A.2)} \\ \leq\ & \mathbb{P}\left(\left\{ \sup_{s \in [\tau, 1-\tau]} |\Upsilon_n(s)| > \eta \right\} \cap E_{\gamma n} \cap E_{\delta n} \right) + \mathbb{P}\left(E_{\gamma n}^c\right) + \mathbb{P}\left(E_{\delta n}^c\right) \end{aligned}$$

$$
\begin{aligned}
&\leq \eta^{-1} C n^{1/2-\epsilon} \mathbb{E}\left[\left\| x_i x_i^{\intercal} \left| \mathbf{1}\left[q_i \leq \widehat{\gamma}\right] - \mathbf{1}\left[q_i \leq \gamma_0\right]\right| \mathbf{1}\left[E_{\gamma n}\right]\right\|\right] + 2\varepsilon \\
&\leq \eta^{-1} C n^{1/2-\epsilon} \left\| \sup_{q\in\mathbb{R}} \mathbb{E}\left[x_i x_i^{\intercal} | q_i = q\right]\right\| \left|\mathbb{P}(q_i \leq \gamma_0 + C_{\gamma} n^{-1+2\epsilon}) - \mathbb{P}(q_i \leq \gamma_0)\right| + 2\varepsilon \\
&\leq \eta^{-1} C' n^{-1/2+\epsilon} + 2\varepsilon \\
&\leq 3\varepsilon
\end{aligned}
$$

for some $0 < C' < \infty$. Note that the second inequality is by Markov's inequality; the fourth inequality is by Conditions 1.6, 1.8, and $|\mathbb{P}(q_i \leq \gamma_0 + C_{\gamma} n^{-1+2\epsilon}) - \mathbb{P}(q_i \leq \gamma_0)| = |F(\gamma_0 + C_{\gamma} n^{-1+2\epsilon}) - F(\gamma_0)| \leq f(\gamma_*) C_{\gamma} n^{-1+2\epsilon} = O(n^{-1+2\epsilon})$ for some $\gamma_* \in (\gamma_0, \gamma_0 + C_{\gamma} n^{-1+2\epsilon})$, where $F(\cdot)$ is continuous and $f(\gamma_*) < \infty$ by Condition 1.3. The argument for $\hat{\gamma} \leq \gamma_0$ is identical and hence we have $\sup_{s\in[\tau,1-\tau]} |\Upsilon_n(s)| = o_p(1)$. The desired result follows. ■

We establish the convergence of the key partial sum processes in Section 2.3.

Lemma A.2 *Suppose Condition 1 holds. Then, as* $n \to \infty$

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor rn \rfloor} x_{[i]} u_{[i]} \Rightarrow \int_0^r V(t)^{1/2} dW_k(t) \tag{A.3}
$$

for $r \in [0,1]$ *and*

$$
\sup_{r\in[0,1]} \left\| \frac{1}{n} \sum_{i=1}^{\lfloor rn \rfloor} x_{[i]} x_{[i]}^{\intercal} - \int_0^r D(t)\, dt \right\| \to_p 0, \tag{A.4}
$$

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

Proof of Lemma A.2 We prove the first result (A.3) using Theorem 2 in Bhattacharya (1974). By the Cramér-Wold device, it suffices to show for any $k \times 1$ non-zero vector v,

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor rn \rfloor} v^{\intercal} x_{[i]} u_{[i]} \Rightarrow \int_0^r \left(v^{\intercal} V(t) v\right)^{1/2} dW_1(t). \tag{A.5}
$$

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 $\mathbb{E}[v^{\intercal} x_i u_i | q_i] = 0$ almost surely and

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

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

$$\phi_V(r) = \int_0^r v^\intercal V(t)vdt.$$

By Theorem 2 in Bhattacharya (1974), we have

$$(n\phi_V(1))^{-1/2}\sum_{i=1}^{\lfloor rn \rfloor} v^\intercal x_{[i]}u_{[i]} \Rightarrow W_1\left(\frac{\phi_V(r)}{\phi_V(1)}\right). \tag{A.6}$$

Then (A.5) follows from the continuous mapping theorem and the fact that

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

For the second result (A.4), 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

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

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

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

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, $\sup_{r\in[0,1]} |\phi_{nD}(r) - \phi_D(r)| \to 0$ 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 suffices to show $\sup_{r\in[0,1]} |n^{-1}\sum_{i=1}^{\lfloor rn \rfloor} \xi_{[i]} - \phi_{nD}(r)| \to_p 0$, which is done in a way analogous to (A.6) (e.g., p.1038 in Bhattacharya (1974)). The desired result follows by the Cramér-Wold device. ■

We now show the equivalence results in (15) and (16) in the following lemma, where $n^{-1/2}\sum_{i=1}^{\lfloor rn \rfloor} x_{[i]}u_{[i]} = n^{-1/2}\sum_{i=1}^{n} x_{[i]}u_{[i]}\mathbf{1}[\widehat{F}_n\left(q_{(i)}\right) \le r] = n^{-1/2}\sum_{i=1}^{n} x_i u_i \mathbf{1}[\widehat{F}_n\left(q_i\right) \le r] = n^{-1/2}\sum_{i=1}^{n} x_i u_i \mathbf{1}[q_i \le \widehat{Q}_n(r)]$ and similarly $n^{-1}\sum_{i=1}^{\lfloor rn \rfloor} x_{[i]}x_{[i]}^\intercal = n^{-1}\sum_{i=1}^{n} x_i x_i^\intercal \mathbf{1}[q_i \le \widehat{Q}_n(r)]$.

Lemma A.3 *Under Condition 1,*

$$\sup_{r\in[0,1]}\left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i \left\{\mathbf{1}[q_i \le \widehat{Q}_n(r)] - \mathbf{1}[q_i \le Q\left(r\right)]\right\}\right\| = o_p\left(1\right), \tag{A.7}$$

$$\sup_{r\in[0,1]}\left\|\frac{1}{n}\sum_{i=1}^{n}x_i x_i^{\intercal}\left\{\mathbf{1}[q_i\leq\widehat{Q}_n(r)]-\mathbf{1}[q_i\leq Q(r)]\right\}\right\| = o_p(1), \tag{A.8}$$

where $Q(\cdot)$ and $\widehat{Q}_n(\cdot)$ are quantile and empirical quantile functions of q_i, respectively.

Proof of Lemma A.3 For the first result, we let $J_n(\gamma)=n^{-1/2}\sum_{i=1}^{n}x_i u_i\mathbf{1}[q_i\leq\gamma]$. Lemma A.4 in Hansen (2000) yields that $J_n(\gamma)\Rightarrow\mathcal{J}(\gamma)$, where $\mathcal{J}(\cdot)$ is a mean-zero Gaussian process indexed by $\gamma\in\mathbb{R}$ with almost surely continuous sample paths. Using the change of variables with $\gamma=Q(r)$ and the fact that $\sup_{r\in[\eta,1-\eta]}|\widehat{Q}_n(r)-Q(r)|=o_p(1)$ for any constant $\eta\in(0,1/2)$ by the Glivenko-Cantelli theorem, we obtain that

$$\sup_{r\in[\eta,1-\eta]}\left\|J_n(\widehat{Q}_n(r))-J_n(Q(r))\right\|=o_p(1).$$

