Nonparametric Spatial Threshold and Two-Dimensional Sample Splitting

Yoonseok Lee* Yulong Wang†
Syracuse University Syracuse University

May 2018

Abstract

[Preliminary; please do not circulate without the authors' permission.]

This paper studies a threshold regression model, where the threshold is determined by an unknown relation between two variables. The novel features of this model are in that the threshold is determined by two variables and their relation is nonparametric. Furthermore, we allow the observations to be cross-sectionally dependent and hence the model can be applied to study thresholds over a random field. Empirical relevance is illustrated by estimating an economic border induced by the housing price difference between Queens and Brooklyn in New York City. Such economic border deviates substantially from the administrative one.

Keywords: threshold, spatial, nonparametric, random field.
JEL Classifications: C12, C14, C21, C24.

*[*]*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

This paper studies a threshold regression model, where the threshold is determined by an unknown relation between two variables. More precisely, we consider a model given by

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

for $i = 1, 2, \cdots, n$, in which the marginal effect of x_i to y_i can be different depending on $q_i \leq \gamma_0\left(s_i\right)$ 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 features of this model are in that the threshold is determined by two scalar variables (q_i, s_i) and their relation is nonparametric. Furthermore, we allow that the observations can be cross-sectionally dependent (i.e., they can be strong-mixing random fields as Bolthausen, 1982), and hence the model can be applied to study thresholds over a space.

This paper contributes to the literature as follows. First, we formulate the threshold by some *unknown interaction* between *two variables*: $\mathbf{1}\left[q_i \leq \gamma_0\left(s_i\right)\right]$. Unlike the standard threshold models presuming that the threshold is determined by the level of one variable (e.g., Hansen, 2000), we consider that multiple variables can determine the threshold. Furthermore, the threshold function can be fully nonparametric (but smooth) and hence it can cover many interesting cases that have not been studied. For example, we can consider a model with heterogeneous thresholds if we see $\gamma_0\left(s_i\right)$ as heterogeneous thresholds over i; this specification can cover the case that the threshold is determined by the sign of a conditional moment. Apparently, when $\gamma_0(s) = \gamma_0$ or $\gamma_0(s) = \gamma_0 s$ for some parameter γ_0 and $s \neq 0$, it becomes the standard threshold regression model (where the threshold is determined by the ratio q_i/s_i for the latter case).

Second, this paper allows that the variables are *cross-sectionally dependent*, which has not been considered in the threshold model literature. This generalization allows us to study threshold models over a random field (i.e., space): If we let (q_i, s_i) correspond to the latitude and the longitude on the map, then $\gamma_0(\cdot)$ can be understood as the unknown border that splits the area into two. Examples include identifying the boundary of some airborne pollution (or toxic waste) or some tipping point over an area that segregates population.

The main results of this paper can be summarized in four-folds: First, we apply a two-step estimation for this semiparametric model and derive asymptotic properties of the estimators, where the unknown function $\gamma_0(\cdot)$ is estimated using a kernel method. Provided $\delta_0 = c_0 n^{-\epsilon}$ for some $c_0 \neq 0$ and $\epsilon \in (0, 1/2)$, it is shown that the nonparametric estimator $\widehat{\gamma}(\cdot)$ is uniformly consistent and $(\widehat{\beta}, \widehat{\delta})$ satisfies the $n^{-1/2}$-consistency using asymptotic results of random fields by Bolthausen (1982) and Jenish and Prucha (2009). Limiting distributions of these semiparametric estimators are also derived. Second, we develop a pointwise test of

$H_0 : \gamma_0(s) = \gamma_*(s)$ for given s; simulation studies show its good finite sample performance. Third, as an illustration, we apply this new model to study an unknown spatial threshold. In particular, we estimate an unknown economic border that splits the Queens and the Brooklyn boroughs in New York City, where each region has a different level of elasticity to the house price. Finally, we extend this threshold line model to identify a threshold contour or circle, and we estimate it by rotating the coordinate.

The rest of the paper is organized as follows. Section 2 summarizes the model and our estimation procedure. Section 3 derives limiting properties of the estimators and develops a likelihood ratio test of the threshold function. Section 4 studies small sample properties of the proposes statistics by Monte Carlo simulations. Section 5 applies the results to the housing price data to identify unknown economic border. Section 6 concludes the paper with describing how to extend this idea to estimate a threshold contour. All the mathematical proofs are in the Appendix.

2 Nonparametric Threshold Regression

We consider a threshold regression model given by

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

for $i = 1, 2, \cdots, n$, where $(y_i, x_i, q_i, s_i) \in \mathbb{R}^{1+p+1+1}$ and $\gamma_0(\cdot)$ is an unknown function. The threshold function $\gamma_0(\cdot)$ is unknown and the main parameters of interest are β_0, δ_0, and $\gamma_0(\cdot)$. In this model, the threshold is determined by two scalar variables (q_i, s_i) and their relation is nonparametric. If we see this model as a spatial threshold model over a space, then (q_i, s_i) can be understood as the location index (i.e., latitude and longitude) and hence the threshold $\mathbf{1}\left[q_i \leq \gamma_0\left(s_i\right)\right]$ describes two-dimensional sample splitting.[1]

We estimate the unknown parameters in two steps. More precisely, for given s, we fix $\gamma_0(s) = \gamma$, where γ can depends on s, and we first obtain $\widehat{\beta}(\gamma; s)$ and $\widehat{\delta}(\gamma; s)$ by local least squares conditional on γ:

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

[1] The results of this paper can be generalized for vectors q_i and s_i using multivariate kernel estimation. However, we focus on the scalar case to make the presentation simple. Note that the model (1) is different from Seo and Linton (2007), which specifies linear index form between (q_i, s_i) but assumes a nonparametric smooth transition function instead of $\mathbf{1}[\cdot]$.

where

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

for some kernel function $K(\cdot)$ and a bandwidth parameter b_n. Then, for a compact $\Gamma \subset \mathbb{R}$, $\gamma_0(\cdot)$ is estimated by

$$\widehat{\gamma}(s) = \arg\min_{\gamma \in \Gamma} Q_n(\gamma; s)$$

for given s, where $Q_n(\gamma; s)$ is the concentrated sum of squares defined as

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

Finally, the estimators of β_0 and δ_0 are obtained from

$$(\widehat{\beta}, \widehat{\delta}) = \arg\min_{\beta, \delta} \sum_{i=1}^{n} \left(y_i - x_i'\beta - w_i'\delta\right)^2, \tag{5}$$

where $w_i = x_i \mathbf{1}[q_i \leq \widehat{\gamma}(s_i)]$.

We allow for cross-sectional dependence in $(x_i', q_i, s_i, u_i)'$ in this study. For this purpose, similarly as 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 ℓ-th component of i. We write $|N_n|$ for the number of elements in N_n and we simply let the cardinality of N_n as n (i.e., $|N_n| = n$); the summation in (4) hence can be rewritten as $\sum_{i \in N_n}$. Following Bolthausen (1982) and Jenish and Prucha (2009), we also define a mixing coefficient:

$$\alpha(m) = \sup\left\{|P(A_i \cap A_j) - P(A_i)P(A_j)| : A_i \in \mathcal{F}_i \text{ and } A_j \in \mathcal{F}_j \text{ with } \lambda(i,j) \geq m\right\}, \tag{6}$$

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

We first assume the following conditions. We let $f(q,s)$ be the joint density function of (q_i, s_i), and define

$$D(q,s) = E\left[x_i x_i' | (q_i, s_i) = (q,s)\right], \tag{7}$$
$$V(q,s) = E\left[x_i x_i' u_i^2 | (q_i, s_i) = (q,s)\right]. \tag{8}$$

We also denote $\overline{\mathcal{S}}$ as the support of s_i and $\mathcal{S}$ as a bounded subset in the interior of $\overline{\mathcal{S}}$. In what follows, we only consider $s_i \in \mathcal{S}$.

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 $\lim_{n\to\infty} |\partial N_n| / n = 0$.

(ii) $(x_i', q_i, s_i, u_i)'$ is stationary and α-mixing with bounded $(2+\varphi)$th moments for some $\varphi > 0$ and with the mixing coefficient $\alpha(m)$ defined in (6) that satisfies $\sum_{m=1}^{\infty} m\alpha(m) < \infty$ and $\sum_{m=1}^{\infty} m^2\alpha(m)^{\varphi/(2+\varphi)} < \infty$.

(iii) $\delta_0 = c_0 n^{-\epsilon}$ for some $c_0 \neq 0$ and $\epsilon \in (0, 1/2)$; $\left(c_0', \beta_0'\right) \in \Theta_0$ some compact subset of $\mathbb{R}^{2p}$.

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

(v) $\gamma_0 : \mathcal{S} \mapsto \Gamma$ is twice continuously differentiable and $0 < P(q_i \leq \gamma_0(s_i)) < 1$, where Γ is a compact subset of the support of q_i.

(vi) Uniformly in (q,s), there exists some constant $0 < C < \infty$ such that $E[||x_i||^{8+\tau} | (q_i, s_i) = (q,s)] < C$ and $E[||x_i u_i||^{8+\tau} | (q_i, s_i) = (q,s)] < C$ for some $\tau > 0$.

(vii) $D(q,s)$, $V(q,s)$, and $f(q,s)$ are bounded, continuous in q, and twice continuously differentiable in $s \in \mathcal{S}$ with bounded derivatives.

(viii) $c_0' D(\gamma_0(s), s) c_0 > 0$, $c_0' V(\gamma_0(s), s) c_0 > 0$, and $f(\gamma_0(s), s) > 0$ for all $s \in \mathcal{S}$.

(ix) $E[x_i x_i' | s_i = s]$ is positive definite and bounded for any $s \in \mathcal{S}$.

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

(xi) $K(\cdot)$ is uniformly bounded, continuous, and symmetric around zero with satisfying $\int K(v)\, dv = 0$, $\int v^2 K(v)\, dv < 0$, $\kappa_2 = \int K(v)^2 dv < \infty$, $\lim_{v\to\infty} |v| K(v) = 0$, and $\lim_{v\to\infty} |v| K(v)^2 = 0$.

Most of these conditions are similar to Assumption 1 of Hansen (2000). Note that λ_0 in Assumption A-(i) can be any strictly positive value, but we can impose $\lambda_0 > 1$ without loss of generality. The conditions in Assumption A-(ii) are required to establish CLT for spatially dependent random field by Bolthausen (1982). 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 bandwidth in kernel estimation. Note that when $\alpha(m)$ decays at an exponential rate, these conditions are readily satisfied. On the other hand, 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 $k > 3(2+\varphi)/\varphi$. Assumption A-(x) and (xi) are standard in the kernel estimation literature (e.g., Li and Racine, 2007), except that the magnitude of the bandwidth b_n depends on ϵ.

Given $\gamma_0(\cdot)$, the parameters β_0 and δ_0 are well identified provided that $0 < P(q_i \leq \gamma_0(s_i)) < 1$ and hence $E[z_i z_i' | s_i = s]$ is positive definite under Assumption A-(ix), where $z_i = [x_i', x_i' \mathbf{1}[q_i \leq \gamma_0(s_i)]]'$. Assumption A-(v) restricts that the threshold $\gamma_0(s_i)$ lies in the interior of the support of q_i. The unknown border $\gamma_0(\cdot)$ is also well identified because for any $\gamma \neq \gamma_0$,

$$
\begin{aligned}
& E\left[\left(y_i - x_i'\beta_0 - x_i'\delta_0 \mathbf{1}[q_i \leq \gamma(s_i)]\right)^2 \middle| s_i = s \right] \\
& - E\left[\left(y_i - x_i'\beta_0 - x_i'\delta_0 \mathbf{1}[q_i \leq \gamma_0(s_i)]\right)^2 \middle| s_i = s \right] \\
& = \delta_0' E\left[x_i x_i' \left(\mathbf{1}[q_i \leq \gamma(s_i)] - \mathbf{1}[q_i \leq \gamma_0(s_i)]\right)^2 \middle| s_i = s \right] \delta_0 \\
& = n^{-2\epsilon} c_0' E\left[x_i x_i' \mathbf{1}[\min\{\gamma(s_i), \gamma_0(s_i)\} < q_i \leq \max\{\gamma(s_i), \gamma_0(s_i)\}] \middle| s_i = s \right] c_0 \\
& > 0
\end{aligned}
$$

under Assumptions A-(iii), (v), and (ix), where the condition $0 < P(q_i \leq \gamma_0(s_i)) < 1$ ensures that there exist q_i such that $\mathbf{1}[\min\{\gamma(s_i), \gamma_0(s_i)\} < q_i \leq \max\{\gamma(s_i), \gamma_0(s_i)\}] = 1$.

