CENSORED QUANTILE REGRESSION WITH VARYING COEFFICIENTS

Guosheng Yin[1], Donglin Zeng[2], and Hui Li[3]

The University of Hong Kong[1]
The University of North Carolina at Chapel Hill[2]
Beijing Normal University[3]

Supplementary Material

This supplementary material gives the detailed proofs of the theorems in the paper.

S1 Useful Lemmas

We define a VW_n class as a class of functions $\mathcal{F}_n$ indexed by $\theta \in \Theta$ which satisfies the conditions in (2.11.21) and the conditions of Theorem 2.11.22 in van der Vaart and Wellner (1996). That is, the envelope function F_n satisfies

$$P^* F_n^2 = O(1),$$

$$P^* F_n^2 I\left(F_n > \sqrt{n}\eta\right) \to 0, \quad \forall\ \eta > 0,$$

$$\sup_{\rho(\theta_1,\theta_2)<\delta_n} P(f_{n,\theta_1} - f_{n,\theta_2})^2 \to 0, \quad \forall\ \delta_n \downarrow 0,$$

where ρ is a metric for θ so that (Θ, ρ) is total bounded. Moreover,

$$\int \sup_{\mathcal{Q}} \sqrt{\log \mathcal{N}(\epsilon \|F_n\|_{\mathcal{Q},2}, \mathcal{F}_n, L_2(\mathcal{Q}))} \to 0, \quad \forall\ \delta_n \downarrow 0,$$

where $\mathcal{N}$ is the covering number and $\mathcal{Q}$ is the discrete probability measure.

We divide the proofs of Theorems 1 and 2 into a sequence of steps. Without loss of generality, we assume that $\mathbf{Z}$ and W are bounded by 1. Let $\mathbb{P}_n$ and $\mathbb{P}$ denote the empirical measure of n i.i.d. observations and the expectation, respectively; i.e., for any measurable function $f(Y)$ in $L_2(P)$,

$$\mathbb{P}_n f(Y) = n^{-1} \sum_{i=1}^{n} f(Y_i), \quad \mathbb{P} f(Y) = E f(Y).$$

Let $\mathbb{G}_n$ denote the empirical process based on n i.i.d. observations, i.e., $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - \mathbb{P})$.

For $k = 0, \ldots, r$, we define

$$\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi}, t) = \frac{1}{nh_n}\sum_{i=1}^n K_{h_n}(W_i - w_0)\left(\frac{W_i - w_0}{h_n}\right)^k \mathbf{Z}_i I(\log X_i - \boldsymbol{\xi}^T\mathbf{Z}_i^* \geq t),$$

and

$$\mathbf{A}_n^{(k)}(\boldsymbol{\xi}, t) = \frac{1}{h_n}E\left\{K_{h_n}(W - w_0)\left(\frac{W - w_0}{h_n}\right)^k \mathbf{Z} I(\log X - \boldsymbol{\xi}^T\mathbf{Z}^* \geq t)\right\}.$$

Clearly, we can write

$$\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi}, t) = \frac{1}{\sqrt{nh_n}}\mathbb{G}_n \mathbf{f}_n^{(k)}(\Delta, X, W, \mathbf{Z}; \boldsymbol{\xi}, t) + \mathbf{A}_n^{(k)}(\boldsymbol{\xi}, t), \tag{A.1}$$

where

$$\mathbf{f}_n^{(k)}(\Delta, X, W, \mathbf{Z}; \boldsymbol{\xi}, t) = \frac{1}{\sqrt{h_n}}K_{h_n}(W - w_0)\left(\frac{W - w_0}{h_n}\right)^k \mathbf{Z} I(\log X - \boldsymbol{\xi}^T\mathbf{Z}^* \geq t).$$

Let t_0 be the upper bound of $(\log X - \boldsymbol{\xi}_0^T\mathbf{Z}^*)$, and we consider a class

$$\mathcal{F}_{1k}^n = \left\{\mathbf{f}_n^{(k)}(\Delta, X, W, \mathbf{Z}; \boldsymbol{\xi}, t) : \|\boldsymbol{\xi} - \boldsymbol{\xi}_0\| < \delta_n, t \in (-\infty, t_0)\right\},$$

for some small constant δ_n. We first state two lemmas followed with their proofs, respectively.

