Supplementary Material on "The Bootstrap in Threshold Regression"

1. No Error Term: An Illustration

This section uses a simple example to illustrate the main results of this paper. The speciality of this example is that finite-sample distributions of all estimators are available, but in the general case, such distributions are hard to derive. In construction of this paper, this example provides the direction of conjecturing the general results. We put this example here only to make the results in the main text more expected. Nevertheless, connecting threshold regression with the existing boundary literature in the bootstrap environment seems novel. As in the main text, we still use Z and Z^* to represent the asymptotic distribution and the asymptotic bootstrap distribution of the γ estimators, respectively. Subscripts L and M are used to indicate the objects based on the LLSE and the MLSE, respectively. We adopt an unusual notation of conditioning: $|F_n$, $|\widehat{\gamma}$ and $|D(\cdot)$ all indicate conditioning on the original sample path instead of on the σ-field generated by F_n, $\widehat{\gamma}$ and $D(\cdot)$ (which are defined below). Such notations are used in some statistical literature such as Bickel and Freedman (1981).

We simplify (1) in the main text to the extreme case as follows:

$$y = \mathbf{1}(q \leq \gamma), \ q \sim U[0,1]. \tag{1}$$

This corresponds to $x = 1$, $\beta_{1,0} = 1$, $\beta_{2,0} = 0$, $\sigma_{1,0} = \sigma_{2,0} = 0$ in (1) of the main text. Here, q follows a uniform distribution on $[0,1]$, and $\gamma_0 = 1/2$ is of main interest. Note that there is no error term ε in (1), so the observed y values can only be 0 or 1. In this case, the threshold point is essentially a "middle" boundary of q because there is a sharp change in y values when q switches from the left to the right side of γ_0; see Section 2 of Yu (2012) for a detailed analysis of this point. It is easy to see that the LLSE and LMLE are the same, denoted as $\widehat{\gamma}_L$, and equals the q_i closest to 1/2 from the left.[1] For $t < 0$,

$$P\left(n\left(\widehat{\gamma}_L - \gamma_0\right) \leq t\right) = P\left(q_i \notin (\gamma_0 + \frac{t}{n}, \gamma_0] \text{ for all } i\right) = \left(1 + \frac{t}{n}\right)^n \to e^t, \tag{2}$$

so the asymptotic distribution of $n\left(\widehat{\gamma}_L - \gamma_0\right)$ is a negative standard exponential, and there is no density on the positive axis. Note further that $\widehat{\gamma}_L$ is a nondecreasing function of n conditional on the original sample path, and there is no data point between $\widehat{\gamma}_L$ and γ_0. Similarly, the MLSE and MMLE are the same, denoted as $\widehat{\gamma}_M$, and equal the average of the two q_i's closest to 1/2 from the left and right. Suppose there are m y_i's taking value 1, and the remaining $(n-m)$ y_i's take value 0, then $\widehat{\gamma}_L = q_{(m)}$ and $\widehat{\gamma}_M = \frac{q_{(m)}+q_{(m+1)}}{2}$, where

[1] Strictly speaking, $\widehat{\gamma}_L$ depends on whether there exists q_i no greater than 1/2 or not. If there is q_i no greater than 1/2, then $\widehat{\gamma}_L$ equals the q_i closest to 1/2 from the left. Otherwise, $\widehat{\gamma}_L$ equals the q_i closest to 1/2 from the right. Since the probability that all q_i's are greater than 1/2 equals $\left(\frac{1}{2}\right)^n$ which converges to zero, we can assume $\widehat{\gamma}_L \leq \frac{1}{2}$ in the following discussion.