3 Asymptotic Results

We first obtain the asymptotic properties of the nonparametric estimator $\widehat{\gamma}(s)$. The first theorem shows that $\widehat{\gamma}(s)$ is consistent and derives the limiting distribution of $\widehat{\gamma}(s)$. Similar to Hansen (2000), we let $W(\cdot)$ be a two-sided Brownian motion.

Theorem 1 *Under Assumption A, $\widehat{\gamma}(s) \to_p \gamma_0(s)$ as $n \to \infty$ for any fixed $s \in \mathcal{S}$. Furthermore, if $n^{1-2\epsilon} b_n^3 \to 0$,*

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

as $n \to \infty$, where

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

and $\kappa_2 = \int K(v)^2 dv$.

Note that the distribution of $\arg\max_{r \in \mathbb{R}} (W(r) - |r|/2)$ is known (e.g., Bhattacharya and Brockwell, 1976), which is also described in Hansen (2000, p.581). The constant term $\xi(s)$

determines the scale of the distribution at given s, which increases in the conditional variance $E\left[u_i^2|x_i,q_i,s_i\right]$; but decreases in the size of the threshold constant $|c_0|$ and the density of (q_i,s_i) near the threshold.

Theorem 1 also shows that the pointwise rate of convergence of $\widehat{\gamma}(s)$ is $n^{1-2\epsilon}b_n$, which depend on two parameters, ϵ and b_n. It is decreasing in ϵ like the parametric case. As noted in Hansen (2000), a larger ϵ reduces the threshold effect $\delta_0 = c_0 n^{-\epsilon}$ and hence decreases effective sampling information on the threshold. Since we estimate $\gamma(\cdot)$ using the kernel estimation method, the rate of convergence depends on the bandwidth size b_n as well. Like the standard kernel estimator cases, smaller bandwidth decreases effective local sample size, which reduces the precision of estimators of $\gamma(\cdot)$. 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 seems to be large and close to 1/2). For instance, by balancing the square of conventional $O(b_n^2)$ bias and the $O((n^{1-2\epsilon}b_n)^{-1})$ precision from Theorem 1, the optimal bandwidth satisfies $b_n^* = c^* n^{-(1-2\epsilon)/5}$ for some constant $0 < c^* < \infty$.[2] However, it does not mean that we can always choose b_n as large as possible, as well documented in the standard kernel estimation. The choice needs to be such that $n^{1-2\epsilon}b_n^3 \to 0$, which is required to control for the $O(b_n^2)$ bias term in the kernel estimator and hence the limiting distribution of $n^{1-2\epsilon}b_n\left(\widehat{\gamma}(s) - \gamma_0(s)\right)$ has mean zero.

From Theorem 1, we can consider a pointwise likelihood ratio test statistic for

$$H_0 : \gamma_0(s) = \gamma_*(s) \quad \text{for some } s \in \mathcal{S}, \tag{9}$$

which is given as

$$LR_n(s) = \left(\sum_{i=1}^{n} K\left(\frac{s_i - s}{b_n}\right)\right) \times \frac{Q_n\left(\gamma_*(s), s\right) - Q_n\left(\widehat{\gamma}(s), s\right)}{Q_n\left(\widehat{\gamma}(s), s\right)}. \tag{10}$$

The following theorem obtains the null limiting distribution of this test statistic.

Theorem 2 *Under the same condition in Theorem 1, for any fixed $s \in \mathcal{S}$, the test statistic in (10) under the hull hypothesis (9) satisfies*

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

[2] It is the standard problem in the kernel estimation studies that the optimal bandwidth parameter selection based on this expression is not feasible in practice since the constant term c^* is unknown. In our case, unfortunately, it is even more infeasible because the choice of the bansdwidth parameter depends on the nuisance parameter ϵ as well, which is not even estimable. We can use the cross-validation approach in practice, though its statistical properties need to be studied further.

as $n \to \infty$, where

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

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

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

Table I: Asymptotic Critical Values (Gaussian Kernel)

$P(\zeta^* > cv)$	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

For the general cases, $\xi_{LR}(s)$ can be estimated as

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

where $\widehat{\sigma}^2(s) = \sum_{i=1}^n \omega_{1i}(s) \widehat{u}_i^2$, and $\widehat{D}(\widehat{\gamma}(s), s)$ and $\widehat{V}(\widehat{\gamma}(s), s)$ are the standard Nadaraya-Watson estimators. Recall that

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

with $\widehat{u}_i = y_i - x_i'\widehat{\beta} - w_i'\widehat{\delta}$ from (5) and

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

for some bivariate kernel function $\mathbb{K}(\cdot,\cdot)$ and bandwidth parameters b_n', b_n''. Note that we can also form an asymptotic confidence interval for $\widehat{\gamma}(s)$ using the likelihood test inversion method advocated by Hansen (2000).

Finally, once we construct the estimator $\widehat{\gamma}(s_i)$ for all $s_i \in \mathcal{S}$, we can plug it in (5) to obtain the $\sqrt{n}$-consistency of $\widehat{\beta}$ and $\widehat{\delta}$, which is simply regressing y_i on x_i and $x_i \mathbf{1}_i(\widehat{\gamma}(s_i))$. For this purpose, we obtain the uniform rate of convergence of $\widehat{\gamma}(s)$ as follows, which requires more conditions on the bandwidth b_n. More precisely, similarly as Carbon et al. (2007), we suppose that either

$$n^{1-2\epsilon} b_n / (\log n)^3 \to \infty \quad \text{and } n b_n^3 / \log n \to 0 \tag{11}$$

if the mixing coefficient $\alpha(m)$ decays at an exponential rate; or

$$n^{1-2\epsilon} b_n / \log n \to \infty \quad \text{and } n^{1-2\epsilon} b_n^{(k+4)/(k-4)} / (\log n)^{(k-2)/(k-4)} \to \infty \tag{12}$$

for some $k > 4$, if the mixing coefficient $\alpha(m)$ decays at a polynomial rate (i.e., $\alpha(m) \leq C_\alpha m^{-k}$ for some $k > 0$).

Theorem 3 *Under the same condition in Theorem 1 and if either (11) or (12) hold,*

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

Using this uniform convergence of $\widehat{\gamma}(s)$, the following theorem formalizes the $\sqrt{n}$-consistency as well as the joint limiting distribution of $\widehat{\beta}$ and $\widehat{\delta}$.

Theorem 4 *Let $\widehat{\theta} = (\widehat{\beta}', \widehat{\delta})'$ and $\theta_0 = (\beta_0', \delta_0)'$. Suppose the same condition in Theorem 1 holds. Then, under $n^{1-2\epsilon} b_n^2 \to \infty$ and either (11) or (12),*

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \to_d \mathcal{N}\left(0, M^{*-1} V^* M^{*-1}\right) \quad \text{as } n \to \infty,$$

where $M^* = E[z_i z_i']$ and $V^* = \lim_{n\to\infty} Var[n^{-1/2} \sum_{i=1}^n z_i u_i]$ with $z_i = [x_i', x_i' \mathbf{1}_i(\gamma_0(s_i))]'$.

Note that we need a smaller bandwidth parameter b_n (i.e., $n^{1-2\epsilon} b_n^2 \to \infty$) in order to achieve the $\sqrt{n}$-consistency of $\widehat{\theta}$ in Theorem 4. This additional condition is required to satisfy the asymptotic orthogonality condition between $\widehat{\theta}$ and $\widehat{\gamma}$ (e.g., Assumption N(c) in Andrews (1994)), that is, replacing $\widehat{\gamma}$ by γ_0 in (5) has an effect at most $o_p(n^{-1/2})$. Given that we recover the $n^{-1/2}$ rate by using all the observations (except those with $s_i \notin \mathcal{S}$), the asymptotic variance now involves cross-sectional dependence as indicated by the long-run variance form

V^*. This can be consistently estimated by the spatial HAC estimator proposed by Conley (1999) and Conley and Molinari (2007) using $\widehat{u}_i = y_i - x_i'\widehat{\beta} - w_i'\widehat{\delta}$ and $\widehat{\gamma}(\cdot)$, with a slightly stronger condition on the mixing coefficient $\alpha(m)$.

4 Monte Carlo Experiments

In order to study small sample performance of the likelihood test, we conduct Monte Carlo simulations as follows. We consider the threshold regression in (1) with

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

where $\beta_{01} = \beta_{02} = 0$, $\gamma_0\left(s\right) = \sin(s)/2$, and δ takes different values. For the dependence structure in $(x_i, q_i, s_i, u_i)'$, we consider the following data generating process:

$$\begin{cases} (q_i, s_i)' \sim iid\mathcal{N}\left(0, I_2\right); \\ x_i|\left(q_i, s_i\right) \sim iid\mathcal{N}\left(0, (1 + \rho\left(s_i^2 + q_i^2\right))^{-1}\right); \\ \underline{\mathbf{u}}|\{(x_i, q_i, s_i)\}_{i=1}^n \sim \mathcal{N}\left(0, \Omega\right), \end{cases}$$

where $\underline{\mathbf{u}} = (u_1, \cdots, u_n)'$. The (i, j)th element of Ω is $\Omega_{ij} = \rho^{\lfloor \ell_{ij} n \rfloor}\mathbf{1}\left(\ell_{ij} < m/n\right)$, where $\ell_{ij} = \left((s_i - s_j)^2 + (q_i - q_j)^2\right)^{1/2}$ is the L^2-distance between the i and j observation and $\lfloor A \rfloor$ denotes the largest integer smaller than A. The diagonal elements of Ω are normalized as $\Omega_{ii} = 1$. 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 m distance, the dependence also decays in a polynomial rate as indicated by $\rho^{\lfloor \ell_{ij} n \rfloor}$. The single parameter ρ describes the cross-sectional dependence in the way that a larger ρ leads to stronger dependence relative to the unit standard deviation. In particular, we consider $\rho = 0$ (which is for i.i.d. observations), 0.5 and 1.

Tables II to V report the small sample rejection probabilities of the LR test (10) at 5% nominal level over three locations $s = 0$, 0.5, and 1. In general, the test for γ_0 performs better when (i) the sample size is larger; (ii) the coefficient change at the threshold is larger; and (iii) the cross-sectional dependence is weaker. In particular, we follow Conley and Molinari (2007) to use $m = 3$ in the first three tables, while in Table V we examine the effect of slower dependence decay by setting $m = 10$. For each location, we consider 9 cases with $n = 100, 200, 500$ and $\delta_{10} = \delta_{20} = \delta_0 = 1, 2, 3$ (cf. Hansen (2000)). For the bandwidth parameter, we simply normalize s_i and q_i to have mean zero and unit standard deviation and choose $b_n = n^{-2/5}$ in the main regression. To estimate D and V, we use the rule-of-thumb bandwidths $b_n' = n^{-1/5}$ and $b_n'' = n^{-1/6}$. Note that each combination of (n, δ_0) determines ϵ

for a fixed c_0 as $\epsilon = (\log c_0 - \log \delta_0)/\log n$. All the results are based on 2,000 simulations.

Table II: Rej. Prob. with i.i.d. data
($\rho = 0,\ m = 3$)

	$s = 0.0$			$s = 0.5$			$s = 1.0$		
$n \setminus \delta_0$	1	2	3	1	2	3	1	2	3
100	0.08	0.06	0.05	0.09	0.06	0.05	0.10	0.06	0.05
200	0.07	0.05	0.04	0.07	0.05	0.04	0.07	0.04	0.03
500	0.06	0.04	0.03	0.05	0.03	0.03	0.07	0.03	0.03

Table III: Rej. Prob. with spatially correlated data
($\rho = 0.5,\ m = 3$)

	$s = 0.0$			$s = 0.5$			$s = 1.0$		
$n \setminus \delta_0$	1	2	3	1	2	3	1	2	3
100	0.08	0.06	0.06	0.09	0.06	0.05	0.10	0.08	0.05
200	0.06	0.05	0.05	0.07	0.04	0.04	0.07	0.06	0.03
500	0.05	0.04	0.04	0.08	0.04	0.04	0.07	0.03	0.02

Table IV: Rej. Prob. with spatially correlated data
($\rho = 1,\ m = 3$)

	$s = 0.0$			$s = 0.5$			$s = 1.0$		
$n \setminus \delta_0$	1	2	3	1	2	3	1	2	3
100	0.09	0.07	0.06	0.09	0.07	0.06	0.10	0.08	0.05
200	0.08	0.06	0.05	0.08	0.05	0.03	0.08	0.06	0.03
500	0.06	0.04	0.04	0.07	0.03	0.03	0.06	0.03	0.02

Table V: Rej. Prob. with spatially correlated data
($\rho = 1,\ m = 10$)

	$s = 0.0$			$s = 0.5$			$s = 1.0$		
$n \setminus \delta_0$	1	2	3	1	2	3	1	2	3
100	0.13	0.07	0.07	0.10	0.09	0.06	0.11	0.09	0.06
200	0.09	0.05	0.06	0.08	0.06	0.04	0.08	0.06	0.05
500	0.05	0.04	0.04	0.06	0.04	0.03	0.08	0.04	0.02

Lastly, Table VI depicts the finite sample coverage properties of the 95% confidence intervals for the parametric components β_{02} and δ_{02} (i.e., the pre-break value and the change size associate with x_i, respectively). We use the same DGP as above with $\rho = 0.5$ and $m = 3$. Regarding the lag number required for the HAC estimator, we follow Conley and Molinari (2007) and use the spatial lag order of 5. Results with other choices are similar and available upon request. The numbers in different columns suggest that the asymptotic normality is better approximated with lower dependence and larger change size, while comparing rows indicates that the sample size has to be large enough for the asymptotics to perform satisfactorily.

