TESTING FOR HOMOGENEOUS THRESHOLDS IN THRESHOLD REGRESSION MODELS

Yoonseok Lee
Syracuse University

Yulong Wang
Syracuse University

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 also be 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, *Quarterly Journal of Economics* 123, 177–218) and statistically justify that the location of the tipping point varies across tracts.

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 test of threshold constancy).

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:

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

608

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, $(\beta_{01}, \beta_{02}^\mathsf{T}, \delta_{01})^\mathsf{T}$ 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 et al. (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 et al. (2021) and Yu and Fan (2021) or apply the method 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 aforementioned 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) and Elliott and Müller (2007, 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 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), 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 et al. (2020). In addition, Seo and Linton (2007), Lee et al. (2021), and Yu and Fan (2021) 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 time, our method essentially becomes the structural break model. See, among many others, Bai, Lumsdaine, and Stock (1998), Bai and Perron (1998), and Elliott and Müller (2007, 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 (2018, 2020) recently develop methods under the infill asymptotics. Casini and Perron (2021a,

2021b, 2022) 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

$$y_i = x_i^{\mathsf{T}} \beta_0 + x_i^{\mathsf{T}} \delta_0 \mathbf{1}[q_i \leq \gamma_{0i}] + u_i, \tag{2}$$

for $i = 1, \ldots, n$. The variables $(y_i, x_i^{\mathsf{T}}, q_i)^{\mathsf{T}} \in \mathbb{R}^{1+k+1}$ are observed, but the threshold parameter $\gamma_{0i} \in \mathbb{R}$ and the regression coefficients $\theta_0 = (\beta_0^{\mathsf{T}}, \delta_0^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{2k}$ 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; Hansen, 2000). Specifically, under the homogeneous threshold restriction, we estimate γ_0 by minimizing

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

in γ, where $(\widehat{\beta}^{\mathsf{T}}(\gamma), \widehat{\delta}^{\mathsf{T}}(\gamma))^{\mathsf{T}}$ are the least-squares estimators of (2) with a fixed γ. Once $\widehat{\gamma}$ is obtained, we let $\widehat{\theta} = (\widehat{\beta}^{\mathsf{T}}, \widehat{\delta}^{\mathsf{T}})^{\mathsf{T}} = (\widehat{\beta}^{\mathsf{T}}(\widehat{\gamma}), \widehat{\delta}^{\mathsf{T}}(\widehat{\gamma}))^{\mathsf{T}}$ and write $\widehat{u}_i = y_i - x_i^{\mathsf{T}} \widehat{\beta} - x_i^{\mathsf{T}} \widehat{\delta} \mathbf{1}[q_i \leq \widehat{\gamma}]$ 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

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

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 encompasses many different cases. For example, the threshold varies across i, where γ_{0i} can be either a discrete or continuous random variable. The threshold can be a nonconstant function of some random variables z_i, such as $\gamma_{0i} = z_i^\mathsf{T} \gamma$ for some parameter γ, as in Lee et al. (2021) and Yu and Fan (2021).

It is worthy to note that the alternative hypothesis in (3) 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, assume that γ_{0i} is i.i.d. and independent from $(q_i, x_i^\mathsf{T}, u_i)^\mathsf{T}$ and let

$$\gamma_{0i} = \begin{cases} \gamma_{0,1}, & \text{with probability } p_0, \\ \gamma_{0,2}, & \text{with probability } 1 - p_0, \end{cases} \tag{4}$$

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

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

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

$$y_i = x_i^\mathsf{T}\beta_0 + x_i^\mathsf{T}\delta_0\mathbf{1}\big[q_i \leq \gamma_{0,1}\big]\big(p_0 + \lambda_{i,1}\big) + x_i^\mathsf{T}\delta_0\mathbf{1}\big[q_i \leq \gamma_{0,2}\big]\big(1 - p_0 + \lambda_{i,2}\big) + u_i$$
$$= x_i^\mathsf{T}\beta_0 + x_i^\mathsf{T}\delta_{0,1}^*\mathbf{1}\big[q_i \leq \gamma_{0,1}\big] + x_i^\mathsf{T}\delta_{0,2}^*\mathbf{1}\big[q_i \leq \gamma_{0,2}\big] + u_i^*, \tag{5}$$

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

$$u_i^* = u_i + x_i^\mathsf{T}\delta_0\big\{\mathbf{1}\big[q_i \leq \gamma_{0,1}\big]\lambda_{i,1} + \mathbf{1}\big[q_i \leq \gamma_{0,2}\big]\lambda_{i,2}\big\}.$$

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

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. 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 et al., 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 $\widehat{\theta} = (\widehat{\beta}^{\mathsf{T}}, \widehat{\delta}^{\mathsf{T}})^{\mathsf{T}}$. Furthermore, it holds that (e.g., equation (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 $G_n(r) = n^{-1/2} \sum_{i=1}^{n} x_i \widehat{u}_i \mathbf{1}[q_i \leq Q(r)]$. Also define r_0 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; Kosorok, 2008), we can obtain that

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

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

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

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. Define[2]

$$D(r) = \mathbb{E}\left[x_i x_i^{\mathsf{T}} | q_i = Q(r)\right],$$
$$V(r) = \mathbb{E}\left[x_i x_i^{\mathsf{T}} u_i^2 | q_i = Q(r)\right],$$

---

[1] See Lemma A.4 in Hansen (2000).

[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 $[0, 1]$.

and monotonically increasing functions

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

for $r \in [\tau, 1-\tau]$ with some $\tau \in (0, 1/2)$ and for any $k \times 1$ vector v that satisfies $v^\mathsf{T} v = 1$.[3] Furthermore, let $F(\cdot)$ be the distribution function of q_i and define the $k \times 1$ vector of weight

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

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

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

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 twofold: 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,

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

for $s \in [0, 1]$, as $n \to \infty$. 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

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

where

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

$$\mathcal{G}_{2n}^*(s) = \frac{1}{\sqrt{1-g(r_0)}} \left\{(\mathcal{G}_n(1) - \mathcal{G}_n(s)) - \frac{1-s}{1-g(r_0)} (\mathcal{G}_n(1) - \mathcal{G}_n(g(r_0)))\right\}.$$

We suppose $r_0 \in (\tau, 1-\tau)$ so that we avoid the threshold $\gamma_0 = Q(r_0)$ being close to the boundary, and hence $g(r_0) \in (0, 1)$ holds by construction. Then $\mathcal{G}_{1n}^*(\cdot)$ and $\mathcal{G}_{2n}^*(\cdot)$ are, respectively, properly standardized, and both weakly converge to two independent standard Brownian Bridge processes. Based on this observation, we

---

[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 $v = (1, 0, \dots, 0)^\mathsf{T}$.

construct the test statistic as

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

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 $\mathcal{G}_{2n}^* (\cdot)$ is not properly centered, and the test statistic diverges as $n \to \infty$ because of the nonzero drift. See Section 3 for details.

2.3. Interpretation from Time Series Perspectives

As discussed above, our test uses the idea of Nyblom (1989) and Elliott and Müller (2007), which was originally developed in the time series context where q_i is 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 et al., 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 to the order of q_i. By sorting the random sample $\{q_i\}_{i=1}^n$ into the order statistics $q_{[1:n]} \leq q_{[2:n]} \leq \cdots \leq q_{[n:n]}$ and rearranging the observations according to the rank of q_i, we denote the reordered observations $(y_i, x_i^\mathsf{T})^\mathsf{T}$ associated with $q_{[i:n]}$ as $(y_{[i:n]}, x_{[i:n]}^\mathsf{T})^\mathsf{T}$, that is, $(y_{[i:n]}, x_{[i:n]}^\mathsf{T})^\mathsf{T} = (y_j, x_j^\mathsf{T})^\mathsf{T}$ if $q_{[i:n]} = q_j$.[4] These reordered statistics are called induced-order statistics or concomitants (e.g., Bhattachary, 1974; Sen, 1976; 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 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 "time." For the empirical distribution $\widehat{F}_n(\cdot)$, $\widehat{F}_n(q_{[i]}) = i/n$, hence, 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

---

[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.

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 $n^{-1} \sum_{i=1}^{n} x_i x_i^{\mathsf{T}} \mathbf{1} [q_i \leq Q(r)] = n^{-1} \sum_{i=1}^{n} x_i x_i^{\mathsf{T}} \mathbf{1} [F(q_i) \leq r]$ in (7) are asymptotically negligible in the sense that

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

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

where $n^{-1/2} \sum_{i=1}^{\lfloor rn \rfloor} x_{[i]} u_{[i]} = n^{-1/2} \sum_{i=1}^{n} x_{[i]} u_{[i]} \mathbf{1} [\widehat{F}_n(q_{[i]}) \leq r] = n^{-1/2} \sum_{i=1}^{n} x_i u_i \mathbf{1} [\widehat{F}_n(q_i) \leq r]$ and similarly for $n^{-1} \sum_{i=1}^{\lfloor rn \rfloor} x_{[i]} x_{[i]}^{\mathsf{T}}$. Therefore, it is asymptotically equivalent to rewrite $\mathcal{G}_n(\cdot)$ 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 $\mathcal{G}_n(\cdot)$ by

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

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

$$= \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} \sqrt{h(1-\tau)} g^{(1)}(i/n) v^{\mathsf{T}} D(i/n)^{-1} x_{[i]} u_{[i]}$$

$$- \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} \sqrt{h(1-\tau)} g^{(1)}(i/n) v^{\mathsf{T}} D(i/n)^{-1} x_{[i]} x_{[i]}^{\mathsf{T}} (\widehat{\beta} - \beta_0)$$

$$- \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor+1}^{\min\{\lfloor g^{-1}(s)n \rfloor, \lfloor r_0 n \rfloor\}} \sqrt{h(1-\tau)} g^{(1)}(i/n) v^{\mathsf{T}} D(i/n)^{-1} x_{[i]} x_{[i]}^{\mathsf{T}} (\widehat{\delta} - \delta_0)$$

$$\Rightarrow \sqrt{h(1-\tau)} \int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t) v^{\mathsf{T}} D(t)^{-1} V(t)^{1/2} dW_k(t) \tag{18}$$

$$- \sqrt{h(1-\tau)} \int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t) dt \cdot v^{\mathsf{T}} \Phi_\beta \tag{19}$$

$$- \sqrt{h(1-\tau)} \int_{g^{-1}(0)}^{\min\{g^{-1}(s), r_0\}} g^{(1)}(t) dt \cdot v^{\mathsf{T}} \Phi_\delta, \tag{20}$$

where $D(i/n)^{-1}$ asymptotically cancels out with $\mathbb{E}[x_{[i]} x_{[i]}^{\mathsf{T}}]$ by construction and W_k is the $k \times 1$ vector standard Wiener process defined on $[0, 1]$. Then, by the facts that $g(\tau) = 0$ 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, $v^{\mathsf{T}} D(\cdot)^{-1} V(\cdot) D(\cdot)^{-1} v$. Then, the first term becomes the standard Wiener process, and the limit of $\mathcal{G}_n(s)$ is obtained as (11). 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 is time, and the functions $D(\cdot)$ and $V(\cdot)$ are, respectively, constant matrices $\bar{D}$ and $\bar{V}$ under the piecewise stationarity assumption (e.g., Bai and Perron, 1998). Then $g(\cdot)$ reduces to the identity function, and the weight w_i becomes the constant $(v^\mathsf{T} \bar{D}^{-1} \bar{V} \bar{D}^{-1} v (1 - 2\tau))^{-1/2} \bar{D}^{-1} v$. 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, compared with (10), the partial sum process in (17) does not require to estimate the distribution function F directly. Therefore, the implementation of our test, as well as the derivation of its limiting distribution, becomes 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. $(x_i^\mathsf{T}, u_i, q_i)^\mathsf{T}$ *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* $\epsilon \in (0, 1/2)$; $(c_0^\mathsf{T}, \beta_0^\mathsf{T})^\mathsf{T}$ *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. $\mathbb{E}[x_i x_i^\mathsf{T}] > \mathbb{E}[x_i x_i^\mathsf{T} \mathbf{1}[q_i \leq Q(r)]] > 0$, *for any $r \in (0, 1)$.*
8. $\sup_{q \in \mathbb{R}} \mathbb{E}[||x_i u_i||^4 | q_i = q] < \infty$ *and* $\sup_{q \in \mathbb{R}} \mathbb{E}[||x_i||^4 | q_i = q] < \infty$.