Lemma S.1 $\mathcal{F}_{1k}^n$ is a VW_n class.

Proof. Clearly, an envelope function is given by

$$F_{nk} = \frac{1}{\sqrt{h_n}}K_{h_n}(W - w_0),$$

and thus, $E(F_{nk}^2) \leq O(1)h_n^{-1}E\{K_{h_n}(W - w_0)^2\} = O(1)$. In addition,

$$E\left\{F_{nk}^2 I(F_{nk} \geq \sqrt{n}\eta)\right\} \to 0,$$

as $\sqrt{nh_n} \to \infty$. It holds that for any two functions f_{n1} and f_{n2} indexed by $(\boldsymbol{\xi}_1, t_1)$ and $(\boldsymbol{\xi}_2, t_2)$,

$$\begin{aligned}
& P(f_{n1} - f_{n2})^2 \\
\leq\ & h_n^{-1}E\left\{K_{h_n}(W - w_0)^2\|\mathbf{Z}\|^2 |I(\log X - \boldsymbol{\xi}_1^T\mathbf{Z}^* \geq t_1) - I(\log X - \boldsymbol{\xi}_2^T\mathbf{Z}^* \geq t_2)|\right\} \\
\leq\ & O(1)E\left\{|I(\log X - \boldsymbol{\xi}_1^T\mathbf{Z}^* \geq t_1) - I(\log X - \boldsymbol{\xi}_2^T\mathbf{Z}^* \geq t_2)|\right\},
\end{aligned}$$

according to condition (C1). We can choose $\rho((\boldsymbol{\xi}_1, t_1), (\boldsymbol{\xi}_2, t_2))$ as defined at the right-hand side of the above inequality, which is a totally bounded semimetric as implied in the following arguments. On the other hand,

$$\left\{\log X - \boldsymbol{\xi}^T \mathbf{Z}^* - t : \|\boldsymbol{\xi} - \boldsymbol{\xi}_0\| < \delta_n,\ t \in [0, t_0]\right\}$$

is a VC class, and so is $\left\{\log X - \boldsymbol{\xi}^T \mathbf{Z}^* - t \geq 0 : \|\boldsymbol{\xi} - \boldsymbol{\xi}_0\| < \delta_n, t \in [0, t_0]\right\}$. Therefore, the uniform entropy integral condition holds for this class, and it also holds for $\mathcal{F}_{1k}^n$ by a factor $\|F_{nk}\|_{L_2(\mathcal{Q})}$ for any probability measure $\mathcal{Q}$. In conclusion, as the covariance for any two functions in $\mathcal{F}_{1k}^n$ has a limit, we have that $\sqrt{nh_n}\left\{\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi}, t) - \mathbf{A}_n^{(k)}(\boldsymbol{\xi}, t)\right\}$ converges weakly to a tight Gaussian process uniformly in $\boldsymbol{\xi}$ and t.

We now consider the $(k+1)$th component of $\mathbf{U}_n(\boldsymbol{\xi})$,

$$\mathbf{U}_n^{(k)}(\boldsymbol{\xi}) = \frac{1}{nh_n} \sum_{i=1}^n \frac{K_{h_n}(W_i - w_0)((W_i - w_0)/h_n)^k \mathbf{Z}_i \Delta_i I(\log X_i - \boldsymbol{\xi}^T \mathbf{Z}_i^* \leq 0)}{\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi}, \log X_i - \boldsymbol{\xi}^T \mathbf{Z}_i^*)} + \log(1-\tau),$$

where hereafter the division and multiplication of vectors are componentwise, unless otherwise noted.