For (A.7), therefore, it is sufficient to show that for any $\varepsilon>0$, we can pick η such that for a sufficiently large n,

$$\mathbb{P}\left(\sup_{r\in[0,\eta]}\left\|J_n(\widehat{Q}_n(r))\right\|>\varepsilon\right)<\varepsilon \quad\text{and}\quad \mathbb{P}\left(\sup_{r\in[0,\eta]}\|J_n(Q(r))\|>\varepsilon\right)<\varepsilon, \tag{A.9}$$

and the same results for $r\in[1-\eta,1]$. To establish the first one in (A.9), we use (A.3) to obtain that

$$J_n(\widehat{Q}_n(r))=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i u_i\mathbf{1}[\widehat{F}_n(q_i)\leq r]=\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor rn\rfloor}x_{[i]}u_{[i]}\Rightarrow\int_0^r V(t)^{1/2}dW_k(t)$$

for $r\in[0,1]$, and hence

$$\mathbb{P}\left(\sup_{r\in[0,\eta]}\left\|J_n(\widehat{Q}_n(r))\right\|>\varepsilon\right)\to\mathbb{P}\left(\sup_{r\in[0,\eta]}\left\|\int_0^r V(t)^{1/2}dW_k(t)\right\|>\varepsilon\right)$$

as $n\to\infty$. However, since the process $\mathcal{J}_Q(r)=\int_0^r V(t)^{1/2}dW_k(t)$ indexed by r satisfies $\mathcal{J}_Q(0)=0$ almost surely and has an almost surely continuous sample path, the above probability can be smaller than ε if η is sufficiently small. The second one in (A.9) can be similarly shown since $J_n(Q(r))\Rightarrow\mathcal{J}(Q(r))$ by Lemma A.4 in Hansen (2000), where $\mathcal{J}(Q(0))=0$ almost surely and has an almost surely continuous sample path as well. The same results as (A.9) can be shown for $r\in[1-\eta,1]$ symmetrically and hence omitted. Therefore, (A.7) is established.

For (A.8), note that $n^{-1}\sum_{i=1}^{n}x_i x_i^{\intercal}\mathbf{1}[q_i\leq\gamma]\to_p\mathcal{M}(\gamma)$ uniformly in $\gamma\in\mathbb{R}$ by Lemma 1 of Hansen (1996), where $\mathcal{M}(\gamma)$ is continuous in γ. The desired result can be shown by a similar argument as (A.7) using (A.4). ■

A.2 Proofs of the Results in Section 3

Proof of Lemma 1 Note that

$$\begin{aligned}
\widehat{G}_n(r) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor rn\rfloor} x_{[i]}u_{[i]} \\
&\quad -\frac{1}{n}\sum_{i=1}^{\lfloor rn\rfloor} x_{[i]}x_{[i]}^{\intercal}\left\{\sqrt{n}(\widehat{\beta}-\beta_0)+\mathbf{1}[\widehat{F}_n(q_{(i)})\le r_0]\sqrt{n}(\widehat{\delta}-\delta_0)\right\} \\
&\quad -\frac{1}{n}\sum_{i=1}^{\lfloor rn\rfloor} x_{[i]}x_{[i]}^{\intercal}\left\{\mathbf{1}\left[F(q_{(i)})\le r_0\right]-\mathbf{1}[\widehat{F}_n(q_{(i)})\le r_0]\right\}\sqrt{n}(\widehat{\delta}-\delta_0) \\
&\quad -\frac{1}{n}\sum_{i=1}^{\lfloor rn\rfloor} x_{[i]}x_{[i]}^{\intercal}\left\{\mathbf{1}[\widehat{F}_n(q_{(i)})\le \widehat{r}]-\mathbf{1}\left[F(q_{(i)})\le r_0\right]\right\}\sqrt{n}\widehat{\delta} \\
&\equiv G_{n1}(r)-G_{n2}(r)-G_{n3}(r)-G_{n4}(r),
\end{aligned}$$

where the continuous mapping theorem yields

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

from Lemma A.2 and (6), since $\mathbf{1}[\widehat{F}_n(q_{(i)})\le r_0]=\mathbf{1}[i/n\le r_0]$. For the third term, $\sup_{r\in[0,1]}\|G_{n3}(r)\|=o_p(1)$ by (A.8) in Lemma A.3. Finally, for the last term,

$$\|G_{n4}(r)\|\le\left\|\frac{1}{n}\sum_{i=1}^{n}x_ix_i^{\intercal}\left|\mathbf{1}\left[q_i\le\widehat{\gamma}\right]-\mathbf{1}\left[q_i\le\gamma_0\right]\right|\right\|\left\|\sqrt{n}\widehat{\delta}\right\|$$

for any $r\in[0,1]$, where the inequality is because the summands are nonnegative. Note that the bound does not depend on r. By Lemma A.12 in Hansen (2000) and Condition 1.4, we have that $||\sqrt{n}\widehat{\delta}||\le||\sqrt{n}(\widehat{\delta}-\delta_0)||+||\sqrt{n}\delta_0||=O_p(1)+O_p(n^{1/2-\epsilon})$ with $\epsilon\in(0,1/2)$. Let $E_{\delta n}$ be the event that $||\sqrt{n}\widehat{\delta}||\le C_\delta n^{1/2-\epsilon}$ for some $0<C_\delta<\infty$ and then $\mathbb{P}(E_{\delta n}^c)\le\varepsilon$ for any $\varepsilon>0$ if n is sufficiently large. Let $E_{\gamma n}$ be the event that $\widehat{\gamma}\in(\gamma_0-C_\gamma n^{-1+2\epsilon},\gamma_0+C_\gamma n^{-1+2\epsilon})$ for some $0<C_\gamma<\infty$. Then, using the same argument in (A.2), for any $\eta>0$ and any $\varepsilon>0$, if n is sufficiently large,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{r\in[0,1]}\|G_{n4}(r)\|>\eta\right) \\
&\le\mathbb{P}\left(\left\{\sup_{r\in[0,1]}\|G_{n4}(r)\|>\eta\right\}\cap E_{\gamma n}\cap E_{\delta n}\right)+\mathbb{P}\left(E_{\gamma n}^c\right)+\mathbb{P}\left(E_{\delta n}^c\right)
\end{aligned}$$

$$\le \eta^{-1} C n^{1/2-\epsilon} \mathbb{E}\left[\|x_i x_i^{\intercal} \left|\mathbf{1}\left[q_i \le \widehat{\gamma}\right] - \mathbf{1}\left[q_i \le \gamma_0\right]\right| \mathbf{1}\left[E_{\gamma n}\right]\|\right] + 2\varepsilon$$
$$\le \eta^{-1} C' n^{-1/2+\epsilon} + 2\varepsilon$$
$$\le 3\varepsilon$$

for some $0 < C, C' < \infty$, where the second inequality is by Markov's inequality and by Condition 1.4 with $\epsilon \in (0, 1/2)$; the third inequality is by Conditions 1.3, 1.6, and 1.8. Hence

$$\sup_{r \in [0,1]} \|G_{n4}(r)\| = o_p(1), \tag{A.10}$$

and the desired result follows. ■

Lemma A.4 *Let*

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

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.4 For expositional simplicity, we only present the case with scalar x_i. As $\widehat{u}_i = u_i - x_i(\widehat{\beta} - \beta_0) - x_i(\widehat{\delta} - \delta_0)\mathbf{1}[q_i \le \gamma_0] - x_i\widehat{\delta}(\mathbf{1}[q_i \le \widehat{\gamma}] - \mathbf{1}[q_i \le \gamma_0])$, we have

$$\begin{aligned}
\left|\widehat{V}(r) - \widehat{V}^0(r)\right| &= \left|\frac{1}{n}\sum_{i=1}^{n} x_{[i]}^2 \left(\widehat{u}_{[i]} + u_{[i]}\right)\left(\widehat{u}_{[i]} - u_{[i]}\right) K_i(r)\right| \qquad \text{(A.11)} \\
&\le \frac{1}{n}\sum_{i=1}^{n} \left|x_i^3 (\widehat{u}_i + u_i)(\widehat{\beta} - \beta_0) K_i(r)\right| \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \left|x_i^3 (\widehat{u}_i + u_i)(\widehat{\delta} - \delta_0)\mathbf{1}[q_i \le \gamma_0] K_i(r)\right| \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \left|x_i^3 (\widehat{u}_i + u_i)\widehat{\delta}\{(\mathbf{1}[q_i \le \widehat{\gamma}] - \mathbf{1}[q_i \le \gamma_0])\} K_i(r)\right| \\
&\equiv V_{1n}(r) + V_{2n}(r) + V_{3n}(r).
\end{aligned}$$