Condition 1.1 assumes 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) and Lemma 3.2 in Bhattacharya (1984)) and 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 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 et al., 2016; Yu and Phillips, 2018). 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 regression 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; 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*

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

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. Moreover, the process $W_k(\cdot)$ and the normal variables Φ_β and Φ_δ are correlated since they both depend on the limit of the summation of x_iu_i. The exact correlation structure also involves $D(\cdot)$ and $V(\cdot)$ and hence can be complicated in general. Fortunately, the transformation (10) eliminates the effect of $V(\cdot)$ and $D(\cdot)$, and the self-normalization in (13) eliminates Φ_β and Φ_δ asymptotically under the null hypothesis. Then the test statistic (14) only involves the modified process $\mathcal{G}_n^*(\cdot)$ and becomes pivotal. We proceed to obtain its feasible sample analog and study its asymptotic properties. To this end, we first estimate

$D(r)$ and $V(r)$ as (e.g., Yang, 1985)

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

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

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

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

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, Cor. 1).

CONDITION 2. 1. *$K(\cdot)$ is Lipschitz continuous, continuously differentiable with bounded derivative, and symmetric around zero, which satisfies $\int K(t)\,dt = 1$, $\int tK(t)\,dt = 0$, $0 < \int t^2 K(t)\,dt < \infty$, $\lim_{t \to \infty} |t|K(t) = 0$, and $\lim_{t \to \infty} t^2 (\partial K(t)/\partial t) = 0$.*
2. *$b_n \to 0$, $nb_n/\log n \to \infty$, and $n^{1/4}b_n \to \infty$ as $n \to \infty$.*

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

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

where $\widehat{g}^{(1)}(i/n) = \{v^{\mathsf{T}}\widehat{D}(i/n)^{-1}\widehat{V}(i/n)\widehat{D}(i/n)^{-1}vh(1-\tau)\}^{-1}$, 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 $v^{\mathsf{T}}v = 1$, 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*

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

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

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

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

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 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.[5]

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

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

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), where γ_{0i} is i.i.d. and independent of $(q_i, x_i^\mathsf{T}, u_i)^\mathsf{T}$.

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 $v^\mathsf{T}v = 1$. 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.

---

[5]We focus on the alternative model such that the threshold is exogenous. More precisely, γ_{0i} in (3) is i.i.d. and independent of $(q_i, x_i^\mathsf{T}, u_i)^\mathsf{T}$. Such an assumption does not change the null distribution of our test but substantially simplifies the power analysis.

TABLE 1. Simulated critical values of the CT test

$\mathbb{P}(\int_0^1 \mathcal{B}_2(t)^\mathsf{T} \mathcal{B}_2(t)\,dt > \mathrm{cv})$	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 50,000 replications and 5,000 step approximations to the continuous-time process.

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 $r \in \{(\lfloor \tau n \rfloor + 1)/n, (\lfloor \tau n \rfloor + 2)/n, \ldots, \lfloor (1-\tau)n \rfloor/n\}$, 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 $s \in \{1/n, 2/n, \ldots, 1\}$ 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 the local alternative model given as

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

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

Though appearing differently, (30) is essentially equivalent to the local alternative of (3). Recall that the alternative hypothesis is $\mathbb{P}(\gamma_{0i} = \gamma_0) < 1$ for any $\gamma_0 \in \Gamma$. Hence, a genuine way of constructing the local alternative is to consider $\mathbb{P}(\gamma_{0i} = \gamma_0) = 1 - n^{-\ell}$, for some $\ell > 0$. To this end, we let

$$\gamma_{0i}^* = \begin{cases} \gamma_0, & \text{with probability } 1 - n^{-(1/2-\epsilon)}, \\ \gamma_0 + \Delta_i, & \text{with probability } n^{-(1/2-\epsilon)}, \end{cases} \tag{31}$$

where Δ_i is some random variable that describes the local deviation. We consider $\Delta_i > 0$ without loss of generality. Now, define $\lambda_{i,1}^* = \mathbf{1}\left[\gamma_{0i}^* = \gamma_0\right] - (1 - n^{-(1/2-\epsilon)})$ and $\lambda_{i,2}^* = \mathbf{1}\left[\gamma_{0i}^* = \gamma_0 + \Delta_i\right] - n^{-(1/2-\epsilon)}$, which yield that

$$\mathbf{1}\left[q_i \le \gamma_{0i}^*\right] = \mathbf{1}[q_i \le \gamma_0]\mathbf{1}\left[\gamma_{0i}^* = \gamma_0\right] + \mathbf{1}[q_i \le \gamma_0 + \Delta_i]\mathbf{1}\left[\gamma_{0i}^* = \gamma_0 + \Delta_i\right]$$
$$= \mathbf{1}[q_i \le \gamma_0](1 - n^{-(1/2-\epsilon)} + \lambda_{i,1}^*) + \mathbf{1}[q_i \le \gamma_0 + \Delta_i](n^{-(1/2-\epsilon)} + \lambda_{i,2}^*).$$

We assume Δ_i is i.i.d. with the distribution function F_Δ and independent of $(q_i, x_i^\mathsf{T}, u_i)^\mathsf{T}$ as in Theorem 1. We allow Δ_i to be a constant. Recall $\delta_0 = c_0 n^{-\epsilon}$ for nonzero c_0.

Then, the threshold regression model with (31) can be rewritten as (30) in the following way:

$$y_i = x_i^\mathsf{T}\beta_0 + x_i^\mathsf{T}\delta_0\mathbf{1}\left[q_i \le \gamma_{0i}^*\right] + u_i$$
$$= x_i^\mathsf{T}\beta_0 + x_i^\mathsf{T}\delta_0\mathbf{1}[q_i \le \gamma_0] + x_i^\mathsf{T}\delta_0 n^{-(1/2-\epsilon)}\mathbf{1}[\gamma_0 < q_i \le \gamma_0 + \Delta_i]$$
$$\quad + x_i^\mathsf{T}\delta_0\left\{\mathbf{1}[q_i \le \gamma_0]\lambda_{i,1}^* + \mathbf{1}[q_i \le \gamma_0 + \Delta_i]\lambda_{i,2}^*\right\} + u_i$$
$$= x_i^\mathsf{T}\beta_0 + x_i^\mathsf{T}\delta_0\mathbf{1}[q_i \le \gamma_0] + x_i^\mathsf{T}\{n^{-1/2}\alpha(q_i)\} + u_i^*,$$

where

$$\alpha(q) = c_0\mathbb{E}\left[\mathbf{1}[\gamma_0 < q_i \le \gamma_0 + \Delta_i]|x_i, q_i = q\right] \tag{32}$$

$$= \begin{cases} c_0(1 - F_\Delta(q - \gamma_0)), & \text{if } q > \gamma_0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$u_i^* = u_i + x_i^\mathsf{T}\delta_0\left\{\mathbf{1}[q_i \le \gamma_0]\lambda_{i,1}^* + \mathbf{1}[q_i \le \gamma_0 + \Delta_i]\lambda_{i,2}^*\right\}$$
$$\quad + x_i^\mathsf{T}\delta_0 n^{-(1/2-\epsilon)}\{\mathbf{1}[\gamma_0 < q_i \le \gamma_0 + \Delta_i] - \mathbb{E}[\mathbf{1}[\gamma_0 < q_i \le \gamma_0 + \Delta_i]|x_i, q_i]\}.$$

Note that $\mathbb{E}\left[u_i^*|x_i, q_i\right] = 0$ by construction and $\alpha(q)$ cannot be constant over $q > \gamma_0$.

The shrinking magnitude of the local deviation in (30) is of the order $n^{-1/2}$, with which the *CT* test has a nontrivial 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 $\theta_0 = (\beta_0^\mathsf{T}, \delta_0^\mathsf{T})^\mathsf{T}$. 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 the conditions in Theorem 1 hold. Then, under the local alternative in (30),*

$$CT_n \to_d \int_0^1 (\mathcal{B}_2(t) + \mu(t))^\mathsf{T}(\mathcal{B}_2(t) + \mu(t))\, dt,$$

as $n \to \infty$, where $\mu(t) = (\mu_1(t), \mu_2(t))^\top$ with

$$
\mu_1(t) = \sqrt{\frac{h(1-\tau)}{g(r_0)}} \left\{ \int_\tau^{g^{-1}(t)} \Psi_v(s)ds - \frac{t}{g(r_0)} \int_\tau^{r_0} \Psi_v(s)ds \right\},
$$

$$
\mu_2(t) = \sqrt{\frac{h(1-\tau)}{1-g(r_0)}} \left\{ \int_{g^{-1}(t)}^{1-\tau} \Psi_v(s)ds - \frac{1-t}{1-g(r_0)} \int_{r_0}^{1-\tau} \Psi_v(s)ds \right\},
$$

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

The local deviation $n^{-1/2}\alpha(\cdot)$ introduces a potentially nonzero drift function $\mu(\cdot)$ to the standard Brownian bridge. For any given $t \in (0,1)$, the scaled integrand $(\mathcal{B}_2(t) + \mu(t))^\top (\mathcal{B}_2(t) + \mu(t))/(t(1-t))$ has a noncentral chi-square distribution with two degrees of freedom and the noncentrality parameter given by $\mu(t)^\top \mu(t) /(t(1-t))$ (e.g., Andrews, 1993, p. 842). As long as $\alpha(\cdot)$ is nonconstant either before or after the first break, at least one component of $\mu(\cdot)$ is not uniformly zero and then leads to a nontrivial local power.

