Nonparametric Sample Splitting*

YOONSEOK LEE† — Syracuse University

YULONG WANG‡ — Syracuse University

First version: May 2019
This version: November 2019

Abstract

This paper develops a threshold regression model where an unknown relationship between two variables nonparametrically determines the threshold. We allow the observations to be cross-sectionally dependent so that the model can be applied to determine an unknown spatial border for sample splitting over a random field. We derive the uniform rate of convergence and the nonstandard limiting distribution of the nonparametric threshold estimator. We also obtain the root-n consistency and the asymptotic normality of the regression coefficient estimator. Our model has broad empirical relevance as illustrated by estimating the tipping point in social segregation problems as a function of demographic characteristics; and determining metropolitan area boundaries using nighttime light intensity collected from satellite imagery. We find that the new empirical results are substantially different from the existing studies.

Keywords: sample splitting, threshold, nonparametric, random field, tipping point, metropolitan area boundary.
JEL Classifications: C14, C21, C24, R1

---

*We thank Bo Honoré, Sokbae Lee, Yuan Liao, Myung Seo, Ping Yu, and participants at numerous seminar/conference presentations for very helpful comments. Financial supports from the Appleby-Mosher grant and the CUSE grant are highly appreciated.

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

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

1 Introduction

Sample splitting and threshold regression models have spawned a vast literature in econometrics and statistics. Existing studies parametrically specify the splitting criteria as whether a single random variable or a linear combination of variables crosses some unknown threshold. See, for example, Hansen (2000), Caner and Hansen (2004), Seo and Linton (2007), Lee, Seo, and Shin (2011), Li and Ling (2012), Yu (2012), Lee, Liao, Seo, and Shin (2018), Hidalgo, Lee, and Seo (2019), and Yu and Fan (2019). In this paper, we study a novel extension to consider a *nonparametric* sample splitting model. Such an extension leads to new theoretical results and substantially generalizes the applicability of threshold models.

Specifically, we consider a model given by

$$y_i = x_i^\top \beta_0 + x_i^\top \delta_0 \mathbf{1}\left[q_i \leq \gamma_0\left(s_i\right)\right] + u_i \tag{1}$$

for $i = 1, \ldots, n$, in which the marginal effect of x_i to y_i can be different as β_0 or $(\beta_0+\delta_0)$ depending on whether $q_i \leq \gamma_0(s_i)$ or not. The threshold function $\gamma_0(\cdot)$ is unknown, and the main parameters of interest are β_0, δ_0, and $\gamma_0(\cdot)$. The novel feature of this model is that the sample splitting is determined by an unknown relationship between two variables q_i and s_i, and their relationship is characterized by the nonparametric threshold function $\gamma_0(\cdot)$. In contrast, the classical threshold regression models assume $\gamma_0(\cdot)$ to be a constant or linear index. This specification can cover interesting cases that have not been studied. For example, we can consider the model that the threshold is heterogeneous and specific to each observation i if we see $\gamma_0(s_i) = \gamma_{0i}$; or the model that the threshold is determined by the direction of some moment conditions $\gamma_0(s_i) = \mathbb{E}[q_i|s_i]$. Apparently, when $\gamma_0(s) = \gamma_0$ or $\gamma_0(s) = \gamma_0 s$ for some parameter γ_0 and $s \neq 0$, it reduces to the standard threshold regression model.

To illustrate the empirical significance of the nonparametric threshold model (1), we revisit two important questions in public/labor and urban economics, respectively. The first one is about the tipping point model proposed by Schelling (1971), who analyzes the phenomenon that a neighborhood's white population substantially decreases once the minority share exceeds a certain threshold, called the tipping point. Card, Mas, and Rothstein (2008) empirically estimate the tipping point model by considering the constant threshold regression, $y_i = \beta_{10} + \delta_{10}\mathbf{1}\left[q_i > \gamma_0\right] + x_{2i}^\top \beta_{20} + u_i$, where y_i and q_i denote the white population change in a decade and the initial minority share in the

ith tract, respectively. The parameters δ_{10} and γ_0 denote the change size and the threshold, respectively. In Section VII of Card, Mas, and Rothstein (2008), however, they find that the tipping point γ_0 varies across cities depending on the white's attitudes toward the minority. This finding raises the concern on the constant threshold model and motivates the more general model (1) by specifying the tipping point γ_0 as a nonparametric function of local demographic characteristics as demonstrated in Section 6.1.

For the second application, we use model (1) to define metropolitan area boundaries, which is a fundamental problem in urban economics. Recently, many studies propose to use nighttime light intensity collected from satellite imagery to define the metropolitan area. They set an ad hoc level of light intensity as a threshold and categorize a pixel in the satellite imagery as a part of the metropolitan area if the light intensity of that pixel is higher than the threshold. See, for example, Rozenfeld, Rybski, Gabaix, and Makse (2011), Henderson, Storeygard, and Weil (2012), Dingel, Miscio, and Davis (2019) and Vogel, Goldblatt, Hanson, and Khandelwal (2019). In contrast, the model (1) can provide a guidance of choosing the intensity threshold from the econometric perspective, if we let y_i as the light intensity in the ith pixel and (q_i, s_i) as the location information of that pixel (more precisely, the radius and the angle relative to some city center in the polar coordinate). In Section 6.2, we estimate the metropolitan area of Dallas, Texas, especially its development from 1995 to 2010, and find a substantial difference from the conventional approaches. To the best of our knowledge, this is the first paper to nonparametrically determine the metropolitan area using a threshold model.

We develop a two-step estimator of (1), where we estimate $\gamma_0(\cdot)$ using local constant estimation. Under the shrinking threshold setup (e.g., Bai (1997), Bai and Perron (1998), and Hansen (2000)) with $\delta_0 = c_0 n^{-\epsilon}$ for some $c_0 \neq 0$ and $\epsilon \in (0, 1/2)$, we show that the nonparametric estimator $\widehat{\gamma}(\cdot)$ is uniformly consistent and $(\widehat{\beta}^\top, \widehat{\delta}^\top)^\top$ satisfies the root-n-consistency. The uniform rate of convergence and the pointwise limiting distribution of $\widehat{\gamma}(\cdot)$ are also derived. We also develop a pointwise specification test of $\gamma_0(s)$ for any given s (i.e., a test for the null hypothesis $H_0 : \gamma_0(s) = \gamma_*(s)$).

We can highlight some novel technical features of the new estimator as follows. First, since the nonparametric function $\gamma_0(\cdot)$ is inside the indicator function, deriving the asymptotic properties requires nonstandard proof. In particular, we establish the uniform rate of convergence of $\widehat{\gamma}(\cdot)$, which involves substantially more complicated derivations than the standard (constant) threshold regression model (e.g., Hansen (2000)). Second, we find that, unlike the standard local constant estimator, $\widehat{\gamma}(\cdot)$ is asymptotically unbiased even if the optimal bandwidth is used. Also, when the change size δ_0 is large (i.e., ϵ is close to 0), the optimal rate of convergence of $\widehat{\gamma}(\cdot)$ is close to $n^{-1/2}$. In the standard kernel regression, this root-n rate is obtained when the unknown function is infinitely differentiable, while we only require the second-order differentiability of $\gamma_0(\cdot)$. Third, to achieve the Neyman orthogonality in semiparametric estimation (e.g., Andrews (1994)), we propose to use the observations that are far away from the estimated threshold function in the second step estimation. The choice of this distance is obtained by the uniform convergence rate of $\widehat{\gamma}(\cdot)$. Finally, we let the variables be cross-sectionally dependent by considering the strong-mixing random field as Bolthausen (1982). This generalization allows us to study sample splitting in spatial data. For instance, if we let (q_i, s_i) correspond to the geographical location (i.e., latitude and longitude on the map), then the threshold $\mathbf{1}[q_i \leq \gamma_0(s_i)]$ identifies the unknown border yielding a two-dimensional sample splitting. In more general contexts, the model can be applied to identify social or economic segregation over interacting agents.

The rest of the paper is organized as follows. Section 2 sets up the model, previews our estimators, and establishes the identification. Section 3 further derives the asymptotic properties of the estimators and develops a likelihood ratio test of the threshold function. Section 4 describes how to extend the main model to a threshold contour. Section 5 studies small sample properties of the proposed statistics by Monte Carlo simulations. Section 6 applies the new method to estimate the tipping point function and to determine metropolitan areas. Section 7 concludes this paper with some remarks. The main proofs are in the Appendix, and all the omitted proofs are collected in the supplementary material.

We use the following notations. Let $\rightarrow_p$ denote convergence in probability, $\rightarrow_d$ convergence in distribution, and $\Rightarrow$ weak convergence of the underlying probability measure as $n \rightarrow \infty$. Let $\lfloor r \rfloor$ denote the biggest integer smaller than or equal to r and $\mathbf{1}[A]$ the indicator function of a generic event A. Let $\|B\|$ denote the Euclidean norm of a vector or matrix B, and C a generic constant that may vary over different lines.

2 Nonparametric Sample Splitting

We consider a threshold regression model given by (1), which is

$$y_i = x_i^\top \beta_0 + x_i^\top \delta_0 \mathbf{1}\left[q_i \le \gamma_0\left(s_i\right)\right] + u_i$$

for $i = 1, \ldots, n$, where $(y_i, x_i^\top, q_i, s_i)^\top \in \mathbb{R}^{1+p+1+1}$ are observed but the threshold function $\gamma_0 : \mathbb{R} \to \mathbb{R}$ as well as the regression coefficients $\theta_0 = (\beta_0^\top, \delta_0^\top)^\top \in \mathbb{R}^{2p}$ are unknown.[1] The parameters of interest are θ_0 and $\gamma_0(\cdot)$. Denote $\mathcal{Q} \subset \mathbb{R}$ and $\mathcal{S} \subset \mathbb{R}$ as the supports of q_i and s_i, respectively.

We estimate this semiparametric model in two steps. First, for given $s \in \mathcal{S}$, we fix $\gamma_0(s) = \gamma$ and obtain $\widehat{\beta}\left(\gamma; s\right)$ and $\widehat{\delta}\left(\gamma; s\right)$ by local least squares estimation conditional on γ:

$$(\widehat{\beta}\left(\gamma; s\right), \widehat{\delta}\left(\gamma; s\right)) = \arg\min_{\beta,\delta} Q_n\left(\beta, \delta, \gamma; s\right), \tag{2}$$

where

$$Q_n\left(\beta, \delta, \gamma; s\right) = \sum_{i=1}^{n} K\left(\frac{s_i - s}{b_n}\right)\left(y_i - x_i^\top \beta - x_i^\top \delta \mathbf{1}\left[q_i \le \gamma\right]\right)^2 \tag{3}$$

for some kernel function $K\left(\cdot\right)$ and a bandwidth parameter b_n. Suppose the space of $\gamma_0\left(s\right)$ for any s is a compact set Γ that is strictly within $\mathcal{Q}$,[2] then $\gamma_0(s)$ is estimated by

$$\widehat{\gamma}\left(s\right) = \arg\min_{\gamma \in \Gamma_n} Q_n\left(\gamma; s\right)$$

for given s, where $\Gamma_n = \Gamma \cap \{q_1, \ldots, q_n\}$ and $Q_n\left(\gamma; s\right)$ is the concentrated sum of squares defined as

$$Q_n\left(\gamma; s\right) = Q_n\left(\widehat{\beta}\left(\gamma; s\right), \widehat{\delta}\left(\gamma; s\right), \gamma; s\right). \tag{4}$$

The nonparametric estimator $\widehat{\gamma}\left(s\right)$ can be seen as a local version of the standard (constant) threshold regression estimator. Comparing to local linear estimation, this local constant estimation substantially reduces the computational burden since it requires computing the criteria function for only n times. If we implement local linear estimation by considering $\mathbf{1}\left[q_i \le \gamma_1 + \gamma_2(s_i - s)\right]$ in (3), we have to numerically de-

[1] The main results of this paper can be extended to consider multi-dimensional s_i using multivariate kernels. However, we only consider the scalar case for the expositional simplicity. Furthermore, the results are readily generalized to the case where only a subset of parameters differ between regimes.

[2] When the space of $\gamma_0(s)$ varies over s, we let Γ be the smallest compact set that includes $\cup_{s \in \mathcal{S}} \Gamma(s)$, where $\gamma_0(s) \in \Gamma(s)$ for each s.

termine γ_1 and γ_2 simultaneously, which is very difficult to solve by grid search as illustrated by Yu and Fan (2019). Also, to avoid additional technical complexity, we focus on estimation of $\gamma_0(s)$ at $s \in \mathcal{S}_0 \subset \mathcal{S}$ for some compact interior subset $\mathcal{S}_0$ of the support, say the middle 70% quantiles.

In the second step, we estimate the parametric components β_0 and δ_0. Different from existing threshold literature, we cannot treat $\widehat{\gamma}(\cdot)$ as the known threshold and simply regress y_i on x_i and $x_i \mathbf{1}[q_i \leq \widehat{\gamma}(s_i)]$ because the bias of $\widehat{\gamma}(\cdot)$ in the first stage estimation can be large. As an alternative, we estimate β_0 and $\delta_0^* = \beta_0 + \delta_0$ using the observations that are far away from the estimated $\gamma_0(s_i)$. This is implemented by considering

$$\widehat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - x_i^\top \beta\right)^2 \mathbf{1}\left[q_i > \widehat{\gamma}_{-i}(s_i) + \Delta_n\right] \mathbf{1}[s_i \in \mathcal{S}_0], \quad (5)$$

$$\widehat{\delta}^* = \arg\min_{\delta^*} \sum_{i=1}^{n} \left(y_i - x_i^\top \delta^*\right)^2 \mathbf{1}\left[q_i < \widehat{\gamma}_{-i}(s_i) - \Delta_n\right] \mathbf{1}[s_i \in \mathcal{S}_0] \quad (6)$$

for some constant $\Delta_n > 0$ satisfying $\Delta_n \to 0$ as $n \to \infty$, which is defined later. We use the leave-one-out estimator $\widehat{\gamma}_{-i}(s)$ in the first step. The change size δ can be estimated as $\widehat{\delta} = \widehat{\delta}^* - \widehat{\beta}$.

We now introduce the conditions for identification.

Assumption ID

(i) $\mathbb{E}[u_i x_i | q_i, s_i] = 0$.

(ii) $\mathbb{E}\left[x_i x_i^\top\right] > \mathbb{E}\left[x_i x_i^\top \mathbf{1}[q_i \leq \gamma]\right] > 0$ *for any* $\gamma \in \Gamma$.

(iii) $(\beta_0^\intercal, \delta_0^\intercal)^\intercal$ *are in the interior of some compact subsets of* $\mathbb{R}^{2p}$.

(iv) For any $s \in \mathcal{S}$, *there exists* $\varepsilon(s) > 0$ *such that* $\varepsilon(s) < \mathbb{P}(q_i \leq \gamma_0(s_i) | s_i = s) < 1 - \varepsilon(s)$ *and* $\delta_0^\top \mathbb{E}\left[x_i x_i^\top | q_i = q, s_i = s\right] \delta_0 > 0$ *for all* $(q, s) \in \mathcal{Q} \times \mathcal{S}$.

(v) q_i *is continuously distributed with its conditional density* $f(q|s)$ *satisfying that* $0 < C_1 < f(q|s) < C_2 < \infty$ *for all* $(q, s) \in \Gamma \times \mathcal{S}$ *and some constants* C_1 and C_2.

Assumption ID is mild. In particular, the condition (i) excludes endogeneity, and (ii) is the full rank condition. Assumption ID-(iv) restricts that the threshold $\gamma_0(s)$

lies in the interior of the support of q_i for any $s \in \mathcal{S}$ and the coefficient change exists (e.g., $\delta_0 \neq 0$). Assumption ID-(v) requires the conditional density of q_i given any s_i is positive. Under these conditions, the following theorem establishes the identification of all unknown parameters.[3]

Theorem 1 *Under Assumption ID, the threshold function $\gamma_0(\cdot)$ and the parameters $(\beta_0^\top, \delta_0^\top)^\top$ are uniquely identified.*

For asymptotic derivation, we allow for cross-sectional dependence to study the spatial sampling splitting. More precisely, we suppose $\left(x_i^\top, q_i, s_i, u_i\right)^\top$ is generated from a strictly stationary and α-mixing random field. See, for example, pp.1047 in Bolthausen (1982) and Assumption 1 in Jenish and Prucha (2009). We consider the samples over a random expanding lattice $N_n \subset \mathbb{R}^2$ endowed with a metric $\lambda(i,j) = \max_{1\leq \ell \leq 2} |i_\ell - j_\ell|$ and the corresponding norm $\max_{1\leq \ell\leq 2}|i_\ell|$, where i_ℓ denotes the ℓth component of i. We denote $|N_n|$ as the cardinality of N_n and $\partial N_n = \{i \in N_n\colon$ there exists $j \notin N_n$ with $\lambda(i,j) = 1\}$. Let $|N_n| = n$ and then the summation in (3) can be written as $\sum_{i\in N_n}$. We also define a mixing coefficient:

$$\alpha(m) = \sup\left\{\left|\mathbb{P}\left(A_i \cap A_j\right) - \mathbb{P}\left(A_i\right)\mathbb{P}\left(A_j\right)\right| : A_i \in \mathcal{F}_i \text{ and } A_j \in \mathcal{F}_j \text{ with } \lambda(i,j) \geq m\right\}, \tag{7}$$

where $\mathcal{F}_i$ is the σ-algebra generated by $\left(x_i^\top, q_i, s_i, u_i\right)^\top$.

The following conditions are imposed for deriving the asymptotic properties of our two-step estimator. Let $f(q,s)$ be the joint density function of (q_i, s_i) and

$$D(q,s) = \mathbb{E}[x_i x_i^\top | (q_i, s_i) = (q,s)], \tag{8}$$

$$V(q,s) = \mathbb{E}[x_i x_i^\top u_i^2 | (q_i, s_i) = (q,s)]. \tag{9}$$

Assumption A

(i) The lattice $N_n \subset \mathbb{R}^2$ is infinite countable; all the elements in N_n are located at distances at least $\lambda_0 > 1$ from each other, i.e., for any $i, j \in N_n : \lambda(i,j) \geq \lambda_0$; and $\lim_{n\to\infty} |\partial N_n|/n = 0$.

[3] Since the last condition in Assumption ID-(iv) does not require the strict positive definiteness of $\mathbb{E}\left[x_i x_i^\top | q_i = q, s_i = s\right]$, q_i or s_i can be one of the elements of x_i (e.g., threshold autoregressive model, Tong (1983)) or a linear combination of x_i, even when x_i includes a constant term.

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

(iii) $\left(x_i^\top, q_i, s_i, u_i\right)^\top$ *is strictly stationary and* α*-mixing with bounded* $(2+\varphi)$*th moments for some* $\varphi > 0$*; the mixing coefficient* $\alpha(m)$ *defined in (7) satisfies* $\sum_{m=1}^{\infty} m\alpha(m) < \infty$ *and* $\sum_{m=1}^{\infty} m^2 \alpha(m)^{\varphi/(2+\varphi)} < \infty$ *for some* $\varphi \in (0, 2)$.

(iv) $0 < \mathbb{E}\left[u_i^2 | x_i, q_i, s_i\right] < \infty$ *almost surely.*

(v) Uniformly in (q, s)*, there exists some constant* $C < \infty$ *such that* $\mathbb{E}[||x_i||^8 \,|(q_i, s_i) = (q, s)] < C$ *and* $\mathbb{E}[||x_i u_i||^8 \,|(q_i, s_i) = (q, s)] < C$.

(vi) $\gamma_0 : \mathcal{S} \mapsto \Gamma$ *is a twice continuously differentiable function with bounded derivatives.*

(vii) $D(q, s)$, $V(q, s)$, *and* $f(q, s)$ *are bounded, continuous in* q*, and twice continuously differentiable in* s *with bounded derivatives.*

(viii) $c_0^\top D(\gamma_0(s), s)\, c_0 > 0$, $c_0^\top V(\gamma_0(s), s)\, c_0 > 0$, *and* $0 < C_1 < f(\gamma_0(s), s) < C_2 < \infty$ *for all* $s \in \mathcal{S}$ and some constants C_1 and C_2.

(ix) As $n \to \infty$, $b_n \to 0$ *and* $n^{1-2\epsilon} b_n \to \infty$.

(x) $K(\cdot)$ *is uniformly bounded, continuous, symmetric around zero, and satisfies* $\int K(v)\, dv = 0$, $\int v^2 K(v)\, dv \in (0, \infty)$, $\int K^2(v) dv \in (0, \infty)$, *and* $\lim_{v \to \infty} |v| K(v) = 0$.

We provide some discussions about these assumptions. First, we assume that q_i and s_i are continuous random variables to characterize the threshold model as in the example in Section 6.1. However, this setup can cover the two-dimensional "structural break" model as a special case, where q_i and s_i are non-random indices on a two-dimensional grid, respectively, as the geographic location in Section 6.2. In this case, we denote n_1 and n_2 as the numbers of rows (latitudes) and columns (longitudes) in the grid of pixels, and we normalize q and s in the way that $q \in \{1/n_1, 2/n_1, \ldots, 1\}$ and $s \in \{1/n_2, 2/n_2, \ldots, 1\}$. Under similar regularity conditions as Assumption A, we can show that the asymptotic results in the following sections are the same as if $(q_i, s_i)^\top$ were independently uniformly distributed over $[0, 1]^2$. This similarity is also found in the standard structural break and the threshold regression models (e.g., Proposition 5 in Bai and Perron (1998) and Theorem 1 in Hansen (2000)). We provide more details in the supplementary material.

Second, Assumption A is mild and common in the existing literature. In particular, Assumption A-(i) is the same as in Bolthausen (1982) to define the latent random field. Note that λ_0 in Assumption A-(i) can be any strictly positive value, and hence we can impose $\lambda_0 > 1$ without loss of generality. In Assumption A-(ii), we adopt the widely used shrinking change size setup as in Bai (1997), Bai and Perron (1998), and Hansen (2000) to obtain a simple limiting distribution. In contrast, a constant change size ($\epsilon = 0$) leads to a complicated asymptotic distribution of the threshold estimator, which depends on nuisance parameters (e.g., Chan (1993)). The conditions in Assumption A-(iii) are required to establish the central limit theorem (CLT) for the spatially dependent random field. The condition on the mixing coefficient is slightly stronger than that of Bolthausen (1982), which is because we need to control for the dependence within the local neighborhood in kernel estimation. When $\alpha(m)$ decays at an exponential rate, these conditions are readily satisfied. When $\alpha(m)$ decays at a polynomial rate (i.e., $\alpha(m) \leq C_\alpha m^{-k}$ for some $k > 0$), we need some restrictions on k and φ to satisfy these conditions, such as $k > 3(2+\varphi)/\varphi$. Assumptions-(iv) to (viii) are similar to Assumption 1 of Hansen (2000). Assumptions A-(ix) and (x) are standard in the kernel estimation literature, except that the magnitude of the bandwidth b_n depends on both n and ϵ. The conditions in A-(x) hold for many commonly used kernel functions, such as the Gaussian kernel and the uniform kernel.

Third, it is important to note that we assume γ_0 to be a function from $\mathcal{S}$ to Γ in Assumption A-(vi), which is not necessarily one-to-one. For this reason, sample splitting based on $\mathbf{1}[q_i \leq \gamma_0(s_i)]$ can be different from that based on $\mathbf{1}[s_i \geq \breve{\gamma}_0(q_i)]$ for some function $\breve{\gamma}_0$. Instead of restricting γ_0 to be one-to-one in this paper, for the identification purpose, we presume that we know which variables should be respectively assigned as q_i and s_i from the context. In Section 4, however, we discuss how to relax this point to identify a threshold contour as an extreme case.

3 Asymptotic Results

We first obtain the asymptotic properties of $\widehat{\gamma}(s)$. The following theorem derives the pointwise consistency and the pointwise rate of convergence at the interior points of $\mathcal{S}$.

Theorem 2 *For a given $s \in \mathcal{S}_0$, under Assumptions ID and A, $\widehat{\gamma}(s) \to_p \gamma_0(s)$ as $n \to \infty$. Furthermore,*

$$\widehat{\gamma}(s) - \gamma_0(s) = O_p\left(\frac{1}{n^{1-2\epsilon} b_n}\right)$$

provided that $n^{1-2\epsilon} b_n^2$ does not diverge.

The pointwise rate of convergence of $\widehat{\gamma}(s)$ depends on two parameters, ϵ and b_n. It is decreasing in ϵ like the parametric (constant) threshold case: a larger ϵ reduces the threshold effect $\delta_0 = c_0 n^{-\epsilon}$ and hence decreases the effective sampling information on the threshold. Since we estimate $\gamma_0(\cdot)$ using the kernel estimation method, the rate of convergence depends on the bandwidth b_n as well. As in the standard kernel estimator case, a smaller bandwidth decreases the effective local sample size, which reduces the precision of the estimator $\widehat{\gamma}(s)$. Therefore, in order to have a sufficient level of rate of convergence, we need to choose b_n large enough when the threshold effect δ_0 is expected to be small (i.e., when ϵ is close to 1/2).

Unlike the standard kernel estimator, there appears no bias-variance trade-off in $\widehat{\gamma}(s)$ as we further discuss after Theorem 3. It thus seems like that we can improve the rate of convergence by choosing a larger bandwidth b_n. However, b_n cannot be chosen too large to result in $n^{1-2\epsilon} b_n^2 \to \infty$, because otherwise $n^{1-2\epsilon} b_n(\widehat{\gamma}(s) - \gamma_0(s))$ is no longer $O_p(1)$. Therefore, we can use the restriction $n^{1-2\epsilon} b_n^2 \to \varrho$ for some $\varrho \in (0, \infty)$ to obtain the optimal bandwidth.

Under the choice that $n^{1-2\epsilon} b_n^2 \to \varrho \in (0, \infty)$, the optimal bandwidth can be chosen as $b_n^* = n^{-(1-2\epsilon)/2} c^*$ for some constant $0 < c^* < \infty$. This b_n^* provides the fastest convergence rate. Using this optimal bandwidth, the optimal pointwise rate of convergence of $\widehat{\gamma}(s)$ is then given as $n^{-(1-2\epsilon)/2}$. However, such a bandwidth choice is not feasible in practice since the constant term c^* is unknown, which also depends on the nuisance parameter ϵ that is not estimable. In practice, we suggest cross-validation as we implement in Section 6, although its statistical properties need to be studied further.[4]

[4] If ϵ is close to zero, the optimal rate of convergence of $\widehat{\gamma}(s)$ is close to $n^{-1/2}$ when the optimal bandwidth b_n^* is used. Such a fast convergence rate requires infinite order of smoothness in the standard kernel regressions with the MSE-optimal bandwidth. In contrast, we only require the second-order differentiability in this nonparametric threshold model.

The next theorem derives the limiting distribution of $\widehat{\gamma}(s)$. We let $W(\cdot)$ be a two-sided Brownian motion defined as in Hansen (2000):

$$W(r) = W_1(-r)\mathbf{1}[r<0] + W_2(r)\mathbf{1}[r>0], \tag{10}$$

where $W_1(\cdot)$ and $W_2(\cdot)$ are independent standard Brownian motions on $[0,\infty)$.

Theorem 3 *Under Assumptions ID and A, for a given $s \in \mathcal{S}_0$, if $n^{1-2\epsilon}b_n^2 \to \varrho \in (0,\infty)$,*

$$n^{1-2\epsilon}b_n\left(\widehat{\gamma}(s) - \gamma_0(s)\right) \to_d \xi(s) \arg\max_{r\in\mathbb{R}} \left(W(r) + \mu(r,\varrho;s)\right) \tag{11}$$

as $n \to \infty$, where

$$\begin{aligned}
\mu(r,\varrho;s) &= -|r|\,\psi_1(r,\varrho;s) + \varrho\psi_2(r,\varrho;s),\\
\psi_1(r,\varrho;s) &= \int_0^{\xi(s)|r|/(\varrho|\dot{\gamma}_0(s)|)} K(t)\,dt,\\
\psi_2(r,\varrho;s) &= \xi(s)\,|\dot{\gamma}_0(s)| \int_0^{\xi(s)|r|/(\varrho|\dot{\gamma}_0(s)|)} tK(t)\,dt,
\end{aligned}$$

and

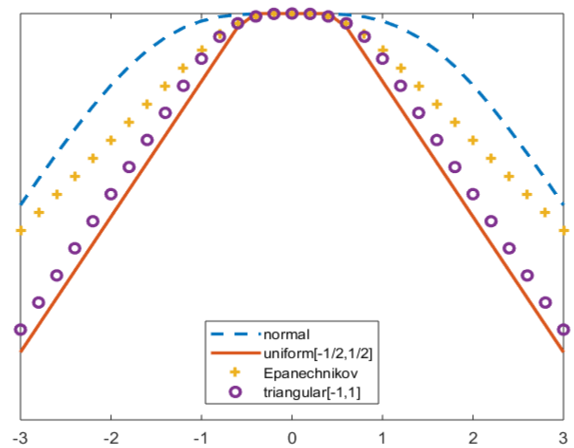
$$\xi(s) = \frac{\kappa_2 c_0^\top V(\gamma_0(s), s)\, c_0}{\left(c_0^\top D(\gamma_0(s), s)\, c_0\right)^2 f(\gamma_0(s), s)}$$

with $\kappa_2 = \int K(v)^2 dv$ and $\dot{\gamma}_0(s)$ is the first derivative of γ_0 at s. Furthermore,

$$\mathbb{E}\left[\arg\max_{r\in\mathbb{R}} \left(W(r) + \mu(r,\varrho;s)\right)\right] = 0.$$

The drift term $\mu(r,\varrho;s)$ in (11) depends on ϱ, the limit of $n^{1-2\epsilon}b_n^2 = (n^{1-2\epsilon}b_n)b_n$, and $|\dot{\gamma}_0(s)|$, the steepness of $\gamma_0(\cdot)$ at s. Interestingly, it resembles the typical $O(b_n)$ boundary bias of the standard local constant estimator even when s belongs to the interior of the support of s_i. This bias is from the inequality restriction in the indicator function of the threshold regression. Derivation of this result is non-standard and substantially different from that in Hansen (2000), as presented in Lemmas A.2 and A.13 in the Appendix.

Figure 1: Plot of drift functions with different kernels (color online)



However, having this non-zero drift term in the limiting expression does not mean that the limiting distribution of $\widehat{\gamma}(s)$ itself has a non-zero mean even when we use the optimal bandwidth $b_n^* = O(n^{-(1-2\epsilon)/2})$, which satisfies $n^{1-2\epsilon}b_n^{*2} \to \varrho \in (0,\infty)$. This is mainly because the drift function $\mu(r,\varrho;s)$ is symmetric about zero and hence the limiting random variable $\arg\max_{r\in\mathbb{R}}(W(r)+\mu(r,\varrho;s))$ is mean zero. In particular, we can show that the random variable $\arg\max_{r\in\mathbb{R}}(W(r)+\mu(r,\varrho;s))$ always has zero mean if $\mu(r,\varrho;s)$ is a non-random function that is symmetric about zero and monotonically decreasing fast enough. This result might be of independent research interest and is summarized in Lemma A.9 in the Appendix. Figure 1 depicts the drift function $\mu(r,\varrho;s)$ for various kernels when $|\dot{\gamma}_0(s)| = \xi(s) = \varrho = 1$.

Since the limiting distribution in (11) depends on unknown components, like ϱ and $\dot{\gamma}_0(s)$, it is hard to use this result for further inference. We instead suggest undersmoothing for practical use. More precisely, if we suppose $n^{1-2\epsilon}b_n^2 \to 0$ as $n \to \infty$, then the limiting distribution in (11) simplifies to[5]

$$n^{1-2\epsilon}b_n(\widehat{\gamma}(s)-\gamma_0(s)) \to_d \xi(s)\arg\max_{r\in\mathbb{R}}\left(W(r)-\frac{|r|}{2}\right) \tag{12}$$

as $n\to\infty$, which appears the same as in the parametric case in Hansen (2000) except for the scaling factor $n^{1-2\epsilon}b_n$. The distribution of $\arg\max_{r\in\mathbb{R}}(W(r)-|r|/2)$ is known

[5] We let $\psi_1(r,0;s) = \int_0^\infty K(t)\,dt = 1/2$.

(e.g., Bhattacharya and Brockwell (1976) and Bai (1997)), which is also described in Hansen (2000, p.581). The term $\xi(s)$ determines the scale of the distribution at given s: it increases in the conditional variance $\mathbb{E}[u_i^2|x_i, q_i, s_i]$; and decreases in the size of the threshold constant c_0 and the density of (q_i, s_i) near the threshold.

Even when $n^{1-2\epsilon}b_n^2 \to 0$ as $n \to \infty$, the asymptotic distribution in (12) still depends on the unknown parameter ϵ (or equivalently c_0) in $\xi(s)$ that is not estimable. Thus, this result cannot be directly used for inference of $\gamma_0(s)$. Alternatively, given any $s \in \mathcal{S}_0$, we can consider a pointwise likelihood ratio test statistic for

$$H_0 : \gamma_0(s) = \gamma_*(s) \quad \text{against} \quad H_1 : \gamma_0(s) \neq \gamma_*(s), \tag{13}$$

which is given as

$$LR_n(s) = \sum_{i=1}^{n} K\left(\frac{s_i - s}{b_n}\right) \frac{Q_n(\gamma_*(s), s) - Q_n(\widehat{\gamma}(s), s)}{Q_n(\widehat{\gamma}(s), s)}. \tag{14}$$

The following corollary obtains the limiting null distribution of this test statistic that is free of nuisance parameters. By inverting the likelihood ratio statistic, we can form a pointwise asymptotic confidence interval of $\gamma_0(s)$.

