Online Supplement for "Threshold Regression with Endogeneity"*

Ping Yu†
University of Hong Kong

Peter C. B. Phillips‡
Yale University, University of Auckland
University of Southampton & Singapore Management University

May 2017

1. Difficulties in Applying the DKE

When there are no other covariates besides q, the DKE is a popular procedure for estimating γ. Porter and Yu (2015) provide some discussion and references to the related literature. In this simple case, we have the model $y = g(q) + (1,q)\,\delta 1\,(q \le \gamma) + e$ with $\mathbb{E}[e|q] = 0$. The DKE is defined as the extremum estimator

$$\widehat{\gamma}_{DKE} = \arg\max_{\gamma} \widehat{\Delta}^2(\gamma), \tag{1}$$

where $\widehat{\Delta}(\gamma) = \widehat{\mathbb{E}}[y|q=\gamma-] - \widehat{\mathbb{E}}[y|q=\gamma+]$ with $\widehat{\mathbb{E}}[y|q=\gamma-] = n^{-1}\sum_{j=1}^{n} w_j(\gamma) y_j d_j(\gamma)$ and $\widehat{\mathbb{E}}[y|q=\gamma+] = n^{-1}\sum_{j=1}^{n} w_j(\gamma) y_j (1-d_j(\gamma))$ being estimators of $\mathbb{E}[y|q=\gamma-]$ and $\mathbb{E}[y|q=\gamma+]$, and $d_j(\gamma) = 1\,(q_j \le \gamma)$. In the definition of $\widehat{\mathbb{E}}[y|q=\gamma\pm]$, the weight function $w_j(\gamma) = k_h(q_j-\gamma) \Big/ \sum_{l=1}^{n} k_h(q_l-\gamma)$, where $k_h(\cdot) = k(\cdot/h)/h$ is a rescaled kernel density, and h is the bandwidth. Due to the weighted average nature of kernel smoothers, $\widehat{\Delta}(\gamma)$ would be near zero if there were no jump at γ. Otherwise, the difference would be near the magnitude of the jump $\Delta_0 = (1,\gamma_0)\,\delta_0$ which is assumed to be nonzero. This difference ensures that the estimator $\widehat{\gamma}_{DKE}$ is consistent. Porter and Yu (2015) have recently shown that $\widehat{\gamma}_{DKE}$ converges at rate n and the asymptotic distribution is related to a compound Poisson process. This limit theory is explained by interpreting γ as a 'middle' boundary point of q (see Yu, 2012). For boundary point estimation, it is well-known that only data in an $O\left(n^{-1}\right)$ neighborhood is informative, so the h neighborhood in the construction of the DKE is typically large enough to ensure the n-consistency of $\widehat{\gamma}_{DKE}$. Given $\widehat{\gamma}_{DKE}$, the literature has also considered the estimation of the jump magnitude Δ_0. But no estimator of δ_0 is presently available.

When there are additional covariates, Delgado and Hidalgo (2000) suggested that the DKE continue to be used to estimate γ.[1] In this case, the procedure can be employed by fixing some point (say x_o) in the support of x and redefining $\widehat{\Delta}(\gamma)$ as $\widehat{\mathbb{E}}[y|x_o, q=\gamma-] - \widehat{\mathbb{E}}[y|x_o, q=\gamma+]$, where $\widehat{\mathbb{E}}[y|x_o, q=\gamma\pm]$ is an estimator

*Phillips acknowledges support from the NSF under Grant No. SES 1258258.

†School of Economics and Finance, The University of Hong Kong, Pokfulam Road, Hong Kong; corresponding author email: pingyu@hku.hk.

‡Cowles Foundation for Research in Economics, Yale University, POBox 208281, New Haven, CT, USA; email: peter.phillips@yale.edu

[1] One might consider neglecting the data of x, and using only the data of q and y to estimate γ. This will generate the DKE of Porter and Yu (2015). Now, the jump size $E[\mathbf{x}'\delta|q=\gamma_0]$ is an average of the jumps at all x values, so may be zero or small, which results in identification failure or weak identification. Even if $E[\mathbf{x}'\delta|q=\gamma_0]$ is large, this DKE might be less efficient than the IDKE because the jump information at γ_0 is not fully explored; see footnote 6 for further analysis.

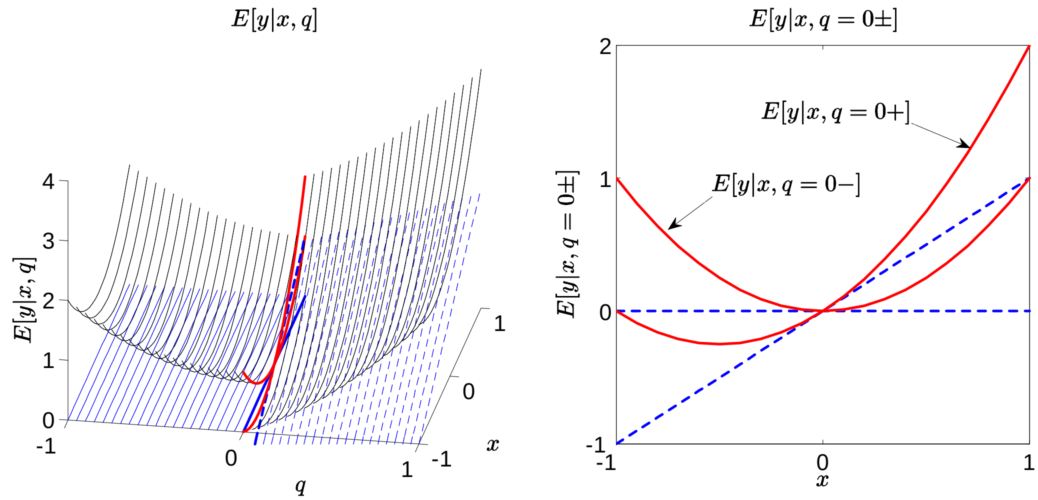

Figure 1: $E[y|x,q]$ and $E[y|x,q=\gamma_0\pm]$ With Endogeneity: The Blue Lines Represent the Case Without Endogeneity

of the conditional mean of y given $x=x_o$ and $q=\gamma\pm$. The objective function converges to zero when $\gamma\neq\gamma_0$, and to $\Delta_o^2=(\mathbb{E}[y|x_o,q=\gamma_0-]-\mathbb{E}[y|x_o,q=\gamma_0+])^2$ when $\gamma=\gamma_0$, so $\widehat{\gamma}_{DKE}$ is consistent if $\Delta_o\neq 0$.

There are several difficulties in applying the DKE in this way. First, the selection of x_o raises difficulties, as shown in the following example. Suppose $y=(x+q)\,1\,(q>\gamma)+\varepsilon$, where $\gamma_0=0$, the supports of x and q are both $[-1,1]$, and endogeneity takes the form $\mathbb{E}\,[\varepsilon|x,q]=x^2+q^2$. Figure 1 shows $\mathbb{E}[y|x,q]$ and $\mathbb{E}[y|x,q=\gamma_0\pm]$. To identify γ successfully we need to select x_o so that Δ_o^2 is large, which means that x_o should be on the boundary of x's support. On the other hand, we also need $f_{x|q}(x_o|\gamma_0)$ to be large so that there is sufficient data to identify γ. When the density of $f_{x|q}(x|\gamma_0)$ takes on a bell shape, as in a typical case, x_o should ideally be in the middle of x's support. Hence, the selection of x_o poses a dilemma and a potential tradeoff that is presently unresolved from both theory and practical perspectives. Second, consistency of $\widehat{\gamma}_{DKE}$ requires that $\Delta_o\neq 0$, but Δ_o can be 0 as shown in the example of Figure 1. Delgado and Hidalgo (2000) apply the DKE to estimate γ, assuming that $\left(\delta_{x0}',\delta_{q0}\right)'=\mathbf{0}$ and $\delta_{\alpha 0}\neq 0$ so that $\Delta_o=\delta_{\alpha 0}\neq 0$ does not depend on the choice of x_o. Moreover, their kernel function uses data in the neighborhood of $q=\gamma_0$ inefficiently, so that the convergence rate of $\widehat{\gamma}_{DKE}$ is quite slow, as discussed further in Section 2.2. Furthermore, given $\widehat{\gamma}_{DKE}$, the induced estimator of $\delta_{\alpha 0}$ uses only data in the neighborhood of $(x_o',\widehat{\gamma}_{DKE})'$, so the convergence rate of $\widehat{\delta}_{\alpha,DKE}$ is also very slow, especially when the dimension of x is large.

2. Difficulties in Applying Two Alternative Estimators

It is known (Appendix A of Porter and Yu, 2015) that the DKE is asymptotically equivalent to the LSE and the PLE when q is a single covariate. In what follows, we define the LSE and the PLE when other covariates are present and discuss the difficulties that arise in deriving their asymptotic distributions.

Define the nonparametric LSE of γ in the general case as follows,

$$\widehat{\gamma}_{LSE}^{N}=\arg\min_{\gamma}\frac{1}{n}\sum_{i=1}^{n}\left[y_i\widehat{f}_i-\widehat{mf}_{-}^{\gamma}(x_i,q_i)1(q_i\leq\gamma)-\widehat{mf}_{+}^{\gamma}(x_i,q_i)1(q_i>\gamma)\right]^2,$$

where $\widehat{f}_i = \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} K_{h,ij}$ with $K_{h,ij} = K_{h,ij}^x \cdot k_h(q_j - q_i)$ is the kernel estimator of $f_i \equiv f(x_i, q_i)$,

$$\begin{aligned}
\widehat{mf}_-^{\gamma}(x_i, q_i) &= \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} y_j K_{h,ij}^x k_h^{\gamma-}(q_j - q_i, q_i), \\
\widehat{mf}_+^{\gamma}(x_i, q_i) &= \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} y_j K_{h,ij}^x k_h^{\gamma+}(q_j - q_i, q_i),
\end{aligned}$$

with

$$\begin{aligned}
k_h^{\gamma-}(u,t) &= \begin{cases} \frac{1}{h} k\left(\frac{u}{h}\right), & \text{if } t \leq \gamma - h, \\ \frac{1}{h} k_-\left(\frac{u}{h}, \frac{\gamma - t}{h}\right), & \text{if } \gamma - h \leq t \leq \gamma, \end{cases} \\
k_h^{\gamma+}(u,t) &= \begin{cases} \frac{1}{h} k\left(\frac{u}{h}\right), & \text{if } t \geq \gamma + h, \\ \frac{1}{h} k_+\left(\frac{u}{h}, \frac{t - \gamma}{h}\right), & \text{if } \gamma \leq t \leq \gamma + h. \end{cases}
\end{aligned}$$

In the construction of $\widehat{\gamma}_{LSE}^N$, we eliminate the random denominator as in $\widehat{\gamma}$. We next define the PLE as

$$\widehat{\gamma}_{PLE} = \arg\min_{\gamma} \frac{1}{n} \sum_{i=1}^{n} \left[y_i \widehat{f}_i - \mathbf{x}_i' \delta 1(q_i \leq \gamma) \widehat{f}_i - \widehat{g}_f(x_i, q_i; \delta, \gamma) \right]^2, \tag{2}$$

where

$$\widehat{g}_f(x_i, q_i; \delta, \gamma) = \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} \left(y_j - \mathbf{x}_j' \delta 1(q_j \leq \gamma) \right) K_{h,ij}.$$

This density-weighted objective function of the PLE was suggested in Li (1996) without considering the threshold effects. In both the LSE and the PLE, γ is estimated by finding the best fit between y_i and an estimator of $\mathbb{E}[y_i|x_i, q_i]$; the difference lies in that different estimators of $\mathbb{E}[y_i|x_i, q_i]$ are used.

The objective function of the IDKE is superior to that of the LSE and the PLE in two respects. First, according to Yu (2008, 2012), only the information around the threshold point is informative for γ_0, so $\widehat{\Delta}_i(\gamma)$ in the objective function of the IDKE is constructed using only data in the neighborhood of γ. In contrast, the objective functions of the LSE and the PLE use information in other areas, and the resulting biases need to be handled carefully. The objective function of the IDKE therefore takes advantage of its local construction, whereas the global objective functions of the LSE and the PLE are influenced by the effects of information throughout the distribution. Second, since $\widehat{\Delta}_i(\gamma)$ in the objective function of the IDKE is linear in $k_+\left(\frac{q_j - \gamma}{h}\right)$ and $k_-\left(\frac{q_j - \gamma}{h}\right)$, it is easy to localize in the neighborhood of γ, which is key to deriving the convergence rate and the asymptotic distribution of $\widehat{\gamma}$. However, the objective functions of the LSE and PLE are complicated nonlinear functions of γ, which makes localization extremely hard. In addition, the objective function of the IDKE does not rely on the assumption that $\delta(x, q) = \mathbf{x}'\delta$, whereas that of the PLE does.

3. Proofs for the Theorems and Corollaries

In the following proofs, some steps are omitted for brevity whenever they are available in the literature and references are provided. This simplification makes the proofs cleaner and more readable. Derivations that differ from the existing literature are given in full detail. Propositions that are used in these derivations are given in the following Section 4 and additional lemmas that are needed are given in the following Section 5.

First, we introduce the new notations K_h^x and K^x through $K_{h,ij}^x =: K_h^x(x_j - x_i, x_i) =: \frac{1}{h^{d-1}} K^x\left(\frac{x_j - x_i}{h}, x_i\right)$, where $K_{h,ij}^x$ was introduced in the definition of the IDKE of γ.

Proof of Theorem 1. Proposition 1 proves the consistency of $\widehat{\gamma}$, and Proposition 2 proves $\widehat{\gamma} - \gamma_0 = O_p(n^{-1})$, so we can apply the argmax continuous mapping theorem (see, e.g., Theorem 3.2.2 of Van der Vaart and Wellner (1996)) to establish the asymptotic distribution of $n(\widehat{\gamma} - \gamma_0)$. From Proposition 3, for v in any compact set of $\mathbb{R}$,

$$nh\left(\widehat{Q}_n\left(\gamma_0 + \frac{v}{n}\right) - \widehat{Q}_n(\gamma_0)\right)\Big/ 2k_+(0) = -\sum_{i=1}^n \overline{z}_{1i} 1\left(\gamma_0 - \frac{v}{n} < q_i \le \gamma_0\right) - \sum_{i=1}^n \overline{z}_{2i} 1\left(\gamma_0 < q_i \le \gamma_0 + \frac{v}{n}\right) + o_p(1),$$

where $\overline{z}_{1i}$ and $\overline{z}_{2i}$ are defined in the main text. Now, we can obtain the asymptotic distribution of $n(\widehat{\gamma} - \gamma_0)$ by applying the same argument as in the proofs of Theorem 1 and 2 in Yu (2012). The only difference lies in the definitions of $\overline{z}_{1i}$ and $\overline{z}_{2i}$. ■

Proof of Corollary 1. The proofs of the consistency of $\widetilde{\gamma}$ and $nh^{d-1}(\widetilde{\gamma} - \gamma_0) = O_p(1)$ are similar to Theorem 1, so are omitted here. We concentrate on deriving the weak limit of the localized process $nh^d\left(\widehat{\Delta}_o^2(\gamma) - \widehat{\Delta}_o^2(\gamma_0)\right)$ for γ in an $\left(nh^{d-1}\right)^{-1}$ neighborhood of γ_0 .