Moreover, when $\alpha(\cdot)$ is a $k \times 1$ vector of constants, say α, we have $\Psi_v(\cdot) = g^{(1)}(\cdot) v^\top \alpha$ and it is readily verified that $\mu(t) = 0$, for all $t \in (0,1)$. In fact, in view of (32), $\alpha(\cdot)$ can be a constant only when F_Δ is the step function that jumps from zero to one at γ_0, which corresponds to the null model. From this observation, we can see that the *CT* test has a nontrivial asymptotic local power under (30) for any nonconstant function $\alpha(\cdot)$.

To better understand the drift function $\mu(\cdot)$, we illustrate the case with two thresholds. Assume that the local alternative model has a second threshold $\gamma_1 < \gamma_0$. Accordingly, we let $\alpha(q_i) = \alpha_0 \mathbf{1}[q_i \leq \gamma_1]$ with some nonzero vector α_0. Then, $\mu_2(t)$ is zero for $t \in [g(r_0), 1)$. Denote $\gamma_1 = Q(r_1)$ 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, $\sqrt{h(1-\tau)} \min\{s, g(r_1)\} v^\top \alpha_0$. This nonzero drift term cannot be removed by the standardization in (13), and thus we have

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

over the region $t \in (0, g(r_0))$ in the limit experiment, which yields nontrivial powers. The optimal choice of v could be obtained by maximizing the local power (cf. Andrews, 1993; 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 $y_i = x_i^{\mathsf{T}} \beta_0 + x_i^{\mathsf{T}} \delta_0 \mathbf{1}[q_i \leq 0] + u_i$;

DGP-2 $y_i = x_i^{\mathsf{T}} \beta_0 + x_i^{\mathsf{T}} \delta_0 \mathbf{1}[q_i \leq 0] + x_i^{\mathsf{T}} \alpha(q_i) + u_i$;

DGP-3 $y_i = x_i^{\mathsf{T}} \beta_0 + x_i^{\mathsf{T}} \delta_0 (\mathbf{1}[q_i \leq 0] + \mathbf{1}[q_i > 0.5]) + u_i$,

where $x_i = (x_{1i}, x_{2i})^{\mathsf{T}} \in \mathbb{R}^2$ with the first element $x_{1i} = 1$ and x_{2i} 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 $\iota_2 = (1,1)^{\mathsf{T}}$. In DGP-2, we set $\alpha(q) = |q| \iota_2$.

These DGPs correspond to each of the following three different threshold specifications: (i) one single threshold at zero; (ii) one first threshold at zero and an additional drift function $\alpha(\cdot)$; and (iii) two thresholds at 0 and 0.5. In particular, DGP-1 corresponds to the null hypothesis of the homogeneous threshold in (3). DGP-2 corresponds to an alternative model as in (30), and DGP-3 corresponds to the alternative model discussed in the end of Section 3.2. We set $\tau = 0.1$ and $v = (v_1, v_2)^{\mathsf{T}}$ to be proportional to $(1, 1/\mathbb{E}[x_{2i}^2])^{\mathsf{T}}$ with $v^{\mathsf{T}} v = 1$.[6] We use the rule-of-thumb choice of the bandwidth $b_n = (1/12)^{1/2} n^{-1/5}$ and the Gaussian kernel. Other choices of bandwidth, kernel, and τ are also implemented, which lead to negligible changes. The sample sizes are $n = 500, 1,000$, and $1,500$, and the significance level is 5%. The results are based on $1,000$ 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

---

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

TABLE 2. Rejection probabilities when q and x are independent

		DGP-1			DGP-2			DGP-3		
δ	$n=$	500	1,000	1,500	500	1,000	1,500	500	1,000	1,500
CT test										
0.25		0.03	0.05	0.05	0.72	1.00	1.00	0.24	0.50	0.66
0.50		0.03	0.05	0.05	0.94	1.00	1.00	0.70	0.96	1.00
0.75		0.03	0.05	0.05	0.97	1.00	1.00	0.92	1.00	1.00
1.00		0.04	0.05	0.05	0.94	1.00	1.00	0.94	1.00	1.00
$F(2|1)$ test										
0.25		0.01	0.01	0.01	1.00	1.00	1.00	0.15	0.50	0.76
0.50		0.01	0.01	0.01	1.00	1.00	1.00	0.86	1.00	1.00
0.75		0.01	0.01	0.02	1.00	1.00	1.00	1.00	1.00	1.00
1.00		0.00	0.01	0.01	1.00	1.00	1.00	1.00	1.00	1.00
BIC1										
0.25		0.24	0.04	0.01	1.00	1.00	1.00	0.93	0.94	0.96
0.50		0.05	0.03	0.02	1.00	1.00	1.00	0.98	1.00	1.00
0.75		0.07	0.03	0.03	1.00	1.00	1.00	1.00	1.00	1.00
1.00		0.06	0.04	0.03	1.00	1.00	1.00	1.00	1.00	1.00
BIC3										
0.25		0.97	0.74	0.34	1.00	1.00	1.00	1.00	1.00	1.00
0.50		0.04	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
0.75		0.00	0.00	0.00	0.99	1.00	1.00	1.00	1.00	1.00
1.00		0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00

Notes: 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 1,000 simulations. The significance level is 5%. Data are generated from three DGPs with $(q_i, u_i, x_{2i}) \sim iid \mathcal{N}(0, I_3)$.

count the mis-selection probability when $m = 1$ as the rejection probability. We follow Gonzalo and Pitarakis (2002) to choose the Bayesian information criterion (BIC) approach by setting $\varphi_n = \log n$ and $3 \log n$, denoted by BIC1 and BIC3, respectively, in Tables 2 and 3. 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 $(q_i, u_i, x_{2i}) \sim \mathcal{N}(0, I_3)$. 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, whereas the *CT* test has

TABLE 3. Rejection probabilities when q and x are dependent

δ	$n =$	DGP-1			DGP-2			DGP-3		
		500	1,000	1,500	500	1,000	1,500	500	1,000	1,500
CT test										
0.25		0.03	0.03	0.04	0.25	0.50	0.80	0.09	0.12	0.23
0.50		0.04	0.03	0.05	0.35	0.74	0.93	0.28	0.53	0.69
0.75		0.04	0.05	0.05	0.50	0.83	0.97	0.50	0.81	0.95
1.00		0.05	0.04	0.04	0.56	0.85	0.96	0.72	0.97	1.00
$F(2\vert1)$ test										
0.25		0.11	0.16	0.17	1.00	1.00	1.00	0.19	0.36	0.55
0.50		0.11	0.16	0.15	0.99	1.00	1.00	0.56	0.93	0.99
0.75		0.10	0.14	0.14	0.98	1.00	1.00	0.91	1.00	1.00
1.00		0.10	0.13	0.14	1.00	1.00	1.00	0.99	1.00	1.00
BIC1										
0.25		0.61	0.43	0.32	1.00	1.00	1.00	0.90	0.89	0.91
0.50		0.33	0.30	0.28	1.00	1.00	1.00	0.94	0.98	1.00
0.75		0.32	0.29	0.28	1.00	1.00	1.00	0.99	1.00	1.00
1.00		0.35	0.31	0.30	1.00	1.00	1.00	1.00	1.00	1.00
BIC3										
0.25		0.99	0.96	0.86	0.90	1.00	1.00	1.00	1.00	1.00
0.50		0.59	0.07	0.00	0.69	1.00	1.00	1.00	0.98	0.98
0.75		0.02	0.00	0.00	0.55	0.99	1.00	0.98	0.99	1.00
1.00		0.00	0.00	0.00	0.61	1.00	1.00	0.99	1.00	1.00

Notes: Entries are rejection probabilities under the null hypothesis in (3) of the *CT* test, the $F(2\vert1)$ test by Bai and Perron (1998), the model selection using the BIC by Gonzalo and Pitarakis (2002). The results are based on 1,000 simulations. The significance level is 5%. Data are generated from three DGPs with $q_i \sim iid\mathcal{N}(0,1)$, $x_{2i}\vert q_i = q \sim iid\mathcal{N}\left(0, 1/(1+q^2)\right)$, and $u_i\vert x_{2i} = x \sim iid\mathcal{N}(0, 1+x^2)$.

approximately the correct size. The middle three columns show the rejection probabilities under the alternative model with an additional drift function $\alpha(\cdot)$. The *CT* test and the $F(2\vert1)$ test have similar powers. Third, the next three columns show the powers under the alternative with two thresholds. This is the exact alternative that the $F(2\vert1)$ test is designed for, whereas our *CT* test still achieves comparable powers. Fourth, the model selection based on BIC1 or BIC3 has good selection probabilities. However, its performance is very sensitive to the choice of the tuning parameter as we compare the results between BIC1 and BIC3. In particular, BIC3 uses a larger tuning parameter (i.e., heavier penalty) than BIC1, which leads to substantially lower rejection probabilities. This feature is also seen in Table 3.

In Table 3, we introduce some dependence between q_i and x_i, and conditional heteroskedastic errors u_i. In particular, we generate data from each DGP with $q_i \sim iid\mathcal{N}(0,1)$, $x_{2i}|q_i = q \sim iid\mathcal{N}(0,1/(1+q^2))$, and $u_i|x_{2i} = x \sim iid\mathcal{N}(0,1+x^2)$. Several findings can be summarized as follows. First, as expected, the *CT* test is the only one that satisfies the size constraint under the null hypothesis. It also has nontrivial powers under the alternative models, especially when δ and n are large. Second, the $F(2|1)$ test fails to control size since its asymptotic distribution is contaminated by the rank-varying moments. Third, the mis-selection probabilities from BIC1 are far from 5%. This issue can be alleviated by choosing a larger tuning parameter as in BIC3, which again leads to severe under-rejections.

5. APPLICATION: TIPPING POINT AND SOCIAL SEGREGATION

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

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

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 is 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. For most cities and all three periods, they find that white population flows exhibit the tipping point behavior, with the estimated tipping points γ_0 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 et al. (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 $v = (1,0,\ldots,0)^\mathsf{T}$ since only the constant term involves a coefficient change. We also follow Card et al. (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,

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	1,263	17.47	0.000
New York	315	16.08	0.000
Washington D.C.	719	15.54	0.000

Notes: 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 et al. (2008).

which are, hence, not reported. These results suggest that we need to use a more flexible form of threshold in the tipping point analysis.

6. CONCLUSION

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

Although 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 et al., 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, although 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. We leave these questions for future research.