In addition, some unreported results suggest that if we increase δ_{01} (i.e., the change size associated with the constant), the coverage for β_{02} and δ_{02} can be improved. This is because the larger δ_{01} provides more information on the threshold $\gamma(\cdot)$ estimation that in turn results in better estimation of other parameters.

Table VI-a: Coverage Prob. of β_{02}
($\rho = 0.5$, $m = 3$)

	$\rho = 0.0$			$\rho = 0.5$			$\rho = 1.0$		
$n \setminus \delta_0$	1	2	3	1	2	3	1	2	3
100	0.85	0.90	0.92	0.86	0.90	0.91	0.84	0.90	0.92
200	0.88	0.92	0.93	0.88	0.91	0.93	0.86	0.91	0.92
500	0.89	0.93	0.93	0.89	0.92	0.92	0.88	0.93	0.93

Table VI-b: Coverage Prob. of δ_{02}

100	0.82	0.88	0.90	0.82	0.87	0.90	0.82	0.89	0.90
200	0.83	0.89	0.90	0.82	0.89	0.90	0.82	0.88	0.89
500	0.85	0.89	0.89	0.82	0.89	0.90	0.82	0.90	0.89

5 Empirical illustration

As an illustration, we study the housing price of the Queens and the Brooklyn boroughs in New York City, using the single family house sales data in the year 2017. The data set (*Rolling Sales Data*) is available at http://www1.nyc.gov/site/finance/taxes/property-rolling-sales-data.page. In the threshold regression model (1), we consider the following

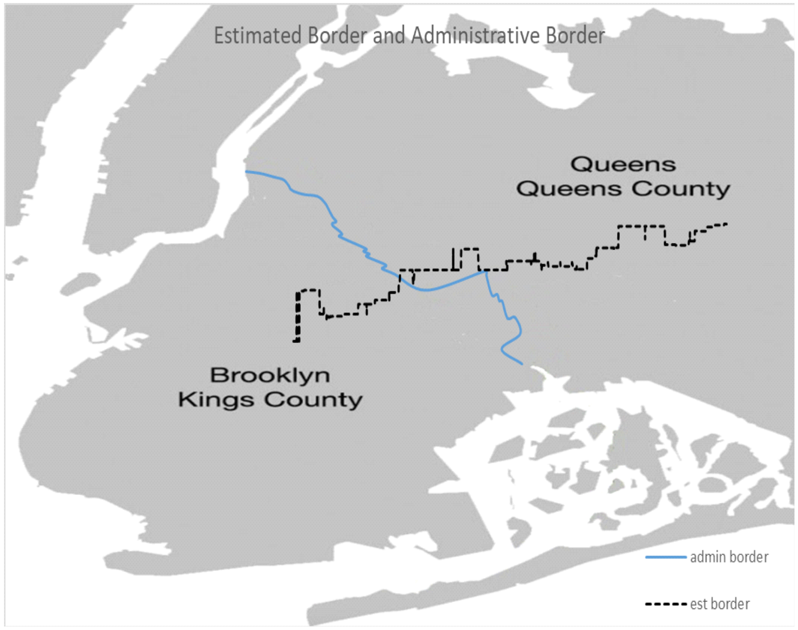

Figure 1: Threshold Function Estimate

variables:[3]

y_i	x_i	q_i	s_i
log house price ($)	constant	latitude	longitude
	log of Gross Square Footage (ft^2)		
	dummy for built before 1945 (WWII)		

In this exercise, since the pair (q_i, s_i) corresponds to the latitude and the longitude on the map, "above the threshold" means the region on the northern side of the economic border, whereas "below the threshold" means the region on the southern side of the economic border. We focus on single family houses under property tax class 1, accounting for 57.9% of the original sample, and drop duplicate observations. The sample size is $n = 8121$ (5962 observations in Queens; 2159 observations in Brooklyn).

Figure 1 depicts the nonparametric threshold function estimates $\widehat{\gamma}$, which is the "unknown" *economic* border that splits the Queens and the Brooklyn boroughs in New York City. The estimated border (black dash line) is found to be substantively different from the administrative border between these two boroughs (blue solid line). Somewhat surprisingly,

[3] "Gross Square Footage" is the total area of all the floors of a building as measured from the exterior surfaces of the outside walls of the building, including the land area and space within any building or structure on the property. (Source: http://www1.nyc.gov/assets/finance/downloads/pdf/07pdf/glossary_rsf071607.pdf)

the estimated border in Queens approximately coincides with the Long Island Rail Road (LIRR) route.

The bandwidth b_n in the regression (4) is chosen by the cross validation. In particular, we choose the constant c in $cn^{-2/5}$ that minimizes the sum of $(y_i - \widehat{y}_i)^2 \mathbf{1}[s_i \in \mathcal{S}]$, where $\mathcal{S}$ includes the observations between 15th and 85th percentiles of $\{s_i\}$, $\widehat{y}_i = x_i'\widehat{\beta}_{-i} + x_i'\widehat{\delta}_{-i}\mathbf{1}\left[q_i \leq \widehat{\gamma}_{-i}(s_i)\right]$, and $(\widehat{\beta}_{-i}, \widehat{\delta}_{-i}, \widehat{\gamma}_{-i}(\cdot))'$ are obtained without the i-th observation (i.e., leave-one-out estimators). Table VII summarizes the coefficient estimates for the parametric components, $\widehat{\beta}$ and $\widehat{\delta}$. The standard errors reported in the parentheses are computed by the Conley (1999)'s HAC estimator with 5 spatial lags.

Table VII: Estimation Result

	$\widehat{\beta}$	$\widehat{\delta}$
constant	9.91	-1.08
	(0.01)***	(0.01)***
log of Gross Square Footage	0.40	-0.01
	(0.05)***	(0.04)
dummy for built before 1945	-0.07	-0.06
	(0.01)***	(0.01)***

Note: *** indicates significant at 1%.

The average housing price on the southern side of the threshold (or economic border) is lower than that on north. The elasticity of the Gross Square Footage remains similar across the economic border but the negative effect of the house age is larger on the southern side.

6 Extension and Concluding Remarks

The threshold model (1) can be generalized to allow for the following unknown *contour threshold* model:

$$y_i = x_i'\beta_0 + x_i'\delta_0 \cdot \mathbf{1}[\mu(q_i, s_i) \leq 0] + u_i, \tag{13}$$

where the unknown function μ of q_i and s_i 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 identifying a group or a region in which the agents share common demographic/political/economic characteristics.

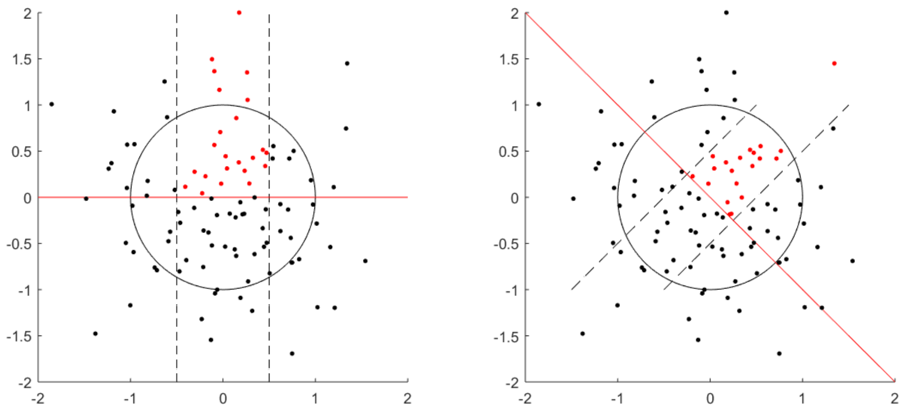

Figure 2: Illustration of Rotation

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 $\mu(q_i^*, s_i^*) < 0$. Without loss of generality, we can normalize (q_i^*, s_i^*) to be $(0, 0)$ and the 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 $(\mathbf{I}_i, \mathbf{II}_i, \mathbf{III}_i, \mathbf{IV}_i)$ respectively denotes the indicator function that the ith observation locates in the first, second, third, or forth quadrant. We suppose that there is only one breakpoint at any angle.[4] For each fixed $a^\circ \in [0^\circ, 360^\circ)$, we then can rotate the original coordinate counterclockwise and implement the least squares estimator (4) with the observations in the first two quadrants after rotation.

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

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

[4] If we further suppose there exists a continuous function $g : [0^\circ, 360^\circ) \to \mathbb{R}$ such that $\mathbf{1}[\mu(q_i, s_i) \le 0] = \mathbf{1}[l_i - g(a_i^\circ) \le 0]$, then the function $g(\cdot)$ describes a two-dimensional contour (e.g., a circle), which induces a coefficient change if the subject steps from the inside to the outside. As we suppose there is only one breakpoint at any angle a°, it reduces to our baseline model (1).

After this rotation, we estimate the following threshold model:

$$y_i = x_i'\beta_0 + x_i'\delta_0 \cdot \mathbf{1}\left[q_i\left(a^{\circ}\right) \leq \gamma\left(s_i\left(a^{\circ}\right)\right)\right] + u_i \tag{14}$$

using only the observations satisfying $q_i\left(a^{\circ}\right) \geq 0$, where $\gamma\left(\cdot\right)$ serves as the unknown threshold line as in the model (1) in the rotated coordinate. Such reparameterization guarantees that $\gamma\left(\cdot\right)$ is always positive and we are estimating its value pointwisely at 0. Figure 2 illustrates the idea of such rotation and pointwise estimation with a bounded the kernel function so that only the red points are included for estimation at different angles. Thus, the estimation and inference procedure developed before is directly applicable.

In this paper, we propose a two-dimensional sample splitting model which captures the fact that two variables jointly determine the separation boundary. We illustrate the empirical relevance in a simple spatial context where the housing price is different across an unknown economic border between Brooklyn and Queens. Potentially more interesting applications are being explored, including investigating how congressional district is determined and drawing the economic boundary between two regions by Satellite data.

A Appendix

Throughout the proof, we denote $K_i(s) = K((s_i - s)/b_n)$ and $\mathbf{1}_i(\gamma) = \mathbf{1}[q_i \leq \gamma]$. $C \in (0, \infty)$ stands for a generic constant term that may vary, which can depend on the location s.