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 $0 < C < \infty$, where $\mathcal{B}_r(x)$ denotes a generic open ball centered at x with radius r. Lemmas A.9 and A.12 in Hansen (2000) imply $\mathbb{P}(E_{\theta n}^c) \le \varepsilon$ and $\mathbb{P}\left(E_{\gamma n}^c\right) \le \varepsilon$ for any $\varepsilon > 0$ if C and n are large enough. Then, for any $\eta > 0$,

$$\mathbb{P}\left(\sup_{r \in [\tau, 1-\tau]} |V_{1n}(r)| > \eta\right)$$

$$
\begin{aligned}
&\leq \mathbb{P}\left(\left\{\sup_{r\in[\tau,1-\tau]}|V_{1n}(r)|>\eta\right\}\cap E_{\gamma n}\cap E_{\theta n}\right)+\mathbb{P}(E_{\gamma n}^{c}\cup E_{\theta n}^{c})\\
&\leq \eta^{-1}\max_{1\leq i\leq n}\sup_{r\in[0,1]}K_i(r)\times\mathbb{E}\left[\left|x_i^3\left(\widehat{u}_i+u_i\right)(\widehat{\beta}-\beta_0)\right||E_{\theta n}\right]+2\varepsilon\\
&\leq \eta^{-1}\max_{1\leq i\leq n}\sup_{r\in[0,1]}K_i(r)\times\Big\{2\mathbb{E}\left[\left|x_i^3u_i(\widehat{\beta}-\beta_0)\right||E_{\theta n}\right]+\mathbb{E}\left[\left|x_i^4(\widehat{\beta}-\beta_0)^2\right||E_{\theta n}\right]\\
&\quad+\mathbb{E}\left[\left|x_i^4\mathbf{1}\left[q_i\leq\gamma_0\right](\widehat{\delta}-\delta_0)(\widehat{\beta}-\beta_0)^2\right||E_{\theta n}\right]\\
&\quad+\mathbb{E}\left[\left|x_i^4\widehat{\delta}(\widehat{\beta}-\beta_0)\left(\mathbf{1}\left[q_i\leq\widehat{\gamma}\right]-\mathbf{1}\left[q_i\leq\gamma_0\right]\right)\right||E_{\gamma n}\cap E_{\theta n}\right]\Big\}+2\varepsilon\\
&\leq C\eta^{-1}n^{-1/2}b_n^{-1}\left(2\mathbb{E}\left[\left|x_i^3u_i\right|\right]+\mathbb{E}\left[\left|x_i^4\right|\right]\right)+2\varepsilon\\
&\leq 3\varepsilon
\end{aligned}
$$

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 $\mathbf{1}[\cdot]\leq 1$; and the last inequality follows from Conditions 1.8 and 2.2. For $V_{2n}(r)$ and $V_{3n}(r)$, the same argument yields that $\sup_{r\in[\tau,1-\tau]}|V_{2n}(r)|=o_p(1)$ and $\sup_{r\in[\tau,1-\tau]}|V_{3n}(r)|=o_p(1)$ as well because $\widehat{\delta}=O_p(n^{-\epsilon})=o_p(1)$. Hence, the desired result follows. ■

Lemma A.5 *Suppose Conditions 1 and 2 hold. Then under the null hypothesis in (3),* $\sup_{r\in[\tau,1-\tau]}||\widehat{D}(r)-D(r)||=o_p(1)$*,* $\sup_{r\in[\tau,1-\tau]}||\widehat{V}(r)-V(r)||=o_p(1)$*,* $\sup_{r\in[\tau,1-\tau]}|\widehat{h}(r)-h(r)|=o_p(1)$*, and* $\sup_{r\in[\tau,1-\tau]}|\widehat{g}(r)-g(r)|=o_p(1)$*.*

Proof of Lemma A.5 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.4, 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. Note that

$$
V(r)=\mathbb{E}\left[x_i^2u_i^2|F(q_i)=r\right]=\frac{1}{f_\nu(r)}\iint x^2u^2f_{x,u,\nu}(x,u,r)dxdu,
$$

where $\nu_i=F(q_i)$ is the standard uniform random variable and hence $f_\nu(r)=1$.

The triangular inequality yields

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

where the first item is $o_p(1)$ as established in equations (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

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

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

$$\begin{aligned}
\mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\widehat{V}^0(r)-\mathbb{E}[\widehat{V}^0(r)]\right|>\eta\right) &\leq \mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\widehat{V}^0(r)-\widehat{V}^\kappa(r)\right|>\eta/3\right) \quad \text{(A.12)}\\
&\quad +\mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\mathbb{E}[\widehat{V}^0(r)]-\mathbb{E}[\widehat{V}^\kappa(r)]\right|>\eta/3\right)\\
&\quad +\mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\widehat{V}^\kappa(r)-\mathbb{E}[\widehat{V}^\kappa(r)]\right|>\eta/3\right)\\
&\equiv P_{n1}+P_{n2}+P_{n3}.
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\left[\sup_{r\in[\tau,1-\tau]}\left|\widehat{V}^0(r)-\widehat{V}^\kappa(r)\right|\right] &\leq \mathbb{E}\left[\frac{C_1}{nb_n}\sum_{i=1}^n x_{[i]}^2u_{[i]}^2\mathbf{1}[x_{[i]}^2u_{[i]}^2>\kappa_n]\right] \quad \text{(A.13)}\\
&\leq b_n^{-1}\kappa_n^{-1}C_1\sup_{q\in\mathbb{R}}\mathbb{E}\left[x_i^4u_i^4|q_i=q\right]\\
&\leq C_1b_n^{-1}\kappa_n^{-1},
\end{aligned}$$

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,

$$\sup_{r\in[\tau,1-\tau]}\left|\mathbb{E}[\widehat{V}^0(r)]-\mathbb{E}[\widehat{V}^\kappa(r)]\right|\leq b_n^{-1}\kappa_n^{-1}C_1\sup_{q\in\mathbb{R}}\mathbb{E}\left[x_i^4u_i^4|q_i=q\right]\leq C_1b_n^{-1}\kappa_n^{-1}$$

and hence $P_{n2}\leq 3C_1/(\eta b_n\kappa_n)$ as well. For P_{n3}, Lemma A.6 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 when $b_n=O\left(n^{-1/5}\right)$. By combining these results, it follows that

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

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.6 *Under the same condition as in Lemma A.5, for any $\eta>0$, P_{n3} in (A.12) satisfies that $P_{n3}\le(\eta/3)^{-1}C(\log n/(nb_n))^{1/2}$ for some $0<C<\infty$.*