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. Proof of the Results in Section 2

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

$$\mathcal{G}_n(s)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i^\mathsf{T} x_i u_i \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right]$$

$$- \frac{1}{n} \sum_{i=1}^n w_i^\mathsf{T} x_i x_i^\mathsf{T} \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right]\{\sqrt{n}(\widehat{\beta} - \beta_0) + \mathbf{1}[q_i \le Q(r_0)]\sqrt{n}(\widehat{\delta} - \delta_0)\}$$

$$- \frac{1}{n} \sum_{i=1}^n w_i^\mathsf{T} x_i x_i^\mathsf{T} \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right]\{\mathbf{1}[q_i \le \widehat{\gamma}] - \mathbf{1}[q_i \le Q(r_0)]\}\sqrt{n}\widehat{\delta}.$$

$$\tag{A.1}$$

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

$$\mathcal{G}_{nA}(s) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i^\mathsf{T} x_i u_i \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right] \Rightarrow W_1(s),$$

$$\mathcal{G}_{nB}(s) \equiv \frac{1}{n} \sum_{i=1}^n w_i^\mathsf{T} x_i x_i^\mathsf{T} d_0 \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right] \to_p s v^\mathsf{T} d_0 h_\tau^{1/2}$$

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

$$Cov\left[\mathcal{G}_{nA}(s), \mathcal{G}_{nA}(s')\right]$$

$$= \mathbb{E}\left[\left(w_i^\mathsf{T} x_i u_i\right)^2 \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right]\right]$$

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

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

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

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

$$\mathbb{E}[\mathcal{G}_{nB}(s)] = \mathbb{E}\left[w_i^\mathsf{T} x_i x_i^\mathsf{T} d_0 \mathbf{1}\left[Q(\tau) \le q_i \le Q(g^{-1}(s))\right]\right]$$

$$= h_\tau^{1/2} \int_\tau^{g^{-1}(s)} \mathbb{E}\left[\left.\frac{v^\mathsf{T} D(F(q_i))^{-1} x_i x_i^\mathsf{T} d_0}{v^\mathsf{T} D(F(q_i))^{-1} V(F(q_i)) D(F(q_i))^{-1} v h_\tau}\right| F(q_i) = r\right] dr$$

$$= h_\tau^{1/2} v^\mathsf{T} d_0 \int_\tau^{g^{-1}(s)} \frac{1}{v^\mathsf{T} D(r)^{-1} V(r) D(r)^{-1} v h_\tau} dr = s v^\mathsf{T} d_0 h_\tau^{1/2}.$$

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

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

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

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

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

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 $\|\sqrt{n\delta}\| \leq \|\sqrt{n}(\hat{\delta} - \delta_0)\| + \|\sqrt{n}\delta_0\| = O_p(1) + O_p(n^{1/2-\epsilon})$ with $\epsilon \in (0, 1/2)$. Let $E_{\delta n}$ be the event that $\|\sqrt{n\delta}\| \leq C_\delta n^{1/2-\epsilon}$, for some $0 < C_\delta < \infty$, and then $\mathbb{P}\left(E_{\delta n}^c \right) \leq \varepsilon$, for any $\varepsilon > 0$, if n is sufficiently large. Now, let $E_{\gamma n}$ be the event that $\hat{\gamma} \in (\gamma_0 - C_\gamma n^{-1+2\epsilon}, \gamma_0 + C_\gamma n^{-1+2\epsilon})$, 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,

$$\mathbb{P} \left(\sup_{s \in [\tau, 1-\tau]} |\Upsilon_n(s)| > \eta \right) \tag{A.2}$$

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

$$\leq \eta^{-1} C n^{1/2-\epsilon} \mathbb{E} \left[\left\| x_i x_i^{\mathsf{T}} \left| \mathbf{1}[q_i \leq \hat{\gamma}] - \mathbf{1}[q_i \leq \gamma_0] \right| \mathbf{1}[E_{\gamma n}] \right\| \right] + 2\varepsilon$$

$$\leq \eta^{-1} C n^{1/2-\epsilon} \left\| \sup_{q \in \mathbb{R}} \mathbb{E} \left[x_i x_i^{\mathsf{T}} | q_i = q \right] \right\| \left| \mathbb{P}(q_i \leq \gamma_0 + C_\gamma n^{-1+2\epsilon}) - \mathbb{P}(q_i \leq \gamma_0) \right| + 2\varepsilon$$

$$\leq \eta^{-1} C' n^{-1/2+\epsilon} + 2\varepsilon$$

$$\leq 3\varepsilon,$$

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

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

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 that, for any $k \times 1$ nonzero vector v,

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

Note that $v^{\mathsf{T}} x_{[i]} u_{[i]}$ is a scalar random variable and is the induced-order statistics of $v^{\mathsf{T}} x_i u_i$ associated with q_i. We now check Conditions 1–3 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 $\mathbb{E}[v^{\mathsf{T}} x_i u_i | q_i] = 0$ almost surely and

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

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

$$\phi_V(r) = \int_0^r v^{\mathsf{T}} V(t) v\, dt.$$

By Theorem 2 in Bhattacharya (1974), we have

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

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 $\xi_i = v^{\mathsf{T}} x_i x_i^{\mathsf{T}} v$ 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., page 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 $n^{-1/2}\sum_{i=1}^{\lfloor rn \rfloor}x_{[i]}u_{[i]}=n^{-1/2}\sum_{i=1}^{n}x_{[i]}u_{[i]}\mathbf{1}[\widehat{F}_n(q_{[i]})\leq r]=n^{-1/2}\sum_{i=1}^{n}x_i u_i\mathbf{1}[\widehat{F}_n(q_i)\leq r]=n^{-1/2}\sum_{i=1}^{n}x_i u_i\mathbf{1}[q_i\leq\widehat{Q}_n(r)]$ and similarly $n^{-1}\sum_{i=1}^{\lfloor rn \rfloor}x_{[i]}x_{[i]}^{\mathsf{T}}=n^{-1}\sum_{i=1}^{n}x_i x_i^{\mathsf{T}}\mathbf{1}[q_i\leq\widehat{Q}_n(r)]$.

LEMMA A.3. *Under Condition 1,*

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

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

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,

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

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 $n^{-1} \sum_{i=1}^{n} x_i x_i^{\mathsf{T}} \mathbf{1}[q_i \leq \gamma] \to_p \mathcal{M}(\gamma)$ 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 in (A.7) using (A.4).   ∎

A.2. Proofs of the Results in Section 3

Proof of Lemma A.1.   Note that

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

where the continuous mapping theorem yields

$$
G_{n1}(r) \Rightarrow \int_0^r V(t)^{1/2} dW_k(t),
$$

$$
G_{n2}(r) \Rightarrow \left(\int_0^r D(t) \, dt \right) \Phi_\beta - \left(\int_0^{\min\{r, r_0\}} D(t) \, dt \right) \Phi_\delta
$$

from Lemma A.2 and (6), since $\mathbf{1}[\widehat{F}_n(q_{[i]}) \leq r_0] = \mathbf{1}[i/n \leq r_0]$. 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,

$$
\|G_{n4}(r)\| \leq \left\| \frac{1}{n} \sum_{i=1}^{n} x_i x_i^{\mathsf{T}} \left| \mathbf{1}[q_i \leq \widehat{\gamma}] - \mathbf{1}[q_i \leq \gamma_0] \right| \right\| \| \sqrt{n}\delta \|,
$$

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 $\|\sqrt{n}\delta\| \leq \|\sqrt{n}(\widehat{\delta} - \delta_0)\| + \|\sqrt{n}\delta_0\| = O_p(1) + O_p(n^{1/2-\epsilon})$ with $\epsilon \in (0, 1/2)$. Let $E_{\delta n}$ be the event that $\|\sqrt{n}\delta\| \leq C_\delta n^{1/2-\epsilon}$ 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. 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,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{r \in [0,1]} \|G_{n4}(r)\| > \eta \right) \\
&\leq \mathbb{P}\left(\left\{ \sup_{r \in [0,1]} \|G_{n4}(r)\| > \eta \right\} \cap E_{\gamma n} \cap E_{\delta n} \right) + \mathbb{P}\left(E_{\gamma n}^c \right) + \mathbb{P}\left(E_{\delta n}^c \right) \\
&\leq \eta^{-1} C n^{1/2-\epsilon} \mathbb{E}\left[\left\| x_i x_i^{\mathsf{T}} \left| \mathbf{1}[q_i \leq \widehat{\gamma}] - \mathbf{1}[q_i \leq \gamma_0] \right| \mathbf{1}[E_{\gamma n}] \right\| \right] + 2\varepsilon \\
&\leq \eta^{-1} C' n^{-1/2+\epsilon} + 2\varepsilon \\
&\leq 3\varepsilon,
\end{aligned}
$$

for some $0 < C, C' < \infty$, where the second inequality is by Markov's inequality and 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*

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

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 $\widehat{u}_i = u_i - x_i(\widehat{\beta} - \beta_0) - x_i(\widehat{\delta} - \delta_0)\mathbf{1}[q_i \leq \gamma_0] - x_i\widehat{\delta}(\mathbf{1}[q_i \leq \widehat{\gamma}] - \mathbf{1}[q_i \leq \gamma_0])$, we have

$$\left|\widehat{V}(r) - \widehat{V}^0(r)\right| = \left|\frac{1}{n} \sum_{i=1}^{n} x_{[i]}^2 \left(\widehat{u}_{[i]} + u_{[i]}\right)\left(\widehat{u}_{[i]} - u_{[i]}\right) K_i(r)\right| \tag{A.11}$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \left|x_i^3 \left(\widehat{u}_i + u_i\right)(\widehat{\beta} - \beta_0) K_i(r)\right|$$

$$+ \frac{1}{n} \sum_{i=1}^{n} \left|x_i^3 \left(\widehat{u}_i + u_i\right)(\widehat{\delta} - \delta_0)\mathbf{1}[q_i \leq \gamma_0] K_i(r)\right|$$

$$+ \frac{1}{n} \sum_{i=1}^{n} \left|x_i^3 \left(\widehat{u}_i + u_i\right)\widehat{\delta}\left\{\left(\mathbf{1}[q_i \leq \widehat{\gamma}] - \mathbf{1}[q_i \leq \gamma_0]\right)\right\} K_i(r)\right|$$

$$\equiv V_{1n}(r) + V_{2n}(r) + V_{3n}(r).$$

Let $E_{\theta n}$ be the event that $\widehat{\theta} = (\widehat{\beta}^\mathsf{T}, \widehat{\delta}^\mathsf{T})^\mathsf{T} \in \mathcal{B}_{Cn^{-1/2}}(\theta_0)$ 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 $\mathbb{P}(E_{\theta n}^c) \leq \varepsilon$ and $\mathbb{P}\left(E_{\gamma n}^c\right) \leq \varepsilon$ for any $\varepsilon > 0$ if C and n are large enough. Then, for any $\eta > 0$,

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

$$\leq \mathbb{P}\left(\left\{\sup_{r \in [\tau, 1-\tau]} |V_{1n}(r)| > \eta\right\} \cap E_{\gamma n} \cap E_{\theta n}\right) + \mathbb{P}(E_{\gamma n}^c \cup E_{\theta n}^c)$$

$$\leq \eta^{-1} \max_{1 \leq i \leq n} \sup_{r \in [0,1]} K_i(r) \times \mathbb{E}\left[\left|x_i^3 \left(\widehat{u}_i + u_i\right)(\widehat{\beta} - \beta_0)\right| \Big| E_{\theta n}\right] + 2\varepsilon$$

$$\leq \eta^{-1} \max_{1 \leq i \leq n} \sup_{r \in [0,1]} K_i(r) \times \left\{2\mathbb{E}\left[\left|x_i^3 u_i(\widehat{\beta} - \beta_0)\right| \Big| E_{\theta n}\right] + \mathbb{E}\left[\left|x_i^4(\widehat{\beta} - \beta_0)^2\right| \Big| E_{\theta n}\right]\right.$$

$$+\mathbb{E}\left[\left|x_i^4\mathbf{1}[q_i\leq\gamma_0]\,(\widehat{\delta}-\delta_0)(\widehat{\beta}-\beta_0)^2\right||E_{\theta n}\right]$$

$$+\mathbb{E}\left[\left|x_i^4\widehat{\delta}(\widehat{\beta}-\beta_0)\left(\mathbf{1}[q_i\leq\widehat{\gamma}]-\mathbf{1}[q_i\leq\gamma_0]\right)\right||E_{\gamma n}\cap E_{\theta n}\right]\right\}+2\varepsilon$$

$$\leq Cn^{-1}n^{-1/2}b_n^{-1}\left(2\mathbb{E}\left[\left|x_i^3u_i\right|\right]+\mathbb{E}\left[\left|x_i^4\right|\right]\right)+2\varepsilon$$

$$\leq 3\varepsilon$$

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. Denote f_v as the density of $v_i=F(q_i)$ and $f_{x,u,v}$ as the joint density of (x_i,u_i,v_i). Note that

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

where $f_v(r)=1$ since v_i is standard uniform.