Corollary 1 *Suppose $n^{1-2\epsilon}b_n^2 \to 0$ as $n \to \infty$. Under the same condition in Theorem 3, for any fixed $s \in \mathcal{S}_0$, the test statistic in (14) satisfies*

$$LR_n(s) \to_d \xi_{LR}(s) \max_{r \in \mathbb{R}} (2W(r) - |r|) \tag{15}$$

as $n \to \infty$ under the null hypothesis (13), where

$$\xi_{LR}(s) = \frac{\kappa_2 c_0^\top V(\gamma_0(s), s) c_0}{\sigma^2(s) c_0^\top D(\gamma_0(s), s) c_0}$$

with $\sigma^2(s) = \mathbb{E}[u_i^2|s_i = s]$ and $\kappa_2 = \int K(v)^2 dv$.

When $\mathbb{E}[u_i^2|x_i, q_i, s_i] = \mathbb{E}[u_i^2|s_i]$, which is the case of local conditional homoskedasticity, the scale parameter $\xi_{LR}(s)$ is simplified as κ_2, and hence the limiting null distribution of $LR_n(s)$ becomes free of nuisance parameters and the same for all $s \in \mathcal{S}_0$. Though this limiting distribution is still nonstandard, the critical values in this case can

Table 1: Simulated Critical Values of the LR Test (Gaussian Kernel)

$\mathbb{P}(\zeta^* > cv)$	0.800	0.850	0.900	0.925	0.950	0.975	0.990
cv	1.268	1.439	1.675	1.842	2.074	2.469	2.988

Note: ζ^* is the limiting distribution of $LR_n(s)$ under the local conditional homoskedasticity. The Gaussian kernel is used.

be obtained using the same method as Hansen (2000, p.582) with the scale adjusted by κ_2. More precisely, since the distribution function of $\zeta = \max_{r\in\mathbb{R}}(2W(r) - |r|)$ is given as $\mathbb{P}(\zeta \leq z) = (1 - e^{-z/2})^2 \mathbf{1}[z \geq 0]$, the distribution function of $\zeta^* = \kappa_2\zeta$ is $\mathbb{P}(\zeta^* \leq z) = (1 - e^{-z/2\kappa_2})^2 \mathbf{1}[z \geq 0]$, where ζ^* is the limiting random variable of $LR_n(s)$ given in (15) under the local conditional homoskedasticity. By inverting it, we can obtain the asymptotic critical values given a choice of $K(\cdot)$. For instance, the asymptotic critical values for the Gaussian kernel is reported in Table 1, where $\kappa_2 = (2\sqrt{\pi})^{-1} \simeq 0.2821$ in this case.

In general, we can estimate $\xi_{LR}(s)$ by

$$\widehat{\xi}_{LR}(s) = \frac{\kappa_2 \widehat{\delta}^\top \widehat{V}(\widehat{\gamma}(s), s) \widehat{\delta}}{\widehat{\sigma}^2(s) \widehat{\delta}^\top \widehat{D}(\widehat{\gamma}(s), s) \widehat{\delta}},$$

where $\widehat{\delta}$ is from (5) and (6), and $\widehat{\sigma}^2(s)$, $\widehat{D}(\widehat{\gamma}(s), s)$, and $\widehat{V}(\widehat{\gamma}(s), s)$ are the standard Nadaraya-Watson estimators. In particular, we let $\widehat{\sigma}^2(s) = \sum_{i=1}^n \omega_{1i}(s)\widehat{u}_i^2$ with $\widehat{u}_i = y_i - x_i^\top\widehat{\beta} - x_i^\top\widehat{\delta}\mathbf{1}[q_i \leq \widehat{\gamma}_{-i}(s_i)]$,

$$\widehat{D}(\widehat{\gamma}(s), s) = \sum_{i=1}^n \omega_{2i}(s) x_i x_i^\top, \text{ and } \widehat{V}(\widehat{\gamma}(s), s) = \sum_{i=1}^n \omega_{2i}(s) x_i x_i^\top \widehat{u}_i^2,$$

where

$$\omega_{1i}(s) = \frac{K((s_i - s)/b_n)}{\sum_{j=1}^n K((s_j - s)/b_n)} \text{ and } \omega_{2i}(s) = \frac{\mathbb{K}((q_i - \widehat{\gamma}(s))/b_n', (s_i - s)/b_n'')}{\sum_{j=1}^n \mathbb{K}((q_j - \widehat{\gamma}(s))/b_n', (s_j - s)/b_n'')}$$

for some bivariate kernel function $\mathbb{K}(\cdot,\cdot)$ and bandwidth parameters (b_n', b_n'').

Finally, we show the $\sqrt{n}$-consistency of the semiparametric estimators $\widehat{\beta}$ and $\widehat{\delta}^*$ in (5) and (6). For this purpose, we first obtain the uniform rate of convergence of $\widehat{\gamma}(s)$.

Theorem 4 *Under Assumptions ID and A,*

$$\sup_{s \in \mathcal{S}_0} |\widehat{\gamma}(s) - \gamma_0(s)| = O_p\left(\frac{\log n}{n^{1-2\epsilon} b_n}\right)$$

provided that $n^{1-2\epsilon} b_n^2$ does not diverge.

Apparently, the uniform consistency of $\widehat{\gamma}(s)$ follows provided $\log n/(n^{1-2\epsilon} b_n) \to 0$. Based on this uniform convergence, the following theorem derives the joint limiting distribution of $\widehat{\beta}$ and $\widehat{\delta}^*$. We let $\widehat{\theta}^* = (\widehat{\beta}^\top, \widehat{\delta}^{*\top})^\top$ and $\theta_0^* = (\beta_0^\top, \delta_0^{*\top})^\top$.

Theorem 5 *Suppose the conditions in Theorem 4 hold and $\log n/(n^{1-2\epsilon} b_n) \to 0$ as $n \to \infty$. If we let $\Delta_n > 0$ such that $\Delta_n \to 0$, $\{\log n/(n^{1-2\epsilon} b_n)\}/\Delta_n \to 0$ as $n \to \infty$, we have*

$$\sqrt{n}\left(\widehat{\theta}^* - \theta_0^*\right) \to_d \mathcal{N}\left(0, \Lambda^{*-1} \Omega^* \Lambda^{*-1}\right) \tag{16}$$

as $n \to \infty$, where

$$\Lambda^* = \begin{bmatrix} \mathbb{E}\left[x_i x_i^\top \mathbf{1}_i^+\right] & 0 \\ 0 & \mathbb{E}\left[x_i x_i^\top \mathbf{1}_i^-\right] \end{bmatrix} \quad \text{and} \quad \Omega^* = \lim_{n \to \infty} n^{-1} Var \begin{bmatrix} \sum_{i=1}^n x_i u_i \mathbf{1}_i^+ \\ \sum_{i=1}^n x_i u_i \mathbf{1}_i^- \end{bmatrix}$$

with $\mathbf{1}_i^+ = \mathbf{1}[q_i > \gamma_0(s_i)]\mathbf{1}[s_i \in \mathcal{S}_0]$ and $\mathbf{1}_i^- = \mathbf{1}[q_i < \gamma_0(s_i)]\mathbf{1}[s_i \in \mathcal{S}_0]$.

Note that we do not use the conventional plug-in estimator, $\arg\min_{\beta,\delta} \sum_{i=1}^n (y_i - x_i^\top \beta - x_i^\top \delta \mathbf{1}\left[q_i \leq \widehat{\gamma}_{-i}(s_i)\right])^2 \mathbf{1}[s_i \in \mathcal{S}_0]$, in our second step. The reason is that this estimator may not be asymptotically orthogonal to the first-step nonparametric estimator $\widehat{\gamma}(s)$ when $n^{1-2\epsilon} b_n^2 \to \varrho \in (0, \infty)$ as $n \to \infty$, though they are still consistent. This is further because $\widehat{\gamma}(s)$ could have very slow rate of convergence, and its estimation error will affect the limiting distribution of the second step estimator. Besides, unlike the standard semiparametric literature, the asymptotic effect of $\hat{\gamma}(s)$ to the second step estimation cannot be easily derived due to the discontinuity. The new estimation idea above, however, only uses the observations that are not affected by the estimation error in the first-step nonparametric estimator. This is done by choosing a large enough Δ_n in (5) and (6) such that the observations are outside the uniform convergence bound of $|\widehat{\gamma}(s) - \gamma_0(s)|$. Thanks to the threshold regression structure, we can estimate the parameters on each side of the threshold even using these subsamples. However, we also want $\Delta_n \to 0$ fast enough so that more observations are included in the estimation.

The estimator $(\widehat{\beta}^\top, \widehat{\delta}^{*\top})^\top$ thus satisfies the Neyman orthogonality condition (e.g., Assumption N(c) in Andrews (1994)), that is, replacing $\widehat{\gamma}$ by the true γ_0 in estimating the parametric component has an effect at most $o_p(n^{-1/2})$ in their limiting distribution. Though we lose some efficiency in finite samples, we can derive the asymptotic normality of $(\widehat{\beta}^\top, \widehat{\delta}^\top)^\top$ that has mean zero and achieves the same asymptotic variance as if $\gamma_0(\cdot)$ was known.

Using the delta method, we can readily obtain the limiting distribution of $\widehat{\theta} = (\widehat{\beta}^\top, \widehat{\delta}^\top)^\top$ as

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \to_d \mathcal{N}\left(0, \Lambda^{-1}\Omega\Lambda^{-1}\right) \quad \text{as } n \to \infty, \tag{17}$$

where

$$\Lambda = \mathbb{E}\left[z_i z_i^\top \mathbf{1}\left[s_i \in \mathcal{S}_0\right]\right] \quad \text{and} \quad \Omega = \lim_{n\to\infty} n^{-1} Var\left[\sum_{i=1}^{n} z_i u_i \mathbf{1}\left[s_i \in \mathcal{S}_0\right]\right]$$

with $z_i = \left[x_i^\top, x_i^\top \mathbf{1}\left[q_i \leq \gamma_0\left(s_i\right)\right]\right]^\top$. The asymptotic variance expressions in (16) and (17) allow for cross-sectional dependence as they have the long-run variance (LRV) forms Ω^* and Ω. They can be consistently estimated by the spatial HAC estimator of Conley and Molinari (2007) using $\widehat{u}_i = (y_i - x_i^\top \widehat{\beta} - x_i^\top \widehat{\delta} \mathbf{1}\left[q_i \leq \widehat{\gamma}_{-i}\left(s_i\right)\right])\mathbf{1}[s_i \in \mathcal{S}_0]$. The terms Λ^* and Λ can be estimated by their sample analogues.

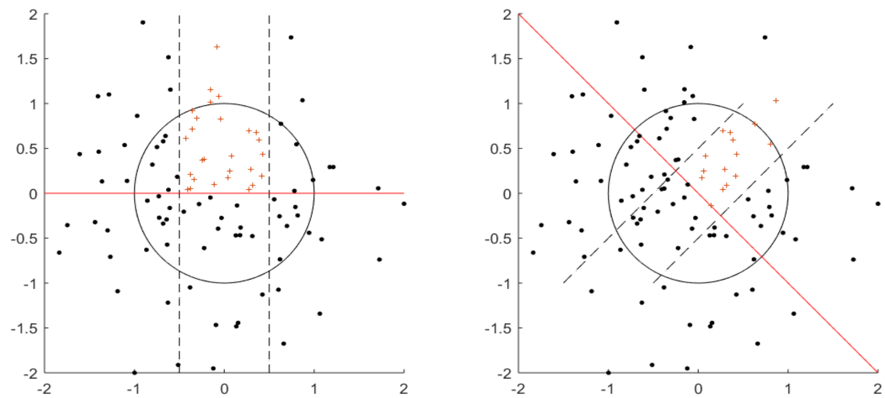
4 Threshold Contour

When we consider sample splitting over a two-dimensional space (i.e., q_i and s_i respectively correspond to the latitude and longitude on the map), the threshold model (1) can be generalized to estimate a nonparametric contour threshold model:

$$y_i = x_i^\top \beta_0 + x_i^\top \delta_0 \mathbf{1}\left[m_0\left(q_i, s_i\right) \leq 0\right] + u_i, \tag{18}$$

where the unknown function $m_0 : \mathcal{Q} \times \mathcal{S} \mapsto \mathbb{R}$ determines the contour on a random field. An interesting example includes identifying an unknown closed boundary over the map, such as a city boundary relative to some city center, and an area of a disease outbreak or airborne pollution. In social science, it can identify a group boundary

Figure 2: Illustration of rotation (color online)



or a region in which the agents share common demographic, political, or economic characteristics.

To relate this generalized form to the original threshold model (1), we suppose there exists a known center at (q_i^*, s_i^*) such that $m_0(q_i^*, s_i^*) < 0$. Without loss of generality, we can normalize (q_i^*, s_i^*) to be $(0,0)$ and re-center all other observations $\{q_i, s_i\}_{i=1}^n$ accordingly. In addition, we define the radius distance l_i and angle a_i° of the ith observation relative to the origin as

$$\begin{aligned} l_i &= \sqrt{q_i^2 + s_i^2}, \\ a_i^\circ &= \bar{a}_i^\circ \mathbf{I}_i + (180^\circ - \bar{a}_i^\circ)\,\mathbf{II}_i + (180^\circ + \bar{a}_i^\circ)\,\mathbf{III}_i + (360^\circ - \bar{a}_i^\circ)\,\mathbf{IV}_i, \end{aligned}$$

where $\bar{a}_i^\circ = \arctan(|q_i/s_i|)$, and each of $(\mathbf{I}_i, \mathbf{II}_i, \mathbf{III}_i, \mathbf{IV}_i)$ respectively denotes the indicator that the ith observation locates in the first, second, third, and forth quadrant.

We suppose that there is only one threshold at any angle and the threshold contour is star-shaped. For each fixed $a^\circ \in [0^\circ, 360^\circ)$, we rotate the original coordinate counterclockwise and implement the least squares estimation (4) only using the observations in the first two quadrants after rotation. Doing so ensures that the threshold mapping after rotation is a well-defined function.

In particular, the angle relative to the origin is $a_i^\circ - a^\circ$ after rotating the coordinate by a° degrees counterclockwise, and the new location (after the rotation) is given as

$(q_i(a^\circ), s_i(a^\circ))$, where

$$\begin{pmatrix} q_i(a^\circ) \\ s_i(a^\circ) \end{pmatrix} = \begin{pmatrix} q_i \cos(a^\circ) - s_i \sin(a^\circ) \\ s_i \cos(a^\circ) + q_i \sin(a^\circ) \end{pmatrix}.$$

After this rotation, we estimate the following nonparametric threshold model:

$$y_i = x_i^\top \beta_0 + x_i^\top \delta_0 \mathbf{1}[q_i(a^\circ) \leq \gamma_{a^\circ}(s_i(a^\circ))] + u_i \tag{19}$$

using only the observations satisfying $q_i(a^\circ) \geq 0$, where $\gamma_{a^\circ}(\cdot)$ serves as the unknown threshold line as in the model (1) in the a°-degree-rotated coordinate. Such reparametrization guarantees that $\gamma_{a^\circ}(\cdot)$ is always positive and we estimate its value pointwisely at 0. Figure 2 illustrates the idea of such rotation and pointwise estimation over a bounded support so that only the red cross points are included for estimation at different angles. Thus, the estimation and inference procedure developed before is directly applicable, though we expect efficiency loss as we only use a subsample in estimation at each rotated coordinate.

This rotating coordinate idea can be a quick solution when we do not know which variables should be assigned as q_i versus s_i, in the original model (1). As an extreme example, if γ_0 is the vertical line, the original model does not work. In this case, we can check if γ_0 is (near) the vertical line by investigating the estimates among different rotations; when γ_0 is suspected as the vertical line or has a very steep slope, we can switch q_i and s_i in the original model (1) to improve the local constant fitting.

5 Monte Carlo Experiments

We examine the small sample performance of the semiparametric threshold regression estimator by Monte Carlo simulations. We generate n draws from

$$y_i = x_i^\top \beta_0 + x_i^\top \delta_0 \mathbf{1}[q_i \leq \gamma_0(s_i)] + u_i, \tag{20}$$

where $x_i = (1, x_{2i})^\top$ and $x_{2i} \in \mathbb{R}$. We let $\beta_0 = (\beta_{10}, \beta_{20})^\top = 0\iota_2$ and consider three different values of $\delta_0 = (\delta_{10}, \delta_{20})^\top = \delta\iota_2$ with $\delta = 1, 2, 3,$ and 4 where $\iota_2 = (1, 1)^\top$. For the threshold function, we let $\gamma_0(s) = \sin(s)/2$. We consider the cross-sectional

Table 2: Rej. Prob. of the LR Test with i.i.d. Data

n	$\delta =$	$s = 0.0$				$s = 0.5$				$s = 1.0$			
		1	2	3	4	1	2	3	4	1	2	3	4
100		0.14	0.06	0.05	0.05	0.16	0.07	0.05	0.05	0.25	0.18	0.14	0.13
200		0.08	0.03	0.02	0.02	0.08	0.04	0.02	0.02	0.15	0.10	0.06	0.06
500		0.05	0.01	0.02	0.02	0.05	0.02	0.02	0.02	0.09	0.05	0.03	0.01

Note: Entries are rejection probabilities of the LR test (14) when data are generated from (20) with $\gamma_0(s) = \sin(s)/2$. The dependence structure is given in (21) with $\rho = 0$. The significance level is 5% and the results are based on 1000 simulations.

dependence structure in $(x_{2i}, q_i, s_i, u_i)^\top$ as follows:

$$\begin{cases} (q_i, s_i)^\top \sim iid\mathcal{N}(0, I_2); \\ x_{2i} | (q_i, s_i) \sim iid\mathcal{N}(0, (1 + \rho(s_i^2 + q_i^2))^{-1}); \\ \underline{\mathbf{u}} | \{(x_i, q_i, s_i)\}_{i=1}^n \sim \mathcal{N}(0, \Sigma), \end{cases} \tag{21}$$

where $\underline{\mathbf{u}} = (u_1, \ldots, u_n)^\top$. The (i,j)th element of Σ is $\Sigma_{ij} = \rho^{\lfloor \ell_{ij} n \rfloor} \mathbf{1}[\ell_{ij} < m/n]$, where $\ell_{ij} = \{(s_i - s_j)^2 + (q_i - q_j)^2\}^{1/2}$ is the L^2-distance between the ith and jth observations. The diagonal elements of Σ are normalized as $\Sigma_{ii} = 1$. This m-dependent setup follows from the Monte Carlo experiment in Conley and Molinari (2007) in the sense that there are roughly at most $2m^2$ observations that are correlated with each observation. Within the m distance, the dependence decays at a polynomial rate as indicated by $\rho^{\lfloor \ell_{ij} n \rfloor}$. The parameter ρ describes the strength of cross-sectional dependence in the way that a larger ρ leads to stronger dependence relative to the unit standard deviation. In particular, we consider the cases with $\rho = 0$ (i.e., i.i.d. observations), 0.5, and 1. We consider the sample size $n = 100$, 200, and 500 and set $\mathcal{S}_0$ to include the middle 70% observations of s_i.

First, Tables 2 and 3 report the small sample rejection probabilities of the LR test in (14) for $H_0 : \gamma_0(s) = \sin(s)/2$ against $H_1 : \gamma_0(s) \neq \sin(s)/2$ at the 5% nominal level at three different locations $s = 0$, 0.5, and 1. In particular, Table 2 examines the case with no cross-sectional dependence ($\rho = 0$), while Table 3 examines the case with cross-sectional dependence whose dependence decays slowly with $\rho = 1$ and $m = 10$. For the bandwidth parameter, we normalize s_i and q_i to have mean zero and unit standard deviation and choose $b_n = 0.5n^{-1/2}$ in the main regression. This choice is for undersmoothing as $n^{1-2\epsilon} b_n^2 = n^{-2\epsilon} \to 0$. To estimate $D(\gamma_0(s), s)$ and $V(\gamma_0(s), s)$,

Table 3: Rej. Prob. of the LR Test with Cross-sectionally Correlated Data

		$s=0.0$				$s=0.5$				$s=1.0$			
n	$\delta=$	1	2	3	4	1	2	3	4	1	2	3	4
100		0.19	0.10	0.07	0.03	0.20	0.10	0.08	0.07	0.28	0.19	0.17	0.11
200		0.10	0.04	0.03	0.03	0.12	0.07	0.04	0.04	0.21	0.11	0.08	0.04
500		0.05	0.02	0.02	0.02	0.06	0.03	0.02	0.02	0.14	0.05	0.03	0.03

Note: Entries are rejection probabilities of the LR test (14) when data are generated from (20) with $\gamma_0(s)=\sin(s)/2$. The dependence structure is given in (21) with $\rho=1$ and $m=10$. The significance level is 5% and the results are based on 1000 simulations.

Table 4: Coverage Prob. of the Plug-in Confidence Interval

		β_{20}				$\beta_{20}+\delta_{20}$				δ_{20}			
n	$\delta=$	1	2	3	4	1	2	3	4	1	2	3	4
100		0.85	0.89	0.91	0.87	0.87	0.87	0.89	0.90	0.85	0.87	0.93	0.91
200		0.86	0.90	0.93	0.93	0.89	0.92	0.94	0.93	0.85	0.90	0.93	0.92
500		0.83	0.92	0.95	0.96	0.84	0.90	0.93	0.94	0.78	0.88	0.93	0.95

Note: Entries are coverage probabilities of 95% confidence intervals for β_{20}, $\beta_{20}+\delta_{20}$, and δ_{20}. Data are generated from (20) with $\gamma_0(s)=\sin(s)/2$, where the dependence structure is given in (21) with $\rho=0.5$ and $m=3$. The results are based on 1000 simulations.

we use the rule-of-thumb bandwidths from the standard kernel regression satisfying $b_n'=O(n^{-1/5})$ and $b_n''=O(n^{-1/6})$. All the results are based on 1000 simulations. In general, the test for γ_0 performs better as (i) the sample size gets larger; (ii) the coefficient change gets more significant; (iii) the cross-sectional dependence gets weaker; and (iv) the target gets closer to the mid-support of s. When δ_0 and n are large, the LR test is conservative, which is also found in the classic threshold regression case (Hansen (2000)).

Second, Table 4 shows the finite sample coverage properties of the 95% confidence intervals for the parametric components β_{20}, $\delta_{20}^*=\beta_{20}+\delta_{20}$, and δ_{20}. The results are based on the same simulation design as above with $\rho=0.5$ and $m=3$. Regarding the tuning parameters, we use the same bandwidth choice $b_n=0.5n^{-1/2}$ as before and set the truncation parameter $\Delta_n=(nb_n)^{-1/2}$. Unreported results suggest that choice of the constant in the bandwidth matters particularly with small samples like $n=100$, but such effect quickly decays as the sample size gets larger. For the lag number required for the HAC estimator, we use the spatial lag order of 5 following Conley

Table 5: Coverage Prob. of the Plug-in Confidence Interval (w/ LRV adj.)

		β_{20}				$\beta_{20}+\delta_{20}$				δ_{20}			
n	$\delta=$	1	2	3	4	1	2	3	4	1	2	3	4
100		0.92	0.95	0.94	0.95	0.91	0.95	0.94	0.95	0.93	0.95	0.95	0.95
200		0.93	0.95	0.97	0.96	0.94	0.94	0.95	0.96	0.90	0.93	0.97	0.94
500		0.89	0.95	0.97	0.97	0.89	0.96	0.97	0.97	0.84	0.92	0.95	0.97

Note: Entries are coverage probabilities of 95% confidence intervals for β_{20}, $\beta_{20}+\delta_{20}$, and δ_{20} with a small sample adjustment of the LRV estimator. Data are generated from (20) with $\gamma_0(s)=\sin(s)/2$, where the dependence structure is given in (21) with $\rho=0.5$ and $m=3$. The results are based on 1000 simulations.

and Molinari (2007). Results with other lag choices are similar and hence omitted. The result suggests that the asymptotic normality is better approximated with larger samples and larger change sizes. Table 5 shows the same results with a small sample adjustment of the LRV estimator for Ω^* by dividing it by the sample truncation fraction $\sum_{i=1}^{n}(\mathbf{1}[q_i>\widehat{\gamma}_{-i}(s_i)+\Delta_n]+\mathbf{1}[q_i<\widehat{\gamma}_{-i}(s_i)-\Delta_n])\mathbf{1}[s_i\in\mathcal{S}_0]/\sum_{i=1}^{n}\mathbf{1}[s_i\in\mathcal{S}_0]$. This ratio enlarges the LRV estimator and hence the coverage probabilities, especially when the change size is small. It only affects the finite sample performance as it approaches one in probability as $n\to\infty$.

6 Applications

6.1 Tipping point and social segregation

The first example is about the tipping point problem in social segregation, which stimulates a vast literature in labor/public and political economics. Schelling (1971) initially proposes the tipping point model to study the fact that the white population decreases substantially once the minority share exceeds a certain tipping point. Card, Mas, and Rothstein (2008) empirically estimate this model and find strong evidence for such a tipping point phenomenon. In particular, they specify the threshold regression model as

$$y_i=\beta_{10}+\delta_{10}\mathbf{1}\left[q_i>\gamma_0\right]+x_{2i}^{\top}\beta_{20}+u_i,$$

where for tract i in a certain city, q_i denotes the minority share in percentage at the beginning of a certain decade, y_i the normalized white population change in percentage within this decade, and x_{2i} denotes a vector of control variables. They apply the least squares method to estimate the tipping point γ_0. For most cities and for the periods 1970-80, 1980-90, and 1990-2000, they find that white population flows exhibit the tipping-like behavior, with the estimated tipping points ranging approximately from 5% to 20% across cities.

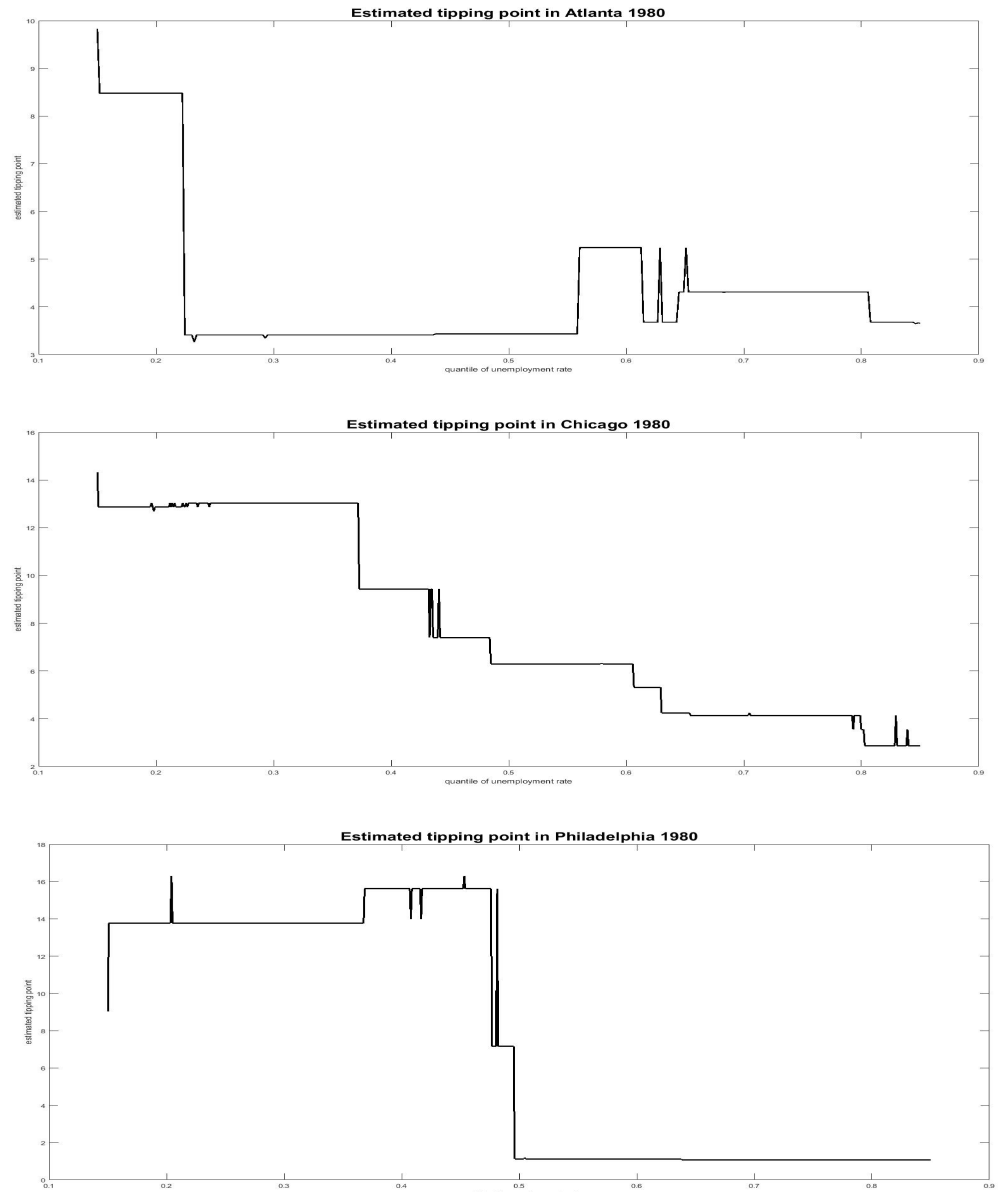
In Section VII of Card, Mas, and Rothstein (2008), they also find that the location of the tipping point substantially depends on white people's attitudes toward the minority. Specifically, they first construct a city-level index that measures white attitudes and regress the estimated tipping point from each city on this index. The regression coefficient is significantly different from zero, suggesting that the tipping point should be modeled as a function of the index. In this regards, a more robust model in the tract level can be written as

$$y_i = \beta_{10} + \delta_{10}\mathbf{1}\left[q_i > \gamma_0(s_i)\right] + x_{2i}^{\top}\beta_{20} + u_i,$$

where $\gamma_0(\cdot)$ denotes an unknown tipping point function, and s_i denotes the attitude index.

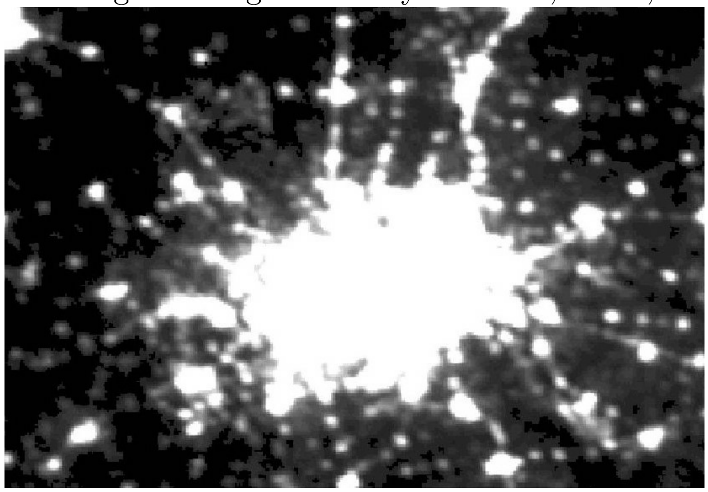
The attitude index by Card, Mas, and Rothstein (2008) is available only at the city-level, hence we cannot use it to analyze the census tract-level observations. Instead, we use the tract-level unemployment rate as s_i to illustrate the nonparametric threshold function. We use the data provided by Card, Mas, and Rothstein (2008) and estimate the tipping point function $\gamma_0(\cdot)$ over census tracts by the method introduced in Section 2. We use five control variables as x_{2i}, including the logarithm of mean family income, the fractions of single-unit, vacant, and renter-occupied housing units, and the fraction of workers who use public transport to travel to work. The bandwidth is chosen as $b_n = cn^{-1/2}$, with c being obtained from the leave-one-out cross-validation. Figure 3 depicts the estimated tipping point in the years 1980-90 in Atlanta, Chicago, and Philadelphia, where the sample sizes are relatively large. The pattern clearly shows that the tipping point varies substantially in the unemployment rate even within the city. Therefore, the standard constant tipping point model is insufficient to characterize the segregation fully.

Figure 3: Estimate of the tipping point as a function of the unemployment rate



Note: The figure depicts the point estimate of the tipping points as a function of the unemployment rate, using the data in Atlanta, Chicago, and Philadelphia in 1980-1990. Data are available from Card, Mas, and Rothstein (2008).

Figure 4: Nighttime light intensity in Dallas, Texas, in 2010



Note: The figure depicts the intensity of the stable nighttime light in Dallas 2010. Data are available from https://www.ncei.noaa.gov/.

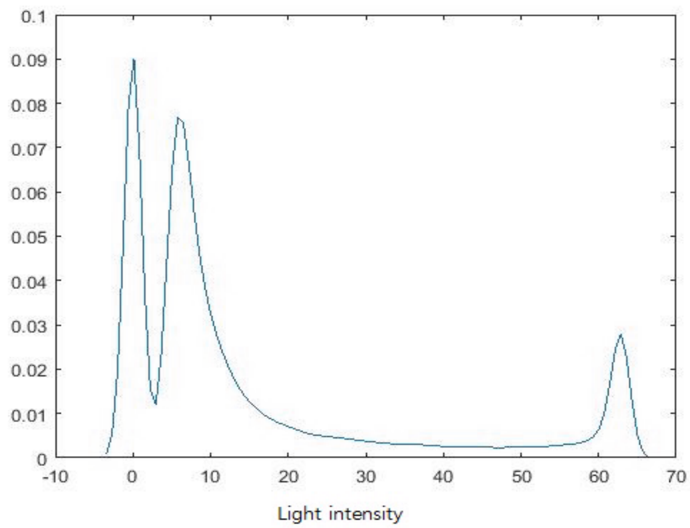
6.2 Metropolitan area determination

The second application is about determining the boundary of metropolitan areas, which is a fundamental question in urban economics. Recently, researchers propose to use nighttime light intensity obtained by satellite imagery to define the metropolitan boundary. The intuition is straightforward: metropolitan areas are bright at night while rural areas are dark.

