Supplement to "Root-n Consistency of the Intercept of a Binary Response Model under Tail Restrictions"*

Lili Tan† Yichong Zhang‡

September 21, 2017

Abstract

This paper gathers the supplementary material to Tan and Zhang (2017). Section 1 establishes the connection between our tail restrictions and the information of α. Section 2 derives consistent estimators of the EV indices of ε. Section 3 collects additional simulation results. Section 4 and 5 contain the proofs of Theorems 1.1 and 2.1 in this supplementary material, respectively. Section 6 contains the proof of Theorem 4.1 in Tan and Zhang (2017). All technical lemmas are collected in Section 7.

1 Tail Conditions and the Efficiency Bound

Khan and Tamer (2010) pointed out that α is irregularly identified because its Fisher information is zero. Following Chamberlain (1986), the next theorem recalculates the Fisher information of α, using a different subpath from the one used in Khan and Tamer (2010). The new result sheds light on the direct connection between the Fisher information and the tail behaviors of V and ε.

Theorem 1.1. *If Assumptions 1(2)–1(4) hold, and for any $\delta > 0$, there exists a function $C_\delta(\cdot) : \Re \mapsto \Re$ such that*

$$\mathbb{E}(1 - C_\delta(\alpha + V))^2 \leq \delta \tag{1.1}$$

and

$$\mathbb{E}C_\delta(\varepsilon) = 0, \tag{1.2}$$

then α has zero information.

*We thank Shakeeb Khan, Federico Bugni, Arnaud Maurel, Xavier D'Haultfœuille, the editor, the co-editor, three anonymous referees, and the participants in the 2014 Econometric Society China Summer Meeting and the Duke microeconometrics lunch group for their comments. The first author acknowledges financial support from Project 71563062 supported by National Natural Science Foundation of China and Key Project of the National Social Science Foundation of China (Grant No. 13AZD082). The usual disclaimer applies.

†Yunnan University. E-mail address: tanlili@ynu.edu.cn

‡Corresponding author. Singapore Management University. E-mail address: yczhang@smu.edu.sg.

Theorem 1.1 illustrates the direct link between the information and the tail conditions. Let $C_\delta(t) = 1$ when $|t| < M_\delta$ for some $M_\delta \to \infty$ as $\delta \to 0$. If $C_\delta(t)$ was bounded from below when $|t| \geq M_\delta$, then (1.2) would not hold because $P(|\varepsilon| > M_\delta) \to 0$ as $M_\delta \to \infty$. Therefore, we need $C_\delta(t) \to -\infty$ as $|t| \to \infty$ so that the negative part of $C_\delta(t)$ at the tails ($|t| \geq M_\delta$) can cumulate and cancel the positive part of $C_\delta(t)$ in the middle ($|t| < M_\delta$). On the other hand, (1.1) implies $C_\delta^2(t)$ diverges to $-\infty$ slower than the decay rate of $f_V(t)$ as $t \to \infty$. Therefore, the existence of such C_δ implies that, heuristically,

$$\begin{aligned} \text{decay rate of } f_V(t) &> \text{divergence rate of } C_\delta^2(t) \\ &> \text{divergence rate of } C_\delta(t) = \text{ decay rate of } f_\varepsilon(t), \end{aligned}$$

i.e., V has thinner tails than ε does. This case is ruled out by our Assumption 5. Intuitively, Assumption 5 requires the tails of V to be thicker than the tails of ε.

2 Estimating the Tail Index of the Unobservable

In this section, we propose an estimator of the EV index of the unobservable ε, and show it is consistent. For simplicity, we consider the model without covariates:

$$Y = \mathbb{1}\{\alpha + V - \varepsilon \geq 0\}$$

where $V \perp\!\!\!\perp \varepsilon$. Denote the CDF of $\varepsilon - \alpha$ as $F_{\varepsilon-\alpha}$. Then, the conditional expectation of Y given $V = v$ is $F_{\varepsilon-\alpha}(v)$. By finding $q(\tau_n)$ such that $F_{\varepsilon-\alpha}(q(\tau_n)) = \tau_n$, we obtain the τ_n-th quantile $F_\varepsilon^{\leftarrow}(\tau_n)$ of ε as $q(\tau_n) + \alpha$. Note that, for any $m > 0$, $l > 0$, and some positive integer r, $F_\varepsilon^{\leftarrow}(ml^r\tau_n) - F_\varepsilon^{\leftarrow}(l^r\tau_n) = q(ml^r\tau_n) - q(l^r\tau_n)$. By replacing the CDF of $\varepsilon - \alpha$ by its nonparametric estimator $\widehat{F}_{\varepsilon-\alpha}$, we can estimate $q(\tau_n)$ by $\hat{q}(\tau_n) = \widehat{F}_{\varepsilon-\alpha}^{\leftarrow}(\tau_n)$. We estimate $\widehat{F}_{\varepsilon-\alpha}(v)$ by

$$\widehat{F}_{\varepsilon-\alpha}(v) = \Big(\frac{1}{nh}\sum_{i=1}^{n} Y_i K(\frac{V_i - v}{h})\Big) / \Big(\frac{1}{nh}\sum_{i=1}^{n} K(\frac{V_i - v}{h})\Big)$$

for $v \in \hat{S}_n$, where $\hat{S}_n$ is the same as is defined in Section 3.2 of Tan and Zhang (2017).

We focus on the left EV index λ_l. Let R and $\{\omega_r\}_{r=1}^R$ be some positive integer and a set of positive weights, respectively. Then, we estimate λ_l by

$$\hat{\lambda}_l = \sum_{r=1}^{R} \frac{-\omega_r}{\log(l)} \log\Big(\frac{\hat{q}(ml^r\tau_n) - \hat{q}(l^r\tau_n)}{\hat{q}(ml^{r-1}\tau_n) - \hat{q}(l^{r-1}\tau_n)}\Big).$$

Theorem 2.1. *If Assumptions 1–4 and 6 hold, the PDF of ε is monotone in the lower tail, and τ_n is chosen such that $\tau_n \to 0$, $\tau_n n \to \infty$, and $(L_n n^{(1+\sigma)\rho}/(nh))^{1/2}/\tau_n \to 0$ where $\rho = \rho_l(1 + \xi_l)$, then $\hat{\lambda}_l \xrightarrow{p} \lambda_l$.*

Two comments are in order. First, under Assumption 6, $(L_n n^{(1+\sigma)\rho}/(nh))^{1/2} \to 0$. Therefore, there always exists a sequence of τ_n that satisfies the condition in Theorem 2.1. Second, in order to construct a rigorous statistical test for the EV index, one has to establish the limiting distribution of $\hat{\lambda}_l$. This is left as a useful research direction.

3 Additional Simulations

In this section, we report the coverage of the true parameter α using t-statistics constructed based on our estimator $\hat{\alpha}$. The simulation designs and choices of tuning parameters are exactly the same as in Section 5 of Tan and Zhang (2017). We found that for the first five designs, in which the tail restrictions hold and there exists a regular estimator of the intercept, the coverage rate of our estimators, Ex and L, is close to the nominal rate as well as those of the untrimmed estimator (L4). In design 6, in which there does not exist any regular estimator, our estimators as well as other trimmed and untrimmed estimators all perform poorly even when the sample size is as large as 6,400. The coverage rate of the untrimmed estimator is closer to the nominal rate than those of the rest. Note that the asymptotic bias is defined as the ratio between the raw bias and the standard error of the estimator. Then the smaller the asymptotic bias, the better the coverage. When there is no trimming at all, the bias becomes the smallest, while the variance becomes the largest. Both facts help in terms of reducing the asymptotic bias, and improving the coverage. However, the root-mean-square errors in Table 6 of Tan and Zhang (2017) indicate that the untrimmed estimator achieves the best coverage by sacrificing precision, i.e., the confidence interval for the untrimmed estimator is the widest among all estimators.

Design 1

N	Ex	L	L1	L2	L3	L4
200	0.910	0.933	0.921	0.951	0.951	0.951
400	0.933	0.941	0.904	0.951	0.951	0.951
800	0.962	0.969	0.886	0.968	0.969	0.969
1,600	0.962	0.965	0.798	0.965	0.966	0.966
3,200	0.964	0.964	0.631	0.963	0.965	0.965
6,400	0.964	0.964	0.362	0.962	0.964	0.964

Table 1: 95% coverage

Design 2

N	Ex	L	L1	L2	L3	L4
200	0.863	0.886	0.856	0.906	0.907	0.907
400	0.887	0.894	0.816	0.902	0.907	0.907
800	0.913	0.923	0.662	0.924	0.928	0.928
1,600	0.903	0.910	0.510	0.905	0.912	0.912
3,200	0.937	0.947	0.209	0.920	0.951	0.951
6,400	0.930	0.932	0.021	0.889	0.930	0.932

Table 2: 95% coverage

Design 3

N	Ex	L	L1	L2	L3	L4
200	0.947	0.942	0.776	0.948	0.948	0.948
400	0.965	0.964	0.716	0.964	0.965	0.965
800	0.978	0.979	0.575	0.979	0.979	0.979
1,600	0.983	0.983	0.375	0.983	0.983	0.983
3,200	0.989	0.989	0.143	0.988	0.989	0.989
6,400	0.985	0.984	0.016	0.984	0.984	0.984

Table 3: 95% coverage

Design 4

N	Ex	L	L1	L2	L3	L4
200	0.897	0.920	0.947	0.958	0.958	0.958
400	0.939	0.950	0.955	0.971	0.971	0.971
800	0.968	0.969	0.958	0.971	0.971	0.971
1,600	0.981	0.981	0.951	0.981	0.981	0.981
3,200	0.970	0.970	0.937	0.970	0.970	0.970
6,400	0.978	0.978	0.913	0.978	0.978	0.978

Table 4: 95% coverage

Design 5

N	Ex	L	L1	L2	L3	L4
200	0.870	0.873	0.592	0.874	0.874	0.874
400	0.908	0.908	0.438	0.905	0.908	0.908
800	0.916	0.916	0.244	0.911	0.916	0.916
1,600	0.940	0.940	0.052	0.932	0.940	0.940
3,200	0.943	0.943	0.006	0.938	0.943	0.943
6,400	0.943	0.943	0.000	0.927	0.943	0.943

Table 5: 95% coverage

Design 6

N	Ex	L	L1	L2	L3	L4
200	0.745	0.805	0.791	0.900	0.900	0.900
400	0.787	0.840	0.656	0.894	0.905	0.905
800	0.731	0.788	0.371	0.856	0.891	0.891
1,600	0.719	0.781	0.129	0.780	0.881	0.883
3,200	0.776	0.827	0.010	0.742	0.886	0.900
6,400	0.749	0.810	0.000	0.521	0.854	0.885

Table 6: 95% coverage

4 Proof of Theorem 1.1

Let $g_0(t) = P(\varepsilon \leq t)$ and the subpath be

$$g_\delta(t) = g_0(t) + \delta\eta(t),$$

where η is a function from $\Re$ to $\Re$ such that $\eta(+\infty) = \eta(-\infty) = 0$. In addition, we assume $\eta(t) = o(\frac{1}{|t|})$ as $|t| \to \infty$. Since $\mathbb{E}\varepsilon = 0$, we need $\int t g_\delta'(t)dt = 0$, or equivalently,

$$\int t\eta'(t)dt = 0.$$

By integration by parts, we have

$$\int t\eta'(t)dt = -\int \eta(t)dt = 0.$$

Denote $\psi_\alpha(y, v)$ and $\psi_\delta(y, v)$ as the scores for α and δ, respectively. Khan and Tamer (2010) showed that

$$\psi_\alpha(y, v) = \frac{1}{2}\{y g_0(\alpha + v)^{-1/2} - y(1 - g_0(\alpha + v))^{-1/2}\} g_0'(\alpha + v)$$

and

$$\psi_\delta(y,v)=\frac{1}{2}\{yg_0(\alpha+v)^{-1/2}-y(1-g_0(\alpha+v))^{-1/2}\}\eta(\alpha+v).$$

Then, as Equation (5) of Chamberlain (1986), I_α is the information of α where

$$\begin{aligned}I_\alpha&=\int(\psi_\alpha(y,v)-\psi_\delta(y,v))^2d\mu(y,v)\\&=\int\frac{(g_0'(\alpha+v)-\eta(\alpha+v))^2}{g_0(\alpha+v)(1-g_0(\alpha+v))}f_V(v)dv.\end{aligned}$$

Let $\eta(t)=g_0'(t)C(t)$ for some $C(t)$ such that

$$\int C(t)g_0'(t)dt=\mathbb{E}C(\varepsilon)=0.$$

Then

$$I_\alpha=\int\frac{(g_0'(\alpha+v))^2}{g_0(\alpha+v)(1-g_0(\alpha+v))}(1-C(\alpha+v))^2f_V(v)dv.$$

Khan and Tamer (2010) has shown that $\frac{(g_0'(\alpha+v))^2}{g_0(\alpha+v)(1-g_0(\alpha+v))}$ is bounded by some constant M uniformly over $v\in\Re$. Therefore, by letting $C(t)=C_\delta(t)$ which is assumed to exist by the assumption in Theorem 1.1, we have

$$I_\alpha\le M\int(1-C_\delta(t+\alpha))^2f_V(t)dt=M\mathbb{E}(1-C_\delta(V+\alpha))^2\le M\delta.$$

Therefore, we can conclude the proof by letting $\delta\to0$.

5 Proof of Theorem 2.1

By Lemma 7.4, we know $\hat{S}_n$ is nested by S_n^+ w.p.a.1. Next, we divide the proof into three steps. First, we bound $\widehat{F}_{\varepsilon-\alpha}(v)-F_{\varepsilon-\alpha}(v)$ uniformly over $v\in S_n^+$. Second, we derive the order of magnitude of $\hat{q}(\tau_n)-q(\tau_n)$. Last, we prove the desired result.

Step 1

Denote $F_1(v)$ and $\hat{F}_1(v)$ as $F_{\varepsilon-\alpha}(v)f(v)$ and $\frac{1}{nh}\sum_{i=1}^n Y_iK(\frac{V_i-v}{h})$, respectively, where $f(\cdot)$ is the PDF of V. Then we have

$$\widehat{F}_{\varepsilon-\alpha}(v)-F_{\varepsilon-\alpha}(v)=\frac{\hat{F}_1(v)-F_1(v)}{\hat{f}(v)}+F_{\varepsilon-\alpha}(v)\frac{f(v)-\hat{f}(v)}{\hat{f}(v)}.\tag{5.1}$$

By Lemma 7.9,

$$\sup_{v\in S_n^+}\frac{|f(v)-\hat{f}(v)|}{\hat{f}(v)}=O_p\Big((L_nn^{(1+\sigma)\rho}/(nh))^{1/2}\Big).$$

For the first term of (5.1), denote $\mathcal{G}=\{YK(\frac{V-v}{h})/(hf(v)^{(1-\sigma)/2}):v\in S_n^+\}$ with envelope $CL_nn^{(1-\sigma)\rho/2}/h$.