Using the empirical process notation and expansion (A.1), we can write

$$\begin{aligned}
\mathbf{U}_n^{(k)}(\boldsymbol{\xi}) &= \frac{1}{\sqrt{nh_n}} \mathbb{G}_n \mathbf{S}_n^{(k)}(\Delta, X, W, \mathbf{Z}; \boldsymbol{\xi}) \\
&\quad + \frac{1}{h_n} E\left\{\frac{K_{h_n}(W - w_0)((W - w_0)/h_n)^k \mathbf{Z} \Delta I(\log X - \boldsymbol{\xi}^T \mathbf{Z}^* \leq 0)}{\mathbf{A}_n^{(k)}(\boldsymbol{\xi}, \log X - \boldsymbol{\xi}^T \mathbf{Z}^*)}\right\} + \log(1-\tau),
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{S}_n^{(k)}(\Delta, X, W, \mathbf{Z}; \boldsymbol{\xi}) &= \frac{1}{\sqrt{h_n}} \frac{K_{h_n}(W - w_0)((W - w_0)/h_n)^k \mathbf{Z} \Delta I(\log X - \boldsymbol{\xi}^T \mathbf{Z}^* \leq 0)}{\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi}, \log X - \boldsymbol{\xi}^T \mathbf{Z}^*)} \\
-\tilde{E}&\left\{\frac{1}{h_n} \frac{K_{h_n}(\tilde{W} - w_0)((\tilde{W} - w_0)/h_n)^k \tilde{\mathbf{Z}} \tilde{\Delta} I(\log \tilde{X} - \boldsymbol{\xi}^T \tilde{\mathbf{Z}} \leq 0)}{\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi}, \log \tilde{X} - \boldsymbol{\xi}^T \tilde{\mathbf{Z}}) \mathbf{A}_n^{(k)}(\boldsymbol{\xi}, \log \tilde{X} - \boldsymbol{\xi}^T \tilde{\mathbf{Z}})} \mathbf{f}_n^{(k)}(\Delta, X, W, \mathbf{Z}; \boldsymbol{\xi}, \log \tilde{X} - \boldsymbol{\xi}^T \tilde{\mathbf{Z}})\right\},
\end{aligned}$$

and $\tilde{E}$ takes the expectation with respect to $(\tilde{\Delta}, \tilde{X}, \tilde{W}, \tilde{\mathbf{Z}})$. Let $\mathcal{F}_{2k}^n$ be a class that contains all the functions $\mathbf{S}_n^{(k)}(\Delta, X, W, \mathbf{Z}; \boldsymbol{\xi})$ for $\|\boldsymbol{\xi} - \boldsymbol{\xi}_0\| < \delta_n$.

Lemma S.2 $\mathcal{F}_{2k}^n$ is a VW_n class.

Proof. To prove this lemma, we first show

$$\mathcal{F}_{3k}^n = \left\{\frac{1}{\sqrt{h_n}} \frac{K_{h_n}(W - w_0)((W - w_0)/h_n)^k \mathbf{Z} \Delta I(\log X_i - \boldsymbol{\xi}^T \mathbf{Z}_i^* \leq 0)}{\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi}, \log X - \boldsymbol{\xi}^T \mathbf{Z}^*)} : \|\boldsymbol{\xi} - \boldsymbol{\xi}_0\| < \delta_n\right\}$$

is a VW_n class. Following similar arguments as in the proof of Lemma S.1, we can easily verify that $\mathcal{F}_{3k}^n$ satisfies condition (2.11.21) in van der Vaart and Wellner (1996) with an envelope function $O(h_n^{-1/2}K_{h_n}(W-w_0))$, by noting that both $\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi},t)$ and $\mathbf{A}_n^{(k)}(\boldsymbol{\xi},t)$ converge to

$$\mu_k E\left\{\mathbf{Z}I(\log X-\boldsymbol{\xi}^T\mathbf{Z}^*\geq t)|W=w_0\right\},$$

uniformly in $\boldsymbol{\xi}$ and t with the limit bounded away from zero.

