Nonparametric Spatial Threshold and Two-Dimensional Sample Splitting

Yoonseok Lee* Yulong Wang†
Syracuse University Syracuse University

March 2018

Abstract

[Very Preliminary; please do not distribute 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 that the observations can be spatially correlated and hence the model can be applied to study thresholds over a random field. We derive the limiting distributions of the semiparametric estimators and develop a likelihood ratio test on the nonparametric threshold. As an empirical illustration, 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 the square-footage elasticity to the house price.

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, this paper formulates the threshold by some *unknown interactions* 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 enough) 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 $\gamma_0(s) = \gamma$ or $\gamma_0(s) = \gamma s$ for some parameter γ 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 three-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 a given s in the support of s_i; 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 the square-footage elasticity to the house price.

The rest of the paper is organized as follows. Section 2 summarizes the model and our estimation procedure. Section 3 derived limiting properties of the estimators. Section 4 develops a likelihood ratio test of the threshold function and studies its small sample performance by Monte Carlo simulations. Section 5 applies the results to the housing price data to identify unknown economic border. 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 a 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:

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

where

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

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

[1] This model 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 $\mathbf{1}[\cdot]$.

$\gamma_0(\cdot)$ is estimated by

$$\widehat{\gamma}(s) = \arg\min_{\gamma\in\Gamma} Q_n(\gamma; s) \quad \text{for a given } s \in \mathcal{S},$$

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

$$Q_n(\gamma; s) = Q_n\left(\widehat{\beta}(\gamma; s), \widehat{\delta}(\gamma; s), \gamma; s\right). \tag{3}$$

Finally, the estimators of β_0 and δ_0 are obtained from

$$(\widehat{\beta}, \widehat{\delta}) = \arg\min_{\beta,\delta} \sum_{i=1}^{n} \left(y_i - x_i'\beta - w_i'\delta\right)^2, \tag{4}$$

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

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 $L_n \subset \mathbb{R}^2$ endowed with a metric $\rho(i,j) = \max_{1\leq\ell\leq 2}|i_\ell - j_\ell|$ and the corresponding norm $\max_{1\leq\ell\leq 2}|i_\ell|$, where i_ℓ denotes the ℓ-th component of i. We write $|L_n|$ for the number of elements in L_n and we simply let the cardinality of L_n as n (i.e., $|L_n| = n$); the summation in (3) hence can be rewritten as $\sum_{i\in L_n}$. Following Bolthausen (1982) and Jenish and Prucha (2009), we also define a mixing coefficient:

$$\alpha(m) = \sup\left\{\left|P(A_i \cap A_j) - P(A_i)P(A_j)\right| : A_i \in \mathcal{A}_i \text{ and } A_j \in \mathcal{A}_j \text{ with } \rho(i,j) \geq m\right\}, \tag{5}$$

where $\mathcal{A}_i$ 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

$$\begin{aligned} D(q,s) &= E\left[x_i x_i' \middle| (q_i, s_i) = (q,s)\right] \quad \text{and} \qquad (6)\\ V(q,s) &= E\left[x_i x_i' u_i^2 \middle| (q_i, s_i) = (q,s)\right]. \qquad (7)\end{aligned}$$

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 $L_n \subset \mathbb{R}^2$ is infinite countable; all the elements in L_n are located at distances at least $\rho_0 > 1$ from each other, i.e., for any $i, j \in L_n : \rho(i,j) \geq \rho_0$; and

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

(ii) $(x_i', q_i, s_i, u_i)'$ *is stationary and* α*-mixing with the mixing coefficient* $\alpha(m)$ *satisfying* $\sum_{m=1}^{\infty} m\alpha(m) < \infty$ *and* $\sum_{m=1}^{\infty} m^2\alpha(m)^{\varphi/(2+\varphi)} < \infty$ *for some* $\varphi > 0$.

(iii) $\delta_0 = c_0 n^{-\epsilon}$ *for some* $c_0 \neq 0$ *and* $\epsilon \in (0, 1/2)$.

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

(v) $\gamma : S \mapsto \Gamma$ *is twice continuously differentiable, where* Γ *is a compact subset of* $\mathbb{R}$.

(vi) Uniformly in (q, s)*, there exists* $R < \infty$ *such that* $E\left[||x_i||^{8+\tau} |(q_i, s_i) = (q, s)\right] < R$ *and* $E\left[||x_i u_i||^{8+\tau} |(q_i, s_i) = (q, s)\right] < R$ *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* $s \in \mathcal{S}$.

(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) $E[x_i x_i' \mathbf{1}[q_i \leq \gamma] | s_i = s]$ *is positive definite for any* $\gamma \in \Gamma$ *and 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 ρ_0 in Assumption A-(i) can be any strictly positive value, but we can impose $\rho_0 > 1$ without loss of generality. The mixing condition in Assumption A-(ii) is from Bolthausen (1982). 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.

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 uniformly consistent.

Theorem 1 *Under Assumption A,* $\sup_{s\in\mathcal{S}} |\widehat{\gamma}(s) - \gamma_0(s)| \to_p 0$ *as* $n \to \infty$.

The second result derives the limiting distribution of $\widehat{\gamma}(s)$. Similar to Hansen (2000), we let $W(\cdot)$ be a two-sided Brownian motion.