Let $a_n = nh^{d-1}(= o(h))$, then

$$nh^d\left(\widehat{\Delta}_o^2\left(\gamma_0 + \frac{v}{a_n}\right) - \widehat{\Delta}_o^2(\gamma_0)\right) = \left(\widehat{\Delta}_o\left(\gamma_0 + \frac{v}{a_n}\right) + \widehat{\Delta}_o(\gamma_0)\right) nh^d\left(\widehat{\Delta}_o\left(\gamma_0 + \frac{v}{a_n}\right) - \widehat{\Delta}_o(\gamma_0)\right).$$

It is easy to show that $\widehat{\Delta}_o\left(\gamma_0 + \frac{v}{a_n}\right) - \mathbb{E}\left[\widehat{\Delta}_o\left(\gamma_0 + \frac{v}{a_n}\right)\right] \xrightarrow{p} 0$ for v in any compact set. Without loss of generality, let $\gamma > \gamma_0$ or $v > 0$. Then

$$\begin{aligned}
\mathbb{E}\left[\widehat{\Delta}_o(\gamma)\right] &= \int_{-1}^0 \int K^x(u_x, x_o) k_-(u_q) g(x_o + u_x h, \gamma + u_q h) f(x_o + u_x h, \gamma + u_q h) du_x du_q \\
&+ \int_{-1}^{\frac{\gamma_0 - \gamma}{h}} \int K^x(u_x, x_o) k_-(u_q) \left(1, (x_o + u_x h)', \gamma + u_q h\right) \delta_0 f(x_o + u_x h, \gamma + u_q h) du_x du_q \\
&- \int_0^1 \int K^x(u_x, x_o) k_+(u_q) g(x_o + u_x h, \gamma + u_q h) f(x_o + u_x h, \gamma + u_q h) du_x du_q \\
&= (1, x_o', \gamma_0)\, \delta_0 f(x_o, \gamma_0) + O(h).
\end{aligned}$$

Now, we need only consider the behavior of $nh^d\left(\widehat{\Delta}_o\left(\gamma_0 + \frac{v}{a_n}\right) - \widehat{\Delta}_o(\gamma_0)\right)$. Proposition 4 shows that

$$nh^d\left(\widehat{\Delta}_o\left(\gamma_0 + \frac{v}{a_n}\right) - \widehat{\Delta}_o(\gamma_0)\right) \Rightarrow D_o(v),$$

where $\Rightarrow$ signifies the weak convergence on a compact set of v,

$$D_o(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} z_{1i}, \text{ if } v \le 0, \\ \sum_{i=1}^{N_2(v)} z_{2i}, \text{ if } v > 0, \end{cases}$$

is a cadlag process with $D_o(0) = 0$,

$$z_{1i} = \left(-2e_i^- - (1, x_o', \gamma_0)\,\delta_0\right) K(U_i^-) k_-(0), \; z_{2i} = \left(2e_i^+ - (1, x_o', \gamma_0)\,\delta_0\right) K(U_i^+) k_+(0),$$

and the distributions of $e_i^-, e_i^+, U_i^-, U_i^+$ are defined in the corollary. So

$$nh^d\left(\widehat{\Delta}_o^2\left(\gamma_0+\frac{v}{a_n}\right)-\widehat{\Delta}_o^2(\gamma_0)\right)\Rightarrow\overline{D}(v),$$

where $\overline{D}(v)$ takes a similar form to $D_o(v)$, but now

$$z_{1i}=2\left(-2\left(1,x_o',\gamma_0\right)\delta_0e_i^- -\delta_0'\left(1,x_o',\gamma_0\right)'\left(1,x_o',\gamma_0\right)\delta_0\right)K(U_i^-)f(x_o,\gamma_0)k_-(0),$$

and

$$z_{2i}=2\left(-2\left(1,x_o',\gamma_0\right)\delta_0e_i^+ -\delta_0'\left(1,x_o',\gamma_0\right)'\left(1,x_o',\gamma_0\right)\delta_0\right)K(U_i^+)f(x_o,\gamma_0)k_+(0).$$

Given the weak limit of $nh^d\left(\widehat{\Delta}_o^2\left(\gamma_0+\frac{v}{a_n}\right)-\widehat{\Delta}_o^2(\gamma_0)\right)$, we can apply the argmax continuous mapping theorem (Theorem 3.2.2 in Van der Vaart and Wellner, 1996) to obtain the asymptotic distribution of $\widetilde{\gamma}$. We need to check four conditions, just as in the proof of Theorem 2 of Yu (2012). Since these checks are all similar, we omit the details here and only note that $\arg\max_v\overline{D}(v)=\arg\min_v D(v)$, given that $k_-(0)=k_+(0)>0$ and $f(x_o,\gamma_0)>0$. ■

Proof of Theorem 2. We first derive the formula for $\widehat{\delta}$. Following Appendix A.1 of Heckman et al. (1998), we have

$$\begin{aligned}\left(\widehat{a}_+(x_i),\widehat{b}_+'(x_i)\right)' &= \left(I_{d+1},\mathbf{0}\right)\left(X_i'W_i^+X_i\right)^{-1}X_i'W_i^+Y\\ &= \left(I_{d+1},\mathbf{0}\right)H^{-1}\left(H^{-1}X_i'W_i^+X_iH^{-1}\right)^{-1}H^{-1}X_i'W_i^+Y\\ &= \left(I_{d+1},\mathbf{0}\right)H^{-1}\left(Z_i'W_i^+Z_i\right)^{-1}Z_i'W_i^+Y\\ &= \left(I_{d+1},\mathbf{0}\right)H^{-1}\left(\frac{1}{n}\sum_{j=1,j\neq i}^n z_j^iz_j^{i\prime}w_j^{i+}\right)^{-1}\left(\frac{1}{n}\sum_{j=1,j\neq i}^n z_j^iw_j^{i+}y_j\right)\\ &\equiv \left(I_{d+1},\mathbf{0}\right)H^{-1}\left(M_i^+\right)^{-1}r_i^+,\end{aligned}$$

where

$$\begin{aligned}X_i &= \begin{pmatrix}(x_1-x_i,q_1-\widehat{\gamma})^{S_p}\\ \vdots\\ (x_{i-1}-x_i,q_{i-1}-\widehat{\gamma})^{S_p}\\ \mathbf{0}\\ (x_{i+1}-x_i,q_{i+1}-\widehat{\gamma})^{S_p}\\ \vdots\\ (x_n-x_i,q_n-\widehat{\gamma})^{S_p}\end{pmatrix}_{n\times\sum_{\nu=0}^p(\nu+d-1)!/\nu!(d-1)!},\ Y=\begin{pmatrix}y_1\\ \vdots\\ y_{i-1}\\ 0\\ y_{i+1}\\ \vdots\\ y_n\end{pmatrix}_{n\times1},\\ H &= \text{diag}\left\{1,hI_d,\cdots,hI_{(p+d-1)!/p!(d-1)!}\right\},Z_i=X_iH^{-1},z_j^{i\prime}=(x_j-x_i,q_j-\widehat{\gamma})^{S_p}H^{-1},\\ W_i^+ &= \text{diag}\left\{K_h^x(x_1-x_i,x_i)k_h^+(q_1-\widehat{\gamma}),\cdots,K_h^x(x_n-x_i,x_i)k_h^+(q_n-\widehat{\gamma})\right\}=\text{diag}\left\{w_1^{i+},\cdots,w_n^{i+}\right\}_{n\times n},\end{aligned}$$

and $\left(\widehat{a}_-(x_i),\widehat{b}_-'(x_i)\right)'$ are similarly defined with

$$W_i^-=diag\left\{K_h^x(x_1-x_i,x_i)k_h^-(q_1-\widehat{\gamma}),\cdots,K_h^x(x_n-x_i,x_i)k_h^-(q_n-\widehat{\gamma})\right\}$$

replacing W_i^+. Next

$$\begin{pmatrix} \widehat{\Delta} \\ \widehat{\delta}_{xq} \end{pmatrix} = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{q_i-\widehat{\gamma}}{h}\right)\begin{pmatrix} \widehat{a}_-(x_i)-\widehat{a}_+(x_i) \\ \widehat{b}_-(x_i)-\widehat{b}_+(x_i) \end{pmatrix} \Bigg/ \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{q_i-\widehat{\gamma}}{h}\right),$$

and

$$\begin{pmatrix} \widehat{\delta}_{\alpha} \\ \widehat{\delta}_{xq} \end{pmatrix} = \begin{pmatrix} \widehat{\Delta} - \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{q_i-\widehat{\gamma}}{h}\right)(x_i',\widehat{\gamma})\left(\widehat{b}_-(x_i)-\widehat{b}_+(x_i)\right) \Big/ \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{q_i-\widehat{\gamma}}{h}\right) \\ \widehat{\delta}_{xq} \end{pmatrix}.$$

The first step in deriving the asymptotic distribution of $\widehat{\delta}$ is to show that $\widehat{\gamma}$ can be replaced by γ_0. In other words,

$$\sqrt{nh}h\begin{pmatrix} \widehat{\delta}_{\alpha}-\widehat{\delta}_{\alpha}^0 \\ \widehat{\delta}_{xq}-\widehat{\delta}_{xq}^0 \end{pmatrix} \xrightarrow{p} 0,$$

where estimators with superscript 0 denotes the original estimators but with $\widehat{\gamma}$ replaced by γ_0. Of course, we need only show that $\begin{pmatrix} \sqrt{nh}\left(\widehat{\Delta}-\widehat{\Delta}^0\right) \\ \sqrt{nh}h\left(\widehat{\delta}_{xq}-\widehat{\delta}_{xq}^0\right) \end{pmatrix} \xrightarrow{p} 0$ since $\sqrt{nh}h\begin{pmatrix} \widehat{\delta}_{\alpha}-\widehat{\delta}_{\alpha}^0 \\ \widehat{\delta}_{xq}-\widehat{\delta}_{xq}^0 \end{pmatrix}$ is just a linear combination of $\begin{pmatrix} \sqrt{nh}\left(\widehat{\Delta}-\widehat{\Delta}^0\right) \\ \sqrt{nh}h\left(\widehat{\delta}_{xq}-\widehat{\delta}_{xq}^0\right) \end{pmatrix}$. Proposition 5 gives this result. Now,

$$\begin{aligned} &\begin{pmatrix} \sqrt{nh}\left(\widehat{\Delta}^0-\overline{\Delta}^0\right) \\ \sqrt{nh}h\left(\widehat{\delta}_{xq}^0-\delta_{xq0}\right) \end{pmatrix} \\ =\ & \frac{1}{\sqrt{nh}}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right)\begin{pmatrix} \widehat{a}_-^0(x_i)-\widehat{a}_+^0(x_i)-\left(a_-^0(x_i)-a_+^0(x_i)\right) \\ h\left(\widehat{b}_-^0(x_i)-\widehat{b}_+^0(x_i)-\delta_{xq0}\right) \end{pmatrix} \Bigg/ \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right) \\ =\ & \frac{1}{\sqrt{nh}}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right)\begin{pmatrix} \left(\widehat{a}_-^0(x_i)-a_-^0(x_i)\right)-\left(\widehat{a}_+^0(x_i)-a_+^0(x_i)\right) \\ h\left[\left(\widehat{b}_-^0(x_i)-b_-^0(x_i)\right)-\left(\widehat{b}_+^0(x_i)-b_+^0(x_i)\right)\right] \end{pmatrix} \Bigg/ \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right), \end{aligned}$$

where $\left(a_+^0(x_i), b_+^0(x_i)'\right)$ is defined by $a_+^0(x_i) = m_+(x_i) \equiv \lim_{\gamma\to\gamma_0+} m(x_i,\gamma)$ and by $b_+^0(x_i) = \nabla m_+(x_i) \equiv \lim_{\gamma\to\gamma_0+} \left(\partial m(x_i,\gamma)/\partial x_i',\ \partial m(x_i,\gamma)/\partial\gamma\right)'$, with a similar definition for $\left(a_-^0(x_i), b_-^0(x_i)'\right)$, and

$$\overline{\Delta}^0 = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right)\left(a_-^0(x_i)-a_+^0(x_i)\right) \Bigg/ \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right).$$

Note that, under our specification, $b_-^0(x_i)-b_+^0(x_i) = \delta_{xq0}$ and $a_-^0(x_i)-a_+^0(x_i)-(x_i',\gamma_0)(b_-^0(x_i)-b_+^0(x_i)) = \delta_{\alpha 0}$ for any x_i. Also,

$$\begin{aligned} &\sqrt{nh}h\left(\widehat{\delta}_{\alpha}^0-\delta_{\alpha 0}\right) \\ =\ & \sqrt{nh}h\left(\widehat{\Delta}^0-\overline{\Delta}^0\right) - \frac{1}{\sqrt{nh}}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right)(x_i',\gamma_0)\left[h\left(\widehat{b}_-^0(x_i)-b_-^0(x_i)\right)-h\left(\widehat{b}_+^0(x_i)-b_+^0(x_i)\right)\right] \Bigg/ \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right). \end{aligned}$$

We first derive the asymptotic distribution of $\begin{pmatrix} \sqrt{nh}\left(\widehat{\Delta}^0-\overline{\Delta}^0\right) \\ \sqrt{nh}h\left(\widehat{\delta}_{xq}^0-\delta_{xq0}\right) \end{pmatrix}$ and then consider $\sqrt{nh}h\left(\widehat{\delta}_{\alpha}^0-\delta_{\alpha 0}\right)$.

Given assumptions E, F, G$'$, and H$'$, we can apply the arguments in Theorem 3 of Heckman et al. (1998) to obtain

$$\begin{pmatrix} \sqrt{nh}\left(\widehat{\Delta}^0 - \overline{\Delta}^0\right) \\ \sqrt{nh}h\left(\widehat{\delta}^0_{xq} - \delta_{xq0}\right) \end{pmatrix}$$

$$= -\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right)(I_{d+1},\mathbf{0})\left\{ \begin{array}{c} h^{p+1}\left[\left(\overline{M}^+_{i0}\right)^{-1}\overline{r}^{m+}_{i0} - \left(\overline{M}^-_{i0}\right)^{-1}\overline{r}^{m-}_{i0}\right] \\ +\left[\left(\overline{M}^+_{i0}\right)^{-1} r^{e+}_{i0} - \left(\overline{M}^-_{i0}\right)^{-1} r^{e-}_{i0}\right] + R^+_i - R^-_i \end{array}\right\} \Big/ \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right),$$

where $\overline{M}^+_{i0}$ is the square matrix of size $\sum_{\nu=0}^{p}(\nu+d-1)!/\nu!(d-1)!$ with the l-th row, t-th column "block" being, for $0 \le l, t \le p$,

$$\int_0^\infty \int \left(u_x', u_q\right)^{S(l)\prime}\left(u_x', u_q\right)^{S(t)} K^x\left(u_x, x_i\right) k_+(u_q) f(x_i + u_x h, \gamma_0 + u_q h) du_x du_q,$$

$\overline{r}^{m+}_{i0}$ is a $\sum_{\nu=0}^{p}(\nu+d-1)!/\nu!(d-1)!$ by 1 vector with the t-th block being the transpose of

$$\int \left(u_x', u_q\right)^{S(t)}\left[\left(u_x', u_q\right)^{S(p+1)} m_+^{(p+1)}(x_i)\right] K^x\left(u_x, x_i\right) k_+(u_q) f(x_i, \gamma_0) du_x du_q,$$

and $m_+^{(p+1)}(x)$ being a $(p+d)!/(p+1)!(d-1)! \times 1$ vector of the partial derivatives of $m(x,q)$ at $q=\gamma_0+$,

$$r^{e+}_{i0} = \frac{1}{n}\sum_{j=1, j\neq i}^{n} z^i_{j0} w^{i+}_{j0} e_j,$$

with z^i_{j0} and w^{i+}_{j0} being z^i_j and w^{i+}_j but having $\widehat{\gamma}$ replaced by γ_0,

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right)(I_{d+1},\mathbf{0}) R^+_i = o_p(1),$$

and the objects with superscript $-$ are similarly defined. It turns out that the terms associated with $\overline{r}^{m\pm}_{i0}$ will contribute to the bias and the terms associated with $\overline{r}^{e\pm}_{i0}$, which is a U-statistic, will contribute to the variance. Given that $\frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right) \xrightarrow{p} f_q(\gamma_0)$, we need only concentrate on the numerator.