It remains to verify the uniform entropy condition in Theorem 2.11.22 in van der Vaart and Wellner (1996). We first argue that $\left\{\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi},\log X-\boldsymbol{\xi}^T\mathbf{Z}^*):\|\boldsymbol{\xi}-\boldsymbol{\xi}_0\|<\delta_n\right\}$ has a finite uniform entropy integral. To see this, we note that

$$\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi},\log X-\boldsymbol{\xi}^T\mathbf{Z}^*)\left\{\frac{1}{nh_n}\sum_{i=1}^n K_{h_n}(W_i-w_0)\left(\frac{W_i-w_0}{h_n}\right)^k\right\}^{-1}$$

is a convex combination of function $I(\log X-\boldsymbol{\xi}^T\mathbf{Z}^*\leq t)$ for $t=\log X_i-\boldsymbol{\xi}^T\mathbf{Z}_i^*$. Moreover,

$$\frac{1}{nh_n}\sum_{i=1}^n K_{h_n}(W_i-w_0)\left(\frac{W_i-w_0}{h_n}\right)^k$$

has a non-zero limit. Thus, $\left\{\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi},\log X-\boldsymbol{\xi}^T\mathbf{Z}^*):\|\boldsymbol{\xi}-\boldsymbol{\xi}_0\|<\delta_n\right\}$ has a finite uniform entropy integral according to Theorem 2.6.9 in van der Vaart and Wellner (1996). Since the limit of $\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi},t)$ is bounded from zero, the same conclusion holds for $\left\{\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi},\log X-\boldsymbol{\xi}^T\mathbf{Z}^*)^{-1}:\|\boldsymbol{\xi}-\boldsymbol{\xi}_0\|<\delta_n\right\}$. It is then easy to see that $\mathcal{F}_{3k}^n$ has a finite uniform entropy integral.

The same arguments apply to the class containing

$$\tilde{E}\left\{\frac{K_{h_n}(\tilde{W}-w_0)((\tilde{W}-w_0)/h_n)^k\tilde{\mathbf{Z}}\tilde{\Delta}I(\log\tilde{X}-\boldsymbol{\xi}^T\tilde{\mathbf{Z}}\leq 0)}{h_n\hat{\mathbf{A}}_n^{(k)}(\boldsymbol{\xi},\log\tilde{X}-\boldsymbol{\xi}^T\tilde{\mathbf{Z}})\mathbf{A}_n^{(k)}(\boldsymbol{\xi},\log\tilde{X}-\boldsymbol{\xi}^T\tilde{\mathbf{Z}})}\mathbf{f}_n^{(k)}(\Delta,X,W,\mathbf{Z};\boldsymbol{\xi},\log\tilde{X}-\boldsymbol{\xi}^T\tilde{\mathbf{Z}})\right\}.$$

Particularly, from condition (C2), the function in this class is Lipschitz with respect to ξ with the Lipschitz coefficient bounded by $O(h_n^{-1/2}K_{h_n}(W-w_0))$. Thus, the uniform entropy integral condition holds for this class, and Lemma S.2 holds.

Next we will prove Theorems 1 and 2, respectively.

S2 Proof of Theorem 1

We have that

$$\mathbf{U}_n(\boldsymbol{\xi})=\frac{1}{\sqrt{nh_n}}\mathbb{G}_n\mathbf{S}_n(\boldsymbol{\xi})+\mathbf{R}_n(\boldsymbol{\xi}),$$

where $\mathbf{S}_n(\boldsymbol{\xi}) = \{\mathbf{S}_n^{(0)}(\boldsymbol{\xi})^T, \ldots, \mathbf{S}_n^{(r)}(\boldsymbol{\xi})^T\}^T$, and $\mathbf{R}_n(\boldsymbol{\xi}) = \{\mathbf{R}_n^{(0)}(\boldsymbol{\xi})^T, \ldots, \mathbf{R}_n^{(r)}(\boldsymbol{\xi})^T\}^T$ with

$$\mathbf{R}_n^{(k)}(\boldsymbol{\xi}) = \frac{1}{h_n} E\left\{\frac{K_{h_n}(W-w_0)((W-w_0)/h_n)^k \mathbf{Z}\Delta I(\log X - \boldsymbol{\xi}^T\mathbf{Z}^* \leq 0)}{\mathbf{A}_n^{(k)}(\boldsymbol{\xi}, \log X - \boldsymbol{\xi}^T\mathbf{Z}^*)}\right\} + \log(1-\tau).$$