Theorem 2 *Under Assumption A and* $n^{1-2\epsilon}b_n^3 \to 0$*, for any fixed* $s \in \mathcal{S}$*,*

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

as $n \to \infty$*, where*

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

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

Note that the distribution of $\arg\max_{r\in\mathbb{R}}\left(W\left(r\right)-|r|/2\right)$ is known (e.g., Bhattacharya and Brockwell, 1976), which is also described in Hansen (2000, p.581). $\xi\left(s\right)$ term 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 2 also shows that the (pointwise) rate of convergence of $\widehat{\gamma}\left(s\right)$ is $n^{1-2\epsilon}b_n$, which depends 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 b_n^2-rate bias and the $(n^{1-2\epsilon}b_n)^{-1}$-rate precision from Theorem 2, the optimal bandwidth satisfies $b_n^* = cn^{-(1-2\epsilon)/5}$ for some constant $0 < c < \infty$.[2] However, it does not mean that we can always choose b_n as large as possible, which is also common 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 $O_p(b_n^2)$ bias term in the kernel estimator and hence the limiting distribution of $n^{1-2\epsilon}b_n\left(\widehat{\gamma}\left(s\right)-\gamma_0\left(s\right)\right)$ has mean zero.

The next result derives the limiting distribution of the parameter estimators $\widehat{\beta}$ and $\widehat{\delta}$, where they satisfy the conventional $n^{-1/2}$-consistency. We denote $z_i = [x_i', x_i'\mathbf{1}_i\left(\gamma_0\left(s_i\right)\right)]'$.

Theorem 3 *Let* $\widehat{\theta} = (\widehat{\beta}', \widehat{\delta})'$ *and* $\theta_0 = (\beta_0', \delta_0)'$*. Under the same condition in Theorem 2 and*

[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.

$n^{1-2\epsilon}b_n^2 \to \infty$,

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

as $n \to \infty$, where

$$\begin{aligned} M^* &= E\left[z_i z_i'\right] = \begin{pmatrix} E\left[x_i x_i'\right] & E\left[x_i x_i' \mathbf{1}_i\left(\gamma_0\left(s_i\right)\right)\right] \\ E\left[x_i x_i' \mathbf{1}_i\left(\gamma_0\left(s_i\right)\right)\right] & E\left[x_i x_i' \mathbf{1}_i\left(\gamma_0\left(s_i\right)\right)\right] \end{pmatrix}, \\ V^* &= Var\left[z_i u_i\right] = \begin{pmatrix} E\left[x_i x_i' u_i^2\right] & E\left[x_i x_i' u_i^2 \mathbf{1}_i\left(\gamma_0\left(s_i\right)\right)\right] \\ E\left[x_i x_i' u_i^2 \mathbf{1}_i\left(\gamma_0\left(s_i\right)\right)\right] & E\left[x_i x_i' u_i^2 \mathbf{1}_i\left(\gamma_0\left(s_i\right)\right)\right] \end{pmatrix}, \end{aligned}$$

which are positive definite by construction.

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 $n^{-1/2}$-consistency of $\widehat{\theta}$. in Theorem 3. 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)), and hence the replacement of $\widehat{\gamma}$ by γ_0 in (4) has an effect at most $o_p(n^{-1/2})$. The asymptotic variance can be estimated by the same analogue of M^* and V^* using $\widehat{u}_i = y_i - x_i'\widehat{\beta} - w_i'\widehat{\delta}$ and $\widehat{\gamma}$.

4 Likelihood Ratio Test

From Theorem 2, we can consider a pointwise likelihood ratio test statistic for

$$H_0 : \gamma_0\left(s\right) = \gamma_*\left(s\right) \quad \text{for some } s \in \mathcal{S}, \tag{8}$$

which is given as

$$LR_n(s) = \left(\sum_{i=1}^{n} K\left(\frac{s_i - s}{b_n}\right)\right) \times \frac{Q_n\left(\gamma_*\left(s\right), s\right) - Q_n\left(\widehat{\gamma}\left(s\right), s\right)}{Q_n\left(\widehat{\gamma}\left(s\right), s\right)}. \tag{9}$$

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

Theorem 4 *Under the same condition in Theorem 2, for any fixed $s \in \mathcal{S}$, the test statistic in (9) under the hull hypothesis (8) satisfies*

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

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.

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 $\widehat{\sigma}^2(s) = \sum_{i=1}^n \omega_{1i}(s) \widehat{u}_i^2$, $\widehat{D}(\widehat{\gamma}(s), s) = \sum_{i=1}^n \omega_{2i}(s) x_i x_i'$, and $\widehat{V}(\widehat{\gamma}(s), s) = \sum_{i=1}^n \omega_{2i}(s) x_i x_i' \widehat{u}_i^2$ are the standard Nadaraya-Watson estimators with $\widehat{u}_i = y_i - x_i'\widehat{\beta} - w_i'\widehat{\delta}$ from (4) and

$$\begin{aligned} \omega_{1i}(s) &= K\left(\frac{s_i - s}{b_n}\right) \Big/ \sum_{i=1}^n K\left(\frac{s_i - s}{b_n}\right), \\ \omega_{2i}(s) &= \mathbb{K}\left(\frac{q_i - \widehat{\gamma}(s)}{b_n'}, \frac{s_i - s}{b_n''}\right) \Big/ \sum_{i=1}^n \mathbb{K}\left(\frac{q_i - \widehat{\gamma}(s)}{b_n'}, \frac{s_i - s}{b_n''}\right), \end{aligned}$$

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).