Lemma A.1 *For given $s \in \mathcal{S}$, let*

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

Under Assumption A,

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

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, Assumption A-(vii) and Taylor expansion, we have

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

where $D(q, s)$ is defined in (7). For the variance, we have

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

where the order of the first term is from the standard kernel estimation result; and for the second term we use Assumption A-(vi), (vii), and Lemma 1 of Bolthausen (1982) to obtain that

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

for some finite $\varphi > 0$, where $\alpha(m)$ is the mixing coefficient defined in (6) 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 from a similar argument as in Lemma 4.6 of Zhu and Lahiri (2007) and the uniform convergence follows from standard argument. For $J_n(\gamma; s)$, since $E[u_i x_i | q_i, s_i] = 0$, the proof for $\sup_{\gamma\in\Gamma} |n^{-1/2} b_n^{-1/2} J_n(\gamma, s)| \xrightarrow{p} 0$ 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, Theorem of Bolthausen (1982) implies that $J_n(\gamma; s) \Rightarrow J(\gamma; s)$ under Assumption A-(ii). 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 $\Delta > 0$ such that if n is large enough,

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

for any ζ. To this end, we consider a fine enough grid over $[\zeta, \zeta + \Delta]$ such that $\zeta = \zeta_0 < \zeta_1 < \zeta_2 < \cdots < \zeta_{G_n-1} < \zeta_{G_n} = \zeta+\Delta$, where $nb_n\Delta/2 \leq G_n \leq nb_n\Delta$ and $\max_{1\leq g\leq G_n}\left(\zeta_g - \zeta_{g-1}\right) \leq \Delta/G_n$. We define $h_{ig}(s) = x_i u_i K_i(s)\, \mathbf{1}\left[\zeta_{g-1} < q_i \leq \zeta_g\right]$ and $H_{ng}(s) = n^{-1}b_n^{-1}\sum_{i=1}^{n} |h_{ig}(s)|$ for $1 \leq g \leq G_n$. Then for any $\gamma \in \left[\zeta_{g-1}, \zeta_g\right]$,

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

and hence

$$\begin{aligned}
& \sup_{\gamma\in[\zeta,\zeta+\Delta]} \left|J_n\left(\gamma;s\right)-J_n\left(\zeta;s\right)\right| \\
\leq\ & \max_{1\leq g\leq G_n} \left|J_n\left(\zeta_g;s\right)-J_n\left(\zeta;s\right)\right| \\
& +\max_{1\leq g\leq G_n} \sqrt{nb_n}\left|H_{ng}(s)-E\left[H_{ng}(s)\right]\right| + \max_{1\leq g\leq G_n} \sqrt{nb_n}E\left[H_{ng}(s)\right] \\
\equiv\ & \Psi_1(s)+\Psi_2(s)+\Psi_3(s).
\end{aligned}$$

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

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

where each term's bound is obtained as follows. For $\Psi_{11}(s)$, a straightforward calculation and Assumption A-(vi) 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.2),

$$\begin{aligned}
\Psi_{12}(s) &\leq \frac{2}{n^2b_n^2}\sum_{i<j}^n\left(E\left[h_i^2(s)\right]E\left[h_j^2(s)\right]+\left|Cov\left[h_i^2(s),h_j^2(s)\right]\right|\right) \qquad \text{(A.3)} \\
&\leq 2\left(E\left[\widetilde{h}_i^2\right]\right)^2+\frac{2}{nb_n^2}\left\{Cb_n^2\sum_{m=1}^\infty m\alpha\left(m\right)^{\varphi/(2+\varphi)}\left(E\left[\widetilde{h}_i^{4+2\varphi}\right]\right)^{2/(2+\varphi)}+O\left(nb_n^4\right)\right\}
\end{aligned}$$

for some $\varphi>0$ that depends on s, where we let $\widetilde{h}_i=x_iu_iK\left(t_i\right)\mathbf{1}\left[\zeta_g<q_i\leq\zeta_k\right]$ 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

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

for some constant $0<C'<\infty$. Using the same argument as the second component in (A.3), we can also show that $\Psi_{13}(s) = O(n^{-1}) + O(b_n^2)$. For $\Psi_{14}(s)$, by stationarity,

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

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

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

where the first equality is by the change of variables ($t_i = (s_i - s)/b_n$) and by Assumption A-(xi). It follows that the first term in (A.4) satisfies

$$
\begin{aligned}
& \frac{4!}{nb_n^2} \sum_{i=1}^{n} \sum_{j,k\leq i} |cov[h_1(s), h_{i+1}(s)h_{i+j+1}(s)h_{i+j+k+1}(s)]| \\
\leq\ & \frac{C4!}{nb_n^{2-(4/(2+\varphi))}} \sum_{i=1}^{\infty} i^2\alpha(i)^{\varphi/(2+\varphi)} \\
& \times \left\{\left(E\left[\widetilde{h}_1^{2+\varphi}\right]\right)^{1/(2+\varphi)} \left(E\left[\left(\widetilde{h}_{i+1}\widetilde{h}_{i+j+1}\widetilde{h}_{i+j+k+1}\right)^{2+\varphi}\right]\right)^{1/(2+\varphi)} + O\left(b_n^2\right)\right\} \\
=\ & O\left(\frac{1}{nb_n^{2\varphi/(2+\varphi)}}\right) + O\left(\frac{b_n^{4/(2+\varphi)}}{n}\right)
\end{aligned}
\tag{A.5}
$$

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

$$
\frac{2\varphi}{2+\varphi} \leq \frac{1}{1-2\epsilon}, \tag{A.6}
$$

then $nb_n^{2\varphi/(2+\varphi)} = (n^{1-2\epsilon}b_n^{(2\varphi/(2+\varphi))(1-2\epsilon)})^{1/(1-2\epsilon)} \to \infty$ by Assumption A-(x), which yields (A.5) becomes $o(1)$. Using the same argument, we can also verify that the rest of terms in (A.4) 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

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

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

$$E\left[\left|J_n\left(\zeta_g; s\right) - J_n\left(\zeta_k; s\right)\right|^4\right] \leq C_1(s)\left(\zeta_k - \zeta_g\right)^2$$

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

$$P\left(\max_{1\leq g\leq G_n}\left|J_n\left(\zeta_g; s\right) - J_n\left(\zeta; s\right)\right| > \eta(s)\right) \leq \frac{C_1(s)\Delta^2}{\eta^4(s)b_n}, \tag{A.7}$$

which bounds $\Psi_1(s)$.

To bound $\Psi_2(s)$, the standard result (e.g., Li and Racine, 2007, Ch.1) yields that $E\left[h_{ik}^2\right] \leq C_2(s)b_n$ for some constant $0 < C_2(s) < \infty$ given s. Then by Lemma 1 of Bolthausen (1982), we have

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

and hence by Markov's inequality,

$$P\left(\max_{1\leq g\leq G_n}\sqrt{nb_n}\left|H_{ng}(s) - E\left[H_{ng}(s)\right]\right| > \eta(s)\right) \leq \frac{C_2(s)\Delta}{\eta^2(s)}. \tag{A.8}$$

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

$$\sqrt{nb_n}E\left[H_{ng}(s)\right] = n^{1/2}b_n^{1/2}C_3(s)\Delta/G_n \leq 2C_3(s)n^{-1/2}b_n^{-1/2} \tag{A.9}$$

for some constant $0 < C_3(s) < \infty$ given s, where $\Delta/G_n \leq 2/nb_n$. So tightness is proved by combining (A.7), (A.8), and (A.9), and hence the weak convergence follows from Theorem 15.5 of Billingsley (1968). ■

Lemma A.2 *Under Assumption A, for given $s \in \mathcal{S}$, $\widehat{\gamma}(s) \to_p \gamma_0(s)$ as $n \to \infty$.*

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

$$\widehat{\theta}(\gamma;s)=(\widetilde{Z}(\gamma;s)'\widetilde{Z}(\gamma;s))^{-1}\widetilde{Z}(\gamma;s)'\widetilde{y}(s), \tag{A.10}$$

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

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

where $P_{\widetilde{Z}}(\gamma;s)=\widetilde{Z}(\gamma;s)(\widetilde{Z}(\gamma;s)'\widetilde{Z}(\gamma;s))^{-1}\widetilde{Z}(\gamma;s)'$ and I 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 $\widetilde{X}(\gamma;s)'(\widetilde{X}(s)-\widetilde{X}(\gamma;s))=0$. Furthermore, for $\gamma\geq\gamma_0(s_i)$, $\widetilde{X}(\gamma_0(s_i);s)'(\widetilde{X}(s)-\widetilde{X}(\gamma;s))=0$ and $\widetilde{X}(\gamma_0(s_i);s)'\widetilde{X}(\gamma;s)=\widetilde{X}(\gamma_0(s_i);s)'\widetilde{X}(\gamma_0(s_i);s)$. Since

$$M_n(\gamma;s)=\frac{1}{nb_n}\sum_{i=1}^{n}\widetilde{x}_i(\gamma;s)\widetilde{x}_i(\gamma;s)' \text{ and } J_n(\gamma;s)=\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{n}\widetilde{x}_i(\gamma;s)\widetilde{u}_i(s),$$

Lemma A.1 hence yields that

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

for given s. It follows that

$$\begin{aligned}
&\frac{1}{n^{1-2\epsilon}b_n}\left(Q_n(\gamma;s)-\widetilde{u}(s)'\widetilde{u}(s)\right) \qquad\qquad (A.11)\\
=\ & O_p\left(\frac{1}{n^{1-2\epsilon}b_n}\right)+O_p\left(\sqrt{\frac{1}{n^{1-2\epsilon}b_n}}\right)+\frac{1}{nb_n}c_0'\widetilde{X}(\gamma_0(s_i);s)'\left(I-P_{\widetilde{Z}}(\gamma;s)\right)\widetilde{X}(\gamma_0(s_i);s)c_0\\
=\ & \frac{1}{nb_n}c_0'\widetilde{X}(\gamma_0(s_i);s)'\left(I-P_{\widetilde{Z}}(\gamma;s)\right)\widetilde{X}(\gamma_0(s_i);s)c_0+o_p(1)
\end{aligned}$$

for $n^{1-2\epsilon}b_n\to 0$ as $n\to\infty$. Hence, similarly as Lemma A.1, it can be verified that

$$\begin{aligned}
E\left[M_n(\gamma_0(s_i);s)\right] &= \frac{1}{b_n}\iint E[x_i^2|q,v]\mathbf{1}[q\leq\gamma_0(v)]K\left(\frac{v-s}{b_n}\right)f(q,v)\,dqdv \qquad (A.12)\\
&= E\left[M_n(\gamma_0;s)\right]+\int\left(\int_{\gamma_0(s)}^{\gamma_0(s+b_nt)}D(q,s+b_nt)f(q,s+b_nt)dq\right)K(t)dt\\
&= E\left[M_n(\gamma_0;s)\right]+\int\left(\int_{\gamma_0(s)}^{\gamma_0(s+b_nt)}D(q,s)f(q,s)dq\right)\left(1+C_1b_n^2t^2\right)K(t)dt\\
&= E\left[M_n(\gamma_0;s)\right]+\int\left(C_{21}b_nt+C_{22}b_n^2t^2\right)\left(1+C_1b_n^2t^2\right)K(t)dt\\
&= E\left[M_n(\gamma_0;s)\right]+O\left(b_n^2\right),
\end{aligned}$$

for some $C_1, C_{21}, C_{22} < \infty$, where the fourth equality is by the Leibniz integral rule under Assumption A-(v). It follows that, uniformly over $\gamma \in \Gamma \cap [\gamma_0(s_i), \infty)$,

$$\begin{aligned} & \frac{1}{nb_n} c_0' \widetilde{X}(\gamma_0(s_i); s)' \left(I - P_{\widetilde{Z}}(\gamma; s)\right) \widetilde{X}(\gamma_0(s_i); s) c_0 \\ & \rightarrow_p c_0' M(\gamma_0; s) c_0 - c_0' M(\gamma_0; s)' M(\gamma; s)^{-1} M(\gamma_0; s) c_0 < \infty, \end{aligned} \tag{A.13}$$

from Lemma A.1 and Assumptions A-(viii) and (ix), as $O\left(b_n^2\right) = o(1)$. Note that $M(\gamma; s) = E[x_i x_i' \mathbf{1}_i(\gamma) | s_i = s] f_s(s)$ is positive definite from Assumptions A-(viii) and (ix), where $f_s(s)$ is the marginal density of s_i. The pointwise consistency follows using the same argument as the proof of Lemma A.5 of Hansen (2000). ■

Lemma A.3 *Define $a_n = n^{1-2\epsilon} b_n$, where ϵ is given in Assumption A-(iii). For given $s \in \mathcal{S}$, let $\gamma_n(s) = \gamma_0(s) + r/a_n$ with some $|r| < \infty$, and*

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

Then,

$$A_n^*(r, s) \rightarrow_p |r| \, c_0' D(\gamma_0(s), s) c_0 f(\gamma_0(s), s)$$

and

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

as $n \rightarrow \infty$ under Assumption A, where $\kappa_2 = \int K(v)^2 dv$.