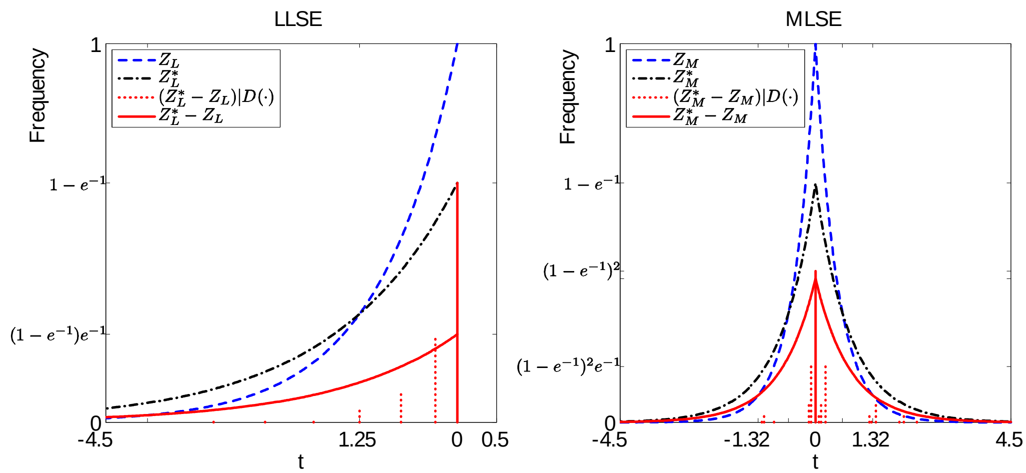

Figure 1: Z, Z^*, $(Z^* - Z)\,|D(\cdot)$ and $Z^* - Z$ for the LLSE and MLSE

$q_{(m)}$ is the mth order statistic of $\{q_i\}_{i=1}^n$. By a similar analysis as in (2),

$$P\left(n\left(q_{(m+1)} - \gamma_0\right) > t\right) = \left(1 - \frac{t}{n}\right)^n \to e^{-t}, \tag{3}$$

for $t \geq 0$. So for $t < 0$,

$$\begin{aligned} P\left(n\left(\widehat{\gamma}_M - \gamma_0\right) \leq t\right) &= P\left(\frac{n\left(q_{(m)} - \gamma_0\right) + n\left(q_{(m+1)} - \gamma_0\right)}{2} \leq t\right) \\ &= P\left(n\left(q_{(m)} - \gamma_0\right) \leq 2t - n\left(q_{(m+1)} - \gamma_0\right)\right) \to \int_0^\infty e^{2t-s} d\left(1 - e^{-s}\right) = \frac{e^{2t}}{2}, \end{aligned} \tag{4}$$

where the convolution form in the convergence step is from the fact that $q_{(m+1)}$ and $q_{(m)}$ are independent. Similarly, for $t \geq 0$, $P\left(n\left(\widehat{\gamma}_M - \gamma_0\right) > t\right) \to \frac{e^{-2t}}{2}$. So $n\left(\widehat{\gamma}_M - \gamma_0\right)$ converges to the double exponential distribution with scale $1/2$.

1.1 Invalidity of the Nonparametric Bootstrap

The objective function of the least squares estimation is $\sum_{i=1}^n \left(y_i - \mathbf{1}(q_i \leq \gamma)\right)^2$. To consider the ability of the bootstrap to approximate the distribution of $n\left(\widehat{\gamma}_L - \gamma_0\right)$, we need to obtain the asymptotic distribution of $n\left(\widehat{\gamma}_L^* - \widehat{\gamma}_L\right)$, where $\widehat{\gamma}_L^* = \arg\min_{\gamma} \sum_{i=1}^n \left(y_i^* - \mathbf{1}(q_i^* \leq \gamma)\right)^2$ is the closest q_i^* to $1/2$ from the left, and $(y_i^*, q_i^*)'$ follows the empirical distribution F_n. Since γ is essentially a boundary, the following derivation is similar to that in Example 3 of Bickel et al. (1997).

According to Chan (1993), $\widehat{\gamma}_L$ converges to $1/2$ as n goes to infinity for almost every sample point ω in Ω. In bootstrap sampling, as long as $\left(q_{(m)}, y_{(m)}\right)$ is drawn, $\widehat{\gamma}_L^* = q_{(m)}$. So $P^*\left(n\left(\widehat{\gamma}_L^* - \widehat{\gamma}_L\right) = 0|F_n\right) =$

$1 - P^*\left(\left(q_{(m)}, y_{(m)}\right) \text{ is not drawn}\right) = 1 - \left(1 - \frac{1}{n}\right)^n \to 1 - e^{-1} > 0,$[2] while $P\left(n\left(\widehat{\gamma}_L - 1/2\right) = 0\right) \to 0$ since the asymptotic distribution of $n\left(\widehat{\gamma}_L - 1/2\right)$ is continuous. Therefore, the bootstrap is not consistent. In the general case, there is no explicit form for the limit of $P^*\left(n\left(\widehat{\gamma}_L^* - \widehat{\gamma}_L\right) = 0 | F_n\right)$ to check whether the bootstrap is valid, so here we analyze the whole distribution of $n\left(\widehat{\gamma}_L^* - \widehat{\gamma}_L\right)|F_n$ to provide more intuitions.