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

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 $x_i \in \mathbb{R}^1$, $\beta_0 = 0$, and $\gamma_0(s) = \sin(s)/2$. For the dependence structure in $(x_i, q_i, s_i, u_i)'$, we consider the following two cases with $\lambda = 0.1$:

- DGP1: $(x_i, q_i, s_i, u_i)' \sim iid\mathcal{N}(0, I_4)$
- DGP2: $\begin{cases} (q_i, s_i)' \sim iid\mathcal{N}(0, I_2); \\ x_i | (q_i, s_i) \sim iid\mathcal{N}\left(0, [1 + \lambda\left(q_i^2 + s_i^2\right)]^{-1}\right); \\ \underline{u} | \{(x_i, q_i, s_i)\}_{i=1}^n \sim \mathcal{N}(0, \Omega), \end{cases}$

where the (i, j)th element of Ω is $\Omega_{ij} = [1 + \lambda((q_i - q_j)^2 + (s_i - s_j)^2)]^{-1}$ for $i, j = 1, 2, \cdots, n$, and $\underline{u} = (u_1, \cdots, u_n)'$. DGP1 is the case with i.i.d. observations, whereas DGP2 is the case with spatially correlated observations. For the bandwidth parameter, we simply select $b_n = b_n' = b_n'' = n^{-1/2}\sigma_s$, where σ_s is the standard deviation of s_i. Tables II and III summarize the rejection probabilities of $LR_n(s)$ at 5% nominal size over three different locations $s = 0.0$, 0.5, and 1.0 for these two DGP's. For each location, we consider nine cases with $n = 100, 200, 500$ and $\delta_0 = 1, 2, 3$. Note that each combination of (u, δ_0) determines ϵ for a fixed c_0 as $\epsilon = (\log c_0 - \log \delta_0)/\log n$. (See Hansen (2000) for a similar simulation design.) The result shows that the performance of the likelihood ratio test improves with n and the size of threshold δ_0.

Table II: Rej. Prob. with i.i.d. data (DGP1)

	$s = 0.0$			$s = 0.5$			$s = 1.0$		
$n \setminus \delta_0$	1	2	3	1	2	3	1	2	3
100	0.14	0.08	0.05	0.11	0.08	0.04	0.10	0.06	0.04
200	0.17	0.09	0.05	0.18	0.08	0.06	0.16	0.10	0.03
500	0.26	0.08	0.06	0.28	0.11	0.05	0.27	0.11	0.05

Table III: Rej. Prob. with spatially correlated data (DGP2)

	$s = 0.0$			$s = 0.5$			$s = 1.0$		
$n \setminus \delta_0$	1	2	3	1	2	3	1	2	3
100	0.08	0.06	0.04	0.06	0.05	0.05	0.06	0.04	0.03
200	0.11	0.05	0.04	0.12	0.06	0.05	0.11	0.07	0.04
500	0.22	0.12	0.08	0.27	0.13	0.09	0.27	0.14	0.10

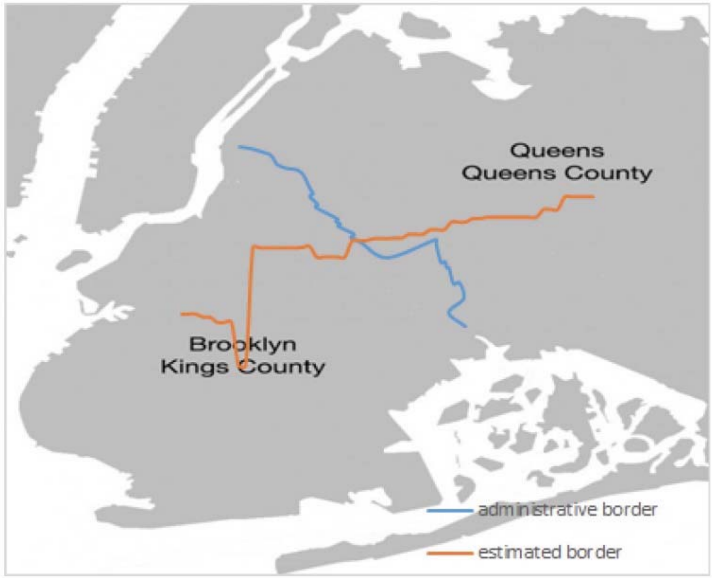

Figure 1: Threshold Function Estimate

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 variables:[3]

y_i	x_i	q_i	s_i
house price (\$)	constant	latitude	longitude
	log of Gross Square Footage (ft^2)		
	log of Land 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. The sample size is $n = 51,387$ ($27,233$ observations in Queens; $24,154$ observations in Brooklyn).