Specifically, the National Oceanic and Atmospheric Administration (NOAA) collects satellite imagery of nighttime lights at approximately 1-kilometer resolution continuously since 1992. From there, NOAA further constructs several indices measuring the annual light intensity. Following convention, we choose the "average visible, stable lights" index that ranges from 0 (dark) to 63 (bright). For illustration, we focus on Dallas, Texas, and use the data for the years 1995, 2000, 2005, and 2010. In each year, the data are recorded as a 240×360 grid that covers the latitudes from 32°N to 34°N and the longitudes from 98.5°W to 95.5°W. The total sample size is 240×360=86400. These data are available at NOAA's website and also provided on the authors' website. Figure 4 depicts the data in 2010, which suggests a bright metropolitan area in the center of Dallas. Let y_i denote the intensity and (q_i, s_i) the latitude and longitude of the ith pixel (normalized into equally-spaced grids on $[0, 1]$).

To define the metropolitan area, existing literature in urban economics first chooses

Figure 5: Kernel density estimate of nighttime light intensity, Dallas 2010



Note: The figure depicts the kernel density estimate of the strength of the stable nighttime light in Dallas 2010. Data are available from https://www.ncei.noaa.gov/.

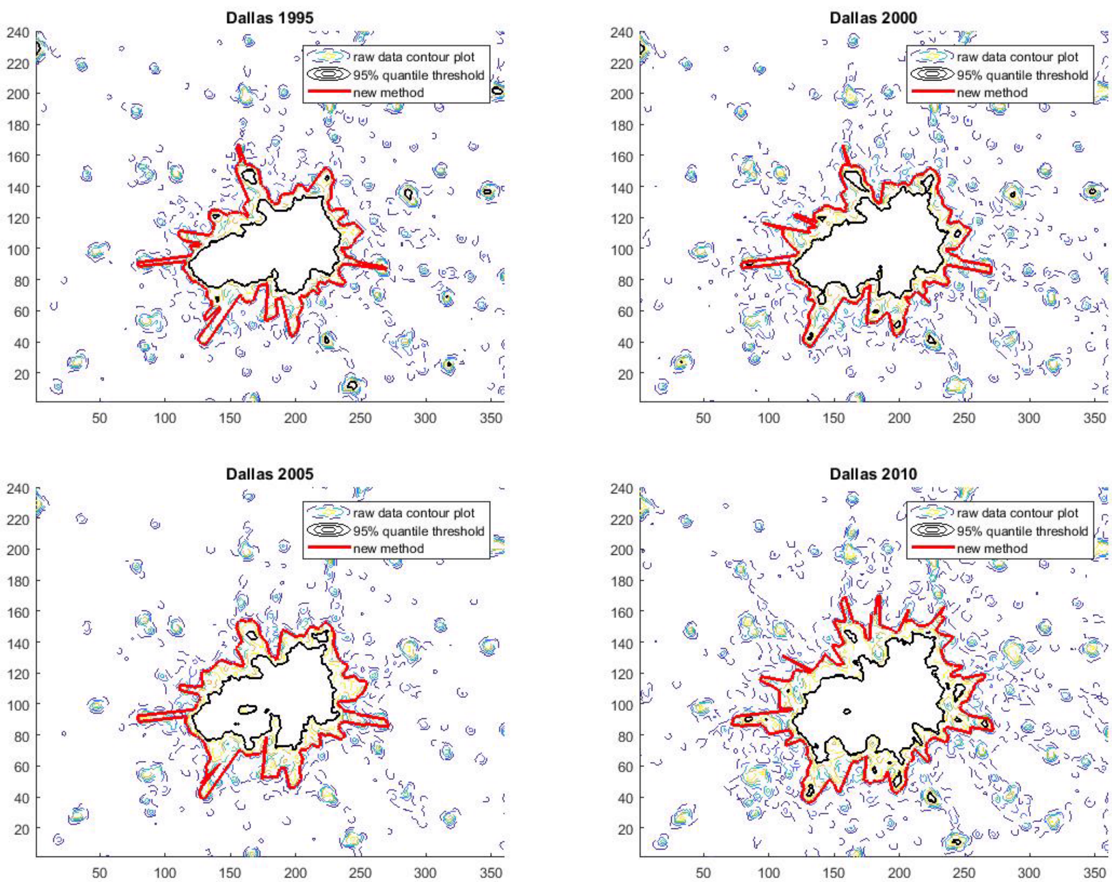
an ad hoc intensity threshold, say 95% quantile of y_i, and categorizes the ith pixel as a part of the metropolitan area if y_i is larger than the threshold. For example, see Dingel, Miscio, and Davis (2019), Vogel, Goldblatt, Hanson, and Khandelwal (2019), and references therein. On p.3 in Dingel, Miscio, and Davis (2019), they note that "... the choice of the light-intensity threshold, which governs the definitions of the resulting metropolitan areas, is not pinned down by economic theory or prior empirical research." Such arbitrariness can be solved using our new estimator.

We first examine whether the light intensity data exhibits a clear threshold-type pattern. To this end, we plot the kernel density estimates of y_i in the year 2010 in Figure 5. The bandwidth is the standard rule-of-thumb one. The estimated density exhibits three peaks at around 0, 8, and 63. They respectively correspond to the rural area, small towns, and the central metropolitan area. Therefore, the threshold model is appropriate in characterizing such a mean-shift pattern.

Now we implement the rotation and estimation method introduced in Section 4. In particular, we start with the center point in the bright middle area as our metropolitan center.[6] Then for each a° in the 500 equally-spaced grid on $[0, 360^\circ]$, we rotate the

[6]This corresponds to the pixel in the 181st column from the left and the 100th row from the bottom.

Figure 6: Metropolitan area determination in Dallas (color online)



Note: The figure depicts the city boundary determined by either the new method or by taking the 0.95 quantile of nighttime light strength as the threshold, using the satellite imagery data for Dallas in the years 1995, 2000, 2005, and 2010. Data are available from https://www.ncei.noaa.gov/.

data by a° degrees counterclockwise and estimate the model (19) with $x_i = 1$. The bandwidth is chosen as $cn^{-1/2}$ with $c = 1$. Other choices of c lead to almost identical results, given the large sample size. Figure 6 presents the estimated metropolitan area (red) and that determined by the 95% quantile of y_i (black).

Several interesting findings are summarized as follows. First, the estimated boundary is highly nonlinear as a function of the angle. Therefore, any parametric threshold model could lead to a substantially misleading result. Second, our estimated area is larger than that determined by the 95% quantile by 80.31%, 81.56%, 106.46%, and 102.09% in the years 1995, 2000, 2005, and 2010, respectively. In particular, our estimator tends to include some suburban areas that exhibit strong light intensity and that are geographically close to the city center. For example, the very left stretch-out point in the estimated boundary corresponds to Fort Worth, which is 30 miles from downtown Dallas. Residents can easily commute by train or driving on the interstate highway 30. It is then reasonable to include Fort Worth as a part of the metropolitan Dallas for further economic analysis. Finally, the estimated $\beta_0 + \delta_0$ is approximately 53, which corresponds to the 89% quantile of y_i. This finding provides a rule-of-thumb choice of the intensity threshold from the econometric point of view.

7 Concluding Remarks

This paper proposes a novel approach to conduct sample splitting. In particular, we develop a nonparametric threshold regression model where two variables can jointly determine the unknown splitting boundary. Our approach can be easily generalized so that the sample splitting depends on more numbers of variables, though such an extension is subject to the curse of dimensionality, as usually observed in the kernel regression literature. The main interest is in identifying the threshold function that determines how to split the sample. Thus our model should be distinguished from the smoothed threshold regression model or the random coefficient regression model.

This new approach is empirically relevant in broad areas studying sample splitting (e.g., segregations and group-formation) and heterogeneous effects over different subsamples. We illustrate this with the tipping point problem in social segregation and metropolitan area determination using satellite imagery datasets.

There are theoretical extensions and empirical applications of our method, which we suppress in the current paper due to space limitations. We list a few here. First, we omit an application where we use housing prices to determine the economic border between Brooklyn and Queens boroughs in New York City. The estimated border is substantially different from the existing administrative border, which was determined in 1931 and cannot reflect the dramatic city development. Besides, the estimated border coincides with the Jackson Robinson Parkway and the Long Island Railroad. This finding provides new evidence that local transportation corridors could increase community segregation (cf. Ananat (2011) and Heilmann (2018)). Second, as mentioned in Section 2, we focus on the local constant threshold regression model for computational simplicity. A natural extension is to consider the local linear one by using $\mathbf{1}\left[q_i \leq \gamma_1 + \gamma_2(s - s_i)\right]$ in (3). Although grid search is almost infeasible in determining the two threshold parameters (γ_1 and γ_2), we could use the MCMC algorithm developed by Yu and Fan (2019) and the mixed integer optimization (MIO) algorithms developed by Lee, Liao, Seo, and Shin (2018). Besides the computational challenge, asymptotic derivation in this setup is more involved since we need to consider higher-order expansions of the objective function. Third, our nonparametric setup focuses on the threshold function while some recent literature studies the model

$$y_i = \begin{cases} m_1(x_i) + u_i & \text{if} \quad q_i \leq \gamma_0 \\ m_2(x_i) + u_i & \text{if} \quad q_i > \gamma_0, \end{cases}$$

where $m_1(\cdot)$ and $m_2(\cdot)$ are different nonparametric functions. See, for example, Henderson, Parmeter, and Su (2017), Chiou, Chen, and Chen (2018), Yu and Phillips (2018), and Yu, Liao, and Phillips (2019). One could imagine that the regression function and the threshold function are both nonparametric to allow for more flexible models.

A Appendix

Throughout the proof, we denote $K_i(s) = K((s_i - s)/b_n)$ and $\mathbf{1}_i(\gamma) = \mathbf{1}[q_i \leq \gamma]$. We let $C \in (0, \infty)$ stand for a generic constant term that may vary, which can depend on the location s. We also let $a_n = n^{1-2\epsilon} b_n$. All the additional lemmas in the proof assume the conditions in Assumptions ID and A hold. Omitted proofs for some lemmas are all collected in the supplementary material.

A.1 Proof of Theorem 1

Proof of Theorem 1 We first establish the identification of $\left(\beta_0^\top, \delta_0^\top\right)^\top$ and then the identification of $\gamma_0(s)$ for each $s \in \mathcal{S}$. To this end, we consider two cases separately: (a) $\left(\beta^\top, \delta^\top\right)^\top \neq \left(\beta_0^\top, \delta_0^\top\right)^\top$ and (b) $\left(\beta^\top, \delta^\top\right)^\top = \left(\beta_0^\top, \delta_0^\top\right)^\top$ but $\gamma(s) \neq \gamma_0(s)$.

For case (a), for any $\gamma(s) \in \Gamma$ with given $s \in \mathcal{S}$, we define

$$\begin{aligned} R(\beta, \delta, \gamma; s) &= \mathbb{E}\left[\left(y_i - x_i^\top \beta - x_i^\top \delta \mathbf{1}\left[q_i \leq \gamma(s_i)\right]\right)^2 \middle| s_i = s\right] \\ &\quad - \mathbb{E}\left[\left(y_i - x_i^\top \beta_0 - x_i^\top \delta_0 \mathbf{1}\left[q_i \leq \gamma_0(s_i)\right]\right)^2 \middle| s_i = s\right]. \end{aligned}$$

Then,

$$R(\beta, \delta, \gamma; s) = \begin{cases} \mathbb{E}\left[\left(x_i^\top((\beta + \delta) - (\beta_0 + \delta_0)\right)^2 \middle| s_i = s\right] & \text{on } \{q_i \leq \gamma(s)\} \cap \{q_i \leq \gamma_0(s)\}; \\ \mathbb{E}\left[\left(x_i^\top(\beta - \beta_0)\right)^2 \middle| s_i = s\right] & \text{on } \{q_i > \gamma(s)\} \cap \{q_i > \gamma_0(s)\}. \end{cases}$$

Therefore, by integrating over s_i and Assumption ID-(ii), we have

$$\begin{aligned} & \mathbb{E}\left[R(\beta, \delta, \gamma; s_i)\right] \\ \geq\ & ||(\beta + \delta) - (\beta_0 + \delta_0)||^2 \, \mathbb{E}\left[||x_i x_i^\intercal|| \, \mathbf{1}\left[q_i \leq \underline{\gamma}\right]\right] + ||\beta - \beta_0||^2 \, \mathbb{E}\left[||x_i x_i^\intercal|| \, \mathbf{1}\left[q_i > \overline{\gamma}\right]\right] \\ >\ & 0, \end{aligned}$$

where $\underline{\gamma}$ and $\overline{\gamma}$ denote the lower and upper bounds of Γ, respectively. Therefore, $\left(\beta_0^\top, \delta_0^\top\right)^\top$ are identified as the unique minimizer of $\mathbb{E}[\left(y_i - x_i^\top \beta - x_i^\top \delta \mathbf{1}\left[q_i \leq \gamma\right]\right)^2]$ for any given $\gamma \in \Gamma$.

For case (b), the function $\gamma_0(\cdot)$ is pointwisely identified as the minimizer of

$$\mathbb{E}[\left(y_i - x_i^\top \beta_0 - x_i^\top \delta_0 \mathbf{1}\left[q_i \leq \gamma(s_i)\right]\right)^2 | s_i = s]$$

for each $s \in \mathcal{S}$. This is because for any $\gamma(s) \neq \gamma_0(s)$ at $s_i = s$ and given $(\beta_0^\top, \delta_0^\top)^\top$,

$$\begin{aligned} & R(\beta_0, \delta_0, \gamma; s) \\ =\ & \mathbb{E}\left[\left(y_i - x_i^\top \beta_0 - x_i^\top \delta_0 \mathbf{1}\left[q_i \leq \gamma(s_i)\right]\right)^2 \middle| s_i = s\right] \\ & - \mathbb{E}\left[\left(y_i - x_i^\top \beta_0 - x_i^\top \delta_0 \mathbf{1}\left[q_i \leq \gamma_0(s_i)\right]\right)^2 \middle| s_i = s\right] \\ =\ & \delta_0^\top \mathbb{E}\left[x_i x_i^\top \left(\mathbf{1}\left[q_i \leq \gamma(s_i)\right] - \mathbf{1}\left[q_i \leq \gamma_0(s_i)\right]\right)^2 \middle| s_i = s\right] \delta_0 \end{aligned}$$

$$
\begin{aligned}
&= \delta_0^\top \mathbb{E}\left[x_i x_i^\top \mathbf{1}\left[\min\{\gamma(s_i),\gamma_0(s_i)\} < q_i \le \max\{\gamma(s_i),\gamma_0(s_i)\}\right] \middle| s_i = s\right] \delta_0 \\
&= \int_{\min\{\gamma(s),\gamma_0(s)\}}^{\max\{\gamma(s),\gamma_0(s)\}} \delta_0^\top \mathbb{E}\left[x_i x_i^\top \middle| q_i = q, s_i = s\right] \delta_0 f(q|s) dq \\
&\ge C(s)\mathbb{P}\left(\min\{\gamma(s_i),\gamma_0(s_i)\} < q_i \le \max\{\gamma(s_i),\gamma_0(s_i)\} \middle| s_i = s\right) \\
&> 0
\end{aligned}
$$

from Assumptions ID-(i) and (iii), where $C(s) = \inf_{q\in\mathcal{Q}} \delta_0^\top \mathbb{E}\left[x_i x_i^\top \middle| q_i = q, s_i = s\right] \delta_0 > 0$. Note that the last probability is strictly positive because we assume $f(q|s) > 0$ for any $(q,s) \in \mathcal{Q}\times\mathcal{S}$ and $\gamma_0(s)$ is not located on the boundary of $\mathcal{Q}$ as $\varepsilon(s) < \mathbb{P}(q_i \le \gamma_0(s_i)|s_i = s) < 1-\varepsilon(s)$ for some $\varepsilon(s) > 0$. The identification follows since $R(\beta_0,\delta_0,\gamma;s)$ is continuous at $\gamma = \gamma_0(s)$ from Assumption ID-(v). ■

A.2 Proof of Theorem 2

For a given $s \in \mathcal{S}_0$, we define

$$
\begin{aligned}
M_n(\gamma;s) &= \frac{1}{nb_n}\sum_{i=1}^n x_i x_i^\top \mathbf{1}_i(\gamma) K_i(s), \\
J_n(\gamma;s) &= \frac{1}{\sqrt{nb_n}}\sum_{i=1}^n x_i u_i \mathbf{1}_i(\gamma) K_i(s).
\end{aligned}
$$

Lemma A.1

$$
\begin{aligned}
&\sup_{\gamma\in\Gamma} \|M_n(\gamma;s) - M(\gamma;s)\| \to_p 0, \\
&\sup_{\gamma\in\Gamma} \left\| n^{-1/2} b_n^{-1/2} J_n(\gamma;s) \right\| \to_p 0
\end{aligned}
$$

as $n\to\infty$, where

$$
M(\gamma;s) = \int_{-\infty}^{\gamma} D(q,s) f(q,s)\, dq
$$

and

$$
J_n(\gamma;s) \Rightarrow J(\gamma;s)
$$

a mean-zero Gaussian process indexed by γ.

Proof of Lemma A.1 For expositional simplicity, we only present the case of scalar x_i. We first prove the pointwise convergence of $M_n(\gamma;s)$. By stationarity, Assumptions A-(vii), (x), and Taylor expansion, we have

$$
\begin{aligned}
\mathbb{E}[M_n(\gamma;s)] &= \frac{1}{b_n}\iint \mathbb{E}[x_i^2|q,v]\mathbf{1}[q\le\gamma] K\left(\frac{v-s}{b_n}\right) f(q,v)\, dq dv \qquad \text{(A.1)} \\
&= \iint D(q,s+b_n t)\mathbf{1}[q\le\gamma] K(t) f(q,s+b_n t)\, dq dt \\
&= \int_{-\infty}^{\gamma} D(q,s) f(q,s)\, dq + O\left(b_n^2\right),
\end{aligned}
$$

where $D(q,s)$ is defined in (8). For the variance, we have

$$
\begin{aligned}
Var\left[M_n(\gamma;s)\right] &= \frac{1}{n^2 b_n^2}\mathbb{E}\left[\left(\sum_{i=1}^{n}\left\{x_i^2\mathbf{1}_i(\gamma)K_i(s)-\mathbb{E}\left[x_i^2\mathbf{1}_i(\gamma)K_i(s)\right]\right\}\right)^2\right] \qquad (A.2)\\
&= \frac{1}{nb_n^2}\mathbb{E}\left[\left\{x_i^2\mathbf{1}_i(\gamma)K_i(s)-\mathbb{E}\left[x_i^2\mathbf{1}_i(\gamma)K_i(s)\right]\right\}^2\right]\\
&\quad +\frac{2}{n^2b_n^2}\sum_{i<j}^{n}Cov\left[x_i^2\mathbf{1}_i(\gamma)K_i(s),x_j^2\mathbf{1}_j(\gamma)K_j(s)\right]\\
&= O\left(\frac{1}{nb_n}\right)+O\left(\frac{1}{n}+b_n^2\right)\to 0,
\end{aligned}
$$

where the order of the first term is from the standard kernel estimation result. For the second term, we use Assumptions A-(v), (vii), (x), and Lemma 1 of Bolthausen (1982) to obtain that

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{i<j}^{n}Cov\left[x_i^2\mathbf{1}_i(\gamma)K_i(s),x_j^2\mathbf{1}_j(\gamma)K_j(s)\right]\right| \qquad (A.3)\\
\leq\ & \frac{1}{n}\sum_{i<j}^{n}\left|Cov\left[x_i^2\mathbf{1}_i(\gamma)K\left(\frac{s_i-s}{b_n}\right),x_j^2\mathbf{1}_j(\gamma)K\left(\frac{s_j-s}{b_n}\right)\right]\right|\\
=\ & \frac{b_n^2}{n}\sum_{i<j}^{n}\left|Cov\left[x_i^2\mathbf{1}_i(\gamma)K(t_i),x_j^2\mathbf{1}_j(\gamma)K(t_j)\right]+O\left(b_n^2\right)\right|\\
\leq\ & Cb_n^2\sum_{m=1}^{\infty}m\alpha(m)^{\varphi/(2+\varphi)}\left(\mathbb{E}\left[x_i^{4+2\varphi}\mathbf{1}_i(\gamma)K(t_i)^{2+\varphi}\right]\right)^{2/(2+\varphi)}+O\left(nb_n^4\right)\\
=\ & O\left(b_n^2+nb_n^4\right)
\end{aligned}
$$

for some finite $\varphi>0$, where $\alpha(m)$ is the mixing coefficient defined in (7) and the first equality is by the change of variables $t_i=(s_i-s)/b_n$ in the covariance operator. Hence, the pointwise convergence is established. For given s, the uniform tightness of $M_n(\gamma;s)$ in γ follows similarly as (and even simpler than) that of $J_n(\gamma;s)$ below, and the uniform convergence follows from standard argument. For $J_n(\gamma;s)$, since $\mathbb{E}[u_i x_i|q_i,s_i]=0$, the proof for $\sup_{\gamma\in\Gamma}|(nb_n)^{-1/2}J_n(\gamma,s)|\xrightarrow{p}0$ is identical as $M_n(\gamma;s)$ and hence omitted.

Next, we derive the weak convergence of $J_n(\gamma;s)$. For any fixed s and γ, the Theorem of Bolthausen (1982) implies that $J_n(\gamma;s)\Rightarrow J(\gamma;s)$ under Assumption A-(iii). Because γ is in the indicator function, such pointwise convergence in γ can be generalized into any finite collection of γ to yield the finite dimensional convergence in distribution. By theorem 15.5 of Billingsley (1968), it remains to show that, for each positive $\eta(s)$ and $\varepsilon(s)$ at given s, there exist $\varpi>0$ such that if n is large enough,

$$
\mathbb{P}\left(\sup_{\gamma\in[\gamma_1,\gamma_1+\varpi]}\left|J_n(\gamma;s)-J_n(\gamma_1;s)\right|>\eta(s)\right)\leq\varepsilon(s)\varpi
$$

for any γ_1. To this end, we consider a fine enough grid over $[\gamma_1,\gamma_1+\varpi]$ such that $\gamma_g=$

$\gamma_1+(g-1)\varpi/\overline{g}$ for $g=1,\ldots,\overline{g}+1$, where $nb_n\varpi/2\leq\overline{g}\leq nb_n\varpi$ and $\max_{1\leq g\leq\overline{g}}\left(\gamma_g-\gamma_{g-1}\right)\leq\varpi/\overline{g}$. We define $h_{ig}(s)=x_iu_iK_i\left(s\right)\mathbf{1}\left[\gamma_g<q_i\leq\gamma_{g+1}\right]$ and $H_{ng}(s)=n^{-1}b_n^{-1}\sum_{i=1}^{n}\left|h_{ig}(s)\right|$ for $1\leq g\leq\overline{g}$. Then for any $\gamma\in\left[\gamma_g,\gamma_{g+1}\right]$,

$$
\begin{aligned}
\left|J_n\left(\gamma;s\right)-J_n\left(\gamma_g;s\right)\right| &\leq \sqrt{nb_n}H_{ng}(s)\\
&\leq \sqrt{nb_n}\left|H_{ng}(s)-\mathbb{E}\left[H_{ng}(s)\right]\right|+\sqrt{nb_n}\mathbb{E}\left[H_{ng}(s)\right]
\end{aligned}
$$

and hence

$$
\begin{aligned}
&\sup_{\gamma\in[\gamma_1,\gamma_1+\varpi]}\left|J_n\left(\gamma;s\right)-J_n\left(\gamma_1;s\right)\right|\\
\leq\ &\max_{2\leq g\leq\overline{g}+1}\left|J_n\left(\gamma_g;s\right)-J_n\left(\gamma_1;s\right)\right|\\
&+\max_{1\leq g\leq\overline{g}}\sqrt{nb_n}\left|H_{ng}(s)-\mathbb{E}\left[H_{ng}(s)\right]\right|+\max_{1\leq g\leq\overline{g}}\sqrt{nb_n}\mathbb{E}\left[H_{ng}(s)\right]\\
\equiv\ &\Psi_1(s)+\Psi_2(s)+\Psi_3(s).
\end{aligned}
$$

In what follows, we simply denote $h_i(s)=x_iu_iK_i\left(s\right)\mathbf{1}\left[\gamma_g<q_i\leq\gamma_k\right]$ for any given $1\leq g<k\leq\overline{g}$ and for fixed s. First, for $\Psi_1(s)$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\left|J_n\left(\gamma_g;s\right)-J_n\left(\gamma_k;s\right)\right|^4\right]\\
=\ &\frac{1}{n^2b_n^2}\sum_{i=1}^{n}\mathbb{E}\left[h_i^4(s)\right]+\frac{1}{n^2b_n^2}\sum_{i\neq j}^{n}\mathbb{E}\left[h_i^2(s)h_j^2(s)\right]+\frac{1}{n^2b_n^2}\sum_{i\neq j}^{n}\mathbb{E}\left[h_i^3(s)h_j(s)\right]\\
&+\frac{1}{n^2b_n^2}\sum_{i\neq j\neq k\neq l}^{n}\mathbb{E}\left[h_i(s)h_j(s)h_k(s)h_l(s)\right]+\frac{1}{n^2b_n^2}\sum_{i\neq j\neq k}^{n}\mathbb{E}\left[h_i^2(s)h_j(s)h_k(s)\right]\\
\equiv\ &\Psi_{11}(s)+\Psi_{12}(s)+\Psi_{13}(s)+\Psi_{14}(s)+\Psi_{15}(s),
\end{aligned}
$$

where each term's bound is obtained as follows. For $\Psi_{11}(s)$, a straightforward calculation and Assumptions A-(v) and (x) yield $\Psi_{11}(s)\leq C_1(s)n^{-1}b_n^{-1}+O(b_n/n)=O(n^{-1}b_n^{-1})$ for some constant $0<C_1(s)<\infty$. For $\Psi_{12}(s)$, similarly as (A.3),

$$
\begin{aligned}
\Psi_{12}(s) &\leq \frac{2}{n^2b_n^2}\sum_{i<j}^{n}\left(\mathbb{E}\left[h_i^2(s)\right]\mathbb{E}\left[h_j^2(s)\right]+\left|Cov\left[h_i^2(s),h_j^2(s)\right]\right|\right) \qquad \text{(A.4)}\\
&\leq 2\left(\mathbb{E}\left[\widetilde{h}_i^2\right]\right)^2+\frac{2}{nb_n^2}\left\{Cb_n^2\sum_{m=1}^{\infty}m\alpha\left(m\right)^{\varphi/(2+\varphi)}\left(\mathbb{E}\left[\widetilde{h}_i^{4+2\varphi}\right]\right)^{2/(2+\varphi)}+O\left(nb_n^4\right)\right\}
\end{aligned}
$$

for some $\varphi>0$ that depends on s, where we let $\widetilde{h}_i=x_iu_iK\left(t_i\right)\mathbf{1}\left[\gamma_g<q_i\leq\gamma_k\right]$ from the change of variables $t_i=(s_i-s)/b_n$. Then, by the stationarity, Cauchy-Schwarz inequality, and Lemma 1 of Bolthausen (1982), we have

$$
\Psi_{12}(s)\leq C'\left(\gamma_k-\gamma_g\right)^2+O(n^{-1})+O(b_n^2).
$$

for some constant $0<C'<\infty$. Using the same argument as the second component in (A.4), we can also show that $\Psi_{13}(s) = O(n^{-1}) + O(b_n^2)$. For $\Psi_{14}(s)$, by stationarity,

$$\begin{aligned}
\Psi_{14}(s) &\leq \frac{4!n}{n^2 b_n^2} \sum_{1<i<j<k}^{n} \left| \mathbb{E}\left[h_1(s) h_i(s) h_j(s) h_k(s) \right] \right| \\
&\leq \frac{4!}{n b_n^2} \sum_{i=1}^{n} \sum_{j,k \leq i} \left| Cov\left[h_1(s), h_{i+1}(s) h_{i+j+1}(s) h_{i+j+k+1}(s) \right] \right| \\
&\quad + \frac{4!}{n b_n^2} \sum_{j=1}^{n} \sum_{i,k \leq j} \left| Cov\left[h_1(s) h_{i+1}(s), h_{i+j+1}(s) h_{i+j+k+1}(s) \right] \right| \\
&\quad + \frac{4!}{n b_n^2} \sum_{k=1}^{n} \sum_{i,j \leq k} \left| Cov\left[h_1(s), h_{i+1}(s) h_{i+j+1}(s), h_{i+j+k+1}(s) \right] \right|
\end{aligned} \tag{A.5}$$

similarly as Billingsley (1968), p.173. By Assumptions A-(v), (vii), (x), and Lemma 1 of Bolthausen (1982),

$$\begin{aligned}
& \left| Cov\left[h_1(s), h_{i+1}(s) h_{i+j+1}(s) h_{i+j+k+1}(s) \right] \right| \\
\leq\ & C\alpha(i)^{\varphi/(2+\varphi)} \\
& \times \left(\mathbb{E}\left[h_1(s)^{2+\varphi} \right] \right)^{1/(2+\varphi)} \left(\mathbb{E}\left[\left(h_{i+1}(s) h_{i+j+1}(s) h_{i+j+k+1}(s) \right)^{2+\varphi} \right] \right)^{1/(2+\varphi)} \\
=\ & C\alpha(i)^{\varphi/(2+\varphi)} \\
& \times \left(b_n \left\{ \mathbb{E}\left[\widetilde{h}_1^{2+\varphi} \right] + O\left(b_n^2\right) \right\} \right)^{1/(2+\varphi)} \left(b_n^3 \left\{ \mathbb{E}\left[\left(\widetilde{h}_{i+1} \widetilde{h}_{i+j+1} \widetilde{h}_{i+j+k+1} \right)^{2+\varphi} \right] + O\left(b_n^2\right) \right\} \right)^{1/(2+\varphi)} \\
=\ & C b_n^{4/(2+\varphi)} \alpha(i)^{\varphi/(2+\varphi)} \\
& \times \left\{ \left(\mathbb{E}\left[\widetilde{h}_1^{2+\varphi} \right] \right)^{1/(2+\varphi)} \left(\mathbb{E}\left[\left(\widetilde{h}_{i+1} \widetilde{h}_{i+j+1} \widetilde{h}_{i+j+k+1} \right)^{2+\varphi} \right] \right)^{1/(2+\varphi)} + O\left(b_n^2\right) \right\},
\end{aligned}$$

where the first equality is by the change of variables $t_i = (s_i - s)/b_n$. It follows that the first term in (A.5) satisfies

$$\begin{aligned}
& \frac{4!}{n b_n^2} \sum_{i=1}^{n} \sum_{j,k \leq i} \left| Cov\left[h_1(s), h_{i+1}(s) h_{i+j+1}(s) h_{i+j+k+1}(s) \right] \right| \\
\leq\ & \frac{C4!}{n b_n^{2-(4/(2+\varphi))}} \sum_{i=1}^{\infty} i^2 \alpha(i)^{\varphi/(2+\varphi)} \\
& \times \left\{ \left(\mathbb{E}\left[\widetilde{h}_1^{2+\varphi} \right] \right)^{1/(2+\varphi)} \left(\mathbb{E}\left[\left(\widetilde{h}_{i+1} \widetilde{h}_{i+j+1} \widetilde{h}_{i+j+k+1} \right)^{2+\varphi} \right] \right)^{1/(2+\varphi)} + O\left(b_n^2\right) \right\} \\
=\ & O\left(\frac{1}{n b_n^{2\varphi/(2+\varphi)}} \right) + O\left(\frac{b_n^{4/(2+\varphi)}}{n} \right)
\end{aligned} \tag{A.6}$$

by Assumption A-(iii). However, we select φ small enough such that

$$\frac{2\varphi}{2+\varphi} \leq \frac{1}{1-2\epsilon}, \tag{A.7}$$

which holds for $\varphi \in (0,2)$ in Assumption A-(iii). Then (A.6) becomes $o(1)$ because $nb_n^{2\varphi/(2+\varphi)} = (n^{1-2\epsilon} b_n^{(2\varphi/(2+\varphi))(1-2\epsilon)})^{1/(1-2\epsilon)} \to \infty$ by Assumption A-(ix). Using the same argument, we can also verify that the rest of terms in (A.5) are all $o(1)$ and hence $\Psi_{14}(s) = o(1)$. For $\Psi_{15}(s)$, we can similarly show that it is $o(1)$ as well because

$$\begin{aligned}\Psi_{15}(s) &\leq \frac{3!}{nb_n^2} \sum_{i=1}^{n} \sum_{j \leq i} \left| Cov\left[h_1^2(s), h_{i+1}(s) h_{i+j+1}(s)\right] \right| \\ &\quad + \frac{3!}{nb_n^2} \sum_{j=1}^{n} \sum_{i \leq j} \left| Cov\left[h_1^2(s) h_{i+1}(s), h_{i+j+1}(s)\right] \right|.\end{aligned}$$

By combining these results for $\Psi_{11}(s)$ to $\Psi_{15}(s)$, we thus have

$$\mathbb{E}\left[\left|J_n\left(\gamma_g; s\right) - J_n\left(\gamma_k; s\right)\right|^4\right] \leq C_1(s) \left(\gamma_k - \gamma_g\right)^2$$

for some constant $0 < C_1(s) < \infty$ given s, and Theorem 12.2 of Billingsley (1968) yields

$$\mathbb{P}\left(\max_{1 \leq g \leq \overline{g}} \left|J_n\left(\gamma_g; s\right) - J_n\left(\gamma_1; s\right)\right| > \eta(s)\right) \leq \frac{C_1(s)\varpi^2}{\eta^4(s) b_n}, \tag{A.8}$$

which bounds $\Psi_1(s)$.