Actually, the limit distribution of $n\left(\widehat{\gamma}_L^* - \widehat{\gamma}_L\right)|F_n$ does not exist. Suppose it indeed exists, then for any $t < 0$,

$$P^*\left(n\left(\widehat{\gamma}_L^* - \widehat{\gamma}_L\right) < t \middle| F_n\right) = P_n^*\left(\text{no } q_i^* \text{ is sampled from } \left[\widehat{\gamma}_L + \frac{t}{n}, \widehat{\gamma}_L\right] \middle| F_n\right) = \left(1 - \frac{k}{n}\right)^n$$

must converge to a fixed value, where $k = \sum_{i=1}^n \mathbf{1}\left(\widehat{\gamma}_L + \frac{t}{n} \le q_i \le \widehat{\gamma}_L\right)$. This means that k must converge to a fixed value. But $\overline{\lim}_{n\to\infty} k = \infty$ and $\underline{\lim}_{n\to\infty} k = 0$ for any ω, so $P^*\left(n\left(\widehat{\gamma}_L^* - \widehat{\gamma}_L\right) < t | F_n\right)$ cannot converge. Nevertheless, we can find the weak limit of $n\left(\widehat{\gamma}_L^* - \widehat{\gamma}_L\right)$ under P_r. Conditional on $\widehat{\gamma}_L$, $k \sim 1 + \text{Bin}\left(n - 1, \frac{|t|/n}{1 - (1/2 - \widehat{\gamma}_L)}\right)$ converges weakly to $1 + N(|t|)$ for any $\widehat{\gamma}_L$, where $N(\cdot)$ is a standard Poisson process. So $P_r\left(n\left(\widehat{\gamma}_L^* - \widehat{\gamma}_L\right) < t\right)$ converges to $E\left[e^{-(1+N(|t|))}\right]$, which is the average of $e^{-(1+N(|t|))}$ for all realizations of $N(|t|)$, or the asymptotic distribution of the average bootstrap. For a given realization of $N(|t|)$, the implied distribution by the component measure $e^{-(1+N(|t|))}$ is discrete. A new jump in $e^{-(1+N(|t|))}$ happens as $|t|$ gets larger such that the expanding interval $\left[\widehat{\gamma}_L + \frac{t}{n}, \widehat{\gamma}_L\right]$ covers a new q_i. Because $e^{-(1+N(|t|))}|_{t=0} + \left(1 - e^{-1}\right) = 1$, there is no probability on the positive axis, which is similar to the asymptotic distribution. The left panel of Figure 1 shows a typical realization of $e^{-(1+N(|t|))}$. The density function of $E\left[e^{-(1+N(|t|))}\right]$ is also shown in Figure 1. It is obviously different from the asymptotic distribution since there is a point mass $1 - e^{-1}$ at zero.

The above results are surprising in two aspects. First, while the asymptotic distribution of the threshold estimator exists and is continuous, the conditional weak limit of the bootstrap estimator does not exist. This is different from conventional models where the asymptotic distribution is normal. Essentially, this is because the asymptotic bootstrap distribution of the threshold point relies on bootstrap sampling on the local data (i.e., $\widehat{\gamma} + \frac{t}{n} \le q_i \le \widehat{\gamma}$) rather than sampling on the whole dataset in conventional models. Second, the asymptotic distribution of the average bootstrap is a genuine mixture of discrete measures instead of a fixed measure as required by the bootstrap validity. Although the point mass at zero is always $1 - e^{-1}$ for all discrete measures, how to distribute the remaining e^{-1} probability depends on how the original data are sampled. When more data are sampled in the left neighborhood of 1/2, the point masses on the negative axis in Figure 1 will be closer to zero.