[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. "Land Square Footage" is the land area of the property listed in square feet. (Source: http://www1.nyc.gov/assets/finance/downloads/pdf/07pdf/glossary_rsf071607.pdf)

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 (orange line) is found to be substantively different from the administrative border between these two boroughs (blue line). One interesting note is that the big drop-down of the red line in the middle of Brooklyn is where the Brooklyn College is located. Table IV summarizes the coefficient estimates for the parametric components, $\widehat{\beta}$ and $\widehat{\delta}$.

Table IV: Estimation Result

	$\widehat{\beta}$	$\widehat{\delta}$
constant	8.837	−2.367
log of Gross Square Footage	0.200	−0.506
log of Land Square Footage	0.418	0.824
dummy for built before 1945	0.119	0.025

We can find that the housing price on the southern side of the threshold (or economic border) is lower than that on north. The (semi-) elasticity of the Gross Square Footage is higher on the northern side, whereas the (semi-) elasticity of the Land Square Footage is higher on the southern side. It is also found that houses on the southern side are older than the north.

A Appendix

Throughout the proof, we denote $K_i(s) = K((s_i - s)/b_n)$ and $\mathbf{1}_i(\gamma) = \mathbf{1}[q_i \leq \gamma]$ for any $\gamma : \mathcal{S} \mapsto \Gamma$.

A.1 Useful Lemmas

Lemma A.1 *For a 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. Throughout the proof, $C \in (0, \infty)$ stands for a generic constant term that may vary, which can depend on the location s. We first prove the pointwise convergence of $M_n(\gamma; s)$. By stationarity, Assumption A-(vii) and Taylor expansion, we have

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

where $D(q,s)$ is defined in (6). However, we have

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

Note that we use Assumption A-(vi), (vii), and Lemma 1 of Bolthausen (1982) to show that

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

for some finite $\varphi>0$, where $\alpha(m)$ is the mixing coefficient defined in (5) and the equality is by the change of variables ($t_i=(s_i-s)/b_n$) in the covariance operator. The equality is from the changes of variables and Taylor expansion like in $E\left[M_n(\gamma;s)\right]$ above. Hence, the pointwise convergence is established. For a 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). Then the uniform convergence follows from standard argument. Since $E\left[u_ix_i|q_i,s_i\right]=0$, the proof for $\sup_{\gamma\in\Gamma}|n^{-1/2}b_n^{-1/2}J_n(\gamma,s)|\xrightarrow{p}0$ is identical 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,

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

for any ζ. To this end, we consider a fine enough grid 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 $H_{ng}(s) = n^{-1}b_n^{-1}\sum_{i=1}^{n}|h_{ig}(s)|$. 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

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

In what follows, we denote $h_i(s) = x_i u_i K_i(s)\mathbf{1}\left[\zeta_g < q_i \leq \zeta_k\right]$ for any given $1 \leq g < k \leq G_n$ and for a 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.1),

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

for some $\varphi > 0$ that depends on s, where we let $\widetilde{h}_i = x_i u_i K(t_i)\mathbf{1}\left[\zeta_g < q_i \leq \zeta_k\right]$ 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 $0 < C' < \infty$ similarly as Hansen (2000). Using the same argument as the second component in (A.2), we can also show that $\Psi_{13}(s) = O(n^{-1}) + O(b_n^2)$. For $\Psi_{14}(s)$, by stationarity,

$$\begin{aligned}
\Psi_{14}(s) &\leq \frac{4!n}{n^2 b_n^2} \sum_{1<i<j<k}^{n} \left| E\left[h_1(s)h_i(s)h_j(s)h_k(s)\right]\right| \\
&\leq \frac{4!}{nb_n^2} \sum_{i=1}^{n} \sum_{j,k\leq i} \left| cov\left[h_1(s), h_{i+1}(s)h_{i+j+1}(s)h_{i+j+k+1}(s)\right]\right| \\
&\quad + \frac{4!}{nb_n^2} \sum_{j=1}^{n} \sum_{i,k\leq j} \left| cov\left[h_1(s)h_{i+1}(s), h_{i+j+1}(s)h_{i+j+k+1}(s)\right]\right| \qquad \text{(A.3)} \\
&\quad + \frac{4!}{nb_n^2} \sum_{k=1}^{n} \sum_{i,j\leq k} \left| cov\left[h_1(s), h_{i+1}(s)h_{i+j+1}(s), h_{i+j+k+1}(s)\right]\right|
\end{aligned}$$

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

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

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.3) satisfies

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

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

$$\frac{2\varphi}{2+\varphi} \geq \frac{1}{1-2\epsilon},$$

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.4) becomes $o(1)$. Using the same argument, we can also verify that the rest of terms in (A.3) 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

$$P\left(\max_{1 \leq g \leq G_n} \left|J_n\left(\zeta_g; s\right) - J_n\left(\zeta; s\right)\right| > \eta(s)\right) \leq \frac{C_1(s)\Delta^2}{\eta^4(s) b_n}, \tag{A.5}$$

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,

$$P\left(\max_{1 \leq g \leq G_n} \sqrt{nb_n}\left|H_{ng}(s) - E\left[H_{ng}(s)\right]\right| > \eta(s)\right) \leq \frac{C_2(s)\Delta}{\eta^2(s)}. \tag{A.6}$$