To bound $\Psi_2(s)$, the standard result of kernel estimation yields that $\mathbb{E}\left[h_{ik}^2\right] \leq C_2(s) b_n$ by Assumption A-(x) for some constant $0 < C_2(s) < \infty$ given s. Then by Lemma 1 of Bolthausen (1982), we have

$$\begin{aligned}\mathbb{E}\left[\left(\sqrt{nb_n}\left|H_{ng}(s) - \mathbb{E}\left[H_{ng}(s)\right]\right|\right)^2\right] &= \frac{1}{nb_n} Var\left[\sum_{i=1}^{n} \left|h_{ig}(s)\right|\right] \\ &\leq \frac{1}{b_n}\mathbb{E}\left[h_{ig}^2(s)\right] + \frac{2}{nb_n} \sum_{i<j} \left|Cov\left(\left|h_{ig}(s)\right|, \left|h_{jg}(s)\right|\right)\right| \\ &\leq C_2(s)\varpi/\overline{g}\end{aligned}$$

and hence by Markov's inequality,

$$\mathbb{P}\left(\max_{1 \leq g \leq \overline{g}} \sqrt{nb_n}\left|H_{ng}(s) - \mathbb{E}\left[H_{ng}(s)\right]\right| > \eta(s)\right) \leq \frac{C_2(s)\varpi}{\eta^2(s)}. \tag{A.9}$$

Finally, to bound $\Psi_3(s)$, note that

$$\sqrt{nb_n}\mathbb{E}\left[H_{ng}(s)\right] = \sqrt{nb_n} C_3(s) \varpi/\overline{g} \leq 2C_3(s)/\sqrt{nb_n} \tag{A.10}$$

for some constant $0 < C_3(s) < \infty$ given s, where $\varpi/\overline{g} \leq 2/nb_n$. So tightness is proved by combining (A.8), (A.9), and (A.10), and hence the weak convergence follows from Theorem 15.5 of Billingsley (1968). ■

Lemma A.2 *Uniformly over $s \in \mathcal{S}_0$,*

$$\Delta M_n\left(s\right) \equiv \frac{1}{nb_n} \sum_{i=1}^{n} x_i x_i^\top \left\{\mathbf{1}_i\left(\gamma_0\left(s_i\right)\right) - \mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right\} K_i\left(s\right) = O_p\left(b_n\right). \tag{A.11}$$

Lemma A.3 *For a given $s \in \mathcal{S}_0$, $\widehat{\gamma}(s) \to_p \gamma_0(s)$ as $n \to \infty$.*

Proof of Lemma A.3 For given $s \in \mathcal{S}_0$, we let $\widetilde{y}_i(s) = K_i(s)^{1/2} y_i$, $\widetilde{x}_i(s) = K_i(s)^{1/2} x_i$, $\widetilde{u}_i(s) = K_i(s)^{1/2} u_i$, $\widetilde{x}_i(\gamma; s) = K_i(s)^{1/2} x_i \mathbf{1}_i(\gamma)$, and $\widetilde{x}_i(\gamma_0(s_i); s) = K_i(s)^{1/2} x_i \mathbf{1}_i(\gamma_0(s_i))$; we denote $\widetilde{y}(s)$, $\widetilde{X}(s)$, $\widetilde{u}(s)$, $\widetilde{X}(\gamma; s)$, and $\widetilde{X}(\gamma_0(s_i); s)$ as their corresponding matrices of n-stacks. Then $\widehat{\theta}(\gamma; s) = (\widehat{\beta}(\gamma; s)^\top, \widehat{\delta}(\gamma; s)^\top)^\top$ in (2) is given as

$$\widehat{\theta}(\gamma; s) = (\widetilde{Z}(\gamma; s)^\top \widetilde{Z}(\gamma; s))^{-1} \widetilde{Z}(\gamma; s)^\top \widetilde{y}(s), \tag{A.12}$$

where $\widetilde{Z}(\gamma; s) = [\widetilde{X}(s), \widetilde{X}(\gamma; s)]$. Therefore, since $\widetilde{y}(s) = \widetilde{X}(s)\beta_0 + \widetilde{X}(\gamma_0(s_i); s)\delta_0 + \widetilde{u}(s)$ and $\widetilde{X}(s)$ lies in the space spanned by $\widetilde{Z}(\gamma; s)$, we have

$$\begin{aligned}
Q_n(\gamma; s) - \widetilde{u}(s)^\top \widetilde{u}(s) &= \widetilde{y}(s)^\top \left(I_n - P_{\widetilde{Z}}(\gamma; s)\right) \widetilde{y}(s) - \widetilde{u}(s)^\top \widetilde{u}(s) \\
&= -\widetilde{u}(s)^\top P_{\widetilde{Z}}(\gamma; s) \widetilde{u}(s) + 2\delta_0^\top \widetilde{X}(\gamma_0(s_i); s)^\top \left(I_n - P_{\widetilde{Z}}(\gamma; s)\right) \widetilde{u}(s) \\
&\quad + \delta_0^\top \widetilde{X}(\gamma_0(s_i); s)^\top \left(I_n - P_{\widetilde{Z}}(\gamma; s)\right) \widetilde{X}(\gamma_0(s_i); s)\delta_0,
\end{aligned}$$

where $P_{\widetilde{Z}}(\gamma; s) = \widetilde{Z}(\gamma; s)(\widetilde{Z}(\gamma; s)^\top \widetilde{Z}(\gamma; s))^{-1} \widetilde{Z}(\gamma; s)^\top$ and I_n is the identity matrix of rank n. Note that $P_{\widetilde{Z}}(\gamma; s)$ is the same as the projection onto $[\widetilde{X}(s) - \widetilde{X}(\gamma; s), \widetilde{X}(\gamma; s)]$, where $\widetilde{X}(\gamma; s)^\top (\widetilde{X}(s) - \widetilde{X}(\gamma; s)) = 0$. Furthermore, for $\gamma \geq \gamma_0(s_i)$, $\widetilde{x}_i(\gamma_0(s_i); s)^\top (\widetilde{x}_i(s) - \widetilde{x}_i(\gamma; s)) = 0$ and hence $\widetilde{X}(\gamma_0(s_i); s)^\top \widetilde{X}(\gamma; s) = \widetilde{X}(\gamma_0(s_i); s)^\top \widetilde{X}(\gamma_0(s_i); s)$. Since

$$\begin{aligned}
M_n(\gamma; s) &= \frac{1}{nb_n} \sum_{i=1}^n \widetilde{x}_i(\gamma; s) \widetilde{x}_i(\gamma; s)^\top \quad \text{and} \\
J_n(\gamma; s) &= \frac{1}{\sqrt{nb_n}} \sum_{i=1}^n \widetilde{x}_i(\gamma; s) \widetilde{u}_i(s),
\end{aligned}$$

Lemma A.1 yields that

$$\begin{aligned}
\widetilde{Z}(\gamma; s)^\top \widetilde{u}(s) &= [\widetilde{X}(s)^\top \widetilde{u}(s), \widetilde{X}(\gamma; s)^\top \widetilde{u}(s)] = O_p\left((nb_n)^{1/2}\right) \\
\widetilde{Z}(\gamma; s)^\top \widetilde{X}(\gamma_0(s_i); s) &= [\widetilde{X}(s)^\top \widetilde{X}(\gamma_0(s_i); s), \widetilde{X}(\gamma; s)^\top \widetilde{X}(\gamma_0(s_i); s)] \\
&= [\widetilde{X}(s)^\top \widetilde{X}(\gamma_0(s_i); s), \widetilde{X}(\gamma_0(s_i); s)^\top \widetilde{X}(\gamma_0(s_i); s)] = O_p(nb_n)
\end{aligned}$$

for given s. It follows that

$$\begin{aligned}
&\frac{1}{a_n}\left(Q_n(\gamma; s) - \widetilde{u}(s)^\top \widetilde{u}(s)\right) \qquad \text{(A.13)} \\
&= O_p\left(\frac{1}{a_n}\right) + O_p\left(\frac{1}{a_n^{1/2}}\right) + \frac{1}{nb_n} c_0^\top \widetilde{X}(\gamma_0(s_i); s)^\top \left(I_n - P_{\widetilde{Z}}(\gamma; s)\right) \widetilde{X}(\gamma_0(s_i); s) c_0 \\
&= \frac{1}{nb_n} c_0^\top \widetilde{X}(\gamma_0(s_i); s)^\top \left(I - P_{\widetilde{Z}}(\gamma; s)\right) \widetilde{X}(\gamma_0(s_i); s) c_0 + o_p(1)
\end{aligned}$$

for $a_n = n^{1-2\epsilon} b_n \to \infty$ as $n \to \infty$. Moreover, we have

$$M_n(\gamma_0(s_i); s) = \frac{1}{nb_n} \sum_{i=1}^n \widetilde{x}_i(\gamma_0(s_i); s) \widetilde{x}_i(\gamma_0(s_i); s)^\top \tag{A.14}$$

$$
\begin{aligned}
&= M_n(\gamma_0(s); s) + \Delta M_n(s) \\
&= M_n(\gamma_0(s); s) + O_p(b_n)
\end{aligned}
$$

from Lemma A.2, where $\Delta M_n(s)$ is defined in (A.11). It follows that

$$
\begin{aligned}
&\frac{1}{nb_n} c_0^\top \widetilde{X}(\gamma_0(s_i); s)^\top \left(I_n - P_{\widetilde{Z}}(\gamma; s)\right) \widetilde{X}(\gamma_0(s_i); s) c_0 \qquad \text{(A.15)} \\
&\to_p c_0^\top M(\gamma_0(s); s) c_0 - c_0^\top M(\gamma_0(s); s)^\top M(\gamma; s)^{-1} M(\gamma_0(s); s) c_0 \equiv \Upsilon(\gamma; s) < \infty
\end{aligned}
$$

uniformly over $\gamma \in \Gamma \cap [\gamma_0(s), \infty)$, from Lemma A.1 and Assumptions ID-(ii) and A-(viii), as $b_n \to 0$ as $n \to \infty$. However,

$$
d\Upsilon(\gamma; s)/d\gamma = c_0^\top M(\gamma_0(s); s)^\top M(\gamma; s)^{-1} D(\gamma, s) f(\gamma, s) M(\gamma; s)^{-1} M(\gamma_0(s); s) c_0 \geq 0
$$

and

$$
d\Upsilon(\gamma_0(s); s)/d\gamma = c_0^\top D(\gamma_0(s), s) f(\gamma_0(s), s) c_0 > 0
$$

from Assumption A-(viii), which implies that $\Upsilon(\gamma; s)$ is continuous, non-decreasing, and uniquely minimized at $\gamma_0(s)$ given $s \in \mathcal{S}_0$.

We can symmetrically show that the probability limit of (A.15) for $\gamma \in \Gamma \cap (-\infty, \gamma_0(s)]$ is continuous, non-increasing, and uniquely minimized at $\gamma_0(s)$ as well. Therefore, given $s \in \mathcal{S}_0$, uniformly over Γ, the probability limit of $a_n^{-1}\left(Q_n(\gamma; s) - \widetilde{u}(s)^\top \widetilde{u}(s)\right)$ in (A.13) is continuous and uniquely minimized at $\gamma_0(s)$. Since $\widehat{\gamma}(s)$ is the minimizer of $a_n^{-1}\left(Q_n(\gamma; s) - \widetilde{u}(s)^\top \widetilde{u}(s)\right)$, the pointwise consistency follows as the proof of Lemma A.5 of Hansen (2000). ■

We let $\phi_{1n} = a_n^{-1}$, where $a_n = n^{1-2\epsilon} b_n$ and ϵ is given in Assumption A-(ii). For a given $s \in \mathcal{S}_0$, we define

$$
\begin{aligned}
T_n(\gamma; s) &= \frac{1}{nb_n} \sum_{i=1}^n \left(c_0^\top x_i\right)^2 \left|\mathbf{1}_i(\gamma(s)) - \mathbf{1}_i(\gamma_0(s))\right| K_i(s), \\
\overline{T}_n(\gamma, s) &= \frac{1}{nb_n} \sum_{i=1}^n \|x_i\|^2 \left|\mathbf{1}_i(\gamma(s)) - \mathbf{1}_i(\gamma_0(s))\right| K_i(s), \\
L_n(\gamma; s) &= \frac{1}{\sqrt{nb_n}} \sum_{i=1}^n c_0^\top x_i u_i \left\{\mathbf{1}_i(\gamma(s)) - \mathbf{1}_i(\gamma_0(s))\right\} K_i(s) \\
\overline{L}_n(\gamma; s) &= \frac{1}{\sqrt{nb_n}} \sum_{i=1}^n \|x_i u_i\| \left\{\mathbf{1}_i(\gamma(s)) - \mathbf{1}_i(\gamma_0(s))\right\} K_i(s).
\end{aligned}
$$

Lemma A.4 *For a given $s \in \mathcal{S}_0$, for any $\eta(s) > 0$ and $\varepsilon(s) > 0$, there exist constants $0 < C_T(s), C_{\overline{T}}(s), \overline{C}(s), \overline{r}(s) < \infty$ such that for all n,*

$$
\mathbb{P}\left(\inf_{\overline{r}(s)\phi_{1n} < |\gamma(s) - \gamma_0(s)| < \overline{C}(s)} \frac{T_n(\gamma; s)}{|\gamma(s) - \gamma_0(s)|} < C_T(1 - \eta(s))\right) \leq \varepsilon(s), \qquad \text{(A.16)}
$$

$$
\mathbb{P}\left(\sup_{\overline{r}(s)\phi_{1n} < |\gamma(s) - \gamma_0(s)| < \overline{C}(s)} \frac{\overline{T}_n(\gamma; s)}{|\gamma(s) - \gamma_0(s)|} > C_{\overline{T}}(1 + \eta(s))\right) \leq \varepsilon(s), \qquad \text{(A.17)}
$$

$$\mathbb{P}\left(\sup_{\overline{r}(s)\phi_{1n}<|\gamma(s)-\gamma_0(s)|<\overline{C}(s)} \frac{L_n(\gamma;s)}{\sqrt{a_n}\,|\gamma(s)-\gamma_0(s)|} > \eta(s)\right) \leq \varepsilon(s), \tag{A.18}$$

$$\mathbb{P}\left(\sup_{\overline{r}(s)\phi_{1n}<|\gamma(s)-\gamma_0(s)|<\overline{C}(s)} \frac{\overline{L}_n(\gamma;s)}{\sqrt{a_n}\,|\gamma(s)-\gamma_0(s)|} > \eta(s)\right) \leq \varepsilon(s), \tag{A.19}$$

if $n^{1-2\epsilon}b_n^2 \to \varrho < \infty$.

For a given $s \in \mathcal{S}_0$, we let $\widehat{\theta}(\widehat{\gamma}(s)) = (\widehat{\beta}(\widehat{\gamma}(s))^\top, \widehat{\delta}(\widehat{\gamma}(s))^\top)^\top$ and $\theta_0 = (\beta_0^\top, \delta_0^\top)^\top$.

Lemma A.5 *For a given* $s \in \mathcal{S}_0$, $n^\epsilon(\widehat{\theta}(\widehat{\gamma}(s)) - \theta_0) = o_p(1)$.

Proof of Theorem 2 The consistency is proved in Lemma A.3 above. For given $s \in \mathcal{S}_0$, we let

$$\begin{aligned} Q_n^*(\gamma(s);s) &= Q_n(\widehat{\beta}(\widehat{\gamma}(s)), \widehat{\delta}(\widehat{\gamma}(s)), \gamma(s); s) \\ &= \sum_{i=1}^n \left\{ y_i - x_i^\top \widehat{\beta}(\widehat{\gamma}(s)) - x_i^\top \widehat{\delta}(\widehat{\gamma}(s)) \mathbf{1}_i(\gamma(s)) \right\}^2 K_i(s) \end{aligned} \tag{A.20}$$

for any $\gamma(\cdot)$, where $Q_n(\beta,\delta,\gamma;s)$ is the sum of squared errors function in (3). Consider $\gamma(s)$ such that $\gamma(s) \in \left[\gamma_0(s) + \overline{r}(s)\phi_{1n}, \gamma_0(s) + \overline{C}(s)\right]$ for some $0 < \overline{r}(s), \overline{C}(s) < \infty$ that are chosen in Lemma A.4. We let $\Delta_i(\gamma;s) = \mathbf{1}_i(\gamma(s)) - \mathbf{1}_i(\gamma_0(s))$; $\widehat{c}_j(\widehat{\gamma}(s))$ and c_{0j} be the jth element of $\widehat{c}(\widehat{\gamma}(s)) \in \mathbb{R}^p$ and $c_0 \in \mathbb{R}^p$, respectively. Then, since $y_i = \beta_0^\top x_i + \delta_0^\top x_i \mathbf{1}_i(\gamma_0(s_i)) + u_i$,

$$\begin{aligned}
& Q_n^*(\gamma(s);s) - Q_n^*(\gamma_0(s);s) \\
= & \sum_{i=1}^n \left(\widehat{\delta}(\widehat{\gamma}(s))^\top x_i\right)^2 \Delta_i(\gamma;s) K_i(s) \\
& -2\sum_{i=1}^n \left(y_i - \widehat{\beta}(\widehat{\gamma}(s))^\top x_i - \widehat{\delta}(\widehat{\gamma}(s))^\top x_i \mathbf{1}_i(\gamma_0(s))\right)\left(\widehat{\delta}(\widehat{\gamma}(s))^\top x_i\right) \Delta_i(\gamma;s) K_i(s) \\
= & \sum_{i=1}^n \left(\delta_0^\top x_i\right)^2 \Delta_i(\gamma;s) K_i(s) + \sum_{i=1}^n \left\{ \left(\widehat{\delta}(\widehat{\gamma}(s))^\top x_i\right)^2 - \left(\delta_0^\top x_i\right)^2 \right\} \Delta_i(\gamma;s) K_i(s) \\
& -2\sum_{i=1}^n \delta_0^\top x_i u_i \Delta_i(\gamma;s) K_i(s) - 2\sum_{i=1}^n \left(\widehat{\delta}(\widehat{\gamma}(s)) - \delta_0\right)^\top x_i u_i \Delta_i(\gamma;s) K_i(s) \\
& -2\sum_{i=1}^n \left(\widehat{\beta}(\widehat{\gamma}(s)) - \beta_0\right)^\top x_i x_i^\top \widehat{\delta}(\widehat{\gamma}(s)) \Delta_i(\gamma;s) K_i(s) \\
& -2\sum_{i=1}^n \delta_0^\top x_i x_i^\top \delta_0 \left\{\mathbf{1}_i(\gamma_0(s_i)) - \mathbf{1}_i(\gamma_0(s))\right\} \Delta_i(\gamma;s) K_i(s) \qquad \text{(A.21)} \\
& -2\sum_{i=1}^n \delta_0^\top x_i x_i^\top \left(\widehat{\delta}(\widehat{\gamma}(s)) - \delta_0\right) \left\{\mathbf{1}_i(\gamma_0(s_i)) - \mathbf{1}_i(\gamma_0(s))\right\} \Delta_i(\gamma;s) K_i(s) \qquad \text{(A.22)}
\end{aligned}$$

$$-2\sum_{i=1}^{n}\left(\widehat{\delta}\left(\widehat{\gamma}\left(s\right)\right)-\delta_{0}\right)^{\top}x_{i}x_{i}^{\top}\widehat{\delta}\left(\widehat{\gamma}\left(s\right)\right)\mathbf{1}_{i}\left(\gamma_{0}\left(s\right)\right)\Delta_{i}(\gamma;s)K_{i}\left(s\right), \tag{A.23}$$

where the absolute values of the last two summations (A.22) and (A.23) are bounded by

$$\sum_{i=1}^{n}\delta_{0}^{\top}x_{i}x_{i}^{\top}\left(\widehat{\delta}\left(\widehat{\gamma}\left(s\right)\right)-\delta_{0}\right)\left|\Delta_{i}(\gamma;s)\right|K_{i}\left(s\right) \quad \text{and}$$
$$\sum_{i=1}^{n}\left(\widehat{\delta}\left(\widehat{\gamma}\left(s\right)\right)-\delta_{0}\right)^{\top}x_{i}x_{i}^{\top}\widehat{\delta}\left(\widehat{\gamma}\left(s\right)\right)\left|\Delta_{i}(\gamma;s)\right|K_{i}\left(s\right),$$

respectively, since $|\mathbf{1}_i(\gamma_0(s))| \leq 1$ and $|\mathbf{1}_i(\gamma_0(s_i)) - \mathbf{1}_i(\gamma_0(s))| \leq 1$. Moreover, for the term in (A.21), we have

$$\begin{aligned}&\frac{1}{a_n}\sum_{i=1}^{n}\delta_{0}^{\top}x_{i}x_{i}^{\top}\delta_{0}\left\{\mathbf{1}_{i}\left(\gamma_{0}\left(s_{i}\right)\right)-\mathbf{1}_{i}\left(\gamma_{0}\left(s\right)\right)\right\}\Delta_{i}(\gamma;s)K_{i}\left(s\right)\\ \leq\;&\frac{1}{a_n}\sum_{i=1}^{n}\delta_{0}^{\top}x_{i}x_{i}^{\top}\delta_{0}\left|\mathbf{1}_{i}\left(\gamma_{0}\left(s_{i}\right)\right)-\mathbf{1}_{i}\left(\gamma_{0}\left(s\right)\right)\right|K_{i}\left(s\right)=C^{*}(s)b_{n}\end{aligned}$$

for some $C^*(s) = O_p(1)$ as in (A.14). It follows that

$$\begin{aligned}&\frac{Q_n^*(\gamma(s);s)-Q_n^*(\gamma_0(s);s)}{a_n(\gamma(s)-\gamma_0(s))} \qquad \text{(A.24)}\\ \geq\;&\frac{T_n\left(\gamma;s\right)}{\gamma(s)-\gamma_0(s)}-\left\|\widehat{c}\left(\widehat{\gamma}\left(s\right)\right)-c_0\right\|\left\|\widehat{c}\left(\widehat{\gamma}\left(s\right)\right)+c_0\right\|\frac{\overline{T}_n(\gamma,s)}{\gamma(s)-\gamma_0(s)}\\ &-2\frac{L_n\left(\gamma;s\right)}{\sqrt{a_n}(\gamma(s)-\gamma_0(s))}-2\max_{1\leq j\leq p}\left|\widehat{c}_j(\widehat{\gamma}\left(s\right))-c_{0j}\right|\frac{\overline{L}_n\left(\gamma;s\right)}{\sqrt{a_n}\left(\gamma(s)-\gamma_0(s)\right)}\\ &-2\left\|n^{\epsilon}(\widehat{\beta}\left(\widehat{\gamma}\left(s\right)\right)-\beta_0)\right\|\left\|\widehat{c}(\widehat{\gamma}\left(s\right))\right\|\frac{\overline{T}_n(\gamma,s)}{\gamma(s)-\gamma_0(s)}\\ &-2\frac{C^*(s)b_n}{\gamma(s)-\gamma_0(s)}\\ &-2\left\|c_0\right\|\left\|\widehat{c}\left(\widehat{\gamma}\left(s\right)\right)-c_0\right\|\frac{\overline{T}_n(\gamma,s)}{\gamma(s)-\gamma_0(s)}\\ &-2\left\|n^{\epsilon}(\widehat{\delta}\left(\widehat{\gamma}\left(s\right)\right)-\delta_0)\right\|\left\|\widehat{c}(\widehat{\gamma}\left(s\right))\right\|\frac{\overline{T}_n(\gamma,s)}{\gamma(s)-\gamma_0(s)}\\ =\;&\frac{T_n\left(\gamma;s\right)}{\gamma(s)-\gamma_0(s)}-\frac{2L_n\left(\gamma;s\right)}{\sqrt{a_n}\left(\gamma(s)-\gamma_0(s)\right)}-\frac{2C^*(s)b_n}{\gamma(s)-\gamma_0(s)}+o_p(1),\end{aligned}$$

where the last line follows from Lemma A.5. Then given Lemma A.4 and the Markov's inequality, there exist $0 < C(s), \overline{C}(s), \overline{r}(s), \eta(s), \varepsilon(s) < \infty$ such that

$$\mathbb{P}\left(\inf_{\overline{r}(s)\phi_{1n}<|\gamma(s)-\gamma_0(s)|<\overline{C}(s)}\frac{T_n\left(\gamma;s\right)}{\left|\gamma\left(s\right)-\gamma_0\left(s\right)\right|}<\left(1-\eta(s)\right)C(s)\right) \;\leq\; \frac{\varepsilon(s)}{3},$$

$$\mathbb{P}\left(\sup_{\overline{r}(s)\phi_{1n}<|\gamma(s)-\gamma_0(s)|<\overline{C}(s)} \frac{2L_n(\gamma;s)}{\sqrt{a_n}\,|\gamma(s)-\gamma_0(s)|} > \eta(s)\right) \leq \frac{\varepsilon(s)}{3}.$$

In addition, for $\gamma(s) \in \left[\gamma_0(s) + \overline{r}(s)\phi_{1n}, \gamma_0(s) + \overline{C}(s)\right]$, since

$$\sup_{\overline{r}(s)\phi_{1n}<|\gamma(s)-\gamma_0(s)|<\overline{C}(s)} \frac{C^*(s)b_n}{\gamma(s)-\gamma_0(s)} < \frac{C^*(s)b_n}{\overline{r}(s)\phi_{1n}} = a_n b_n \frac{C^*(s)}{\overline{r}(s)} < \infty$$

provided $n^{1-2\epsilon}b_n^2 \to \varrho < \infty$, we also have

$$\mathbb{P}\left(\sup_{\overline{r}(s)\phi_{1n}<|\gamma(s)-\gamma_0(s)|<\overline{C}(s)} \frac{2C^*(s)b_n}{|\gamma(s)-\gamma_0(s)|} > \eta(s)\right) \leq \frac{\varepsilon(s)}{3}$$

by choosing $\overline{r}(s)$ large enough. Thus for any $\varepsilon(s) > 0$ and $\eta(s) > 0$, we have

$$\mathbb{P}\left(\inf_{\overline{r}(s)\phi_{1n}<|\gamma(s)-\gamma_0(s)|<\overline{C}(s)} \{Q_n^*(\gamma(s);s) - Q_n^*(\gamma_0(s);s)\} > \eta(s)\right) \geq 1 - \varepsilon(s),$$

which yields $\mathbb{P}\left(Q_n^*(\gamma(s);s) - Q_n^*(\gamma_0(s);s) > 0\right) \to 1$ as $n \to \infty$. We can similarly show the same result when $\gamma(s) \in \left[\gamma_0(s) - \overline{C}(s), \gamma_0(s) - \overline{r}(s)\phi_{1n}\right]$. Therefore, with probability approaching to one, it should hold that $|\widehat{\gamma}(s) - \gamma_0(s)| \leq r(s)\phi_{1n}$ since $Q_n^*(\widehat{\gamma}(s);s) - Q_n^*(\gamma_0(s);s) \leq 0$ for any $s \in \mathcal{S}_0$ by construction. ■

A.3 Proof of Theorem 3 and Corollary 1

For a given $s \in \mathcal{S}_0$, we let $\gamma_n(s) = \gamma_0(s) + r/a_n$ with some $|r| < \infty$, where $a_n = n^{1-2\epsilon}b_n$ and ϵ is given in Assumption A-(ii). We define

$$\begin{aligned} A_n^*(r,s) &= \sum_{i=1}^n \left(\delta_0^\top x_i\right)^2 |\mathbf{1}_i(\gamma_n(s)) - \mathbf{1}_i(\gamma_0(s))| K_i(s), \\ B_n^*(r,s) &= \sum_{i=1}^n \delta_0^\top x_i u_i \{\mathbf{1}_i(\gamma_n(s)) - \mathbf{1}_i(\gamma_0(s))\} K_i(s). \end{aligned}$$

Lemma A.6 *If $n^{1-2\epsilon}b_n^2 \to \varrho < \infty$,*

$$A_n^*(r,s) \to_p |r|\, c_0^\top D(\gamma_0(s), s)\, c_0 f(\gamma_0(s), s)$$

and

$$B_n^*(r,s) \Rightarrow W(r)\sqrt{c_0^\top V(\gamma_0(s), s)\, c_0 f(\gamma_0(s), s)\,\kappa_2}$$

as $n \to \infty$, where $\kappa_2 = \int K(v)^2 dv$ and $W(r)$ is the two-sided Brownian Motion defined in (10).

Proof of Lemma A.6 Let $\Delta_i(\gamma_n;s) = \mathbf{1}_i(\gamma_n(s)) - \mathbf{1}_i(\gamma_0(s))$. First, for $A_n^*(r,s)$, consider the case with $r > 0$. Note that $\delta_0 = c_0 n^{-\epsilon} = c_0(a_n/(nb_n))^{1/2}$. By change of variables and Taylor expansion, Assumptions A-(v), (viii), and (x) imply that

$$\begin{aligned}
\mathbb{E}\left[A_n^*(r,s)\right] &= \frac{a_n}{nb_n}\sum_{i=1}^{n}\mathbb{E}\left[\left(c_0^\top x_i\right)^2\Delta_i(\gamma_n;s)K_i(s)\right] \qquad \text{(A.25)}\\
&= a_n\iint_{\gamma_0(s)}^{\gamma_0(s)+r/a_n}\mathbb{E}\left[\left(c_0^\top x_i\right)^2|q,s+b_nt\right]K(t)\,f(q,s+b_nt)\,dqdt\\
&= rc_0^\top D(\gamma_0(s),s)\,c_0 f(\gamma_0(s),s)+O\left(\frac{1}{a_n}+b_n^2\right),
\end{aligned}$$

where the third equality holds under Assumption A-(vi). Next, we have

$$\begin{aligned}
Var\left[A_n^*(r,s)\right] &= \frac{a_n^2}{n^2b_n^2}Var\left[\sum_{i=1}^{n}\left(c_0^\top x_i\right)^2\Delta_i(\gamma_n;s)K_i(s)\right] \qquad \text{(A.26)}\\
&= \frac{a_n^2}{nb_n^2}Var\left[\left(c_0^\top x_i\right)^2\Delta_i(\gamma_n;s)K_i(s)\right]\\
&\quad+\frac{2a_n^2}{n^2b_n^2}\sum_{i<j}^{n}Cov\left[\left(c_0^\top x_i\right)^2\Delta_i(\gamma_n;s)K_i(s),\left(c_0^\top x_j\right)^2\Delta_j(\gamma_n;s)K_j(s)\right]\\
&\equiv \Psi_{A1}(r,s)+\Psi_{A2}(r,s).
\end{aligned}$$

Similarly as (A.25), Taylor expansion and Assumptions A-(vii), (viii), and (x) lead to

$$\begin{aligned}
\Psi_{A1}(r,s) &= \frac{a_n}{nb_n}\left(\frac{a_n}{b_n}\mathbb{E}\left[\left(c_0^\top x_i\right)^4\Delta_i(\gamma_n;s)K_i^2(s)\right]\right)\\
&\quad-\frac{1}{n}\left(\frac{a_n}{b_n}\mathbb{E}\left[\left(c_0^\top x_i\right)^2\Delta_i(\gamma_n;s)K_i(s)\right]\right)^2\\
&= O\left(n^{-2\epsilon}+\frac{1}{n}\right)
\end{aligned}$$

since $\{\Delta_i(\gamma_n;s)\}^2=\Delta_i(\gamma_n;s)$ for $r>0$. Furthermore, by change of variables $t_i=(s_i-s)/b_n$ in the covariance operator and Lemma 1 of Bolthausen (1982),

$$\begin{aligned}
\Psi_{A2}(r,s) &\le \frac{2a_n^2}{n^2}\sum_{i<j}^{n}Cov\left[\left(c_0^\top x_i\right)^2\Delta_i(\gamma_n;s)K(t_i),\left(c_0^\top x_j\right)^2\Delta_j(\gamma_n;s)K(t_j)\right]\\
&\le \frac{2a_n^2}{n}\sum_{m=1}^{\infty}m\alpha(m)^{\varphi/(2+\varphi)}\left(\mathbb{E}\left[\left|\left(c_0^\top x_i\right)^2\Delta_i(\gamma_n;s)K(t_i)\right|^{2+\varphi}\right]\right)^{2/(2+\varphi)}\\
&= O(a_n^{2-2/(2+\varphi)}n^{-1})=O(n^{-2\epsilon}),
\end{aligned}$$

where the last line follows from the conditions that $\varphi\in(0,2)$ in Assumption A-(iii) and $n^{1-2\epsilon}b_n^2\to\varrho<\infty$. Hence, the pointwise convergence of $A_n^*(r,s)$ is obtained. Since $rc_0^\top D(\gamma_0(s),s)\,c_0f(\gamma_0(s),s)$ is strictly increasing and continuous in r, the convergence holds uniformly on any compact set. Symmetrically, we can show that $\mathbb{E}[A_n^*(r,s)] = -rc_0^\top D(\gamma_0(s),s)\,c_0f(\gamma_0(s),s)+O\left(a_n^{-1}+b_n^2\right)$ when $r<0$. The uniform convergence also holds in this case using the same argument as above, which completes the proof for $A_n^*(r,s)$.