Proof of Lemma A.6 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

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

where

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

We first bound T_{3n}^κ. Let

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

and then

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

Recall that κ_n is some large truncation parameter satisfying $\kappa_n\to\infty$ as $n\to\infty$. Note that, similarly as (A.13), $\sup_{r\in[\tau,1-\tau]}x_{[i]}^2u_{[i]}^2K_i(r)\,\mathbf{1}[x_{[i]}^2u_{[i]}^2\le\kappa_n]$ is bounded by $C_2\kappa_nb_n^{-1}$ for some constant $C_2\in(0,\infty)$ and hence $|Z_{n,i}^\kappa(r)|\le 2C_2\kappa_n/(nb_n)$ for all $i=1,\ldots,n$. Define $\psi_n=(nb_n\log n)^{1/2}/\kappa_n$. Then $\psi_n|Z_{n,i}^\kappa(r)|\le 2C_2(\log n/(nb_n))^{1/2}\le 1/2$ 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 $\exp(\psi_n |Z^{\kappa}_{n,i}(r)|) \leq 1+\psi_n |Z^{\kappa}_{n,i}(r)| + \psi_n^2 |Z^{\kappa}_{n,i}(r)|^2$. Hence

$$\mathbb{E}[\exp(\psi_n \left|Z^{\kappa}_{n,i}(r)\right|)] \leq 1 + \psi_n^2 \mathbb{E}\left[(Z^{\kappa}_{n,i}(r))^2\right] \leq \exp\left(\psi_n^2 \mathbb{E}\left[(Z^{\kappa}_{n,i}(r))^2\right]\right) \tag{A.14}$$

since $\mathbb{E}[Z^{\kappa}_{n,i}(r)] = 0$ and $1 + x \leq \exp(x)$ for $x \geq 0$. By the Markov's inequality, $\mathbb{P}(X > c) \leq \mathbb{E}[\exp(Xa)]/\exp(ac)$ holds for any nonnegative random variable X and positive constants a and c. Then we have, for some constant η_n to be specified later,

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

for some sequence $\eta_n \to 0$ as $n \to \infty$, where the second inequality is by (A.14) and 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.,

$$\sup_{r \in [\tau, 1-\tau]} \mathbb{P}\left(\left|\widehat{V}^{\kappa}(r) - \mathbb{E}[\widehat{V}^{\kappa}(r)]\right| > \eta_n\right) \leq 2\exp\left(-\psi_n \eta_n + \psi_n^2 C_3 \kappa_n^2/(nb_n)\right). \tag{A.15}$$

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.15) is summable. 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.15), we have

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

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

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

by the Borel-Cantelli lemma.

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

$$\begin{aligned}\mathbb{E}\left|\widehat{V}^{\kappa}\left(r\right) - \widehat{V}^{\kappa}\left(r_j\right)\right| &= \mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n} x_{[i]}^2 u_{[i]}^2 \left(K_i\left(r\right) - K_i\left(r_j\right)\right)\mathbf{1}[x_{[i]}^2 u_{[i]}^2 \leq \kappa_n]\right|\right] \\ &\leq C_4\left(1-2\tau\right)\kappa_n/\left(m_n b_n^2\right)\end{aligned}$$

for some constant $C_4 < \infty$ given $r \in \mathcal{I}_j$. This bound does not depend on j and hence $T_{1n}^{\kappa} = O_{a.s.}(\kappa_n/\left(m_n b_n^2\right))$. The same argument yields that

$$\begin{aligned}\left|\mathbb{E}[\widehat{V}^{\kappa}\left(r\right)] - \mathbb{E}[\widehat{V}^{\kappa}\left(r_j\right)]\right| &\leq \mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n} x_{[i]}^2 u_{[i]}^2 \left(K_i\left(r\right) - K_i\left(r_j\right)\right)\mathbf{1}[x_{[i]}^2 u_{[i]}^2 \leq \kappa_n]\right|\right] \\ &\leq C_4\left(1-2\tau\right)\kappa_n/\left(m_n b_n^2\right),\end{aligned}$$

which does not depend on j, and hence it gives the uniform bound $T_{2n}^{\kappa} = O(\kappa_n/\left(m_n b_n^2\right))$ as well. Therefore, by choosing $m_n = (\log n b_n^3/n)^{-1/2}\kappa_n$, we have that T_{1n}^{κ} and T_{2n}^{κ} 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.7 *Suppose Conditions 1 and 2 hold. For*

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

we have $\widetilde{\mathcal{G}}_n(\cdot) \Rightarrow \mathcal{G}(\cdot)$ as $n \to \infty$ under the null hypothesis in (3).

Proof of Lemma A.7 We let $\pi(\cdot) \equiv h_{\tau}^{1/2} g^{(1)}(\cdot) v^{\intercal} D(\cdot)^{-1}$. Similarly as in Lemma 1, we decompose

$$\begin{aligned}\widetilde{\mathcal{G}}_n(s) &= \frac{1}{\sqrt{n}}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} \pi(i/n) x_{[i]} u_{[i]} \\ &\quad -\frac{1}{n}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} \pi(i/n) x_{[i]} x_{[i]}^{\intercal}\sqrt{n}(\widehat{\beta} - \beta_0) \\ &\quad -\frac{1}{n}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor} \pi(i/n) x_{[i]} x_{[i]}^{\intercal}\mathbf{1}[\widehat{F}_n(q_{(i)}) \leq r_0]\sqrt{n}(\widehat{\delta} - \delta_0)\end{aligned} \tag{A.16}$$

$$
\begin{aligned}
& -\frac{1}{n}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor}\pi(i/n)x_{[i]}x_{[i]}^{\intercal}\left\{\mathbf{1}\left[F(q_{(i)})\leq r_0\right]-\mathbf{1}[\widehat{F}_n(q_{(i)})\leq r_0]\right\}\sqrt{n}(\widehat{\delta}-\delta_0)\\
& -\frac{1}{n}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor}\pi(i/n)x_{[i]}x_{[i]}^{\intercal}\left\{\mathbf{1}[\widehat{F}_n(q_{(i)})\leq \widehat{r}]-\mathbf{1}\left[F(q_{(i)})\leq r_0\right]\right\}\sqrt{n}\widehat{\delta}\\
\equiv\; & A_{1n}(s)-A_{2n}(s)-A_{3n}(s)-A_{4n}(s)-A_{5n}(s).
\end{aligned}
$$

First, we derive the limit of $A_{1n}(s)$ by applying Corollary 29.14 in Davidson (1994).[7] To this end, we let $U_{n,i}=h_\tau^{1/2}n^{-1/2}g^{(1)}(i/n)v^{\intercal}D(i/n)^{-1}x_{[i]}u_{[i]}$ and $\overrightarrow{\mathrm{q}}=\{q_i\}_{i=1}^n$, and check Conditions 29.6(a) to (f′) in the corollary. Condition (a) is satisfied since $\mathbb{E}[U_{n,i}]=\mathbb{E}[\mathbb{E}[U_{n,i}|\overrightarrow{\mathrm{q}}]]=0$ 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

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

where $||\cdot||_p$ 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 $K_n(t)=\lfloor g^{-1}(t)n\rfloor$, and from the fact that $g^{-1}(\cdot)$ 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

$$
\begin{aligned}
\mathbb{E}\left[x_{[i]}x_{[i]}^{\intercal}u_{[i]}^2\right] &= \mathbb{E}\left[\mathbb{E}\left[x_jx_j^{\intercal}u_j^2|q_j=q_{(i)}\right]\right]\\
&= \mathbb{E}\left[V(F(q_{(i)}))\right]\\
&= V(i/n)+\mathbb{E}\left[\frac{\partial V(t_i)}{\partial t}\left(F\left(q_{(i)}\right)-i/n\right)\right]\\
&= V(i/n)+O\left(n^{-1/2}\right),
\end{aligned}
\tag{A.17}
$$

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

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

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

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

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

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.17); the second expression from the bottom is by Riemann integral as $n\to\infty$; the last expression is by the definition of $g^{(1)}\left(\cdot\right)$ 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.2, Lemma A.12 in Hansen (2000), and the continuous mapping theorem to obtain that