Then, by Lemma 7.9,

$$\sup_{v\in S_n^+}\frac{|\hat F_1(v)-F_1(v)|}{\hat f(v)}\le\sup_{v\in S_n^+}\frac{|\hat F_1(v)-F_1(v)|}{f(v)(1+o_p(1))}$$
$$\le n^{(1+\sigma)\rho/2}L_n||\mathcal{P}_n-\mathcal{P}||_{\mathcal{G}}+n^{\rho}L_nh^{\nu}=O_p\Big((L_nn^{(1+\sigma)\rho}/(nh))^{1/2}\Big)$$

where the last equality is by Corollary 5.1 of Chernozhukov, Chetverikov, and Kato (2014) and the fact that

$$\sup_{g\in\mathcal{G}}\mathbb{E}g^2\le\sup_{v\in S_n^+}\mathbb{E}K^2(\frac{V-v}{h})/(h^2f(v)^{1-\sigma})\lesssim h^{-1}.$$

Therefore, denote $r_n=(L_nn^{(1+\sigma)\rho}/(nh))^{1/2}$,

$$\sup_{v\in S_n^+}|\widehat{F}_{\varepsilon-\alpha}(v)-F_{\varepsilon-\alpha}(v)|=O_p(r_n).$$

Step 2

Next, we invert $\widehat{F}_{\varepsilon-\alpha}(\cdot)$ at τ_n, i.e., finding $\hat q(\tau_n)\in\hat S_n$[1] such that

$$\widehat{F}_{\varepsilon-\alpha}(\hat q_n)\ge\tau_n\ge\widehat{F}_{\varepsilon-\alpha}(\hat q_n-r_n^2).$$

Let $q(\tau_n)=F^{\leftarrow}_{\varepsilon-\alpha}(\tau_n)$. Then

$$F_{\varepsilon-\alpha}(\hat q(\tau_n))-F_{\varepsilon-\alpha}(q(\tau_n))\le F_{\varepsilon-\alpha}(\hat q(\tau_n))-\hat F_{\varepsilon-\alpha}(\hat q(\tau_n)-r_n^2)$$
$$\le F_{\varepsilon-\alpha}(\hat q(\tau_n))-F_{\varepsilon-\alpha}(\hat q(\tau_n)-r_n^2)+\sup_{v\in S_n^+}|F_{\varepsilon-\alpha}(v)-\hat F_{\varepsilon-\alpha}(v)|$$
$$=O_p(r_n).$$

Similarly, we can obtain the inequality for the other direction that

$$F_{\varepsilon-\alpha}(\hat q(\tau_n))-F_{\varepsilon-\alpha}(q(\tau_n))\ge O_p(r_n).$$

This implies that, for $R_n=F_{\varepsilon-\alpha}(\hat q(\tau_n))-F_{\varepsilon-\alpha}(q(\tau_n))$, $R_n=O_p(r_n)$. Therefore,

$$\hat q(\tau_n)-q(\tau_n)=F^{\leftarrow}_{\varepsilon-\alpha}(\tau_n+R_n)-F^{\leftarrow}_{\varepsilon-\alpha}(\tau_n)=\tau_n\int_1^{1+R_n/\tau_n}\frac{dm}{f_{\varepsilon-\alpha}(F^{\leftarrow}_{\varepsilon-\alpha}(m\tau_n))}.$$

Multiplying $f_{\varepsilon-\alpha}(F^{\leftarrow}_{\varepsilon-\alpha}(\tau_n))$ on both sides, we have

$$f_{\varepsilon-\alpha}(F^{\leftarrow}_{\varepsilon-\alpha}(\tau_n))(\hat q(\tau_n)-q(\tau_n))=\tau_n\int_1^{1+R_n/\tau_n}\frac{f_{\varepsilon-\alpha}(F^{\leftarrow}_{\varepsilon-\alpha}(\tau_n))dm}{f_{\varepsilon-\alpha}(F^{\leftarrow}_{\varepsilon-\alpha}(m\tau_n))}\sim R_n.$$

[1] By choosing a proper τ_n, such $\hat q(\tau_n)$ exists w.p.a.1. This is because $F_{\varepsilon-\alpha}(l_n^+)\to 0$ where l_n^+ is the lower endpoint of S_n^+. So $\tau_n>F_{\varepsilon-\alpha}(l_n^+)$ suffices.

In addition, by Proposition 0.7 of Resnick (2007) (for the regularly varying case) or Lemma 4.2 of D'Haultfoeuille, Maurel, and Zhang (2016) (for the rapidly varying case), $f_{\varepsilon-\alpha}(F^{\leftarrow}_{\varepsilon-\alpha}(\tau_n)) = O(\tau_n/(F^{\leftarrow}_{\varepsilon-\alpha}(e\tau_n) - F^{\leftarrow}_{\varepsilon-\alpha}(\tau_n)))$. Therefore,

$$\hat{q}(\tau_n) - q(\tau_n) = O_p(\frac{r_n}{\tau_n}(F^{\leftarrow}_{\varepsilon-\alpha}(e\tau_n) - F^{\leftarrow}_{\varepsilon-\alpha}(\tau_n))) = O_p(\frac{r_n}{\tau_n}(q(e\tau_n) - q(\tau_n))).$$

Step 3

For any $m > 0$ and $l > 0$,

$$\begin{aligned}\frac{\hat{q}(m\tau_n) - \hat{q}(l\tau_n)}{q(e\tau_n) - q(\tau_n)} =& \frac{q(m\tau_n) - q(l\tau_n)}{q(e\tau_n) - q(\tau_n)} + \frac{\hat{q}(m\tau_n) - q(m\tau_n)}{q(e\tau_n) - q(\tau_n)} + \frac{q(l\tau_n) - \hat{q}(l\tau_n)}{q(e\tau_n) - q(\tau_n)} \\ =& \frac{q(m\tau_n) - q(l\tau_n)}{q(e\tau_n) - q(\tau_n)} + O_p(r_n/\tau_n)\end{aligned}$$

So $\hat{\lambda}_l = \sum_{r=1}^{R} \frac{-\omega_r}{\log(l)} \log\left(\frac{q(ml^r\tau_n) - q(l^r\tau_n)}{q(ml^{r-1}\tau_n) - q(l^{r-1}\tau_n)}\right) + O_p(r_n/\tau_n) = \lambda_l + o_p(1)$.

6 Proof of Tan and Zhang (2017, Theorem 4.1)

We decompose $\hat{\Phi}$ which is defined in (4.3).

$$\hat{\Phi} = \frac{1}{n}\sum_{i=1}^{n} \frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})}{f(U_i)} I_{n,i} + \frac{1}{n}\sum_{i=1}^{n} \frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})}{f(U_i)^2} I_{n,i}(f(U_i) - \tilde{f}(\hat{U}_i)) + R_{n,1} + R_{n,2} + R_{n,3}$$

in which $\hat{U}_i = V_i - S_i'\hat{\gamma}$,

$$\begin{aligned}R_{n,1} &= \frac{1}{n}\sum_{i=1}^{n} \left(\frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})}{f(U_i)}\right)(\tilde{I}_{n,i} - I_{n,i}), \\ R_{n,2} &= \frac{1}{n}\sum_{i=1}^{n} \left(\frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})}{f(U_i)}\right)\left(\frac{f(U_i) - \tilde{f}(\hat{U}_i)}{\tilde{f}(\hat{U}_i)}\right)(\tilde{I}_{n,i} - I_{n,i}), \\ R_{n,3} &= \frac{1}{n}\sum_{i=1}^{n} \left[\frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})}{f(U_i)}\right]\left[\frac{(f(U_i) - \tilde{f}(\hat{U}_i))^2}{f(U_i)\tilde{f}(\hat{U}_i)}\right] I_{n,i}.\end{aligned}$$

By Lemma 7.10,

$$\sum_{j=1}^{3} R_{n,j} = o_p(\frac{1}{\sqrt{n}}).$$

Hence,

$$\hat{\Phi} = \frac{1}{n}\sum_{i=1}^{n} \frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})}{f(U_i)} I_{n,i} + \frac{1}{n}\sum_{i=1}^{n} \frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})}{f(U_i)^2} I_{n,i}(f(U_i) - \tilde{f}(\hat{U}_i)) + o_p(\frac{1}{\sqrt{n}}). \quad (6.1)$$

Next, we further decompose $f(U_i) - \tilde{f}(\hat{U}_i)$. Define $\hat{f}(U_i) = \frac{1}{(n-1)h}\sum_{j\neq i} K(\frac{U_j - U_i}{h})$ in which h is the tuning parameter defined in Assumption 12. Then, we have

$$\begin{aligned}
f(U_i) - \tilde{f}(\hat{U}_i) &= f(U_i) - \hat{f}(U_i) + \hat{f}(U_i) - \tilde{f}(\hat{U}_i) \\
&= f(U_i) - \frac{1}{(n-1)h}\sum_{j\neq i} K(\frac{U_j - U_i}{h}) + \frac{1}{(n-1)h}\sum_{j\neq i}(K(\frac{U_j - U_i}{h}) - k(\frac{\hat{U}_j - \hat{U}_i}{h})) \\
&= f(U_i) - \frac{1}{(n-1)h}\sum_{j\neq i} K(\frac{U_j - U_i}{h}) + \frac{1}{(n-1)h^2}\sum_{j\neq i} K'(\frac{U_j - U_i}{h})(U_j - U_i - (\hat{U}_j - \hat{U}_i)) \\
&+ \frac{1}{2(n-1)h^3}\sum_{j\neq i} K''(\frac{\tilde{U}_j - \tilde{U}_i}{h})(U_j - U_i - (\hat{U}_j - \hat{U}_i))^2,
\end{aligned}$$

where $\tilde{U}_j - \tilde{U}_i$ is between $\hat{U}_j - \hat{U}_i$ and $U_j - U_i$.

Since

$$\max_{1\le i\le n} |U_i - \hat{U}_i| = \max_{1\le i\le n} |Z_i'(\hat{\gamma} - \gamma)| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

and $h = n^{-H}$ for $H < \frac{1}{4}$, we have, uniformly over $1 \le i \le n$,

$$\begin{aligned}
&\max_{i\le n} |\frac{1}{2(n-1)h^3}\sum_{j\neq i} K''(\frac{\tilde{U}_j - \tilde{U}_i}{h})(U_j - U_i - (\hat{U}_j - \hat{U}_i))^2| \\
\lesssim &\frac{1}{n^2h^3}\sum_{j\neq i}\left(|K''(\frac{\tilde{U}_j - \tilde{U}_i}{h}) - K''(\frac{U_j - U_i}{h})| + |K''(\frac{U_j - U_i}{h})|\right) \\
= &O_p(\frac{1}{n^{3/2}h^4}) + \frac{1}{n^2h^3}\sup_u \sum_{i=1}^n \left(|K''(\frac{U_j - u}{h})| - \mathbb{E}|K''(\frac{U_j - u}{h})|\right) + \frac{1}{nh^3}\sup_u \mathbb{E}|K''(\frac{U_j - u}{h})| \\
= &O_p(\frac{1}{n^{3/2}h^4} + \frac{1}{n^{3/2}h^{7/2}} + \frac{1}{nh^2}) \\
= &o_p(\frac{1}{\sqrt{n}}).
\end{aligned}$$

This implies

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^n \left[\frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})}{f(U_i)^2} I_{n,i}\left(\frac{1}{2(n-1)h^3}\sum_{j\neq i} K''(\frac{\tilde{U}_j - \tilde{U}_i}{h})(U_j - U_i - (\hat{U}_j - \hat{U}_i))^2\right)\right]\right| \\
\lesssim &\frac{1}{n}\sum_{i=1}^n \frac{|Y_i - \mathbb{1}\{V_i > 0\}|}{f(U_i)^2} I_{n,i}\, o_p(\frac{1}{\sqrt{n}}) \lesssim o_p(\frac{1}{\sqrt{n}}),
\end{aligned}$$

in which the last inequality is because $\frac{1}{n}\sum_{i=1}^n \left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(U_i)^2}\right| = O_p(1)$ by Lemma 7.12(2) or 7.13(2).

Then, we can simplify (6.1) as

$$
\begin{aligned}
\hat{\Phi} =& \frac{1}{n}\sum_{i=1}^{n}\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})}{f(U_i)}I_{n,i} + \frac{1}{n}\sum_{i=1}^{n}\Big[\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})}{f(U_i)^2}I_{n,i}\Big(f(U_i)-\frac{1}{(n-1)h}\sum_{j\neq i}K(\frac{U_j-U_i}{h})\Big)\Big] \\
&+\frac{1}{n}\sum_{i=1}^{n}\Big[\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})}{f(U_i)^2}I_{n,i}\Big(\frac{1}{(n-1)h^2}\sum_{j\neq i}K'(\frac{U_j-U_i}{h})(U_j-U_i-(\hat{U}_j-\hat{U}_i))\Big)\Big]+o_p(\frac{1}{\sqrt{n}}) \\
&=\tilde{\delta}_{n,1}+\tilde{\delta}_{n,2}+\tilde{\delta}_{n,3}+o_p(\frac{1}{\sqrt{n}}).
\end{aligned}
$$

To compute $\tilde{\delta}_{n,2}$, we follow the same steps in the proof of Theorem 3.1 which we will not repeat. The key condition for applying the same argument is that $G(u)\in\mathcal{L}^2(f(u)^{1-\sigma}du)$ where

$$G(u)=\frac{\mathbb{E}(Z_i(Y_i-\mathbb{1}\{V_i>0\})|U_i=u)}{f(u)}\mathbb{1}\{u\in S_n\}.$$

To see this, we note that

$$
\begin{aligned}
\Big|\int G^2(u)f(u)^{1-\sigma}du\Big| &= \Big|\mathbb{E}\frac{\mathbb{E}(Z_i(Y_i-\mathbb{1}\{V_i>0\})|U_i)}{f(U_i)^{2+\sigma}}\mathbb{1}\{U_i\in S_n\}\Big| \\
&\lesssim \mathbb{E}\Big|\frac{Y_i-\mathbb{1}\{V_i>0\}}{f(U_i)^{2+\sigma}}\Big|<\infty,
\end{aligned}
$$

in which the second last inequality is because Z_i is bounded and the last inequality is by Lemma 7.12(2) or 7.13(2). Then, we obtain

$$\tilde{\delta}_{n,2}=-\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{E}(Z_i(Y_i-\mathbb{1}\{V_i>0\})|U_i)}{f(U_i)}I_{n,i}+\mathbb{E}\frac{\mathbb{E}(Z_i(Y_i-\mathbb{1}\{V_i>0\})|U_i)}{f(U_i)}I_{n,i}+o_p(\frac{1}{\sqrt{n}}).$$