One important similarity between the discrete component measure in the average bootstrap and the asymptotic distribution is that both of them critically depend on the local information around the threshold point. The asymptotic distribution depends on the density of q at 1/2, while those discrete measures depend on the local data around $\widehat{\gamma}$ in the original sample. This is not difficult to understand by noting that the true distribution of q in the asymptotic theory is $f_q(\cdot)$ ($U[0,1]$ in this example), and the true value of γ is 1/2, while in the bootstrap, the true distribution of q is the empirical distribution of $\{q_i\}_{i=1}^n$, and the true value

[2] In the general case, this probability is not constant from Figure 1 in the main text. So from Hewitt-Savage zero-one law, $P^*\left(n\left(\widehat{\gamma}_L^* - \widehat{\gamma}_L\right) = 0 | F_n\right)$ cannot have a weak limit.

of γ is $\widehat{\gamma}$.

In summary, although this example is very simple, it shows one general feature of the bootstrap for the threshold point: the local information around γ_0 (or $\widehat{\gamma}$) is most important for the bootstrap inference. As a result, the conditional weak limit of the bootstrap estimator does not exist, and the component measures in the average bootstrap are discrete and depend on the original data. Therefore, the bootstrap of the threshold point is invalid.

Putting this example in the general framework of threshold regression,

$$D(v) = \begin{cases} N_1(|v|), & \text{if } v \le 0; \\ N_2(v), & \text{if } v > 0; \end{cases} \quad \text{and } D^*(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} N_{i-}^*, \text{ if } v \le 0; \\ \sum_{i=1}^{N_2(v)} N_{i+}^*, \text{ if } v > 0. \end{cases}$$

Now,

$$P^*\left(n\left(\widehat{\gamma}_L^* - \widehat{\gamma}_L\right) = 0 | F_n\right) \longrightarrow P^*\left(N_{1-}^* > 0\right) = 1 - e^{-1},$$

and for $t \le 0$, $P_r\left(n\left(\widehat{\gamma}_L^* - \widehat{\gamma}_L\right) < t\right)$ is the average of the following probabilities for different realizations of $N(|t|)$:

$$\begin{cases} P^*\left(N_{1-}^* = 0\right) = e^{-1}, & \text{if } N(|t|) = 0, \\ P^*\left(N_{1-}^* = 0, N_{2-}^* = 0\right) = e^{-2}, & \text{if } N(|t|) = 1, \\ \vdots & \vdots \\ P^*\left(\text{Poisson}(k+1) = 0\right) = e^{-(k+1)}, & \text{if } N(|t|) = k, \\ \vdots & \vdots \end{cases} = e^{-(1+N(|t|))},$$

where $N(|t|)$ is a *truncated Poisson process* starting from $t_0 \equiv \inf\{t : N_1(|t|) = 0\}$. This $N(|t|)$ is the same as the $N(|t|)$ above. The point mass implied in $e^{-(1+N(|t|))}$ at the kth jump of $D(v)$ on $v \le 0$ is

$$P^*\left(N_{j-}^* = 0 \text{ for } j \le k \text{ and } N_{(k+1)-}^* > 0\right) = e^{-k} \cdot \left(1 - e^{-1}\right),$$

which is exponentially decaying. Note also that under P_r, Z_L has a thinner tail than Z_L^* and $Z_L^* - Z_L$. For comparison, their densities on $t < 0$ are listed below:

$$f_{Z_L}(t) = e^t; \; f_{Z_L^*}(t) = \exp\{t - t/e\}\left(1 - e^{-1}\right); \; f_{Z_L^* - Z_L}(t) = \exp\{t - t/e - 1\}\left(1 - e^{-1}\right).$$

In the general case, the bootstrap still fails as shown in Section 3.2, but there are some differences in the component measure. First, there is positive probability on the positive axis. Second, not every jumping location of $D(\cdot)$ on $v \le 0$ corresponds to a point mass. Third, the probability mass function is not necessarily monotone on the negative axis. Fourth, the point mass at zero is not fixed as $1 - e^{-1}$.