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) and (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 $\eta>0$,

$$\mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\widehat{V}^0(r)-\mathbb{E}[\widehat{V}^0(r)]\right|>\eta\right)\leq\mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\widehat{V}^0(r)-\widehat{V}^\kappa(r)\right|>\eta/3\right)\quad\textbf{(A.12)}$$

$$+\mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\mathbb{E}[\widehat{V}^0(r)]-\mathbb{E}[\widehat{V}^\kappa(r)]\right|>\eta/3\right)$$

$$+\mathbb{P}\left(\sup_{r\in[\tau,1-\tau]}\left|\widehat{V}^\kappa(r)-\mathbb{E}[\widehat{V}^\kappa(r)]\right|>\eta/3\right)$$

$$\equiv P_{n1}+P_{n2}+P_{n3}.$$

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

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

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,

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

and hence $P_{n2} \leq 3C_1/(\eta b_n \kappa_n)$ as well. For P_{n3}, Lemma A.6 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

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

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 $\widehat{g}(r)$ 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 $P_{n3} \leq (\eta/3)^{-1}C(\log n/(nb_n))^{1/2}$, 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 by $\mathcal{I}_j$, for $j = 1, \ldots, m_n$, and choose m_n later. The triangular inequality yields

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

where

$$T_{1n}^\kappa = \max_{1 \leq j \leq m_n} \sup_{r \in \mathcal{I}_j} \left| \widehat{V}^\kappa (r) - \widehat{V}^\kappa (r_j) \right|,$$

$$T_{2n}^\kappa = \max_{1 \leq j \leq m_n} \sup_{r \in \mathcal{I}_j} \left| \mathbb{E}[\widehat{V}^\kappa (r)] - \mathbb{E}[\widehat{V}^\kappa (r_j)] \right|,$$

$$T_{3n}^\kappa = \max_{1 \leq j \leq m_n} \left| \widehat{V}^\kappa (r_j) - \mathbb{E}[\widehat{V}^\kappa (r_j)] \right|.$$

We first bound T_{3n}^κ. Let

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

and then

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

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

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

since $\mathbb{E}[Z_{n,i}^\kappa (r)] = 0$ and $1 + x \leq \exp(x)$, for $x \geq 0$. By 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,

$$\mathbb{P}\left(\left| \widehat{V}^\kappa (r) - \mathbb{E}[\widehat{V}^\kappa (r)] \right| > \eta_n \right) = \mathbb{P}\left(\widehat{V}^\kappa (r) - \mathbb{E}[\widehat{V}^\kappa (r)] > \eta_n \right) + \mathbb{P}\left(-\widehat{V}^\kappa (r) + \mathbb{E}[\widehat{V}^\kappa (r)] > \eta_n \right)$$

$$\leq \frac{\mathbb{E}\left[\exp\left(\psi_n \sum_{i=1}^n Z_{n,i}^\kappa (r) \right) \right] + \mathbb{E}\left[\exp\left(-\psi_n \sum_{i=1}^n Z_{n,i}^\kappa (r) \right) \right]}{\exp(\psi_n \eta_n)}$$

$$\leq 2 \exp(-\psi_n \eta_n) \exp\left(\psi_n^2 \sum_{i=1}^n \mathbb{E}\left[(Z_{n,i}^\kappa (r))^2 \right] \right)$$

$$\leq 2 \exp(-\psi_n \eta_n) \exp\left(\psi_n^2 C_3 \kappa_n^2 / (n b_n) \right)$$

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

$$\sum_{i=1}^n \mathbb{E}\left[(Z_{n,i}^\kappa (r))^2 \right] \leq n^{-2} \sum_{i=1}^n \mathbb{E}\left[x_{[i]}^4 u_{[i]}^4 K_i^2 (r) \mathbf{1}[x_{[i]}^2 u_{[i]}^2 \leq \kappa_n] \right] \leq C_3 \kappa_n^2 (n b_n)^{-1}$$

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.,

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

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

$$\mathbb{P}(T_{3n}^\kappa > \eta_n) = \mathbb{P}\left(\max_{1 \leq j \leq m_n} |\widehat{V}^\kappa(r_j) - \mathbb{E}[\widehat{V}^\kappa(r_j)]| > \eta_n\right)$$

$$\leq m_n \sup_{s \in [\tau, 1-\tau]} \mathbb{P}(|\widehat{V}^\kappa(r) - \mathbb{E}[\widehat{V}^\kappa(r)]| > \eta_n) \leq 2 \frac{m_n}{n^{C^* - C_4}}.$$

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

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

by the Borel–Cantelli lemma.

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

$$\mathbb{E}|\widehat{V}^\kappa(r) - \widehat{V}^\kappa(r_j)| = \mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^n x_{[i]}^2 u_{[i]}^2 (K_i(r) - K_i(r_j)) \mathbf{1}[x_{[i]}^2 u_{[i]}^2 \leq \kappa_n]\right|\right]$$

$$\leq C_4 (1 - 2\tau) \kappa_n / \left(m_n b_n^2\right)$$

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

$$|\mathbb{E}[\widehat{V}^\kappa(r)] - \mathbb{E}[\widehat{V}^\kappa(r_j)]| \leq \mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^n x_{[i]}^2 u_{[i]}^2 (K_i(r) - K_i(r_j)) \mathbf{1}[x_{[i]}^2 u_{[i]}^2 \leq \kappa_n]\right|\right]$$

$$\leq C_4 (1 - 2\tau) \kappa_n / \left(m_n b_n^2\right),$$

which does not depend on j, and hence it gives the uniform bound $T_{2n}^\kappa = O(\kappa_n / (m_n b_n^2))$ 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*

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

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 $\pi(\cdot) \equiv h_\tau^{1/2} g^{(1)}(\cdot) v^\mathsf{T} D(\cdot)^{-1}$. Similarly to the one in Lemma 1, we decompose

$$\widetilde{G}_n(s) = \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} \pi(i/n) x_{[i]} u_{[i]} \tag{A.16}$$

$$- \frac{1}{n} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} \pi(i/n) x_{[i]} x_{[i]}^\mathsf{T} \sqrt{n}(\widehat{\beta} - \beta_0)$$

$$- \frac{1}{n} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} \pi(i/n) x_{[i]} x_{[i]}^\mathsf{T} \mathbf{1}[\widehat{F}_n(q_{[i]}) \le r_0] \sqrt{n}(\widehat{\delta} - \delta_0)$$

$$- \frac{1}{n} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} \pi(i/n) x_{[i]} x_{[i]}^\mathsf{T} \left\{ \mathbf{1}\left[F(q_{[i]}) \le r_0\right] - \mathbf{1}[\widehat{F}_n(q_{[i]}) \le r_0]\right\} \sqrt{n}(\widehat{\delta} - \delta_0)$$

$$- \frac{1}{n} \sum_{i=\lfloor \tau n \rfloor+1}^{\lfloor g^{-1}(s)n \rfloor} \pi(i/n) x_{[i]} x_{[i]}^\mathsf{T} \left\{ \mathbf{1}[\widehat{F}_n(q_{[i]}) \le \widehat{r}] - \mathbf{1}\left[F(q_{[i]}) \le r_0\right]\right\} \sqrt{n}\widehat{\delta}$$

$$\equiv A_{1n}(s) - A_{2n}(s) - A_{3n}(s) - A_{4n}(s) - A_{5n}(s).$$

First, we derive the limit of $A_{1n}(s)$ by applying Corollary 29.14 in Davidson (1994).[7] To this end, we let $U_{n,i} = h_\tau^{1/2} n^{-1/2} g^{(1)}(i/n) v^\mathsf{T} D(i/n)^{-1} x_{[i]} u_{[i]}$ and $\overrightarrow{q} = \{q_i\}_{i=1}^n$, and check Condition 29.6(a)–(f′) in the corollary. Condition (a) is satisfied since $\mathbb{E}\left[U_{n,i}\right] = \mathbb{E}[\mathbb{E}\left[U_{n,i} | \overrightarrow{q}\right]] = 0$ given our Conditions 1.1 and 1.2. Condition (b) is implied by our Conditions 1.6 and 1.8 by setting $c_{n,i} = 1$ in the corollary as seen by

$$\sup_{i/n \in [\tau, 1-\tau]} \left\| U_{n,i} \right\|_4 \le \frac{h_\tau^{1/2}}{\sqrt{n}} \sup_{r \in [\tau, 1-\tau]} \left\| v^\mathsf{T} D(r)^{-1} \right\|_4 \sup_{r \in [\tau, 1-\tau]} \left| g^{(1)}(r) \right|$$

$$\times \left(\sup_{q \in \mathbb{R}} \mathbb{E}\left[||x_i u_i||^4 | q_i = q \right] \right)^{1/4} < \infty,$$

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 near-epoch dependence (NED) condition is satisfied. Condition (d) holds by setting $c_{n,i} = 1$ and $K_n(t) = \left\lfloor g^{-1}(t)n \right\rfloor$, and from the fact that $g^{-1}(\cdot)$ 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 $\frac{i}{n}$ yield that

$$\mathbb{E}\left[x_{[i]} x_{[i]}^\mathsf{T} u_{[i]}^2\right] = \mathbb{E}\left[\mathbb{E}\left[x_j x_j^\mathsf{T} u_j^2 | q_j = q_{[i]}\right]\right]$$

$$= \mathbb{E}\left[V(F(q_{[i]}))\right]$$

$$= V(i/n) + \mathbb{E}\left[\frac{\partial V(t_i)}{\partial t} \left(F\left(q_{[i]}\right) - i/n\right)\right]$$

$$= V(i/n) + O\left(n^{-1/2}\right), \tag{A.17}$$

---

[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 $\{g^{(1)}(i/n) v^\mathsf{T} D(i/n)^{-1} x_{[i]} u_{[i]}\}_{i=1}^n$ is $\{g^{(1)}(R_i/n) v^\mathsf{T} D(R_i/n)^{-1} x_i u_i\}_{i=1}^n$, which is no longer i.i.d. given the rank statistics R_i.