First, analyze the bias.

$$\begin{aligned} & \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right)(I_{d+1},\mathbf{0})\left[\left(\overline{M}^+_{i0}\right)^{-1}\overline{r}^{m+}_{i0} - \left(\overline{M}^-_{i0}\right)^{-1}\overline{r}^{m-}_{i0}\right]\right] \\ = \; & (I_{d+1},\mathbf{0})\int\int\left[\left(\overline{M}^+_{i0}\right)^{-1}\overline{r}^{m+}_{i0} - \left(\overline{M}^-_{i0}\right)^{-1}\overline{r}^{m-}_{i0}\right] f(x_i|q_i) dx_i k_h\left(q_i - \gamma_0\right) f(q_i) dq_i \\ \to \; & (I_{d+1},\mathbf{0})\left[\left(M^+_o\right)^{-1} B^+ - \left(M^-_o\right)^{-1} B^-\right] \mathbb{E}[g^{(p+1)}(x,\gamma_0)\big| q = \gamma_0] f_q(\gamma_0), \end{aligned}$$

where $M^\pm_o$ and $B^\pm$ are defined in the main text, and $m_+^{(p+1)}(x_i) = m_-^{(p+1)}(x_i) = g^{(p+1)}(x_i,\gamma_0)$ under Assumption G$'$. Note here that the kernel K^x is replaced by K because the data in the h neighborhood of the boundary of $\mathcal{X}$ can be neglected asymptotically. Also, we can calculate that the variance of this term is $O\left(\frac{1}{nh}\right) = o(1)$, so it converges in probability to its expectation. Second, analyze the variance. Taking the

lth element of $\begin{pmatrix} \sqrt{nh}\left(\widehat{\Delta}^0 - \overline{\Delta}^0\right) \\ \sqrt{nh}h\left(\widehat{\delta}^0_{xq} - \delta_{xq0}\right) \end{pmatrix}$, we consider

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_l'\left[\left(\overline{M}_{i0}^{+}\right)^{-1} r_{i0}^{e+} - \left(\overline{M}_{i0}^{-}\right)^{-1} r_{i0}^{e-}\right],$$

which is a second-order U-statistic. From Lemma 8.4 of Newey and McFadden (1994), this U-statistic is asymptotically equivalent to $\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} m_n(x_i, q_i, e_i)$, where

$$\begin{aligned} m_n(x_j, q_j, e_j) &= \mathbb{E}\left[k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_l'\left[\left(\overline{M}_{i0}^{+}\right)^{-1} z_{j0}^{i} w_{j0}^{i+} e_j - \left(\overline{M}_{i0}^{-}\right)^{-1} z_{j0}^{i} w_{j0}^{i-} e_j\right]\middle| x_j, q_j, e_j\right] \\ &= e_j \int k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_l'\left[\left(\overline{M}_{i0}^{+}\right)^{-1} z_{j0}^{i} w_{j0}^{i+} - \left(\overline{M}_{i0}^{-}\right)^{-1} z_{j0}^{i} w_{j0}^{i-}\right] f(x_i, q_i) dx_i dq_i, \end{aligned}$$

We apply the Liapunov central limit theorem to derive the asymptotic distribution. It is standard to check that the Liapunov condition is satisfied, so we concentrate on calculating the asymptotic variance as follows.

$$\begin{aligned} & \frac{1}{h}\mathbb{E}\left[e_j^2\left(\int k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_l'\left[\left(\overline{M}_{i0}^{+}\right)^{-1} z_{j0}^{i} w_{j0}^{i+} - \left(\overline{M}_{i0}^{-}\right)^{-1} z_{j0}^{i} w_{j0}^{i-}\right] f(x_i, q_i) dx_i dq_i\right)^2\right] \\ \approx\ & \frac{1}{h}\mathbb{E}\left[e_j^2\left(\int k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_l'\left(\overline{M}_{i0}^{+}\right)^{-1} z_{j0}^{i} w_{j0}^{i+} f(x_i, q_i) dx_i dq_i\right)^2\right] \\ & + \frac{1}{h}\mathbb{E}\left[e_j^2\left(\int k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_l'\left(\overline{M}_{i0}^{-}\right)^{-1} z_{j0}^{i} w_{j0}^{i-} f(x_i, q_i) dx_i dq_i\right)^2\right] \\ & - \frac{2}{h}\mathbb{E}\left[\begin{array}{c} e_j^2\left(\int k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_l'\left(\overline{M}_{i0}^{+}\right)^{-1} z_{j0}^{i} w_{j0}^{i+} f(x_i, q_i) dx_i dq_i\right) \\ \left(\int k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_l'\left(\overline{M}_{i0}^{-}\right)^{-1} z_{j0}^{i} w_{j0}^{i-} f(x_i, q_i) dx_i dq_i\right) \end{array}\right] \\ \equiv\ & T1 + T2 + T3. \end{aligned}$$

We analyze $T1$, $T2$ and $T3$ in turn.

$$\begin{aligned} T1 &\approx \frac{1}{h}\mathbb{E}\left[e_j^2\left(k_+\left(\frac{q_j-\gamma_0}{h}\right)\int k(u_q)\mathbf{e}_l'\left(M_o^{+}\right)^{-1}\left[\left(u_x', \frac{q_j-\gamma_0}{h}\right)^{S_p}\right]' K(u_x) du_x du_q\right)^2\right] \\ &\approx \int\int \sigma_+^2(x_j)\left(k_+(v_q)\int k(u_q)\mathbf{e}_l'\left(M_o^{+}\right)^{-1}\left[(u_x', v_q)^{S_p}\right]' K(u_x) du_x du_q\right)^2 f(x_j, \gamma_0) dx_j dv_q \\ &= \mathbb{E}\left[\int k_+^2(v_q)\sigma_+^2(x_j) C_l^{+}(v_q)^2 dv_q | q_j = \gamma_0\right] f_q(\gamma_0). \end{aligned}$$

Similarly,

$$T2 \approx \mathbb{E}\left[\int k_-^2(v_q)\sigma_-^2(x_j) C_l^{-}(v_q)^2 dv_q | q_j = \gamma_0\right] f_q(\gamma_0),$$

and $T3 = 0$ since $k_+\left(\frac{q_j-\gamma_0}{h}\right)k_-\left(\frac{q_j-\gamma_0}{h}\right) = 0$. In summary,

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n} m_n(x_i, q_i, e_i) \xrightarrow{d} N\left(0, \mathbb{E}\left[\int\left[k_+^2(v_q)\sigma_+^2(x) C_l^{+}(v_q)^2 + k_-^2(v_q)\sigma_-^2(x) C_l^{-}(v_q)^2\right] dv_q \middle| q = \gamma_0\right] f_q(\gamma_0)\right),$$

and the asymptotic distribution of $\sqrt{nh}h\left(\widehat{\delta}^0_{xq}-\delta_{xq0}\right)$ follows as in the theorem.

We next derive the asymptotic distribution of $\sqrt{nh}h\left(\widehat{\delta}^0_{\alpha}-\delta_{\alpha 0}\right)$. Given that $\sqrt{nh}\left(\widehat{\Delta}^0-\overline{\Delta}^0\right)=O_p(1)$ under Assumption H$'$, the term $\sqrt{nh}h\left(\widehat{\Delta}^0-\overline{\Delta}^0\right)$ can be neglected, and $\sqrt{nh}h\left(\widehat{\delta}^0_{\alpha}-\delta_{\alpha 0}\right)$ has the same asymptotic distribution as

$$-\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}k\left(\frac{q_i-\gamma_0}{h}\right)(x_i',\gamma_0)\left[h\left(\widehat{b}^0_-(x_i)-b^0_-(x_i)\right)-h\left(\widehat{b}^0_+(x_i)-b^0_+(x_i)\right)\right]\Bigg/\frac{1}{nh}\sum_{i=1}^{n}k\left(\frac{q_i-\gamma_0}{h}\right).$$

For the bias, note that

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}k\left(\frac{q_i-\gamma_0}{h}\right)(x_i',\gamma_0)\left(\mathbf{0},I_d,\mathbf{0}\right)\left[\left(\overline{M}^+_{i0}\right)^{-1}\overline{r}^{m+}_{i0}-\left(\overline{M}^-_{i0}\right)^{-1}\overline{r}^{m-}_{i0}\right]\right]\\
=&\int\int(x_i',\gamma_0)\left(\mathbf{0},I_d,\mathbf{0}\right)\left[\left(\overline{M}^+_{i0}\right)^{-1}\overline{r}^{m+}_{i0}-\left(\overline{M}^-_{i0}\right)^{-1}\overline{r}^{m-}_{i0}\right]f(x_i|q_i)dx_ik_h\left(q_i-\gamma_0\right)f(q_i)dq_i\\
\to&\ \mathbb{E}\left[(x',\gamma_0)\left(\mathbf{0},I_d,\mathbf{0}\right)\left[\left(M^+_o\right)^{-1}B^+-\left(M^-_o\right)^{-1}B_-\right]g^{(p+1)}(x,\gamma_0)\Big|q=\gamma_0\right]f_q(\gamma_0).
\end{aligned}$$

For the variance, the corresponding U-projection $m_n(x_i,q_i,e_i)$ is

$$e_j\int k_h\left(\frac{q_i-\gamma_0}{h}\right)(x_i',\gamma_0)\left(\mathbf{0},I_d,\mathbf{0}\right)\left[\left(\overline{M}^+_{i0}\right)^{-1}z^i_{j0}w^{i+}_{j0}-\left(\overline{M}^-_{i0}\right)^{-1}z^i_{j0}w^{i-}_{j0}\right]f(x_i,q_i)dx_idq_i.$$

We can proceed in a similar fashion to the above in deriving the asymptotic variance. For example, the corresponding form to $T1$ is

$$\begin{aligned}
T1\ \approx&\ \frac{1}{h}\mathbb{E}\left[e_j^2\left(k^+_h\left(q_j-\gamma_0\right)\int k\left(u_q\right)(x_j',\gamma_0)\left(\mathbf{0},I_d,\mathbf{0}\right)\left(M^+_o\right)^{-1}\left[\left(u_x',\frac{q_j-\gamma_0}{h}\right)^{S_p}\right]'K\left(u_x\right)du_xdu_q\right)^2\right]\\
\approx&\int\int\sigma^2_+(x_j)\left(k_+\left(v_q\right)\int k\left(u_q\right)(x_j',\gamma_0)\left(\mathbf{0},I_d,\mathbf{0}\right)\left(M^+_o\right)^{-1}\left[\left(u_x'\ v_q\right)^{S_p}\right]'K\left(u_x\right)du_xdu_q\right)^2f(x_j,\gamma_0)dx_jdv_q\\
=&\ \mathbb{E}\left[\int k^2_+\left(v_q\right)\sigma^2_+(x_j)C^+(x_j,v_q)^2dv_q|q_j=\gamma_0\right]f_q(\gamma_0).
\end{aligned}$$

■

Proof of Corollary 2. The asymptotic distribution of $\sqrt{nh}\left(\widehat{\Delta}^0-\Delta_0\right)$ is more involved since it includes variations from two components as in

$$\sqrt{nh}\left(\widehat{\Delta}^0-\Delta_0\right)=\sqrt{nh}\left(\widehat{\Delta}^0-\overline{\Delta}^0\right)+\sqrt{nh}\left(\overline{\Delta}^0-\Delta_0\right).$$

First note that

$$\begin{aligned}
\sqrt{nh}\left(\widehat{\Delta}^0-\Delta_0\right)=&\ \sqrt{nh}\left(\frac{\widehat{\Delta}^0_N}{\widehat{f}_q(\gamma_0)}-\frac{\Delta_0f_q(\gamma_0)}{f_q(\gamma_0)}\right)\\
\approx&\ \frac{\sqrt{nh}\left[\widehat{\Delta}^0_N-\Delta_0f_q(\gamma_0)\right]-\sqrt{nh}\Delta_0\left[\widehat{f}_q(\gamma_0)-f_q(\gamma_0)\right]}{f_q(\gamma_0)}\\
=&\ \frac{\sqrt{nh}\left[\widehat{\Delta}^0_N-\overline{\Delta}^0_N\right]+\sqrt{nh}\left[\overline{\Delta}^0_N-\Delta_0f_q(\gamma_0)\right]-\sqrt{nh}\Delta_0\left[\widehat{f}_q(\gamma_0)-f_q(\gamma_0)\right]}{f_q(\gamma_0)}
\end{aligned}$$

where $\widehat{\Delta}_N^0$ and $\overline{\Delta}_N^0$ are the numerators of $\widehat{\Delta}^0$ and $\overline{\Delta}^0$, and $\widehat{f}_q(\gamma_0) = \frac{1}{nh}\sum_{i=1}^n k\left(\frac{q_i-\gamma_0}{h}\right)$. From the earlier analysis in the proof of Theorem 2, $\widehat{\Delta}_N^0 - \overline{\Delta}_N^0$ satisfies

$$\begin{aligned}
&\sqrt{nh}\left(\widehat{\Delta}_N^0 - \overline{\Delta}_N^0 - h^{p+1}\mathbf{e}_1\left[\left(M_o^-\right)^{-1}B^- - \left(M_o^+\right)^{-1}B^+\right]\mathbb{E}[g^{(p+1)}(x,\gamma_0)\Big| q=\gamma_0]f_q(\gamma_0)\right)\\
\approx\ & \frac{1}{\sqrt{nh}}\sum_{i=1}^n k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_1'\left[\left(\overline{M}_{i0}^-\right)^{-1}r_{i0}^{e-} - \left(\overline{M}_{i0}^+\right)^{-1}r_{i0}^{e+}\right]\\
\approx\ & \frac{1}{\sqrt{nh}}\sum_{j=1}^n e_j\int k_h\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_1'\left[\left(\overline{M}_{i0}^-\right)^{-1}z_{j0}^i w_{j0}^{i-} - \left(\overline{M}_{i0}^+\right)^{-1}z_{j0}^i w_{j0}^{i+}\right]f(x_i,q_i)dx_i dq_i,
\end{aligned}$$

and also

$$\begin{aligned}
&\sqrt{nh}\left(\overline{\Delta}_N^0 - \mathbb{E}\left[\overline{\Delta}_N^0\right] - \Delta_0\left[\widehat{f}_q(\gamma_0) - \mathbb{E}\left[\widehat{f}_q(\gamma_0)\right]\right]\right)\\
=\ & \frac{1}{\sqrt{nh}}\sum_{i=1}^n\left\{k\left(\frac{q_i-\gamma_0}{h}\right)\left(a_-^0(x_i)-a_+^0(x_i)\right) - \mathbb{E}\left[k\left(\frac{q_i-\gamma_0}{h}\right)\left(a_-^0(x_i)-a_+^0(x_i)\right)\right]\right\}\\
&-\frac{\Delta_0}{\sqrt{nh}}\sum_{i=1}^n\left\{k\left(\frac{q_i-\gamma_0}{h}\right) - \mathbb{E}\left[k\left(\frac{q_i-\gamma_0}{h}\right)\right]\right\}.
\end{aligned}$$

It is not hard to see that these two influence functions are uncorrelated, so the variance of $\sqrt{nh}\left(\widehat{\Delta}^0 - \Delta_0\right)$ is the sum of the variances of these two parts. The variance of the first part is derived in the proof of Theorem 2. As to the second part, note that

$$\begin{aligned}
&\frac{1}{h}\mathbb{E}\left[k\left(\frac{q_i-\gamma_0}{h}\right)^2\left(a_-^0(x_i)-a_+^0(x_i)\right)^2\right]\\
=\ & \frac{1}{h}\int k\left(\frac{q_i-\gamma_0}{h}\right)^2\left(a_-^0(x_i)-a_+^0(x_i)\right)^2 f(x_i,q_i)dx_i dq_i\\
\approx\ & \int k\left(v_q\right)^2\left(a_-^0(x_i)-a_+^0(x_i)\right)^2 f(x_i,\gamma_0)dx_i dv_q\\
=\ & \int k\left(v_q\right)^2 dv_q\mathbb{E}[\left(a_-^0(x_i)-a_+^0(x_i)\right)^2|q_i=\gamma_0]f_q(\gamma_0).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{1}{h}\mathbb{E}\left[k\left(\frac{q_i-\gamma_0}{h}\right)^2\left(a_-^0(x_i)-a_+^0(x_i)\right)\right] &\approx \int k\left(v_q\right)^2 dv_q\Delta_0 f_q(\gamma_0),\\
\frac{1}{h}\mathbb{E}\left[k\left(\frac{q_i-\gamma_0}{h}\right)^2\right] &\approx \int k\left(v_q\right)^2 dv_q f_q(\gamma_0),
\end{aligned}$$

so the variance of the second part is approximately

$$\begin{aligned}
&\int k\left(v_q\right)^2 dv_q\mathbb{E}[\left(a_-^0(x_i)-a_+^0(x_i)\right)^2|q_i=\gamma_0]f_q(\gamma_0) + \Delta_0^2\int k\left(v_q\right)^2 dv_q f_q(\gamma_0) - 2\Delta_0\int k\left(v_q\right)^2 dv_q\Delta_0 f_q(\gamma_0)\\
=\ & \int k\left(v_q\right)^2 dv_q\left(\mathbb{E}[\left(a_-^0(x_i)-a_+^0(x_i)\right)^2|q_i=\gamma_0] - \Delta_0^2\right)f_q(\gamma_0).
\end{aligned}$$