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

and

$$
A_{3n}\left(s\right) \to_p \left(\int_{g^{-1}(0)}^{\min(g^{-1}(s),r_0)} g^{(1)}(t) v^\intercal D\left(t\right)^{-1} D\left(t\right) dt\right)\Phi_\delta h_\tau^{1/2} = \min\{s, g\left(r_0\right)\} v^\intercal \Phi_\delta h_\tau^{1/2}.
$$

For $A_{4n}(s)$, since $g^{-1}(1)=1-\tau$, we have

$$
|A_{4n}(s)| \le \sup_{r\in[\tau,1-\tau]} \|\pi(r)\| \left\| \frac{1}{n}\sum_{i=\lfloor \tau n\rfloor+1}^{\lfloor (1-\tau)n\rfloor} x_i x_i^\intercal \left\{\mathbf{1}\left[F(q_i)\le r_0\right] - \mathbf{1}[\widehat{F}_n(q_i)\le r_0]\right\}\right\| \left\|\sqrt{n}(\widehat{\delta}-\delta_0)\right\|,
$$

and hence $\sup_{s\in[0,1]}|A_{4n}(s)| = o_p(1)$ by (A.8) in Lemma A.3.

Finally, for $A_{5n}(s)$, let $E_{\delta n}$ be the event that $||\sqrt{n}\widehat{\delta}|| \le C_\delta$ for some $0<C_\delta<\infty$ and $E_{\gamma n}$ be the event that $\widehat{\gamma}\in(\gamma_0 - C_\gamma n^{-1+2\epsilon}, \gamma_0 + C_\gamma n^{-1+2\epsilon})$ for some $0<C_\gamma<\infty$. Then, using the same argument in (A.2), for any $\eta > 0$ and $\varepsilon > 0$, if n is sufficiently large,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{s\in[0,1]} |A_{5n}(s)| > \eta\right) \\
&\leq \mathbb{P}\left(\left\{\sup_{s\in[0,1]} |A_{5n}(s)| > \eta\right\} \cap E_{\gamma n} \cap E_{\delta n}\right) + 2\varepsilon \\
&\leq \eta^{-1} C n^{1/2-\epsilon} \mathbb{E}\left[\left\| \frac{1}{n} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor (1-\tau)n \rfloor} \pi(i/n) x_{[i]} x_{[i]}^{\intercal} \left\{\mathbf{1}\left[q_{(i)} \leq \widehat{\gamma}\right] - \mathbf{1}\left[q_{(i)} \leq \gamma_0\right]\right\} \mathbf{1}\left[E_{\gamma n}\right] \right\|\right] + 2\varepsilon \\
&\leq \eta^{-1} C n^{1/2-\epsilon} \sup_{r\in[\tau,1-\tau]} \|\pi(r)\| \, \mathbb{E}\left[\|x_i x_i^{\intercal} \left|\mathbf{1}\left[q_i \leq \widehat{\gamma}\right] - \mathbf{1}\left[q_i \leq \gamma_0\right]\right| \mathbf{1}\left[E_{\gamma n}\right]\|\right] + 2\varepsilon \\
&\leq \eta^{-1} C' n^{-1/2+\epsilon} + 2\varepsilon \\
&\leq 3\varepsilon
\end{aligned}
$$

for some $0 < C, C' < \infty$, where the second inequality is by Markov's inequality and the fourth inequality is by Conditions 1.3, 1.6, and 1.8. Thus, $\sup_{s\in[0,1]} |A_{5n}(s)| = o_p(1)$. The desired result follows by combining these results. ■

Proof of Lemma 2 The first result follows from Lemma A.5. For the second result, given Lemma A.7, it suffices to establish

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

We first consider the case with $g^{-1}(s) > \widehat{g}^{-1}(s)$. we let $\pi(\cdot) \equiv h_{\tau}^{1/2} g^{(1)}(\cdot) v^{\intercal} D(\cdot)^{-1}$ and $\widehat{\pi}(\cdot) \equiv \widehat{h}_{\tau}^{1/2} \widehat{g}^{(1)}(\cdot) v^{\intercal} \widehat{D}(\cdot)^{-1}$. Note that, for any $s \in [0,1]$,

$$
\begin{aligned}
\widehat{\mathcal{G}}_n(s) - \widetilde{\mathcal{G}}_n(s) &= \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor \widehat{g}^{-1}(s) n \rfloor} \widehat{\pi}(i/n) x_{[i]} \widehat{u}_{[i]} - \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor g^{-1}(s) n \rfloor} \pi(i/n) x_{[i]} \widehat{u}_{[i]} \\
&= \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor \widehat{g}^{-1}(s) n \rfloor} \left\{\widehat{\pi}(i/n) - \pi(i/n)\right\} x_{[i]} \widehat{u}_{[i]} + \frac{1}{\sqrt{n}} \sum_{i=\lfloor g^{-1}(s) n \rfloor+1}^{\lfloor \widehat{g}^{-1}(s) n \rfloor} \pi(i/n) x_{[i]} \widehat{u}_{[i]} \\
&\equiv B_{1n}(s) + B_{2n}(s).
\end{aligned}
$$

For expositional simplicity, we only present the case with scalar x_i.

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

$$
B_{1n}(s) = \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor \widehat{g}^{-1}(s) n \rfloor} \left\{\widehat{\pi}(i/n) - \pi(i/n)\right\} x_{[i]} u_{[i]}
$$

$$
\begin{aligned}
& +\frac{1}{\sqrt{n}} \sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor\widehat{g}^{-1}(s)n\rfloor} \{\widehat{\pi}(i/n)-\pi(i/n)\} x_{[i]}\left(\widehat{u}_{[i]}-u_{[i]}\right) \\
\equiv\ & B_{11n}(s)+B_{12n}(s). \qquad \text{(A.18)}
\end{aligned}
$$

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.8 below. For $B_{12n}(s)$, define the event $E_{\theta n}=\{\widehat{\theta}\in\mathcal{B}_{C_\theta n^{-1/2}}(\theta_0)\}$ for some $0<C_\theta<\infty$. Lemma A.12 in Hansen (2000) 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

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

where the second inequality is by Lemma A.5, and the last equality follows from Lemma A.2 and (A.10). Therefore, $B_{1n}(s)$ is uniformly $o_p(1)$.

For $B_{2n}(s)$, we write

$$
\begin{aligned}
B_{2n}(s) &= \frac{1}{\sqrt{n}}\sum_{i=\lfloor g^{-1}(s)n\rfloor+1}^{\lfloor\widehat{g}^{-1}(s)n\rfloor}\pi(i/n)x_{[i]}u_{[i]}+\frac{1}{\sqrt{n}}\sum_{i=\lfloor g^{-1}(s)n\rfloor+1}^{\lfloor\widehat{g}^{-1}(s)n\rfloor}\pi(i/n)x_{[i]}\left(\widehat{u}_{[i]}-u_{[i]}\right) \\
&\equiv B_{21n}(s)+B_{22n}(s). \qquad \text{(A.19)}
\end{aligned}
$$