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

$$\sqrt{nb_n} E\left[H_{ng}(s)\right] = n^{1/2} b_n^{1/2} C_3(s)\Delta/G_n \leq 2C_3(s) n^{-1/2} b_n^{-1/2} \tag{A.7}$$

for some constant $0 < C_3(s) < \infty$ given s, where $\Delta/G_n \leq 2/nb_n$. So tightness is complete by combining (A.5), (A.6), and (A.7), and hence the weak convergence follows from Theorem

15.5 of Billingsley (1968). ■

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

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

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 $n \to \infty$ 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.2 First consider $r > 0$. By change of variables and Taylor expansion, Assumption A-(vii) and (viii) imply that

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

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

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

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

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

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

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

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

$$\begin{aligned}&Cov\left[c_0'x_iu_i\left(\mathbf{1}_i\left(\gamma_0+r/a_n\right)-\mathbf{1}_i\left(\gamma_0\right)\right)K_i\left(s\right),\right.\\ &\qquad\left.c_0'x_ju_j\left(\mathbf{1}_j\left(\gamma_0+r/a_n\right)-\mathbf{1}_j\left(\gamma_0\right)\right)K_j\left(s\right)\right]\leq b_n^2a_n^{-1}C\end{aligned}\tag{A.8}$$

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

$$\begin{aligned}Var[B_n^*\left(r,s\right)] &= \frac{a_n}{b_n}Var\left[c_0'x_iu_i\left|\mathbf{1}_i\left(\gamma_0+r/a_n\right)-\mathbf{1}_i\left(\gamma_0\right)\right|K_i\left(s\right)\right]+O\left(b_n\right)\\ &= rc_0'V\left(\gamma_0\left(s\right),s\right)c_0f\left(\gamma_0(s),s\right)\kappa_2+o\left(1\right),\end{aligned}$$

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

$$B_n^*\left(r,s\right)\Rightarrow W\left(r\right)\sqrt{c_0'V\left(\gamma_0\left(s\right),s\right)c_0f\left(\gamma_0\left(s\right),s\right)\kappa_2}$$

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 $r_1<r_2$, $Cov\left[B_n^*\left(r_1,s\right),B_n^*\left(r_2,s\right)\right]=Var\left[B_n^*\left(r_1,s\right)\right]+o\left(1\right)$ since $\left(\mathbf{1}_i\left(\gamma_0+r_2/a_n\right)-\mathbf{1}_i\left(\gamma_0+r_1/a_n\right)\right)\mathbf{1}_i\left(\gamma_0+r_1/a_n\right)=0$ and (A.8). The tightness follows

from a similar argument as in Lemma A.1 and the desired result follows by Theorem 15.5 in Billingsley (1968). ■

A.2 Proof of Main Theorems

Proof of Theorem 1 We first show pointwise convergence. 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 $\widetilde{x}_i(\gamma; s) = K_i(s)^{1/2} x_i \mathbf{1}_i(\gamma(s))$; 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

$$\widehat{\theta}(\gamma; s) = (\widetilde{Z}(\gamma; s)'\widetilde{Z}(\gamma; s))^{-1}\widetilde{Z}(\gamma; s)'\widetilde{y}(s), \tag{A.9}$$

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. Because

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

however, Lemma A.1 yields that

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

for given s. It follows that

$$\begin{aligned} &\frac{1}{n^{1-2\epsilon}b_n}\left(Q_n(\gamma; s) - \widetilde{u}(s)'\widetilde{u}(s)\right) \qquad \text{(A.10)} \\ &= O_p\left(\frac{1}{n^{1-2\epsilon}b_n}\right) + O_p\left(\sqrt{\frac{1}{n^{1-2\epsilon}b_n}}\right) + \frac{1}{nb_n}c_0'\widetilde{X}(\gamma_0(s_i); s)'\left(I - P_{\widetilde{Z}}(\gamma; s)\right)\widetilde{X}(\gamma_0(s_i); s)c_0 \\ &= \frac{1}{nb_n}c_0'\widetilde{X}(\gamma_0(s_i); s)'\left(I - P_{\widetilde{Z}}(\gamma; s)\right)\widetilde{X}(\gamma_0(s_i); s)c_0 + o_p(1) \end{aligned}$$

as $n^{1-2\epsilon}b_n \to 0$ with $n \to \infty$. 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 $\widetilde{X}(\gamma_0(s_i); s)'\widetilde{X}(\gamma; s) = \widetilde{X}(\gamma_0(s_i); s)'\widetilde{X}(\gamma_0(s_i); s)$. Hence, similarly as Lemma A.1, it can be verified that

$$
\begin{aligned}
E\left[M_n\left(\gamma_0(s_i);s\right)\right] &= \frac{1}{b_n}\iint E[x_i^2|q,v]\mathbf{1}[q\leq\gamma_0(v)]K\left(\frac{v-s}{b_n}\right)f\left(q,v\right)dqdv \qquad \text{(A.11)}\\
&= E\left[M_n\left(\gamma_0;s\right)\right]+\int\left(\int_{\gamma_0(s)}^{\gamma_0(s+b_nt)}D(q,s+b_nt)f(q,s+b_nt)dq\right)K(t)dt\\
&= E\left[M_n\left(\gamma_0;s\right)\right]+\int\left(\int_{\gamma_0(s)}^{\gamma_0(s+b_nt)}D(q,s)f(q,s)dq\right)\left(1+C_1b_n^2t^2\right)K(t)dt\\
&= E\left[M_n\left(\gamma_0;s\right)\right]+\int\left(C_{21}b_nt+C_{22}b_n^2t^2\right)\left(1+C_1b_n^2t^2\right)K(t)dt\\
&= E\left[M_n\left(\gamma_0;s\right)\right]+O\left(b_n^2\right),
\end{aligned}
$$