For the bias of the second part, note that

$$
\begin{aligned}
& \mathbb{E}\left[\overline{\Delta}_N^0\right]-\Delta_0 f_q(\gamma_0) \\
= & \int k_h\left(q_i-\gamma_0\right)\left(a_-^0(x_i)-a_+^0(x_i)\right) f(x_i,q_i)dx_i dq_i-\Delta_0 f_q(\gamma_0) \\
= & \int k\left(v_q\right)\left(a_-^0(x_i)-a_+^0(x_i)\right) f(x_i,\gamma_0+v_q h)dx_i dv_q-\Delta_0 f_q(\gamma_0) \\
\approx & \int k\left(v_q\right)\left(a_-^0(x_i)-a_+^0(x_i)\right)\sum\nolimits_{l=1}^{p+1}\frac{1}{l!}f_\gamma^{(l)}(x_i,\gamma_0)\left(v_q h\right)^l dx_i dv_q \\
= & \sum\nolimits_{l=1}^{p+1}\frac{h^l}{l!}\left[\int k\left(v_q\right)v_q^l dv_q\right]\int\left(a_-^0(x_i)-a_+^0(x_i)\right) f_\gamma^{(l)}(x_i,\gamma_0)dx_i
\end{aligned}
$$

where $f_\gamma^{(l)}(x_i,\gamma_0)$ is the lth order partial derivative of $f(x_i,\gamma)$ with respect to γ evaluated at $\gamma=\gamma_0$, and

$$
\begin{aligned}
\mathbb{E}\left[\widehat{f}_q(\gamma_0)\right]-f_q(\gamma_0) &= \int k_h\left(q_i-\gamma_0\right)f(q_i)dq_i-f_q(\gamma_0) \\
&= \int k\left(v_q\right)f(\gamma_0+v_q h)dv_q-f_q(\gamma_0) \\
&\approx \int k\left(v_q\right)\sum\nolimits_{l=1}^{p+1}\frac{1}{l!}f_\gamma^{(l)}(\gamma_0)\left(v_q h\right)^l dv_q \\
&= \sum\nolimits_{l=1}^{p+1}\frac{h^l}{l!}\left[\int k\left(v_q\right)v_q^l dv_q\right]f_\gamma^{(l)}(\gamma_0),
\end{aligned}
$$

where $f_\gamma^{(l)}(\gamma_0)$ is the lth order derivative of $f_q(\gamma)$ with respect to γ evaluated at $\gamma=\gamma_0$. Under Assumptions F$'$ and K$'$, all the terms except the $(p+1)$th term in $\mathbb{E}\left[\overline{\Delta}_N^0\right]-\Delta_0 f_q(\gamma_0)$ and $\mathbb{E}\left[\widehat{f}_q(\gamma_0)\right]-f_q(\gamma_0)$ would disappear. In sum, the asymptotic distribution of $\sqrt{nh}\left(\widehat{\Delta}^0-\Delta_0\right)$ is as stated in the theorem. ■

Proof of Theorem 3. First note the following explicit formula for $\left(\overline{\delta}_\alpha,\widetilde{\delta}_x'\right)'$ from the extremum estimation problem (equation (7) in the main text)

$$
\left(\overline{\delta}_\alpha,\widetilde{\delta}_x'\right)'=\left(\frac{1}{n}\sum_{i=1}^n(1,x_i')'(1,x_i')k_h\left(q_i-\widehat{\gamma}\right)\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n(1,x_i')'k_h\left(q_i-\widehat{\gamma}\right)\left(\widehat{a}_-(x_i)-\widehat{a}_+(x_i)\right)\right).
$$

By similar analysis to the proof of Theorem 2, $\widehat{\gamma}$ in $\left(\overline{\delta}_\alpha,\widetilde{\delta}_x'\right)'$ can be replaced by γ_0 without affecting its asymptotic distribution. Also, $\widehat{a}_-(x_i)-\widehat{a}_+(x_i)$ can be replaced by its linear approximation with no asymptotic impact. In summary,

$$
\begin{aligned}
& \sqrt{nh}\left(\left(\overline{\delta}_\alpha,\widetilde{\delta}_x'\right)'-\left(\delta_{\alpha 0}+\gamma_0\delta_{q0},\delta_{x0}'\right)'\right) \\
\approx & \left(\frac{1}{n}\sum_{i=1}^n(1,x_i')'(1,x_i')k_h\left(q_i-\gamma_0\right)\right)^{-1}\cdot\left(\frac{1}{\sqrt{nh}}\sum_{i=1}^n(1,x_i')'k\left(\frac{q_i-\gamma_0}{h}\right)\right. \\
& \left.\mathbf{e}_1\left\{h^{p+1}\left[\left(\overline{M}_{i0}^-\right)^{-1}\overline{r}_{i0}^{m-}-\left(\overline{M}_{i0}^+\right)^{-1}\overline{r}_{i0}^{m+}\right]+\left[\left(\overline{M}_{i0}^-\right)^{-1}r_{i0}^{e-}-\left(\overline{M}_{i0}^+\right)^{-1}r_{i0}^{e+}\right]\right\}\right),
\end{aligned}
$$

where $\overline{M}_{i0}^\pm$, $\overline{r}_{i0}^{m\pm}$ and $r_{i0}^{e\pm}$ are defined in the proof of Theorem 2.

By standard methods, the denominator converges in probability to $M\cdot f_q(\gamma_0)$, where M is defined in the main text, so we concentrate on the numerator. First, consider the bias term. From the proof of Theorem 2,

$$\frac{1}{nh}\sum_{i=1}^{n}(1,x_i)'\,k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_1\left[\left(\overline{M}_{i0}^{-}\right)^{-1}\overline{r}_{i0}^{m-}-\left(\overline{M}_{i0}^{+}\right)^{-1}\overline{r}_{i0}^{m+}\right]$$
$$\xrightarrow{p}\mathbb{E}\left[(1,x')'\,\mathbf{e}_1\left[\left(M_o^{-}\right)^{-1}B^{-}-\left(M_o^{+}\right)^{-1}B^{+}\right]g^{(p+1)}(x,\gamma_0)\Big|\,q=\gamma_0\right]f_q(\gamma_0).$$

Next consider the variance. We need to calculate the covariance between the lth and tth element of the numerator, $l,t=1,\cdots,d$. Taking the $(l+1)$th element of the numerator, $l=1,\cdots .d-1$, we consider

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}x_{li}k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_1'\left[\left(\overline{M}_{i0}^{+}\right)^{-1}r_{i0}^{e+}-\left(\overline{M}_{i0}^{-}\right)^{-1}r_{i0}^{e-}\right],$$

which is a second-order U-statistic. From Lemma 8.4 of Newey and McFadden (1994), this U-statistic is asymptotically equivalent to $\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}m_n^l(x_i,q_i,e_i)$, where

$$\begin{aligned}m_n^l(x_j,q_j,e_j) &= \mathbb{E}\left[x_{li}k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_1'\left[\left(\overline{M}_{i0}^{+}\right)^{-1}z_{j0}^{i}w_{j0}^{i+}e_j-\left(\overline{M}_{i0}^{-}\right)^{-1}z_{j0}^{i}w_{j0}^{i-}e_j\right]\Big|\,x_j,q_j,e_j\right]\\ &= e_j\int x_{li}k\left(\frac{q_i-\gamma_0}{h}\right)\mathbf{e}_1'\left[\left(\overline{M}_{i0}^{+}\right)^{-1}z_{j0}^{i}w_{j0}^{i+}-\left(\overline{M}_{i0}^{-}\right)^{-1}z_{j0}^{i}w_{j0}^{i-}\right]f(x_i,q_i)dx_idq_i.\end{aligned}$$

It is not hard to show that

$$\frac{1}{nh}\sum_{i=1}^{n}m_n^l(x_i,q_i,e_i)m_n^t(x_i,q_i,e_i)\xrightarrow{p}\mathbb{E}\left[x_lx_t\int\left[k_+^2\left(v_q\right)\sigma_+^2(x)C_1^+(v_q)^2+k_-^2\left(v_q\right)\sigma_-^2(x)C_1^-(v_q)^2\right]dv_q\Big|\,q=\gamma_0\right]f_q(\gamma_0).$$

Then, applying the Liapunov central limit theorem, the asymptotic distribution of $\sqrt{nh}\left(\overline{\delta}_\alpha-\delta_{\alpha0}-\gamma_0\delta_{q0}\right)$ and $\sqrt{nh}\left(\widetilde{\delta}_{x_l}-\delta_{x_l0}\right)$, $l=1,\cdots,d-1$, follows as in the theorem.

When $\gamma_0=0$, $\sqrt{nh}\left(\widetilde{\delta}_\alpha-\delta_{\alpha0}\right)=\sqrt{nh}\left(\overline{\delta}_\alpha-\widehat{\gamma}\widehat{\delta}_q-\delta_{\alpha0}\right)=\sqrt{nh}\left(\overline{\delta}_\alpha-\delta_{\alpha0}\right)-\sqrt{nh}O_p(n^{-1})O_p(\left(\sqrt{nh}h\right)^{-1}+h^p)=\sqrt{nh}\left(\overline{\delta}_\alpha-\delta_{\alpha0}\right)+o_p(1)$, so $\sqrt{nh}\left(\widetilde{\delta}_\alpha-\delta_{\alpha0}\right)$ have the same asymptotic distribution as $\sqrt{nh}\left(\overline{\delta}_\alpha-\delta_{\alpha0}\right)$. When $\gamma_0\neq0$, the convergence rate of $\widetilde{\delta}_\alpha-\delta_{\alpha0}$ is $\sqrt{nh}h$. It is obvious that $\sqrt{nh}h\left(\overline{\delta}_\alpha-\widehat{\gamma}\widehat{\delta}_q-\delta_{\alpha0}\right)-\sqrt{nh}h\left(\overline{\delta}_\alpha-\gamma_0\widehat{\delta}_q-\delta_{\alpha0}\right)=\sqrt{nh}hO_p(n^{-1})O_p(\left(\sqrt{nh}h\right)^{-1}+h^p)=o_p(1)$. Also,

$$\begin{aligned}\sqrt{nh}h\left(\overline{\delta}_\alpha-\gamma_0\widehat{\delta}_q-\delta_{\alpha0}\right) &= \sqrt{nh}h\left(\overline{\delta}_\alpha-\delta_{\alpha0}-\gamma_0\delta_{q0}\right)-\gamma_0\sqrt{nh}h\left(\widehat{\delta}_q-\delta_{q0}\right)\\ &= o_p(1)-\gamma_0\sqrt{nh}h\left(\widehat{\delta}_q-\delta_{q0}\right).\end{aligned}$$

So $\sqrt{nh}h\left(\widetilde{\delta}_\alpha-\delta_{\alpha0}\right)$ has the same asymptotic distribution as $-\gamma_0\sqrt{nh}h\left(\widehat{\delta}_q-\delta_{q0}\right)$. ■

Proof of Theorem 4. Assume the densities of $(x',q)'$ and e are known. Since the minimax risk for a larger class of probability models must not be smaller than that for a smaller class of probability models, the lower bound for a particular distributional assumption also holds for a wider class of distributions. To simplify the calculation, assume e_i is iid $N(0,1)$ and $(x_i',q_i)'$ is iid uniform on $\mathcal{X}\times\mathcal{N}$, where $\mathcal{N}$ is specified as $[-\zeta,\zeta]$. Such a specification also appears in Fan (1993) where it is called the assumption of richness of joint densities. We will use the technique in Sun (2005) to develop our results. This technique is also implicitly used in Stone (1980) and the essential part of the technique can be cast in the language of Neyman-Pearson testing.

Let P, Q be probability measures defined on the same measurable space $(\Omega, \mathcal{A})$ with the affinity between the two measures defined as usual to be

$$\pi(P,Q) = \inf\left(\mathbb{E}_P[\phi] + \mathbb{E}_Q[1-\phi]\right),$$

where the infimum is taken over the measurable function ϕ such that $0 \leq \phi \leq 1$. In other words, $\pi(P,Q)$ is the smallest sum of type I and type II errors of any test between P and Q. It is a natural measure of the difficulty of distinguishing P and Q. Suppose μ is a measure dominating both P and Q with corresponding densities p and q. It follows from the Neyman-Pearson lemma that the infimum is achieved by setting $\phi = 1(p \leq q)$ and then

$$\begin{aligned}\pi(P,Q) &= \int 1(p \leq q) p d\mu + \int 1(p > q) q d\mu \\ &= 1 - \frac{1}{2}\int |p-q|\, d\mu \equiv 1 - \frac{1}{2}\|P-Q\|_1,\end{aligned}$$

where $\|\cdot\|_1$ is the L_1 distance between two probability measures. Now consider a pair of probability models $P, Q \in \mathcal{P}(s,B)$ such that $|\delta_\alpha(P) - \delta_\alpha(Q)| \geq \epsilon$. For any estimator $\widehat{\delta}$, we have

$$1\left(\left\|\widehat{\delta}_\alpha - \delta_\alpha(P)\right\| > \epsilon/2\right) + 1\left(\left\|\widehat{\delta}_\alpha - \delta_\alpha(Q)\right\| > \epsilon/2\right) \geq 1.$$

Let

$$\phi = \frac{1\left(\left|\widehat{\delta}_\alpha - \delta_\alpha(P)\right| > \epsilon/2\right)}{1\left(\left|\widehat{\delta}_\alpha - \delta_\alpha(P)\right| > \epsilon/2\right) + 1\left(\left|\widehat{\delta}_\alpha - \delta_\alpha(Q)\right| > \epsilon/2\right)}.$$

Then $0 \leq \phi \leq 1$ and

$$\begin{aligned}\sup_{\mathbb{P} \in \mathcal{P}(s,B)} \mathbb{P}\left(\left|\widehat{\delta}_\alpha - \delta_\alpha(\mathbb{P})\right| > \epsilon/2\right) &\geq \frac{1}{2}\left\{P\left(\left|\widehat{\delta}_\alpha - \delta_\alpha(P)\right| > \epsilon/2\right) + Q\left(\left|\widehat{\delta}_\alpha - \delta_\alpha(Q)\right| > \epsilon/2\right)\right\} \\ &\geq \frac{1}{2}\mathbb{E}_P[\phi] + \frac{1}{2}\mathbb{E}_Q[1-\phi].\end{aligned}$$

Therefore

$$\inf_{\widehat{\delta}_\alpha} \sup_{\mathbb{P} \in \mathcal{P}(s,B)} \mathbb{P}\left(\left|\widehat{\delta}_\alpha - \delta_\alpha(\mathbb{P})\right| > \epsilon/2\right) \geq \frac{1}{2}\pi(P,Q)$$

for any P and Q such that $|\delta_\alpha(P) - \delta_\alpha(Q)| \geq \epsilon$. So we need only search for the pair (P,Q) which minimize $\pi(P,Q)$ subject to the constraint $|\delta_\alpha(P) - \delta_\alpha(Q)| \geq \epsilon$. To obtain a lower bound with a sequence of independent observations, let $(\Omega, \mathcal{A})$ be the product space and $\mathcal{P}(s,B)$ be the family of product probabilities on such a space. Then for any pair of finite-product measures $P = \prod_{i=1}^n P_i$ and $Q = \prod_{i=1}^n Q_i$, the minimax risk satisfies

$$\inf_{\widehat{\delta}_\alpha} \sup_{P \in \mathcal{P}(s,B)} P\left(\left|\widehat{\delta}_\alpha - \delta_\alpha(P)\right| > \epsilon/2\right) \geq \frac{1}{2}\left(1 - \frac{1}{2}\left\|\prod_{i=1}^n P_i - \prod_{i=1}^n Q_i\right\|_1\right)$$

provided that $|\delta_\alpha(P) - \delta_\alpha(Q)| \geq \epsilon$. From Pollard (1993), if $dQ_i/dP_i = 1 + \Delta_i(\cdot)$, then

$$\left\|\prod_{i=1}^n P_i - \prod_{i=1}^n Q_i\right\|_1 \leq \exp\left(\sum_{i=1}^n \nu_i^2\right) - 1,$$

where $\nu_i^2 = \mathbb{E}_{P_i}[\Delta_i^2(\cdot)]$ is finite. So

$$\inf_{\widehat{\delta}_\alpha} \sup_{P\in\mathcal{P}(s,B)} P\left(\left|\widehat{\delta}_\alpha - \delta_\alpha(P)\right| > \epsilon/2\right) \geq \frac{1}{2}\left(\frac{3}{2} - \exp\left(\sum_{i=1}^n \nu_i^2\right)\right) \tag{3}$$

provided that $|\delta_\alpha(P) - \delta_\alpha(Q)| \geq \epsilon$.