For $B_n^*(r,s)$, Assumption ID-(i) leads to $\mathbb{E}[B_n^*(r,s)] = 0$. Then, similarly as for $A_n^*(r,s)$, for any $i \neq j$, we have

$$Cov\left[c_0^\top x_i u_i \Delta_i(\gamma_n; s) K_i(s), c_0^\top x_j u_j \Delta_j(\gamma_n; s) K_j(s)\right] \leq C b_n^2 a_n^{-1} \tag{A.27}$$

for some positive constant $C < \infty$, by the change of variables in the covariance operator and Lemma 1 of Bolthausen (1982). It follows that, similarly as (A.25),

$$\begin{aligned} Var[B_n^*(r,s)] &= \frac{a_n}{b_n} Var\left[c_0^\top x_i u_i \Delta_i(\gamma_n; s) K_i(s)\right] + O(b_n) \\ &= |r|\, c_0^\top V(\gamma_0(s), s)\, c_0 f(\gamma_0(s), s)\, \kappa_2 + o(1), \end{aligned}$$

where $\kappa_2 = \int K(v)^2 dv$. Then by the CLT for stationary and mixing random field (e.g. Bolthausen (1982); Jenish and Prucha (2009)), we have

$$B_n^*(r,s) \Rightarrow W(r) \sqrt{c_0^\top V(\gamma_0(s), s)\, c_0 f(\gamma_0(s), s)\, \kappa_2}$$

as $n \to \infty$, where $W(r)$ is the two-sided Brownian Motion defined in (10). This pointwise convergence in r can be extended to any finite-dimensional convergence in r by the fact that for any $r_1 < r_2$, $Cov[B_n^*(r_1,s), B_n^*(r_2,s)] = Var[B_n^*(r_1,s)] + o(1)$, which is because $(\mathbf{1}_i(\gamma_0 + r_2/a_n) - \mathbf{1}_i(\gamma_0 + r_1/a_n))\, \mathbf{1}_i(\gamma_0 + r_1/a_n) = 0$ and (A.27). The tightness follows from a similar argument as $J_n(\gamma; s)$ in Lemma A.1 and the desired result follows by Theorem 15.5 in Billingsley (1968). ■

For a given $s \in \mathcal{S}_0$, we let $\widehat{\theta}(\gamma_0(s)) = (\widehat{\beta}(\gamma_0(s))^\top, \widehat{\delta}(\gamma_0(s))^\top)^\top$. Recall that $\theta_0 = (\beta_0^\top, \delta_0^\top)^\top$ and $\widehat{\theta}(\widehat{\gamma}(s)) = (\widehat{\beta}(\widehat{\gamma}(s))^\top, \widehat{\delta}(\widehat{\gamma}(s))^\top)^\top$.

Lemma A.7 *For a given $s \in \mathcal{S}_0$, $\sqrt{nb_n}(\widehat{\theta}(\widehat{\gamma}(s)) - \theta_0) = O_p(1)$, if $n^{1-2\epsilon} b_n^2 \to \varrho < \infty$ as $n \to \infty$.*

Proof of Theorem 3 From Theorem 2, we define a random variable $r^*(s)$ such that

$$r^*(s) = a_n(\widehat{\gamma}(s) - \gamma_0(s)) = \arg\max_{r \in \mathbb{R}} \left\{ Q_n^*(\gamma_0(s); s) - Q_n^*\left(\gamma_0(s) + \frac{r}{a_n}; s\right)\right\},$$

where $Q_n^*(\gamma(s); s)$ is defined in (A.20). We let $\Delta_i(s) = \mathbf{1}_i(\gamma_0(s) + (r/a_n)) - \mathbf{1}_i(\gamma_0(s))$. We then have

$$\begin{aligned} &\Delta Q_n^*(r; s) \\ &= Q_n^*(\gamma_0(s); s) - Q_n^*\left(\gamma_0(s) + \frac{r}{a_n}; s\right) \\ &= -\sum_{i=1}^n \left(\widehat{\delta}(\widehat{\gamma}(s))^\top x_i\right)^2 |\Delta_i(s)|\, K_i(s) \\ &\quad +2\sum_{i=1}^n \left(y_i - \widehat{\beta}(\widehat{\gamma}(s))^\top x_i - \widehat{\delta}(\widehat{\gamma}(s))^\top x_i \mathbf{1}_i(\gamma_0(s))\right)\left(\widehat{\delta}(\widehat{\gamma}(s))^\top x_i\right) \Delta_i(s) K_i(s) \end{aligned} \tag{A.28}$$

$$\equiv \quad -A_n(r;s) + 2B_n(r;s).$$

For $A_n(r;s)$, Lemmas A.6 and A.7 yield

$$\begin{aligned}
& A_n(r;s) && \text{(A.29)}\\
= \quad & \sum_{i=1}^{n} \left(\left(\delta_0 + n^{-1/2} b_n^{-1/2} C_\delta(s) + o_p(n^{-1/2} b_n^{-1/2}) \right)^\top x_i \right)^2 |\Delta_i(s)| K_i(s) \\
= \quad & A_n^*(r,s) + \frac{1}{n^{1-2\epsilon} b_n} \sum_{i=1}^{n} \left(n^{-\epsilon} C_\delta(s) \right)^\top x_i x_i^\top \left(n^{-\epsilon} C_\delta(s) \right) |\Delta_i(s)| K_i(s) + o_p\left(a_n^{-1} \right) \\
= \quad & A_n^*(r,s) + O_p(a_n^{-1})
\end{aligned}$$

for some $p \times 1$ vector $C_\delta(s) = O_p(1)$, since $\sum_{i=1}^{n} n^{-2\epsilon} C_\delta^\top(s) x_i x_i^\top C_\delta(s) |\Delta_i(s)| K_i(s) = O_p(1)$ from Lemma A.6 and $a_n = n^{1-2\epsilon} b_n \to \infty$. Note that $\widehat{\delta}(\widehat{\gamma}(s)) - \delta_0 = O_p((nb_n)^{-1/2})$ from Lemma A.7. Similarly, for $B_n(r;s)$, since $y_i = \beta_0^\top x_i + \delta_0^\top x_i \mathbf{1}_i(\gamma_0(s_i)) + u_i$, we have for some $p \times 1$ vector $C_\beta(s) = O_p(1)$,

$$\begin{aligned}
& B_n(r;s) && \text{(A.30)}\\
= \quad & \sum_{i=1}^{n} \Bigg(u_i + \delta_0^\top x_i \left\{ \mathbf{1}_i(\gamma_0(s_i)) - \mathbf{1}_i(\gamma_0(s)) \right\} - \left(\widehat{\beta}(\widehat{\gamma}(s)) - \beta_0 \right)^\top x_i \\
& \qquad - \left(\widehat{\delta}(\widehat{\gamma}(s)) - \delta_0 \right)^\top x_i \mathbf{1}_i(\gamma_0(s)) \Bigg) \widehat{\delta}(\gamma_0(s))^\top x_i \Delta_i(s) K_i(s) \\
= \quad & \sum_{i=1}^{n} \Big(u_i + \delta_0^\top x_i \left\{ \mathbf{1}_i(\gamma_0(s_i)) - \mathbf{1}_i(\gamma_0(s)) \right\} - n^{-1/2} b_n^{-1/2} C_\beta^\top(s) x_i \\
& \qquad - n^{-1/2} b_n^{-1/2} C_\delta^\top(s) x_i \mathbf{1}_i(\gamma_0(s)) \Big) \left(\delta_0 + n^{-1/2} b_n^{-1/2} C_\delta(s) \right)^\top x_i \Delta_i(s) K_i(s) + o_p(1) \\
= \quad & B_n^*(r,s) \\
& + \frac{1}{\sqrt{n^{1-2\epsilon} b_n}} \sum_{i=1}^{n} u_i x_i \left(n^{-\epsilon} C_\delta(s) \right) \Delta_i(s) K_i(s) \\
& + \sum_{i=1}^{n} \delta_0^\top x_i x_i^\top \delta_0 \left(\Delta_i(s) \left\{ \mathbf{1}_i(\gamma_0(s_i)) - \mathbf{1}_i(\gamma_0(s)) \right\} \right) K_i(s) && \text{(A.31)}\\
& + \frac{1}{\sqrt{n^{1-2\epsilon} b_n}} \sum_{i=1}^{n} \delta_0^\top x_i x_i^\top \left(n^{-\epsilon} C_\delta(s) \right) \left(\Delta_i(s) \left\{ \mathbf{1}_i(\gamma_0(s_i)) - \mathbf{1}_i(\gamma_0(s)) \right\} \right) K_i(s) \\
& + \frac{1}{\sqrt{n^{1-2\epsilon} b_n}} \sum_{i=1}^{n} \delta_0^\top x_i x_i^\top \left(n^{-\epsilon} C_\beta(s) \right) \Delta_i(s) K_i(s) \\
& + \frac{1}{n^{1-2\epsilon} b_n} \sum_{i=1}^{n} \left(n^{-\epsilon} C_\beta(s) \right)^\top x_i x_i^\top \left(n^{-\epsilon} C_\delta(s) \right) \Delta_i(s) K_i(s) \\
& + \frac{1}{\sqrt{n^{1-2\epsilon} b_n}} \sum_{i=1}^{n} \delta_0^\top x_i x_i^\top \left(n^{-\epsilon} C_\delta(s) \right) \left\{ \Delta_i(s) \mathbf{1}_i(\gamma_0(s)) \right\} K_i(s)
\end{aligned}$$

$$+\frac{1}{n^{1-2\epsilon}b_n}\sum_{i=1}^{n}\left(n^{-\epsilon}C_\delta(s)\right)^\top x_i x_i^\top\left(n^{-\epsilon}C_\delta(s)\right)\{\Delta_i(s)\mathbf{1}_i(\gamma_0(s))\}K_i(s)$$
$$+o_p(\left(n^{1-2\epsilon}b_n\right)^{-1/2}),$$

where all the terms are $O_p\left(\left(n^{1-2\epsilon}b_n\right)^{-1/2}\right)=O_p(a_n^{-1/2})$ except for the first term $B_n^*(r,s)$ and the third term in the line of (A.31) that we denote $B_{n3}^*(r,s)$. In Lemma A.8 below, we show that, if $n^{1-2\epsilon}b_n^2\to\varrho\in(0,\infty)$,

$$B_{n3}^*(r,s)\to_p |r|\,c_0^\top D(\gamma_0(s),s)\,c_0 f(\gamma_0(s),s)\left\{\frac{1}{2}-\mathcal{K}_0(r,\varrho;s)\right\}$$
$$+\varrho c_0^\top D(\gamma_0(s),s)\,c_0 f(\gamma_0(s),s)\,|\dot\gamma_0(s)|\,\mathcal{K}_1(r,\varrho;s)$$

as $n\to\infty$, where $\dot\gamma_0(\cdot)$ is the first derivatives of $\gamma_0(\cdot)$ and $\mathcal{K}_j(r,\varrho;s)=\int_0^{|r|/(\varrho|\dot\gamma_0(s)|)}t^j K(t)\,dt$ for $j=0,1$.

From Lemma A.6, it follows that

$$\begin{aligned}
\Delta Q_n^*(r;s) &= -A_n^*(r,s)+2B_{n3}^*(r,s)+2B_n^*(r,s)\\
&= -|r|\,c_0^\top D(\gamma_0(s),s)\,c_0 f(\gamma_0(s),s)\\
&\quad +|r|\,c_0^\top D(\gamma_0(s),s)\,c_0 f(\gamma_0(s),s)\{1-2\mathcal{K}_0(r,\varrho;s)\}\\
&\quad +2\varrho c_0^\top D(\gamma_0(s),s)\,c_0 f(\gamma_0(s),s)\,|\dot\gamma_0(s)|\,\mathcal{K}_1(r,\varrho;s)\\
&\quad +2W(r)\sqrt{c_0^\top V(\gamma_0(s),s)\,c_0 f(\gamma_0(s),s)\,\kappa_2}+O_p(a_n^{-1/2}+b_n),\\
&= -2|r|\,\ell_D(s)\widetilde\psi_1(r,\varrho;s)+2\varrho\ell_D(s)\widetilde\psi_2(r,\varrho;s)\\
&\quad +2W(r)\sqrt{\ell_V(s)}+O_p(a_n^{-1/2}+b_n),
\end{aligned}$$

where

$$\begin{aligned}
\ell_D(s) &= c_0^\top D(\gamma_0(s),s)\,c_0 f(\gamma_0(s),s),\\
\ell_V(s) &= c_0^\top V(\gamma_0(s),s)\,c_0 f(\gamma_0(s),s)\,\kappa_2,\\
\widetilde\psi_1(r,\varrho;s) &= \mathcal{K}_0(r,\varrho;s),\\
\widetilde\psi_2(r,\varrho;s) &= |\dot\gamma_0(s)|\,\mathcal{K}_1(r,\varrho;s).
\end{aligned}$$

However, if we let $\xi(s)=\ell_V(s)/\ell_D^2(s)$ and $r=\xi(s)\nu$, we have

$$\begin{aligned}
&\arg\max_{r\in\mathbb{R}}\left(2W(r)\sqrt{\ell_V(s)}-2|r|\,\ell_D(s)\widetilde\psi_1(r,\varrho;s)+2\varrho\ell_D(s)\widetilde\psi_2(r,\varrho;s)\right)\\
=\;&\xi(s)\arg\max_{\nu\in\mathbb{R}}\left(W(\xi(s)\nu)\sqrt{\ell_V(s)}-|\xi(s)\nu|\,\ell_D(s)\widetilde\psi_1(\xi(s)\nu,\varrho;s)+\varrho\ell_D(s)\widetilde\psi_2(\nu,\varrho;s)\right)\\
=\;&\xi(s)\arg\max_{\nu\in\mathbb{R}}\left(W(\nu)\frac{\ell_V(s)}{\ell_D(s)}-|\nu|\frac{\ell_V(s)}{\ell_D(s)}\widetilde\psi_1(\xi(s)\nu,\varrho;s)+\varrho\frac{\ell_V(s)}{\ell_D(s)}\xi(s)\widetilde\psi_2(\xi(s)\nu,\varrho;s)\right)\\
=\;&\xi(s)\arg\max_{\nu\in\mathbb{R}}\left(W(\nu)-|\nu|\,\widetilde\psi_1(\xi(s)\nu,\varrho;s)+\varrho\xi(s)\widetilde\psi_2(\xi(s)\nu,\varrho;s)\right)
\end{aligned}$$

similar to the proof of Theorem 1 in Hansen (2000). By Theorem 2.7 of Kim and Pollard

(1990), it follows that (rewriting ν as r)

$$n^{1-2\epsilon} b_n \left(\widehat{\gamma}(s) - \gamma_0(s)\right) \to_d \xi(s) \arg\max_{r\in\mathbb{R}} \left(W(r) - |r| \psi_1(r, \varrho; s) + \varrho \psi_2(r, \varrho; s)\right)$$

as $n \to \infty$, where

$$\begin{aligned}\psi_1(r, \varrho; s) &= \widetilde{\psi}_1(\xi(s) r, \varrho; s) = \int_0^{\xi(s)|r|/(\varrho|\dot{\gamma}_0(s)|)} K(t)\, dt,\\ \psi_2(r, \varrho; s) &= \xi(s)\widetilde{\psi}_2(\xi(s) r, \varrho; s) = \xi(s) |\dot{\gamma}_0(s)| \int_0^{\xi(s)|r|/(\varrho|\dot{\gamma}_0(s)|)} t K(t)\, dt.\end{aligned}$$

Note that when $\varrho = 0$, we let $\psi_1(r, 0; s) = \int_0^\infty K(t)\, dt = 1/2$. Finally, letting

$$\mu(r, \varrho; s) = -|r| \psi_1(r, \varrho; s) + \varrho \psi_2(r, \varrho; s), \tag{A.32}$$

$\mathbb{E}\left[\arg\max_{r\in\mathbb{R}} \left(W(r) + \mu(r, \varrho; s)\right)\right] = 0$ follows from Lemmas A.9 and A.10 below. ■

Lemma A.8 *For a given $s \in \mathcal{S}_0$, let r be the same term used in Lemma A.6. If $n^{1-2\epsilon} b_n^2 \to \varrho \in (0, \infty)$,*

$$\begin{aligned}B_{n3}^*(r, s) &\equiv \sum_{i=1}^n \left(\delta_0^\top x_i\right)^2 \left\{\mathbf{1}_i(\gamma_0(s) + r/a_n) - \mathbf{1}_i(\gamma_0(s))\right\} \left\{\mathbf{1}_i(\gamma_0(s_i)) - \mathbf{1}_i(\gamma_0(s))\right\} K_i(s)\\ &\to_p |r| c_0^\top D(\gamma_0(s), s) c_0 f(\gamma_0(s), s) \left\{\frac{1}{2} - \mathcal{K}_0(r, \varrho; s)\right\}\\ &\quad + \varrho c_0^\top D(\gamma_0(s), s) c_0 f(\gamma_0(s), s) |\dot{\gamma}_0(s)| \mathcal{K}_1(r, \varrho; s)\end{aligned}$$

as $n \to \infty$, where $\dot{\gamma}_0(\cdot)$ is the first derivatives of $\gamma_0(\cdot)$ and

$$\mathcal{K}_j(r, \varrho; s) = \int_0^{|r|/(\varrho|\dot{\gamma}_0(s)|)} t^j K(t)\, dt$$

for $j = 0, 1$.

Lemma A.9 *Let $\tau = \arg\max_{r\in\mathbb{R}} (W(r) + \mu(r))$, where $W(r)$ is a two-sided Brownian motion in (10) and $\mu(r)$ is a continuous drift function satisfying: $\mu(0) = 0$, $\mu(-r) = \mu(r)$, $\mu(r)$ is monotonically decreasing on $\mathbb{R}\backslash[-\underline{r}, \underline{r}]$ for some $\underline{r} > 0$, and $\lim_{|r|\to\infty} |r|^{-((1/2)+\varepsilon)} \mu(r) = -\infty$ for some $\varepsilon > 0$. Then, $\mathbb{E}[\tau] = 0$.*

Lemma A.10 *For given (ϱ, s), $\mu(r, \varrho; s)$ in (A.32) satisfies conditions in Lemma A.9*

Proof of Corollary 1 From (A.13) and (A.15), we have

$$\frac{1}{nb_n} Q_n(\widehat{\gamma}(s), s) = \frac{1}{nb_n} \sum_{i=1}^n u_i^2 K_i(s) + o_p(1) \to_p \mathbb{E}\left[u_i^2 | s_i = s\right] f_s(s),$$

where $f_s(s)$ is the marginal density of s_i. In addition, from Theorem 3 and the proof of Lemma A.7, we have

$$Q_n(\gamma_0(s), s) - Q_n(\widehat{\gamma}(s), s) = Q_n^*(\gamma_0(s), s) - Q_n^*(\widehat{\gamma}(s), s) + o_p(1)$$

since $\widehat{\theta}\left(\widehat{\gamma}\left(s\right)\right)-\widehat{\theta}\left(\gamma_0\left(s\right)\right)=o_p((nb_n)^{-1/2})$. Similar to Theorem 2 of Hansen (2000), the rest of the proof follows from the change of variables and the continuous mapping theorem because $(nb_n)^{-1}\sum_{i=1}^{n}K_i\left(s\right)\to_p f_s\left(s\right)$ by the standard result of the kernel density estimator. ■

A.4 Proof of Theorem 4

We let $\phi_{2n}=\log n/a_n$, where $a_n=n^{1-2\epsilon}b_n$ and ϵ is given in Assumption A-(ii).

Lemma A.11 *For a given $s\in\mathcal{S}_0$, let $\gamma(s)=\gamma_0(s)+r(s)\phi_{2n}$ for some continuously differentiable $r(s)$ satisfying $0<\underline{r}=\inf_{s\in\mathcal{S}_0}r(s)\leq\sup_{s\in\mathcal{S}_0}r(s)=\overline{r}<\infty$. Then there exist constants $0<C_T,C_{\overline{T}}<\infty$ such that for any $\eta>0$,*

$$\begin{aligned}\mathbb{P}\left(\sup_{s\in\mathcal{S}_0}\left|T_n\left(\gamma;s\right)-\mathbb{E}\left[T_n\left(\gamma;s\right)\right]\right|>\eta\right)&\leq\frac{C_T}{\eta}\left(\phi_{2n}\frac{\log n}{nb_n}\right)^{1/2},\\ \mathbb{P}\left(\sup_{s\in\mathcal{S}_0}\left|\overline{T}_n\left(\gamma;s\right)-\mathbb{E}\left[\overline{T}_n\left(\gamma;s\right)\right]\right|>\eta\right)&\leq\frac{C_{\overline{T}}}{\eta}\left(\phi_{2n}\frac{\log n}{nb_n}\right)^{1/2}\end{aligned}$$

if n is large enough.

Lemma A.12 *For a given $s\in\mathcal{S}_0$, let $\gamma(s)=\gamma_0(s)+r(s)\phi_{2n}$, where $r(s)$ is defined in Lemma A.11. Then there exists a constant $0<C_L,C_{\overline{L}}<\infty$ such that for any $\eta>0$,*

$$\begin{aligned}\mathbb{P}\left(\sup_{s\in\mathcal{S}_0}\left\|L_n\left(\gamma;s\right)\right\|>\eta\right)&\leq\frac{C_L}{\eta}\left(\phi_{2n}\log n\right)^{1/2},\\ \mathbb{P}\left(\sup_{s\in\mathcal{S}_0}\left\|\overline{L}_n\left(\gamma;s\right)\right\|>\eta\right)&\leq\frac{C_{\overline{L}}}{\eta}\left(\phi_{2n}\log n\right)^{1/2}\end{aligned}$$

if n is large enough.

Lemma A.13 *For any $\eta>0$ and $\varepsilon>0$, there exist constants $0<\overline{C},\overline{r},C_T,C_{\overline{T}}<\infty$ such that*

$$\mathbb{P}\left(\inf_{\overline{r}\phi_{2n}<\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|<\overline{C}}\frac{\sup_{s\in\mathcal{S}_0}T_n\left(\gamma;s\right)}{\sup_{s\in\mathcal{S}_0}\left|\gamma\left(s\right)-\gamma_0\left(s\right)\right|}<C_T(1-\eta)\right)\leq\varepsilon,\quad\text{(A.33)}$$

$$\mathbb{P}\left(\sup_{\overline{r}\phi_{2n}<\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|<\overline{C}}\frac{\sup_{s\in\mathcal{S}_0}\overline{T}_n\left(\gamma;s\right)}{\sup_{s\in\mathcal{S}_0}\left|\gamma\left(s\right)-\gamma_0\left(s\right)\right|}>C_{\overline{T}}(1+\eta)\right)\leq\varepsilon,\quad\text{(A.34)}$$

$$\mathbb{P}\left(\sup_{\overline{r}\phi_{2n}<\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|<\overline{C}}\frac{\sup_{s\in\mathcal{S}_0}\left\|L_n\left(\gamma;s\right)\right\|}{\sqrt{a_n}\sup_{s\in\mathcal{S}_0}\left|\gamma\left(s\right)-\gamma_0\left(s\right)\right|}>\eta\right)\leq\varepsilon,\quad\text{(A.35)}$$

$$\mathbb{P}\left(\sup_{\overline{r}\phi_{2n}<\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|<\overline{C}}\frac{\sup_{s\in\mathcal{S}_0}\left\|\overline{L}_n\left(\gamma;s\right)\right\|}{\sqrt{a_n}\sup_{s\in\mathcal{S}_0}\left|\gamma\left(s\right)-\gamma_0\left(s\right)\right|}>\eta\right)\leq\varepsilon,\quad\text{(A.36)}$$

if $n^{1-2\epsilon}b_n^2\to\varrho<\infty$.

Lemma A.14 $n^{\epsilon}\sup_{s\in\mathcal{S}_0}\left\|\widehat{\theta}(\widehat{\gamma}(s))-\theta_0\right\|=o_p(1)$.

Proof of Theorem 4 Since $\sup_{s\in\mathcal{S}_0}(Q_n^*(\widehat{\gamma}(s);s)-Q_n^*(\gamma_0(s);s))\leq 0$ by construction, where $Q_n^*(\gamma(s);s)$ is defined in (A.20), it suffices to show that as $n\to\infty$,

$$\mathbb{P}\left(\sup_{s\in\mathcal{S}_0}\{Q_n^*(\gamma(s);s)-Q_n^*(\gamma_0(s);s)\}>0\right)\to 1$$

for any $\gamma(s)$ such that $\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|>\overline{r}\phi_{2n}$ where $\overline{r}$ is chosen in Lemma A.13.

To this end, consider γ such that $\overline{r}\phi_{2n}\leq\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|\leq\overline{C}$ for some $0<\overline{r},\overline{C}<\infty$. Then, using (A.24) and Lemma A.14, we have

$$\begin{aligned}&\frac{Q_n^*(\gamma(s);s)-Q_n^*(\gamma_0(s);s)}{a_n\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|}\\ \geq\;&\frac{T_n(\gamma;s)}{\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|}-2\frac{2L_n(\gamma;s)}{\sqrt{a_n}\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|}-\frac{2C^*(s)b_n}{\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|}+o_p(1)\end{aligned}$$

for some $C^*(s)=O_p(1)$. Furthermore, Lemma A.2 gives that $\sup_{s\in\mathcal{S}_0}C^*(s)$ is also $O_p(1)$, and hence

$$\begin{aligned}\sup_{\overline{r}\phi_{2n}<|\gamma(s)-\gamma_0(s)|<\overline{C}}\frac{\sup_{s\in\mathcal{S}_0}C^*(s)b_n}{\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|}&<\frac{\sup_{s\in\mathcal{S}_0}C^*(s)b_n}{\overline{r}\phi_{2n}}\\&=\frac{\sup_{s\in\mathcal{S}_0}C^*(s)}{\overline{r}}\left(\frac{a_nb_n}{\log n}\right)\\&=O_p(1)\end{aligned}$$

given $a_nb_n\to\varrho<\infty$. Thus, we have

$$\mathbb{P}\left(\sup_{\overline{r}\phi_{2n}<|\gamma(s)-\gamma_0(s)|<\overline{C}}\frac{2\sup_{s\in\mathcal{S}_0}C^*(s)b_n}{\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|}>\eta(s)\right)\leq\frac{\varepsilon}{3}$$

when n is sufficiently large. Therefore, Lemma A.13 yields that, for $\varepsilon>0$ and $\eta>0$,

$$\mathbb{P}\left(\inf_{\overline{r}\phi_{2n}<\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|<\overline{C}}\sup_{s\in\mathcal{S}_0}\{Q_n^*(\gamma(s);s)-Q_n^*(\gamma_0(s);s)\}>\eta\right)\geq 1-\varepsilon,$$

which completes the proof by the same argument as Theorem 2. ■

A.5 Proof of Theorem 5

Proof of Theorem 5 We simply denote the leave-one-out estimator $\widehat{\gamma}_{-i}(s_i)$ as $\widehat{\gamma}(s_i)$ in this proof. We let $\mathbf{1}_{\mathcal{S}_0}=\mathbf{1}[s_i\in\mathcal{S}_0]$ and consider a sequence $\Delta_n>0$ such that $\Delta_n\to 0$ as $n\to\infty$. Then,

$$\begin{aligned}\sqrt{n}\left(\widehat{\beta}-\beta_0\right)&=\left(\frac{1}{n}\sum_{i=1}^n x_ix_i^\top\mathbf{1}\left[q_i>\widehat{\gamma}(s_i)+\Delta_n\right]\mathbf{1}_{\mathcal{S}_0}\right)^{-1}\\&\quad\times\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n x_iu_i\mathbf{1}\left[q_i>\gamma_0(s_i)+\Delta_n\right]\mathbf{1}_{\mathcal{S}_0}\right.\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i \left\{\mathbf{1}\left[q_i > \widehat{\gamma}\left(s_i\right)+\Delta_n\right] - \mathbf{1}\left[q_i > \gamma_0\left(s_i\right)+\Delta_n\right]\right\}\mathbf{1}_{\mathcal{S}_0}\Bigg\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i x_i^\top \delta_0 \mathbf{1}\left[q_i \le \gamma_0\left(s_i\right)\right]\mathbf{1}\left[q_i > \widehat{\gamma}\left(s_i\right)+\Delta_n\right]\mathbf{1}_{\mathcal{S}_0}\Bigg\} \\
\equiv\ & \Xi_{n00}^{-1}\left\{\Xi_{n01}+\Xi_{n02}+\Xi_{n03}\right\} \qquad \text{(A.37)}
\end{aligned}
$$

and

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\delta}^* - \delta_0^*\right) &= \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top \mathbf{1}\left[q_i < \widehat{\gamma}\left(s_i\right)-\Delta_n\right]\mathbf{1}_{\mathcal{S}_0}\right)^{-1} \\
&\quad \times\Bigg\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i \mathbf{1}\left[q_i < \gamma_0\left(s_i\right)-\Delta_n\right]\mathbf{1}_{\mathcal{S}_0} \\
&\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i \left\{\mathbf{1}\left[q_i < \widehat{\gamma}\left(s_i\right)-\Delta_n\right] - \mathbf{1}\left[q_i < \gamma_0\left(s_i\right)-\Delta_n\right]\right\}\mathbf{1}_{\mathcal{S}_0}\Bigg\} \\
&\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i x_i^\top \delta_0 \mathbf{1}\left[q_i \le \gamma_0\left(s_i\right)\right]\mathbf{1}\left[q_i < \widehat{\gamma}\left(s_i\right)-\Delta_n\right]\mathbf{1}_{\mathcal{S}_0}\Bigg\} \\
&\equiv \Xi_{n10}^{-1}\left\{\Xi_{n11}+\Xi_{n12}+\Xi_{n13}\right\}, \qquad \text{(A.38)}
\end{aligned}
$$

where Ξ_{n02}, Ξ_{n03}, Ξ_{n12}, and Ξ_{n13} are all $o_p(1)$ from Lemma A.15 below. Therefore,

$$
\sqrt{n}\left(\widehat{\theta}^* - \theta_0^*\right) = \begin{pmatrix} \Xi_{n00}^{-1}\Xi_{n01} \\ \Xi_{n10}^{-1}\Xi_{n11} \end{pmatrix} + o_p(1) = \begin{pmatrix} \Xi_{n00} & 0 \\ 0 & \Xi_{n10} \end{pmatrix}^{-1} \begin{pmatrix} \Xi_{n01} \\ \Xi_{n11} \end{pmatrix} + o_p(1)
$$

and the desired result follows since

$$
\Xi_{n00} \to_p \mathbb{E}\left[x_i x_i^\top \mathbf{1}\left[q_i > \gamma_0\left(s_i\right)\right]\mathbf{1}_{\mathcal{S}_0}\right], \qquad \text{(A.39)}
$$

$$
\Xi_{n10} \to_p \mathbb{E}\left[x_i x_i^\top \mathbf{1}\left[q_i < \gamma_0\left(s_i\right)\right]\mathbf{1}_{\mathcal{S}_0}\right], \qquad \text{(A.40)}
$$

and

$$
\begin{pmatrix} \Xi_{n01} \\ \Xi_{n11} \end{pmatrix} \to_d \mathcal{N}\left(0, \lim_{n\to\infty}\frac{1}{n}Var\left[\sum_{i=1}^{n}\begin{pmatrix} x_i u_i \mathbf{1}\left[q_i > \gamma_0\left(s_i\right)\right]\mathbf{1}_{\mathcal{S}_0} \\ x_i u_i \mathbf{1}\left[q_i < \gamma_0\left(s_i\right)\right]\mathbf{1}_{\mathcal{S}_0} \end{pmatrix}\right]\right) \qquad \text{(A.41)}
$$

as $n \to \infty$.