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.5, $\mathbb{P}(E_{gn}^c)\leq\varepsilon$ for any $\varepsilon>0$ and $\eta>0$ as $n\to\infty$. On the event E_{gn} and using the same argument as in proving Lemma A.7, we then have that for any given value $\widehat{g}^{-1}(s)=\varrho(s)$,

$$
\sup_{s\in[0,1]}|B_{21n}(s)|\leq\sup_{s\in[0,1]}\sup_{|\varrho(s)-g^{-1}(s)|<\eta}\left|\frac{1}{\sqrt{n}}\sum_{i=\lfloor\varrho(s)n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor}\pi(i/n)x_{[i]}u_{[i]}\right|
$$

$$
\begin{aligned}
&\Rightarrow \sup_{s\in[0,1]} \sup_{|\varrho(s)-g^{-1}(s)|<\eta} \left| h_\tau^{1/2} \int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t) D(t)^{-1} V(t)^{1/2} dW_k(t) \right. \\
&\qquad\qquad\qquad \left. - h_\tau^{1/2} \int_{g^{-1}(0)}^{\varrho(s)} g^{(1)}(t) D(t)^{-1} V(t)^{1/2} dW_k(t) \right| \\
&=_d \sup_{s\in[0,1]} \sup_{|\varrho(s)-g^{-1}(s)|<\eta} \left| W_1(s) - W_1(g(\varrho(s))) \right|.
\end{aligned}
$$

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 $\mathbb{E}[\sup_{s\in[0,t]} |W_1(s)|] \leq \sqrt{2t/\pi}$; 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 these two events, using the same decomposition with the $A_{2n}(s)$, $A_{3n}(s)$, and $A_{4n}(s)$ terms as in (A.16), we have that

$$
\begin{aligned}
&\sup_{s\in[0,1]} |B_{22n}(s)| \\
&\leq \sup_{r\in[\tau,1-\tau]} |\pi(r)| \sup_{s\in[0,1]} \frac{1}{\sqrt{n}} \sum_{i=\lfloor \widehat{g}^{-1}(s)n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} \left| x_{[i]} (\widehat{u}_{[i]} - u_{[i]}) \right| \\
&\leq C \sup_{s\in[0,1]} \frac{1}{\sqrt{n}} \sum_{i=\lfloor (g^{-1}(s)-\eta)n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} x_{[i]}^2 \left\{ |\widehat{\beta} - \beta_0| + |\widehat{\delta} - \delta_0| \mathbf{1}\left[q_{(i)} \leq \gamma_0\right] + \widehat{\delta} \left| \mathbf{1}\left[q_{(i)} \leq \widehat{\gamma}\right] - \mathbf{1}\left[q_{(i)} \leq \gamma_0\right] \right| \right\} \\
&\leq C' \sup_{s\in[0,1]} \frac{1}{n} \sum_{i=\lfloor (g^{-1}(s)-\eta)n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} x_{[i]}^2 \\
&\to_p C' \sup_{s\in[0,1]} \int_{g^{-1}(s)-\eta}^{g^{-1}(s)} D(t)\, dt
\end{aligned}
$$

for some constant $0 < C, C' < \infty$, 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, the result in (A.10), and by conditioning on the events $E_{\theta n}$ and E_{gn}; the last convergence is from Lemma A.2. By choosing a sufficiently small η, therefore, $\sup_{s\in[0,1]} |B_{22n}(s)| = o_p(1)$. The proof for $g(s) \leq \widehat{g}^{-1}(s)$ is identical and hence omitted. The desired result thus follows. ■

Lemma A.8 *Under the same condition as in Lemma 2,* $\sup_{s\in[0,1]} |B_{11n}(s)| = o_p(1)$*, where* $B_{11n}(\cdot)$ *is defined in (A.18).*

Proof of Lemma A.8 Note that for each n, $\{x_{[i]}u_{[i]}\}_{i=1}^n$ are independent conditional on $\overrightarrow{\mathrm{q}} = \{q_i\}_{i=1}^n$ 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 $\pi(\cdot) = h_\tau^{1/2} g^{(1)}(\cdot) v^\intercal D(\cdot)^{-1}$ and the stochastic process

$$\mathbb{L}_\ell(\pi) = \sum_{i=\lfloor \tau n\rfloor+1}^{\ell} L_{ni}(\pi),$$

where $L_{ni}(\pi) = n^{-1/2}\pi(i/n)x_{[i]}u_{[i]}$. Define the semi-metric $\rho(\pi_1, \pi_2) = \sup_{r\in[\tau,1-\tau]} |\pi_1(r) - \pi_2(r)|$. 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 $\varepsilon > 0$,

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

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

$$\sup_{\rho(\pi,\pi_1)\leq\varepsilon_n}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor(1-\tau)n\rfloor}\mathbb{E}\left[\left(L_{ni}(\pi)-L_{ni}(\pi_1)\right)^2|\overrightarrow{\mathrm{q}}\right] \;\leq\; \bar{C}^2\varepsilon_n n^{-1}\sum_{i=\lfloor\tau n\rfloor+1}^{\lfloor(1-\tau)n\rfloor}\mathbb{E}\left[\left|x_{[i]}u_{[i]}\right|^2\middle|\overrightarrow{\mathrm{q}}\right] \;\to\; 0 \text{ a.s.}$$

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 $\overrightarrow{\mathrm{q}}$, the process $\mathbb{L}_n(\cdot)$ is asymptotically tight, that is, for any $\varepsilon>0$, there exists some η such that if n is large enough,

$$\mathbb{P}\left(\sup_{\rho(\pi_1,\pi_2)\leq\eta}|\mathbb{L}_n(\pi_1)-\mathbb{L}_n(\pi_2)|>\varepsilon\middle|\overrightarrow{\mathrm{q}}\right)\leq\varepsilon \text{ a.s.} \tag{A.20}$$

Define $E_{\pi n}=\{\rho(\widehat{\pi},\pi)\leq\eta_n\}$ for $\eta_n>0$, where $\widehat{\pi}(\cdot)=\widehat{h}_\tau^{1/2}\widehat{g}^{(1)}(\cdot)v^\intercal\widehat{D}(\cdot)^{-1}$. Then, for any $\varepsilon>0$, we have

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

The second inequality is from Lemma A.5 that implies $\mathbb{P}(E^c_{\pi n})\leq\varepsilon$ 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 $\overrightarrow{\mathrm{q}}$. The last inequality is from (A.20) and the steps in p.227 in van der Vaart and Wellner (1996). In particular, for some $1\leq n_0\leq n$,

$$\max_{1\leq\ell\leq n}\mathbb{P}\left(\sqrt{\ell/n}\sup_{\rho(\widehat{\pi},\pi)\leq\eta}|\mathbb{L}_\ell(\pi)-\mathbb{L}_\ell(\widehat{\pi})|>\varepsilon\middle|\overrightarrow{\mathrm{q}}\right)$$

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

where the second inequality follows from Markov's inequality, (A.20), and setting a large enough n_0 satisfying $n_0 \to \infty$ and $n_0 n^{-1/2} \to 0$. ■

Proof of Theorem 1 We first prove (29) under the null hypothesis. To this end, define

$$\widehat{\mathcal{G}}_n^{0*}(s) = \begin{cases} \widehat{\mathcal{G}}_{1n}^*(s) & \text{if } s \leq g(r_0) \\ \widehat{\mathcal{G}}_{2n}^*(s) & \text{otherwise,} \end{cases}$$

which is different from $\widehat{\mathcal{G}}_n^*(\cdot)$ only in a neighborhood of $g(r_0)$. Then, since the empirical distribution function is uniformly consistent, Lemma A.5 yields $\widehat{g}(\widehat{r}) - g(r_0) = o_p(1)$. It yields that $\int_{g(r_0)-\varepsilon_n}^{g(r_0)+\varepsilon_n} |\widehat{\mathcal{G}}_n^*(t) - \widehat{\mathcal{G}}_n^{0*}(t)| dt = o_p(1)$ for some $\varepsilon_n \to 0$ with $n \to \infty$, which is implied by the fact that both $\sup_{s\in[0,1]} |\widehat{\mathcal{G}}_n^*(s)|$ and $\sup_{s\in[0,1]} |\widehat{\mathcal{G}}_n^{0*}(s)|$ are $O_p(1)$ given Lemma 2. It follows that $\int_0^1 |\widehat{\mathcal{G}}_n^*(s) - \widehat{\mathcal{G}}_n^{0*}(s)| ds = o_p(1)$.