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

$$\sup_{i/n \in [\tau, 1-\tau]} \left\| \mathbb{E}\left[\frac{\partial V(t_i)}{\partial t} \left(F\left(q_{[i]}\right) - \widehat{F}_n(q_{[i]}) \right) \right] \right\|$$

$$\leq \sup_{t \in [\tau, 1-\tau]} \left\| \frac{\partial V(t)}{\partial t} \right\| \mathbb{E}\left[\sup_{t \in [\tau, 1-\tau]} \left| F(t) - \widehat{F}_n(t) \right| \right]$$

$$= O\left(n^{-1/2}\right),$$

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

$$\mathbb{E}\left[\left(\sum_{i=\lfloor \tau n \rfloor + 1}^{K_n(s)} U_{n,i} \right)^2 \right]$$

$$= \mathbb{E}\left[\sum_{i=\lfloor \tau n \rfloor + 1}^{K_n(s)} U_{n,i}^2 \right]$$

$$= \frac{h_\tau}{n} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor g^{-1}(s)n \rfloor} \left(g^{(1)}(i/n) \right)^2 v^{\mathsf{T}} D(i/n)^{-1} \mathbb{E}\left[x_{[i]} x_{[i]}^{\mathsf{T}} u_{[i]}^2 \right] D(i/n)^{-1} v$$

$$= \frac{h_\tau}{n} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor g^{-1}(s)n \rfloor} \left(g^{(1)}(i/n) \right)^2 v^{\mathsf{T}} D(i/n)^{-1} V(i/n) D(i/n)^{-1} v + O(n^{-1/2})$$

$$\to h_\tau \int_\tau^{g^{-1}(s)} \left(g^{(1)}(t) \right)^2 v^{\mathsf{T}} D(t)^{-1} V(t) D(t)^{-1} v \, dt$$

$$= \int_{g^{-1}(0)}^{g^{-1}(s)} g^{(1)}(t) \, dt = s,$$

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 $n \to \infty$; and the last expression is by the definition of $g^{(1)}(\cdot)$ and $g^{-1}(0) = \tau$. Therefore, Davidson (1994, Cor. 29.14) implies that $A_{1n}(s) = \sum_{i=\lfloor \tau n \rfloor + 1}^{K_n(s)} U_{n,i}^2 \Rightarrow W_1(s)$, 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

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

and

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

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

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

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 $||\sqrt{n}\widehat{\delta}|| \leq C_\delta$ 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,

$$
\mathbb{P}\left(\sup_{s\in[0,1]} |A_{5n}(s)| > \eta \right)
$$

$$
\leq \mathbb{P}\left(\left\{ \sup_{s\in[0,1]} |A_{5n}(s)| > \eta \right\} \cap E_{\gamma n} \cap E_{\delta n} \right) + 2\varepsilon
$$

$$
\leq \eta^{-1} C n^{1/2-\epsilon} \mathbb{E}\left[\left\| \frac{1}{n} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor (1-\tau)n \rfloor} \pi(i/n) x_{[i]} x_{[i]}^\mathsf{T} \left\{ \mathbf{1}\left[q_{[i]} \leq \widehat{\gamma} \right] - \mathbf{1}\left[q_{[i]} \leq \gamma_0 \right] \right\} \mathbf{1}\left[E_{\gamma n} \right] \right\| \right] + 2\varepsilon
$$

$$
\leq \eta^{-1} C n^{1/2-\epsilon} \sup_{r\in[\tau,1-\tau]} \|\pi(r)\| \, \mathbb{E}\left[\|x_i x_i^\mathsf{T}| \, |\mathbf{1}[q_i \leq \widehat{\gamma}] - \mathbf{1}[q_i \leq \gamma_0]| \mathbf{1}\left[E_{\gamma n} \right] \right] + 2\varepsilon
$$

$$
\leq \eta^{-1} C' n^{-1/2+\epsilon} + 2\varepsilon
$$

$$
\leq 3\varepsilon,
$$

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 $g^{-1}(s) > \widehat{g}^{-1}(s)$. Let $\pi(\cdot) \equiv h_\tau^{1/2} g^{(1)}(\cdot) v^\mathsf{T} D(\cdot)^{-1}$ and $\widehat{\pi}(\cdot) \equiv \widehat{h}_\tau^{1/2} \widehat{g}^{(1)}(\cdot) v^\mathsf{T} \widehat{D}(\cdot)^{-1}$. Note that, for any $s \in [0,1]$,

$$
\widehat{\mathcal{G}}_n(s) - \widetilde{\mathcal{G}}_n(s) = \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor \widehat{g}^{-1}(s)n \rfloor} \widehat{\pi}(i/n) x_{[i]} \widehat{u}_{[i]} - \frac{1}{\sqrt{n}} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor g^{-1}(s)n \rfloor} \pi(i/n) x_{[i]} \widehat{u}_{[i]}
$$

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

$$
\equiv B_{1n}(s) + B_{2n}(s).
$$

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

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

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

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

$$
\equiv B_{11n}(s) + B_{12n}(s). \tag{A.18}
$$

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. 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

$$\sup_{s\in[0,1]} |B_{12n}(s)|$$

$$\leq \sup_{r\in[\tau,1-\tau]} |\widehat{\pi}(r) - \pi(r)| \sup_{r\in[\tau,1-\tau]} \frac{1}{\sqrt{n}} \left| \sum_{i=\lfloor \tau n\rfloor +1}^{\lfloor rn\rfloor} x_{[i]}(\widehat{u}_{[i]} - u_{[i]}) \right|$$

$$\leq o_p(1) \left\{ \sup_{r\in[\tau,1-\tau]} \frac{1}{n} \sum_{i=\lfloor \tau n\rfloor +1}^{\lfloor rn\rfloor} x_{[i]}^2 |\sqrt{n}(\widehat{\beta} - \beta_0)| \right.$$

$$+ \sup_{r\in[\tau,1-\tau]} \frac{1}{n} \sum_{i=\lfloor \tau n\rfloor +1}^{\lfloor rn\rfloor} x_{[i]}^2 \mathbf{1}\left[q_{(i)} \leq \gamma_0\right] |\sqrt{n}(\widehat{\delta} - \delta_0)|$$

$$\left. + \sup_{r\in[\tau,1-\tau]} \frac{1}{n} \sum_{i=\lfloor \tau n\rfloor +1}^{\lfloor rn\rfloor} x_{[i]}^2 \left|\mathbf{1}\left[q_{(i)} \leq \gamma_0\right] - \mathbf{1}\left[q_{(i)} \leq \widehat{\gamma}\right]\right| |\sqrt{n}\delta| \right\}$$

$$= o_p(1),$$

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

$$B_{2n}(s) = \frac{1}{\sqrt{n}} \sum_{i=\lfloor g^{-1}(s)n\rfloor +1}^{\lfloor \widehat{g}^{-1}(s)n\rfloor} \pi(i/n)x_{[i]}u_{[i]} + \frac{1}{\sqrt{n}} \sum_{i=\lfloor g^{-1}(s)n\rfloor +1}^{\lfloor \widehat{g}^{-1}(s)n\rfloor} \pi(i/n)x_{[i]} \left(\widehat{u}_{[i]} - u_{[i]}\right)$$

$$\equiv B_{21n}(s) + B_{22n}(s). \tag{A.19}$$

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 $\widehat{g}^{-1}(s) = \varrho(s)$,

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

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

$$\left. - h_\tau^{1/2} \int_{g^{-1}(0)}^{\varrho(s)} g^{(1)}(t)D(t)^{-1}V(t)^{1/2} \, dW_k(t) \right|$$

$$=_d \sup_{s\in[0,1]} \sup_{|\varrho(s)-g^{-1}(s)|<\eta} |W_1(s) - W_1(g(\varrho(s)))|.$$

Then, we can choose η small enough to obtain that, for any $\varepsilon > 0$,

$$
\mathbb{P}\left(\sup_{s\in[0,1]}|B_{21n}(s)| > \varepsilon\right) \le \mathbb{P}\left(\left\{\sup_{s\in[0,1]}|B_{21n}(s)| > \varepsilon\right\} \cap E_{gn}\right) + \mathbb{P}(E_{gn}^c)
$$

$$
\to \mathbb{P}\left(\sup_{s\in[0,1]}\sup_{|\varrho(s)-g^{-1}(s)|<\eta}|W_1(s) - W_1(g(\varrho(s)))| > \varepsilon\right) + \varepsilon
$$

$$
\le \varepsilon^{-1}\mathbb{E}\left[\sup_{s\in[0,1]}\sup_{|\varrho(s)-g^{-1}(s)|<\eta}|W_1(s) - W_1(g(\varrho(s)))|\right] + \varepsilon
$$

$$
\le \varepsilon^{-1}\eta^{1/2}C + \varepsilon
$$

$$
\le 2\varepsilon,
$$

where the second inequality is by Markov's inequality; the third inequality follows from the continuity of $g(\cdot)$ and from the fact that $\mathbb{E}[\sup_{s\in[0,t]}|W_1(s)|] \le \sqrt{2t/\pi}$; 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

$$
\sup_{s\in[0,1]}|B_{22n}(s)|
$$

$$
\le \sup_{r\in[\tau,1-\tau]}|\pi(r)| \sup_{s\in[0,1]}\frac{1}{\sqrt{n}}\sum_{i=\lfloor \widehat{g}^{-1}(s)n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor}\left|x_{[i]}(\widehat{u}_{[i]} - u_{[i]})\right|
$$

$$
\le C \sup_{s\in[0,1]}\frac{1}{\sqrt{n}}\sum_{i=\lfloor (g^{-1}(s)-\eta)n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor}
$$
$$
\times x_{[i]}^2\left\{|\widehat{\beta} - \beta_0| + |\widehat{\delta} - \delta_0|\mathbf{1}\left[q_{(i)} \le \gamma_0\right] + \widehat{\delta}\left|\mathbf{1}\left[q_{(i)} \le \widehat{\gamma}\right] - \mathbf{1}\left[q_{(i)} \le \gamma_0\right]\right|\right\}
$$

$$
\le C' \sup_{s\in[0,1]}\frac{1}{n}\sum_{i=\lfloor (g^{-1}(s)-\eta)n\rfloor+1}^{\lfloor g^{-1}(s)n\rfloor}x_{[i]}^2
$$

$$
\to_p C' \sup_{s\in[0,1]}\int_{g^{-1}(s)-\eta}^{g^{-1}(s)}D(t)\,dt
$$

for some constant $0 < C, C' < \infty$, where the second inequality is from Condition 1.6; the third inequality is from the fact that $\mathbf{1}\left[q_{[i]} \le \gamma\right] \le 1$ for any γ, the result in (A.10), and by conditioning on the events $E_{\theta n}$ and E_{gn}; and 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 $g(s) \le \widehat{g}^{-1}(s)$ 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 $\overrightarrow{q} = \{q_i\}_{i=1}^n$ 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 $\pi(\cdot) = h_\tau^{1/2} g^{(1)}(\cdot) v^\mathsf{T} D(\cdot)^{-1}$ 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 by $C^1[\tau, 1-\tau]$, 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 $\varepsilon > 0$,