It remains to find probabilities P and Q that are difficult to distinguish by the data set $\{(x_i', q_i, y_i)\}_{i=1}^n$. First assume $\gamma_0 \neq 0$. Without loss of generality, let $\gamma_0 > 0$. Under P, the data is generated according to

$$y_i = g_P(x_i, q_i) + (\delta_{\alpha P} + x_i'\delta_{xP} + q_i\delta_{qP})\, 1(q_i \leq \gamma_0) + e_i,$$

and under Q, g_P and δ_P are changed to g_Q and δ_Q, respectively. We now specify g and δ for each model. For P, let $g_P = 0$ and $\delta_P = 0$; for Q, let

$$g_Q(x,q) = -\xi\eta^s\varphi_q\left(\frac{q-\gamma_0}{\eta}\right),\ \delta_{\alpha Q} = -\xi\gamma_0\eta^{s-1},\ \delta_{xQ} = 0, \text{ and } \delta_{qQ} = \xi\eta^{s-1},$$

where ξ is a positive constant, $\eta = n^{-1/(2s+1)}$, φ_q is an infinitely differentiable function in q satisfying (i) $\varphi_q(v) = 0$ for $v \geq 0$, (ii) $\varphi_q(v) = v$, for $v \leq -\zeta$, and (iii) $v - \varphi_q(v) \in (0,1)$ for $v \in (-\zeta, 0)$. It is not hard to check that $g_Q(x,q) \in C(s,B)$ for some $B > 0$, so it remains to compute the L_1 distance between the two measures. Let the density of Q_i with respect to P_i be $1 + \Delta_i(\cdot)$, then

$$\Delta_i(x_i, q_i, y_i) = \begin{cases} \phi(y_i - g_Q(x_i,q_i) - \delta_{\alpha Q} - q_i\delta_{qQ})/\phi(y_i) - 1, & \text{if } q_i \in [\gamma_0 - \zeta\eta, \gamma_0], \\ 0, & \text{otherwise} \end{cases}$$

where $\phi(\cdot)$ is the standard normal pdf. Therefore,

$$\begin{aligned}
\mathbb{E}_{P_i}[\Delta_i^2] &= \int_{\gamma_0-\zeta\eta}^{\gamma_0}\int_0^1\cdots\int_0^1\int_{-\infty}^{\infty}\left[\phi(y - g_Q(x,q) - \delta_{\alpha Q} - q\delta_{qQ})/\phi(y) - 1\right]^2\phi(y)f(x,q)dydxdq \\
&= \frac{1}{2\zeta}\int_{\gamma_0-\zeta\eta}^{\gamma_0}\int_0^1\cdots\int_0^1\int_{-\infty}^{\infty}\phi(y - g_Q(x,q) - \delta_{\alpha Q} - q\delta_{qQ})^2/\phi(y)dydxdq \\
&\quad -\frac{1}{\zeta}\int_{\gamma_0-\zeta\eta}^{\gamma_0}\int_0^1\cdots\int_0^1\int_{-\infty}^{\infty}\phi(y - g_Q(x,q) - \delta_{\alpha Q} - q\delta_{qQ})dydxdq + \frac{\eta}{2} \\
&= \frac{1}{2\zeta}\int_{\gamma_0-\zeta\eta}^{\gamma_0}\int_0^1\cdots\int_0^1\int_{-\infty}^{\infty}\phi(y - g_Q(x,q) - \delta_{\alpha Q} - q\delta_{qQ})^2/\phi(y)dydxdq - \frac{\eta}{2}.
\end{aligned}$$

Plugging in the standard normal pdf yields

$$\begin{aligned}
\mathbb{E}_{P_i}[\Delta_i^2] &= \frac{1}{2\zeta}\int_{\gamma_0-\zeta\eta}^{\gamma_0}\int_0^1\cdots\int_0^1\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{2(y - g_Q(x,q) - \delta_{\alpha Q} - q\delta_{qQ})^2}{2} + \frac{y^2}{2}\right\}dydxdq - \frac{\eta}{2} \\
&= \frac{1}{2\zeta}\int_{\gamma_0-\zeta\eta}^{\gamma_0}\int_0^1\cdots\int_0^1\exp\left\{\left[g_Q(x,q) + \delta_{\alpha Q} + q\delta_{qQ}\right]^2\right\}dxdq - \frac{\eta}{2} \\
&= \frac{1}{2\zeta}\int_{\gamma_0-\zeta\eta}^{\gamma_0}\exp\left\{\xi^2\eta^{2s}\left[\frac{q-\gamma_0}{\eta} - \varphi_q\left(\frac{q-\gamma_0}{\eta}\right)\right]^2\right\}dq - \frac{\eta}{2} \\
&\leq \frac{\eta}{2}\exp\left(\xi^2\eta^{2s}\right) - \frac{\eta}{2} = \frac{\eta}{2}\left(\exp\left(\xi^2\eta^{2s}\right) - 1\right) = \frac{\xi^2}{2}\eta^{2s+1}(1 + o(1)) \leq \frac{\xi^2}{2n},
\end{aligned}$$

when n is large enough.

When ξ is small enough, say $\xi^2/2 \le \log(5/4)$, we have

$$\exp\left(\sum_{i=1}^{n}\nu_i^2\right) \le \exp\left(\frac{\xi^2}{2}\right) < \frac{5}{4}.$$

It follows from (3) that

$$\inf_{\widehat{\delta}_\alpha}\sup_{P\in\mathcal{P}(s,B)} P\left(\left|\widehat{\delta}_\alpha - \delta_\alpha(P)\right| > \frac{\epsilon}{2}n^{-\frac{s-1}{2s+1}}\right) \ge \frac{1}{2}\left(\frac{3}{2}-\frac{5}{4}\right) = \frac{1}{8} \ge C,$$

on choosing $C \le 1/8$, where $\frac{\epsilon}{2}n^{-\frac{s-1}{2s+1}}$ appears because $|\delta_\alpha(P) - \delta_\alpha(Q)| = \gamma_0\xi n^{-\frac{s-1}{2s+1}} \ge \epsilon n^{-\frac{s-1}{2s+1}}$ for a small ϵ.

When $\gamma_0 = 0$, we choose

$$g_Q(x,q) = -\xi\eta^s\varphi_q\left(\frac{q}{\eta}\right),\ \delta_{\alpha Q} = \xi\eta^s,\ \delta_{xQ} = 0, \text{ and } \delta_{qQ} = 0,$$

where φ_q is an infinitely differentiable function in q satisfying (i) $\varphi_q(v) = 0$ for $v \ge 0$, (ii) $\varphi_q(v) = 1$, for $v \le -\zeta$, and (iii) $\varphi_q(v) \in (0,1)$ for $v \in (-\zeta, 0)$, then

$$\mathbb{E}_{P_i}[\Delta_i^2] = \frac{1}{2\zeta}\int_{-\zeta\eta}^{0}\exp\left\{\xi^2\eta^{2s}\left[1-\varphi_q\left(\frac{q}{\eta}\right)\right]^2\right\}dq - \frac{\eta}{2} \le \frac{\eta}{2}\exp\left(\xi^2\eta^{2s}\right) - \frac{\eta}{2},$$

and following similar steps to those above we have $\inf_{\widehat{\delta}_\alpha}\sup_{P\in\mathcal{P}(s,B)} P\left(\left|\widehat{\delta}_\alpha - \delta_\alpha(P)\right| > \frac{\epsilon}{2}n^{-\frac{s}{2s+1}}\right) \ge C$ for some ϵ and C.

The above argument also shows that the optimal rate of convergence for δ_q is $n^{-\frac{s-1}{2s+1}}$. As for δ_x, we need only choose another pair of probabilities P and Q. To simplify notation, let $d-1=1$ so that x is only one-dimensional. Let P be the same as above, and

$$g_Q(x,q) = -\xi\eta^s\varphi_q\left(\frac{q-\gamma_0}{\eta}\right)x,\ \delta_{\alpha Q} = 0,\ \delta_{xQ} = \xi\eta^s, \text{ and } \delta_{qQ} = 0,$$

where φ_q is an infinitely differentiable function in q satisfying (i) $\varphi_q(v) = 0$ for $v \ge 0$, (ii) $\varphi_q(v) = 1$, for $v \le -\zeta$, and (iii) $\varphi_q(v) \in (0,1)$ for $v \in (-\zeta, 0)$. Then

$$\mathbb{E}_{P_i}[\Delta_i^2] = \frac{1}{2\zeta}\int_{\gamma_0-\zeta\eta}^{\gamma_0}\int_0^1\exp\left\{\xi^2\eta^{2s}x^2\left[1-\varphi_q\left(\frac{q}{\eta}\right)\right]^2\right\}dxdq - \frac{\eta}{2} \le \frac{\eta}{2}\exp\left(\xi^2\eta^{2s}\right) - \frac{\eta}{2},$$

and it follows that $\inf_{\widehat{\delta}_x}\sup_{P\in\mathcal{P}(s,B)} P\left(\left|\widehat{\delta}_x - \delta_x(P)\right| > \frac{\epsilon}{2}n^{-\frac{s}{2s+1}}\right) \ge C$ for some ϵ and C. ■

Proof of Theorem 5. Note that

$$\sqrt{n}\begin{pmatrix}\widehat{\beta}_{GMM}-\beta_0\\ \widehat{\delta}_{GMM}-\delta_0\end{pmatrix} = \left(\widehat{G}'\widehat{\Omega}^{-1}\widehat{G}\right)^{-1}\widehat{G}'\widehat{\Omega}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}z_i 1\left(q_i \le \widehat{\gamma}\right)\\ z_i' 1\left(q_i > \widehat{\gamma}\right)\end{pmatrix}\left(\varepsilon_i + \mathbf{x}_i'\delta_0 1\left(\widehat{\gamma} < q_i \le \gamma_0\right)\right).$$

By the consistency of $\widehat{\gamma}$ and Glivenko-Cantelli, $\widehat{G} \xrightarrow{p} G$. Following the proof of Theorem 3 of Caner and Hansen (2004), we can show that $\widehat{\Omega} \xrightarrow{p} \Omega$ under the moment restrictions on x, q, ε and z. We still need to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} z_i 1\left(q_i \le \widehat{\gamma}\right) \\ z_i' 1\left(q_i > \widehat{\gamma}\right)\end{pmatrix}\mathbf{x}_i'\delta_0 1\left(\widehat{\gamma} < q_i \le \gamma_0\right)$$

$$= \begin{pmatrix} \mathbf{0} \\ \frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n} z_i'\mathbf{x}_i'\delta_0 1\left(\widehat{\gamma} < q_i \le \gamma_0\right)\end{pmatrix} \xrightarrow{p} 0,$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} z_i 1\left(\gamma_0 < q_i \le \widehat{\gamma}\right) \\ z_i' 1\left(\widehat{\gamma} < q_i \le \gamma_0\right)\end{pmatrix}\varepsilon_i \xrightarrow{p} 0.$$

For these two results, consistency of $\widehat{\gamma}$ is not enough; we need $n^{1/2}\left(\widehat{\gamma}-\gamma_0\right) \xrightarrow{p} 0$. But in this case, $\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n} z_i'\mathbf{x}_i'\delta_0 1\left(\widehat{\gamma} < q_i \le \gamma_0\right) = o_p\left(\frac{1}{n}\sum\limits_{i=1}^{n} z_i'\mathbf{x}_i'\delta_0\right) = o_p(1)$, and the second result holds similarly. Given these two results, standard arguments yield the asymptotic distribution of the GMM estimator. ■

4. Proofs for the Propositions

The following four propositions are needed in the proof of Theorem 1 and Corollary 1 and hold under the conditions of that theorem.

Proposition 1 $\widehat{\gamma}-\gamma_0 = O_p(h)$.

Proof. We apply Lemma 4 of Porter and Yu (2015) to prove this result. Define $Q_n(\gamma)$ as the probability limit of $\widehat{Q}_n(\gamma)$. Lemma 1 shows that

$$\sup_{\gamma\in\Gamma}\left|\widehat{Q}_n(\gamma)-Q_n(\gamma)\right| \xrightarrow{p} 0,$$

where

$$Q_n(\gamma) = \int\left[\begin{array}{c} \int_{-1}^{0}\int K^x(u_x,x)k_-(u_q)m(x+u_xh,\gamma+u_qh)f(x+u_xh,\gamma+u_qh)du_xdu_q \\ -\int_{0}^{1}\int K^x(u_x,x)k_+(u_q)m(x+u_xh,\gamma+u_qh)f(x+u_xh,\gamma+u_qh)du_xdu_q \end{array}\right]^2 f(x)dx.$$

Let $\mathcal{N}_n = [\gamma_0-h,\gamma_0+h]$ and $\gamma_n = \arg\max\limits_{\gamma\in\Gamma} Q_n(\gamma)$, then it remains to show that $\sup\limits_{\gamma\in\Gamma\backslash\mathcal{N}_n} Q_n(\gamma) < Q_n(\gamma_n)-C$ for some positive constant C. It is easy to show that $\sup\limits_{\gamma\in\Gamma\backslash\mathcal{N}_n} Q_n(\gamma) = O(h^2)$. On the contrary, for $\gamma\in\mathcal{N}_n$, $Q_n(\gamma)$ behaves quite differently. Specifically, let $\gamma=\gamma_0+ah$, $a\in(0,1)$, then

$$Q_n(\gamma) = \int\left[\begin{array}{c} \int_{-1}^{0}\int K^x(u_x,x)k_-(u_q)g(x+u_xh,\gamma+u_qh)f(x+u_xh,\gamma+u_qh)du_xdu_q \\ +\int_{-1}^{-a}\int K^x(u_x,x)k_-(u_q)\left(1,x'+hu_x',\gamma+u_qh\right)\delta_0 f(x+u_xh,\gamma+u_qh)du_xdu_q \\ -\int_{0}^{1}\int K^x(u_x,x)k_+(u_q)g(x+u_xh,\gamma+u_qh)f(x+u_xh,\gamma+u_qh)du_xdu_q \end{array}\right]^2 f(x)dx.$$

The difference of the first and the third terms in brackets is $O(h^2)$, so the second term will dominate. From Assumption I, $\left(1,x',\gamma_0\right)\delta_0 \ne 0$ for some $x\in\mathcal{X}$, so $\int\left[\int K^x(u_x,x)\left(1,x',\gamma_0\right)\delta_0 f(x,\gamma_0)du_x\right]^2 f(x)dx > C$ for some positive constant C. Because $k_-(0)>0$ and $k_-(\cdot)\ge 0$, $\int_{-1}^{-a}k_-(u_q)du_q < 1$ and is a decreasing function of a. As a result, $Q_n(\gamma)$ is a decreasing function of a for $a\in(0,1)$ up to $O(h^2)$. Similarly, it is an increasing function of a for $a\in(-1,0)$. So $Q_n(\gamma)$ is maximized at some $\gamma_n\in\mathcal{N}_n$ such that $Q_n(\gamma_n) > \sup\limits_{\gamma\in\Gamma\backslash\mathcal{N}_n}|Q_n(\gamma)| + C/2$ for n large enough. The required result follows. ■

Proposition 2 $\widehat{\gamma}-\gamma_0=O_p(n^{-1})$.

Proof. We use the standard shelling method (see, e.g., Theorem 3.2.5 of Van der Vaart and Wellner (1996)) to prove this result.