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

Proof of Lemma A.3 First consider $r > 0$. By change of variables and Taylor expansion, Assumption A-(vii) and (viii) imply that

$$\begin{aligned} E\left[A_n^*(r, s)\right] &= \frac{a_n}{nb_n} \sum_{i=1}^{n} E\left[\left(c_0' x_i\right)^2 \left(\mathbf{1}_i(\gamma_0(s) + r/a_n) - \mathbf{1}_i(\gamma_0(s))\right) K_i(s)\right] \\ &= a_n \int\int_{\gamma_0(s)}^{\gamma_0(s) + r/a_n} E\left[\left(c_0' x_i\right)^2 | v, s + b_n t\right] K(t) f(v, s + b_n t) \, dv dt \\ &= r c_0' D(\gamma_0(s), s) c_0 f(\gamma_0(s), s) + o(1). \end{aligned}$$

Next, given that $\left(\mathbf{1}_i\left(\gamma_0\left(s\right)+r/a_n\right)-\mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right)^2=\mathbf{1}_i\left(\gamma_0\left(s\right)+r/a_n\right)-\mathbf{1}_i\left(\gamma_0\left(s\right)\right)$ for $r>0$, we have

$$\begin{aligned}
Var\left[A_n^*\left(r,s\right)\right] &= \frac{a_n^2}{n^2b_n^2}Var\left[\sum_{i=1}^n\left(c_0'x_i\right)^2\left(\mathbf{1}_i\left(\gamma_0\left(s\right)+r/a_n\right)-\mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right)K_i\left(s\right)\right]\\
&= \frac{a_n^2}{nb_n^2}Var\left[\left(c_0'x_i\right)^2\left(\mathbf{1}_i\left(\gamma_0\left(s\right)+r/a_n\right)-\mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right)K_i\left(s\right)\right]\\
&\quad+\frac{2a_n^2}{n^2b_n^2}\sum_{i<j}^n Cov\left[\left(c_0'x_i\right)^2\left(\mathbf{1}_i\left(\gamma_0\left(s\right)+r/a_n\right)-\mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right)K_i\left(s\right),\right.\\
&\qquad\qquad\left.\left(c_0'x_j\right)^2\left(\mathbf{1}_j\left(\gamma_0\left(s\right)+r/a_n\right)-\mathbf{1}_j\left(\gamma_0\left(s\right)\right)\right)K_j\left(s\right)\right]\\
&\equiv \Psi_{A1}(r,s)+\Psi_{A2}(r,s).
\end{aligned}$$

Taylor expansion and Assumption A-(vii) and (viii) lead to that

$$\begin{aligned}
\Psi_{A1}(r,s) &= \frac{a_n^2}{nb_n^2}E\left[\left(c_0'x_i\right)^4\left(\mathbf{1}_i\left(\gamma_0\left(s\right)+r/a_n\right)-\mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right)K_i^2\left(s\right)\right]\\
&\quad-\frac{a_n^2}{nb_n^2}\left(E\left[\left(c_0'x_i\right)^2\left(\mathbf{1}_i\left(\gamma_0\left(s\right)+r/a_n\right)-\mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right)K_i\left(s\right)\right]\right)^2\\
&= O\left(n^{-2\epsilon}\right)\to 0.
\end{aligned}$$

Furthermore, by change of variables ($t_i=(s_i-s)/b_n$) in the covariance operator and Lemma 1 of Bolthausen (1982), for some $\varphi>0$,

$$\begin{aligned}
&\Psi_{A2}(r,s)\\
\le\;& \frac{2a_n^2}{n^2}\sum_{i<j}^n Cov\left[\left(c_0'x_i\right)^2\left(\mathbf{1}_i\left(\gamma_0\left(s\right)+r/a_n\right)-\mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right)K\left(t_i\right),\right.\\
&\qquad\qquad\left.\left(c_0'x_j\right)^2\left(\mathbf{1}_j\left(\gamma_0\left(s\right)+r/a_n\right)-\mathbf{1}_j\left(\gamma_0\left(s\right)\right)\right)K\left(t_j\right)\right]\\
\le\;& \frac{2a_n^2}{n}\sum_{m=1}^\infty m\alpha\left(m\right)^{2/(2+\varphi)}\left(E\left[\left|\left(c_0'x_i\right)^2\left(\mathbf{1}_i\left(\gamma_0\left(s\right)+r/a_n\right)-\mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right)K\left(t_i\right)\right|^{2+\varphi}\right]\right)^{2/(2+\varphi)}\\
=\;& O\left(n^{-1}\right)\to 0.
\end{aligned}$$

Hence, the pointwise convergence of $A_n^*\left(r,s\right)$ is obtained. Since $rc_0'D\left(\gamma_0\left(s\right),s\right)c_0f\left(\gamma_0\left(s\right),s\right)$ is strictly increasing and continuous in r, the convergence holds uniformly on any compact set. The same argument holds for negative r, which completes the proof for $A_n^*\left(r,s\right)$.

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

$$\begin{aligned}
&Cov\left[c_0'x_iu_i\left(\mathbf{1}_i\left(\gamma_0+r/a_n\right)-\mathbf{1}_i\left(\gamma_0\right)\right)K_i\left(s\right),\right.\\
&\qquad\left.c_0'x_ju_j\left(\mathbf{1}_j\left(\gamma_0+r/a_n\right)-\mathbf{1}_j\left(\gamma_0\right)\right)K_j\left(s\right)\right]\le Cb_n^2a_n^{-1}
\end{aligned}\tag{A.14}$$

for some positive constant $C<\infty$, by the change of variables in the covariance operator and

Lemma 1 of Bolthausen (1982). It follows that

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

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

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

as $n \to \infty$. This pointwise convergence in r can be extended to any finite-dimensional convergence in r by the fact that for any $r_1 < r_2$, $Cov[B_n^*(r_1,s), B_n^*(r_2,s)] = Var[B_n^*(r_1,s)] + o(1)$ since $(\mathbf{1}_i(\gamma_0 + r_2/a_n) - \mathbf{1}_i(\gamma_0 + r_1/a_n))\mathbf{1}_i(\gamma_0 + r_1/a_n) = 0$ and (A.14). 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). ■

Lemma A.4 *Let* $\widehat{\theta}(\widehat{\gamma}(s)) = (\widehat{\beta}(\widehat{\gamma}(s))', \widehat{\delta}(\widehat{\gamma}(s))')'$, $\widehat{\theta}(\gamma_0(s)) = (\widehat{\beta}(\gamma_0(s))', \widehat{\delta}(\gamma_0(s))')'$, *and* $\theta_0 = (\beta_0', \delta_0')'$ *for given* $s \in \mathcal{S}$. *Then, under Assumption A,*

$$
\sqrt{nb_n}\left(\widehat{\theta}(\gamma_0(s)) - \theta_0\right) = O_p(1) \quad \text{and} \quad \sqrt{nb_n}\left(\widehat{\theta}(\widehat{\gamma}(s)) - \widehat{\theta}(\gamma_0(s))\right) = o_p(1).
$$

Proof of Lemma A.4 For the first result, from (A.10), we have

$$
\begin{aligned}
& \sqrt{nb_n}\left(\widehat{\theta}(\gamma_0(s)) - \theta_0\right) \\
=& \left(\frac{1}{nb_n}\widetilde{Z}(\gamma_0; s)'\widetilde{Z}(\gamma_0; s)\right)^{-1}\left(\frac{1}{\sqrt{nb_n}}\widetilde{Z}(\gamma_0; s)'\widetilde{u}(s)\right) \\
=& \begin{pmatrix} \frac{1}{nb_n}\sum_{i=1}^n x_i x_i' K_i(s) & M_n(\gamma_0; s) \\ M_n(\gamma_0; s) & M_n(\gamma_0; s) \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{nb_n}}\sum_{i=1}^n x_i' u_i K_i(s) \\ J_n(\gamma_0; s) \end{pmatrix} \\
=& O_p(1)
\end{aligned}
$$

from Lemma A.1, where $(nb_n)^{-1}\sum_{i=1}^n x_i x_i' K_i(s) \to_p M(s) < \infty$ for some positive definite $M(s)$ and $(nb_n)^{-1/2}\sum_{i=1}^n x_i' u_i K_i(s) = O_p(1)$.

For the second result, we let $\widehat{z}_i(s) = [x_i', x_i'\mathbf{1}_i(\widehat{\gamma}(s))]'$, $z_i(s) = [x_i', x_i'\mathbf{1}_i(\gamma_0(s))]'$, and

$z_i = [x_i', x_i'\mathbf{1}_i(\gamma_0(s_i))]'$. Then, $y_i = z_i'\theta_0 + u_i$. Using a similar expression as above, we have

$$\begin{aligned}
& \sqrt{nb_n}\left(\widehat{\theta}(\widehat{\gamma}(s)) - \widehat{\theta}(\gamma_0(s))\right) \qquad \text{(A.15)} \\
= & \sqrt{nb_n}\left(\widehat{\theta}(\widehat{\gamma}(s)) - \theta_0\right) - \sqrt{nb_n}\left(\widehat{\theta}(\gamma_0(s)) - \theta_0\right) \\
= & \left(C + o_p\left(\frac{1}{nb_n}\right)\right)^{-1} \frac{1}{\sqrt{nb_n}} \sum_{i=1}^{n} \left\{\widehat{z}_i(s)\left(y_i - \widehat{z}_i(s)'\theta_0\right) - z_i(s)\left(y_i - z_i(s)'\theta_0\right)\right\} K_i(s) \\
= & \left(C + o_p\left(\frac{1}{nb_n}\right)\right)^{-1} \left\{\frac{1}{\sqrt{nb_n}} \sum_{i=1}^{n} \left(\widehat{z}_i(s) - z_i\right) u_i K_i(s)\right. \\
& \left. -\frac{1}{\sqrt{nb_n}} \sum_{i=1}^{n} \widehat{z}_i(s)\left(\widehat{z}_i(s) - z_i\right)' \theta_0 K_i(s) - \frac{1}{\sqrt{nb_n}} \sum_{i=1}^{n} z_i(s)\left(z_i - z_i(s)\right)' \theta_0 K_i(s)\right\}
\end{aligned}$$

for some $0 < C < \infty$. However, since $\widehat{z}_i(s) - z_i = [0, x_i'(\mathbf{1}_i(\widehat{\gamma}(s)) - \mathbf{1}_i(\gamma_0(s_i)))]'$,

$$\begin{aligned}
& \frac{1}{\sqrt{nb_n}} \sum_{i=1}^{n} \left(\widehat{z}_i(s) - z_i\right) u_i K_i(s) \\
= & \left[0, \frac{1}{\sqrt{nb_n}} \sum_{i=1}^{n} x_i u_i \left(\mathbf{1}_i(\widehat{\gamma}(s)) - \mathbf{1}_i(\gamma_0(s))\right) K_i(s)\right]' \\
= & \left[0, \frac{1}{\sqrt{n^{1-2\epsilon} b_n}} \sum_{i=1}^{n} n^{-\epsilon} x_i u_i \left(\mathbf{1}_i(\widehat{\gamma}(s)) - \mathbf{1}_i(\gamma_0(s))\right) K_i(s)\right]' \\
= & O_p((n^{1-2\epsilon} b_n)^{-1/2}) \to 0
\end{aligned}$$

by Lemma A.3 and Assumption A-(x). Similarly, we also have

$$\begin{aligned}
& \frac{1}{\sqrt{nb_n}} \sum_{i=1}^{n} \widehat{z}_i(s)\left(\widehat{z}_i(s) - z_i\right)' \theta_0 K_i(s) \\
= & \left[\begin{array}{c} (n^{1-2\epsilon} b_n)^{-1/2} \sum_{i=1}^{n} n^{-\epsilon} x_i x_i' \delta_0 \left(\mathbf{1}_i(\widehat{\gamma}(s)) - \mathbf{1}_i(\gamma_0(s_i))\right) K_i(s) \\ (n^{1-2\epsilon} b_n)^{-1/2} \sum_{i=1}^{n} n^{-\epsilon} x_i x_i' \delta_0 \left(\mathbf{1}_i(\widehat{\gamma}(s)) - \mathbf{1}_i(\min\{\widehat{\gamma}(s), \gamma_0(s_i)\})\right) K_i(s) \end{array}\right] \\
= & O_p((n^{1-2\epsilon} b_n)^{-1/2}) \to 0
\end{aligned}$$

since $(\widehat{z}_i(s) - z_i)' \theta_0 = \delta_0' x_i (\mathbf{1}_i(\widehat{\gamma}(s)) - \mathbf{1}_i(\gamma_0(s_i)))$. The last component in (A.15) can be shown to be $O_p((n^{1-2\epsilon} b_n)^{-1/2})$ as well using the same argument, which completes the proof. ■

Lemma A.5 *Define $a_n = n^{1-2\epsilon} b_n$, where ϵ is given in Assumption A-(iii). For any fixed $s \in \mathcal{S}$,*

$$\widehat{\gamma}(s) - \gamma_0(s) = O_p(a_n^{-1})$$

under Assumption A.