$$\sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor (1-\tau)n \rfloor} \mathbb{E} \left[\sup_{\pi \in \mathcal{F}} |L_{ni}(\pi)| \, \mathbf{1} \left[\sup_{\pi \in \mathcal{F}} |L_{ni}(\pi)| > \varepsilon \right] \middle| \vec{\mathsf{q}} \right]$$

$$\leq \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor (1-\tau)n \rfloor} \mathbb{E} \left[\sup_{\pi \in \mathcal{F}} |L_{ni}(\pi)|^2 \middle| \vec{\mathsf{q}} \right]^{1/2} \mathbb{P} \left(\sup_{\pi \in \mathcal{F}} |L_{ni}(\pi)| > \varepsilon \middle| \vec{\mathsf{q}} \right)^{1/2}$$

$$\leq \varepsilon^{-4} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor (1-\tau)n \rfloor} \mathbb{E} \left[\sup_{\pi \in \mathcal{F}} |L_{ni}(\pi)|^2 \middle| \vec{\mathsf{q}} \right]^{1/2} \mathbb{E} \left[\sup_{\pi \in \mathcal{F}} |L_{ni}(\pi)|^4 \middle| \vec{\mathsf{q}} \right]^{1/2}$$

$$\leq \bar{C}^3 n^{-3/2} \varepsilon^{-4} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor (1-\tau)n \rfloor} \mathbb{E} \left[\left\| x_{[i]} u_{[i]} \right\|^2 \middle| \vec{\mathsf{q}} \right]^{1/2} \mathbb{E} \left[\left\| x_{[i]} u_{[i]} \right\|^4 \middle| \vec{\mathsf{q}} \right]^{1/2}$$

$$\to 0 \text{ a.s.,}$$

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

$$\sup_{\rho(\pi, \pi_1) \leq \varepsilon_n} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor (1-\tau)n \rfloor} \mathbb{E} \left[(L_{ni}(\pi) - L_{ni}(\pi_1))^2 \middle| \vec{\mathsf{q}} \right] \leq \bar{C}^2 \varepsilon_n n^{-1} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor (1-\tau)n \rfloor} \mathbb{E} \left[\left| x_{[i]} u_{[i]} \right|^2 \middle| \vec{\mathsf{q}} \right]$$

$$\to 0 \text{ a.s.}$$

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 $\vec{\mathsf{q}}$, 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,

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

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

$$\mathbb{P}\left(\sup_{s \in [0, 1]} |B_{11n}(s)| > \varepsilon \right)$$

$$\leq \mathbb{E}\left[\mathbb{P}\left(\left\{ \sup_{s \in [0, 1]} |B_{11n}(s)| > \varepsilon \right\} \cap E_{\pi n} \middle| \overrightarrow{\mathsf{q}} \right) \right] + \mathbb{P}\left(E_{\pi n}^c \right)$$

$$\leq \mathbb{E}\left[\mathbb{P}\left(\max_{1 \leq \ell \leq n} \sup_{\rho(\widehat{\pi}, \pi) \leq \eta} |\mathbb{L}_\ell(\pi) - \mathbb{L}_\ell(\widehat{\pi})| > \varepsilon \middle| \overrightarrow{\mathsf{q}} \right) \right] + \varepsilon$$

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

$$\leq C\varepsilon.$$

The second inequality is from Lemma A.5 that implies $\mathbb{P}(E_{\pi n}^c) \leq \varepsilon$ if n is large enough, and from the law of iterated expectations. The third inequality is from the Ottaviani's inequality (e.g., A.1.1 in van der Vaart and Wellner (1996)) and the fact that $\{x_{[i]} u_{[i]}\}_{i=1}^n$ are independent conditional on $\overrightarrow{\mathsf{q}}$. The last inequality is from (A.20) and the steps on page 227 in van der Vaart and Wellner (1996). In particular, for some $1 \leq n_0 \leq n$,

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

$$\leq \max_{\ell \leq n_0} \mathbb{P}\left(n^{-1/2} \sum_{i=\lfloor \tau n \rfloor + 1}^{n_0} C \left\| x_{[i]} u_{[i]} \right\| > \varepsilon \middle| \overrightarrow{\mathsf{q}} \right)$$

$$+ \max_{n_0 \leq \ell} \mathbb{P}\left(\sup_{\rho(\widehat{\pi}, \pi) \leq \eta} |\mathbb{L}_\ell(\pi) - \mathbb{L}_\ell(\widehat{\pi})| > \varepsilon \middle| \overrightarrow{\mathsf{q}} \right)$$

$$\leq C\varepsilon \quad \text{a.s.},$$

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 the 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

$$
\widehat{\mathcal{G}}_{1n}^*(s) \Rightarrow \frac{1}{\sqrt{g(r_0)}} \left\{ \mathcal{G}(s) - \frac{s}{g(r_0)} \mathcal{G}(g(r_0)) \right\}
$$

$$
=_d \frac{1}{\sqrt{g(r_0)}} \left\{ W_1(s) - \frac{s}{g(r_0)} W_1(g(r_0)) \right\} \equiv \mathcal{G}_1(s), \qquad \text{(A.21)}
$$

$$
\widehat{\mathcal{G}}_{2n}^*(s) \Rightarrow \frac{1}{\sqrt{1-g(r_0)}} \left\{ (\mathcal{G}(1) - \mathcal{G}(s)) - \frac{1-s}{1-g(r_0)} (\mathcal{G}(1) - \mathcal{G}(g(r_0))) \right\}
$$

$$
=_d \frac{1}{\sqrt{1-g(r_0)}} \left\{ (W_1(1) - W_1(s)) - \frac{1-s}{1-g(r_0)} (W_1(1) - W_1(g(r_0))) \right\} \equiv \mathcal{G}_2(s).
$$
$$\text{(A.22)}$$

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

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

$$
=_d \int_0^1 \{W_1(t) - t W_1(1)\}^2 dt
$$

and

$$
\frac{1}{1-g(r_0)} \int_{g(r_0)}^1 \mathcal{G}_2(s)^2 ds
$$

$$
=_d \frac{1}{(1-g(r_0))} \int_0^1 \{(W_1(1) - W_1(1 - (1-g(r_0))t)) - t(W_1(1) - W_1(g(r_0)))\}^2 dt
$$

$$
=_d \frac{1}{(1-g(r_0))} \int_0^1 \{W_1((1-g(r_0))t) - t W_1(1-g(r_0))\}^2 dt
$$

$$
=_d \int_0^1 \{W_1(t) - t W_1(1)\}^2 dt.
$$

Therefore, the limiting null distribution of CT_n is obtained as

$$
CT_n = \frac{1}{g(r_0)} \times \frac{1}{n} \sum_{i=1}^{\lfloor g(r_0)n \rfloor} \widehat{\mathcal{G}}_{1n}^*(i/n)^2 + \frac{1}{1-g(r_0)} \times \frac{1}{n} \sum_{i=\lfloor g(r_0)n \rfloor + 1}^{n} \widehat{\mathcal{G}}_{2n}^*(i/n)^2 + o_p(1)
$$

$$
\to_d \frac{1}{g(r_0)} \int_0^{g(r_0)} \mathcal{G}_1(s)^2 ds + \frac{1}{1-g(r_0)} \int_{g(r_0)}^1 \mathcal{G}_2(s)^2 ds
$$

$$
=_d \int_0^1 \mathcal{B}_2(t)^\mathsf{T} \mathcal{B}_2(t) dt
$$

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_{0i} (or r_{0i}) is not equal to γ_0 almost surely. 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 $\widehat{\theta} = (\widehat{\beta}^{\mathsf{T}}, \widehat{\delta}^{\mathsf{T}})^{\mathsf{T}}$, 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 $X_i(\gamma) = (x_i^{\mathsf{T}}, x_i^{\mathsf{T}} \mathbf{1}[q_i < \gamma])^{\mathsf{T}}$ and $X_i(\gamma_i) = (x_i^{\mathsf{T}}, x_i^{\mathsf{T}} \mathbf{1}[q_i < \gamma_i])^{\mathsf{T}}$. Given $\widehat{\gamma} = \gamma$, for any γ,

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

Similarly to the one in Lemma 1 in Hansen (1996), we have $\widehat{\Theta}_{n1} \to_p \Theta_1 = \mathbb{E}\left[X_i(\gamma) X_i(\gamma)^{\mathsf{T}} \right]$, 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, although it can be $o_p(1)$ under some special circumstances.

Let $r_{[i]}$ be the induced-order statistics of $\widehat{F}_n(\gamma_{0i})$ associated with $q_{[i]}$, and $\widehat{\pi}(\cdot) = \widehat{h}_{\tau}^{1/2} \widehat{g}^{(1)}(\cdot) v^{\mathsf{T}} \widehat{D}(\cdot)^{-1}$. We decompose

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

and denote their rescaled and demeaned terms as in (27) 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 $n \to \infty$, although $\bar{g}$ 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 in Lemma A.7 implies that, for any $s \in [\tau, 1 - \tau]$,

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

$$\frac{1}{n} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor \widehat{g}^{-1}(s)n \rfloor} \widehat{g}^{(1)}(i/n) \widehat{D}(i/n)^{-1} x_{[i]} x_{[i]}^\mathsf{T} \mathbf{1}[i/n \leq \widehat{r}] \to_p \min\{s, \bar{g}\} I_k, \tag{A.24}$$

as $n \to \infty$, where I_k denotes the $k \times k$ identity matrix. If follows that