For each n, the parameter space can be partitioned into the "shells" $S_{l,n}=\left\{\pi: 2^{l-1}<n\left|\gamma-\gamma_0\right|\leq 2^l\right\}$ with l ranging over the integers. If $n\left|\widehat{\gamma}-\gamma_0\right|$ is larger than 2^L for a given integer L, then $\widehat{\gamma}$ is in one of the shells $S_{l,n}$ with $l\geq L$. In that case the supremum of the map $\gamma\mapsto\widehat{Q}_n(\gamma)-\widehat{Q}_n(\gamma_0)$ over this shell is nonnegative by the property of $\widehat{\gamma}$. Note that

$$
\begin{aligned}
& P\left(n\left|\widehat{\gamma}-\gamma_0\right|>2^L\right) \\
\leq\ & P\left(\sup_{2^L<n|\gamma-\gamma_0|\leq nh}\left(\frac{1}{n}\sum_{i=1}^n\widehat{\Delta}_i^2(\gamma)-\frac{1}{n}\sum_{i=1}^n\widehat{\Delta}_i^2(\gamma_0)\right)\geq 0\right)+P\left(\left|\widehat{\gamma}-\gamma_0\right|\geq h\right) \\
\leq\ & \sum_{l=L}^{\log_2(nh)}P\left(\sup_{S_{l,n}}\frac{1}{n}\sum_{i=1}^n\widehat{\Delta}_i^2(\gamma)\geq\frac{1}{n}\sum_{i=1}^n\widehat{\Delta}_i^2(\gamma_0)\right)+P\left(\left|\widehat{\pi}-\pi_0\right|\geq h\right) \\
\leq\ & \sum_{l=L}^{\log_2(nh)}P\left(\sup_{S_{l,n}}\frac{1}{n}\sum_{i=1}^n\widehat{\Delta}_i^2(\gamma)1(\Delta(x_i)>0)\geq\frac{1}{n}\sum_{i=1}^n\widehat{\Delta}_i^2(\gamma_0)1(\Delta(x_i)>0)\right) \\
& +\sum_{l=L}^{\log_2(nh)}P\left(\sup_{S_{l,n}}\frac{1}{n}\sum_{i=1}^n\widehat{\Delta}_i^2(\gamma)1(\Delta(x_i)<0)\geq\frac{1}{n}\sum_{i=1}^n\widehat{\Delta}_i^2(\gamma_0)1(\Delta(x_i)<0)\right) \\
& +P\left(\left|\widehat{\pi}-\pi_0\right|\geq h\right) \\
\equiv\ & T1+T2+T3,
\end{aligned}
$$

where $\Delta(x_i)\equiv\left(1,x_i',\gamma_0\right)\delta_0$. $T3$ converges to zero by the last proposition, so we concentrate on the first two terms. $T2$ can be analyzed similar to $T1$, so we only consider $T1$ in the following discussion.

$$
\begin{aligned}
T1\ \leq\ & \sum_{l=L}^{\log_2(nh)}P\left(\sup_{S_{l,n}}\frac{1}{n}\sum_{i=1}^n\left(\widehat{\Delta}_i(\gamma)-\widehat{\Delta}_i(\gamma_0)\right)1(\Delta(x_i)>0)>0\right) \\
& +\sum_{l=L}^{\log_2(nh)}P\left(\sup_{S_{l,n}}\frac{1}{n}\sum_{i=1}^n\left(\widehat{\Delta}_i(\gamma)+\widehat{\Delta}_i(\gamma_0)\right)1(\Delta(x_i)>0)<0\right).
\end{aligned}
$$

We concentrate on the first term since the second term is easier to analyze given that $\Delta(x_i)>0$. To simplify notations, we neglect $1(\Delta(x_i)>0)$ in the following discussion.

Note that

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^n\left(\widehat{\Delta}_i(\gamma)-\widehat{\Delta}_i(\gamma_0)\right) \\
=\ & \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\left(y_jK_{h,ij}^{\gamma-}-y_jK_{h,ij}^{\gamma+}\right)-\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\left(y_jK_{h,ij}^{\gamma_0^-}-y_jK_{h,ij}^{\gamma_0^+}\right) \\
=\ & \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\left[\left(m_jK_{h,ij}^{\gamma-}-m_jK_{h,ij}^{\gamma+}\right)-\left(m_jK_{h,ij}^{\gamma_0^-}-m_jK_{h,ij}^{\gamma_0^+}\right)\right] \\
& +\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\left(e_jK_{h,ij}^{\gamma-}-e_jK_{h,ij}^{\gamma+}\right)-\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\left(e_jK_{h,ij}^{\gamma_0^-}-e_jK_{h,ij}^{\gamma_0^+}\right) \\
\equiv\ & D1+D2,
\end{aligned}
$$

where $m_j = g_j + (1, x_j', q_j)\delta_0 1(q_j \leq \gamma_0)$ with $g_j = g(x_j, q_j)$. Suppose $\gamma_0 < \gamma < \gamma_0 + h$. Then

$$\begin{aligned} D1 &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} g_j \left(K_{h,ij}^{\gamma_0^+} - K_{h,ij}^{\gamma+} \right) + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} g_j \left(K_{h,ij}^{\gamma-} - K_{h,ij}^{\gamma_0^-} \right) \\ &\quad + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} (1, x_j', q_j)\delta_0 \left(K_{h,ij}^{\gamma-} - K_{h,ij}^{\gamma_0^-} \right) 1(q_j \leq \gamma_0) \\ &\leq -C\frac{|\gamma - \gamma_0|}{h}, \end{aligned}$$

for some $C > 0$ with probability approaching 1 by calculating the mean and variance of $D1$ in its U-projection, where the first two terms contribute only $O_p(|\gamma - \gamma_0|)$, and the third term contributes to $-C\frac{|\gamma-\gamma_0|}{h}$ because for each i, $K_{h,ij}^{\gamma-}$ covers less j terms than $K_{h,ij}^{\gamma_0^-}$ given that $\gamma > \gamma_0$ and $k_{\pm}(0) > 0$. In consequence, for $\eta \leq h$,

$$P\left(\sup_{|\gamma-\gamma_0|<\eta} \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{\Delta}_i(\gamma) - \widehat{\Delta}_i(\gamma_0) \right) > 0 \right) \leq P\left(\sup_{|\gamma-\gamma_0|<\eta} D2 > C\frac{|\gamma_0 - \gamma|}{h} \right).$$

Notice that

$$\begin{aligned} D2 &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} e_j \left(K_{h,ij}^{\gamma-} - K_{h,ij}^{\gamma_0^-} \right) 1(q_j \leq \gamma_0) + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} e_j \left(K_{h,ij}^{\gamma_0^+} - K_{h,ij}^{\gamma+} \right) 1(q_j > \gamma) \\ &\quad + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} e_j \left(K_{h,ij}^{\gamma-} + K_{h,ij}^{\gamma_0^+} \right) 1\left(\gamma_0 < q_j \leq \gamma\right) \equiv D_{21} + D_{22} + D_{23}. \end{aligned}$$

By Lemma 8.4 of Newey and McFadden (1994), we can show

$$D_{21} \approx \frac{1}{n} \sum_{j=1}^{n} \frac{e_j}{h} \left[k_-\left(\frac{q_j - \gamma}{h}\right) - k_-\left(\frac{q_j - \gamma_0}{h}\right) \right] 1(q_j \leq \gamma_0),$$

so $Var\left(D_{21}\right) = O\left(\frac{1}{nh}\left(\frac{\gamma-\gamma_0}{h}\right)^2\right)$. Similarly, $Var(D_{22}) = O\left(\frac{1}{nh}\left(\frac{\gamma-\gamma_0}{h}\right)^2\right)$. As to D_{23}, we can show

$$D_{23} \approx \frac{1}{n} \sum_{j=1}^{n} \frac{e_j}{h} \left[k_-\left(\frac{q_j - \gamma}{h}\right) + k_+\left(\frac{q_j - \gamma_0}{h}\right) \right] 1\left(\gamma_0 < q_j \leq \gamma\right),$$

so $Var\left(D_{23}\right) = O\left(\frac{1}{nh}\frac{|\gamma-\gamma_0|}{h}\right)$. By the independence of U-projections of D_{21}, D_{22} and D_{23}, we have

$$Var\left(D2\right) = O\left(\frac{1}{nh}\left(\frac{\gamma-\gamma_0}{h}\right)^2 + \frac{1}{nh}\frac{|\gamma-\gamma_0|}{h}\right) = O\left(\frac{1}{nh}\frac{|\gamma-\gamma_0|}{h}\right).$$

In consequence,

$$\begin{aligned} &P\left(\sup_{|\gamma-\gamma_0|<\eta} \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{\Delta}_i(\gamma) - \widehat{\Delta}_i(\gamma_0) \right) > 0 \right) \leq C\, \mathbb{E}\left[\left(\sup_{|\gamma-\gamma_0|<\eta} D2 \right)^2 \right] \Big/ \left(\frac{|\gamma-\gamma_0|}{h} \right)^2 \\ &\leq \frac{C\,|\gamma-\gamma_0|}{nh^2} \Big/ \frac{(\gamma-\gamma_0)^2}{h^2} \leq \frac{C}{n\,|\gamma-\gamma_0|}, \end{aligned}$$

by Markov's inequality. So

$$\sum_{l=L}^{\log_2(nh)} P\left(\sup_{S_{l,n}} \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\Delta}_i(\gamma)-\widehat{\Delta}_i(\gamma_0)\right)>0\right) \leq \sum_{l\geq L}\frac{C}{n\cdot 2^l/n}=C\sum_{l\geq L}\frac{1}{2^l}\to 0$$

as $L\to\infty$, and the proof is complete. ■

Proposition 3 *For v in any compact set of $\mathbb{R}$,*

$$\begin{aligned} & nh\left(\widehat{Q}_n\left(\gamma_0+\frac{v}{n}\right)-\widehat{Q}_n(\gamma_0)\right)\Big/ 2k_+(0) \\ = & -\sum_{i=1}^{n}\overline{z}_{1i}1\left(\gamma_0-\frac{v}{n}<q_i\leq\gamma_0\right)-\sum_{i=1}^{n}\overline{z}_{2i}1\left(\gamma_0<q_i\leq\gamma_0+\frac{v}{n}\right)+o_p(1). \end{aligned}$$

Proof. We use the same notation as the last proposition and denote $\gamma_0+\frac{v}{n}$ as γ_0^v. Then

$$\begin{aligned} nh\left(\widehat{Q}_n\left(\gamma_0^v\right)-\widehat{Q}_n(\gamma_0)\right) &= \sum_{i=1}^{n}h\widehat{\Delta}_i(\gamma_0^v)^2-\sum_{i=1}^{n}h\widehat{\Delta}_i(\gamma_0)^2 \\ &= \sum_{i=1}^{n}\left(\widehat{\Delta}_i(\gamma_0^v)+\widehat{\Delta}_i(\gamma_0)\right)h\left(\widehat{\Delta}_i(\gamma_0^v)-\widehat{\Delta}_i(\gamma_0)\right). \end{aligned}$$

Following Lemma B.1 of Newey (1994), we can show that $\widehat{\Delta}_i(\gamma_0^v)\xrightarrow{p}(1,x_i',\gamma_0)\,\delta_0 f(x_i,\gamma_0)\equiv\Delta_f(x_i)=O_p(1)$ uniformly in i and v, so $\widehat{\Delta}_i(\gamma_0^v)+\widehat{\Delta}_i(\gamma_0)\xrightarrow{p}2\Delta_f(x_i)$ uniformly in i and v. We concentrate on $h\left(\widehat{\Delta}_i(\gamma_0^v)-\widehat{\Delta}_i(\gamma_0)\right)$. For simplicity, let $v>0$. Now,

$$\begin{aligned} & h\left(\widehat{\Delta}_i(\gamma_0^v)-\widehat{\Delta}_i(\gamma_0)\right) \\ = & \left(\frac{h}{n-1}\sum_{j=1,j\neq i}^{n}y_jK_{h,ij}^{\gamma_0^v-}-\frac{h}{n-1}\sum_{j=1,j\neq i}^{n}y_jK_{h,ij}^{\gamma_0^v+}\right)-\left(\frac{h}{n-1}\sum_{j=1,j\neq i}^{n}y_jK_{h,ij}^{\gamma_0-}-\frac{h}{n-1}\sum_{j=1,j\neq i}^{n}y_jK_{h,ij}^{\gamma_0+}\right) \\ = & \left[\begin{array}{c} \frac{h}{n-1}\sum_{j=1,j\neq i}^{n}\left(g(x_j,q_j)+\left(1,x_j',q_j\right)\delta_0+e_j\right)1(q_j\leq\gamma_0)K_{h,ij}^{\gamma_0^v-}-\frac{h}{n-1}\sum_{j=1,j\neq i}^{n}\left(g(x_j,q_j)+e_j\right)K_{h,ij}^{\gamma_0^v+} \\ +\frac{h}{n-1}\sum_{j=1,j\neq i}^{n}\left(g(x_j,q_j)+e_j\right)1(\gamma_0<q_j\leq\gamma_0^v)K_{h,ij}^{\gamma_0^v-} \end{array}\right] \\ & -\left[\frac{h}{n-1}\sum_{j=1,j\neq i}^{n}\left(g(x_j,q_j)+\left(1,x_j',q_j\right)\delta_0+e_j\right)K_{h,ij}^{\gamma_0-}-\frac{h}{n-1}\sum_{j=1,j\neq i}^{n}\left(g(x_j,q_j)+e_j\right)K_{h,ij}^{\gamma_0+}\right] \\ = & T_{1i}+T_{2i}+T_{3i}+T_{4i}+T_{5i}+T_{6i}, \end{aligned}$$

where

$$\begin{aligned}
T_{1i} &= -\frac{h}{n-1}\sum_{j=1,j\neq i}^{n} g(x_j,q_j)\left(K_{h,ij}^{\gamma_0^v+}-K_{h,ij}^{\gamma_0+}\right),\\
T_{2i} &= \frac{h}{n-1}\sum_{j=1,j\neq i}^{n}\left[g(x_j,q_j)+\left(1,x_j',q_j\right)\delta_0\right]\left(K_{h,ij}^{\gamma_0^v-}-K_{h,ij}^{\gamma_0-}\right),\\
T_{3i} &= -\frac{h}{n-1}\sum_{j=1,j\neq i}^{n} e_j 1(q_j>\gamma_0^v)\left(K_{h,ij}^{\gamma_0^v+}-K_{h,ij}^{\gamma_0+}\right),\\
T_{4i} &= \frac{h}{n-1}\sum_{j=1,j\neq i}^{n} e_j 1(q_j\leq\gamma_0)\left(K_{h,ij}^{\gamma_0^v-}-K_{h,ij}^{\gamma_0-}\right),\\
T_{5i} &= \frac{h}{n-1}\sum_{j=1,j\neq i}^{n} e_j 1(\gamma_0<q_j\leq\gamma_0^v)K_{h,ij}^{\gamma_0+},(*)\\
T_{6i} &= -\frac{h}{n-1}\sum_{j=1,j\neq i}^{n}\left[\left(1,x_j',q_j\right)\delta_0-e_j\right]1(\gamma_0<q_j\leq\gamma_0^v)K_{h,ij}^{\gamma_0^v-}.(*)
\end{aligned}$$

Our target is to show that

$$\sum_{i=1}^{n}\left(T_{1i}+T_{2i}+T_{3i}+T_{4i}\right)=o_p(1),$$

and

$$\begin{aligned}
\sum_{i=1}^{n}\left(T_{5i}+T_{6i}\right)\Delta_f(x_i) &= k_+(0)\sum_{i=1}^{n}\left[-\left(1,x_i',\gamma_0\right)\delta_0+2e_i\right]f(x_i)\Delta_f(x_i)1\left(\gamma_0<q_i\leq\gamma_0^v\right)+o_p(1)\\
&= -k_+(0)\sum_{i=1}^{n}\overline{z}_{2i}1\left(\gamma_0<q_i\leq\gamma_0^v\right)+o_p(1).
\end{aligned}$$

The first result is shown in Lemma 2, and the second is shown in Lemma 3. ■

Proposition 4 *On any compact set of v, $nh^d\left(\widehat{\Delta}_o\left(\gamma_0+\frac{v}{a_n}\right)-\widehat{\Delta}_o\left(\gamma_0\right)\right)\Rightarrow D_o(v)$.*

Proof. The proof proceeds by establishing convergence of the finite dimensional distributions of $R(v)\equiv nh^d\left(\widehat{\Delta}_o\left(\gamma_0^v\right)-\widehat{\Delta}_o\left(\gamma_0\right)\right)$ to those of $D_o(v)$ and then showing that $R(v)$ is tight, where $\gamma_0^v=\gamma_0+\frac{v}{a_n}$.