Note that, under the null hypothesis, Lemmas A.2, A.5, and the continuous mapping theorem yield that

$$\begin{aligned}
\widehat{\mathcal{G}}_{1n}^*(s) &\Rightarrow \frac{1}{\sqrt{g(r_0)}} \left\{ \mathcal{G}(s) - \frac{s}{g(r_0)} \mathcal{G}(g(r_0)) \right\} \\
&=_d \frac{1}{\sqrt{g(r_0)}} \left\{ W_1(s) - \frac{s}{g(r_0)} W_1(g(r_0)) \right\} \equiv \mathcal{G}_1(s), && \text{(A.21)} \\
\widehat{\mathcal{G}}_{2n}^*(s) &\Rightarrow \frac{1}{\sqrt{1-g(r_0)}} \left\{ (\mathcal{G}(1) - \mathcal{G}(s)) - \frac{1-s}{1-g(r_0)} (\mathcal{G}(1) - \mathcal{G}(g(r_0))) \right\} \\
&=_d \frac{1}{\sqrt{1-g(r_0)}} \left\{ (W_1(1) - W_1(s)) - \frac{1-s}{1-g(r_0)} (W_1(1) - W_1(g(r_0))) \right\} \equiv \mathcal{G}_2(s). && \text{(A.22)}
\end{aligned}$$

Moreover, by change of variables with $t = s/g(r_0)$ and $t = (1-s)/(1-g(r_0))$, respectively, we have

$$\begin{aligned}
\frac{1}{g(r_0)} \int_0^{g(r_0)} \mathcal{G}_1(s)^2 ds &=_d \frac{1}{g(r_0)} \int_0^1 \{W_1(g(r_0)t) - tW_1(g(r_0))\}^2 dt \\
&=_d \int_0^1 \{W_1(t) - tW_1(1)\}^2 dt
\end{aligned}$$

and

$$
\begin{aligned}
&\frac{1}{1-g(r_0)}\int_{g(r_0)}^{1}\mathcal{G}_2(s)^2ds\\
&=_d \frac{1}{(1-g(r_0))}\int_0^1\left\{(W_1(1)-W_1(1-(1-g(r_0))t))-t(W_1(1)-W_1(g(r_0)))\right\}^2dt\\
&=_d \frac{1}{(1-g(r_0))}\int_0^1\left\{W_1((1-g(r_0))t)-tW_1(1-g(r_0))\right\}^2dt\\
&=_d \int_0^1\left\{W_1(t)-tW_1(1)\right\}^2dt.
\end{aligned}
$$

Therefore, the limiting null distribution of CT_n is obtained as

$$
\begin{aligned}
CT_n &= \frac{1}{g(r_0)}\times\frac{1}{n}\sum_{i=1}^{\lfloor g(r_0)n\rfloor}\widehat{\mathcal{G}}_{1n}^*(i/n)^2+\frac{1}{1-g(r_0)}\times\frac{1}{n}\sum_{i=\lfloor g(r_0)n\rfloor+1}^{n}\widehat{\mathcal{G}}_{2n}^*(i/n)^2+o_p(1)\\
&\to_d \frac{1}{g(r_0)}\int_0^{g(r_0)}\mathcal{G}_1(s)^2ds+\frac{1}{1-g(r_0)}\int_{g(r_0)}^1\mathcal{G}_2(s)^2ds\\
&=_d \int_0^1\mathcal{B}_2(t)^\intercal\mathcal{B}_2(t)\,dt
\end{aligned}
$$

where $\mathcal{B}_2(t)$ is the 2×1 standard Brownian bridge on $[0,1]$.

We now examine the limit of CT_n 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

$$
n^\epsilon(\widehat{\theta}-\theta_0)=O_p(1) \tag{A.23}
$$

for any given $\widehat{\gamma}$ (or $\widehat{r}$). To see this, 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,

$$
\begin{aligned}
n^\epsilon\left(\widehat{\theta}-\theta_0\right) &= n^\epsilon\left(\sum_{i=1}^n X_i(\gamma)X_i(\gamma)^\intercal\right)^{-1}\left(\sum_{i=1}^n X_i(\gamma)\left\{y_i-X_i(\gamma)^\intercal\theta_0\right\}\right)\\
&= \left(\frac{1}{n}\sum_{i=1}^n X_i(\gamma)X_i(\gamma)^\intercal\right)^{-1}\left(\frac{n^\epsilon}{n}\sum_{i=1}^n X_i(\gamma)u_i+\frac{n^\epsilon}{n}\sum_{i=1}^n X_i(\gamma)x_i^\intercal\delta_0(\mathbf{1}[q_i<\gamma]-\mathbf{1}[q_i<\gamma_i])\right)\\
&\equiv \widehat{\Theta}_{n1}^{-1}\left(\widehat{\Theta}_{n2}+\widehat{\Theta}_{n3}\right).
\end{aligned}
$$

Similarly as Lemma 1 in Hansen (1996), we have $\widehat{\Theta}_{n1}\to_p\Theta_1=\mathbb{E}[X_i(\gamma)X_i(\gamma)^\intercal]$, 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 $\widehat{\Theta}_{n2} = O_p(n^{-1/2+\epsilon}) = o_p(1)$ 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 $q_{(i)}$, and $\widehat{\pi}(\cdot) = \widehat{h}_\tau^{1/2}\widehat{g}^{(1)}(\cdot)v^\intercal \widehat{D}(\cdot)^{-1}$. We decompose

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

and denote their re-scaled and demeaned terms as in (12) 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 Lemma 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.7 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

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

since $\widehat{\theta}-\theta_0 = O_p(n^{-\epsilon})$ from (A.23). However, as $\widehat{C}_{2n}(s)$ is linear in s, the re-scaling and demeaning procedure eliminates the leading term and hence we have $\widehat{C}_{2n}^*(s) = o_p\left(n^{1/2-\epsilon}\right)$.

Lastly, for $\widehat{C}^{*}_{3n}(s)$, the fact that $r_{[i]}$ is a non-degenerate random variable implies

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

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 $\widehat{C}_{3n}(s)=O_p\left(n^{1/2-\epsilon}\right)$ as $\delta_0=c_0n^{-\epsilon}$ with $c_0\neq0$ 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.[8] It follows that $\widehat{C}^{*}_{3n}(s)=O_p\left(n^{1/2-\epsilon}\right)$, which dominates $\widehat{\mathcal{G}}^{*}_n(\cdot)$. Since $\epsilon\in(0,1/2)$, therefore, $\widehat{\mathcal{G}}^{*}_n(\cdot)$ diverges and hence $CT_n\to\infty$ with probability approaching to one under the alternative hypothesis. ■

Proof of Theorem 2 Under the local alternative, the error term is now defined as $\widetilde{u}_i=u_i+n^{-1/2}x_i^{\intercal}\alpha\left(q_i\right)$. However, Lemma A.5 in Hansen (2000) still implies that $n^{-1}\sum_{i=1}^{n}x_i\widetilde{u}_i\to_p\mathbb{E}\left[x_iu_i\right]$, which yields $\widehat{\gamma}\to_p\gamma_0$. We can also show $\widehat{\gamma}-\gamma_0=O_p\left(n^{-1+2\epsilon}\right)$ by the same argument as Lemmas A.6-A.9 in Hansen (2000). We only present the different part, which shows up in the proof of Lemma A.9. In particular, eq. (43) in Hansen (2000) now involves the following additional term

$$M_n^{\alpha}=\frac{1}{n}\sum_{i=1}^{n}x_ix_i^{\intercal}\alpha\left(q_i\right).$$