First, by Assumptions A-(v) and (ix), (A.39) can be readily verified since we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top \mathbf{1}\left[q_i > \widehat{\gamma}\left(s_i\right)+\Delta_n\right]\mathbf{1}_{\mathcal{S}_0} \\
=\ &\frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top \mathbf{1}\left[q_i > \gamma_0\left(s_i\right)+\Delta_n\right]\mathbf{1}_{\mathcal{S}_0} \\
&+\frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top \left\{\mathbf{1}\left[q_i > \widehat{\gamma}\left(s_i\right)+\Delta_n\right] - \mathbf{1}\left[q_i > \gamma_0\left(s_i\right)+\Delta_n\right]\right\}\mathbf{1}_{\mathcal{S}_0}
\end{aligned}
$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i x_i^{\top} \mathbf{1}\left[q_i > \gamma_0(s_i) + \Delta_n\right] \mathbf{1}_{\mathcal{S}_0} + O_p(\phi_{2n})$$

with $\Delta_n \to 0$ as $n \to \infty$. More precisely, given Theorem 4, we consider $\widehat{\gamma}(s)$ in a neighborhood of $\gamma_0(s)$ with distance at most $\overline{r}\phi_{2n}$ for some large enough constant $\overline{r}$. We define a non-random function $\widetilde{\gamma}(s) = \gamma_0(s) + \overline{r}\phi_{2n}$ and $\widetilde{\Delta}_i(s_i) = \mathbf{1}\left[q_i > \widetilde{\gamma}(s_i) + \Delta_n\right] - \mathbf{1}\left[q_i > \gamma_0(s_i) + \Delta_n\right]$. Then, on the event $E_n^* = \{\sup_{s \in \mathcal{S}_0} |\widehat{\gamma}(s) - \gamma_0(s)| \leq \overline{r}\phi_{2n}\}$,

$$\begin{aligned} \mathbb{E}\left[x_i x_i^{\top} \widehat{\Delta}_i(s_i) \mathbf{1}_{\mathcal{S}_0}\right] &\leq \mathbb{E}\left[x_i x_i^{\top} \widetilde{\Delta}_i(s_i) \mathbf{1}_{\mathcal{S}_0}\right] \\ &= \int_{\mathcal{S}_0} \int_{\gamma_0(v)+\Delta_n}^{\widetilde{\gamma}(v)+\Delta_n} D(q,v) f(q,v)\, dq dv \\ &= \int_{\mathcal{S}_0} \left\{ D(\gamma_0(v), v) f(\gamma_0(v), v) (\widetilde{\gamma}(v) - \gamma_0(v)) + o_p(\phi_{2n}) \right\} dv \\ &\leq \overline{r}\phi_{2n} \int D(\gamma_0(v), v) f(\gamma_0(v), v)\, dv \\ &= O_p(\phi_{2n}) = o_p(1) \end{aligned} \tag{A.42}$$

from Theorem 4, Assumptions A-(v), (vii), and (ix). (A.40) can be verified symmetrically. Using a similar argument, since $\mathbb{E}\left[x_i u_i \mathbf{1}\left[q_i > \gamma_0(s_i)\right] \mathbf{1}_{\mathcal{S}_0}\right] = \mathbb{E}\left[x_i u_i \mathbf{1}\left[q_i < \gamma_0(s_i)\right] \mathbf{1}_{\mathcal{S}_0}\right] = 0$ from Assumption ID-(i), asymptotic normality in (A.41) follows by the Theorem of Bolthausen (1982) under Assumption A-(iii), which completes the proof. ■

Lemma A.15 *When $\phi_{2n} \to 0$ as $n \to \infty$, if we let $\Delta_n > 0$ such that $\Delta_n \to 0$ and $\phi_{2n}/\Delta_n \to 0$ as $n \to \infty$, then it holds that Ξ_{n02}, Ξ_{n03}, Ξ_{n12}, and Ξ_{n13} in (A.37) and (A.38) are all $o_p(1)$.*

References

Ananat, E. O. (2011): "The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality," *American Economic Journal: Applied Economics*, 3(2), 34–66.

Andrews, D. W. K. (1994): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 62(1), 43–72.

Bai, J. (1997): "Estimation of a Change Point in Multiple Regressions," *Review of Economics and Statistics*, 79, 551–563.

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

Bhattacharya, P. K., and P. J. Brockwell (1976): "The minimum of an additive process with applications to signal estimation and storage theory," *Z. Wahrsch. Verw. Gebiete,*, 37, 51–75.

Billingsley, P. (1968): *Convergence of Probability Measure*. Wiley, New York.

Bolthausen, E. (1982): "On the central limit theorem for stationary mixing random fields," *The Annuals of Probability*, 10(4), 1047–1050.

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

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

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

Chiou, Y., M. Chen, and J. Chen (2018): "Nonparametric Regression with Multiple Thresholds: Estimation and Inference," *Journal of Econometrics*, 206, 472–514.

Conley, T. G., and F. Molinari (2007): "Spatial correlation robust inference with errors in location or distance," *Journal of Econometrics*, 140(1), 76–96.

Dingel, J. I., A. Miscio, and D. R. Davis (2019): "Cities, Lights, and Skills in Developing Economics," *Forthcoming in Journal of Urban Economics*.

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

Heilmann, K. (2018): "Transit Access and Neighborhood Segregation. Evidence from the Dallas Light Rail System," *Regional Science and Urban Economics*, 73, 237–250.

Henderson, D. J., C. F. Parmeter, and L. Su (2017): "Nonparametric Threshold Regression: Estimation and Inference," *Working Paper*.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012): "Measuring Economic Growth from Outer Space," *American Economic Review*, 102(2), 994–1028.

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

Jenish, N., and I. R. Prucha (2009): "Central limit theorems and uniform laws of large numbers for arrays of random fields," *Journal of Econometrics*, 150, 86–98.

Kim, J., and D. Pollard (1990): "Cube Root Asymptotics," *The Annals of Statistics*, 18, 191–219.

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

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

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

Rozenfeld, H. D., D. Rybski, X. Gabaix, and H. A. Makse (2011): "The Area and Population of Cities: New Insights from a Di erent Perspective on Cities," *American Economic Review*, 101(5), 2205–25.

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

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

Tong, H. (1983): *Threshold models in nonlinear time series analysis (Lecture Notes in Statistics No. 21)*. New York: Springer-Verlag.

Vogel, K. B., R. Goldblatt, G. Hanson, and A. K. Khandelwal (2019): "Detecting Urban Markets with Satellite Imagery: An Application to India," *forthcoming in Journal of Urban Economics*.

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

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

Yu, P., Q. Liao, and P. Phillips (2019): "Inferences and Specification Testing in Threshold Regression with Endogeneity," *Working Paper*.

Yu, P., and P. Phillips (2018): "Threshold Regression with Endogeneity," *Journal of Econometrics*, 203, 50–68.

Supplement Material to "Nonparametric Sample Splitting"

By Yoonseok Lee and Yulong Wang

This supplementary material contains omitted proofs of some lemmas in Section S.1 and more details about the case where q_i and s_i are non-random in Section S.2 as noted in Section 2.

S.1 Omitted Proofs of Lemmas

Proof of Lemma A.2 We first show the pointwise convergence. For expositional simplicity, we only present the case of scalar x_i. Similarly as (A.1), we have

$$\begin{aligned}&\mathbb{E}\left[\Delta M_n(s)\right]\\ =&\iint D(q,s+b_nt)f(q,s+b_nt)\left\{\mathbf{1}\left[q<\gamma_0(s+b_nt)\right]-\mathbf{1}\left[q<\gamma_0(s)\right]\right\}K(t)dqdt,\end{aligned}$$

which is non-zero only when (i) $\gamma_0(s)<q<\gamma_0(s+b_nt)$ if $\gamma_0(s)<\gamma_0(s+b_nt)$; or (ii) $\gamma_0(s+b_nt)<q<\gamma_0(s)$ if $\gamma_0(s)>\gamma_0(s+b_nt)$. We suppose $\gamma_0(\cdot)$ is increasing around s. Then, for the case (i), since $0<\gamma_0(s+b_nt)-\gamma_0(s)$, it restricts $t>0$. For the case (ii), however, it restricts $t<0$. Therefore, if we let $m(q,s)=D(q,s)f(q,s)<\infty$, by Taylor expansion,

$$\begin{aligned}&\mathbb{E}\left[\Delta M_n(s)\right]\\ =&\int_0^\infty\int_{\gamma_0(s)}^{\gamma_0(s+b_nt)}m(q,s+b_nt)K(t)dqdt+\int_{-\infty}^0\int_{\gamma_0(s+b_nt)}^{\gamma_0(s)}m(q,s+b_nt)K(t)dqdt\\ =&\ m(\gamma_0(s),s)\dot\gamma_0(s)b_n\int_0^\infty tK(t)dt-m(\gamma_0(s),s)\dot\gamma_0(s)b_n\int_{-\infty}^0 tK(t)dt+O\left(b_n^2\right)\\ =&\ m(\gamma_0(s),s)\dot\gamma_0(s)b_n+O\left(b_n^2\right),\end{aligned}$$

where $\int_0^\infty tK(t)dt=-\int_{-\infty}^0 tK(t)dt$ and $\dot\gamma_0(s)=d\gamma_0(s)/ds>0$ in this case.

Symmetrically, we can also derive $\mathbb{E}\left[\Delta M_n(s)\right]=-m(\gamma_0(s),s)\dot\gamma_0(s)b_n+O\left(b_n^2\right)$ when $\gamma_0(\cdot)$ is decreasing around s. Therefore, $\mathbb{E}\left[\Delta M_n(s)\right]=m(\gamma_0(s),s)\left|\dot\gamma_0(s)\right|b_n=O(b_n)$ because $m(\gamma_0(s),s)\left|\dot\gamma_0(s)\right|<\infty$ from Assumptions A-(vi) and (vii). The desired result follows since $Var\left[\Delta M_n(s)\right]\leq 2Var\left[M_n(\gamma_0(s_i);s)\right]+2Var\left[M_n(\gamma_0(s);s)\right]=o(1)$ from (A.2).

Given the pointwise rate, it suffices to show $\Delta M_n(s)$ is uniformly tight. This is implied by the tightness of $M_n(s)$ in Lemma A.1 since $\gamma_0(\cdot)$ is continuous. The proof is complete. ■

Proof of Lemma A.4 We first show (A.16). We consider the case with $\gamma(s)>\gamma_0(s)$, and the other direction can be shown symmetrically. In this case, since $T_n(\gamma;s)=c_0^\top(M_n(\gamma(s);s)-M_n(\gamma_0(s);s))c_0$ where $\partial\mathbb{E}\left[T_n(\gamma;s)\right]/\partial\gamma(s)=c_0^\top D(\gamma(s),s)c_0f(\gamma(s),s)$ is continuous at $\gamma_0(s)$

and $c_0^\top D(\gamma_0(s), s)c_0 f(\gamma_0(s), s) > 0$ from Assumptions A-(vii) and (viii), there exists a sufficiently small $\overline{C}(s) > 0$ such that

$$\underline{\ell}_D(s) = \inf_{|\gamma(s)-\gamma_0(s)|<\overline{C}(s)} c_0^\top D(\gamma(s), s)c_0 f(\gamma(s), s) > 0.$$

By Taylor expansion, we have

$$\begin{aligned}\mathbb{E}\left[T_n(\gamma; s)\right] &= \int\int_{\gamma_0(s)}^{\gamma(s)} \mathbb{E}\left[\left(c_0^\top x_i\right)^2 |q, s+b_n t\right] f(q, s+b_n t)K(t)\, dqdt \\ &= \{\gamma(s) - \gamma_0(s)\}\left\{c_0^\top D(\gamma, s)c_0 f(\gamma, s) + C_1(s)b_n^2\right\}\end{aligned}$$

for some $C_1(s) < \infty$, which yields

$$\mathbb{E}\left[T_n(\gamma; s)\right] \geq \{\gamma(s) - \gamma_0(s)\}\,(\underline{\ell}_D(s) + C_1(s)b_n^2), \tag{B.1}$$

since $\mathbb{E}\left[T_n(\gamma_0; s)\right] = 0$. Furthermore, if we let $\Delta_i(\gamma; s) = \mathbf{1}_i(\gamma(s)) - \mathbf{1}_i(\gamma_0(s))$ and $Z_{n,i}(s) = \left(c_0^\top x_i\right)^2 \Delta_i(\gamma; s)K_i(s) - \mathbb{E}[\left(c_0^\top x_i\right)^2 \Delta_i(\gamma; s)K_i(s)]$, using a similar argument as (A.2), we have

$$\begin{aligned}& \mathbb{E}\left[\left(T_n(\gamma; s) - \mathbb{E}\left[T_n(\gamma; s)\right]\right)^2\right] \qquad \text{(B.2)}\\ =\ & \frac{1}{n^2 b_n^2}\sum_{i=1}^n \mathbb{E}\left[Z_{n,i}^2(s)\right] + \frac{1}{n^2 b_n^2}\sum_{i\neq j} Cov[Z_{n,i}(s), Z_{n,j}(s)] \\ \leq\ & \frac{C_2(s)}{nb_n}\{\gamma(s) - \gamma_0(s)\}\end{aligned}$$

for some $C_2(s) \in (0, \infty)$ since $\varphi \in (0, 2)$ in Assumption A-(iii).

We suppose n is large enough so that $\overline{r}(s)\phi_{1n} \leq \overline{C}(s)$. Similarly as Lemma A.7 in Hansen (2000), we set γ_g for $g = 1, 2, ..., \overline{g}+1$ such that, for any $s \in \mathcal{S}_0$, $\gamma_g(s) = \gamma_0(s) + 2^{g-1}\overline{r}(s)\phi_{1n}$, where $\overline{g}$ is the integer satisfying $\gamma_{\overline{g}}(s) - \gamma_0(s) = 2^{\overline{g}-1}\overline{r}(s)\phi_{1n} \leq \overline{C}(s)$ and $\gamma_{\overline{g}+1}(s) - \gamma_0(s) = 2^{\overline{g}}\phi_{1n} > \overline{C}(s)$. Then Markov's inequality and (B.2) yield that for any fixed $\eta(s) > 0$,

$$\begin{aligned}& \mathbb{P}\left(\max_{1\leq g\leq \overline{g}}\left|\frac{T_n(\gamma_g; s)}{\mathbb{E}\left[T_n(\gamma_g; s)\right]} - 1\right| > \eta(s)\right) \qquad \text{(B.3)}\\ \leq\ & \mathbb{P}\left(\max_{1\leq g\leq \overline{g}}\left|\frac{T_n(\gamma_g; s) - \mathbb{E}\left[T_n(\gamma_g; s)\right]}{\mathbb{E}\left[T_n(\gamma_g; s)\right]}\right| > \eta(s)\right) \\ \leq\ & \frac{1}{\eta^2(s)}\sum_{g=1}^{\overline{g}} \frac{\mathbb{E}\left[\left(T_n(\gamma_g; s) - \mathbb{E}\left[T_n(\gamma_g; s)\right]\right)^2\right]}{\left|\mathbb{E}\left[T_n(\gamma_g; s)\right]\right|^2} \\ \leq\ & \frac{1}{\eta^2(s)}\sum_{g=1}^{\overline{g}} \frac{C_2(s)\overline{r}(s)\phi_{1n}(nb_n)^{-1}}{|\overline{r}(s)\phi_{1n}(\underline{\ell}_D(s) + C_1(s)b_n^2)|^2}\end{aligned}$$

$$\leq \frac{1}{\eta^2(s)}\sum_{g=1}^{\overline{g}} \frac{C_2(s)(nb_n)^{-1}}{2^{g-1}\underline{\ell}_D^2(s)\overline{r}(s)\phi_{1n}}$$

$$
\begin{aligned}
&\leq \frac{C_2(s)}{\eta^2(s)\overline{r}(s)\underline{\ell}_D^2(s)} \sum_{g=1}^{\infty} \frac{1}{2^{g-1}} \times \frac{1}{n^{2\epsilon}} \\
&\leq \varepsilon(s)
\end{aligned}
$$

for any $\varepsilon(s) > 0$. From eq. (33) of Hansen (2000), for any $\gamma(s)$ such that $\overline{r}(s)\phi_{1n} \leq \gamma(s) - \gamma_0(s) \leq \overline{C}(s)$, there exists some g satisfying $\gamma_g(s) - \gamma_0(s) < \gamma(s) - \gamma_0(s) < \gamma_{g+1}(s) - \gamma_0(s)$, and then

$$
\begin{aligned}
\frac{T_n(\gamma; s)}{|\gamma(s) - \gamma_0(s)|} &\geq \frac{T_n(\gamma_g; s)}{\mathbb{E}\left[T_n(\gamma_g; s)\right]} \times \frac{\mathbb{E}\left[T_n(\gamma_g; s)\right]}{|\gamma_{g+1}(s) - \gamma_0(s)|} \\
&\geq \left\{1 - \max_{1 \leq g \leq \overline{g}} \left| \frac{T_n(\gamma_g; s)}{\mathbb{E}\left[T_n(\gamma_g; s)\right]} - 1 \right| \right\} \frac{\mathbb{E}\left[T_n(\gamma_g; s)\right]}{|\gamma_{g+1}(s) - \gamma_0(s)|}.
\end{aligned}
$$

Hence, we can find $C_T(s) < \infty$ such that

$$
\begin{aligned}
&\mathbb{P}\left(\inf_{\overline{r}(s)\phi_{1n} < |\gamma(s) - \gamma_0(s)| < \overline{C}(s)} \frac{T_n(\gamma; s)}{|\gamma(s) - \gamma_0(s)|} < C_T(s)(1 - \eta(s)) \right) \\
\leq\ &\mathbb{P}\left(\frac{T_n(\gamma_g; s)}{\mathbb{E}\left[T_n(\gamma_g; s)\right]} \times \frac{\mathbb{E}\left[T_n(\gamma_g; s)\right]}{|\gamma_{g+1}(s) - \gamma_0(s)|} < C_T(s)(1 - \eta(s)) \right) \\
\leq\ &\mathbb{P}\left(\left\{1 - \max_{1 \leq g \leq \overline{g}} \left| \frac{T_n(\gamma_g; s)}{\mathbb{E}\left[T_n(\gamma_g; s)\right]} - 1 \right| \right\} \frac{\mathbb{E}\left[T_n(\gamma_g; s)\right]}{|\gamma_{g+1}(s) - \gamma_0(s)|} < C_T(s)(1 - \eta(s)) \right) \\
\leq\ &\varepsilon(s),
\end{aligned}
$$

where the last line follows from (B.1) and (B.3). The proof for (A.17) is similar to that for (A.16) and hence omitted.

For (A.18), $\mathbb{E}\left[L_n(\gamma; s)\right] = 0$ and we have

$$
\mathbb{E}\left[|L_n(\gamma; s)|^2 \right] \leq \phi_{1n} C_3(s) \tag{B.4}
$$

for some $C_3(s) \in (0, \infty)$ similarly as (B.2). By defining γ_g in the same way as above, the Markov's inequality and (B.4) get us that for any fixed $\eta(s) > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1 \leq g \leq \overline{g}} \frac{|L_n(\gamma_g; s)|}{\sqrt{a_n}\left(\gamma_g(s) - \gamma_0(s)\right)} > \eta(s) \right) \qquad \text{(B.5)} \\
\leq\ &\frac{1}{\eta^2(s)} \sum_{g=1}^{\infty} \frac{\mathbb{E}\left[L_n(\gamma_g, s)^2 \right]}{a_n |\gamma_g(s) - \gamma_0(s)|^2} \\
\leq\ &\frac{1}{\eta^2(s)} \sum_{g=1}^{\infty} \frac{\phi_{1n} C_3(s)}{a_n |\gamma_g(s) - \gamma_0(s)|^2} \\
\leq\ &\frac{C_3(s)}{\eta^2(s)\overline{r}(s)} \sum_{g=1}^{\infty} \frac{1}{2^{g-1}}.
\end{aligned}
$$

This probability is arbitrarily close to zero if $\overline{r}(s)$ is chosen large enough. It is worth to note that (B.5) provides the maximal (or sharp) rate of ϕ_{1n} as a_n^{-1} because we need $\phi_{1n}/a_n \left|\gamma_g(s) - \gamma_0(s)\right|^2 = O(\phi_{1n}a_n) = O(1)$ but $\phi_{1n} \to 0$ as $n \to \infty$. This $\phi_{1n}a_n = O(1)$ condition also satisfies (B.3).

Finally, for a given g, we define $\Gamma_g(s)$ as the collection of $\gamma(s)$ satisfying $\overline{r}(s)2^{g-1}\phi_{1n} < \gamma(s) - \gamma_0(s) < \overline{r}(s)2^g\phi_{1n}$ for each $s \in S$. Then,

$$\begin{aligned}
& \mathbb{P}\left(\sup_{\overline{r}(s)\phi_{1n}<|\gamma(s)-\gamma_0(s)|<\overline{C}(s)} \frac{|L_n(\gamma; s)|}{\sqrt{a_n}\,|\gamma(s) - \gamma_0(s)|} > \eta(s)\right) \qquad \text{(B.6)} \\
= \; & \mathbb{P}\left(\max_{1\le g\le \overline{g}} \sup_{\gamma\in\Gamma_g(s)} \frac{|L_n(\gamma; s)|}{\sqrt{a_n}\,(\gamma(s) - \gamma_0(s))} > \eta(s)\right) \\
\le \; & \mathbb{P}\left(\max_{1\le g\le \overline{g}} \frac{\left|L_n\left(\gamma_g; s\right)\right|}{\sqrt{a_n}\left(\gamma_{g+1}(s) - \gamma_0(s)\right)} > \eta(s)\right) \\
\le \; & \frac{C_4(s)}{\eta^2(s)\,\overline{r}(s)}
\end{aligned}$$

for some $C_4(s) \in (0, \infty)$. Combining (B.5) and (B.6), we thus have

$$\begin{aligned}
& \mathbb{P}\left(\sup_{\overline{r}(s)\phi_{1n}<|\gamma(s)-\gamma_0(s)|<\overline{C}(s)} \frac{\left|L_n\left(\gamma_g; s\right)\right|}{\sqrt{a_n}\,(\gamma(s) - \gamma_0(s))} > \eta(s)\right) \\
\le \; & 2\mathbb{P}\left(\max_{1\le g\le \overline{g}} \frac{\left|L_n\left(\gamma_g; s\right)\right|}{\sqrt{a_n}\left(\gamma_g(s) - \gamma_0(s)\right)} > \eta(s)\right) \\
& +2\mathbb{P}\left(\max_{1\le g\le \overline{g}} \sup_{\gamma\in\Gamma_g(s)} \frac{|L_n(\gamma; s)|}{\sqrt{a_n}\,(\gamma(s) - \gamma_0(s))} > \eta(s)\right) \\
\le \; & \varepsilon(s)
\end{aligned}$$

for any $\varepsilon(s) > 0$ if we pick $\overline{r}(s)$ sufficiently large. The proof for (A.19) is similar to that for (A.18) and hence omitted. ■

Proof of Lemma A.5 Using the same notations in Lemma A.3, (A.12) yields

$$\begin{aligned}
& n^{\epsilon}\left(\widehat{\theta}(\widehat{\gamma}(s)) - \theta_0\right) \qquad \text{(B.7)} \\
= \; & \left\{\frac{1}{nb_n}\widetilde{Z}(\widehat{\gamma}(s); s)^{\top}\widetilde{Z}(\widehat{\gamma}(s); s)\right\}^{-1} \\
& \times\left\{\frac{n^{\epsilon}}{nb_n}\widetilde{Z}(\widehat{\gamma}(s); s)^{\top}\widetilde{u}(s) - \frac{n^{\epsilon}}{nb_n}\widetilde{Z}(\widehat{\gamma}(s); s)^{\top}\left(\widetilde{Z}(\widehat{\gamma}(s); s) - \widetilde{Z}(\gamma_0(s_i); s)\right)\theta_0\right\} \\
\equiv \; & \Theta_{A1}^{-1}(s)\left\{\Theta_{A2}(s) - \Theta_{A3}(s)\right\}.
\end{aligned}$$

For the denominator $\Theta_{A1}(s)$, we have

$$\Theta_{A1}(s) = \begin{pmatrix} (nb_n)^{-1}\sum_{i=1}^n x_i x_i^{\top} K_i(s) & M_n(\widehat{\gamma}(s); s) \\ M_n(\widehat{\gamma}(s); s) & M_n(\widehat{\gamma}(s); s) \end{pmatrix} \qquad \text{(B.8)}$$

$$\to_p \begin{pmatrix} M(s) & M(\gamma_0(s); s) \\ M(\gamma_0(s); s) & M(\gamma_0(s); s) \end{pmatrix},$$

where $M_n(\widehat{\gamma}(s); s) \to_p M(\gamma_0(s); s) < \infty$ from Lemma A.1 and the pointwise consistency of $\widehat{\gamma}(s)$ in Lemma A.3. In addition, $(nb_n)^{-1} \sum_{i=1}^{n} x_i x_i^\top K_i(s) \to_p M(s) = \int_{-\infty}^{\infty} D(q,s) f(q,s)\, dq < \infty$ from the standard kernel estimation result. Note that the probability limit is positive definite since both $M(s)$ and $M(\gamma_0(s); s)$ are positive definite and

$$M(s) - M(\gamma_0(s); s) = \int_{\gamma_0(s)}^{\infty} D(q,s) f(q,s)\, dq > 0$$

for any $\gamma_0(s) \in \Gamma$ from Assumption A-(viii).

For the numerator part $\Theta_{A2}(s)$, we have $\Theta_{A2}(s) = O_p(a_n^{-1/2}) = o_p(1)$ because

$$\frac{1}{\sqrt{nb_n}} \widetilde{Z}(\widehat{\gamma}(s); s)^\top \widetilde{u}(s) = \begin{pmatrix} (nb_n)^{-1/2} \sum_{i=1}^{n} x_i u_i K_i(s) \\ J_n(\widehat{\gamma}(s); s) \end{pmatrix} = O_p(1) \tag{B.9}$$

from from Lemma A.1 and the pointwise consistency of $\widehat{\gamma}(s)$ in Lemma A.3. Note that the standard kernel estimation result gives $(nb_n)^{-1/2} \sum_{i=1}^{n} x_i u_i K_i(s) = O_p(1)$. Moreover, we have

$$\Theta_{A3}(s) = \begin{pmatrix} (nb_n)^{-1} \sum_{i=1}^{n} c_0^\top x_i x_i^\top \{\mathbf{1}_i(\widehat{\gamma}(s)) - \mathbf{1}_i(\gamma_0(s_i))\} K_i(s) \\ (nb_n)^{-1} \sum_{i=1}^{n} c_0^\top x_i x_i^\top \mathbf{1}_i(\widehat{\gamma}(s)) \{\mathbf{1}_i(\widehat{\gamma}(s)) - \mathbf{1}_i(\gamma_0(s_i))\} K_i(s) \end{pmatrix} \tag{B.10}$$

and

$$\begin{aligned} & \frac{1}{nb_n} \sum_{i=1}^{n} c_0^\top x_i x_i^\top \{\mathbf{1}_i(\widehat{\gamma}(s)) - \mathbf{1}_i(\gamma_0(s_i))\} K_i(s) \\ \leq \;& \|c_0\| \, \|M_n(\widehat{\gamma}(s); s) - M_n(\gamma_0(s_i); s)\| \\ \leq \;& \|c_0\| \left\{ \|M_n(\widehat{\gamma}(s); s) - M_n(\gamma_0(s); s)\| + O_p(b_n) \right\} \\ = \;& o_p(1), \end{aligned} \tag{B.11}$$

where the second inequality is from (A.14) and the last equality is because $M_n(\gamma; s) \to_p M(\gamma; s)$ is continuous in γ and $\widehat{\gamma}(s) \to_p \gamma_0(s)$ in Lemma A.3. Since

$$\begin{aligned} & \frac{1}{nb_n} \sum_{i=1}^{n} c_0^\top x_i x_i^\top \mathbf{1}_i(\widehat{\gamma}(s)) \{\mathbf{1}_i(\widehat{\gamma}(s)) - \mathbf{1}_i(\gamma_0(s_i))\} K_i(s) \\ \leq \;& \|c_0\| \, \|M_n(\widehat{\gamma}(s); s) - M_n(\gamma_0(s_i); s)\| = o_p(1) \end{aligned} \tag{B.12}$$

from (B.11), we have $\Theta_{A3}(s) = o_p(1)$ as well, which completes the proof. ■

Proof of Lemma A.7 Using the same notations in Lemma A.3, we write

$$\begin{aligned} & \sqrt{nb_n} \left(\widehat{\theta}(\widehat{\gamma}(s)) - \theta_0 \right) \\ = \;& \left\{ \frac{1}{nb_n} \widetilde{Z}(\widehat{\gamma}(s); s)^\top \widetilde{Z}(\widehat{\gamma}(s); s) \right\}^{-1} \end{aligned}$$

$$
\begin{aligned}
&\times \left\{ \frac{1}{\sqrt{nb_n}} \widetilde{Z}(\widehat{\gamma}(s);s)^\top \widetilde{u}(s) - \frac{1}{\sqrt{nb_n}} \widetilde{Z}(\widehat{\gamma}(s);s)^\top \left(\widetilde{Z}(\widehat{\gamma}(s);s) - \widetilde{Z}(\gamma_0(s_i);s) \right) \theta_0 \right\} \\
\equiv \; & \Theta_{B1}^{-1}(s) \left\{ \Theta_{B2}(s) - \Theta_{B3}(s) \right\}
\end{aligned}
$$

similarly as (B.7). For the denominator, since $\Theta_{B1}(s) = \Theta_{A1}(s)$ in (B.7), then $\Theta_{B1}^{-1}(s) = O_p(1)$ from (B.8). For the numerator, we first have $\Theta_{B2}(s) = O_p(1)$ from (B.9). For $\Theta_{B3}(s)$, similarly as (B.10),

$$
\Theta_{B3}(s) = \left(\begin{array}{c} a_n^{-1/2} \sum_{i=1}^n n^{-\epsilon} \delta_0^\top x_i x_i^\top \left\{ \mathbf{1}_i \left(\widehat{\gamma}(s) \right) - \mathbf{1}_i \left(\gamma_0 \left(s_i \right) \right) \right\} K_i \left(s \right) \\ a_n^{-1/2} \sum_{i=1}^n n^{-\epsilon} \delta_0^\top x_i x_i^\top \mathbf{1}_i \left(\widehat{\gamma}(s) \right) \left\{ \mathbf{1}_i \left(\widehat{\gamma}(s) \right) - \mathbf{1}_i \left(\gamma_0 \left(s_i \right) \right) \right\} K_i \left(s \right) \end{array} \right).
$$

However, since $\widehat{\gamma}(s) = \gamma_0(s) + r(s)\phi_{1n}$ for some $r(s) < \infty$ from Theorem 2, similarly as (A.25), we have

$$
\begin{aligned}
& \mathbb{E} \left[\sum_{i=1}^n n^{-\epsilon} \delta_0^\top x_i x_i^\top \left\{ \mathbf{1}_i \left(\widehat{\gamma}(s) \right) - \mathbf{1}_i \left(\gamma_0 \left(s_i \right) \right) \right\} K_i \left(s \right) \right] \\
\leq \; & a_n \left| \iint_{\gamma_0(s+b_n t)}^{\gamma_0(s)+r(s)\phi_{1n}} c_0^\top \mathbb{E} \left[x_i x_i^\top | q, s + b_n t \right] K\left(t\right) f\left(q, s + b_n t\right) dq dt \right| \\
\leq \; & a_n \left| \iint_{\gamma_0(s)}^{\gamma_0(s)+r(s)\phi_{1n}} c_0^\top \mathbb{E} \left[x_i x_i^\top | q, s + b_n t \right] K\left(t\right) f\left(q, s + b_n t\right) dq dt \right| \\
& + a_n \left| \iint_{\gamma_0(s)}^{\gamma_0(s+b_n t)} c_0^\top \mathbb{E} \left[x_i x_i^\top | q, s + b_n t \right] K\left(t\right) f\left(q, s + b_n t\right) dq dt \right| \\
= \; & a_n \phi_{1n} \left| r(s) \right| \left| c_0^\top D\left(\gamma_0\left(s\right), s\right) \right| f\left(\gamma_0\left(s\right), s\right) + O(a_n b_n) \\
= \; & O(1)
\end{aligned}
$$

as $a_n \phi_{1n} = 1$ and $a_n b_n = n^{1-2\epsilon} b_n^2 \to \varrho < \infty$. We also have

$$
Var \left[\sum_{i=1}^n n^{-\epsilon} \delta_0^\top x_i x_i^\top \left\{ \mathbf{1}_i \left(\widehat{\gamma}(s) \right) - \mathbf{1}_i \left(\gamma_0 \left(s_i \right) \right) \right\} K_i \left(s \right) \right] = O(n^{-2\epsilon}) = o(1),
$$

similarly as (A.26). Therefore, from the same reason as (B.12), we have $\Theta_{B3}(s) = O_p(a_n^{-1/2}) = o_p(1)$, which completes the proof. ■

Proof of Lemma A.8 First consider the case with $r > 0$. In this case, we have

$$
\begin{aligned}
& \left\{ \mathbf{1}[q \leq \gamma_0\left(s\right) + r/a_n] - \mathbf{1}[q \leq \gamma_0\left(s\right)] \right\} \left\{ \mathbf{1}[q \leq \gamma_0\left(s + b_n t\right)] - \mathbf{1}[q \leq \gamma_0\left(s\right)] \right\} \\
= \; & \mathbf{1}\left[\gamma_0\left(s\right) < q \leq \gamma_0\left(s + b_n t\right) < \gamma_0\left(s\right) + r/a_n \right] \\
& + \mathbf{1}\left[\gamma_0\left(s\right) < q \leq \gamma_0\left(s\right) + r/a_n < \gamma_0\left(s + b_n t\right) \right].
\end{aligned}
$$

Therefore, if we denote $g(q,s) = c_0^\top D(q,s) c_0 f\left(q,s\right)$,

$$
\mathbb{E}\left[B_{n3}^*(r,s) \right]
$$

$$
\begin{aligned}
&= a_n \iint c_0^\top D(q, s+b_n t) c_0 \left\{\mathbf{1}[q \leq \gamma_0(s) + r/a_n] - \mathbf{1}[q \leq \gamma_0(s)]\right\} \\
&\qquad \times \left\{\mathbf{1}[q \leq \gamma_0(s+b_n t)] - \mathbf{1}[q \leq \gamma_0(s)]\right\} K(t) f(q, s+b_n t)\, dq dt \\
&= a_n \int_{\mathcal{T}_1(r;s)} \int_{\gamma_0(s)}^{\gamma_0(s+b_n t)} g(q, s+b_n t) K(t)\, dq dt \\
&\quad + a_n \int_{\mathcal{T}_2(r;s)} \int_{\gamma_0(s)}^{\gamma_0(s)+r/a_n} g(q, s+b_n t) K(t)\, dq dt \\
&\equiv B_{n31}(r,s) + B_{n32}(r,s),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{T}_1(r;s) &= \left\{\gamma_0(s) < \gamma_0(s+b_n t)\right\} \cap \left\{\gamma_0(s+b_n t) < \gamma_0(s) + r/a_n\right\}, \\
\mathcal{T}_2(r;s) &= \left\{\gamma_0(s) < \gamma_0(s+b_n t)\right\} \cap \left\{\gamma_0(s) + r/a_n < \gamma_0(s+b_n t)\right\}.
\end{aligned}
$$

Note that $\gamma_0(s) < \gamma_0(s) + r/a_n$ always holds for $r > 0$. However, similarly as in the proof of Lemma A.2, when $\gamma_0(\cdot)$ is increasing around s, $\gamma_0(s) < \gamma_0(s+b_n t)$ restricts that $t > 0$. Furthermore, $\gamma_0(s+b_n t) < \gamma_0(s) + r/a_n$ implies that $t < r/(a_n b_n \dot{\gamma}_0(s))$, where $0 < r/(a_n b_n \dot{\gamma}_0(s)) < \infty$. Therefore, $\mathcal{T}_1(r;s) = \{t : t > 0 \text{ and } t < r/(a_n b_n \dot{\gamma}_0(s))\}$. Similarly, since $\gamma_0(s) + r/a_n < \gamma_0(s+b_n t)$ implies $t > r/(a_n b_n \dot{\gamma}_0(s))$, we have $\mathcal{T}_2(r;s) = \{t : t > 0 \text{ and } t > r/(a_n b_n \dot{\gamma}_0(s))\}$. It follows that, by Taylor expansion,

$$
\begin{aligned}
B_{n31}(r,s) &= a_n \int_0^{r/(a_n b_n \dot{\gamma}_0(s))} \int_{\gamma_0(s)}^{\gamma_0(s+b_n t)} g(q, s+b_n t) K(t)\, dq dt \\
&= a_n b_n g(\gamma_0(s), s) \dot{\gamma}_0(s) \int_0^{r/(a_n b_n \dot{\gamma}_0(s))} t K(t)\, dt + a_n b_n O(b_n) \\
&= \varrho g(\gamma_0(s), s) \dot{\gamma}_0(s) \mathcal{K}_1(r, \varrho; s) + O(b_n)
\end{aligned}
$$

as $a_n b_n = n^{1-2\epsilon} b_n^2 \to \varrho \in (0, \infty)$, and

$$
\begin{aligned}
B_{n32}(r,s) &= a_n \int_{r/a_n b_n \dot{\gamma}_0(s)}^{\infty} \int_{\gamma_0(s)}^{\gamma_0(s)+r/a_n} g(q, s+b_n t) K(t)\, dq dt \\
&= r g(\gamma_0(s), s) \int_{r/(a_n b_n \dot{\gamma}_0(s))}^{\infty} K(t)\, dt + O(b_n) \\
&= r g(\gamma_0(s), s) \left\{\frac{1}{2} - \mathcal{K}_0(r, \varrho; s)\right\} + O(b_n)
\end{aligned}
$$

as $|\mathcal{K}_0(r, \varrho; s)| \leq 1/2$ and $|\mathcal{K}_1(r, \varrho; s)| \leq 1/2$.