From the last proposition, $R(v)$ can be written as the sum of six terms:

$$R(v)=\sum_{l=1}^{6}T_l^{+}1(v>0)+\sum_{l=1}^{6}T_l^{-}1(v<0),$$

where T_l^+ is the same as T_{li} except that $\frac{h}{n-1}$ in T_{li} is changed to h^d, x_i is changed to x_o, $\sum_{j=1,j\neq i}^{n}$ changes to $\sum_{j=1}^{n}$, and $K_{h,ij}^{\gamma\pm}$ changes to $K_{h,j}^{\gamma\pm}$, and

$$
\begin{aligned}
T_1^- &= h^d \sum_{j=1}^{n} g(x_j, q_j)\left(K_{h,j}^{\gamma_0+} - K_{h,j}^{\gamma_0^v+}\right), \\
T_2^- &= h^d \sum_{j=1}^{n} \left[g(x_j, q_j) + \left(1, x_j', q_j\right)\delta_0\right]\left(K_{h,j}^{\gamma_0^v-} - K_{h,j}^{\gamma_0-}\right), \\
T_3^- &= -h^d \sum_{j=1}^{n} e_j 1(q_j > \gamma_0)\left(K_{h,j}^{\gamma_0^v+} - K_{h,j}^{\gamma_0+}\right), \\
T_4^- &= h^d \sum_{j=1}^{n} e_j 1(q_j \leq \gamma_0^v)\left(K_{h,j}^{\gamma_0^v-} - K_{h,j}^{\gamma_0-}\right), \\
T_5^- &= -h^d \sum_{j=1}^{n} e_j 1(\gamma_0^v < q_j \leq \gamma_0) K_{h,j}^{\gamma_0-}, (*) \\
T_6^- &= -h^d \sum_{j=1}^{n} \left[\left(1, x_j', q_j\right)\delta_0 + e_j\right] 1(\gamma_0^v < q_j \leq \gamma_0) K_{h,j}^{\gamma_0^v+}. (*)
\end{aligned}
$$

Lemma 4 shows that $\sum_{l=1}^4 T_l^+ + \sum_{l=1}^4 T_l^- = o_p(1)$ uniformly in v, and Lemma 5 shows that for a fixed v,

$$
T_5^+ + T_6^+ + T_5^- + T_6^- \xrightarrow{d} D_o(v).
$$

We next show the tightness of $T_5^+ + T_6^+ + T_5^- + T_6^-$. Take T_5^+ to illustrate the argument. Suppose v_1 and v_2, $0 < v_1 < v_2 < \infty$, are stopping times. Then for any $\epsilon > 0$,

$$
\begin{aligned}
& P\left(\sup_{|v_2 - v_1| < \eta} \left|T_5^+(v_2) - T_5^+(v_1)\right| > \epsilon\right) \\
\leq\ & P\left(\sum_{j=1}^{n} K\left(\frac{x_j - x_o}{h}\right) k_+\left(\frac{q_j - \gamma_0}{h}\right) |e_j| \sup_{|v_2 - v_1| < \eta} 1(\gamma_0^{v_1} < q_j \leq \gamma_0^{v_2}) > \epsilon\right) \\
\leq\ & \sum_{j=1}^{n} \mathbb{E}\left[K\left(\frac{x_j - x_o}{h}\right) k_+\left(\frac{q_j - \gamma_0}{h}\right) |e_j| \sup_{|v_2 - v_1| < \eta} 1(\gamma_0^{v_1} < q_j \leq \gamma_0^{v_2})\right] \Big/ \epsilon \\
\leq\ & C\eta/\epsilon,
\end{aligned}
$$

where the second inequality is from Markov's inequality, and C in the last inequality can take

$$
\sup_{(x,q)\in N} \mathbb{E}\left[|e|\,|x, q\right] f(x, q) \sup_{u_x, u_q} K(u_x) k_+(u_q)
$$

with N being a neighborhood of $(x_o', \gamma_0)'$. The required result now follows. ■

The following Proposition 5 is used in the proof of Theorem 2 and holds under the conditions of that theorem.

Proposition 5 $\begin{pmatrix} \sqrt{nh}\left(\widehat{\Delta}-\widehat{\Delta}^0\right) \\ \sqrt{nh}h\left(\widehat{\delta}_{xq}-\widehat{\delta}^0_{xq}\right)\end{pmatrix} \xrightarrow{p} 0.$

Proof. We need only to show

$$\frac{1}{n}\sum_{i=1}^{n}k_h\left(\frac{q_i-\widehat{\gamma}}{h}\right)\begin{pmatrix} \sqrt{nh}\left(\widehat{a}_-(x_i)-\widehat{a}_+(x_i)\right) \\ \sqrt{nh}h\left(\widehat{b}_-(x_i)-\widehat{b}_+(x_i)\right)\end{pmatrix}-\frac{1}{n}\sum_{i=1}^{n}k_h\left(\frac{q_i-\gamma_0}{h}\right)\begin{pmatrix} \sqrt{nh}\left(\widehat{a}^0_-(x_i)-\widehat{a}^0_+(x_i)\right) \\ \sqrt{nh}h\left(\widehat{b}^0_-(x_i)-\widehat{b}^0_+(x_i)\right)\end{pmatrix} \xrightarrow{p} 0,$$

and

$$\sqrt{nh}\left(\frac{1}{n}\sum_{i=1}^{n}k_h\left(\frac{q_i-\widehat{\gamma}}{h}\right)-\frac{1}{n}\sum_{i=1}^{n}k_h\left(\frac{q_i-\gamma_0}{h}\right)\right)\xrightarrow{p} 0. \tag{4}$$

It is easy to see that the first result is implied by

$$\begin{aligned}
&\sqrt{nh}\left[\left(\widehat{a}_-(x_i)-\widehat{a}_+(x_i)\right)-\left(\widehat{a}^0_-(x_i)-\widehat{a}^0_+(x_i)\right)\right]\xrightarrow{p} 0 \text{ uniformly in } x_i,\\
&\sqrt{nh}h\left[\left(\widehat{b}_-(x_i)-\widehat{b}_+(x_i)\right)-\left(\widehat{b}^0_-(x_i)-\widehat{b}^0_+(x_i)\right)\right]\xrightarrow{p} 0 \text{ uniformly in } x_i.
\end{aligned}$$

Since $\widehat{\gamma}-\gamma_0=O_p(n^{-1})$, $\widehat{\gamma}$ falls into $\left[\gamma_0-\frac{C}{n},\gamma_0+\frac{C}{n}\right]$ for some positive C with any large probability when n is large enough. So we can just prove these results by replacing $\widehat{\gamma}$ by $\gamma_0+\frac{C}{n}\equiv\gamma_0^C$. The corresponding $\widehat{a}_\pm(x_i)$ and $\widehat{b}_\pm(x_i)$ are denoted as $\widehat{a}^C_\pm(x_i)$ and $\widehat{b}^C_\pm(x_i)$. Since the results for $\widehat{a}_-(x_i)$ and $\widehat{b}_-(x_i)$ are similarly proved, we need only prove that

$$\begin{aligned}
&\sqrt{nh}\left[\widehat{a}^C_+(x_i)-\widehat{a}^0_+(x_i)\right]\xrightarrow{p} 0 \text{ uniformly in } x_i,\\
&\sqrt{nh}h\left[\widehat{b}^C_+(x_i)-\widehat{b}^0_+(x_i)\right]\xrightarrow{p} 0 \text{ uniformly in } x_i.
\end{aligned} \tag{5}$$

Without loss of generality, suppose $C>0$. Lemma 6 shows (4), and Lemma 7 shows (5). ∎

5. Proofs for the Lemmas

Lemma 1 $\sup_{\gamma\in\Gamma}\left|\widehat{Q}_n(\gamma)-Q_n(\gamma)\right|\xrightarrow{p} 0.$

Proof. Noting that $\Gamma\times\mathcal{X}$ is compact we have from Lemma B.1 of Newey (1994) that

$$\sup_{\gamma\in\Gamma,x_i\in\mathcal{X}}\left|\widehat{\Delta}_i(\gamma)-\mathbb{E}_i[\widehat{\Delta}_i(\gamma)]\right|=O_p\left(\sqrt{\ln n/nh^d}\right).$$

Given that $\sup_{\gamma\in\Gamma,x_i\in\mathcal{X}}\mathbb{E}_i\left[\widehat{\Delta}_i(\gamma)\right]=O_p(1)$,

$$\begin{aligned}
\widehat{Q}_n(\gamma) &= \frac{1}{n}\sum_{i=1}^{n}\widehat{\Delta}_i^2(\gamma)=\frac{1}{n}\sum_{i=1}^{n}\left[\mathbb{E}_i\left[\widehat{\Delta}_i(\gamma)\right]+\left(\widehat{\Delta}_i(\gamma)-\mathbb{E}_i[\widehat{\Delta}_i(\gamma)]\right)\right]^2\\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_i^2\left[\widehat{\Delta}_i(\gamma)\right]+O_p\left(\sqrt{\ln n/nh^d}\right),
\end{aligned}$$

uniformly in γ. By a Glivenko-Cantelli theorem,

$$\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_i^2\left[\widehat{\Delta}_i(\gamma)\right]-\mathbb{E}\left[\mathbb{E}_i^2\left[\widehat{\Delta}_i(\gamma)\right]\right]\right|\xrightarrow{p} 0.$$

Note that $\mathbb{E}\left[\mathbb{E}_i^2\left[\widehat{\Delta}_i(\gamma)\right]\right]=Q_n(\gamma)$, the result of interest follows. ∎

Lemma 2 $\sum_{i=1}^{n}\sum_{l=1}^{4}T_{li}=o_p(1)$ *uniformly in* v.

Proof. We take T_{4i} to illustrate and have

$$\begin{aligned}
\sum_{i=1}^{n}T_{4i} &= \frac{h}{n-1}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}e_j1(q_j\leq\gamma_0)\left(K_{h,ij}^{\gamma_0^v-}-K_{h,ij}^{\gamma_0-}\right)\\
&= \frac{h}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}e_j1(q_j\leq\gamma_0)n\left(K_{h,ij}^{\gamma_0^v-}-K_{h,ij}^{\gamma_0-}\right)\\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}e_j1(q_j\leq\gamma_0)\frac{1}{h}K_{h,ij}^{x}nh\left[k_-\left(\frac{q_j-\gamma_0^v}{h}\right)-k_-\left(\frac{q_j-\gamma_0}{h}\right)\right]\\
&= O\left(\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}e_j\frac{1(\gamma_0-h\leq q_j\leq\gamma_0)}{h}K_{h,ij}^{x}\right)\equiv O\left(\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}P_n(X_i,X_j)\right)
\end{aligned}$$

uniformly in v, where the second to last equality is from the Lipschitz continuity of $k_-(\cdot)$. By the U-statistic projection, see, e.g., Lemma 8.4 of Newey and McFadden (1994),

$$\begin{aligned}
&\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}P_n(X_i,X_j)\\
=\ &\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[P_n(X_i,X_j)\,|X_j\right]+O_p\left(\frac{1}{n}\mathbb{E}\left[P_n(X_i,X_j)^2\right]^{1/2}\right).
\end{aligned}$$

In our case, $\mathbb{E}\left[P_n(X_i,X_j)\,|X_j\right]=e_j\frac{1(\gamma_0-h\leq q_j\leq\gamma_0)}{h}\int K_{h,ij}^{x}f(x_i)dx_i=O(e_j1(\gamma_0-h\leq q_j\leq\gamma_0)/h)$, and $\mathbb{E}\left[P_n(X_i,X_j)^2\right]\approx\frac{1}{h^d}\int\sigma^2(x_i,\gamma_0)f(x_i)dx_i=O\left(\frac{1}{h^d}\right)$, so

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[P_n(X_i,X_j)\,|X_j\right] &= O_p\left(\frac{1}{nh}\sum_{j=1}e_j1(\gamma_0-h\leq q_j\leq\gamma_0)\right)=o_p(1),\\
\frac{1}{n}\mathbb{E}\left[P_n(X_i,X_j)^2\right]^{1/2} &= O\left(\frac{1}{nh^{d/2}}\right)=o(1).
\end{aligned}$$

∎

Lemma 3 $\sum_{i=1}^{n}(T_{5i}+T_{6i})\Delta_f(x_i)=-k_+(0)\sum_{i=1}^{n}\left[(1,x_i',\gamma_0)\,\delta_0-2e_i\right]1(\gamma_0<q_i\leq\gamma_0^v)f(x_i)\Delta_f(x_i)+o_p(1).$

Proof. $\sum_{i=1}^{n}T_{5i}\Delta_f(x_i)$ is a U-statistic and we write

$$\begin{aligned}
\sum_{i=1}^{n}T_{5i}\Delta_f(x_i) &= \frac{h}{n-1}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}e_j1(\gamma_0<q_j\leq\gamma_0^v)K_{h,ij}^{\gamma_0+}\Delta_f(x_i)\equiv\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}P_n(X_i,X_j)\\
&= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[P_n(X_i,X_j)\,|X_j\right]+O_p\left(\frac{1}{n}\mathbb{E}\left[P_n(X_i,X_j)^2\right]^{1/2}\right),
\end{aligned}$$

where $P_n(X_i, X_j) = nhe_j 1(\gamma_0 < q_j \leq \gamma_0^v) K_{h,ij}^{\gamma_0+} \Delta_f(x_i)$ with $X_i = (x_i', q_i, e_i)'$, and the last equality is from Lemma 8.4 of Newey and McFadden (1994). Then

$$\begin{aligned}
\mathbb{E}\left[P_n(X_i, X_j)|X_j\right] &\approx ne_j 1(\gamma_0 < q_j \leq \gamma_0^v) k_+(0) \frac{1}{h^{d-1}} \int K_h^x(x_i - x_j, x_i) f(x_i) \Delta_f(x_i) dx_i \\
&\approx ne_j 1(\gamma_0 < q_j \leq \gamma_0^v) k_+(0) \int K^x(u_x, x_j + u_x h) f(x_j + u_x h) \Delta_f(x_j + u_x h) du_x \\
&\approx ne_j 1(\gamma_0 < q_j \leq \gamma_0^v) k_+(0) f(x_j) \Delta_f(x_j),
\end{aligned}$$

and

$$\mathbb{E}\left[P_n(X_i, X_j)^2\right] = O\left(n^2 h^2 \frac{1}{n} \frac{1}{h^{2d}} h^{d-1}\right) = O\left(\frac{n}{h^{d-1}}\right),$$

so that

$$\begin{aligned}
\frac{1}{n} \sum_{j=1}^n \mathbb{E}\left[P_n(X_i, X_j)|X_j\right] &= \sum_{i=1}^n e_i 1(\gamma_0 < q_i \leq \gamma_0^v) k_+(0) f(x_i) \Delta_f(x_i), \\
\frac{1}{n} \mathbb{E}\left[P_n(X_i, X_j)^2\right]^{1/2} &= \frac{1}{n} \sqrt{\frac{n}{h^{d-1}}} = \sqrt{\frac{1}{nh^{d-1}}} = o(1).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sum_{i=1}^n T_{6i} \Delta_f(x_i) &= -\frac{h}{n-1} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left[\mathbf{x}_j' \delta_0 - e_j\right] 1(\gamma_0 < q_j \leq \gamma_0^v) K_{h,ij}^{\gamma_0^v -} \Delta_f(x_i) \\
&= -\sum_{i=1}^n \left[(1, x_i', \gamma_0)\delta_0 - e_i\right] 1(\gamma_0 < q_i \leq \gamma_0^v) k_-(0) f(x_i) \Delta_f(x_i) + o_p(1).
\end{aligned}$$

The result follows by noting that $k_-(0) = k_+(0)$. ■

Lemma 4 $\sum_{l=1}^4 T_l^+ + \sum_{l=1}^4 T_l^- = o_p(1)$ *uniformly in* v.