Proof of Lemma A.5 For given $s \in \mathcal{S}$, we let

$$\begin{aligned} Q_n^*(\gamma(s); s) &= Q_n(\widehat{\beta}(\widehat{\gamma}(s)), \widehat{\delta}(\widehat{\gamma}(s)), \gamma(s); s) \\ &= \sum_{i=1}^n K\left(\frac{s_i - s}{b_n}\right)\left(y_i - x_i'\widehat{\beta}(\widehat{\gamma}(s)) - x_i'\widehat{\delta}(\widehat{\gamma}(s))\mathbf{1}[q_i \le \gamma(s)]\right)^2 \end{aligned} \quad (A.16)$$

for any $\gamma(\cdot)$, where $Q_n(\beta, \delta, \gamma; s)$ is the sum of squared errors function in (3). We consider $\gamma(s)$ such that $|\gamma(s) - \gamma_0(s)| \in [r(s)/a_n, C(s)]$ for some $0 < r(s), C(s) < \infty$. Then, given Lemma A.3, it can be verified that $P(Q_n^*(\gamma(s); s) - Q_n^*(\gamma_0(s); s) > 0) \to 1$ as $n \to \infty$ using the standard results in kernel regression (e.g., Li and Racine, 2007, Ch.2) and following Lemma A.9 in Hansen (2000). The detailed proof is similar to the proof of Theorem 3 below and hence omitted. Therefore, with probability approaching to one (w.p.a.1, hereafter), it should hold that $|\widehat{\gamma}(s) - \gamma_0(s)| \le r(s)/a_n$ since $Q_n^*(\widehat{\gamma}(s); s) - Q_n^*(\gamma_0(s); s) \le 0$ for any $s \in \mathcal{S}$ by construction. ■

Proof of Theorem 1 The pointwise consistency is proved in Lemma A.2 above. To derive the limiting distribution, from Lemma A.5, we define a random variable $r^*(s)$ such that

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

where $Q_n^*(\gamma(s); s)$ is defined in (A.16). We let $\Delta_i(r; s) = \mathbf{1}_i(\gamma_0(s) + (r/a_n)) - \mathbf{1}_i(\gamma_0(s))$. We then have

$$\begin{aligned} & Q_n^*(\gamma_0(s); s) - Q_n^*\left(\gamma_0(s) + \frac{r(s)}{a_n}; s\right) \qquad (A.17) \\ = & -\sum_{i=1}^n \left(\widehat{\delta}(\widehat{\gamma}(s))' x_i\right)^2 \Delta_i(r; s) K_i(s) \\ & +2\sum_{i=1}^n \left(y_i - \widehat{\beta}(\widehat{\gamma}(s))' x_i - \widehat{\delta}(\widehat{\gamma}(s))' x_i \mathbf{1}_i(\gamma_0(s))\right)\left(\widehat{\delta}(\widehat{\gamma}(s))' x_i\right)\Delta_i(r; s) K_i(s) \\ \equiv & -A_n(r; s) + 2B_n(r; s). \end{aligned}$$

For $A_n(r; s)$, Lemmas A.3 and A.4 yield

$$\begin{aligned} A_n(r; s) &= \sum_{i=1}^n \left(\left(\delta_0 + n^{-1/2} b_n^{-1/2} C_\delta + o_p(n^{-1/2} b_n^{-1/2})\right)' x_i\right)^2 \Delta_i(r; s) K_i(s) \qquad (A.18) \\ &= A_n^*(r, s) + \frac{1}{n^{1-2\epsilon} b_n} \sum_{i=1}^n \left(n^{-\epsilon} C_\delta\right)' x_i x_i' \left(n^{-\epsilon} C_\delta\right) \Delta_i(r; s) K_i(s) + o_p(1) \\ &= A_n^*(r, s) + O_p\left(\left(n^{1-2\epsilon} b_n\right)^{-1}\right) + o_p(1) \\ &= A_n^*(r, s) + o_p(1) \end{aligned}$$

for some $C_\delta < \infty$, since $n^{1-2\epsilon} b_n \to \infty$ and $\sum_{i=1}^n n^{-2\epsilon} C_\delta' x_i x_i' C_\delta \Delta_i(r; s) K_i(s) = O_p(1)$ from Lemma A.3. Note that $\widehat{\delta}(\widehat{\gamma}(s)) - \delta_0 = (\widehat{\delta}(\widehat{\gamma}(s)) - \widehat{\delta}(\gamma_0(s))) + (\widehat{\delta}(\gamma_0(s)) - \delta_0) = O_p(n^{-1/2} b_n^{-1/2})$ from Lemma A.4. Similarly, for $B_n(r; s)$, since $y_i = \beta_0' x_i + \delta_0' x_i \mathbf{1}_i(\gamma_0(s_i)) + u_i$, we have for some $C_\beta < \infty$

$$
\begin{aligned}
& B_n(r;s) & \text{(A.19)} \\
= \; & \sum_{i=1}^{n} \Big(u_i + \delta_0' 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)' x_i \\
& \qquad - \left(\widehat{\delta}(\widehat{\gamma}(s)) - \delta_0 \right)' x_i \mathbf{1}_i(\gamma_0(s)) \Big) \widehat{\delta}(\gamma_0(s))' x_i \Delta_i(r;s) K_i(s) \\
= \; & \sum_{i=1}^{n} \Big(u_i + \delta_0' 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' x_i - n^{-1/2} b_n^{-1/2} C_\delta' x_i \mathbf{1}_i(\gamma_0(s)) \Big) \\
& \quad \times \left(\delta_0 + n^{-1/2} b_n^{-1/2} C_\delta \right)' x_i \Delta_i(r;s) K_i(s) + o_p(1) \\
= \; & 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 \right) \Delta_i(r;s) K_i(s) \\
& + \sum_{i=1}^{n} \delta_0' x_i x_i' \delta_0 \left(\Delta_i(r;s) \left\{ \mathbf{1}_i(\gamma_0(s_i)) - \mathbf{1}_i(\gamma_0(s)) \right\} \right) K_i(s) & \text{(A.20)} \\
& + \frac{1}{\sqrt{n^{1-2\epsilon} b_n}} \sum_{i=1}^{n} \delta_0' x_i x_i' \left(n^{-\epsilon} C_\delta \right) \left(\Delta_i(r;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\prime} \delta_0' x_i x_i' \left(n^{-\epsilon} C_\beta \right) \Delta_i(r;s) K_i(s) \\
& + \frac{1}{n^{1-2\epsilon} b_n} \sum_{i=1}^{n} \left(n^{-\epsilon} C_\beta \right)' x_i x_i' \left(n^{-\epsilon} C_\delta \right) \Delta_i(r;s) K_i(s) \\
& + \frac{1}{\sqrt{n^{1-2\epsilon} b_n}} \sum_{i=1}^{n\prime} \delta_0' x_i x_i' \left(n^{-\epsilon} C_\delta \right) \left\{ \Delta_i(r;s) \mathbf{1}_i(\gamma_0(s)) \right\} K_i(s) \\
& + \frac{1}{n^{1-2\epsilon} b_n} \sum_{i=1}^{n} \left(n^{-\epsilon} C_\delta \right)' x_i x_i' \left(n^{-\epsilon} C_\delta \right) \left\{ \Delta_i(r;s) \mathbf{1}_i(\gamma_0(s)) \right\} K_i(s) + o_p(1) \\
= \; & B_n^*(r,s) + O_p\left(\left(n^{1-2\epsilon} b_n \right)^{-1/2} \right) + O_p(n^{1-2\epsilon} b_n^3) + O_p\left(\left(n^{1-2\epsilon} b_n \right)^{-1/2} \right) + o_p(1) \\
= \; & B_n^*(r,s) + o_p(1)
\end{aligned}
$$

from Lemma A.3, since we assume $n^{1-2\epsilon} b_n \to \infty$ and $n^{1-2\epsilon} b_n^3 \to 0$. Note that the term on (A.20), denoting $B_{n3}(r;s)$, is $O_p\left(n^{1-2\epsilon} b_n^3\right)$, and hence we need $n^{1-2\epsilon} b_n^3 \to 0$ to make this term negligible. To see this, similarly as $E\left[M_n(\gamma;s)\right]$ in the proof of Lemma A.1, we have

$$
\begin{aligned}
E\left[B_{n3}(r;s)\right] \;=\; & \frac{b_n}{n^{2\epsilon-1}} \iint c_0' D(q, s + b_n t) c_0 \left\{ \mathbf{1}[q \le \gamma_0(s) + (r/a_n)] - \mathbf{1}[q \le \gamma_0(s)] \right\} \\
& \times \left\{ \mathbf{1}[q \le \gamma_0(s + b_n t)] - \mathbf{1}[q \le \gamma_0(s)] \right\} K(t) f(q, s + b_n t) \, dq dt.
\end{aligned}
$$

However, since[5]

$$\begin{aligned}
& \{\mathbf{1}[q \leq \gamma_0(s) + (r/a_n)] - \mathbf{1}[q \leq \gamma_0(s)]\}\{\mathbf{1}[q \leq \gamma_0(s + b_n t)] - \mathbf{1}[q \leq \gamma_0(s)]\} \\
= & \mathbf{1}[\gamma_0(s) < q \leq \min\{\gamma_0(s + b_n t), \gamma_0(s) + (r/a_n)\}] \\
& + \mathbf{1}[\max\{\gamma_0(s + b_n t), \gamma_0(s) + (r/a_n)\} < q \leq \gamma_0(s)] \\
\leq & \mathbf{1}[\gamma_0(s) < q \leq \gamma_0(s + b_n t)] + \mathbf{1}[\gamma_0(s + b_n t) < q \leq \gamma_0(s)],
\end{aligned}$$

we have

$$\begin{aligned}
E[B_{n3}(r;s)] \leq & \frac{b_n}{n^{2\epsilon-1}} \iint_{\gamma_0(s)}^{\gamma_0(s+b_n t)} c_0' D(q, s + b_n t) c_0 K(t) f(q, s + b_n t) \, dq dt \\
& + \frac{b_n}{n^{2\epsilon-1}} \iint_{\gamma_0(s+b_n t)}^{\gamma_0(s)} c_0' D(q, s + b_n t) c_0 K(t) f(q, s + b_n t) \, dq dt \\
= & O\left(n^{1-2\epsilon} b_n^3\right),
\end{aligned}$$

which is because

$$\begin{aligned}
& \iint_{\gamma_0(s)}^{\gamma_0(s+b_n t)} c_0' D(q, s + b_n t) c_0 K(t) f(q, s + b_n t) \, dq dt \\
= & \int \left(\int_{\gamma_0(s)}^{\gamma_0(s+b_n t)} c_0' D(q, s) c_0 f(q, s) \, dq \right) \left(1 + C_1 b_n^2 t^2\right) K(t) dt \\
= & \int \left(C_{21} b_n t + C_{22} b_n^2 t^2\right) \left(1 + C_1 b_n^2 t^2\right) K(t) dt \\
= & O\left(b_n^2\right)
\end{aligned}$$

for some $C_1, C_{21}, C_{22} < \infty$, similarly as (A.12). The other term can be verified symmetrically. It hence follows that

$$Q_n^*(\gamma_0; s) - Q_n^*\left(\gamma_0 + \frac{r}{a_n}; s\right) = -A_n^*(r, s) + 2B_n^*(r, s) + o_p(1)$$

and the desired result follows from Lemma A.3 using the same argument of the proof of Theorem 1 in Hansen (2000). ■

Proof of Theorem 2 From (A.11) and (A.13), we have

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

[5] Note that

$$\begin{aligned}
& \mathbf{1}[r_1 < q \leq \min\{r_2, r_3\}] + \mathbf{1}[\max\{r_2, r_3\} < q \leq r_1] \\
= & \left\{ \begin{array}{ll} \mathbf{1}[r_1 < q \leq r_2] + \mathbf{1}[r_3 < q \leq r_1] & \text{if } r_2 \leq r_3 \\ \mathbf{1}[r_1 < q \leq r_3] + \mathbf{1}[r_2 < q \leq r_1] & \text{if } r_2 > r_3 \end{array} \right\} \\
\leq & \left\{ \begin{array}{ll} \mathbf{1}[r_1 < q \leq r_2] + \mathbf{1}[r_2 < q \leq r_1] & \text{if } r_2 \leq r_3 \\ \mathbf{1}[r_1 < q \leq r_2] + \mathbf{1}[r_2 < q \leq r_1] & \text{if } r_2 > r_3 \end{array} \right\} \\
= & \mathbf{1}[r_1 < q \leq r_2] + \mathbf{1}[r_2 < q \leq r_1].
\end{aligned}$$

where $f_s(s)$ is the marginal density of s_i. In addition, from Theorem 1 and the proof of Lemma A.4, 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 $\widehat{\theta}(\widehat{\gamma}(s)) - \widehat{\theta}(\gamma_0(s)) = o_p((nb_n)^{-1/2})$. Following the proof of Theorem 2 of Hansen (2002), the rest of the proof follows from the change of variables and the continuous mapping theorem because $(nb_n)^{-1}\sum_{i=1}^n K_i(s) \to_p f_s(s)$ by the standard result of kernel density estimation. ■

Lemma A.6 *Define $a_n = n^{1-2\epsilon}b_n$ and $\phi_n = (\log n)/a_n + b_n^2$, where ϵ is given in Assumption A-(iii). For given $s \in \mathcal{S}$, let*

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

Then, for any $\eta > 0$ and $\varepsilon > 0$, there exist constants $0 < \overline{C}, \overline{r} < \infty$ such that

$$\begin{aligned} P\left(\inf_{\overline{r}\phi_n < \sup_{s\in\mathcal{S}}|\gamma(s)-\gamma_0(s)| < \overline{C}} \frac{\sup_{s\in\mathcal{S}} T_n(\gamma; s)}{\sup_{s\in\mathcal{S}}|\gamma(s) - \gamma_0(s)|} < \eta\right) &< \varepsilon, \\ P\left(\sup_{\overline{r}\phi_n < \sup_{s\in\mathcal{S}}|\gamma(s)-\gamma_0(s)| < \overline{C}} \frac{\sup_{s\in\mathcal{S}}|L_n(\gamma; s)|}{\sqrt{a_n}\sup_{s\in\mathcal{S}}|\gamma(s) - \gamma_0(s)|} > \eta\right) &\leq \varepsilon \end{aligned}$$

under the same condition in Theorem 1, provided that either (11) or (12) hold.