By Lemma 7.12(1) or 7.13(1), we have

$$\mathbb{E}\frac{\mathbb{E}(Z_i(Y_i-\mathbb{1}\{V_i>0\})|U_i)}{f(U_i)}I_{n,i}=\mathbb{E}\frac{\mathbb{E}(Z_i(Y_i-\mathbb{1}\{V_i>0\})|U_i)}{f(U_i)}+o(\frac{1}{\sqrt{n}})=\Sigma_{zx}\beta+o(\frac{1}{\sqrt{n}}).$$

So

$$\tilde{\delta}_{n,2}=-\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{E}(Z_i(Y_i-\mathbb{1}\{V_i>0\})|U_i)}{f(U_i)}I_{n,i}+\Sigma_{zx}\beta+o_p(\frac{1}{\sqrt{n}}). \tag{6.2}$$

Now let us turn to $\tilde{\delta}_{n,3}$, whose presence is due to the fact that U_i is estimated by $\hat{U}_i$. Let $W_i=(Y_i,S_i,U_i)$. Then, we write

$$\tilde{\delta}_{n,3}=\mathcal{U}_{n,3}(\hat{\gamma}-\gamma),$$

in which

$$\mathcal{U}_{n,3}=(C_n^2)^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}P_{n,3}(W_i,W_j)$$

and

$$\begin{aligned}&P_{n,3}(W_i,W_j)\\=&\frac{1}{2}\left[\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})}{f(U_i)^2}I_{n,i}\frac{1}{h^2}K'\left(\frac{U_j-U_i}{h}\right)(S_j-S_i)'+\frac{Z_j(Y_j-\mathbb{1}\{V_j>0\})}{f(U_j)^2}I_{n,j}\frac{1}{h^2}K'\left(\frac{U_i-U_j}{h}\right)(S_i-S_j)'\right].\end{aligned}$$

Later we will show

$$\mathcal{U}_{n,3}=-\mathbb{E}\left(\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})f'(U_i)(\mathbb{E}S_i-S_i)'}{f(U_i)^2}\right)+o_p(1) \tag{6.3}$$

and

$$\mathbb{E}\left|\left(\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})f'(U_i)(\mathbb{E}S_i-S_i)'}{f(U_i)^2}\right)\right|<\infty. \tag{6.4}$$

Given (6.3), (6.4), and the fact that

$$\hat{\gamma}-\gamma=\frac{1}{n}\sum_{i=1}^n\phi_i+o_p(\frac{1}{\sqrt{n}}),$$

in which $\phi_i=\Sigma_{ss}^{-1}S_iU_i$, we have

$$\tilde{\delta}_{n,3}=\mathbb{E}(\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})f'(U_i)(S_i-\mathbb{E}S_i)'}{f(U_i)^2})\frac{1}{n}\sum_{i=1}^n\phi_i+o_p(\frac{1}{\sqrt{n}}). \tag{6.5}$$

For (6.4), by Lemma 7.12(2) or 7.13(2), we have

$$\mathbb{E}\left|\left(\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})f'(U_i)(\mathbb{E}S_i-S_i)'}{f(U_i)^2}\right)\right|\lesssim\mathbb{E}\left|\frac{Y_i-\mathbb{1}\{V_i>0\}}{f(U_i)^2}\right|<\infty.$$

To show (6.3), we first show $Var(\mathcal{U}_{n,3})=o(1)$. This implies $\mathcal{U}_{n,3}=\mathbb{E}\mathcal{U}_{n,3}+o_p(1)$. To see that $Var(\mathcal{U}_{n,3})=o(1)$, we note

$$\begin{aligned}&\mathbb{E}|P_{n,3}(W_i,W_j)|^2\\\lesssim&\mathbb{E}\frac{|Z_i(Y_i-\mathbb{1}\{V_i>0\})|^2}{f(U_i)^4}I_{n,i}\frac{1}{h^4}\mathbb{E}\left(|K'(\frac{U_j-U_i}{h})|^2|S_j-S_i|^2|W_i\right)\\\lesssim&\mathbb{E}\frac{|(Y_i-\mathbb{1}\{V_i>0\})|}{f(U_i)^{3+\sigma}}I_{n,i}\frac{1}{h^3}\\\lesssim&\frac{L_n}{h^3}\mathbb{E}\left|\frac{(Y_i-\mathbb{1}\{V_i>0\})}{f(U_i)^{3+\sigma}}\right|\\\lesssim&\frac{L_n}{h^3},\end{aligned}$$

where the second inequality is because

$$\mathbb{E}\left(|K'(\frac{U_j-U_i}{h})|^2|S_j-S_i|^2|W_i\right)\lesssim h\int f(U_i+hu)[K'(u)]^2du\lesssim hf(U_i)^{1-\sigma}$$

and the last inequality is because of Lemma 7.12(2) or 7.13(2). In addition, since $H<1/3$,

$$\frac{L_n}{h^3}=o(n)$$

and thus

$$\mathbb{E}|P_{n,3}(W_i,W_j)|^2=o(n). \tag{6.6}$$

By Lemma A of (Serfling, 2009, Chapter 5), (6.6) implies the desired result

$$Var(\mathcal{U}_{n,3})\leq\frac{2}{n}\mathbb{E}|P_{n,3}(W_i,W_j)|^2=o(1).$$

Next, we compute $\mathbb{E}\mathcal{U}_{n,3}$. Since $U\perp S$, we have $\mathbb{E}(S_j|U_j)=\mathbb{E}(S_j)$ and

$$\begin{aligned}\mathbb{E}\mathcal{U}_{n,3}&=\mathbb{E}P_{n,3}(W_i,W_j)\\&=\mathbb{E}\Big\{\left[\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})}{f(U_i)^2}\right]I_{n,i}\left(\frac{\mathbb{E}S_j-S_i}{h^2}\right)\mathbb{E}\left(K'(\frac{U_j-U_i}{h})|W_i\right)\Big\}.\end{aligned}$$

First, note

$$\begin{aligned}\frac{1}{h^2}\mathbb{E}\left(K'(\frac{U_j-U_i}{h})|W_i\right)&=\frac{1}{h}\int K'(\eta)f(U_i+h\eta)d\eta\\&=-f'(U_i)+R_n(U_i)\end{aligned}$$

in which $|R_n(U_i)|\lesssim h^\nu$. Because $\sqrt{n}h^\nu\to\infty$ and

$$\mathbb{E}\left|\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})(\mathbb{E}S_j-S_i)}{f(U_i)^2}I_{n,i}\right|<\infty,$$

we have

$$\mathbb{E}\mathcal{U}_{n,3}=\mathbb{E}\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})f'(U_i)(S_i-ES_j)'}{f(U_i)^2}I_{n,i}+o(\frac{1}{\sqrt{n}}).$$

In addition, we have

$$\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})f'(U_i)(S_i-\mathbb{E}S_j)'}{f(U_i)^2}I_{n,i}\xrightarrow{p}\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})f'(U_i)(S_i-\mathbb{E}S_j)'}{f(U_i)^2}$$

and

$$\left|\frac{Z_i(Y_i-\mathbb{1}\{V_i>0\})f'(U_i)(S_i-\mathbb{E}S_j)'}{f(U_i)^2}I_{n,i}\right|\lesssim\left|\frac{Y_i-\mathbb{1}\{V_i>0\}}{f(U_i)^2}\right|.$$

Hence, by the dominated convergence theorem,

$$\mathbb{E}\frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})f'(U_i)(S_i - \mathbb{E}S_j)}{f(U_i)^2} I_{n,i} = \mathbb{E}\frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})f'(U_i)(S_i - \mathbb{E}S_j)}{f(U_i)^2} + o(1),$$

and thus,

$$\mathcal{U}_{n,3} = \mathbb{E}\mathcal{U}_{n,3} + o_p(1) = \mathbb{E}\frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})f'(U_i)(S_i - \mathbb{E}S_j)}{f(U_i)^2} + o_p(1).$$

This verifies (6.3). So (6.5) holds. Combining (6.1), (6.2) and (6.5), we have

$$\begin{aligned}\hat{\Phi} - \Sigma_{zx}\beta =& \frac{1}{n}\sum_{i=1}^n \frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})}{f(U_i)} I_{n,i} - \frac{1}{n}\sum_{i=1}^n \frac{\mathbb{E}(Z_i(Y_i - \mathbb{1}\{V_i > 0\})|U_i)}{f(U_i)} I_{n,i} \\ &+ \mathbb{E}\left(\frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})f'(U_i)(S_i - \mathbb{E}S_i)'}{f(U_i)^2}\right)\frac{1}{n}\sum_{i=1}^n \phi_i + o_p(\frac{1}{\sqrt{n}}).\end{aligned}$$

By Lemma 7.12(1) or 7.13(1) and the Markov's inequality,

$$\frac{1}{n}\sum_{i=1}^n \frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})}{f(U_i)}(1 - I_{n,i}) - \frac{1}{n}\sum_{i=1}^n \frac{\mathbb{E}(Z_i(Y_i - \mathbb{1}\{V_i > 0\})|U_i)}{f(U_i)}(1 - I_{n,i}) = o_p(\frac{1}{\sqrt{n}}).$$

Hence, $\hat{\Phi} - \Sigma_{zx}\beta = \frac{1}{n}\sum_{i=1}^n \Psi_i + o_p(\frac{1}{\sqrt{n}})$ where

$$\Psi_i = \frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})}{f(U_i)} - \frac{\mathbb{E}(Z_i(Y_i - \mathbb{1}\{V_i > 0\})|U_i)}{f(U_i)} + \mathbb{E}\left(\frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})f'(U_i)(S_i - \mathbb{E}S_i)'}{f(U_i)^2}\right)\phi_i. \tag{6.7}$$

Last, it is easy to see that $\mathbb{E}\Psi_i = 0$ and

$$\begin{aligned}\mathbb{E}|\Psi_i|^3 &\lesssim \mathbb{E}\left|\frac{Z_i(Y_i - \mathbb{1}\{V_i > 0\})}{f(U_i)}\right|^3 + \mathbb{E}|\phi_i|^3 \\ &\lesssim \mathbb{E}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(U_i)^3}\right| + \mathbb{E}|\phi_i|^3 < \infty.\end{aligned}$$

This implies the Lindeberg condition holds. Then, we have

$$\sqrt{n}(\hat{\beta} - \beta) = (\Sigma_{zx}' W \Sigma_{zx})^{-1}\Sigma_{zx}' W \frac{1}{\sqrt{n}}\sum_{i=1}^n (\Psi_i - (Z_i X_i' - \Sigma_{zx})\beta) + o_p(1) \rightsquigarrow \mathcal{N}(0, \Sigma_\beta),$$

in which

$$\Sigma_\beta = (\Sigma_{zx}' W \Sigma_{zx})^{-1}\Sigma_{zx}' W \Sigma_0 W \Sigma_{zx}(\Sigma_{zx}' W \Sigma_{zx})^{-1}$$

and

$$\Sigma_0 = \mathbb{E}(\Psi_i - (Z_i X_i' - \Sigma_{zx})\beta)(\Psi_i - (Z_i X_i' - \Sigma_{zx})\beta)'.$$

7 Technical Lemmas

7.1 Notations

Throughout this section, we denote C as a generic positive constant, whose value may differ in different contexts. L_n is a generic function of n, which is slowly varying as $n \to \infty$, i.e., $\frac{L_{kn}}{L_n} \to 1$ as $n \to \infty$ for any $k > 0$.

7.2 Lemmas for Tan and Zhang (2017, Theorem 3.1)

Note that $\hat{I}_{n,i} = \mathbb{1}\{V_i \in \hat{S}_n\}$ in which $\hat{S}_n = (\hat{l}_n, \hat{r}_n)$. The unusual feature of our trimming function is that the two endpoints are random. In order to deal with the randomness, we next propose two non-random intervals S_n^- and S_n^+ such that $S_n^- \subset \hat{S}_n \subset S_n^+$ w.p.a.1. We define

$$\rho = \max(\rho_r(1+\xi_r), \rho_l(1+\xi_l)), \quad S_n^+ = (-M_{n,l} + l_n, M_{n,r} + r_n), \text{ and } S_n^- = (M_{n,l} + l_n, -M_{n,r} + r_n),$$

in which the two positive sequences $M_{n,r}$ and $M_{n,l}$ are chosen in Lemma 7.4. Then, by letting

$$A_n = \{|\hat{r}_n - r_n| \le M_{n,r}\} \cap \{|\hat{l}_n - l_n| \le M_{n,l}\},$$

Lemma 7.4(1) shows $P(A_n) \to 1$, and on A_n, $\hat{I}_{n,i} \le \mathbb{1}\{V_i \in S_n^+\}$ and

$$|\hat{I}_{n,i} - I_{n,i}| \le \mathbb{1}\{V_i \in S_n^+\} - \mathbb{1}\{V_i \in S_n^-\} \le \min\Big(1 - \mathbb{1}\{V_i \in S_n^-\}, \mathbb{1}\{V_i \in S_n^+\}\Big). \tag{7.1}$$

We derive bounds for various terms by replacing the random interval $\hat{S}_n$ with two non-random intervals S_n^- and S_n^+. This has been done in Lemmas 7.6–7.9. These bounds are applied to derive the desired results in Lemma 7.1.