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

since $\widehat{\theta} - \theta_0 = O_p(n^{-\epsilon})$ from (A.23). However, since $\widehat{C}_{2n}(s)$ is asymptotically piecewise linear in s, the rescaling 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 $\widehat{C}_{3n}^*(s)$, we denote $F_\gamma(\cdot)$ as the CDF of γ_{0i} and $\bar{F}_\gamma(\cdot) = 1 - F_\gamma(\cdot)$ as its survival function. We note that

$$\frac{1}{n} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor \widehat{g}^{-1}(s)n \rfloor} \widehat{g}^{(1)}(i/n) v^\mathsf{T} \widehat{D}(i/n)^{-1} x_{[i]} x_{[i]}^\mathsf{T} c_0 \mathbf{1}[i/n \leq r_{[i]}]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \widehat{g}^{(1)}(\widehat{F}_n(q_i)) v^\mathsf{T} \widehat{D}(\widehat{F}_n(q_i))^{-1} x_i x_i^\mathsf{T} c_0 \mathbf{1}[q_i \leq \gamma_{0i}] \mathbf{1}\left[\tau \leq \widehat{F}_n(q_i) \leq \widehat{g}^{-1}(s)\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \widehat{g}^{(1)}(\widehat{F}_n(q_i)) v^\mathsf{T} \widehat{D}(\widehat{F}_n(q_i))^{-1} x_i x_i^\mathsf{T} c_0 \bar{F}_\gamma(q_i) \mathbf{1}\left[\tau \leq \widehat{F}_n(q_i) \leq \widehat{g}^{-1}(s)\right]$$

$$+ \frac{1}{n} \sum_{i=1}^{n} \widehat{g}^{(1)}(\widehat{F}_n(q_i)) v^\mathsf{T} \widehat{D}(\widehat{F}_n(q_i))^{-1} x_i x_i^\mathsf{T} c_0 \left\{\mathbf{1}[q_i \leq \gamma_{0i}] - \bar{F}_\gamma(q_i)\right\} \mathbf{1}\left[\tau \leq \widehat{F}_n(q_i) \leq \widehat{g}^{-1}(s)\right]$$

$$\equiv H_{1n}(s) + H_{2n}(s).$$

For $H_{1n}(s)$, since $\widehat{D}(\cdot)$ is still uniformly consistent, a similar argument as in Lemma A.7 implies that

$$H_{1n}(s) = \frac{1}{n} \sum_{i=\lfloor \tau n \rfloor + 1}^{\lfloor \widehat{g}^{-1}(s)n \rfloor} \widehat{g}^{(1)}(i/n) v^\mathsf{T} \widehat{D}(i/n)^{-1} x_{[i]} x_{[i]}^\mathsf{T} c_0 \bar{F}_\gamma\left(\widehat{F}_n^{-1}(i/n)\right)$$

$$\to_p v^\mathsf{T} c_0 \int_0^s \bar{F}_\gamma\left(F^{-1}\left(\tilde{g}^{-1}(t)\right)\right) dt \equiv H(s),$$

where $\tilde{g}(\cdot)$ denotes the probability limit of $\widehat{g}(\cdot)$, which is still monotonically increasing by construction. Recall that $\widehat{F}_n^{-1}(\cdot)$ is the empirical quantile function of q_i, which uniformly converges to the true quantile function $F^{-1}(\cdot)$ over $[\tau, 1 - \tau]$. For $H_{2n}(s)$, since γ_{0i} is independent of $(q_i, x_i^\mathsf{T}, u_i)^\mathsf{T}$, we have that $\mathbb{E}[H_{2n}(s)] = 0$. Similarly, $\mathbb{E}\left[H_{2n}(s)^2\right] = O(1/n)$ since γ_{0i} is i.i.d. and $\|\widehat{g}^{(1)}(\widehat{F}_n(q_i)) \widehat{D}(\widehat{F}_n(q_i))^{-1}\|$ and $\mathbb{E}[\|x_i\|^4]$ are uniformly bounded. We then have $H_{2n}(s) = o_p(1)$. Combining the results of $H_{1n}(s)$ and $H_{2n}(s)$ and using (A.24)

yield that

$$n^{-1/2+\epsilon}\widehat{C}_{3n}(s) \to_p \left(\text{plim}_{n\to\infty}\widehat{h}_\tau^{1/2}\right)\left(v^\mathsf{T}c_0 \min\{s,\bar{g}\} - H(s)\right).$$

Given that $\bar{F}_\gamma(\cdot)$ is nonincreasing and nonconstant and that $F(\cdot)$ and $\tilde{g}(\cdot)$ are strictly increasing, the integrand $\tilde{\eta}(\cdot) \equiv \bar{F}_\gamma\left(F^{-1}\left(\tilde{g}^{-1}(\cdot)\right)\right)$ in $H(\cdot)$ is nonconstant on $[0,1]$. In particular, $\tilde{\eta}(\cdot)$ changes values on $[0,\bar{g}]$ and/or $(\bar{g},1]$. Consider that $\tilde{\eta}(\cdot)$ is nonconstant on $[0,\bar{g}]$. Then, $H(s) = \int_0^s \tilde{\eta}(t)\,dt$ is nonlinear by construction and cannot be equal to the linear function $s\bar{g}^{-1}H(\bar{g})$ for all $s \in [0,\bar{g}]$. Therefore, $\int_0^{\bar{g}}\left(H(s) - s\bar{g}^{-1}H(\bar{g})\right)^2 ds$ is strictly positive. So, there exists some constant $\underline{c} > 0$ such that

$$\mathbb{P}\left(n^{-1+2\epsilon}\left|\int_0^{\widehat{g}(\widehat{r})}\widehat{C}_{3n}^*(s)^2\,ds\right| > \underline{c}\right)$$

$$= \mathbb{P}\left(n^{-1+2\epsilon}\left|\int_0^{\widehat{g}(\widehat{r})}\left(\frac{1}{\sqrt{\widehat{g}(\widehat{r})}}\left\{\widehat{C}_{3n}(s) - \frac{s}{\widehat{g}(\widehat{r})}\widehat{C}_{3n}(\widehat{g}(\widehat{r}))\right\}\right)^2 ds\right| > \underline{c}\right)$$

$$\to \mathbb{P}\left(\frac{\text{plim}_{n\to\infty}\widehat{h}_\tau}{\bar{g}}\left|\int_0^{\bar{g}}\left(H(s) - \frac{s}{g}H(\bar{g})\right)^2 ds\right| > \underline{c}\right)$$

$$= 1,$$

as $n \to \infty$. It follows that $\int_0^{\widehat{g}(\widehat{r})}\widehat{C}_{3n}^*(s)^2\,ds > \underline{c}n^{1-2\epsilon} \to \infty$ with probability approaching to one. Therefore, $\int_0^{\widehat{g}(\widehat{r})}\widehat{C}_{3n}^*(s)^2\,ds$ becomes the leading term in $\int_0^{\widehat{g}(\widehat{r})}\widehat{\mathcal{G}}_{1n}^*(s)^2\,ds$, which diverges with probability approaching to one. The same argument applies to the case when $\tilde{\eta}(\cdot)$ is nonconstant on $(\bar{g},1]$, which also yields that $\int_{\widehat{g}(\widehat{r})}^1\widehat{C}_{3n}^*(s)^2\,ds \to \infty$ and hence becomes the leading term in $\int_{\widehat{g}(\widehat{r})}^1\widehat{\mathcal{G}}_{2n}^*(s)^2\,ds$. Therefore, at least one of $\int_0^{\widehat{g}(\widehat{r})}\widehat{\mathcal{G}}_{1n}^*(s)^2\,ds$ and $\int_{\widehat{g}(\widehat{r})}^1\widehat{\mathcal{G}}_{2n}^*(s)^2\,ds$ diverges with probability approaching to one, yielding the consistency of the test. ∎

Proof of Theorem 2. Under the local alternative, the error term is now defined as $\tilde{u}_i = u_i + n^{-1/2}x_i^\mathsf{T}\alpha(q_i)$. However, Lemma A.5 in Hansen (2000) still implies that $n^{-1}\sum_{i=1}^n x_i\tilde{u}_i \to_p \mathbb{E}[x_iu_i]$, 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, equation (43) in Hansen (2000) now involves the following additional term:

$$M_n^\alpha = \frac{1}{n}\sum_{i=1}^n x_ix_i^\mathsf{T}\alpha(q_i).$$

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

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

Then, the rest of the argument follows from page 597 in Hansen (2000). Note that the additional term $n^{-1/2}x_i^\mathsf{T}\alpha(q_i)$ 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.

Now, given these consistency results, we have

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

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

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

$$\tag{A.25}$$

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

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

and

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

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(4), 821–856.

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

Bae, J., D. Jun, & S. Levental (2010) The uniform CLT for martingale difference arrays under the uniformly Integrable entropy. *Bulletin of the Korean Mathematical Society* 47(1), 39–51.

Bai, J., R.L. Lumsdaine, & J.H. Stock (1998) Testing for and dating common breaks in multivariate time series. *Review of Economic Studies* 65(3), 395–432.

Bai, J. & P. Perron (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66(1), 47–78.

Bhattacharya, P.K. (1974) Convergence of sample paths of normalized sums of induced order statistics. *Annals of Statistics* 2(5), 1034–1039.

Bhattacharya, P.K. (1984) 18 induced order statistics: Theory and applications. *Handbook of Statistics* 4, 383–403.

Caner, M. & B.E. Hansen (2001) Threshold autoregressions with a unit root. *Econometrica* 69(6), 1555–1596.

Caner, M. & B.E. Hansen (2004) Instrumental variable estimation of a threshold model. *Econometric Theory* 20(5), 813–843.

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

Casini, A. & P. Perron (2021a) Continuous record Laplace-based inference about the break date in structural change models. *Journal of Econometrics* 224(1), 3–21.

Casini, A. & P. Perron (2021b) Continuous Record Asymptotics for Change-Point Models. Working paper. Preprint, arXiv:1803.10881

Casini, A. & P. Perron (2022) Generalized Laplace inference in multiple change-points models. *Econometric Theory* 38(1), 35–65.

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

Elliott, G. & U.K. Müller (2007) Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics* 141(2), 1196–1218.

Elliott, G. & U.K. Müller (2014) Pre and post break parameter inference. *Journal of Econometrics* 180(2), 141–157.

Elliott, G., U.K. Müller, & M.W. Watson (2015) Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica* 83(2), 771–811.

Gonzalo, J. & J. Pitarakis (2002) Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics* 110(2), 319–352.

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

Hansen, B.E. (2000) Sample splitting and threshold estimation. *Econometrica* 68(3), 575–603.

Hidalgo, J., J. Lee, & M.H. Seo (2019) Robust inference for threshold regression models. *Journal of Econometrics* 210(2), 291–309.

Jiang, L., X. Wang, & J. Yu (2018) New distribution theory for the estimation of structural break point in mean. *Journal of Econometrics* 205(1), 156–176.

Jiang, L., X. Wang, & S. Yu (2020) In-fill asymptotic theory for structural break point in autoregression: A unified theory. *Econometric Reviews* 40(4), 359–386.

Kosorok, M.R. (2008) *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

Kourtellos, A., T. Stengos, & C.M. Tan (2016) Structural threshold regression. *Econometric Theory* 32(4), 827–860.

Lee, S., Y. Liao, M.H. Seo, & Y. Shin (2021) Factor-driven two-regime regression. *Annals of Statistics* 49(3), 1656–1678.

Lee, S., M.H. Seo, & Y. Shin (2011) Testing for threshold effects in regression models. *Journal of the American Statistical Association* 106(493), 220–231.

Li, D. & S. Ling (2012) On the least squares estimation of multiple-regime threshold autoregressive models. *Journal of Econometrics* 167(1), 240–253.

Li, H. & U.K. Müller (2009) Valid inference in partially unstable general method of moment models. *Review of Economic Studies* 76(1), 343–365.

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

Miao, K., L. Su, & W. Wang (2020) Panel threshold regressions with latent group structures. *Journal of Econometrics* 214(2), 451–481.

Nyblom, J. (1989) Testing for the Constancy of parameters over time. *Journal of the American Statistical Association* 84(405), 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. *Annals of Probability* 4(3), 474–479.

Seo, M.H. & O. Linton (2007) A smooth least squares estimator for threshold regression models. *Journal of Econometrics* 141(2), 704–735.

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

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.

Yang, S.S. (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(1), 274–294.

Yu, P. & X. Fan (2021) Threshold regression with a threshold boundary. *Journal of Business and Economic Statistics* 39(4), 953–971.

Yu, P. & P. Phillips (2018) Threshold regression with Endogeneity. *Journal of Econometrics* 203(1), 50–68.