Furthermore, it is straightforward to verify that the covariance function of $\mathbf{S}_n^{(k)}(\boldsymbol{\xi})$ has a finite limit. Theorem 2.11.22 in van der Vaart and Wellner (1996) yields that $\mathbb{G}_n\mathbf{S}_n(\boldsymbol{\xi})$ converges weakly to a normal distribution with variance-covariance matrix $\boldsymbol{\Sigma}_1$.

We perform the Taylor series expansion of $\mathbf{R}_n(\boldsymbol{\xi})$ in a $(nh_n)^{-1/2}$-neighborhood of $\boldsymbol{\xi}_0$,

$$\begin{aligned}
&\mathbf{R}_n(\boldsymbol{\xi}) \\
= \;& \mathbf{D}\mathbf{H}(\boldsymbol{\xi}-\boldsymbol{\xi}_0) + E\left[\frac{h_n^{-1}K_{h_n}(W-w_0)((W-w_0)/h_n)^k\mathbf{Z}\Delta I(\log X - \boldsymbol{\xi}_0^T\mathbf{Z}^* \leq 0)}{\tilde{E}\{h_n^{-1}K_{h_n}(\tilde{W}-w_0)((\tilde{W}-w_0)/h_n)^k\tilde{\mathbf{Z}}I(\log\tilde{X} - \boldsymbol{\xi}_0^T\tilde{\mathbf{Z}}^* \geq \log X - \boldsymbol{\xi}_0^T\mathbf{Z}^*)\}}\right] \\
&+ \log(1-\tau) + o(\|\mathbf{H}(\boldsymbol{\xi}-\boldsymbol{\xi}_0)\|),
\end{aligned}$$

where

$$\mathbf{D} = \mathbf{I}_{(r+1)\times(r+1)} \otimes \frac{\partial}{\partial\boldsymbol{\beta}^T} E\left[\frac{\mathbf{Z}\Delta I(\log X - \boldsymbol{\beta}^T\mathbf{Z} \leq 0)}{\tilde{E}\{I(\log\tilde{X} - \boldsymbol{\beta}^T\tilde{\mathbf{Z}} \geq \log X - \boldsymbol{\beta}^T\mathbf{Z})|\tilde{W} = w_0\}}|W = w_0\right],$$

evaluated at $\boldsymbol{\beta} = \boldsymbol{\beta}_0(w_0)$. We also note that

$$E\left[\frac{g(W)\mathbf{Z}\Delta I(\log X - \boldsymbol{\beta}_0(W)^T\mathbf{Z} \leq 0)}{\tilde{E}\{g(\tilde{W})\tilde{\mathbf{Z}}I(\log\tilde{X} - \boldsymbol{\beta}_0(\tilde{W})^T\tilde{\mathbf{Z}} \geq \log X - \boldsymbol{\beta}_0(W)^T\mathbf{Z})\}}\right] + \log(1-\tau) = 0,$$

for any measurable function $g(W) \in L_2(P)$, and from condition (C7),

$$\boldsymbol{\xi}_0^T\mathbf{Z}^* - \boldsymbol{\beta}_0(W)^T\mathbf{Z} = -\boldsymbol{\beta}_0^{[r+1]}(w_0)^T\mathbf{Z}(W-w_0)^{r+1} + o(h_n^{r+1}),$$