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 $\gamma \in \Gamma \cap [\gamma_0(s_i), \infty)$,

$$
\begin{aligned}
&\frac{1}{nb_n}c_0'\widetilde{X}(\gamma_0(s_i);s)'\left(I-P_{\widetilde{Z}}(\gamma;s)\right)\widetilde{X}(\gamma_0(s_i);s)c_0 \qquad \text{(A.12)}\\
&\to_p c_0'M(\gamma_0;s)c_0-c_0'M(\gamma_0;s)'M(\gamma;s)^{-1}M(\gamma_0;s)c_0<\infty,
\end{aligned}
$$

from Lemma A.1 and Assumptions A-(viii) and (ix), as $O\left(b_n^2\right) = o(1)$. Note that $M(\gamma; s) = E[x_ix_i'\mathbf{1}_i(\gamma)|s_i = s]f_s(s)$ 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).

Next, to show the uniform consistency over $s \in \mathcal{S}$, it suffices to show the uniform tightness of $\widehat{\gamma}(s)$ on s. For any ε, consider $s_1 < s_2$. By Assumption A-(v), we can pick Δ such that $|\gamma_0(s_2) - \gamma_0(s_1)| \leq \Delta\varepsilon$ for any $|s_2 - s_1| \leq \Delta$. Then, take a fine enough grid $s_1 = t_0 < t_1 < t_2 < \cdots < t_{G-1} < t_G = s_2$ such that $\max_{1\leq g\leq G}|t_g - t_{g-1}| \leq \Delta\varepsilon$. By the pointwise consistency, $\max_{1\leq g\leq G}|\widehat{\gamma}(t_g) - \gamma_0(t_g)| \leq \Delta\varepsilon$ when n is sufficiently large. For any $g \leq G-1$ and any $s \in [t_g, t_{g+1}]$, therefore, Assumption A-(v) and the pointwise convergence imply that with probability greater than $1-\varepsilon$, both $\widehat{\gamma}(s)$ and $\gamma_0(s)$ are in the interval $[\gamma_0(t_g)-\varepsilon, \gamma_0(t_{g+1})+\varepsilon]$, which has length bounded by 3ε. Since s is arbitrary, we have with probability greater than $1-\varepsilon$,

$$
\begin{aligned}
&\sup_{0\leq t\leq\Delta}\left|\widehat{\gamma}(s_1)-\widehat{\gamma}(s_1+t)\right|\\
\leq\ &\max_{1\leq g\leq G}\left|\widehat{\gamma}(t_g)-\gamma_0(t_g)\right|+\max_{1\leq g\leq G}\sup_{s\in[t_g,t_{g+1}]}\left|\widehat{\gamma}(s)-\gamma_0(s)\right|+\sup_{0\leq t\leq\Delta}\left|\gamma_0(s_1)-\gamma_0(s_1+t)\right|\\
\leq\ &4\Delta\varepsilon
\end{aligned}
$$

which completes the proof by Theorem 15.5 of Billingsley (1968). ■

In order to prove Theorem 2, we first show the following lemma.

Lemma A.3 *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 a given* $s \in \mathcal{S}$. *Then, under Assumption A,*

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

Proof of Lemma A.3 For the first result, from (A.9), we have

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

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

For the second result, we let $\widehat{z}_i(s) = [x_i', x_i'\mathbf{1}_i\left(\widehat{\gamma}\left(s\right)\right)]'$, $z_i(s) = [x_i', x_i'\mathbf{1}_i\left(\gamma_0\left(s\right)\right)]'$, and $z_i = [x_i', x_i'\mathbf{1}_i\left(\gamma_0\left(s_i\right)\right)]'$. Then, $y_i = z_i'\theta_0 + u_i$. Using a using a similar expression as above, we have

$$\begin{aligned}
& \sqrt{nb_n}\left(\widehat{\theta}\left(\widehat{\gamma}\left(s\right)\right) - \widehat{\theta}\left(\gamma_0\left(s\right)\right)\right) \qquad \text{(A.13)} \\
= & \sqrt{nb_n}\left(\widehat{\theta}\left(\widehat{\gamma}\left(s\right)\right) - \theta_0\right) - \sqrt{nb_n}\left(\widehat{\theta}\left(\gamma_0\left(s\right)\right) - \theta_0\right) \\
= & \left(C + o_p\left(\frac{1}{nb_n}\right)\right)^{-1}\frac{1}{\sqrt{nb_n}}\sum_{i=1}^n\left\{\widehat{z}_i(s)\left(y_i - \widehat{z}_i(s)'\theta_0\right) - z_i(s)\left(y_i - z_i(s)'\theta_0\right)\right\}K_i\left(s\right) \\
= & \left(C + o_p\left(\frac{1}{nb_n}\right)\right)^{-1}\left\{\frac{1}{\sqrt{nb_n}}\sum_{i=1}^n\left(\widehat{z}_i(s) - z_i\right)u_i K_i\left(s\right)\right. \\
& \left. -\frac{1}{\sqrt{nb_n}}\sum_{i=1}^n\widehat{z}_i(s)\left(\widehat{z}_i(s) - z_i\right)'\theta_0 K_i\left(s\right) - \frac{1}{\sqrt{nb_n}}\sum_{i=1}^n z_i(s)\left(z_i - z_i(s)\right)'\theta_0 K_i\left(s\right)\right\}
\end{aligned}$$