If we consider the MLSE, then the asymptotic distribution of $n(\widehat{\gamma}_M^* - \widehat{\gamma}_M)$ under P_r is the average of the following discrete distributions:

$$P^*(Z_M^* - Z_M = 0|D(\cdot)) = P^*\left(N_{1-}^* > 0, N_{1+}^* > 0\right) = \left(1 - e^{-1}\right)^2,$$

$$P^*\left(Z_M^* - Z_M = \frac{T_{1+} - T_{2-}}{2} - \frac{T_{1+} - T_{1-}}{2}\middle| D(\cdot)\right) = P^*\left(N_{1-}^* = 0, N_{2-}^* > 0, N_{1+}^* > 0\right) = e^{-1}\left(1 - e^{-1}\right)^2,$$

$$P^*\left(Z_M^* - Z_M = \frac{T_{2+} - T_{1-}}{2} - \frac{T_{1+} - T_{1-}}{2}\middle| D(\cdot)\right) = P^*\left(N_{1+}^* = 0, N_{1-}^* > 0, N_{2+}^* > 0\right) = e^{-1}\left(1 - e^{-1}\right)^2,$$

$$P^*\left(Z_M^* - Z_M = \frac{T_{2+} - T_{2-}}{2} - \frac{T_{1+} - T_{1-}}{2}\middle| D(\cdot)\right) = P^*\left(N_{1-}^* = 0, N_{1+}^* = 0, N_{2-}^* > 0, N_{2+}^* > 0\right) = e^{-2}\left(1 - e^{-1}\right)^2,$$

$$\vdots$$

$$P^*\left(Z_M^* - Z_M = \frac{T_{k+} - T_{j-}}{2} - \frac{T_{1+} - T_{1-}}{2}\middle| D(\cdot)\right) = P^*\left(\text{Poisson}(k+j-2) = 0, N_{j-}^* > 0, N_{k+}^* > 0\right)$$
$$= e^{-(k+j-2)}\left(1 - e^{-1}\right)^2,$$

$$\vdots$$

where T_{k+} and T_{j-}, $k, j = 1, 2, \cdots$, are jumping locations of $D(\cdot)$. So the unconditional density of $Z_M^* - Z_M$ at $t \neq 0$ is

$$f_{Z_M^* - Z_M}(t) = \sum_{k,j=1,k+j>2}^{\infty} e^{-(k+j-2)}\left(1 - e^{-1}\right)^2 g_{k,j}(t),$$

where when $k, j \geq 2$, $g_{k,j}(t)$ is the density of $\frac{T_{(k-1)+} - T_{(j-1)-}}{2}$, which is $\frac{2^{k-j}e^{-2t}}{(j-2)!}\sum_{l=0}^{k-2}\frac{2^{-2l}}{l!(k-2-l)!}t^{k-2-l}\Gamma(l+j-1, 0 \vee (-4t))$ with $\Gamma(\cdot,\cdot)$ being the upper incomplete gamma function, and when $k = 1, j \geq 2$, $g_{k,j}(t)$ is $2^{j-1}|t|^{j-2}e^{2t}/(j-2)!\mathbf{1}(t < 0)$ and when $k \geq 2, j = 1$ is $2^{k-1}|t|^{k-2}e^{-2t}/(k-2)!\mathbf{1}(t > 0)$. So the unconditional density of $Z_M^* - Z_M$ at $t \neq 0$ is

$$f_{Z_M^* - Z_M}(t) = \left(1 - e^{-1}\right)^2 e^{-2|t|}\sum_{k=2}^{\infty}\frac{2^{k-1}e^{-(k-1)}}{(k-2)!}|t|^{k-2}$$
$$+ \left(1 - e^{-1}\right)^2 e^{-2t}\sum_{k=2}^{\infty}\sum_{j=2}^{\infty}\frac{2^{k-j}e^{-(k+j-2)}}{(j-2)!}\sum_{l=0}^{k-2}\frac{2^{-2l}}{l!(k-2-l)!}t^{k-2-l}\Gamma(l+j-1, 0 \vee (-4t)),$$

which is symmetric and whose tail is thicker than that of $f_{Z_M}(t) = e^{-2|t|}$. By a similar analysis, we can show that the unconditional density of Z_M^* is

$$f_{Z_M^*}(t) = \sum_{k,j=1}^{\infty} e^{-(k+j-2)}\left(1 - e^{-1}\right)^2 g_{k+1,j+1}(t) = \sum_{k,j=2}^{\infty} e^{-(k+j-4)}\left(1 - e^{-1}\right)^2 g_{k,j}(t)$$
$$= \left(1 - e^{-1}\right)^2 e^{-2t}e^2\sum_{k=2}^{\infty}\sum_{j=2}^{\infty}\frac{2^{k-j}e^{-(k+j-2)}}{(j-2)!}\sum_{l=0}^{k-2}\frac{2^{-2l}}{l!(k-2-l)!}t^{k-2-l}\Gamma(l+j-1, 0 \vee (-4t)),$$

which is also symmetric and whose tail is also thicker than that of f_{Z_M}. The right panel of Figure 1 shows the MLSE counterparts of the LLSE in the left panel.