Proof. Take T_4^+ as an example.

$$\begin{aligned}
T_4^+ &= h^d \sum_{j=1}^n e_j 1(q_j \leq \gamma_0) \left(K_{h,j}^{\gamma_0^v -} - K_{h,j}^{\gamma_0 -}\right) \\
&= \frac{1}{nh^{d-1}} \sum_{j=1}^n e_j 1(q_j \leq \gamma_0) K^x\left(\frac{x_j - x_o}{h}, x_o\right) nh^d \left[k_-\left(\frac{q_j - \gamma_0^v}{h}\right) - k_-\left(\frac{q_j - \gamma_0}{h}\right)\right] \\
&= O\left(\frac{1}{nh^{d-1}} \sum_{j=1}^n e_j 1(\gamma_0 - h \leq q_j \leq \gamma_0) K^x\left(\frac{x_j - x_o}{h}, x_o\right)\right)
\end{aligned}$$

uniformly in v, where the last equality is from the Lipschitz continuity of $k_-(\cdot)$. Since

$$\mathbb{E}[T_4^{+2}] = O\left(\frac{1}{nh^{d-2}}\right) = o(1),$$

$T_4^+ = o_p(1)$. ■

Lemma 5 $T_5^+ + T_6^+ + T_5^- + T_6^- \xrightarrow{d} D_o(v)$.

Proof. Take $T_5^+ + T_5^-$ as an example. We use the characteristic function to find its weak limit. Define $T_5^- = \sum_{j=1}^n T_{5j}^- \equiv \sum_{j=1}^n t_{5j}^- 1(\gamma_0^{v_-} < q_j \le \gamma_0)$ and $T_5^- = \sum_{j=1}^n T_{5j}^+ = \sum_{j=1}^n t_{5j}^+ 1(\gamma_0 < q_j \le \gamma_0^{v_+})$, where $t_{5j}^- = -h^d e_j K_{h,j}^{\gamma_0 -}$, $t_{5j}^+ = h^d e_j K_{h,j}^{\gamma_0 +}$, $v_- < 0$ and $v_+ > 0$. Note that

$$
\begin{aligned}
\exp\left\{\sqrt{-1}s^- T_{5j}^-\right\} &= 1 + 1(\gamma_0^{v_-} < q_j \le \gamma_0)\left[\exp\left\{\sqrt{-1}s^- t_{5j}^-\right\} - 1\right], \\
\exp\left\{\sqrt{-1}s^+ T_{5j}^+\right\} &= 1 + 1(\gamma_0 < q_j \le \gamma_0^{v_+})\left[\exp\left\{\sqrt{-1}s^+ t_{5j}^+\right\} - 1\right].
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left\{\sqrt{-1}\left(s^- T_{5j}^- + s^+ T_{5j}^+\right)\right\}\right] = \mathbb{E}\left[\exp\left\{\sqrt{-1}s^- T_{5j}^-\right\}\right]\mathbb{E}\left[\exp\left\{\sqrt{-1}s^+ T_{5j}^+\right\}\right] \\
&\approx 1 + \mathbb{E}\left[1(\gamma_0^{v_-} < q_j \le \gamma_0)\mathbb{E}\left[\exp\left\{\sqrt{-1}s^- t_{5j}^-\right\} - 1\middle| q_j\right]\right] \\
&\quad + \mathbb{E}\left[1(\gamma_0 < q_j \le \gamma_0^{v_+})\mathbb{E}\left[\exp\left\{\sqrt{-1}s^+ t_{5j}^-\right\} - 1\middle| q_j\right]\right] \\
&\approx 1 + h^{d-1}\int_{\gamma_0^{v_-}}^{\gamma_0}\left[\int\int\left[\exp\left\{-\sqrt{-1}s^- e_j K(u_x) k_-\left(\tfrac{q_j-\gamma_0}{h}\right)\right\} - 1\right] f(e_j, x_o|q_j) de_j du_x\right] f(q_j) dq_j \\
&\quad + h^{d-1}\int_{\gamma_0}^{\gamma_0^{v_+}}\left[\int\int\left[\exp\left\{\sqrt{-1}s^+ e_j K(u_x) k_+\left(\tfrac{q_j-\gamma_0}{h}\right)\right\} - 1\right] f(e_j, x_o|q_j) de_j du_x\right] f(q_j) dq_j \\
&\approx 1 + \tfrac{v_-}{n} f_q(\gamma_0)\int\int\left[\exp\left\{-\sqrt{-1}s^- e_j K(u_x) k_-(0)\right\} - 1\right] f(e_j, x_o|q_j = \gamma_0 -) de_j du_x \\
&\quad + \tfrac{v_+}{n} f_q(\gamma_0)\int\int\left[\exp\left\{\sqrt{-1}s^+ e_j K(u_x) k_+(0)\right\} - 1\right] f(e_j, x_o|q_j = \gamma_0 +) de_j du_x \\
&= 1 + \tfrac{v_-}{n} f_q(\gamma_0) 2^{d-1}\int\int\left[\exp\left\{-\sqrt{-1}s^- e_j K(u_x) k_-(0)\right\} - 1\right]\tfrac{1(K(u_x)>0)}{\text{Vol}(K(u_x)>0)} f(e_j, x_o|q_j = \gamma_0 -) de_j du_x \\
&\quad + \tfrac{v_+}{n} f_q(\gamma_0) 2^{d-1}\int\int\left[\exp\left\{\sqrt{-1}s^+ e_j K(u_x) k_+(0)\right\} - 1\right]\tfrac{1(K(u_x)>0)}{\text{Vol}(K(u_x)>0)} f(e_j, x_o|q_j = \gamma_0 +) de_j du_x \\
&= 1 + \tfrac{v_-}{n} 2^{d-1} f_q(\gamma_0) f_{x|q}(x_o|\gamma_0)\mathbb{E}\left[\exp\left\{-\sqrt{-1}s^- e_j K(U_j^-) k_-(0)\right\} - 1\middle| x_j = x_o, q_j = \gamma_0 -\right] \\
&\quad + \tfrac{v_+}{n} 2^{d-1} f_q(\gamma_0) f_{x|q}(x_o|\gamma_0)\mathbb{E}\left[\exp\left\{\sqrt{-1}s^+ e_j K(U_j^+) k_+(0)\right\} - 1\middle| x_j = x_o, q_j = \gamma_0 +\right] \\
&= 1 + \tfrac{v_-}{n} 2^{d-1} f(x_o, \gamma_0)\mathbb{E}\left[\exp\left\{-\sqrt{-1}s^- e_j K(U_j^-) k_-(0)\right\} - 1\middle| x_j = x_o, q_j = \gamma_0 -\right] \\
&\quad + \tfrac{v_+}{n} 2^{d-1} f(x_o, \gamma_0)\mathbb{E}\left[\exp\left\{\sqrt{-1}s^+ e_j K(U_j^+) k_+(0)\right\} - 1\middle| x_j = x_o, q_j = \gamma_0 +\right],
\end{aligned}
$$

where $\text{Vol}(K(u_x) > 0) = 2^{d-1}$ is the volume of the area of u_x such that $K(u_x) > 0$, and U_j^- and U_j^+ are independent of $(e_j, x_j', q_j)'$ and follow a uniform distribution on the support of $K(\cdot)$. It follows that

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left\{\sqrt{-1}\left(s^-\sum_{j=1}^n T_{5j}^- + s^+\sum_{j=1}^n T_{5j}^+\right)\right\}\right] \\
= &\prod_{j=1}^n \mathbb{E}\left[\exp\left\{\sqrt{-1}\left(s^- T_{5j}^- + s^+ T_{5j}^+\right)\right\}\right] \\
\to &\exp\left\{v_- 2^{d-1} f(x_o, \gamma_0)\mathbb{E}\left[\exp\left\{-\sqrt{-1}s^- e K(U^-) k_-(0)\right\} - 1\middle| x = x_o, q = \gamma_0 -\right]\right. \\
&\left. + v_+ 2^{d-1} f(x_o, \gamma_0)\mathbb{E}\left[\exp\left\{\sqrt{-1}s^+ e K(U^+) k_+(0)\right\} - 1\middle| x = x_o, q = \gamma_0 +\right]\right\}.
\end{aligned}
$$

This is the characteristic function of a compound Poisson process $D_5(\cdot)$ evaluated at v_- and v_+, where

$$
D_5(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} \widetilde{z}_{1i}, \text{ if } v \le 0; \\ \sum_{i=1}^{N_2(v)} \widetilde{z}_{2i}, \text{ if } v > 0, \end{cases}
$$

is a cadlag process with $D_5(0) = 0$, $\widetilde{z}_{1i} = -e_i^- K(U_i^-) k_-(0)$, $\widetilde{z}_{2i} = e_i^+ K(U_i^+) k_-(0)$, and $\left\{e_i^-, e_i^+, U_i^-, U_i^+\right\}_{i \ge 1}$, $N_1(\cdot)$ and $N_2(\cdot)$ are defined in Corollary 1. Generalizing this argument, we get the result of interest. ■

Lemma 6 $\sqrt{nh}\left(\frac{1}{n}\sum_{i=1}^n k_h\left(\frac{q_i - \gamma_0^C}{h}\right) - \frac{1}{n}\sum_{i=1}^n k_h\left(\frac{q_i - \gamma_0}{h}\right)\right) \xrightarrow{p} 0.$

Proof.

$$\begin{aligned}
&\sqrt{nh}\left(\frac{1}{n}\sum_{i=1}^{n}k_h\left(\frac{q_i-\gamma_0^C}{h}\right)-\frac{1}{n}\sum_{i=1}^{n}k_h\left(\frac{q_i-\gamma_0}{h}\right)\right)\\
=\;&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\left[k\left(\frac{q_i-\gamma_0^C}{h}\right)-k\left(\frac{q_i-\gamma_0}{h}\right)\right]\\
\leq\;&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{C}{nh}1(\gamma_0-h\leq q_i\leq\gamma_0^C+h)=O_p\left(\frac{C}{\sqrt{nh}}\right)=o_p(1),
\end{aligned}$$

where the inequality is from the Lipschitz continuity of $k(\cdot)$. ■

Lemma 7 *Uniformly in x_i,*

$$\begin{aligned}
\sqrt{nh}\left[\widehat{a}_+^C(x_i)-\widehat{a}_+^0(x_i)\right]&\xrightarrow{p}0,\\
\sqrt{nh}h\left[\widehat{b}_+^C(x_i)-\widehat{b}_+^0(x_i)\right]&\xrightarrow{p}0.
\end{aligned}$$

Proof. Take the first result as an example. We have

$$\begin{aligned}
&\sqrt{nh}\left[\widehat{a}_+^C(x_i)-\widehat{a}_+^0(x_i)\right]\\
=\;&\sqrt{nh}\left(\mathbf{e}_1'\left(M_{iC}^+\right)^{-1}r_{iC}^+-\mathbf{e}_1'\left(M_{i0}^+\right)^{-1}r_{i0}^+\right)\\
=\;&\sqrt{nh}\left[\mathbf{e}_1'\left(M_{i0}^+\right)^{-1}\left(r_{iC}^+-r_{i0}^+\right)-\mathbf{e}_1'\left(M_{i0}^+\right)^{-1}\left(M_{iC}^+-M_{i0}^+\right)\left(M_{i0}^+\right)^{-1}r_{i0}^+\right]\\
&+\sqrt{nh}\mathbf{e}_1'\left(M_{i0}^+\right)^{-1}\left(M_{iC}^+-M_{i0}^+\right)\left(M_{i0}^+\right)^{-1}\left(M_{iC}^+-M_{i0}^+\right)\left(M_{iC}^+\right)^{-1}r_{i0}^+\\
&-\sqrt{nh}\mathbf{e}_1'\left(M_{i0}^+\right)^{-1}\left(M_{iC}^+-M_{i0}^+\right)\left(M_{iC}^+\right)^{-1}\left(r_{iC}^+-r_{i0}^+\right),
\end{aligned}$$

where M_{iC}^+ and r_{iC}^+ are similarly defined as M_i^+ and r_i^+ but with $\widehat{\gamma}$ replaced by γ_0^C, and the decomposition in the last equality is from Lemma 2 of Yu (2010). Since $\left(M_{i0}^+\right)^{-1}$, $\left(M_{i0}^+\right)^{-1}$ and r_{i0}^+ are $O_p(1)$, we need only to show that

$$\begin{aligned}
\sqrt{nh}\left(M_{iC}^+-M_{i0}^+\right)&\xrightarrow{p}0\text{ uniformly in }i,\\
\sqrt{nh}\left(r_{iC}^+-r_{i0}^+\right)&\xrightarrow{p}0\text{ uniformly in }i.
\end{aligned}$$

Take the second result as an example.

$$\begin{aligned}
&\sqrt{nh}\left(r_{iC}^+-r_{i0}^+\right)'\\
=\;&\frac{1}{\sqrt{nh}}\sum_{j=1,j\neq i}^{n}(x_j'-x_i',q_j-\gamma_0^C)^{S_p}K_h^x\left(x_j-x_i,x_i\right)k_+\left(\frac{q_j-\gamma_0^C}{h}\right)y_j\\
&-\frac{1}{\sqrt{nh}}\sum_{j=1,j\neq i}^{n}(x_j'-x_i',q_j-\gamma_0^C)^{S_p}K_h^x\left(x_j-x_i,x_i\right)k_+\left(\frac{q_j-\gamma_0}{h}\right)y_j.
\end{aligned}$$

Take the following term of $\sqrt{nh}\left(r_{iC}^{+}-r_{i0}^{+}\right)$ as an example since it is the hardest to analyze.

$$\begin{aligned}
&\left|\frac{1}{\sqrt{nh}}\sum_{j=1,j\neq i}^{n}\left(q_j-\gamma_0^C\right)^p K_h^x\left(x_j-x_i,x_i\right)k_+\left(\frac{q_j-\gamma_0^C}{h}\right)y_j\right.\\
&\left.-\frac{1}{\sqrt{nh}}\sum_{j=1,j\neq i}^{n}\left(q_j-\gamma_0\right)^p K_h^x\left(x_j-x_i,x_i\right)k_+\left(\frac{q_j-\gamma_0}{h}\right)y_j\right|\\
\leq\ &\left|\frac{1}{\sqrt{nh}}\sum_{j=1,j\neq i}^{n}K_h^x\left(x_j-x_i,x_i\right)\frac{C}{n}k_+\left(\frac{q_j-\gamma_0^C}{h}\right)y_j\right|\\
&+\left|\frac{1}{\sqrt{nh}}\sum_{j=1,j\neq i}^{n}K_h^x\left(x_j-x_i,x_i\right)\left(q_j-\gamma_0\right)^p 1(\gamma_0-h\leq q_j\leq\gamma_0^C+h)\frac{C}{nh}y_j\right|.
\end{aligned}$$

From Lemma B.1 of Newey (1994), both terms on the right side converge to their expectations uniformly in i, but it is easy to see that these expectations are $O\left(h/\sqrt{nh}\right)=o(1)$. The results of interest follow. ■

Additional References

Fan, J., 1993, Local Linear Regression Smoothers and Their Minimax Efficiency, *Annals of Statistics*, 21, 196-216.

Li, Q., 1996, On the Root-N-Consistent Semiparametric Estimation of Partially Linear Models, *Economics Letters*, 51, 277-285.

Newey, W.K., 1994, Kernel Estimation of Partial Means and a General Variance Estimator, *Econometric Theory*, 10, 233-253.

Newey, W.K. and D.L. McFadden, 1994, Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics*, Vol. 4, R.F. Eagle and D.L. McFadden, eds., New York: Elsevier Science B.V., Ch. 36, 2111-2245.

Pollard, D., 1993, Asymptotics for a Binary Choice Model, Preprint, Department of Statistics, Yale University.

Stone, C.J., 1980, Optimal Rates of Convergence for Nonparametric Estimators, *Annals of Statistics*, 8, 1348-1360.

Sun, Y.X., 2005, Adaptive Estimation of the Regression Discontinuity Model, mimio, Department of Economics, University of California, San Diego.

Van der Vaart, A.W. and J. A. Wellner, 1996, *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York: Springer.