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

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

by Lemma A.2 and Assumption A-(x). Similarly, we also have

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

since $\left(\widehat{z}_i(s) - z_i \right)' \theta_0 = \delta_0' x_i \left(\mathbf{1}_i \left(\widehat{\gamma}(s) \right) - \mathbf{1}_i \left(\gamma_0(s_i) \right) \right)$. The last component in (A.13) 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. ■

Proof of Theorem 2 First, define $a_n = n^{1-2\epsilon} b_n$, where ϵ is given in Assumption A-(iii). For a given $s \in \mathcal{S}$, we consider $\widetilde{\gamma}(s)$ such that $|\widetilde{\gamma}(s) - \gamma_0(s)| \in [\overline{r}(s)/a_n, C]$ for some $0 < \overline{r}(s), C < \infty$. Then, given Lemma A.2, it can be verified that $P(Q_n^*(\widetilde{\gamma}; s) - Q_n^*(\gamma_0; s) > 0) \to 1$ as $n \to \infty$ using the standard results in kernel regression (e.g., Li and Racine, 2007, Ch.2) and following Hansen (2000), where $Q_n^*(\gamma_0; s) = Q_n(\widehat{\beta}(\widehat{\gamma}(s)), \widehat{\delta}(\widehat{\gamma}(s)), \gamma_0; s)$ and $Q_n^*(\widetilde{\gamma}; s) = Q_n(\widehat{\beta}(\widehat{\gamma}(s)), \widehat{\delta}(\widehat{\gamma}(s)), \widetilde{\gamma}; s)$. It follows that with probability approaches to one, since $Q_n^*(\widehat{\gamma}; s) - Q_n^*(\gamma_0; s) \leq 0$ for given s, $|\widehat{\gamma}(s) - \gamma_0(s)| \leq \overline{r}(s)/a_n$ and hence we can define a random variable $r^*(s)$ such that

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

We now let $\Delta_i(r; s) = \mathbf{1}_i \left(\gamma_0(s) + (r/a_n) \right) - \mathbf{1}_i \left(\gamma_0(s) \right)$. We then have

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

For $A_n(r;s)$, Lemmas A.2 and A.3 yield

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

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

$$\begin{aligned}
& B_n(r;s) \\
=& \sum_{i=1}^{n}\Bigg(u_i + \delta_0' x_i\left\{\mathbf{1}_i\left(\gamma_0\left(s_i\right)\right) - \mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right\} - \left(\widehat{\beta}\left(\widehat{\gamma}\left(s\right)\right) - \beta_0\right)' x_i \\
&\qquad - \left(\widehat{\delta}\left(\widehat{\gamma}\left(s\right)\right) - \delta_0\right)' x_i \mathbf{1}_i\left(\gamma_0\left(s\right)\right)\Bigg)\widehat{\delta}\left(\gamma_0\left(s\right)\right)' x_i \Delta_i(r;s)K_i\left(s\right) \\
=& \sum_{i=1}^{n}\left(u_i + \delta_0' x_i\left\{\mathbf{1}_i\left(\gamma_0\left(s_i\right)\right) - \mathbf{1}_i\left(\gamma_0\left(s\right)\right)\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\left(\gamma_0(s)\right)\right) \\
&\quad \times\left(\delta_0 + n^{-1/2}b_n^{-1/2}C_\delta\right)' x_i \Delta_i(r;s)K_i\left(s\right) + o_p(1) \\
=& B_n^*\left(r,s\right) + \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\left(s\right) \\
& + \sum_{i=1}^{n}\delta_0' x_i x_i' \delta_0\left(\Delta_i(r;s)\left\{\mathbf{1}_i\left(\gamma_0\left(s_i\right)\right) - \mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right\}\right)K_i\left(s\right) \qquad \text{(A.14)} \\
& + \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\left(\gamma_0\left(s_i\right)\right) - \mathbf{1}_i\left(\gamma_0\left(s\right)\right)\right\}\right)K_i\left(s\right) \\
& + \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\left(s\right) \\
& + \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\left(s\right) \\
& + \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\left(\gamma_0(s)\right)\right\}K_i\left(s\right) \\
& + \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\left(\gamma_0(s)\right)\right\}K_i\left(s\right) + o_p(1)
\end{aligned}$$

$$
\begin{aligned}
&= B_n^*(r,s) + O_p\left(\left(n^{1-2\epsilon}b_n\right)^{-1/2}\right) + O_p(n^{1-2\epsilon}b_n^3) + O_p\left(\left(n^{1-2\epsilon}b_n\right)^{-1/2}\right) + o_p(1) \\
&= B_n^*(r,s) + o_p(1)
\end{aligned}
$$

from Lemma A.2, since $n^{1-2\epsilon}b_n \to \infty$ and $n^{1-2\epsilon}b_n^3 \to 0$ under Assumption A-(x). Note that the third term on (A.14), denoting $B_{n3}(r;s)$, is $O_p\left(n^{1-2\epsilon}b_n^3\right)$, which is also negligible as $n^{1-2\epsilon}b_n^3 \to 0$. To see this, similarly as $E\left[M_n(\gamma;s)\right]$ in the proof of Lemma A.1, we have