whenever $W - w_0 = O(h_n)$. Hence, we have

$$\begin{aligned}
&E\left[\frac{h_n^{-1}K_{h_n}(W-w_0)((W-w_0)/h_n)^k\mathbf{Z}\Delta I(\log X - \boldsymbol{\xi}_0^T\mathbf{Z}^* \leq 0)}{\tilde{E}\{h_n^{-1}K_{h_n}(\tilde{W}-w_0)((\tilde{W}-w_0)/h_n)^k\tilde{\mathbf{Z}}I(\log\tilde{X} - \boldsymbol{\xi}_0^T\tilde{\mathbf{Z}}^* \geq \log X - \boldsymbol{\xi}_0^T\mathbf{Z}^*)\}}\right] \\
= \;& E\left[\frac{h_n^{-1}K_{h_n}(W-w_0)((W-w_0)/h_n)^k\mathbf{Z}\Delta I(\log X - \boldsymbol{\beta}_0(W)^T\mathbf{Z} \leq 0)}{\tilde{E}\{h_n^{-1}K_{h_n}(\tilde{W}-w_0)((\tilde{W}-w_0)/h_n)^k\tilde{\mathbf{Z}}I(\log\tilde{X} - \boldsymbol{\beta}_0(\tilde{W})^T\tilde{\mathbf{Z}} \geq \log X - \boldsymbol{\beta}_0(W)^T\mathbf{Z})\}}\right] \\
&+ h_n^{r+1}\mathbf{b}_k + o(h_n^{r+1}) \\
= \;& -\log(1-\tau) + h_n^{r+1}\mathbf{b}_k + o(h_n^{r+1}),
\end{aligned}$$

with the constant

$$
\begin{aligned}
&\mathbf{b}_k\\
= \;& \frac{\mu_{k+r+1}}{\mu_k} E\left(\frac{\partial}{\partial\zeta} E\left[\frac{\mathbf{Z}\Delta I(\log X-\zeta\le 0)\boldsymbol{\beta}_0^{[r+1]}(w_0)^T\mathbf{Z}}{\tilde{E}\{I(\log\tilde{X}-\boldsymbol{\beta}_0(w_0)^T\tilde{\mathbf{Z}}\ge\log X-\zeta)|\tilde{W}=w_0\}}\Big|W=w_0,\mathbf{Z}\right]\Big|W=w_0\right)\\
&+\frac{\mu_{k+r+1}}{\mu_k}E\Bigg(\mathbf{Z}\Delta I(\log X-\boldsymbol{\beta}_0(w_0)^T\mathbf{Z}\le 0)\\
&\times\frac{\tilde{E}[\frac{\partial}{\partial\tilde{\zeta}}\tilde{E}\{I(\log\tilde{X}-\tilde{\zeta}\ge\log X-\boldsymbol{\beta}_0(w_0)^T\mathbf{Z})\boldsymbol{\beta}_0^{[r+1]}(w_0)^T\tilde{\mathbf{Z}}|\tilde{W}=w_0,\tilde{\mathbf{Z}}\}|\tilde{W}=w_0]}{\tilde{E}\{I(\log\tilde{X}-\tilde{\zeta}\ge\log X-\boldsymbol{\beta}_0(w_0)^T\mathbf{Z})|\tilde{W}=w_0\}^2}\Big|W=w_0\Bigg),
\end{aligned}
$$

evaluated at $\zeta=\tilde{\zeta}=\boldsymbol{\beta}_0(w_0)^T\mathbf{Z}$. If we define $\mathbf{b}=(\mathbf{b}_0^T,\ldots,\mathbf{b}_r^T)^T$, then

$$\mathbf{R}_n(\boldsymbol{\xi})=\mathbf{DH}(\boldsymbol{\xi}-\boldsymbol{\xi}_0)+h_n^{r+1}\mathbf{b}+o(h_n^{r+1}+\|\mathbf{H}(\boldsymbol{\xi}-\boldsymbol{\xi}_0)\|).$$