When $\gamma_0(\cdot)$ is decreasing around s, $-\infty < r/(a_n b_n \dot{\gamma}_0(s)) < 0$ and we can also derive

$$
\begin{aligned}
B_{n31}(r,s) &= a_n \int_{r/(a_n b_n \dot{\gamma}_0(s))}^{0} \int_{\gamma_0(s)}^{\gamma_0(s+b_n t)} g(q, s+b_n t) K(t)\, dq dt \\
&= -\varrho g(\gamma_0(s), s) \dot{\gamma}_0(s) \mathcal{K}_1(r, \varrho; s) + O(b_n), \\
B_{n32}(r,s) &= a_n \int_{-\infty}^{r/(a_n b_n \dot{\gamma}_0(s))} \int_{\gamma_0(s)}^{\gamma_0(s)+r/a_n} g(q, s+b_n t) K(t)\, dq dt
\end{aligned}
$$

$$= \quad rg(\gamma_0(s), s)\left\{(1/2) - \mathcal{K}_0\left(r, \varrho; s\right)\right\} + O\left(b_n\right),$$

because, when $\dot{\gamma}_0(s) < 0$, we have $\int_{r/(a_n b_n \dot{\gamma}_0(s))}^{0} tK(t)\,dt = -\int_{0}^{r/(a_n b_n(-\dot{\gamma}_0(s)))} tK(t)\,dt$ and $\int_{-\infty}^{r/(a_n b_n \dot{\gamma}_0(s))} K(t)\,dt = \int_{r/(a_n b_n(-\dot{\gamma}_0(s)))}^{\infty} K(t)\,dt$ with $\dot{\gamma}_0(s) < 0$. It follows that, by combining these results, we have

$$\mathbb{E}\left[B_{n3}^*(r,s)\right] = |r|\, g(\gamma_0(s), s)\left\{\frac{1}{2} - \mathcal{K}_0\left(r, \varrho; s\right)\right\} + \varrho g(\gamma_0(s), s)\left|\dot{\gamma}_0(s)\right| \mathcal{K}_1\left(r, \varrho; s\right) + O\left(b_n\right).$$

Furthermore, since $|B_{n3}^*(r,s)| \leq \sum_{i=1}^{n} (\delta_0^\top x_i)^2 \left|\mathbf{1}_i\left(\gamma_0\left(s\right) + r/a_n\right) - \mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right| K_i\left(s\right)$, we have $Var\left[B_{n3}^*(r,s)\right] = O(n^{-2\epsilon}) = o(1)$ from (A.26) in Lemma A.6, which completes the proof. ■

Proof of Lemma A.9 Define $W_\mu(r) = W(r) + \mu(r)$, $\tau^+ = \arg\max_{r\in\mathbb{R}^+} W_\mu(r)$, and $\tau^- = \arg\max_{r\in\mathbb{R}^-} W_\mu(r)$. The process $W_\mu(\cdot)$ is a Gaussian process, and hence Lemma 2.6 of Kim and Pollard (1990) implies that τ^+ and τ^- are unique almost surely. Recall that we define $W(r) = W_1(-r)\mathbf{1}[r < 0] + W_2(r)\mathbf{1}[r > 0]$, where $W_1(\cdot)$ and $W_2(\cdot)$ are two independent standard Wiener processes defined on $\mathbb{R}^+$. We claim that

$$\mathbb{E}[\tau^+] = -\mathbb{E}[\tau^-] < \infty, \tag{B.13}$$

which gives the desired result.

The equality in (B.13) follows directly from the symmetry (i.e., $\mathbb{P}(W_\mu(\tau^+) > W_\mu(\tau^-)) = 1/2$) and the fact that W_1 is independent of W_2. Now, we focus on $r > 0$ and show that $\mathbb{E}[\tau^+] < \infty$. First, for any $r > 0$,

$$\mathbb{P}\left(W_\mu(r) \geq 0\right) = \mathbb{P}\left(W_2(r) \geq -\mu(r)\right) = \mathbb{P}\left(\frac{W_2(r)}{\sqrt{r}} \geq -\frac{\mu(r)}{\sqrt{r}}\right) = 1 - \Phi\left(-\frac{\mu(r)}{\sqrt{r}}\right),$$

where $\Phi(\cdot)$ denotes the standard normal distribution function. Since the sample path of $W_\mu(\cdot)$ is continuous, for some $\underline{r} > 0$, we then have

$$\begin{aligned}
\mathbb{E}[\tau^+] &= \int_0^\infty \left\{1 - \mathbb{P}\left(\tau^+ \leq r\right)\right\} dr \\
&= \int_0^{\underline{r}} \mathbb{P}\left(\tau^+ > r\right) dr + \int_{\underline{r}}^\infty \mathbb{P}\left(\tau^+ > r\right) dr \\
&\leq C_1 + \int_{\underline{r}}^\infty \mathbb{P}\left(W_\mu(\tau^+) \geq 0 \textit{ and } \tau^+ > r\right) dr \\
&\leq C_1 + \int_{\underline{r}}^\infty \mathbb{P}\left(W_\mu(r) \geq 0\right) dr \\
&= C_1 + \int_{\underline{r}}^\infty \left(1 - \Phi\left(-\frac{\mu(r)}{\sqrt{r}}\right)\right) dr \qquad \text{(B.14)}
\end{aligned}$$

for some $C_1 < \infty$, where the first inequality is because $W_\mu(\tau^+) = \max_{r\in\mathbb{R}^+} W_\mu(r) \geq 0$ given $W_\mu(0) = 0$, and the second inequality is because $\mathbb{P}\left(W_\mu(r) \geq 0\right)$ is monotonically decreasing to zero on $\mathbb{R}^+$. The second term in (B.14) can be shown bounded as follows. Using change of variables $t = r^{\varepsilon}$, integral by parts, and the condition that $\lim_{r\to\infty} r^{-((1/2)+\varepsilon)}\mu(r) = -\infty$ for some $\varepsilon > 0$ in turn, we have

$$\begin{aligned}\int_{\underline{r}}^{\infty}\left(1-\Phi\left(-\frac{\mu(r)}{\sqrt{r}}\right)\right)dr &\leq C_2\int_{\underline{r}}^{\infty}\left(1-\Phi\left(r^{\varepsilon}\right)\right)dr\\ &= C_2\int_{\underline{r}^{1/\varepsilon}}^{\infty}\left(1-\Phi\left(t\right)\right)dt^{1/\varepsilon}\\ &= C_2 + C_3\int_{\underline{r}^{1/\varepsilon}}^{\infty}t^{1/\varepsilon}\phi(t)dt<\infty\end{aligned}$$

for some $C_2, C_3 < \infty$ if $\underline{r}$ is large enough, where $\phi(\cdot)$ denotes the standard normal density function and we use $\lim_{t\to\infty} t^{1/\varepsilon}\left(1-\Phi\left(t\right)\right) = 0$. The same result can be obtained for $r < 0$ symmetrically, which completes the proof. ■

Proof of Lemma A.10 For given (ϱ, s), we simply let $\mu(r) = \mu\left(r, \varrho; s\right)$. Then, for the kernel functions satisfying Assumption A-(x), it is readily verified that $\mu(0) = 0$, $\mu(r)$ is continuous in r, and $\mu(r)$ is symmetric about zero. To check other conditions, for $r > 0$, we first write

$$\mu(r) = -r\int_0^{rC_1}K(t)dt + C_2\int_0^{rC_1}tK(t)dt,$$

where C_1 and C_2 are some positive constants depending on $(\varrho, |\dot{\gamma}_0(s)|, \xi(s))$. We consider the two possible cases.

First, if $K(\cdot)$ has a bounded support, say $[-\underline{r}, \underline{r}]$, then $\mu(r) = -rC_3 + C_4$ for $r > \underline{r}$ and some $0 < C_3, C_4 < \infty$. Thus, $\mu(r)$ is monotonically decreasing on $\mathbb{R}\backslash[-\underline{r}, \underline{r}]$ and $\lim_{r\to\infty} r^{-((1/2)+\varepsilon)}\mu(r) = -\infty$ for any $\varepsilon > 0$.

Second, if $K(\cdot)$ has an unbounded support, we have

$$\frac{\partial\mu(r)}{\partial r} = -\int_0^{rC_1}K(t)dt - rC_1K(C_1r) + rC_1^2C_2K(C_1r)$$

by the Leibniz integral rule. However, for $r > \underline{r}$ for some large enough $\underline{r}$, it is strictly negative because $\int_0^{rC_1}K(t)dt > 0$ and $\lim_{r\to\infty} rK(r) = 0$. This proves $\mu(r)$ is monotonically decreasing on $\mathbb{R}\backslash[-\underline{r}, \underline{r}]$. In addition, $\lim_{r\to\infty} r^{-((1/2)+\varepsilon)}\mu(r) = -\infty$ for any $\varepsilon > 0$ because $\int_0^{rC_1}K(t)dt < \int_0^{\infty}K(t)dt < \infty$, $\int_0^{rC_1}tK(t)dt < \int_0^{\infty}tK(t)dt < \infty$. The $r < 0$ case follows symmetrically using the identical argument. ■

Proof of Lemma A.11 We only present the argument for $T_n\left(\gamma; s\right)$ as the proof for $\overline{T}_n\left(\gamma; s\right)$ is identical. Let τ_n be some large truncation parameter to be chosen later, satisfying $\tau_n \to \infty$ as $n \to \infty$. Define $\mathbf{1}_{\tau_n} = \mathbf{1}[\left(c_0^{\top}x_i\right)^2 < \tau_n]$ and

$$T_n^{\tau}(\gamma, s) = \frac{1}{nb_n}\sum_{i=1}^{n}\left(c_0^{\top}x_i\right)^2\left|\Delta_i(\gamma; s)\right|K_i\left(s\right)\mathbf{1}_{\tau_n},$$

where $\Delta_i(\gamma; s) = \mathbf{1}_i(\gamma(s)) - \mathbf{1}_i(\gamma_0(s))$. The triangular inequality gives that, for any η,

$$
\begin{aligned}
& \mathbb{P}\left(\sup_{s\in\mathcal{S}_0} \left|T_n(\gamma; s) - \mathbb{E}\left[T_n(\gamma; s)\right]\right| > \eta\right) \qquad \text{(B.15)}\\
\leq\ & \mathbb{P}\left(\sup_{s\in\mathcal{S}_0} \left|T_n^{\tau}(\gamma; s) - T_n(\gamma; s)\right| > \eta/3\right)\\
& +\mathbb{P}\left(\sup_{s\in\mathcal{S}_0} \left|\mathbb{E}\left[T_n^{\tau}(\gamma; s)\right] - \mathbb{E}\left[T_n(\gamma; s)\right]\right| > \eta/3\right)\\
& +\mathbb{P}\left(\sup_{s\in\mathcal{S}_0} \left|T_n^{\tau}(\gamma; s) - \mathbb{E}\left[T_n^{\tau}(\gamma; s)\right]\right| > \eta/3\right)\\
\equiv\ & P_{T1n} + P_{T2n} + P_{T3n}.
\end{aligned}
$$

For the first one, since $r(s) > 0$ for all s, $\gamma(s) > \gamma_0(s)$ and

$$
\begin{aligned}
& \mathbb{E}\left[\sup_{s\in\mathcal{S}_0} \left|T_n^{\tau}(\gamma; s) - T_n(\gamma; s)\right|\right]\\
\leq\ & \mathbb{E}\left[\left|\frac{1}{nb_n}\sum_{i=1}^{n}\left(c_0^{\top}x_i\right)^2 \mathbf{1}\left[\inf_{s\in\mathcal{S}_0}\gamma_0(s) \leq q_i \leq \sup_{s\in\mathcal{S}_0}\gamma_0(s) + \bar{r}\phi_{2n}\right] K_i(s)\left(1 - \mathbf{1}_{\tau_n}\right)\right|\right]\\
\leq\ & \frac{1}{b_n}\mathbb{E}\left[\left|\left(c_0^{\top}x_i\right)^2 \mathbf{1}\left[\inf_{s\in\mathcal{S}_0}\gamma_0(s) \leq q_i \leq \sup_{s\in\mathcal{S}_0}\gamma_0(s) + \bar{r}\phi_{2n}\right] K_i(s)\left(1 - \mathbf{1}_{\tau_n}\right)\right|\right]\\
=\ & \tau_n^{-1}\int\int_{\inf_{s\in\mathcal{S}_0}\gamma_0(s)}^{\sup_{s\in\mathcal{S}_0}\gamma_0(s)+\bar{r}\phi_{2n}} \mathbb{E}\left[\left(c_0^{\top}x_i\right)^4 | q, s + b_n t\right] f(q, s + b_n t) K(t)\, dq dt\\
\leq\ & C_1\phi_{2n}\tau_n^{-1}
\end{aligned}
$$

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

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

for a generic random variable A. Hence, Markov's inequality yields that $P_{T1n} \leq C\phi_{2n}/(\eta\tau_n)$.

Next, to bound P_{T2n}, note that

$$
\begin{aligned}
& \mathbb{E}\left[T_n^{\tau}(\gamma; s)\right] - \mathbb{E}\left[T_n(\gamma; s)\right]\\
=\ & b_n^{-1}\mathbb{E}\left[\left|\left(c_0^{\top}x_i\right)^2 \mathbf{1}\left[\gamma_0(s) \leq q_i \leq \gamma(s)\right] K_i(s)\left(1 - \mathbf{1}_{\tau_n}\right)\right|\right]\\
\leq\ & \tau_n^{-1}\int\int_{\gamma_0(s)}^{\gamma(s)} \mathbb{E}\left[\left(c_0^{\top}x_i\right)^4 | q, s + b_n t\right] f(q, s + b_n t) K(t)\, dq dt\\
\leq\ & C_2\phi_{2n}\tau_n^{-1}
\end{aligned}
$$

for some $C_2 \in (0, \infty)$. By Assumptions A-(v), (vii), and (viii), the above bound is uniform in s. Hence Markov's inequality yields that $P_{T2n} \leq C_2\phi_{2n}/(\eta\tau_n)$ as well.

Now we bound P_{T3n} and then specify the choice of τ_n. Since $\mathcal{S}_0$ is compact, we can find m_n intervals centered at $s_1, \ldots, s_{m_n}$ with length C_S/m_n that cover $\mathcal{S}_0$ for some $C_S \in (0, \infty)$. We denote these intervals as $\mathcal{I}_k$ for $k = 1, \ldots, m_n$ and choose m_n later. The triangular inequality yields

$$\sup_{s\in\mathcal{S}_0}|T_n^\tau(\gamma;s)-\mathbb{E}[T_n^\tau(\gamma;s)]|\le T_{1n}^*+T_{2n}^*+T_{3n}^*,$$

where

$$\begin{aligned} T_{1n}^* &= \max_{1\le k\le m_n}\sup_{s\in\mathcal{I}_k}|T_n^\tau(\gamma;s)-T_n^\tau(\gamma;s_k)| \\ T_{2n}^* &= \max_{1\le k\le m_n}\sup_{s\in\mathcal{I}_k}|\mathbb{E}[T_n^\tau(\gamma;s)]-\mathbb{E}[T_n^\tau(\gamma;s_k)]| \\ T_{3n}^* &= \max_{1\le k\le m_n}|T_n^\tau(\gamma;s_k)-\mathbb{E}[T_n^\tau(\gamma;s_k)]|. \end{aligned}$$

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

$$Z_{n,i}^\tau(s)=(nb_n)^{-1}\left\{(c_0^\top x_i)^2\Delta_i(\gamma;s)K_i(s)\mathbf{1}_{\tau_n}-\mathbb{E}[(c_0^\top x_i)^2\Delta_i(\gamma;s)K_i(s)\mathbf{1}_{\tau_n}]\right\}$$

and

$$U_n(s)=T_n^\tau(\gamma;s)-\mathbb{E}[T_n^\tau(\gamma;s)]=\sum_{i=1}^n Z_{n,i}^\tau(s).$$

Note that $\sup_{s\in\mathcal{S}_0}|\left(c_0^\top x_i\right)^2\Delta_i(\gamma;s)K_i(s)\mathbf{1}_{\tau_n}|$ is bounded by $C_3\tau_n$ for some constant $C_3\in(0,\infty)$ and hence $\left|Z_{n,i}^\tau(s)\right|\le 2C_3\tau_n/(nb_n)$ for all $i=1,\dots,n$. Define $\lambda_n=(nb_n\log n)^{1/2}/\tau_n$. Then $\lambda_n\left|Z_{n,i}^\tau(s)\right|\le 2C_3(\log n/(nb_n))^{1/2}\le 1/2$ for all $i=1,\dots,n$ when n is sufficiently large. Using the inequality $\exp(v)\le 1+v+v^2$ for $|v|\le 1/2$, we have $\exp(\lambda_n\left|Z_{n,i}^\tau(s)\right|)\le 1+\lambda_n\left|Z_{n,i}^\tau(s)\right|+\lambda_n^2\left|Z_{n,i}^\tau(s)\right|^2$. Hence

$$\mathbb{E}[\exp(\lambda_n\left|Z_{n,i}^\tau(s)\right|)]\le 1+\lambda_n^2\mathbb{E}\left[(Z_{n,i}^\tau(s))^2\right]\le\exp\left(\lambda_n^2\mathbb{E}\left[(Z_{n,i}^\tau(s))^2\right]\right)\tag{B.16}$$

since $\mathbb{E}\left[Z_{n,i}^\tau(s)\right]=0$ and $1+v\le\exp(v)$ for $v\ge 0$. Using the fact that $\mathbb{P}(X>c)\le\mathbb{E}[\exp(Xa)]/\exp(ac)$ for any random variable X and nonrandom constants a and c, we have that

$$\begin{aligned} &\mathbb{P}\left(|U_n(s)|>\phi_{2n}^{1/2}\eta_n\right) \\ =\ &\mathbb{P}\left(\phi_{2n}^{-1/2}U_n(s)>\eta_n\right)+\mathbb{P}\left(-\phi_{2n}^{-1/2}U_n(s)>\eta_n\right) \\ \le\ &\frac{\mathbb{E}\left[\exp\left(\lambda_n\phi_{2n}^{-1/2}\sum_{i=1}^n Z_{n,i}^\tau(s)\right)\right]+\mathbb{E}\left[\exp\left(-\lambda_n\phi_{2n}^{-1/2}\sum_{i=1}^n Z_{n,i}^\tau(s)\right)\right]}{\exp(\lambda_n\eta_n)} \\ \le\ &2\exp(-\lambda_n\eta_n)\exp\left(\lambda_n^2\phi_{2n}^{-1}\sum_{i=1}^n\mathbb{E}\left[(Z_{n,i}^\tau(s))^2\right]\right)\qquad\text{(by (B.16))} \\ \le\ &2\exp(-\lambda_n\eta_n)\exp\left(\lambda_n^2C_4\tau_n^2/(nb_n)\right) \end{aligned}$$

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

$$\mathbb{E}\left[(Z_{n,i}^\tau(s))^2\right]\le(nb_n)^{-2}\mathbb{E}\left[\left(c_0^\top x_i\right)^4\Delta_i(\gamma;s)^2K_i^2(s)\mathbf{1}_{\tau_n}\right]\le C_4\tau_n^2(n^2b_n)^{-1}\phi_{2n}(1+o(1))$$

for some $C_4 \in (0, \infty)$. However, this bound is independent of s given Assumptions A-(v) and (x), and hence it is also the uniform bound, i.e.,

$$\sup_{s \in \mathcal{S}_0} \mathbb{P}\left(|U_n(s)| > \phi_{2n}^{1/2} \eta_n\right) \leq 2 \exp\left(-\lambda_n \eta_n + \lambda_n^2 C_4 \tau_n^2 / (n b_n)\right). \tag{B.17}$$

Now given τ_n, we need to choose $\eta_n \to 0$ as fast as possible, and at the same time we let $\lambda_n \eta_n \to \infty$ at a rate that ensures (B.17) is summable and $\lambda_n \eta_n > \lambda_n^2 \tau_n^2 / (n b_n)$. This is done by choosing $\lambda_n = (n b_n \log n)^{1/2} / \tau_n$ and $\eta_n = C^* \lambda_n^{-1} \log n = C^* \tau_n ((\log n) / (n b_n))^{1/2}$ for some finite constant C^*. This choice yields

$$-\lambda_n \eta_n + \lambda_n^2 C_4 \tau_n^2 / n b_n = -C^* \log n + C_4 \log n = -(C^* - C_4) \log n.$$

Therefore, by substituting this into (B.17), we have

$$\begin{aligned}
\mathbb{P}\left(T_{3n}^* > \phi_{2n}^{1/2} \eta_n\right) &= \mathbb{P}\left(\max_{1 \leq k \leq m_n} |U_n(s_k)| > \phi_{2n}^{1/2} \eta_n\right) \\
&\leq m_n \sup_{s \in \mathcal{S}_0} \mathbb{P}\left(|U_n(s)| > \phi_{2n}^{1/2} \eta_n\right) \leq 2 \frac{m_n}{n^{C^* - C_4}}.
\end{aligned}$$

Now, we can choose C^* sufficiently large so that $\sum_{n=1}^{\infty} \mathbb{P}\left(T_{3n}^* > \phi_{2n}^{1/2} \eta_n\right)$ is summable, from which we have

$$T_{3n}^* = O_{a.s.}(\phi_{2n}^{1/2} \eta_n) = O_{a.s.}\left(\left(\phi_{2n} \frac{\log n}{n b_n}\right)^{1/2}\right)$$

by the Borel-Cantelli lemma.

Next, we consider T_{1n}^*. Note that

$$\begin{aligned}
T_n^{\tau}(\gamma; s) - T_n^{\tau}(\gamma; s_k) &= \frac{1}{n b_n} \sum_{i=1}^{n} \left(c_0^{\top} x_i\right)^2 \Delta_i(\gamma; s) \left(K_i(s) - K_i(s_k)\right) \mathbf{1}_{\tau_n} \\
&\quad + \frac{1}{n b_n} \sum_{i=1}^{n} \left(c_0^{\top} x_i\right)^2 \left(\Delta_i(\gamma; s) - \Delta_i(\gamma; s_k)\right) K_i(s_k) \mathbf{1}_{\tau_n}.
\end{aligned} \tag{B.18}$$

For the first item in (B.18), using a similar derivation as Lemma A.6 yields that if n is sufficiently large,

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{n b_n} \sum_{i=1}^{n} \left(c_0^{\top} x_i\right)^2 \Delta_i(\gamma; s) \left(K_i(s) - K_i(s_k)\right) \mathbf{1}_{\tau_n}\right|\right] \\
\leq\ & b_n^{-1} \tau_n \mathbb{E}\left[\left|\Delta_i(\gamma; s) \left(K_i(s) - K_i(s_k)\right)\right|\right] \\
\leq\ & C_5 C_S \tau_n \phi_{2n} / (m_n b_n).
\end{aligned}$$

for some constant $C_5 < \infty$. For the second item in (B.18), without loss of generality, consider that $\gamma(s) < \gamma(s_k)$ and $\gamma_0(s) < \gamma_0(s_k)$. Then by choosing the covering interval length C_S / m_n smaller than ϕ_{2n}, we have

$$\mathbb{E}\left[\sup_{s \in \mathcal{I}_k} \left|\frac{1}{n b_n} \sum_{i=1}^{n} \left(c_0^{\top} x_i\right)^2 \left(\Delta_i(\gamma; s) - \Delta_i(\gamma; s_k)\right) K_i(s_k) \mathbf{1}_{\tau_n}\right|\right]$$

$$
\begin{aligned}
&\leq\ 2C_6\tau_n\left(\sup_{s\in\mathcal{S}_0}K(s)\right)\mathbb{E}\left[\sup_{s\in\mathcal{I}_k}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left(\gamma_0(s)<q_i\leq\gamma_0(s_k)\right)\right|\right.\\
&\qquad\qquad\qquad\left.+\sup_{s\in\mathcal{I}_k}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left(\gamma(s)<q_i\leq\gamma(s_k)\right)\right|\right]\\
&\leq\ C_6\tau_n\mathbb{P}\left(\inf_{s\in\mathcal{I}_k}\gamma_0(s)<q_i\leq\sup_{s\in\mathcal{I}_k}\gamma_0(s)\right)+C_6\tau_n\mathbb{P}\left(\inf_{s\in\mathcal{I}_k}\gamma(s)<q_i\leq\sup_{s\in\mathcal{I}_k}\gamma(s)\right)\\
&\leq\ C_6C_S\tau_n/m_n,
\end{aligned}
$$

where the last line follows from Taylor expansion and Assumption A-(vi). This bound does not depend on k and hence $T_{1n}^{*}=O_p(\tau_n/m_n)$. Similarly for T_{2n}^{*}, Taylor expansion yields that

$$
\begin{aligned}
\left|\mathbb{E}\left[T_n^{\tau}(\gamma;s)\right]-\mathbb{E}\left[T_n^{\tau}(\gamma;s_k)\right]\right| &\leq b_n^{-1}\tau_n\mathbb{E}\left[\Delta_i(\gamma;s)K_i(s)-\Delta_i(\gamma,s_k)K_i(s_k)\right]\\
&\leq b_n^{-1}\tau_n\mathbb{E}\left[\Delta_i(\gamma;s)\left(K_i(s)-K_i(s_k)\right)\right]\\
&\quad+b_n^{-1}\tau_n\mathbb{E}\left[\left(c_0^{\top}x_i\right)^2\left(\Delta_i(\gamma;s)-\Delta_i(\gamma;s_k)\right)K_i(s_k)\right]\\
&\leq C_7\tau_n/m_n
\end{aligned}
$$

for some $C_7<\infty$, where the last line follows by choosing the covering interval length C_S/m_n smaller than ϕ_{2n}. This bound is also uniform in k and hence $T_{2n}^{*}=O(\tau_n/m_n)$ as well. Therefore, by choosing $m_n=[(\phi_{2n}(\log n)/nb_n)^{1/2}/\tau_n]^{-1}$, we have that T_{1n}^{*} and T_{2n}^{*} are both the order of $(\phi_{2n}(\log n)/nb_n)^{1/2}$. It follows that $P_{T3n}\leq\eta^{-1}C(\phi_{2n}(\log n)/nb_n)^{1/2}$ for some $C\in(0,\infty)$ by Markov's inequality.

Finally, if we choose τ_n such that $\tau_n=O(\phi_{2n}^{1/2}((\log n)/nb_n)^{-1/2})$, we have both P_{T1n} and P_{T2n} are also bounded by $\eta^{-1}C(\phi_{2n}(\log n)/nb_n)^{1/2}$. A possible choice of τ_n is n^{ϵ} or larger. This completes the proof. ■

Proof of Lemma A.12 Since the proof is similar as that in Lemma A.11, we only highlight the different part. We only present the argument for $L_n(\gamma;s)$ as the proof for $\overline{L}_n(\gamma;s)$ is identical. We now define $\mathbf{1}_{\tau_n}=\mathbf{1}[\left|c_0^{\top}x_iu_i\right|<\tau_n]$ for some truncation parameter satisfying $\tau_n\to\infty$ as $n\to\infty$, which can be different from the one chosen in Lemma A.11 above. We let

$$
L_n^{\tau}(\gamma;s)=\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{n}c_0^{\top}x_iu_i\Delta_i(\gamma;s)K_i(s)\,\mathbf{1}_{\tau_n},
$$

and write

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{s\in\mathcal{S}_0}\left|L_n(\gamma;s)\right|>\eta\right)\\
\leq\ &\mathbb{P}\left(\sup_{s\in\mathcal{S}_0}\left|L_n^{\tau}(\gamma;s)-L_n(\gamma;s)\right|>\eta/2\right)+\mathbb{P}\left(\sup_{s\in\mathcal{S}_0}\left|L_n^{\tau}(\gamma;s)\right|>\eta/2\right)\\
\equiv\ &P_{L1n}+P_{L2n},
\end{aligned}
$$

where $\mathbb{E}[L_n^\tau(\gamma, s)] = 0$.

To bound P_{L1n}, similarly as P_{T1n} in the proof of Lemma A.11, note that

$$\begin{aligned}
& \mathbb{E}\left[\sup_{s\in\mathcal{S}_0} \left|L_n^\tau(\gamma; s) - L_n(\gamma; s)\right|\right] \\
\leq\ & \mathbb{E}\left[\frac{1}{\sqrt{nb_n}}\sum_{i=1}^n \left|c_0^\top x_i u_i\right| \mathbf{1}\left[\inf_{s\in\mathcal{S}_0}\gamma_0(s) \leq q_i \leq \sup_{s\in\mathcal{S}_0}\gamma_0(s) + \overline{r}\phi_{2n}\right] K_i(s)\left(1-\mathbf{1}_{\tau_n}\right)\right] \\
\leq\ & (nb_n)^{1/2}\tau_n^{-1}\int_{\inf_{s\in\mathcal{S}_0}\gamma_0(s)}^{\sup_{s\in\mathcal{S}_0}\gamma_0(s)+\overline{r}\phi_{2n}} \mathbb{E}\left[\left(c_0^\top x_i u_i\right)^2 |q, s+tb_n\right] f(q, s+tb_n) K(t)\, dq dt \\
\leq\ & C_1\phi_{2n}(nb_n)^{1/2}\tau_n^{-1}
\end{aligned}$$

for some $C_1 \in (0,\infty)$ and hence $P_{L1n} \leq \eta^{-1}C_1\phi_{2n}(nb_n)^{1/2}\tau_n^{-1}$ by Markov's inequality.