Using Lemma A.2 and the argument in Lemma A.8 in Hansen (2000), for any constants η and ε, there exists some large enough constants a and C such that

$$\mathbb{P}\left(\sup_{\frac{a}{n^{-1+2\epsilon}}\leq|\gamma-\gamma_0|\leq C}\frac{\left|M_n^{\alpha}\right|}{n^{-1+2\epsilon}\left|\gamma-\gamma_0\right|}>\eta\right)\leq\varepsilon.$$

Then the rest of the argument follows from p.597 in Hansen (2000). Note that the additional term $n^{-1/2}x_i^{\intercal}\alpha\left(q_i\right)$ changes the asymptotic distribution of $\widehat{\gamma}$ but not the rate of convergence. Therefore, Lemma A.12 in Hansen (2000) implies that $\widehat{\theta}-\theta_0=O_p\left(n^{-1/2}\right)$. Moreover, given these results, $\widehat{D}(r)$, $\widehat{V}(r)$, $\widehat{h}(r)$, and $\widehat{g}(r)$ are still uniformly consistent on $r\in[\tau,1-\tau]$ under the local alternative in (30), which is implied by the proof of Lemma A.4.

[8] 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 (12) 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.

Now, given these consistency results, we have

$$\begin{aligned}
\widehat{G}_n(r) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor rn\rfloor} x_{[i]}u_{[i]} \\
&\quad -\frac{1}{n}\sum_{i=1}^{\lfloor rn\rfloor} x_{[i]}x_{[i]}^{\intercal}\left\{\sqrt{n}(\widehat{\beta}-\beta_0)+\mathbf{1}[F(q_{(i)})\le r_0]\sqrt{n}(\widehat{\delta}-\delta_0)\right\} \\
&\quad -\frac{1}{n}\sum_{i=1}^{\lfloor rn\rfloor} x_{[i]}x_{[i]}^{\intercal}\left\{\mathbf{1}[\widehat{F}_n(q_{(i)})\le \widehat{r}]-\mathbf{1}\left[F(q_{(i)})\le r_0\right]\right\}\sqrt{n}\widehat{\delta} \\
&\quad +\frac{1}{n}\sum_{i=1}^{\lfloor rn\rfloor} x_{[i]}x_{[i]}^{\intercal}\alpha(q_{(i)}) \\
&\Rightarrow G(r)+\int_0^r D(r)\,\alpha(Q(r))\,dr
\end{aligned}$$

similarly as the proof of Lemma 1. Then the continuous mapping theorem and the same argument as in the proof of Lemma A.7 yield that $\widehat{\mathcal{G}}_n(\cdot)\Rightarrow\mathcal{G}^{\alpha}(\cdot)$, where

$$\mathcal{G}^{\alpha}(s)=W_1(s)-sv^{\intercal}\Phi_{\beta}-\min\{s,g(r_0)\}v^{\intercal}\Phi_{\delta}+h_{\tau}^{1/2}\int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t)v^{\intercal}\alpha(Q(t))\,dt, \tag{A.24}$$

which includes an additional drift term than $\mathcal{G}(s)$. Recall that $g^{-1}(0)=\tau$ and $g^{-1}(1)=1-\tau$. Denoting $\Psi_v(\cdot)=g^{(1)}(\cdot)v^{\intercal}\alpha(Q(\cdot))$, it follows that

$$\frac{1}{\sqrt{g(r_0)}}\left\{\mathcal{G}^{\alpha}(s)-\frac{s}{g(r_0)}\mathcal{G}^{\alpha}(g(r_0))\right\}=_d\mathcal{G}_1(s)+\frac{h_{\tau}^{1/2}}{\sqrt{g(r_0)}}\left\{\int_{\tau}^{g^{-1}(s)}\Psi_v(t)dt-\frac{s}{g(r_0)}\int_{\tau}^{r_0}\Psi_v(t)dt\right\}$$

and

$$\begin{aligned}
&\frac{1}{\sqrt{1-g(r_0)}}\left\{(\mathcal{G}^{\alpha}(1)-\mathcal{G}^{\alpha}(s))-\frac{1-s}{1-g(r_0)}(\mathcal{G}^{\alpha}(1)-\mathcal{G}^{\alpha}(g(r_0)))\right\} \\
&=_d\mathcal{G}_2(s)+\frac{h_{\tau}^{1/2}}{\sqrt{1-g(r_0)}}\left\{\left(\int_{\tau}^{1-\tau}\Psi_v(t)dt-\int_{\tau}^{g^{-1}(s)}\Psi_v(t)dt\right)\right. \\
&\qquad\qquad\left.-\frac{1-s}{1-g(r_0)}\left(\int_{\tau}^{1-\tau}\Psi_v(t)dt-\int_{\tau}^{r_0}\Psi_v(t)dt\right)\right\} \\
&=_d\mathcal{G}_2(s)+\frac{h_{\tau}^{1/2}}{\sqrt{1-g(r_0)}}\left\{\int_{g^{-1}(s)}^{1-\tau}\Psi_v(t)dt-\frac{1-s}{1-g(r_0)}\int_{r_0}^{1-\tau}\Psi_v(t)dt\right\}
\end{aligned}$$

instead of (A.21) and (A.22). Then the desired result follows as in the proof of Theorem 1. ■

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., R. L. Lumsdaine, and J. H. Stock (1998): "Testing for and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies*, 65, 395–432.

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 (2001): "Threshold Autoregressions with a Unit Root," *Econometrica*, 69, 1555–1597.

——— (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.

Casini, A., and P. Perron (2020): "Continous Record Laplace-Based Inference about the Break Date in Structural Change Models," *Forthcoming at Journal of Econometrics.*

——— (2021a): "Continuous Record Asymptotics for Change-point Models," *Working Paper.*

——— (2021b): "Generalized Laplace Inference in Multiple Change-Points Models," *Forthcoming at Econometric Theory.*

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. (1996): "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430.

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

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

Jiang, L., X. Wang, and J. Yu (2018a): "New Distribution Theory for the Estimation of Structural Break Point in Mean," *Journal of Econometrics*, 205(1), 156–176.

Jiang, L., X. Wang, and S. Yu (2018b): "In-fill Asymptotic Theory for Structural Break Point in Autoregression: A Unified Theory," *Working Paper.*

Kosorok, M. R. (2008): *Introduction to Empirical Processes and Semiparametric Inference.* Springer-Verlag, New York.

Kourtellos, A., T. Stengos, and C. M. Tan (2016): "Structural Threshold Regression," *Econometric Theory*, 32(4), 827–860.

Lee, S., Y. Liao, M. H. Seo, and Y. Shin (2020): "Factor-driven Two-regime Regression," *Annuals of Statistics, forthcoming.*

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, D., and S. Ling (2012): "On the Least Squares Estimation of Multiple-Regime Threshold Autoregressive Models," *Journal of Econometrics*, 167, 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.

Miao, K., L. Su, and W. Wang (2020): "Panel Threshold Regressions with Latent Group Structures," *Journal of Econometrics*, 215, 60–83.

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

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

Sen, P. K. (1976): "A Note on Invariance Principles for Induced Order Statistics," *Annuals of Probability*, 4(3), 474–479.

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 (2020): "Threshold Regression with a Threshold Boundary," *Journal of Business & Economic Statistics, forthcoming.*

YU, P., AND P. PHILLIPS (2018): "Threshold Regression with Endogeneity," *Journal of Econometrics*, 203, 50–68.