Lemma 7.1. *Under the assumptions of Theorem 3.1, we have*

$$R_{n,1} + R_{n,2} + R_{n,3} = o_p(\frac{1}{\sqrt{n}}).$$

Proof. First, on A_n, we have

$$|R_{n,1}| \le \frac{1}{n}\sum_{i=1}^{n} \left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)}\right| (1 - \mathbb{1}\{V_i \in S_n^-\}).$$

Lemma 7.6(1) or 7.7(1) shows

$$\mathbb{E}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)}\right| (1 - \mathbb{1}\{V_i \in S_n^-\}) = o(\frac{1}{\sqrt{n}}).$$

Hence, by Markov's inequality, we have

$$R_{n,1} = o_p(\frac{1}{\sqrt{n}}).$$

Similarly, on A_n, we have

$$|R_{n,2}| \leq \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)}\right|(1 - \mathbb{1}\{V_i \in S_n^-\})\left|\frac{f(V_i) - \hat{f}(V_i)}{\hat{f}(V_i)}\right|\mathbb{1}\{V_i \in S_n^+\}.$$

Lemma 7.9 shows

$$\max_{1\leq i\leq n}\left|\frac{f(V_i) - \hat{f}(V_i)}{\hat{f}(V_i)}\right|\mathbb{1}\{V_i \in S_n^+\} \leq \max_{v\in S_n^+}\left|\frac{f(v) - \hat{f}(v)}{f(v)(1 + o_p(1))}\right| = O_p\left((\frac{\log(n)n^{\rho(1+\sigma)}}{nh})^{1/2}\right) = o_p(1).$$

Hence,

$$|R_{n,2}| \leq |\frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)}\right|(1 - \mathbb{1}\{V_i \in S_n^-\})|o_p(1) = o_p(\frac{1}{\sqrt{n}}).$$

For $R_{n,3}$, we have, by Lemma 7.9,

$$\sqrt{n}|R_{n,3}| \leq \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)^{2+\sigma}}\right|\left[\sqrt{n}\left|\frac{(f(V_i) - \hat{f}(V_i))^2}{f(V_i)^{1-\sigma}(1 + o_p(1))}I_{n,i}\right|\right].$$

By Lemma 7.7(2),

$$\frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)^{2+\sigma}}\right| = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)}\right|^{2+\sigma} = O_p(1).$$

In addition, by Lemma 7.9 and the fact that $H < 1/2$,

$$\max_{1\leq i\leq n}\left|\sqrt{n}\frac{(f(V_i) - \hat{f}(V_i))^2}{f(V_i)^{1-\sigma}}I_{n,i}\right| = O_p(n^{-1/2}h^{-1}L_n) = o_p(1).$$

This implies $R_{n,3} = o_p(\frac{1}{\sqrt{n}})$. □

Lemma 7.2.

$$\tilde{\delta}_{n,2} = -\frac{1}{n}\sum_{i=1}^{n}\frac{P(V_i) - \mathbb{1}\{V_i > 0\}}{f(V_i)}I_{n,i} + o_p(\frac{1}{\sqrt{n}}).$$

Proof. We first claim

$$\mathbb{E}(P_n^2(W_i, W_j)) = o(n). \tag{7.2}$$

To see this, first recall that, by Lemma 7.4(2), on $V_i \in S_n \subset S_n^+$, $f(V_i) \geq Cn^{-\rho}L_n$. Second, by

Lemma 7.6(2) or 7.7(2),

$$\mathbb{E}\left|\frac{Y_i - \mathbb{1}\{V > 0\}}{f(V)}\right|^{2+\sigma} = O(1).$$

Last, by Assumption 2(3), $f(v + hu) \le Cf(v)^{1-\sigma}$. Combining the above three facts, we have

$$\begin{aligned}
&\mathbb{E}\left[\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)^2}(f(V_i) - \frac{1}{h}K(\frac{V_i - V_j}{h}))I_{n,i}\right]^2 \\
=&\mathbb{E}\frac{|Y_i - \mathbb{1}\{V_i > 0\}|^2}{f(V_i)^4}\left(f(V_i)^2 - 2f(V_i)\int K(u)f(V_i + hu)du + \int \frac{K^2(u)}{h}f(V_i + hu)du\right) I_{n,i} \\
\le& C\mathbb{E}\frac{(Y_i - \mathbb{1}\{V_i > 0\})^2}{f(V_i)^{3+\sigma}}\frac{1}{h}I_{n,i} \\
=& C\mathbb{E}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)}\right|^{2+\sigma} n^{\rho}\frac{L_n}{h} \\
=& O(\frac{nL_n}{n^{1-\rho}h}) = o(n).
\end{aligned}$$

Define

$$\hat{U}_n = \theta_n + \frac{2}{n}\sum_{i=1}^{n}(r_n(W_i) - \theta_n),$$

in which $r_n(W_i) = \mathbb{E}(P_n(W_i, W_j)|W_i)$, $\theta_n = \mathbb{E}r_n(W_i)$, and $W_i = (Y_i, V_i)$. Because (7.2) holds, we can apply Lemma 3.1 in Powell, Stock, and Stoker (1989) and obtain

$$\sqrt{n}(\tilde{\delta}_{2,n} - \hat{U}_n) = o_p(1). \tag{7.3}$$

Next, we compute $r_n(W_i)$. Note that

$$r_n(W_i) = \frac{1}{2}(r_{n,1}(W_i) + r_{n,2}(W_i)),$$

in which

$$r_{n,1}(W_i) = \mathbb{E}\left(\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)^2}(f(V_i) - \frac{1}{h}K(\frac{V_i - V_j}{h}))I_{n,i}|W_i\right)$$

and

$$r_{n,2}(W_i) = \mathbb{E}\left(\frac{Y_j - \mathbb{1}\{V_j > 0\}}{f(V_j)^2}(f(V_j) - \frac{1}{h}K(\frac{V_i - V_j}{h}))I_{n,j}|W_j\right).$$

By the mean-value theorem and Assumption 2(2), we have

$$r_{n,1}(W_i) = \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)^2}I_{n,i}f(\tilde{V}_i)^{(\nu)}h^{v}.$$

Thus, by Lemma 7.6(2) or Lemma 7.7(2),

$$\sqrt{n}|\theta_n| \le \sqrt{n}\mathbb{E}|r_{n,1}(W_i)| \le C\sqrt{n}h^{v}\mathbb{E}(\frac{Y - \mathbb{1}\{V > 0\}}{f(V)})^2 I_{n,i} = O((nh^{2v})^{\frac{1}{2}}) = o(1), \tag{7.4}$$

where the last equality is because $H > \frac{1+\rho_r(1+\xi_r)}{1+2\nu} > \frac{1+\rho_r}{1+2\nu} > \frac{1}{2\nu}$.

Hence,

$$\frac{1}{n}\sum_{i=1}^{n}(r_{n,1}(W_i) - \theta_n) = o_p(\frac{1}{\sqrt{n}}). \tag{7.5}$$

Now, we define $P(v) = \mathbb{E}(Y_i|V_i = v)$. For $r_{n,2}(W_i)$, we have

$$r_{n,2}(W_i) = \mathbb{E}\frac{Y_j - \mathbb{1}\{V_j > 0\}}{f(V_j)}I_{n,j} - G * K_h(V_i) = \mathbb{E}\frac{Y_j - \mathbb{1}\{V_j > 0\}}{f(V_j)}I_{n,j} - G(V_i) - T_{n,i},$$

in which $*$ means convolution,[2]

$$G(v) = \frac{P(v) - \mathbb{1}\{v > 0\}}{f(v)}\mathbb{1}\{v \in S_n\},$$

$$K_h(v) = \frac{1}{h}K(\frac{v}{h}),$$

and

$$T_{n,i} = G * K_h(V_i) - G(V_i).$$

Then,

$$r_{n,2}(W_i) - \theta_n = -\left[\frac{P(V_i) - \mathbb{1}\{V_i > 0\}}{f(V_i)}\mathbb{1}\{V_i \in S_n\} - \mathbb{E}\left(\frac{P(V_i) - \mathbb{1}\{V_i > 0\}}{f(V_i)}\mathbb{1}\{V_i \in S_n\}\right)\right] - (T_{n,i} - \mathbb{E}(T_{n,i})).$$

If $\mathbb{E}T_{n,i}^2 = o(1)$, then we have

$$\frac{1}{n}\sum_{i=1}^{n}(r_{n,2}(W_i) - \theta_n) = -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{P(V_i) - \mathbb{1}\{V_i > 0\}}{f(V_i)}I_{n,i} - \mathbb{E}\left(\frac{P(V_i) - \mathbb{1}\{V_i > 0\}}{f(V_i)}I_{n,i}\right)\right] + o_p(\frac{1}{\sqrt{n}}). \tag{7.6}$$

Next, we compute $\mathbb{E}T_{n,i}^2$. By Minkowski's inequality, we have

$$\begin{aligned}\mathbb{E}T_{n,i}^2 =& \int (\int (G(v - hu) - G(v))K(u)du)^2 f(v)dv \\ \leq& (\int (\int (G(v - hu) - G(v))^2 f(v)dv)^{\frac{1}{2}} K(u)du)^2.\end{aligned} \tag{7.7}$$

By Lemma 7.8, we have, for each fixed u,

$$\int (G(v - hu) - G(v))^2 f(v)dv \to 0$$

[2] For two generic functions f and g, $f * g(t) = \int_{u\in\Re} f(t - u)g(u)du$.

as $h \to 0$ and

$$\int (G(v-hu)-G(v))^2 f(v)dv \leq \int (G^2(v-hu)+G^2(v))f(v)dv \leq C,$$

where C is a positive constant independent of u. Therefore, by the dominated convergence theorem, the RHS of (7.7) vanishes as $h \to 0$; that is,

$$\left[\int \left(\int (G(v-hu)-G(v))^2 f(v)dv\right)^{\frac{1}{2}} K(u)du\right]^2 \to 0.$$

This concludes the claim $\mathbb{E}T_{n,i}^2 = o(1)$.

□

Lemma 7.3. *Denote*

$$T_{n,1} = \frac{1}{n}\sum_{i=1}^n 2\left(\frac{Y_i - \mathbb{1}\{V_i>0\}}{\hat{f}(V_i)}\right)^2 (\hat{I}_{n,i}-I_{n,i})I_{n,i}, \quad T_{n,2} = \frac{1}{n}\sum_{i=1}^n \left(\frac{Y_i - \mathbb{1}\{V_i>0\}}{\hat{f}(V_i)}\right)^2 \left(\frac{f(V_i)-\hat{f}(V_i)}{\hat{f}(V_i)}\right)\hat{I}_{n,i},$$

$$T_{n,3} = \frac{1}{n}\sum_{i=1}^n \left(\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(V_i)}\right)^2 (\hat{I}_{n,i}-I_{n,i})^2, \quad \textit{and} \quad T_{n,4} = \frac{1}{n}\sum_{i=1}^n \left(\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(V_i)}\right)^2 \left(\frac{f(V_i)-\hat{f}(V_i)}{\hat{f}(V_i)}\right)^2 \hat{I}_{n,i}.$$

Then, $T_{n,j} = o_p(1)$*,* $j = 1, \cdots, 4$*.*

Proof. Since

$$\frac{1}{n}\sum_{i=1}^n \left(\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(V_i)}\right)^2 I_{n,i} \xrightarrow{p} \mathbb{E}\left(\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(V_i)}\right)^2 < \infty,$$

we have

$$\mathbb{E}|T_{n,1}| \to 0 \text{ and } T_{n,1} = o_p(1).$$

Second,

$$|T_{n,2}| \leq \frac{1}{n}\sum_{i=1}^n \left(\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(V_i)}\right)^2 o_p(1) = o_p(1).$$

Similarly, $T_{n,3}$ and $T_{n,4}$ are asymptotically negligible, i.e.,

$$T_{n,3} = o_p(1), \ T_{n,4} = o_p(1).$$

Therefore, we have

$$\frac{1}{n}\sum_{i=1}^n Z_{n,i}^2 \xrightarrow{p} \mathbb{E}\left(\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(V_i)}\right)^2.$$

In addition, $\frac{1}{n}\sum_{i=1}^{n} Z_{n,i} \xrightarrow{p} \alpha$. So we have shown

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} Z_{n,i}^2 - \left[\frac{1}{n}\sum_{i=1}^{n} Z_{n,i}\right]^2 \xrightarrow{p} \Sigma.$$

□

Lemma 7.4. *For $l_n, r_n, \hat{l}_n$ and $\hat{r}_n$ defined in Section 3.2, if Assumption 2 holds, then there exist two positive sequences $M_{n,r}$ and $M_{n,l}$ such that for $S_n^+ = (-M_{n,l} + l_n, M_{n,r} + r_n)$, $S_n^- = (M_{n,l} + l_n, -M_{n,r} + r_n)$, and*

$$A_n = \{|\hat{r}_n - r_n| \leq M_{n,r}\} \cap \{|\hat{l}_n - l_n| \leq M_{n,l}\},$$

we have (1) $P(A_n) \to 1$ and on A_n, $S_n^- \subset \hat{S}_n \subset S_n^+$, and (2) for any $v \in S_n^+$, $f(v) \geq cn^{-\rho} L_n$.

Lemma 7.4(1) states that the feasible random interval $\hat{S}_n$ nests and is nested by two deterministic intervals S_n^- and S_n^+, respectively, w.p.a.1. Lemma 7.4(2) shows that, on interval S_n^+, the decay rate of the density $f(v)$ is controlled by tuning parameters ρ_r and ρ_l.

We first introduce a lemma on the asymptotic properties of extremal quantile estimators $\hat{l}_n$ and $\hat{r}_n$, derived by Dekkers and De Haan (1989). Recall that F is the CDF of the special regressor V. Let $U = (\frac{1}{1-F})^{\leftarrow}$, $V(t) = U(e^t)$. Then, we have $V'(t) = U'(e^t)e^t$. By Assumption 4, $U'(t) \in RV_{\xi_r - 1}(\infty)$. Denote $E_{(1)}^{(n)} \leq E_{(2)}^{(n)} \leq \cdots E_{(n)}^{(n)}$ as the ascending order statistics of $E_1, E_2, \cdots E_n$, where $E_1, E_2, \cdots E_n$ are i.i.d. standard exponential random variables.

Lemma 7.5. *For $m(n) \to \infty$ and $m = \lfloor m(n) \rfloor$, we have*
(1) $\{V_{(n-i+1)}^{(n)}, i = 1, 2, \cdots n\} \stackrel{d}{=} \{V(E_{(n-i+1)}^{(n)}), i = 1, 2, \cdots n\}$.
(2) If $\frac{m(n)}{n} \to 0$, $m(n) \to \infty$, Let $\hat{r}_n = V_{(n-m+1)}^n$, then

$$\sqrt{2m}\left(\frac{\hat{r}_n - r_n}{V_{(n-m+1)}^n - V_{(n-2m+1)}^n}\right) \rightsquigarrow \mathcal{N}(0, \sigma^2(\xi_r))$$

where $\sigma(\xi_r)$ is a constant that only depends on ξ_r.
(3)

$$\sqrt{2m}\left(\frac{V(E_{(n-m+1)}^{(n)}) - V(E_{(n-2m+1)}^{(n)})}{2^{\xi_r} V'(E_{(n-2m+1)}^{(n)})} - \frac{1 - 2^{-\xi_r}}{\xi_r}\right) \rightsquigarrow \mathcal{N}(0, 1).$$

(4)

$$\frac{V'(E_{(n-2m+1)}^{(n)})}{V'(\log(\frac{n}{2m}))} - 1 = \frac{\xi_r(V(E_{(n-2m+1)}^{(n)}) - V(\log(\frac{n}{2m})))}{V'(\log(\frac{n}{2m}))} + o_p(\frac{1}{\sqrt{2m}}) = O_p(\frac{1}{\sqrt{2m}}) = o_p(1).$$

In our definition, for the right tail, $m_r = n^{1-\rho_r}$ for some $0 < \rho_r < 1$. The convergence rate for $\hat{r}_n$ is

$$\frac{\sqrt{2m_r}}{V^n_{(n-m_r+1)} - V^n_{(n-2m_r+1)}} \sim C\frac{\sqrt{2m_r}}{V'(E^{(n)}_{(n-2m_r+1)})} \sim C\frac{\sqrt{2m_r}}{V'(\log(\frac{n}{2m_r}))},$$

where the first and second equivalences are by Lemmas 7.5(3) and 7.5(4), respectively. Similarly, for the left tail, the convergence rate for $\hat{l}_n$ is $\frac{\sqrt{2m_l}}{V'(\log(2m_l))}$ where $m_l = n^{1-\rho_l}$.