Proof of Lemma A.6 We consider the case $\gamma(s) \in \left[\gamma_0(s) + \overline{r}(s)\phi_n, \gamma_0(s) + \overline{C}(s)\right]$ for some $0 < \overline{r}(s), \overline{C}(s) < \infty$, where $\overline{r} = \sup_{s\in\mathcal{S}}\overline{r}(s) < \infty$ and $\overline{C} = \sup_{s\in\mathcal{S}}\overline{C}(s) < \infty$; the other direction can be shown symmetrically. For the first result, from (A.1) we have

$$E\left[T_n(\gamma; s)\right] = c_0'\left(M(\gamma(s); s) - M(\gamma_0(s); s)\right)c_0 + \tau(s)b_n^2$$

for some $0 < \tau(s) < \infty$, where $M(\gamma; s) = \int_{-\infty}^{\gamma} D(q,s)f(q,s)\,dq$. However, for given $s \in \mathcal{S}$, since $\partial E\left[T_n(\gamma; s)\right]/\partial\gamma(s) = c_0'D(\gamma(s), s)c_0 f(\gamma(s), s)$ is continuous in $\gamma(s)$ and $c_0'D(\gamma_0(s), s)c_0 f(\gamma_0(s), s) > 0$ from Assumptions A-(vii) and (viii), there exists $\overline{C}(s) < \infty$ such that

$$\mu_1(s) = \inf_{|\gamma(s)-\gamma_0(s)|<\overline{C}(s)} c_0'D(\gamma(s), s)c_0 f(\gamma(s), s) > 0 \ \text{ and } \ \overline{\mu}_1 = \inf_{s\in\mathcal{S}}\mu_1(s) > 0.$$

Hence, for $E\left[T_n(\gamma_0; s)\right] = 0$, Taylor expansion yields

$$\sup_{s\in\mathcal{S}} E\left[T_n(\gamma; s)\right] \geq \overline{\mu}_1 \sup_{s\in\mathcal{S}}\left(\gamma(s) - \gamma_0(s)\right) + \overline{\tau}b_n^2 > \overline{\mu}_1 \sup_{s\in\mathcal{S}}\left(\gamma(s) - \gamma_0(s)\right), \tag{A.21}$$

where we use the fact that $\overline{\tau} = \sup_{s\in\mathcal{S}} \tau(s) \in (0,\infty)$ from the standard result.[6] Furthermore, by Theorem 2.2 in Carbon et al. (2007), we can similarly show that

$$E\left(\sup_{s\in\mathcal{S}}\left|T_n(\gamma;s) - E\left[T_n(\gamma;s)\right]\right|\right)^2 \le \sup_{s\in\mathcal{S}}\left(\gamma(s)-\gamma_0(s)\right)\overline{\mu}_2\frac{\log n}{nb_n} \tag{A.22}$$

for some $\overline{\mu}_2 \in (0,\infty)$ under Assumptions A-(vii) and (viii) and the conditions on b_n in either (11) or (12). Then following Lemma A.7 in Hansen (2000), set γ_g for $g = 1,2,...,G_n+1$ such that for any $s\in\mathcal{S}$, $\gamma_g(s) = \gamma_0(s) + 2^{g-1}\overline{r}(s)\phi_n$ where G_n is the integer such that $\gamma_{G_n}(s) - \gamma_0(s) = 2^{G_n-1}\overline{r}(s)\phi_n \le \overline{C}$ and $\gamma_{G_n+1}(s) - \gamma_0(s) = 2^{G_n}\overline{r}(s)\phi_n > \overline{C}$. Then Markov's inequality and (A.21), (A.22) yield that for any fixed $\eta > 0$,

$$\begin{aligned}
& P\left(\sup_{1\le g\le G_n}\left|\frac{\sup_{s\in\mathcal{S}} T_n(\gamma_g;s)}{\sup_{s\in\mathcal{S}} E\left[T_n(\gamma_g;s)\right]} - 1\right| > \eta\right) && \text{(A.23)}\\
\le\ & P\left(\sup_{1\le g\le G_n}\left|\frac{\sup_{s\in\mathcal{S}} T_n(\gamma_g;s) - \sup_{s\in\mathcal{S}} E\left[T_n(\gamma_g;s)\right]}{\sup_{s\in\mathcal{S}} E\left[T_n(\gamma_g;s)\right]}\right| > \eta\right)\\
\le\ & P\left(\sup_{1\le g\le G_n}\left|\frac{\sup_{s\in\mathcal{S}} \left|T_n(\gamma_g;s) - E\left[T_n(\gamma_g;s)\right]\right|}{\sup_{s\in\mathcal{S}} E\left[T_n(\gamma_g;s)\right]}\right| > \eta\right)\\
\le\ & \frac{1}{\eta^2}\sum_{g=1}^{G_n}\frac{E\left[\left(\sup_{s\in\mathcal{S}}\left|T_n(\gamma_g;s) - E\left[T_n(\gamma_g;s)\right]\right|\right)^2\right]}{\left|\sup_{s\in\mathcal{S}} E\left[T_n(\gamma_g;s)\right]\right|^2}\\
\le\ & \frac{1}{\eta^2}\sum_{g=1}^{G_n}\frac{\overline{\mu}_2\sup_{s\in\mathcal{S}}\left(\gamma_g(s)-\gamma_0(s)\right)\left((\log n)/nb_n\right)}{\left|\overline{\mu}_1\sup_{s\in\mathcal{S}}\left(\gamma_g(s)-\gamma_0(s)\right)\right|^2}\\
\le\ & \frac{\overline{\mu}_2}{\eta^2\overline{\mu}_1^2\overline{r}}\sum_{g=1}^{\infty}\frac{1}{2^{g-1}}\times\frac{1}{n^{2\epsilon}+(nb_n^3/\log n)}\\
\le\ & \varepsilon
\end{aligned}$$

for any ε. Note that $n^{2\epsilon} + (nb_n^3/\log n) \to \infty$ as $n\to\infty$, which does not require any conditions on $nb_n^3/\log n$. Then following eq. (33) of Hansen (2000), for any $\gamma(\cdot)$ in the set $\{\gamma(\cdot) : \overline{r}\phi_n \le \sup_{s\in\mathcal{S}}(\gamma(s)-\gamma_0(s)) \le \overline{C}\}$, there exist g such that $\gamma_g(s) - \gamma_0(s) < \gamma(s) - \gamma_0(s) < \gamma_{g+1}(s) - \gamma_0(s)$ and

$$\begin{aligned}
\frac{\sup_{s\in\mathcal{S}} T_n(\gamma;s)}{\sup_{s\in\mathcal{S}}\left|\gamma(s)-\gamma_0(s)\right|} &\ge \frac{\sup_{s\in\mathcal{S}} T_n(\gamma_g;s)}{\sup_{s\in\mathcal{S}} E\left[T_n(\gamma_g;s)\right]}\times\frac{\sup_{s\in\mathcal{S}} E\left[T_n(\gamma_g;s)\right]}{\sup_{s\in\mathcal{S}}\left|\gamma_{g+1}(s)-\gamma_0(s)\right|}\\
&\ge (1-o_p(1))\left(\overline{\mu}_1 + O_p\left(b_n^2\right)\right).
\end{aligned}$$

[6] From the standard nonparametric estimation result,

$$\tau(s) = \frac{1}{2}\int t^2K(t)dt\int_{-\infty}^{\gamma(s)}\left(\ddot{D}(q,s)f(q,s) + D(q,s)\ddot{f}(q,s)\right)dq + O\left(b_n^2\right),$$

which satisfies $\sup_{s\in\mathcal{S}}\tau(s)\in(0,\infty)$ from Assumptions A-(vii) and (viii).

This inequality yields that for any ε and η

$$P\left(\inf_{\overline{r}\phi_n<\sup_{s\in\mathcal{S}}|\gamma(s)-\gamma_0(s)|<\overline{C}}\frac{\sup_{s\in\mathcal{S}}T_n(\gamma;s)}{\sup_{s\in\mathcal{S}}|\gamma(s)-\gamma_0(s)|}<\eta\right)<\varepsilon.$$

For the second result, Theorem 2.2 in Carbon et al. (2007) yields that, for a large enough n,

$$E\left(\sup_{s\in\mathcal{S}}|L_n(\gamma;s)|\right)^2\leq\sup_{s\in\mathcal{S}}(\gamma(s)-\gamma_0(s))\,\overline{\mu}_3\log n$$

for some $\overline{\mu}_2\in(0,\infty)$ similarly as above, since $E[L_n(\gamma;s)]=0$. Following Lemma A.8 in Hansen (2000), for $g=1,2,...$, set γ_g such that for any $s\in\mathcal{S}$, $\gamma_g(s)-\gamma_0(s)=2^{g-1}\overline{r}(s)\phi_n$. Then using a similar approach as (A.23), Assumption A-(iii) yield, for any fixed $\eta>0$,

$$\begin{aligned}
& P\left(\sup_{g>0}\frac{\sup_{s\in\mathcal{S}}\left|L_n\left(\gamma_g;s\right)\right|}{\sqrt{a_n}\sup_{s\in\mathcal{S}}\left(\gamma_g(s)-\gamma_0(s)\right)}>\eta\right) \qquad (A.24)\\
\leq\ & \frac{1}{\eta^2}\sum_{g=1}^{\infty}\frac{E\left[\left(\sup_{s\in\mathcal{S}}L_n\left(\gamma_g,s\right)\right)^2\right]}{a_n\left|\sup_{s\in\mathcal{S}}\left(\gamma_g(s)-\gamma_0(s)\right)\right|^2}\\
=\ & \frac{1}{\eta^2}\sum_{g=1}^{\infty}\frac{\sup_{s\in\mathcal{S}}\left(\gamma_g(s)-\gamma_0(s)\right)\overline{\mu}_3\log n}{a_n\left|\sup_{s\in\mathcal{S}}\left(\gamma_g(s)-\gamma_0(s)\right)\right|^2}\\
\leq\ & \frac{\overline{\mu}_3}{\eta^2\overline{r}}\sum_{g=1}^{\infty}\frac{1}{2^{g-1}}\times\frac{1}{1+(n^{1-2\epsilon}b_n^3)/\log n}.
\end{aligned}$$

The above probability is arbitrarily close to 0 if $\overline{r}$ is large enough since $(n^{1-2\epsilon}b_n^3)/\log n\to 0$ under the assumption. Also define Γ_g to be the collection of functions $\{\gamma(\cdot):\overline{r}(s)2^{g-1}\phi_n<\gamma(s)-\gamma_0(s)<\overline{r}(s)2^g\phi_n\}$. By a similar argument as above

$$P\left(\sup_{g>0}\sup_{\gamma\in\Gamma_g}\frac{\sup_{s\in\mathcal{S}}\left|L_n\left(\gamma_g;s\right)\right|}{\sqrt{a_n}\sup_{s\in\mathcal{S}}\left(\gamma_g(s)-\gamma_0(s)\right)}>\eta\right)\leq\frac{C}{\eta^2\overline{r}} \qquad (A.25)$$