To bound P_{L2n}, similarly as P_{T3n} in the proof of Lemma A.11, we write

$$\sup_{s\in\mathcal{S}_0}\left|L_n^\tau(\gamma; s)\right| \leq L_{1n}^* + L_{2n}^*,$$

where

$$\begin{aligned}
L_{1n}^* &= \max_{1\leq k\leq m_n}\sup_{s\in\mathcal{I}_k}\left|L_n^\tau(\gamma; s) - L_n^\tau(\gamma; s_k)\right| \\
L_{2n}^* &= \max_{1\leq k\leq m_n}\left|L_n^\tau(\gamma; s_k)\right|
\end{aligned}$$

and $\{\mathcal{I}_k\}_{k=1}^{m_n}$ denote m_n intervals centered at $s_1, \ldots, s_{m_n}$ with length C_S/m_n that cover $\mathcal{S}_0$ for some $C_S \in (0,\infty)$. (The choices of m_n and C_S can be different from the ones in Lemma A.11 above.) The bound of L_{1n}^* can be obtained similarly as T_{3n}^* above by letting $Z_{n,i}^\tau(s) = (nb_n)^{-1/2} c_0^\top x_i u_i \Delta_i(\gamma; s) K_i(s)\mathbf{1}_{\tau_n}$. In particular, with $\left|Z_{n,i}^\tau(s)\right| \leq C_2\tau_n/(nb_n)^{1/2}$ for all $i = 1, \ldots, n$ and $L_n^\tau(\gamma; s) = \sum_{i=1}^n Z_{n,i}^\tau(s)$, we have

$$\sup_{s\in\mathcal{S}_0}\mathbb{P}\left(\left|L_n^\tau(\gamma; s)\right| > \phi_{2n}^{1/2}\eta_n\right) \leq 2\exp(-\lambda_n\eta_n + \lambda_n^2\tau_n^2 C_3) \tag{B.19}$$

for some $C_3 \in (0,\infty)$. By choosing $\lambda_n = (\log n)^{1/2}/\tau_n$ and $\eta_n = C^*\tau_n(\log n)^{1/2}$ for some finite constant C^*, we get

$$-\lambda_n\eta_n + \lambda_n^2\tau_n^2 C_3 = -(C^* - C_3)\log n.$$

Substituting this into (B.19) gives us

$$\sup_{s\in\mathcal{S}_0}\mathbb{P}\left(\left|L_n^\tau(\gamma; s)\right| > \phi_{2n}^{1/2}\eta_n\right) \leq 2\frac{m_n}{n^{C^*-C_3}},$$

and hence by choosing C^* sufficiently large

$$L_{2n}^* = O_{a.s.}(\phi_{2n}^{1/2}\eta_n) = O_{a.s.}\left((\phi_{2n}\log n)^{1/2}\right)$$

by the Borel-Cantelli lemma. Regarding L_{1n}^*, we choose $m_n = [(\phi_{2n}\log n)^{1/2}/\tau_n]^{-1}$ and use the same argument as bounding T_{1n}^* above to get

$$\mathbb{E}\left[L_{1n}^*\right] = O\left((\phi_{2n}\log n)^{1/2}\right).$$

Therefore, by combining L_{1n}^* and L_{2n}^* and using Markov's inequality, we have $P_{L2n} \leq \eta^{-1}C(\phi_{2n}\log n)^{1/2}$ for some $C \in (0,\infty)$.

Finally, if we choose τ_n such that $\tau_n = O(\phi_{2n}^{1/2}((\log n)/(nb_n))^{-1/2})$, we have $P_{L1n} \leq \eta^{-1}C(\phi_{2n}\log n)^{1/2}$ as well. A possible choice of τ_n is n^{ϵ} or larger. This completes the proof. ■

Proof of Lemma A.13 We first show (A.33). Consider the case with $\gamma(s) - \gamma_0(s) \in [r(s)\phi_{2n}, C(s)]$, where $0 < \underline{r} = \inf_{s\in\mathcal{S}_0} r(s) \leq \sup_{s\in\mathcal{S}_0} r(s) = \overline{r} < \infty$ and $\overline{C} = \sup_{s\in\mathcal{S}_0} C(s) < \infty$; the other direction can be shown symmetrically. Let

$$\underline{\ell}_D(s) = \inf_{|\gamma(s)-\gamma_0(s)|<\overline{C}(s)} c_0^\top D(\gamma(s),s)c_0 f(\gamma(s),s) > 0 \quad \text{and} \quad \underline{\ell} = \inf_{s\in\mathcal{S}_0} \underline{\ell}_D(s) > 0$$

from Assumptions A-(vii) and (viii). Then, from (B.1), we get

$$\begin{aligned}\sup_{s\in\mathcal{S}_0} \mathbb{E}\left[T_n(\gamma;s)\right] &\geq \sup_{s\in\mathcal{S}_0} \left(\gamma(s)-\gamma_0(s)\right)\left(\underline{\ell} + C_1(s)b_n^2\right) \\ &\geq \underline{\ell}\sup_{s\in\mathcal{S}_0} \left(\gamma(s)-\gamma_0(s)\right) = \underline{\ell}\overline{r}\phi_{2n}\end{aligned} \tag{B.20}$$

because $0 < C_1(s) < \infty$ for all $s \in \mathcal{S}_0$ from Assumptions A-(vii) and (viii). Furthermore, Lemma A.11 implies that

$$\mathbb{P}\left(\sup_{s\in\mathcal{S}_0} \left|T_n(\gamma;s) - \mathbb{E}\left[T_n(\gamma;s)\right]\right| > \eta\right) \leq C_2\eta^{-1}\left(\phi_{2n}\frac{\log n}{nb_n}\right)^{1/2} \tag{B.21}$$

for some $C_2 \in (0,\infty)$.

We now set γ_g for $g = 1,\ldots,\overline{g}+1$ such that, for any $s \in \mathcal{S}_0$, $\gamma_g(s) = \gamma_0(s) + 2^{g-1}r(s)\phi_{2n}$ where $\overline{g}$ is the integer satisfying $\gamma_{\overline{g}}(s) - \gamma_0(s) = 2^{\overline{g}-1}r(s)\phi_{2n} \leq \overline{C}$ and $\gamma_{\overline{g}+1}(s) - \gamma_0(s) = 2^{\overline{g}}r(s)\phi_{2n} > \overline{C}$. Then, (B.20) and (B.21) yield that for any fixed $\eta > 0$,

$$\begin{aligned}&\mathbb{P}\left(\max_{1\leq g\leq\overline{g}}\left|\frac{\sup_{s\in\mathcal{S}_0} T_n\left(\gamma_g;s\right)}{\sup_{s\in\mathcal{S}_0}\mathbb{E}\left[T_n\left(\gamma_g;s\right)\right]} - 1\right| > \eta\right) \qquad \text{(B.22)}\\ \leq\ &\mathbb{P}\left(\max_{1\leq g\leq\overline{g}}\frac{\left|\sup_{s\in\mathcal{S}_0} T_n\left(\gamma_g;s\right) - \sup_{s\in\mathcal{S}_0}\mathbb{E}\left[T_n\left(\gamma_g;s\right)\right]\right|}{\left|\sup_{s\in\mathcal{S}_0}\mathbb{E}\left[T_n\left(\gamma_g;s\right)\right]\right|} > \eta\right)\\ \leq\ &\mathbb{P}\left(\max_{1\leq g\leq\overline{g}}\frac{\sup_{s\in\mathcal{S}_0}\left|T_n\left(\gamma_g;s\right) - \mathbb{E}\left[T_n\left(\gamma_g;s\right)\right]\right|}{\left|\sup_{s\in\mathcal{S}_0}\mathbb{E}\left[T_n\left(\gamma_g;s\right)\right]\right|} > \eta\right)\\ \leq\ &\sum_{g=1}^{\overline{g}}\mathbb{P}\left(\sup_{s\in\mathcal{S}_0}\left|T_n\left(\gamma_g;s\right) - \mathbb{E}\left[T_n\left(\gamma_g;s\right)\right]\right| > \eta\left|\sup_{s\in\mathcal{S}_0}\mathbb{E}\left[T_n\left(\gamma_g;s\right)\right]\right|\right)\end{aligned}$$

$$
\begin{aligned}
&\le \sum_{g=1}^{\bar{g}} \frac{C_1\left(\phi_{2n}(\log n)/nb_n\right)^{1/2}}{2^{g-1}\eta\underline{\ell}\overline{r}\phi_{2n}} \\
&\le \frac{C_1}{\eta\underline{\ell}\overline{r}} \sum_{g=1}^{\infty} \frac{1}{2^{g-1}} \times \frac{1}{n^{\epsilon}} \\
&\le \varepsilon
\end{aligned}
$$

for any $\varepsilon > 0$. Then from eq. (33) of Hansen (2000), for any $\gamma(s)$ such that $\overline{r}\phi_{2n} \le \sup_{s\in\mathcal{S}_0}(\gamma(s)-\gamma_0(s)) \le \overline{C}$, there exists some g such that $\gamma_g(s)-\gamma_0(s) < \gamma(s)-\gamma_0(s) < \gamma_{g+1}(s)-\gamma_0(s)$. This implies that

$$
\begin{aligned}
& \frac{\sup_{s\in\mathcal{S}_0} T_n(\gamma;s)}{\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|} \\
\ge\ & \frac{\sup_{s\in\mathcal{S}_0} T_n(\gamma_g;s)}{\sup_{s\in\mathcal{S}_0}\mathbb{E}\left[T_n(\gamma_g;s)\right]} \times \frac{\sup_{s\in\mathcal{S}_0}\mathbb{E}\left[T_n(\gamma_g;s)\right]}{\sup_{s\in\mathcal{S}_0}\left|\gamma_{g+1}(s)-\gamma_0(s)\right|} \\
=\ & \left(1+\left(\frac{\sup_{s\in\mathcal{S}_0} T_n(\gamma_g;s)}{\sup_{s\in\mathcal{S}_0}\mathbb{E}\left[T_n(\gamma_g;s)\right]}-1\right)\right) \times \frac{\sup_{s\in\mathcal{S}_0}\mathbb{E}\left[T_n(\gamma_g;s)\right]}{\sup_{s\in\mathcal{S}_0}\left|\gamma_{g+1}(s)-\gamma_0(s)\right|},
\end{aligned}
$$

and for any $\eta > 0$,

$$
\begin{aligned}
& \mathbb{P}\left(\inf_{\overline{r}\phi_{2n}<\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|<\overline{C}} \frac{\sup_{s\in\mathcal{S}_0} T_n(\gamma;s)}{\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)|} < C(1-\eta)\right) \\
\le\ & \mathbb{P}\left(\left(1-\left|\max_{1\le g\le\bar{g}} \frac{\sup_{s\in\mathcal{S}_0} T_n(\gamma_g;s)}{\sup_{s\in\mathcal{S}_0}\mathbb{E}\left[T_n(\gamma_g;s)\right]}-1\right|\right)\frac{\sup_{s\in\mathcal{S}_0}\mathbb{E}\left[T_n(\gamma_g;s)\right]}{\sup_{s\in\mathcal{S}_0}\left|\gamma_{g+1}(s)-\gamma_0(s)\right|} < C(1-\eta)\right) \\
\le\ & \varepsilon,
\end{aligned}
$$

where the last line follow from (B.20) and (B.22). The proof for (A.34) is similar to that for (A.33) and hence omitted.

For (A.18), Lemma A.12 yields that, for a large enough n,

$$
\mathbb{P}\left(\sup_{s\in\mathcal{S}_0}|L_n(\gamma;s)| > \eta\right) \le \eta^{-1}C_2\phi_{2n}^{1/2}(\log n)^{1/2} \tag{B.23}
$$

for some $C_2 \in (0,\infty)$ similarly as above. Using a similar approach as (B.22), for any fixed $\eta > 0$,

$$
\begin{aligned}
& \mathbb{P}\left(\max_{1\le g\le\bar{g}} \frac{\sup_{s\in\mathcal{S}_0}\left|L_n(\gamma_g;s)\right|}{\sqrt{a_n}\sup_{s\in\mathcal{S}_0}\left(\gamma_g(s)-\gamma_0(s)\right)} > \eta\right) \qquad \text{(B.24)} \\
\le\ & \sum_{g=1}^{\infty}\mathbb{P}\left(\frac{\sup_{s\in\mathcal{S}_0}\left|L_n(\gamma_g;s)\right|}{\sqrt{a_n}\sup_{s\in\mathcal{S}_0}\left(\gamma_g(s)-\gamma_0(s)\right)} > \eta\right) \\
\le\ & \sum_{g=1}^{\infty} \frac{C_2\left(\phi_{2n}\log n\right)^{1/2}}{\eta\sqrt{a_n}2^{g-1}\underline{\mu}\overline{r}\phi_{2n}}
\end{aligned}
$$

$$\leq \frac{C_2}{\eta \underline{\mu} \overline{r}} \sum_{g=1}^{\infty} \frac{1}{2^{g-1}}.$$

from (B.20) and (B.23). This probability is arbitrarily close to 0 if $\overline{r}$ is large enough. Following a similar discussion after (B.5), this result also provides the maximal (or sharp) rate of ϕ_{2n} as $\log n/a_n$ because we need $(\log n/a_n)/\phi_{2n} = O(1)$ but $\phi_{2n} \to 0$ as $\log n/a_n \to 0$ with $n \to \infty$.

Finally, for a given g, we define Γ_g as the collection of $\gamma(s)$ satisfying $\overline{r}2^{g-1}\phi_{2n} < \gamma(s) - \gamma_0(s) < \overline{r}2^g\phi_{2n}$ for all $s \in S$. By a similar argument as (B.24), we have

$$\mathbb{P}\left(\max_{1 \leq g \leq \overline{g}} \sup_{\gamma \in \Gamma_g} \frac{\sup_{s \in \mathcal{S}_0} |L_n(\gamma; s)|}{\sqrt{a_n} \sup_{s \in \mathcal{S}_0} (\gamma(s) - \gamma_0(s))} > \eta \right) \leq \frac{C_3}{\eta \overline{r}} \tag{B.25}$$

for some constant $C_3 < \infty$. Combining (B.24) and (B.25), we thus have

$$\begin{aligned}
& \mathbb{P}\left(\sup_{\overline{r}\phi_{2n} < \sup_{s \in \mathcal{S}_0} |\gamma(s) - \gamma_0(s)| < \overline{C}} \frac{\sup_{s \in \mathcal{S}_0} |L_n(\gamma_g; s)|}{\sqrt{a_n} \sup_{s \in \mathcal{S}_0} (\gamma(s) - \gamma_0(s))} > \eta \right) \\
\leq\ & 2\mathbb{P}\left(\max_{1 \leq g \leq \overline{g}} \frac{\sup_{s \in \mathcal{S}_0} |L_n(\gamma_g; s)|}{\sqrt{a_n} \sup_{s \in \mathcal{S}_0} (\gamma_g(s) - \gamma_0(s))} > \eta \right) \\
& + 2\mathbb{P}\left(\max_{1 \leq g \leq \overline{g}} \sup_{\gamma \in \Gamma_g} \frac{\sup_{s \in \mathcal{S}_0} |L_n(\gamma; s)|}{\sqrt{a_n} \sup_{s \in \mathcal{S}_0} (\gamma(s) - \gamma_0(s))} > \eta \right) \\
\leq\ & \varepsilon
\end{aligned}$$

for any $\varepsilon > 0$ if $\overline{r}$ is sufficiently large. The proof for (A.36) is similar to that for (A.35) and hence omitted. ■

Proof of Lemma A.14 For a given γ, since all the convergence results in Lemma A.5 hold uniformly by Lemma A.1, we only need to show $\sup_{s \in \mathcal{S}_0} |\widehat{\gamma}(s) - \gamma_0(s)| \to_p 0$. To this end, denote $\overline{\Gamma}$ and $\underline{\Gamma}$ as the upper and lower bounds of Γ, respectively, and let $d_\Gamma = \overline{\Gamma} - \underline{\Gamma}$. Since $\mathcal{S}_0$ is compact, it can be covered by the union of a finite number of intervals $\{\mathcal{I}_k\}_{k=1}^m$ with length d_Γ/m and center points $\{s_k\}_{k=1}^m$. On the event E_n^* that $\widehat{\gamma}(s)$ is continuous with probability approaching to one, we can choose a large m such that $\sup_{s \in \mathcal{I}_k} |\widehat{\gamma}(s) - \widehat{\gamma}(s_k)| \leq \eta$ for any η and all k. Such a choice is also valid for $\gamma_0(\cdot)$ since it is also continuous by Assumption A-(vi). Then on the event E_n^*, using triangular inequality and Lemma A.3, for any $\eta > 0$ and any $\varepsilon > 0$, there is a large enough m such that

$$\begin{aligned}
& \mathbb{P}\left(\sup_{s \in \mathcal{S}_0} |\widehat{\gamma}(s) - \gamma_0(s)| > \eta \right) \\
\leq\ & \mathbb{P}\left(\max_{1 \leq k \leq m} \sup_{s \in \mathcal{I}_k} |\widehat{\gamma}(s) - \widehat{\gamma}(s_k)| > \eta/3 \right) + \mathbb{P}\left(\max_{1 \leq k \leq m} \sup_{s \in \mathcal{I}_k} |\gamma_0(s) - \gamma_0(s_k)| > \eta/3 \right) \\
& + \mathbb{P}\left(\max_{1 \leq k \leq m} |\widehat{\gamma}(s_k) - \gamma_0(s_k)| > \eta/3 \right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \quad 2\left(1-\mathbb{P}(E_n^*)\right)+\sum_{k=1}^{m}\mathbb{P}\left(|\widehat{\gamma}(s_k)-\gamma_0(s_k)|>\eta/3\right)\\
&\leq \quad \varepsilon,
\end{aligned}
$$

where the last line follows from that $\mathbb{P}(E_n^*) > 1-\varepsilon$ for any ε. This is because $\widehat{\gamma}(\cdot)$ is a step function taking values in $\{q_i\}_{i=1}^n \cap \Gamma$ and hence is piecewise continuous with countable jump points. ■

Proof of Lemma A.15 We prove $\Xi_{n02} = o_p(1)$ and $\Xi_{n03} = o_p(1)$. The results for Ξ_{n12} and Ξ_{n13} can be shown symmetrically. As in the proof of Theorem 5, we denote the leave-one-out estimator $\widehat{\gamma}_{-i}(s_i)$ as $\widehat{\gamma}(s_i)$ in this proof. For expositional simplicity, we only present the case of scalar x_i.

First, for any continuous function $\gamma(\cdot): \mathcal{S} \to \Gamma$, we define

$$
G_n(\gamma) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i \mathbf{1}\left[q_i > \gamma(s_i)+\Delta_n\right]\mathbf{1}_{\mathcal{S}_0}.
$$

For any fixed $\gamma(\cdot)$, $G_n(\gamma)$ converges to a Gaussian random variable by the random field CLT, where $\mathbb{E}\left[x_i u_i \mathbf{1}\left[q_i > \gamma(s_i)+\Delta_n\right]\mathbf{1}_{\mathcal{S}_0}\right] = 0$ and $\mathbb{E}\left[x_i^2 u_i^2 \mathbf{1}\left[q_i > \gamma(s_i)+\Delta_n\right]\mathbf{1}_{\mathcal{S}_0}\right] < \infty$ from Assumptions ID-(i) and A-(v). Moreover, the convergence holds for any finite collection of $\gamma(\cdot)$ and the process $G_n(\gamma)$ is uniformly tight by a similar argument as Lemma A.1. Therefore, we have $G_n(\gamma) \Rightarrow \mathbb{G}(\gamma)$ as $n \to \infty$, where $\mathbb{G}(\gamma)$ is a Gaussian process with almost surely continuous paths (cf. Lemma A.4 in Hansen (2000)). It follows that, for any $\gamma(s)$ such that $\sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)| \leq \overline{r}\phi_{2n}$ for some $\overline{r} > 0$, we have

$$
G_n(\gamma) - G_n(\gamma_0) \to_p 0
$$

as $G_n(\gamma) - G_n(\gamma_0) \Rightarrow \mathbb{G}(\gamma) - \mathbb{G}(\gamma_0)$. We now denote $\overline{\Gamma}_n$ as the set of continuous functions $\{\gamma(\cdot) : \sup_{s\in\mathcal{S}_0}|\gamma(s)-\gamma_0(s)| \leq \overline{r}\phi_{2n}\}$. If we choose $\overline{r}$ large enough so that $\mathbb{P}(\widehat{\gamma} \notin \overline{\Gamma}_n) < \varepsilon/2$, then for any $\varepsilon > 0$ and $\eta > 0$, we have

$$
\begin{aligned}
& \quad \mathbb{P}\left(|\Xi_{n02}| > \eta\right)\\
&= \quad \mathbb{P}\left(|G_n(\widehat{\gamma}) - G_n(\gamma_0)| > \eta\right)\\
&= \quad \mathbb{P}\left(|G_n(\widehat{\gamma}) - G_n(\gamma_0)| > \eta \text{ and } \widehat{\gamma} \in \overline{\Gamma}_n\right) + \mathbb{P}\left(|G_n(\widehat{\gamma}) - G_n(\gamma_0)| > \eta \text{ and } \widehat{\gamma} \in \overline{\Gamma}_n^c\right)\\
&\leq \quad \mathbb{P}\left(\sup_{\gamma\in E_{n\gamma}}|G_n(\gamma) - G_n(\gamma_0)| > \eta\right) + \mathbb{P}(\widehat{\gamma} \notin \overline{\Gamma}_n)\\
&\leq \quad \varepsilon,
\end{aligned}
$$

which gives the desired result.

Second, we consider $\Delta_n > 0$. On the event E_n^* that $\sup_{s\in\mathcal{S}_0}|\widehat{\gamma}(s)-\gamma_0(s)| \leq \phi_{2n}$, we have

$$
\mathbb{E}\left[|\Xi_{n03}|\right] \quad = \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}\left[\left|x_i^2\delta_0\right|\mathbf{1}\left[q_i \leq \gamma_0(s_i)\right]\mathbf{1}\left[q_i > \widehat{\gamma}(s_i)+\Delta_n\right]\mathbf{1}_{\mathcal{S}_0}\right]
$$

$$
\begin{aligned}
&\leq n^{1/2-\epsilon} C \mathbb{E}\left[\mathbf{1}\left[q_i \leq \gamma_0(s_i)\right] \mathbf{1}\left[q_i > \widehat{\gamma}(s_i) + \Delta_n\right] \mathbf{1}_{\mathcal{S}_0}\right] \\
&\leq n^{1/2-\epsilon} C \mathbb{E}\left[\mathbf{1}\left[q_i \leq \gamma_0(s_i)\right] \mathbf{1}\left[q_i > \gamma_0(s_i) - \phi_{2n} + \Delta_n\right] \mathbf{1}_{\mathcal{S}_0}\right] \\
&= n^{1/2-\epsilon} C \int_{\mathcal{S}_0} \int_{\mathcal{I}(q;s)} f(q,s) dq ds
\end{aligned}
$$

for some constant $0 < C < \infty$, where $\mathcal{I}(q;s) = \{q : q \leq \gamma_0(s) \text{ and } q > \gamma_0(s) - \phi_{2n} + \Delta_n\}$. However, since we set $\Delta_n > 0$ such that $\phi_{2n}/\Delta_n \to 0$, then $\Delta_n - \phi_{2n} > 0$ holds with a sufficiently large n. Therefore, $\mathcal{I}(q;s)$ becomes empty for all s when n is sufficiently large. The desired result follows from Markov's inequality and the fact that $\mathbb{P}(E_n^*) > 1 - \varepsilon$ for any $\varepsilon > 0$. ■

S.2 For Non-random q_i and s_i

The main analysis of the paper assumes that $(q_i, s_i)^\top$ are continuous random variables. It can be easily modified to cover the case where q_i and s_i are non-random integer indices. To fix idea, consider the metropolitan area determination problem where q_i and s_i denote the latitude and longitude, respectively, on an equi-spaced grid in $\mathbb{N}^2$. Denote n_1 and n_2 as the numbers of elements in the latitudes and longitude so that $n = n_1 \times n_2$ is the total sample size. Without loss of generality, we normalize q_i and s_i so that $q_i \in \{1/n_1, 2/n_1, \ldots, 1\}$ and $s_i \in \{1/n_2, 2/n_2, \ldots, 1\}$. We claim that under the following conditions, which are simplified version of Assumptions ID and A, key results in the main context remain unchanged if we treat $(q_i, s_i)^\intercal$ as if they were uniformly distributed over $[0,1]^2$. Accordingly, the density f in Theorem 3 is simply 1. Note that, under strict stationarity, the conditional moments $D(\cdot)$ and $V(\cdot)$ are simplified as $\bar{D} = \mathbb{E}[x_i x_i^\intercal]$ and $\bar{V} = \mathbb{E}\left[x_i x_i^\intercal u_i^2\right]$, respectively.

Assumption ID′

(i) $\mathbb{E}[u_i x_i] = 0$.

(ii) $\mathbb{E}\left[x_i x_i^\top\right] > \mathbb{E}\left[x_i x_i^\top \mathbf{1}\left[q_i \leq \gamma\right]\right] > 0$ *for any* $\gamma \in \Gamma$.

(iii) $(\beta_0^\intercal, \delta_0^\intercal)^\intercal$ *are in the interior of some compact subsets of* $\mathbb{R}^{2p}$.

(iv) $\gamma_0(s)$ *is in the interior of* Γ *for all* $s \in \mathcal{S}$, *where* Γ is a compact subset of $(0,1)$, *and* $\delta_0 \neq 0$.

Assumption A′

(i) The lattice $N_n \subset \mathbb{R}^2$ *is infinite countable; all the elements in* N_n *are located at distances at least* $\lambda_0 > 1$ *from each other, i.e., for any* $i, j \in N_n : \lambda(i,j) \geq \lambda_0$; *and* $\lim_{n\to\infty} |\partial N_n| / n = 0$. $\lim_{n_1 \to \infty, n \to \infty} n_1/n \in (0,1)$.

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

(iii) $\left(x_i^\top, u_i\right)^\top$ *is strictly stationary and α-mixing with bounded $(2+\varphi)$th moments for some $\varphi > 0$; the mixing coefficient $\alpha(m)$ defined in (7) satisfies $\sum_{m=1}^{\infty} m\alpha(m) < \infty$ and $\sum_{m=1}^{\infty} m^2\alpha(m)^{\varphi/(2+\varphi)} < \infty$ for some $\varphi \in (0,2)$.*

(iv) $0 < \mathbb{E}\left[u_i^2|x_i\right] < \infty$ *almost surely.*

(v) $\gamma_0 : \mathcal{S} \mapsto \Gamma$ *is a twice continuously differentiable function with bounded derivatives.*

(vi) $c_0^\top \bar{D} c_0 > 0$, $c_0^\top \bar{V} c_0 > 0$.

(vii) As $n \to \infty$, $b_n \to 0$ *and* $n^{1-2\epsilon} b_n \to \infty$.

(viii) $K(\cdot)$ *is uniformly bounded, continuous, symmetric around zero, and satisfies* $\int K(v)\,dv = 0$, $\int v^2 K(v)\,dv \in (0,\infty)$, $\int K^2(v)dv \in (0,\infty)$, *and* $\lim_{v\to\infty} |v|K(v) = 0$.

We first establish the identification.

Theorem 1′ *Under Assumption ID′, the threshold function $\gamma_0(\cdot)$ and the parameters $(\beta_0^\top, \delta_0^\top)^\top$ are uniquely identified.*

Proof of Theorem 1′ The proof is very similar to that of Theorem 1. First, since q_i and s_i are non-random and take values on equally-spaced grids on $[0,1]$, we can treat them as independently multinomial distributed random variables. Then asymptotically, q_i and s_i are independent and standard uniformly distributed over $[0,1]^2$. Then, the case (a) can be verified from the same argument, directly using

$$\begin{aligned} R(\beta,\delta,\gamma;s) &= \mathbb{E}\left[\left(y_i - x_i^\top\beta - x_i^\top\delta\mathbf{1}\left[q_i \le \gamma(s)\right]\right)^2\right] \\ &\quad -\mathbb{E}\left[\left(y_i - x_i^\top\beta_0 - x_i^\top\delta_0\mathbf{1}\left[q_i \le \gamma_0(s)\right]\right)^2\right] \end{aligned}$$

in this case. For the case (b), for any $\gamma(s) \neq \gamma_0(s)$ at $s_i = s$ and given $(\beta_0^\top, \delta_0^\top)^\top$,

$$R(\beta_0,\delta_0,\gamma;s) = \delta_0^\top \mathbb{E}\left[x_i x_i^\top\right]\delta_0 \left|\gamma(s) - \gamma_0(s)\right| > 0.$$

Hence, we obtain the identification since $R(\beta_0,\delta_0,\gamma;s)$ is continuous at $\gamma = \gamma_0(s)$. ■

Now we establish Lemma A.1, which is the fundamental building block of Theorems 2 and 3.

Lemma A.1′ *Under Assumptions ID′ and A′, for any fixed $s \in S_0 \subset (0,1)$,*

$$\sup_{\gamma\in\Gamma} \left\|M_n(\gamma;s) - M(\gamma;s)\right\| \to_p 0,$$

$$\sup_{\gamma\in\Gamma} \left\|n^{-1/2} b_n^{-1/2} J_n(\gamma;s)\right\| \to_p 0$$

as $n \to \infty$, where $M(\gamma; s) = \gamma \bar{D}$ and $J_n(\gamma; s) \Rightarrow J(\gamma; s)$, a mean-zero Gaussian process indexed by γ.

Proof of Lemma A.1′ In view of the proof of Lemma A.1, the key difference lies on that (q_i, s_i) are now non-random, and hence Taylor expansion in $f(q, v)$ is no longer needed. Alternatively, we use the following two ideas: First, we decompose the summation $\sum_{i=1}^{n}$ into the double summation $\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}$, where $n = n_1 n_2$; second, we use the Taylor expansion and Assumption A′-(viii) to obtain $(n_2 b_n)^{-1}\sum_{j=1}^{n_2} K((j/n_2 - s)/b_n) = O(b_n^2)$.

Then, we have

$$\begin{aligned}
\mathbb{E}[M_n(\gamma; s)] &= \mathbb{E}[x_i^2]\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\mathbf{1}[i/n_1 \le \gamma]\right)\left((n_2 b_n)^{-1}\sum_{j=1}^{n_2} K((j/n_2 - s)/b_n)\right) \\
&= \bar{D}\left(\gamma + O\left(\frac{1}{n_1}\right)\right) O(b_n^2) \\
&= \gamma\bar{D} + O(n^{-1}) + O(b_n^2)
\end{aligned}$$

and

$$\begin{aligned}
Var[M_n(\gamma; s)] &= \frac{1}{n^2 b_n^2}\mathbb{E}\left[\left(\sum_{i=1}^{n}\left\{x_i^2\mathbf{1}_i(\gamma)K_i(s) - \mathbb{E}\left[x_i^2\mathbf{1}_i(\gamma)K_i(s)\right]\right\}\right)^2\right] \\
&= \frac{1}{n^2 b_n^2}\sum_{i=1}^{n}\mathbb{E}\left[\left\{x_i^2\mathbf{1}_i(\gamma)K_i(s) - \mathbb{E}\left[x_i^2\mathbf{1}_i(\gamma)K_i(s)\right]\right\}^2\right] \\
&\quad + \frac{2}{n^2 b_n^2}\sum_{i<j}^{n} Cov\left[x_i^2\mathbf{1}_i(\gamma)K_i(s), x_j^2\mathbf{1}_j(\gamma)K_j(s)\right] \\
&\equiv V_{n1} + V_{n2}.
\end{aligned}$$

To bound V_{n1}, we use the strict stationarity and the aforementioned two ideas to obtain that

$$\begin{aligned}
V_{n1} &= \frac{1}{n^2 b_n^2}\sum_{i=1}^{n}\left(\mathbb{E}\left[x_i^4\mathbf{1}_i(\gamma)K_i(s)\right] - \left\{\mathbb{E}\left[x_i^2\mathbf{1}_i(\gamma)K_i(s)\right]\right\}^2\right) \\
&= \frac{1}{n b_n}\left(\mathbb{E}\left[x_i^4\right]\frac{1}{n_2 b_n}\sum_{i_2=1}^{n_2} K\left(\frac{i_2/n_2 - s}{b_n}\right) - \bar{D}^2\frac{1}{n_2 b_n}\sum_{i_2=1}^{n_2} K^2\left(\frac{i_2/n_2 - s}{b_n}\right)\right) \\
&\quad \times\left(\frac{1}{n_1}\sum_{i_1=1}^{n_1}\mathbf{1}[i_1/n_1 \le \gamma]\right) \\
&= O(1/(n b_n)).
\end{aligned}$$

We also bound V_{n2} as

$$|V_{n2}| \le \left|\frac{1}{n^2 b_n^2}\sum_{i<j}^{n} Cov\left[x_i^2, x_j^2\right]\mathbf{1}_i(\gamma)K\left(\frac{s_i - s}{b_n}\right)\mathbf{1}_j(\gamma)K\left(\frac{s_j - s}{b_n}\right)\right|$$

$$\begin{aligned}
&\leq \frac{1}{n^2 b_n^2} \sum_{i=1}^{n} \mathbf{1}_i(\gamma) K\left(\frac{s_i - s}{b_n}\right) \left| \sum_{m=1}^{n-i} Cov\left[x_i^2, x_{i+m}^2\right] \mathbf{1}_m(\gamma) K\left(\frac{s_m - s}{b_n}\right) \right| \\
&\leq \frac{1}{n^2 b_n^2} \sum_{i_1=1}^{n_1} \mathbf{1}\left[i_1/n_1 \leq \gamma\right] \sum_{i_2=1}^{n_2} K\left(\frac{i_2/n_2 - s}{b_n}\right) \sum_{m=1}^{\infty} m\alpha(m)^{\varphi/(2+\varphi)} \left(\mathbb{E}\left[x_i^{4+2\varphi}\right]\right)^{2/(2+\varphi)} \\
&= O\left(1/\left(nb_n\right)\right).
\end{aligned}$$

Then the pointwise convergence of $M_n(\gamma; s)$ is established. The rest of the proof follows from very similar derivations as in Lemma A.1 and repeatedly using the two ideas aforementioned. ■

Lemma A.1′ establishes the uniform law of large numbers and the central limit theorem required in the rest of the proofs. Using this lemma, we can show that $\widehat{\gamma}(\cdot)$ has the same asymptotic distribution as in Theorem 3 with $\xi(s) = \kappa_2 c_0^\top \bar{V} c_0 / \left(c_0^\top \bar{D} c_0\right)^2$ for all $s \in \mathcal{S}_0$. The proof is again similar as in the main context and hence suppressed to save the space. It is available upon request.

References

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

Kim, J., and D. Pollard (1990): "Cube Root Asymptotics," *The Annals of Statistics*, 18, 191–219.