Proof of Lemma 7.4. We only show the results for the right tail. The argument for the left tail is symmetric.

For (1), by Lemma 7.5, we have $\hat{r}_n - r_n = O_p(\frac{\sqrt{2m_r}}{V'(\log(\frac{n}{2m_r}))})$ where $m_r = n^{1-\rho_r}$. Let M_n be some deterministic sequence such that $M_n \to \infty$. We define $M_{n,r} = \frac{M_n V'(\log(\frac{n}{2m}))}{\sqrt{2m}}$. Then we have $|\hat{r}_n - r_n| = o_p(M_{n,r})$.

For (2), because of the monotonicity of f, for $z \in S_n^+$, we have

$$\frac{f(z)}{f(r_n)} \geq \frac{f(r_n + M_{n,r})}{f(r_n)} = \frac{f((1-F)^{\leftarrow}(1 - F(r_n + M_{n,r})))}{f((1-F)^{\leftarrow}(1 - F(r_n)))}.$$

By Proposition 0.7 of Resnick (2007) (for the regularly varying case) or Lemma 4.2 of D'Haultfoeuille et al. (2016) (for the rapidly varying case), Assumption 4 implies $f((1-F)^{\leftarrow}) \in RV_{\xi_r+1}(0)$. In addition, since V has unbounded support, $\xi_r \geq 0$ and $\xi_r + 1 \geq 1$. If

$$\frac{1 - F(r_n + M_{n,r})}{1 - F(r_n)} \to 1, \tag{7.8}$$

then $\frac{f(r_n+M_{n,r})}{f(r_n)} \to 1$ and thus $f(z) > Cf(r_n) = Cf((1-F)^{\leftarrow}(n^{-\rho_r})) = Cn^{-\rho_r(\xi_r+1)}L_n$. Therefore we only need to verify (7.8). The proof is divided into two cases.

Case (1): $\xi_r > 0$, $1 - F$ is regularly varying. We only have to prove $\frac{r_n+M_{n,r}}{r_n} \to 1$ or equivalently, $\frac{M_{n,r}}{r_n} \to 0$. Note by the choice of $M_{n,r}$, we have

$$\begin{aligned}\frac{M_{n,r}}{r_n} &= M_n \frac{U'(n^{\rho_r})n^{\rho_r}}{n^{\frac{1-\rho_r}{2}}(1-F)^{\leftarrow}(n^{-\rho_r})} \\ &= \frac{M_n}{n^{\frac{1-\rho_r}{2}} L_n}.\end{aligned}$$

Since $\rho_r < 1$, the denominator diverges to infinity. Thus, there exists a sequence M_n such that $M_n \to \infty$ and $\frac{M_{n,r}}{r_n} \to 0$.

Case (2): $\xi_r = 0$, then by Assumption 4, the right tail of F is in the attraction domain of type 1 EV distribution. By Proposition 0.10 of Resnick (2007), it implies $\frac{1}{1-F}$ is Γ-varying with an auxiliary function $f_0(t) = \frac{1-F(t)}{f(t)}$.[3] In addition, we can write $\frac{1-F(r_n+M_{n,r})}{1-F(r_n)} = \frac{1-F(r_n+\frac{M_{n,r}}{f_0(r_n)}f_0(r_n))}{1-F(r_n)}$. If $\frac{M_{n,r}}{f_0(r_n)} \to 0$, then by the definition of Γ-varying function (Equation 0.47 in Resnick (2007)), (7.8) holds. Since $f(1-F)^{\leftarrow} \in RV_1(0)$,

$$f_0(r_n) = f_0((1-F)^{\leftarrow})(n^{-\rho_r}) \sim \frac{(1-F)(1-F)^{\leftarrow}(n^{-\rho_r})}{f(1-F)^{\leftarrow}(n^{-\rho_r})} \sim \frac{1}{L_n},$$

$f_0(r_n)$ is slowly varying. Therefore,

$$\frac{M_{n,r}}{f_0(r_n)} \sim \frac{M_n L_n}{n^{\frac{1-\rho_r}{2}}}. \tag{7.9}$$

Furthermore, $\rho_r < 1$, we have $\frac{L_n}{n^{\frac{1-\rho_r}{2}}} \to 0$. Thus there exists a sequence of M_n, e.g., $M_n = n^{\frac{1-\rho_r}{4}} \to \infty$, such that $\frac{M_{n,r}}{f_0(r_n)} \to 0$. This is the desired result.

□

The next two lemmas verify the high level assumptions that ensure the $\sqrt{n}$-consistency of $\hat{\alpha}$: (1) the asymptotic bias vanishes faster than $\sqrt{n}$ and (2) the Lindeberg condition.

Lemma 7.6. *If Assumption 5(1) or (3) holds and the tuning parameters h and ρ_r are chosen as in Assumption 6, then the following statements hold:*
(1) $\sqrt{n}\mathbb{E}|\frac{Y_i-\mathbb{1}\{V_i>0\}}{f(V_i)^p}|(1-\mathbb{1}\{V_i \in S_n^-\}) \to 0$ *for any* $p>0$;
(2) $\mathbb{E}|\frac{Y_i-\mathbb{1}\{V_i>0\}}{f(V_i)}|^p < \infty$ *for any* $p>0$.

Proof. We only prove the results for the right tail. The proof for the left tail is symmetric. We note that, because the special regressor V is supported on the whole real line, V can only have type 1 or type 2 tails, i.e., its EV indices are nonnegative. If V has type 1 tails, then $(1-F)^{\leftarrow}$ is slowly varying, while if it has type 2 tails, then $(1-F)^{\leftarrow}$ is regularly varying.

For part (1), under Assumption 5(1), $(1-F)^{\leftarrow}(z)$ is a regularly varying function and for any $q_r>0$, $(1-F_\varepsilon)^{\leftarrow}(z^{q_r})$ is slowly varying as $z \to 0$. Therefore, we have $\frac{(1-F)^{\leftarrow}(z)}{(1-F_\varepsilon)^{\leftarrow}(z^{q_r})} \to \infty$ as $z \to 0$. Under Assumption 5(3), for any $q_r > 0$, and as $z \to 0$,

$$\begin{aligned}\frac{(1-F)^{\leftarrow}(z)}{(1-F_\varepsilon)^{\leftarrow}(z^{q_r})} &= \frac{T_r^{\leftarrow}(-\log(z))}{D_r^{\leftarrow}(-q_r\log(z))} \\ &= (-\log(z))^{\frac{1}{d_{1,r}}-\frac{1}{d_{2,r}}} L_n(-\log(z)) \to \infty.\end{aligned}$$

In addition, $(1-F_\varepsilon)^{\leftarrow}(z^{q_r}) \to \infty$ as $z \to 0$. Therefore, under either Assumption 5(1) or Assumption 5(3), for any $q_r>0$ and any constant C independent of n, when z is sufficiently close to zero,

$$\frac{C+(1-F)^{\leftarrow}(z)}{(1-F_\varepsilon)^{\leftarrow}(z^{q_r})} \geq 1. \tag{7.10}$$

[3] A non-decreasing function U is Γ-varying if U is defined on an interval (x_l, x_0), $\lim_{x\uparrow x_0} U(x) = \infty$ and there exists a positive function f_0 defined on (x_l, x_0) such that for all x, $\lim_{t\to x_0}\frac{U(t+xf_0(t))}{U(t)} = \exp(x)$.

We define

$$B_{n,1} = \int_{r_n - m_{n,r}}^{\infty} \frac{1 - F_\varepsilon(\alpha + v)}{f(v)^{p-1}} dv$$

and

$$B_{n,2} = \int_{-\infty}^{-r_{n,l} + m_{n,l}} \frac{F_\varepsilon(\alpha + v)}{f(v)^{p-1}} dv.$$

Then we have $\mathbb{E}|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)^p}|(1 - \mathbb{1}\{V_i \in S_n^-\}) = B_{n,1} + B_{n,2}$. Similar to the proof of Lemma 7.4, we have $\frac{1 - F(r_n - M_{n,r})}{1 - F(r_n)} \to 1$. Therefore, there exists a constant C such that, for n sufficiently large,

$$(1 - F(r_n - M_{n,r})) \leq C(1 - F(r_n)) = C((1 - F)((1 - F)^{\leftarrow}(n^{-\rho_r}))) = Cn^{-\rho_r} L_n. \tag{7.11}$$

(7.11) implies $\{z : (1 - F)^{\leftarrow}(z) \geq r_n - M_{n,r}\} \subset \{z : z \leq Cn^{-\rho_r} L_n\}$. Let $v = (1 - F)^{\leftarrow}(z)$. By the change of variables, we have, for an arbitrary $q_r > 0$,

$$\begin{aligned}
\sqrt{n} B_{n,1} &\leq \sqrt{n} \int_0^{Cn^{-\rho_r} L_n} \frac{(1 - F_\varepsilon)(\alpha + (1 - F)^{\leftarrow}(z))}{f((1 - F)^{\leftarrow}(z))^p} dz \\
&\leq \sqrt{n} \int_0^{Cn^{-\rho_r} L_n} \frac{z^{q_r}}{f((1 - F)^{\leftarrow}(z))^p} dz \\
&\leq \sqrt{n} \int_0^{Cn^{-\rho_r} L_n} z^{q_r - p(\xi_r + 1)} L(z) dz \\
&= O(n^{\frac{1}{2} - \rho_r(q_r - p(\xi_r + 1) + 1)} L_n),
\end{aligned}$$

in which $L(z)$ is a slowly varying function. The second and third inequalities in the above display are by (7.10) and Assumption 4, respectively. Since q_r is arbitrary, we can choose it to be sufficiently large so that $\frac{1}{2} - \rho_r(q_r - p(\xi_r + 1) + 1) < 0$. This means $\sqrt{n} B_{n,1} = o(1)$. Similarly, we can show that $\sqrt{n} B_{n,2} = o(1)$. This concludes lemma 7.6(1).

For part (2), we note

$$\begin{aligned}
\mathbb{E}|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)}|^p &\leq C \int \frac{F_\varepsilon(\alpha + v)(1 - \mathbb{1}\{v > 0\}) + \mathbb{1}\{v > 0\}(1 - F_\varepsilon(\alpha + v))}{f(v)^{p-1}} dv \\
&= C(\int_0^{\infty} \frac{1 - F_\varepsilon(\alpha + v)}{f(v)^{p-1}} dv + \int_{-\infty}^0 \frac{F_\varepsilon(\alpha + v)}{f(v)^{p-1}} dv).
\end{aligned}$$

We now only consider the integrability in the right tail. Let $z = (1 - F)^{\leftarrow}(v)$. By the change of variables, we have

$$\begin{aligned}
\int_0^{\infty} \frac{1 - F_\varepsilon(\alpha + v)}{f(v)^{p-1}} dv &= \int_0^c \frac{1 - F_\varepsilon(\alpha + v)}{f(v)^{p-1}} dv + \int_c^{\infty} \frac{1 - F_\varepsilon(\alpha + v)}{f(v)^{p-1}} dv \\
&\leq C + \int_0^c \frac{(1 - F_\varepsilon)(\alpha + (1 - F)^{\leftarrow}(z))}{(f(1 - F)^{\leftarrow}(z))^p} dz \\
&\leq C + \int_0^c z^{q_r - p(\xi_r + 1)} L(z) dz.
\end{aligned}$$

Since q_r is arbitrary, we can choose q_r sufficiently large so that $q_r - p(\xi_r + 1) + 1 > 0$. This implies the integral is finite. $\square$

Lemma 7.7. *If Assumption 5(2) holds and the tuning parameters h and ρ_r are chosen as in Assumption 6, then the following statements hold:*
(1) $\sqrt{n}\mathbb{E}|\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(V_i)}|(1 - \mathbb{1}\{V_i \in S_n^-\}) \to 0$;
(2) For an arbitrary $\eta > 0$, $\mathbb{E}|\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(V_i)}|^{2+\max(\sigma,\eta)} < \infty$.

Proof. For part (1), we first let q_r be some positive constant such that $q_r < \frac{\xi_r}{\lambda_r}$. Then by Assumption 4(2), for any arbitrary constant C,

$$\frac{C + (1-F)^{\rightarrow}(z)}{(1-F_\varepsilon)^{\rightarrow}(z^{q_r})} \sim z^{q_r\lambda_r - \xi_r} L(z) \to \infty$$

as $z \to \infty$ in which $L(z)$ is a slowly varying function.

In addition, by Assumption 6, we have $\frac{\lambda_r}{2\xi_r(1-\lambda_r)} < \rho_r$ or equivalently, $\rho_r(\frac{\xi_r}{\lambda_r} - \xi_r) > \frac{1}{2}$. It implies that the q_r we previously chose can further satisfy $\rho_r(q_r - \xi_r) > \frac{1}{2}$. Therefore, the same calculation in the proof of Lemma 7.6(1) with the new q_r and $p = 1$ leads to part (1).

For part (2), when $\sigma > 0$, we have

$$\frac{\xi_r}{\lambda_r} - (2+\sigma)(\xi_r + 1) + 1 = (\xi_r + 1)(\frac{\xi_r(1-\lambda_r)}{\lambda_r(1+\xi_r)} - (1+\sigma)) > 0,$$

in which the last inequality is by Assumption 5. Therefore, it implies that the q_r we previously chose can further satisfy $q_r - (2+\sigma)(\xi_r + 1) + 1 > 0$. Then the same argument in the proof of Lemma 7.6(2) with the new q_r and $p = 2 + \sigma$ leads to part (2).

If $\sigma = 0$, then

$$\frac{\xi_r}{\lambda_r} - 2(\xi_r + 1) + 1 = (\xi_r + 1)(\frac{\xi_r(1-\lambda_r)}{\lambda_r(1+\xi_r)} - 1) > 0.$$

Therefore, we can always find an $\eta > 0$ such that

$$\frac{\xi_r}{\lambda_r} - (2+\eta)(\xi_r + 1) + 1 > 0.$$

This concludes the proof. $\square$

Lemma 7.8. *Let $G(v) = \frac{P(v) - \mathbb{1}\{v>0\}}{f(v)}\mathbb{1}\{v \in S_n\}$ where $P(v) = P(Y = 1|V = v)$ and $f(\cdot)$ is the density of V. If Assumption 2(3) holds, then for any fixed $u \in [-1, 1]$ and $h \to 0$, we have*
(1) $\int G^2(v - hu)f(v)dv \le C\mathbb{E}|\frac{Y - \mathbb{1}\{V>0\}}{f(V)}|^{2+\sigma} \le C$;
(2) $\int (G(v - hu) - G(v))^2 f(v)dv \to 0$ as $h \to 0$.