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

However, since[4]

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

we have

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

which is because

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

[4]Note that

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

for some $C_1, C_{21}, C_{22} < \infty$, similarly as (A.11). 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.2 using the same argument of the proof of Theorem 1 of Hansen (2000). ■

Proof of Theorem 3 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 $\Delta_i(s_i) = \mathbf{1}_i(\widehat{\gamma}(s_i)) - \mathbf{1}_i(\gamma_0(s_i))$. Then,

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

and it suffices to establish

$$\frac{1}{n}\sum_{i=1}^n \widehat{z}_i\widehat{z}_i' \to_p M^* \tag{A.15}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n z_iu_i \to_d \mathcal{N}(0, V^*) \tag{A.16}$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{z}_i(\widehat{z}_i - z_i)'\theta_0 = o_p(1) \tag{A.17}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n (z_i - \widehat{z}_i)u_i = o_p(1). \tag{A.18}$$

First, by Assumptions A-(vi), (x), and Theorem 2, (A.15) can be readily verified since $n^{-1}\sum_{i=1}^n x_ix_i'\mathbf{1}_i(\widehat{\gamma}(s_i)) = n^{-1}\sum_{i=1}^n x_ix_i'\mathbf{1}_i(\gamma_0(s_i)) + n^{-1}\sum_{i=1}^n x_ix_i'\Delta_i(s_i)$ and

$$\begin{aligned}E\left[x_ix_i'\Delta_i(s_i)\right] &\leq \int\left|\int_{\gamma_0(v)}^{\widehat{\gamma}(v)} D(q,v)f(q,v)\,dq\right|dv \\ &= \int\left\{|D(\gamma_0(v),v)f(\gamma_0(v),v)|\,O_p\left(\frac{1}{n^{1-2\epsilon}b_n}\right)\right\}dv \\ &= O_p\left(\frac{1}{n^{1-2\epsilon}b_n}\right) \to 0\end{aligned} \tag{A.19}$$

similarly as (A.11). Using a similar argument, asymptotic normality in (A.16) follows by Theorem of Bolthausen (1982) under Assumption A-(ii).

Second, to show (A.17) and (A.18), we consider the case of scalar x_i for expositional simplicity. We first observe that, by Assumptions A-(vi), (x), and (A.19),

$$
\begin{aligned}
E\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i^2\delta_0\Delta_i(s_i)\right)^2\right] &= E\left[x_i^4\delta_0^2\left|\Delta_i(s_i)\right|\right]+2n\left(E\left[x_i^2\delta_0\Delta_i(s_i)\right]\right)^2 \\
&\quad +\frac{2}{n}\sum_{i<j}^{n} cov\left[x_i^2\delta_0\Delta_i(s_i), x_j^2\delta_0\Delta_j(s_i)\right] \\
&\leq n^{-2\epsilon}c_0^2E\left[x_i^4\left|\Delta_i(s_i)\right|\right]+2n^{1-2\epsilon}c_0^2\left(E\left[x_i^2\Delta_i(s_i)\right]\right)^2 \\
&\quad +2n^{-2\epsilon}c_0^2\sum_{m=1}^{\infty}m\alpha(m)^{\varphi/(2+\varphi)}E\left[\left|x_i^2\Delta_i(s_i)\right|^{2+\varphi}\right]^{2/(2+\varphi)} \\
&= O\left(n^{-2\epsilon}\right)+O\left(\frac{1}{n^{1-2\epsilon}b_n^2}\right)+O\left(n^{-2\epsilon}\right)\rightarrow 0,
\end{aligned} \tag{A.20}
$$

provided $n^{1-2\epsilon}b_n^2\rightarrow\infty$. Similarly as (A.19), we can also verify that

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

and hence

$$
E\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_iu_i\Delta_i(s_i)\right)^2\right]=O_p\left(\frac{1}{n^{1-2\epsilon}b_n}\right)\rightarrow 0 \tag{A.21}
$$

since $E[x_iu_i\Delta_i(s_i)]=0$ from Assumption A-(iv). From (A.20) and (A.21), therefore,

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

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

Proof of Theorem 4 From (A.10) and (A.12), we have

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

where $f_s(s)$ is the marginal density of s_i. In addition, from Theorem 2 and the proof of Lemma A.3, we have

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

since $\widehat{\theta}\left(\widehat{\gamma}(s)\right)-\widehat{\theta}\left(\gamma_0(s)\right)=o_p((nb_n)^{-1/2})$. 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. ■

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.

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.