We now study the asymptotic behavior of the covariance matrix $\boldsymbol{\Omega}_n(\boldsymbol{\xi})$. For the first term in $\boldsymbol{\Omega}_n(\boldsymbol{\xi})$, $n^{-1}\sum_{i=1}^n\mathbf{u}_i(\boldsymbol{\xi})\mathbf{u}_i(\boldsymbol{\xi})^T$, each $p\times p$ submatrix has a form of

$$
\begin{aligned}
&(nh_n)^{-1}\sum_{i=1}^n\Bigg[h_n^{-1}K_{h_n}(W_i-w_0)^2\left(\frac{W_i-w_0}{h_n}\right)^{k+j}\Delta_i I(\log X_i-\boldsymbol{\xi}^T\mathbf{Z}_i^*\le 0)\\
&\qquad\times\left\{\frac{\mathbf{Z}_i}{\mathbf{A}_n^{(k)}(\boldsymbol{\xi},\log X_i-\boldsymbol{\xi}^T\mathbf{Z}_i^*)}\right\}\left\{\frac{\mathbf{Z}_i}{\mathbf{A}_n^{(j)}(\boldsymbol{\xi},\log X_i-\boldsymbol{\xi}^T\mathbf{Z}_i^*)}\right\}^T\Bigg],
\end{aligned}
$$

for $k,j=0,\ldots,r$. From Lemma S.1, the first term of $\boldsymbol{\Omega}_n(\boldsymbol{\xi})$ is equivalent to $\boldsymbol{\Sigma}_2(1+o_p(1))/h_n$ for some positive definite matrix $\boldsymbol{\Sigma}_2$. The second term of $\boldsymbol{\Omega}_n(\boldsymbol{\xi})$ is $\mathbf{U}_n(\boldsymbol{\xi})\mathbf{U}_n(\boldsymbol{\xi})^T=o_p(1)$.

Finally, we have that

$$
\begin{aligned}
Q_n(\boldsymbol{\xi}) \;=\; & \left\{(nh_n)^{-1/2}\mathbb{G}_n\mathbf{S}_n(\boldsymbol{\xi}_0)+\mathbf{DH}(\boldsymbol{\xi}-\boldsymbol{\xi}_0)+h_n^{r+1}\mathbf{b}+o(h_n^{r+1}+\|\mathbf{H}(\boldsymbol{\xi}-\boldsymbol{\xi}_0)\|)\right\}^T\\
&\times(\boldsymbol{\Sigma}_2/h_n+o_p(1))^{-1}\\
&\times\left\{(nh_n)^{-1/2}\mathbb{G}_n\mathbf{S}_n(\boldsymbol{\xi}_0)+\mathbf{DH}(\boldsymbol{\xi}-\boldsymbol{\xi}_0)+h_n^{r+1}\mathbf{b}+o(h_n^{r+1}+\|\mathbf{H}(\boldsymbol{\xi}-\boldsymbol{\xi}_0)\|)\right\},
\end{aligned}
$$

in a $(nh_n)^{-1/2}$-neighborhood of $\boldsymbol{\xi}_0$. Therefore, if we let $\boldsymbol{\xi}=\boldsymbol{\xi}_0+(nh_n)^{-1/2}y$, then $(nh_n^2)Q_n(\boldsymbol{\xi}_0+(nh_n)^{-1/2}y)$ converges uniformly for y in any compact set to a quadratic function with a unique minimum at $y=0$. We thus conclude that there exists a local minimum $\hat{\boldsymbol{\xi}}_n$ for minimizing $Q_n(\boldsymbol{\xi})$ and $\hat{\boldsymbol{\xi}}_n=\boldsymbol{\xi}_0+O_p((nh_n)^{-1/2})$. This completes the proof of Theorem 1.

S3 Proof of Theorem 2

Applying the standard M-estimation theory such as Theorem 3.2.16 in van der Vaart and Wellner (1996), we have that

$$(nh_n)^{-1/2}\mathbb{G}_n\mathbf{S}_n(\boldsymbol{\xi}_0) + \mathbf{DH}(\boldsymbol{\xi} - \boldsymbol{\xi}_0) + h_n^{r+1}\mathbf{b} + o\left(h_n^{r+1} + \frac{1}{\sqrt{nh_n}}\right) = o_p(1).$$

Hence, Theorem 2 holds with $\boldsymbol{\Sigma} = \mathbf{D}^{-1}\boldsymbol{\Sigma}_1\mathbf{D}^{-1}$.

References

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* New York: Springer.