Proof. For part (1), we have

$$\begin{aligned}\int G^2(v-hu)f(v)dv &= \int G^2(v)f(v+hu)dv \\ &\le C\int (\frac{P(v)-\mathbb{1}\{v>0\}}{f(v)})^2 f(v)^{1-\sigma}dv \\ &= C\mathbb{E}(\mathbb{E}\frac{Y-\mathbb{1}\{V>0\}}{f(V)^{1+\frac{\sigma}{2}}}|V)^2 \\ &\le C\mathbb{E}\frac{(Y-\mathbb{1}\{V>0\})^2}{f(V)^{2+\sigma}} \\ &= C\mathbb{E}\left|\frac{Y-\mathbb{1}\{V>0\}}{f(V)}\right|^{2+\sigma} \le C,\end{aligned}$$

in which the first equality is by the change of variables and the fact that V has full support, the first inequality is by Assumption 2(3), the second inequality is by Jansen's inequality, the third equality is because $|Y-\mathbb{1}\{V>0\}|$ only take value 0 or 1, and the last inequality is by Lemma 7.6(2) or 7.7(2).

For part (2), we will follow the proof of Lemma (0.12) in Folland (1995). Because $G(v)\in\mathcal{L}^2(f(v)^{1-\sigma}dv)$, for any $\delta>0$, we can pick a continuous function g with compact support, such that

$$\int (G(v)-g(v))^2 f(v)^{1-\sigma}dv < \delta.$$

This implies

$$\int (G(v)-g(v))^2 f(v)dv < c_1\delta$$

where $c_1=\sup_v f(v)^\sigma$. Second, because g is continuous with compact support, we have

$$\int (g(v-hu)-g(v))^2 f(v)dv \le \delta,$$

for h sufficiently small. Last, we have

$$\begin{aligned}\int (G(v-hu)-g(v-hu))^2 f(v)dv &= \int (G(v)-g(v))^2 f(v+hu)dv \\ &\le c_2\int (G(v)-g(v))^2 f(v)^{1-\sigma}dv \\ &\le c_2\delta.\end{aligned}$$

Combining the three inequalities, we have

$$\int (G^2(v-hu)-G^2(v))f(v)dv \le (c_1+1+c_2)\delta.$$

Since δ is arbitrary, this concludes the proof. □

Lemma 7.9. *If Assumptions 1–4 hold, then*

$$\sup_{S_n^+} |\hat{f}(v) - f(v)|/f(v)^{(1-\sigma)/2} = O_p((\frac{\log(n)}{nh})^{\frac{1}{2}}),$$

$$\inf_{S_n^+} \hat{f}(v) f(v)^{-(1-\sigma)/2} \geq cn^{-\rho(1+\sigma)/2} L_n \quad \textit{w.p.a.1},$$

and

$$1 - o_p(1) = \inf_{S_n^+} |\hat{f}(v)/f(v)| \leq \sup_{S_n^+} |\hat{f}(v)/f(v)| = 1 + o_p(1).$$

Proof. Let $\mathcal{G} = \{\frac{1}{h}K(\frac{\cdot - v}{h}) f(v)^{-(1-\sigma)/2} : v \in S_n^+\}$. By Lemma 7.4, for any $v \in S_n^+$, $f(v) \geq cn^{-\rho} L_n$. Therefore, $\mathcal{G}$ has an envelope $G = Cn^{\rho(1-\sigma)/2} L_n h^{-1}$.

$$\begin{aligned}
\sup_{v \in S_n^+} |\hat{f}(v) - f(v)| f(v)^{-(1-\sigma)/2} &\leq ||\mathcal{P}_n - \mathcal{P}||_{\mathcal{G}} + n^{\rho(1-\sigma)/2} L_n \sup_{v \in S_n^+} |\mathbb{E}\frac{1}{h}K(\frac{V_i - v}{h}) - f(v)| \\
&\lesssim ||\mathcal{P}_n - \mathcal{P}||_{\mathcal{G}} + n^{\rho(1-\sigma)/2} L_n h^{\nu}.
\end{aligned}$$

Because $H > \frac{1+\rho}{1+2\nu}$, the second term in the RHS is $O((\frac{\log(n)}{nh})^{\frac{1}{2}})$. We now focus on bounding the first term. Note

$$\sup_{g \in \mathcal{G}} \mathbb{E} g^2 \lesssim h^{-1} \sup_{v \in S_n^+} \int \frac{f(v + hu)}{f(v)^{1-\sigma}} K^2(u) du \lesssim h^{-1}.$$

In addition, $\mathcal{G}$ is a VC-class with a fixed VC-index. Therefore, by Corollary 5.1 of Chernozhukov et al. (2014),

$$\mathbb{E}||\mathcal{P}_n - \mathcal{P}||_{\mathcal{G}} \lesssim (\log(n)/(nh))^{1/2} + \log(n) n^{\rho(1-\sigma)/2}/nh \lesssim (\log(n)/(nh))^{1/2},$$

in which the last inequality holds because $H < 1 - (1+\sigma)\rho \leq 1 - (1-\sigma)\rho$. This leads to the first result.

For the second result, we note that, w.p.a.1,

$$\begin{aligned}
\inf_{S_n^+} |\hat{f}(v)| f(v)^{-(1-\sigma)/2} =& \inf_{S_n^+} \left| f(v)^{(1+\sigma)/2} + f(v)^{-(1-\sigma)/2} (\hat{f}(v) - f(v)) \right| \\
\geq& cn^{-\rho(1+\sigma)/2} L_n - O_p\left((\frac{\log(n)}{nh})^{1/2} \right) \\
\geq& cn^{-\rho(1+\sigma)/2} L_n,
\end{aligned}$$

where the last inequality is due to $H < 1 - (1+\sigma)\rho$.

For the last result, note

$$\sup_{S_n^+} |\frac{\hat{f}(v) - f(v)}{f(v)}| \leq \sup_{S_n^+} \frac{|\hat{f}(v) - f(v)|}{f(v)^{(1-\sigma)/2}} \frac{1}{\inf_{S_n^+} f(v)^{(1+\sigma)/2}} = O_p\left((\frac{\log(n) n^{\rho(1+\sigma)}}{nh})^{1/2} \right) = o_p(1),$$

where the last equality is due to $H < 1-(1+\sigma)\rho$. Therefore,

$$\inf_{S_n^+} |\hat{f}(v)/f(v)| \geq 1 - \sup_{S_n^+} |(\hat{f}(v)-f(v))/f(v)| = 1 - o_p(1).$$

The upper bound can be established by a similar argument. □

7.3 Lemmas for Tan and Zhang (2017, Theorem 4.1)

Lemma 7.10. *If the assumptions in Theorem 4.1 hold, then, for $R_{n,j}$, $j = 1,2,3$ defined in the proof of Theorem 4.1, we have*

$$\sum_{j=1}^{3} R_{n,j} = o_p(\frac{1}{\sqrt{n}}).$$

The key of the proof is to find two deterministic sequences $M_{n,r}$ and $M_{n,l}$ such that for

$$A_n = \{|\tilde{l}_n - l_n| \leq M_{n,l} \cap |\tilde{r}_n - r_n| \leq M_{n,r}\},$$

we have $P(A_n) \to 1$. This has been done in Lemma 7.11.

Next, we define $S_n^+ = (-M_{n,l}+l_n, r_n+M_{n,r})$, $S_n^- = (M_{n,l}+l_n, r_n-M_{n,r})$, $I_{n,i}^+ = \mathbb{1}\{U_i \in S_n^+\}$, and $I_{n,i}^- = \mathbb{1}\{U_i \in S_n^-\}$. On A_n, we have

$$I_{n,i}^- < \tilde{I}_{n,i} < I_{n,i}^+$$

for all $i = 1, \cdots, n$. This implies

$$I_{n,i}^- - 1 \leq I_{n,i}^- - I_{n,i} \leq \tilde{I}_{n,i} - I_{n,i} \leq I_{n,i}^+ - I_{n,i} \leq \min(1 - I_{n,i}^-, I_{n,i}^+). \tag{7.12}$$

We derive bounds for various terms by replacing $\tilde{I}_{n,i}$ by the nonrandom upper and lower bounds. This has been done in Lemmas 7.12–7.14. Given Lemmas 7.11–7.14, we next prove Lemma 7.10.

Proof. By Lemma 7.11(1), on A_n, we have

$$|\tilde{I}_{n,i} - I_{n,i}| \leq 1 - \mathbb{1}\{U_i \in S_n^-\},$$

and $P(A_n) \to 1$. Then, by Lemma 7.12(1) or 7.13(1), on A_n,

$$\mathbb{E}|R_{n,1}| \leq \mathbb{E}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(U_i)}\right|(1 - \mathbb{1}\{U_i \in S_n^-\}) = o(\frac{1}{\sqrt{n}}).$$

This implies $R_{n,1} = o_p(\frac{1}{\sqrt{n}})$. For $R_{n,2}$, on A_n, we have

$$|R_{n,2}| \leq \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(U_i)}\right|(1 - \mathbb{1}\{U_i \in S_n^-\})\left|\frac{f(U_i) - \tilde{f}(\hat{U}_i)}{\tilde{f}(\hat{U}_i)}\right|\mathbb{1}\{U_i \in S_n^+\}.$$

By Lemma 7.14, we have

$$\max_{1\le i\le n}\left|\frac{f(U_i)-\tilde{f}(\hat{U}_i)}{\tilde{f}(\hat{U}_i)}\right|\mathbb{1}\{U_i\in S_n^+\}=o_p(1),$$

Furthermore, we have already shown in Lemma 7.12(1) or 7.13(1) that

$$\frac{1}{n}\sum_{i=1}^n\left|\frac{Y_i-\mathbb{1}\{V_i>0\}}{f(U_i)}\right|(1-\mathbb{1}\{U_i\in S_n^-\})=o_p(\frac{1}{\sqrt{n}}).$$

This implies $R_{n,2}=o_p(\frac{1}{\sqrt{n}})$. For $R_{n,3}$, on A_n, we have

$$|R_{n,3}|\le\frac{1}{n}\sum_{i=1}^n\left|\frac{Y_i-\mathbb{1}\{V_i>0\}}{f(U_i)^3}\right|\left|\frac{\sqrt{n}(f(U_i)-\tilde{f}(\hat{U}_i))^2}{1+o_p(1)}\right|\mathbb{1}\{U_i\in S_n\}.$$

By Lemma 7.12(1) or 7.13(1),

$$\frac{1}{n}\sum_{i=1}^n\left|\frac{Y_i-\mathbb{1}\{V_i>0\}}{f(U_i)^3}\right|=O_p(1)$$

and

$$\begin{aligned}&\max_{1\le i\le n}(f(U_i)-\tilde{f}(\hat{U}_i))^2\\ \le&\max_{1\le i\le n}[f(U_i)-\hat{f}(U_i)]^2+\max_{1\le i\le n}(\tilde{f}(\hat{U}_i)-\hat{f}(U_i))^2\\ \le&O_p(\frac{\log(n)}{nh})+\max_{1\le i\le n}\left(\frac{1}{n-1}\sum_{j\ne i}K'(\frac{\tilde{U}_j-\tilde{U}_i}{h})/(\sqrt{n}h)\right)^2=O_p(\frac{1}{nh^2})=o_p(\frac{1}{\sqrt{n}}).\end{aligned}$$

This concludes that $R_{n,3}=o_p(\frac{1}{\sqrt{n}})$.

□

Lemma 7.11. *There exist positive sequences $M_{n,l}$ and $M_{n,l}$ such that for*

$$A_n=\{|\tilde{l}_n-l_n|\le M_{n,l}\cap|\tilde{r}_n-r_n|\le M_{n,r}\},$$

$S_n^+=(-M_{n,l}+l_n,r_n+M_{n,r})$, $S_n^-=(M_{n,l}+l_n,r_n-M_{n,r})$, $\tilde{S}_n^+=(-2M_{n,l}+l_n,r_n+2M_{n,r})$, *and* $\tilde{S}_n^-=(2M_{n,l}+l_n,r_n-2M_{n,r})$, *we have (1)* $P(A_n)\to 1$ *and on* A_n,

$$\{\hat{U}_i\in\tilde{S}_n^-\}\subset\{U_i\in S_n^-\}\subset\{\hat{U}_i\in\tilde{S}_n\}\subset\{U_i\in S_n^+\}\subset\{\hat{U}_i\in\tilde{S}_n^+\}$$

for $i=1,\cdots,n$, *and (2) for* $u\in S_n^+$, $f(u)\ge cn^{-\rho}L_n$, *where* $\rho=\max(\rho_r(1+\xi_r),\rho_l(1+\xi_l))$.

Proof. The only difference between Lemma 7.11 and Lemma 7.4 is that, here, U_i is unobservable. We propose to replace it by the residual $\hat{U}_i$. Then the feasible trimming points $\tilde{l}_n$ and $\tilde{r}_n$ are computed as order statistics of $\hat{U}_i$. By Assumption 9 and the fact that $\hat{\gamma}$ is $\sqrt{n}$-consistent, we have

$$\max_{1\le i\le n}|\hat{U}_i - U_i| \le \max_{1\le i\le n}|S_i||\hat{\gamma}-\gamma| = O_p(\frac{1}{\sqrt{n}}).$$

Since the convergence rate for the estimator of intermediate order statistics is slower than $\sqrt{n}$, it is expected that the convergence rates of $\tilde{l}_n$ and $\tilde{r}_n$ to their true values will not be affected by using $\hat{U}_i$ for estimation. In particular, for the left tail, let $\tau_{n,l} = n^{-\rho_l}$, $m_l = \lfloor n\tau_{n,l}\rfloor$, $\alpha_{n,l} = \sqrt{\frac{n}{\tau_{n,l}}}f(F^{\leftarrow}(\tau_{n,l}))$. We can show that

$$\hat{U}^{(n)}_{(m_l)} - l_n = O_p(\frac{1}{\alpha_{n,l}}). \tag{7.13}$$

Similarly, for the right tail, we have

$$\hat{U}^{(n)}_{(n-m_r+1)} - r_n = O_p(\frac{1}{\alpha_{n,r}}), \tag{7.14}$$

in which $\tau_{n,r} = n^{-\rho_r}$, $m_r = \lfloor n\tau_{n,r}\rfloor$, and $\alpha_{n,r} = \sqrt{\frac{n}{\tau_{n,r}}}f((1-F)^{\leftarrow}(\tau_{n,r}))$. Given (7.13) and (7.14), for any sequence $M_n \to \infty$, we have

$$\frac{M_n\sqrt{n}}{\alpha_{n,r}} \sim \frac{M_n}{\tau_{n,r}^{\frac{1}{2}+\xi_r}L_n} \to \infty, \quad \frac{M_n\sqrt{n}}{\alpha_{n,l}} \sim \frac{M_n}{\tau_{n,l}^{\frac{1}{2}+\xi_l}L_n} \to \infty.$$

Then w.p.a.1,

$$\begin{aligned}\{\hat{U}_i \le \tilde{r}_n\} &\subset \{\hat{U}_i \le r_n + |r_n - \tilde{r}_n|\}\\ &\subset \{U_i \le r_n + \frac{M_n}{\alpha_{n,r}} + \max_{i\le n}|U_i - \hat{U}_i|\}\\ &\subset \{U_i \le r_n + \frac{M_n}{\alpha_{n,r}} + \frac{C}{\sqrt{n}}\}\\ &\subset \{U_i \le r_n + \frac{2M_n}{\alpha_{n,r}}\}.\end{aligned}$$

Similarly,

$$\{\hat{U}_i \le \tilde{r}_n\} \supset \{U_i \le r_n - \frac{2M_n}{\alpha_{n,r}}\}.$$

We can further show

$$\{U_i \le r_n - \frac{2M_n}{\alpha_{n,r}}\} \supset \{\hat{U}_i \le r_n - \frac{4M_n}{\alpha_{n,r}}\} \quad \text{and} \quad \{U_i \le r_n + \frac{2M_n}{\alpha_{n,r}}\} \subset \{\hat{U}_i \le r_n + \frac{4M_n}{\alpha_{n,r}}\}.$$

This implies we can choose $M_{n,r} = \frac{2M_n}{\alpha_{n,r}}$ for any $M_n \to \infty$ for the right tail. Similarly, we can choose $M_{n,l} = \frac{2M_n}{\alpha_{n,l}}$. Therefore, given (7.13) and (7.14), we have already proven part (1) of the lemma. Part (2) and (3) follow the same argument of Lemma 7.4 which we will not repeat. Now, we turn to (7.13) and (7.14).