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

$$\begin{aligned}
& P\left(\sup_{\overline{r}\phi_n<\sup_{s\in\mathcal{S}}|\gamma(s)-\gamma_0(s)|<\overline{C}}\frac{\sup_{s\in\mathcal{S}}\left|L_n\left(\gamma_g;s\right)\right|}{\sqrt{a_n}\sup_{s\in\mathcal{S}}\left(\gamma(s)-\gamma_0(s)\right)}>\eta\right)\\
\leq\ & 2P\left(\sup_{g>0}\frac{\sup_{s\in\mathcal{S}}\left|L_n\left(\gamma_g;s\right)\right|}{\sqrt{a_n}\sup_{s\in\mathcal{S}}\left(\gamma_g(s)-\gamma_0(s)\right)}>\eta\right)\\
& +2P\left(\sup_{g>0}\sup_{\gamma\in\Gamma_g}\frac{\sup_{s\in\mathcal{S}}\left|L_n\left(\gamma_g;s\right)\right|}{\sqrt{a_n}\sup_{s\in\mathcal{S}}\left(\gamma_g(s)-\gamma_0(s)\right)}>\eta\right)\\
\leq\ & \varepsilon
\end{aligned}$$

for any ε if $\overline{r}$ is sufficiently large. ■

Proof of Theorem 3 Let $a_n = n^{1-2\epsilon} b_n$. Since $\sup_{s\in\mathcal{S}} \left(Q_n^*(\widehat{\gamma}(s); s) - Q_n^*(\gamma_0(s); s)\right) \leq 0$ by construction, where $Q_n^*(\gamma(s); s)$ is defined in (A.16), it suffices to show that as $n \to \infty$, $P\left(\sup_{s\in\mathcal{S}} \left(Q_n^*(\gamma(s); s) - Q_n^*(\gamma_0(s); s)\right) > 0\right) \to 1$ for any $\gamma(s)$ such that $\sup_{s\in\mathcal{S}} |\gamma(s) - \gamma_0(s)| \not< \overline{r}((\log n)/a_n) + b_n^2)$ for some $0 < \overline{r} < \infty$. To this end, for given $s \in \mathcal{S}$, consider $\gamma(s) \in \left[\gamma_0(s) + \overline{r}(s)((\log n)/a_n) + b_n^2), \gamma_0(s) + \overline{C}(s)\right]$ for some $0 < \overline{r}(s), \overline{C}(s) < \infty$, where $\overline{r} = \sup_{s\in\mathcal{S}} \overline{r}(s) < \infty$ and $\overline{C} = \sup_{s\in\mathcal{S}} \overline{C}(s) < \infty$. Then, using a similar decomposition in (A.17), (A.18), and (A.19), since $\gamma(s) > \gamma_0(s)$,

$$
\begin{aligned}
& Q_n^*(\gamma(s); s) - Q_n^*(\gamma_0(s); s) \\
= \; & \sum_{i=1}^n \left(\widehat{\delta}(\widehat{\gamma}(s))' x_i\right)^2 \Delta_i(s) K_i(s) \\
& -2\sum_{i=1}^n \left(y_i - \widehat{\beta}(\widehat{\gamma}(s))' x_i - \widehat{\delta}(\widehat{\gamma}(s))' x_i \mathbf{1}_i(\gamma_0(s))\right) \left(\widehat{\delta}(\widehat{\gamma}(s))' x_i\right) \Delta_i(s) K_i(s) \\
\equiv \; & a_n T_n(\gamma; s) - 2\sqrt{a_n} L_n(\gamma; s) + o_p(1),
\end{aligned}
$$

where $\Delta_i(s) = \mathbf{1}_i(\gamma(s)) - \mathbf{1}_i(\gamma_0(s))$, and $T_n(\gamma; s)$ and $L_n(\gamma; s)$ are defined in Lemma A.6. However, Lemma A.6 yields that, for any $\eta > 0$ and $\varepsilon > 0$,

$$
P\left(\sup_{\overline{r}((\log n)/a_n) + b_n^2) < \sup_{s\in\mathcal{S}} |\gamma(s) - \gamma_0(s)| < \overline{C}} \sup_{s\in\mathcal{S}} \left(Q_n^*(\gamma(s); s) - Q_n^*(\gamma_0(s); s)\right) > \eta\right) \geq 1 - \varepsilon.
$$

We can similarly show that if $\gamma(s) \in \left[\gamma_0(s) - \overline{C}(s), \gamma_0(s) - \overline{r}(s)((\log n)/a_n) + b_n^2)\right]$ then $P\left(\sup_{s\in\mathcal{S}} \left(Q_n^*(\gamma(s); s) - Q_n^*(\gamma_0(s); s)\right) > 0\right) \to 1$ as well, which completes the proof. ■

Proof of Theorem 4 Let $\widehat{z}_i = [x_i', x_i' \mathbf{1}_i(\widehat{\gamma}(s_i))]'$, $z_i = [x_i', x_i' \mathbf{1}_i(\gamma_0(s_i))]'$, and $\widehat{\Delta}_i(s_i) = \mathbf{1}_i(\widehat{\gamma}(s_i)) - \mathbf{1}_i(\gamma_0(s_i))$. Then,

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\theta} - \theta_0\right) &= \left(\frac{1}{n}\sum_{i=1}^n \widehat{z}_i \widehat{z}_i'\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{z}_i \left(u_i - (\widehat{z}_i - z_i)' \theta_0\right)\right) \\
&= \left(\frac{1}{n}\sum_{i=1}^n \widehat{z}_i \widehat{z}_i'\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{z_i u_i + (z_i - \widehat{z}_i) u_i + \widehat{z}_i (\widehat{z}_i - z_i)' \theta_0\right\}\right)
\end{aligned}
$$

and it suffices to establish

$$
\frac{1}{n}\sum_{i=1}^n \widehat{z}_i \widehat{z}_i' \to_p M^* \tag{A.26}
$$

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i u_i \to_d \mathcal{N}(0, V^*) \tag{A.27}
$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{z}_i\left(\widehat{z}_i-z_i\right)'\theta_0=o_p\left(1\right) \tag{A.28}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_i-\widehat{z}_i\right)u_i=o_p\left(1\right). \tag{A.29}$$

First, by Assumptions A-(vi), (x), and Theorem 1, (A.26) can be readily verified since $n^{-1}\sum_{i=1}^{n}x_ix_i'\mathbf{1}_i\left(\widehat{\gamma}\left(s_i\right)\right)=n^{-1}\sum_{i=1}^{n}x_ix_i'\mathbf{1}_i\left(\gamma_0\left(s_i\right)\right)+n^{-1}\sum_{i=1}^{n}x_ix_i'\widehat{\Delta}_i\left(s_i\right)$ and

$$\begin{aligned}
E\left[x_ix_i'\widehat{\Delta}_i\left(s_i\right)\right] &\leq \int\left|\int_{\gamma_0(v)}^{\widehat{\gamma}(v)}D\left(q,v\right)f\left(q,v\right)dq\right|dv \\
&= \int\left\{\left|D\left(\gamma_0\left(v\right),v\right)f\left(\gamma_0\left(v\right),v\right)\right|O_p\left(\frac{1}{n^{1-2\epsilon}b_n}\right)\right\}dv \\
&= O_p\left(\frac{1}{n^{1-2\epsilon}b_n}\right)\to 0
\end{aligned} \tag{A.30}$$

similarly as (A.12). Using a similar argument, asymptotic normality in (A.27) follows by Theorem of Bolthausen (1982) under Assumption A-(ii).

Second, to show (A.28) and (A.29), we consider the case of scalar x_i for expositional simplicity. Given Theorem 3, it suffices to consider $\widehat{\gamma}$ in a neighborhood of γ_0 uniformly with distance at most $\overline{r}\phi_n$ for some large enough $\overline{r}$, where $\phi_n=\log n/a_n+b_n^2$ and $a_n=n^{1-2\epsilon}b_n$. Define $\widetilde{\gamma}\left(s\right)=\gamma_0\left(s\right)+\overline{r}\phi_n$ and $\widetilde{\Delta}_i\left(s_i\right)=\mathbf{1}_i\left(\widetilde{\gamma}\left(s_i\right)\right)-\mathbf{1}_i\left(\gamma_0\left(s_i\right)\right)$. We first observe that, by Assumptions A-(vi), (x), and (A.30), on the event that $\{\sup_{s\in\mathcal{S}}\left|\widehat{\gamma}\left(s\right)-\gamma_0\left(s\right)\right|\leq\overline{r}\phi_n\}$,

$$\begin{aligned}
& E\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i^2\delta_0\widehat{\Delta}_i\left(s_i\right)\right)^2\right] \\
\leq\ & E\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i^2\left|\delta_0\right|\left|\widehat{\Delta}_i\left(s_i\right)\right|\right)^2\right] \\
=\ & E\left[x_i^4\delta_0^2|\widehat{\Delta}_i\left(s_i\right)|\right]+2n\left(E\left[x_i^2\delta_0\widehat{\Delta}_i\left(s_i\right)\right]\right)^2 \\
& +\frac{2}{n}\sum_{i<j}^{n}cov\left[x_i^2\delta_0\widetilde{\Delta}_i\left(s_i\right),x_j^2\delta_0\widetilde{\Delta}_j\left(s_j\right)\right] \\
\leq\ & n^{-2\epsilon}c_0^2E\left[x_i^4|\widehat{\Delta}_i\left(s_i\right)|\right]+2n^{1-2\epsilon}c_0^2\left(E\left[x_i^2\widehat{\Delta}_i\left(s_i\right)\right]\right)^2 \\
& +2n^{-2\epsilon}c_0^2\sum_{m=1}^{\infty}m\alpha\left(m\right)^{\varphi/(2+\varphi)}E\left[\left|x_i^2\widetilde{\Delta}_i\left(s_i\right)\right|^{2+\varphi}\right]^{2/(2+\varphi)} \\
=\ & O\left(n^{-2\epsilon}a_n^{-1}\right)+O\left(\frac{1}{n^{1-2\epsilon}b_n^2}\right)+O\left(n^{-2\epsilon}\phi_n^{-2/(2+\varphi)}\right)=o\left(1\right),
\end{aligned} \tag{A.31}$$

provided $n^{1-2\epsilon}b_n^2\to\infty$, because $E[|x_i^2\widetilde{\Delta}_i\left(s_i\right)|^{2+\varphi}]=O(\phi_n)$ as (A.30). We can also verify that

$$E\left[\left(x_i u_i \widehat{\Delta}_i\left(s_i\right)\right)^2\right] \leq \int\left|\int_{\gamma_0(v)}^{\widehat{\gamma}(v)} V\left(q, v\right) f\left(q, v\right) dq\right| dv = O\left(\frac{1}{n^{1-2\epsilon} b_n}\right)$$

similarly as (A.30) and hence on the event that $\{\sup_{s\in\mathcal{S}}|\widehat{\gamma}\left(s\right)-\gamma_0\left(s\right)| \leq \overline{r}\phi_n\}$

$$\begin{aligned}
& E\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i \widehat{\Delta}_i\left(s_i\right)\right)^2\right] \\
\leq\ & E\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} |x_i u_i|\,|\widehat{\Delta}_i\left(s_i\right)|\right)^2\right] \\
=\ & E\left[x_i^2 u_i^2 |\widehat{\Delta}_i\left(s_i\right)|\right] + \frac{2}{n}\sum_{i<j}^{n} cov\left[|x_i u_i|\,\widetilde{\Delta}_i\left(s_i\right), |x_j u_j|\,\widetilde{\Delta}_j\left(s_j\right)\right] \\
=\ & O\left(\frac{1}{n^{1-2\epsilon} b_n}\right) + O\left(\phi_n^{-2/(2+\varphi)}\right) = o\left(1\right). \qquad \text{(A.32)}
\end{aligned}$$

since $E[x_i u_i \widehat{\Delta}_i\left(s_i\right)] = 0$ and $E[|x_i u_i \widetilde{\Delta}_i\left(s_i\right)|^{2+\varphi}] = O(\phi_n)$. Therefore, from (A.31), (A.32), and the fact that $P(\sup_{s\in\mathcal{S}}|\widehat{\gamma}\left(s\right)-\gamma_0\left(s\right)| \leq \overline{r}\phi_n) \to 1$ as $n\to\infty$ for some $\overline{r}<\infty$,

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{z}_i\left(\widehat{z}_i - z_i\right)'\theta_0 &= \begin{bmatrix} n^{-1/2}\sum_{i=1}^{n} x_i x_i' \delta_0 \widehat{\Delta}_i\left(s_i\right) \\ n^{-1/2}\sum_{i=1}^{n} x_i x_i' \delta_0 \widehat{\Delta}_i\left(s_i\right)\mathbf{1}_i\left(\widehat{\gamma}\left(s_i\right)\right)\end{bmatrix} \\
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_i - \widehat{z}_i\right)u_i &= \begin{bmatrix} n^{-1/2}\sum_{i=1}^{n} x_i u_i \widehat{\Delta}_i\left(s_i\right) \\ n^{-1/2}\sum_{i=1}^{n} x_i u_i \Delta_i\left(s_i\right)\mathbf{1}_i\left(\widehat{\gamma}\left(s_i\right)\right)\end{bmatrix}
\end{aligned}$$

are both $o_p(1)$, which completes the proof. ■

References

Andrews, D.W.K. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity, *Econometrica*, 62, 43-72.

Bhattacharya, P. and P. Brockwell (1976). The minimum of an additive process with applications to signal estimation and storage theory, *Probability Theory and Related Fields*, 37, 51-75.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley.

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields, *Annals of Probability*, 10, 1047-1050.

Carbon, M., C. Francq, and L.T. Tran (2007). Kernel regression estimation for random fields, *Journal of Statistical Planning and Inference*, 137, 778-798.

Conley, T. (1999). GMM estimation with cross sectional dependence, *Journal of Econometrics*, 92, 1-45.

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

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

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

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

Seo, M.H. and O. Linton (2007). A smoothed least squares estimator for threshold regression models, *Journal of Econometrics*, 141, 704-735.

Zhu, J. and S.N. Lahiri (2007). Bootstrapping the empirical distribution function of a spatial process, *Statistical Inference for Stochastic Processes*, 10, 107-145.