We focus on the left tail and show (7.13). (7.14) can be derived in the same manner. Let $\rho_\tau(u)$ be the check function defined as $\rho_\tau(u) = u(\tau - \mathbb{1}\{u \le 0\})$. Then,

$$\hat{U}^{(n)}_{(m_l)} \in \arg\min_q \sum_{i=1}^{n} \rho_{\tau_{n,l}}(\hat{U}_i - q).$$

Define $z = \alpha_{n,l}(q - l_n)$, $\hat{Z}_n = \alpha_{n,l}(\hat{U}^{(n)}_{(m_l)} - l_n)$, and

$$Q(\tau_{n,l}, z) = \frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} [\rho_{\tau_{n,l}}(U_i - l_n - \frac{z}{\alpha_{n,l}} + \hat{U}_i - U_i) - \rho_{\tau_{n,l}}(U_i - l_n)].$$

Then $\hat{Z}_n$ minimizes $Q(\tau_{n,l}, z)$.

In the following, we show that, $Q(\tau_{n,l}, z)$, the rescaled version of the objective function, weakly converges to a limiting process

$$Q_\infty(z) = -zW + \frac{z^2}{2},$$

where $W \sim \mathcal{N}(0, 1)$. Then we can apply the Convexity lemma and the same argument in the proof of Pollard (1991, Theorem 1) to derive the desired result

$$\hat{Z}_n \rightsquigarrow W = O_p(1).$$

By equation (9.44) of Chernozhukov (2005),

$$\begin{aligned} Q(\tau_{n,l}, z) = & \frac{-z}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} (\tau_{n,l} - \mathbb{1}\{U_i \le l_n\}) + \frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} (U_i - \hat{U}_i)(\tau_{n,l} - \mathbb{1}\{U_i \le l_n\}) \\ & + \frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} \int_0^{\frac{z}{\alpha_{n,l}} + U_i - \hat{U}_i} [\mathbb{1}\{U_i - l_n \le s\} - \mathbb{1}\{U_i - l_n \le 0\}] ds. \end{aligned} \tag{7.15}$$

Next, we aim to show

$$\frac{1}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} (\tau_{n,l} - \mathbb{1}\{U_i \le l_n\}) \rightsquigarrow W, \tag{7.16}$$

$$\frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} (U_i - \hat{U}_i)(\tau_{n,l} - \mathbb{1}\{U_i \le l_n\}) = o_p(1), \tag{7.17}$$

and

$$\frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} \int_0^{\frac{z}{\alpha_{n,l}} + U_i - \hat{U}_i} [\mathbb{1}\{U_i - l_n \le s\} - \mathbb{1}\{U_i - l_n \le 0\}] ds \xrightarrow{p} \frac{z^2}{2}. \tag{7.18}$$

(7.16) holds because of the triangular array CLT such as Theorem 3.4.5 in Durrett (2010). Here, the Lyapunov condition for the CLT holds because $n\tau_n \to \infty$.

For (7.17), we have

$$\frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}}\sum_{i=1}^{n}(U_i-\hat{U}_i)(\tau_{n,l}-\mathbb{1}\{U_i\le l_n\}) = \left[\frac{1}{\sqrt{n\tau_{n,l}}}\sum_{i=1}^{n}(\tau_{n,l}-\mathbb{1}\{U_i\le l_n\})Z_i'\right][\alpha_{n,l}(\hat{\gamma}-\gamma)]$$
$$= O_p\left(\frac{\alpha_{n,l}}{\sqrt{n}}\right) = O_p(L(\tau_{n,l})\tau_{n,l}^{\xi_l+0.5}) = o_p(1),$$

in which the second last equality is because $f(F^{\leftarrow}(\tau_{n,l}))\in RV_{\xi_l+1}(0)$, and $L(\tau)$ is a slowly varying function at 0 such that $\frac{L(k\tau)}{L(\tau)}\to 1$ for any $k>0$ as $\tau\to 0$.

Before proving (7.18), we first define $\hat{d}=\sqrt{n}(\hat{\gamma}-\gamma)$. Then $U_i-\hat{U}_i = Z_i\frac{\hat{d}}{\sqrt{n}}$. Since $\hat{\gamma}$ is $\sqrt{n}$-consistent, $\hat{d}=O_p(1)$. Next we consider

$$\Lambda_n(z,d) = \frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}}\sum_{i=1}^{n}\int_0^{\frac{z}{\alpha_{n,l}}+Z_i\frac{d}{\sqrt{n}}}[\mathbb{1}\{U_i-l_n\le s\}-\mathbb{1}\{U_i-l_n\le 0\}]ds$$
$$= \frac{1}{\sqrt{n\tau_{n,l}}}\sum_{i=1}^{n}\int_0^{z+Z_i\frac{d\alpha_{n,l}}{\sqrt{n}}}[\mathbb{1}\{U_i-l_n\le \frac{s}{\alpha_{n,l}}\}-\mathbb{1}\{U_i-l_n\le 0\}]ds,$$

and show $\Lambda_n(z,d)\xrightarrow{p}\frac{z^2}{2}$ uniformly over $|z|\le B, |d|\le B$, for any $B>0$. To see this, we note that

$$\mathbb{E}\Lambda_n(z,d) = \sqrt{\frac{n}{\tau_{n,l}}}\mathbb{E}\int_0^{z+\frac{\alpha_{n,l}}{\sqrt{n}}Z_id}[\mathbb{1}\{U_i-l_n\le \frac{s}{\alpha_{n,l}}\}-\mathbb{1}\{U_i-l_n\le 0\}]ds$$
$$= \sqrt{\frac{n}{\tau_{n,l}}}\mathbb{E}\int_0^{z+\frac{\alpha_{n,l}}{\sqrt{n}}Z_id}[F(l_n+\frac{s}{\alpha_{n,l}})-F(l_n)]ds$$
$$= \sqrt{\frac{n}{\tau_{n,l}}}\mathbb{E}\int_0^{z+\frac{\alpha_{n,l}}{\sqrt{n}}Z_id}f(l_n+\frac{\tilde{s}}{\alpha_{n,l}})\frac{s}{\alpha_{n,l}}ds$$

in which $\tilde{s}$ is between 0 and $z+\frac{\alpha_{n,l}}{\sqrt{n}}Z_id$. Since $|z|<B$, $|d|<B$, $|Z_i|<B$ and $\frac{\alpha_{n,l}}{\sqrt{n}}\to 0$, $\tilde{s}$ is bounded. Then by Equation (9.57) of Chernozhukov (2005), we have

$$f(l_n+\frac{\tilde{s}}{\alpha_{n,l}})\sim f(l_n).$$

Hence we have, uniformly over z,d,

$$\mathbb{E}\Lambda_n(z,d)\sim\sqrt{\frac{\tau_{n,l}}{n}}\frac{f(l_n)}{\alpha_{n,l}}\frac{1}{2}\mathbb{E}(z+\frac{\alpha_{n,l}}{\sqrt{n}}Z_id)^2\to\frac{s^2}{2}.$$

Next, we show

$$\sup_{|d|<B,|z|<B}|\Lambda_n(z,d)-\mathbb{E}\Lambda_n(z,d)|\xrightarrow{p}0.$$

Let us consider the class of functions

$$\mathcal{F} = \left\{ \frac{1}{\sqrt{\tau_{n,l}}} \int_0^{z+Z_i \frac{d\alpha_{n,l}}{\sqrt{n}}} \left[\mathbb{1}\{U_i - l_n \leq \frac{s}{\alpha_{n,l}}\} - \mathbb{1}\{U_i - l_n \leq 0\} \right] ds : |d| < B, |z| < B \right\}$$

with an envelope $F_e = \frac{C}{\sqrt{\tau_{n,l}}}$. It is easy to see that $\mathcal{F}$ satisfies the uniform entropy condition, that is,

$$\sup_Q N(\varepsilon ||F_e||_{Q,2}, \mathcal{F}, ||\cdot||_{Q,2}) \leq \left(\frac{a}{\varepsilon}\right)^v, \quad \forall \varepsilon \in (0,1].$$

In addition, since Z_i is bounded, for $\sigma_n^2 = \sup_{f\in\mathcal{F}} \mathbb{E} f^2$, we have

$$\begin{aligned}
\sigma_n^2 &\lesssim \tau_{n,l}^{-1} \mathbb{E} \left[\mathbb{1}\left\{ U_i \leq l_n + \frac{B + |Z_i| B \frac{\alpha_{n,l}}{\sqrt{n}}}{\alpha_{n,l}} \right\} - \mathbb{1}\left\{ U_i \leq l_n - \frac{B + |Z_i| B \frac{\alpha_{n,l}}{\sqrt{n}}}{\alpha_{n,l}} \right\} \right] \\
&\lesssim \tau_{n,l}^{-1} f(l_n) \frac{1}{\alpha_{n,l}} = \frac{1}{\sqrt{n\tau_{n,l}}},
\end{aligned}$$

in which the second inequality is by Equation (9.57) of Chernozhukov (2005). Then by Corollary 5.1 of Chernozhukov et al. (2014), we have

$$\begin{aligned}
&\mathbb{E} \sup_{|d|<B, |z|<B} |\Lambda_n(z,d) - \mathbb{E}\Lambda_n(z,d)| \\
&= \mathbb{E} ||\sqrt{n}(\mathcal{P}_n - \mathcal{P})||_{\mathcal{F}} \\
&\lesssim \sqrt{\sigma_n^2 \log(\frac{||F_e||_{P,2}}{\sigma_n})} + \frac{1}{\sqrt{n\tau_{n,l}}} \log(\frac{||F_e||_{P,2}}{\sigma_n}) \\
&\lesssim \sqrt{\frac{\log(n)}{\sqrt{n\tau_{n,l}}}} \to 0.
\end{aligned}$$

This implies $\sup_{|d|<B, |z|<B} |\Lambda_n(z,d) - \mathbb{E}\Lambda_n(z,d)| \xrightarrow{p} 0$. Thus

$$\Lambda_n(z,d) \xrightarrow{p} \frac{z^2}{2}$$

uniformly in $|z| < B, |d| < B$. Then uniformly over $|z| < B$,

$$\Lambda_n(z,\hat{d}) \xrightarrow{p} \frac{z^2}{2}$$

and thus (7.18) holds. Combining (7.16)–(7.18), we have

$$Q(\tau_{n,l}, z) \rightsquigarrow -zW + \frac{z^2}{2}.$$

Since the RHS is uniquely minimized at $z = W$, by the same argument of the proof of Pollard (1991, Theorem 1), we have

$$\alpha_{n,l}(\hat{U}^{(n)}_{(m_l)} - l_n) \rightsquigarrow W \quad \text{and thus} \quad (\hat{U}^{(n)}_{(m_l)} - l_n) = O_p(\frac{1}{\alpha_{n,l}}).$$

This concludes the proof. □

Lemma 7.12. *If Assumption 11(1) or (3) holds and the tuning parameters h and (ρ_r, ρ_l) are chosen as in Assumption 12, then the following statements hold:*
(1) $\sqrt{n}\mathbb{E}\left|\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(U_i)^p}\right|(1 - \mathbb{1}\{U_i \in S_n^-\}) \to 0$ for any $p > 0$;
(2) $\mathbb{E}\left|\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(U_i)^p}\right| < \infty$ for any $p > 0$.

Proof. For part (1), we have $1 - \mathbb{1}\{u \in S_n^-\} = \mathbb{1}\{u > r_n - M_{n,r}\} + \mathbb{1}\{u < l_n + M_{n,l}\}$. We focus on the right tail. Since S_i has bounded support, $U_i > r_n - M_{n,r}$, and $\frac{M_{n,r}}{r_n} \to 0$, we have $V_i = U_i + S_i'\gamma > 0$. Therefore,

$$\begin{aligned}
&\sqrt{n}\mathbb{E}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(U_i)^p}\right|\mathbb{1}\{U_i > r_n - M_{n,r}\} \\
\le&\sqrt{n}\mathbb{E}\left(\frac{1 - \mathbb{1}\{\varepsilon_i \le X_i'\beta + Z_i'\gamma + U_i\}}{f(U_i)^p}\right)\mathbb{1}\{U_i > r_n - M_{n,r}\} \\
\le&\sqrt{n}\mathbb{E}\left(\frac{1 - \mathbb{1}\{\varepsilon_i \le U_i - C\}}{f(U_i)^p}\right)\mathbb{1}\{U_i > r_n - M_{n,r}\} \\
\le&\sqrt{n}\int_{r_n - M_{n,r}}^{\infty}\frac{(1 - F_\varepsilon)(u - C)}{f(u)^{p-1}}du.
\end{aligned}$$

The RHS of the above display is $o(1)$ by the same argument in the proof of Lemma 7.6.

For part (2), we can choose C_1 such that when $U_i > C_1$, $V_i > 0$ and when $U_i < -C_1$, $V_i < 0$. On $|U_i| \le C_1$, the integrand $|\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(U_i)^p}|$ is bounded. So we only have to check the integrability at $\pm\infty$. We focus on the right tail.

$$\begin{aligned}
&\mathbb{E}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(U_i)^p}\right|\mathbb{1}\{U_i > C_1\} \\
\le&C + \int_{C_1}^{\infty}\frac{(1 - F_\varepsilon)(u - C)}{f(u)^{p-1}}du \\
\le&C + \int_0^c \frac{(1 - F_\varepsilon)((1 - F)^{\leftarrow}(z) - C)}{f((1 - F)^{\leftarrow}(z))^p}dz < \infty.
\end{aligned}$$

The last inequality holds by the same argument in the proof of part (2) of Lemma 7.6.

□

Lemma 7.13. *If Assumption 11(2) holds and the tuning parameters h and (ρ_r, ρ_l) are chosen as in Assumption 12(2), then the following statements hold:*

(1) $\sqrt{n}\mathbb{E}\left|\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(U_i)}\right|(1-\mathbb{1}\{U_i \in S_n^-\}) \to 0$*;*
(2) $\mathbb{E}\left|\frac{Y_i - \mathbb{1}\{V_i>0\}}{f(U_i)^{3+\sigma}}\right| < \infty$.

Proof. For part (1), by repeating the proof of Lemma 7.12 with $p = 1$, for the right tail, we have

$$\sqrt{n}E\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(U_i)}\right|\mathbb{1}\{U_i > r_n - M_{n,r}\} \le \sqrt{n}\int_{r_n - M_{n,r}}^{\infty}(1-F_\varepsilon)(u-C)du.$$

In order for the RHS to vanish, as in the proof of Lemma 7.7, we need $\rho_r(\frac{\xi_r}{\lambda_r} - \xi_r) > \frac{1}{2}$ which holds by Assumption 12.

For part (2), by repeating the proof of Lemma 7.12 with $p = 3+\sigma$, for the right tail, we have

$$\mathbb{E}\left|\frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(U_i)^{3+\sigma}}\right|\mathbb{1}\{U_i > C_1\} \le \int_0^c \frac{(1-F_\varepsilon)((1-F)^{\leftarrow}(z) - C)}{f((1-F)^{\leftarrow}(z))^{2+\sigma}}dz. \tag{7.19}$$

Then, following the proof of Lemma 7.7, the RHS of (7.19) is finite because $\frac{\xi_r}{\lambda_r} - (3+\sigma)(\xi_r+1)+1 > 0$ by Assumption 11(2). $\square$

Lemma 7.14.

$$\max_{1\le i\le n}|[\tilde{f}(\hat{U}_i) - f(U_i)]/f(U_i)|\mathbb{1}\{U_i \in S_n^+\} = o_p(1),$$

$$1 - o_p(1) \le \min_{1\le i\le n}|\tilde{f}(\hat{U}_i)/f(U_i)|\mathbb{1}\{U_i \in S_n^+\} \le \max_{1\le i\le n}|\tilde{f}(\hat{U}_i)/f(U_i)|\mathbb{1}\{U_i \in S_n^+\} = 1 + o_p(1),$$

and

$$|\tilde{f}(\hat{U}_i)|\mathbb{1}\{U_i \in S_n^+\} \gtrsim n^{-\rho}L_n\mathbb{1}\{U_i \in S_n^+\},$$

where $\rho = \max(\rho_r(1+\xi_r), \rho_l(1+\xi_l))$.

Proof. For the first result, we note

$$\begin{aligned}\tilde{f}(\hat{U}_i) - f(U_i) =& \tilde{f}(\hat{U}_i) - f(\hat{U}_i) + f(\hat{U}_i) - f(U_i)\\ =& \hat{f}(\hat{U}_i) - f(\hat{U}_i) + f(\hat{U}_i) - f(U_i) + \tilde{f}(\hat{U}_i) - \hat{f}(\hat{U}_i),\end{aligned}$$

in which $\hat{f}(\hat{U}_i) = \frac{1}{(n-1)h}\sum_{j\ne i}K(\frac{U_j - \hat{U}_i}{h})$. By Lemma 7.9, we have, w.p.a.1,

$$\max_{1\le i\le n}|\hat{f}(\hat{U}_i) - f(\hat{U}_i)|/f(\hat{U}_i)\mathbb{1}\{U_i \in S_n^+\} \le \sup_{v\in\tilde{S}_n^+}|\hat{f}(v) - f(v)|/f(v) = o_p(1).$$

So we only need to prove

$$\max_{1\le i\le n}|f(\hat{U}_i) - f(U_i)|/f(U_i)\mathbb{1}\{U_i \in S_n^+\} = o_p(1) \tag{7.20}$$

and

$$\max_{1\le i\le n}|\tilde{f}(\hat{U}_i) - \hat{f}(\hat{U}_i)|/f(\hat{U}_i)\mathbb{1}\{U_i \in S_n^+\} = o_p(1). \tag{7.21}$$

We first prove (7.20). Let $U_i = Q(\tau_i)$ where $Q(\cdot)$ is the quantile function of U. By the construction of S_n^+, for any $l < 1$, we have $S_n^+ \in (Q(l\tau_{n,l}), Q(1-l\tau_{n,r}))$. To see this, notice the lower end point of S_n^+ is $Q(\tau_{n,l}) - 2M_n/\alpha_{n,l}$. We claim that $Q(\tau_{n,l}) - 2M_n/\alpha_{n,l} > Q(l\tau_{n,l})$ for any $l < 1$, and therefore, $U_i \geq Q(l\tau_{n,l})$. The claim is equivalent to

$$\alpha_{n,l}(Q(\tau_{n,l}) - Q(l\tau_{n,l})) > 2M_n.$$

Note for the intermediate order quantile, the convergence rate is

$$\alpha_{n,r} \sim c\sqrt{n\tau_{n,l}}/(Q(e\tau_{n,l}) - Q(\tau_{n,l})).$$

Therefore, the above inequality is equivalent to

$$c\sqrt{m_l} > 2M_n \tag{7.22}$$

where c is a generic constant. We can choose M_n in Lemma 7.11 such that (7.22) holds. Similarly, we can show

$$Q(1-\tau_{n,r}) + 2M_n/\alpha_{n,r} \leq Q(1-l\tau_{n,r}).$$

Furthermore, $\max_{1\leq i\leq n}|U_i - \hat{U}_i| \leq \max_{1\leq i\leq n}|S_i||\hat{\gamma}-\gamma| = O_p(\frac{1}{\sqrt{n}})$. Hence, for any δ_n such that $\sqrt{n}\delta_n \to \infty$ and any $l < 1$,

$$\max_{1\leq i\leq n}|f(\hat{U}_i) - f(U_i)|/f(U_i)\mathbb{1}\{U_i \in S_n^+\} \leq \sup_{|\Delta|\leq\delta_n,\tau\in(l\tau_{n,l},1-l\tau_{n,r})}|f(Q(\tau)+\Delta) - f(Q(\tau))|/f(Q(\tau)).$$

We want to show the RHS decays to zero by checking the limits of all possible convergent subsequences. Note that $\tau = \tau_n \in (l\tau_{n,l}, 1-l\tau_{n,r})$. As $n \to \infty$, we have three possibilities: (1) $\tau_n \to \tau_0 \in (0,1)$; (2) $\tau_n \to 0$; and (3) $\tau_n \to 1$. For case (1), by the mean value theorem, we have

$$|f(Q(\tau_n)+\Delta) - f(Q(\tau_n))|/f(Q(\tau_n)) \leq c|\Delta| \to 0.$$

For case (2), we first claim that for any $L > 1$, there exists some sequence δ_n such that $\delta_n \to 0$, $\sqrt{n}\delta_n \to \infty$, and $\Delta \leq \delta_n \leq Q(L\tau_n) - Q(\tau_n)$.

To see the claim, note that, if $\xi_l = 0$, then $Q(L\tau_n) - Q(\tau_n)$ is slowly varying. But δ_n can decay to zero at a rate arbitrarily close to $1/2$. Therefore the claim holds. On the other hand, if $\xi_l > 0$, $Q(L\tau_n) - Q(\tau_n)$ diverges to infinity while δ_n will decay to zero. Again, the claim holds. Similarly, we can show that

$$\Delta \geq Q(\tau_n) - Q(\tau_n/L).$$

Given the density of U is monotonic in the tails, we have

$$\frac{|f(Q(\tau_n)+\Delta_n)-f(Q(\tau_n))|}{f(Q(\tau_n))} \leq \max\Big(\frac{f(Q(L\tau_n))-f(Q(\tau_n))}{f(Q(\tau_n))}, \frac{f(Q(\tau_n))-f(Q(\tau_n/L))}{f(Q(\tau_n))}\Big)$$
$$\to \max(L^{\xi+1}-1, 1-L^{-(\xi+1)}),$$

where the second line is because $f(Q(\tau))$ is regularly varying with varying index $\xi+1>0$. Since $L>1$ is arbitrary, by letting $L\to 1$, we have shown that, as $\tau_n \to 0$,

$$\frac{|f(Q(\tau_n)+\Delta_n)-f(Q(\tau_n))|}{f(Q(\tau_n))} \to 0.$$

Similarly, we can show that, for case (3),

$$\frac{|f(Q(\tau_n)+\Delta_n)-f(Q(\tau_n))|}{f(Q(\tau_n))} \to 0.$$

Hence there exists some sequence $\delta_n \to 0$ for which $\sqrt{n}\delta_n \to \infty$ and

$$\sup_{|\Delta|\leq \delta_n, \tau\in(l\tau_{n,l}, 1-l\tau_{n,r})} |f(Q(\tau)+\Delta)-f(Q(\tau))|/f(Q(\tau)) \to 0.$$

Thus (7.20) holds.

Next, we turn to (7.21). Since $\mathbb{1}\{U_i \in S_n^+\} \leq \mathbb{1}\{\hat{U}_i \in \tilde{S}_n^+\}$ and (7.20), in order to show (7.21), it suffices to show

$$\sup_{v\in\tilde{S}_n^+} |\tilde{f}(v)-\hat{f}(v)|/f(v) = o_p(1).$$

By Lemma 7.9, we have

$$\sup_{v\in\tilde{S}_n^+} |f(v)-\hat{f}(v)|/f(v) = o_p(1).$$

So we can focus on proving

$$\sup_{v\in\tilde{S}_n^+} |\tilde{f}(v)-f(v)|/f(v) = o_p(1).$$

Since $\hat{U}_j = U_j - S_j'(\hat{\gamma}-\gamma)$ and $(\hat{\gamma}-\gamma) = O_p(\frac{1}{\sqrt{n}})$, it suffices to show

$$\sup_{v\in\tilde{S}_n^+, |\pi|\leq M} |\frac{1}{nh}\sum_{j=1}^n K(\frac{U_j - S_j'\pi/\sqrt{n} - v}{h}) - f(v)|/f(v) = o_p(1).$$

Let f_s be the PDF of S. Since $U \perp\!\!\!\perp S$, we have

$$|\mathbb{E}K(\frac{U_j - S_j'\pi/\sqrt{n} - v}{h})/h - f(v)|/f(v) =| \int\int f_s(s)f(v + hu - s'\pi/\sqrt{n})dsK(u)du - f(v)|/f(v)$$
$$\leq \int\int f_s(s)\frac{|f(v + hu - s'\pi/\sqrt{n}) - f(v)|}{f(v)}dsK(u)du.$$

By the same proof of (7.20), it can be shown that

$$\sup_{v\in\tilde{S}_n^+, |\pi|\leq M, s\in \text{Supp}(S)} \frac{|f(v + hu - s'\pi/\sqrt{n}) - f(v)|}{f(v)} \to 0.$$

Therefore,

$$\sup_{v\in\tilde{S}_n^+, |\pi|<M} |\mathbb{E}K(\frac{U_j - S_j'\pi/\sqrt{n} - v}{h})/h - f(v)|/f(v) = o(1).$$

Let $G = \{K(\frac{U - S'\pi/\sqrt{n} - v}{h})/hf(v)^{(1-\sigma)/2} : |\pi| \leq M, v \in \tilde{S}_n^+\}$ with envelope $Ch^{-1}L_n n^{\rho(1-\sigma)/2}$. We can repeat the proof of Lemma 7.9 and show

$$||\mathcal{P}_n - \mathcal{P}||_{\mathcal{G}} = O_p((\log(n)/(nh))^{1/2}).$$

This is because

$$\sup_{\mathcal{G}} \mathbb{E}g^2 \leq h^{-1} \sup_{v\in\tilde{S}_n^+, |\pi|<M, s\in \text{Supp}(S)} \int\int f_s(s)f(v + hu - s'\pi/\sqrt{n})dsK^2(u)du/f(v)^{1-\sigma} \lesssim h^{-1}.$$

Given this and the fact that $1 - H - (1 + \sigma)\rho > 0$, we have

$$\sup_{v\in\tilde{S}_n^+, |\pi|\leq M} |\frac{1}{nh}\sum_{j=1}^{n} K(\frac{U_j - S_j'\pi/\sqrt{n} - v}{h}) - \mathbb{E}K(\frac{U_j - S_j'\pi/\sqrt{n} - v}{h})/h|/f(v) = o_p(1).$$

This implies (7.21) holds, and, thus, the first result of the lemma. The second result follows immediately. The third result holds by noticing $f(v) \geq cn^{-\rho}L_n$ for $v \in S_n^+$. □

References

Chamberlain, G. (1986), "Asymptotic efficiency in semi-parametric models with censoring," *Journal of Econometrics*, 32, 189–218.

Chernozhukov, V. (2005), "Extremal quantile regression," *Annals of Statistics*, 33, 806–839.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014), "Gaussian approximation of suprema of empirical processes," *Annals of Statistics*, 42, 1564–1597.

Dekkers, A. L. and De Haan, L. (1989), "On the estimation of the extreme-value index and large quantile estimation," *Annals of Statistics*, 17, 1795–1832.

D'Haultfoeuille, X., Maurel, A., and Zhang, Y. (2016), "Extremal quantile regressions for selection models and the black-white wage gap," Working Paper.

Durrett, R. (2010), *Probability: Theory and Examples*, Cambridge University Press.

Folland, G. B. (1995), *Introduction to Partial Differential Equations*, Princeton University Press.

Khan, S. and Tamer, E. (2010), "Irregular identification, support conditions, and inverse weight estimation," *Econometrica*, 78, 2021–2042.

Pollard, D. (1991), "Asymptotics for least absolute deviation regression estimators," *Econometric Theory*, 7, 186–199.

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989), "Semiparametric estimation of index coefficients," *Econometrica*, 57, 1403–1430.

Resnick, S. I. (2007), *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*, Springer.

Serfling, R. J. (2009), *Approximation Theorems of Mathematical Statistics*, vol. 162, John Wiley & Sons.

Tan, L. and Zhang, Y. (2017), "Root-n consistency of the intercept of a binary response model under tail restrictions," *Econometric Theory*, (typesetter to insert the DOI).