1.2 Invalidity of the (Parametric) Wild Bootstrap

Because the distribution of the error term is a point mass at zero, each bootstrap sample coincides with the original sample. As a result, each bootstrap estimator is the same as the original MLE or LSE and does not contain any randomness when conditioning on the original data. Consequently, the bootstrap CI only includes the MLE or LSE itself and does not cover γ_0 for almost all original sample pathes. Putting in the general framework of threshold regression, we have $D_{wn}^{*}(v) = D_n(v)$ and $D_{pn}^{*}(v) = D_{pn}(v)$, so $Z^* = Z$ for any $D(\cdot)$ for both estimators.

In this simple example, there is no error term, so a sharp jump in the value of y appears at $q = \gamma_0$, which induces the invalidity of the above bootstrap procedures. When a nondegenerate error term is added in (1), the sharp difference of y values in the left and right neighborhoods of $q = \gamma_0$ is blurred. However, these bootstrap schemes still fail in inference on γ as shown in Section 3 of the main text. The following discusses two valid bootstrap schemes in this simple example: the parametric bootstrap and the smoothed bootstrap.

1.3 Validity of the Parametric Bootstrap and Smoothed Bootstrap

In parametric bootstrap sampling, the following DGP is used: $y = 1(q \le \widehat{\gamma})$, where $q \sim U[0,1]$ and $\widehat{\gamma}$ is the MLE which is also the LSE in this simple case. For any bootstrap sample $\{w_i^*\}_{i=1}^n$ from this DGP, $\widehat{\gamma}^*$ is the MLE using $\{w_i^*\}_{i=1}^n$. The question left is to derive the asymptotic distribution of $n(\widehat{\gamma}^* - \widehat{\gamma})$ conditional on $\widehat{\gamma}$. For this simple setup, the exact distribution of $n(\widehat{\gamma}^* - \widehat{\gamma})$ conditioning on $\widehat{\gamma}$ can be derived explicitly. First consider the LMLE. For any $t < 0$,

$$P^*\left(n\left(\widehat{\gamma}^* - \widehat{\gamma}\right) \le t | \widehat{\gamma}\right) = P^*\left(q_i^* \notin \left(\widehat{\gamma} + \frac{t}{n}, \widehat{\gamma}\right] \text{ for all } i \middle| \widehat{\gamma}\right) = \left(1 + \frac{t}{n}\right)^n \to e^t \text{ for any } \widehat{\gamma},$$

so the parametric bootstrap for γ is consistent P almost surely. Note the similarity of this derivation with (2). Next consider the MMLE. For any $t < 0$,

$$\begin{aligned} P^*\left(n\left(\widehat{\gamma}^* - \widehat{\gamma}\right) \le t | \widehat{\gamma}\right) &= P^*\left(\frac{n\left(q_{(m)}^* - \widehat{\gamma}\right) + n\left(q_{(m+1)}^* - \widehat{\gamma}\right)}{2} \le t \middle| \widehat{\gamma}\right) \\ &= P^*\left(n\left(q_{(m)}^* - \widehat{\gamma}\right) \le 2t - n\left(q_{(m+1)}^* - \widehat{\gamma}\right) \middle| \widehat{\gamma}\right) \\ &= \int_0^\infty \left(1 + \frac{2t - s}{n}\right)^n d\left(1 - \left(1 - \frac{s}{n}\right)^n\right) \to \int_0^\infty e^{2t-s} d\left(1 - e^{-s}\right) = \frac{e^{2t}}{2} \end{aligned}$$

for any a.s. consistent $\widehat{\gamma}$. Similar analysis applies to $t \ge 0$. This derivation is similar to (4).

In the smoothed bootstrap, the only difference is that $q \sim \widehat{f}_q(\cdot)$ which is a consistent estimator of $f_q(\cdot)$.

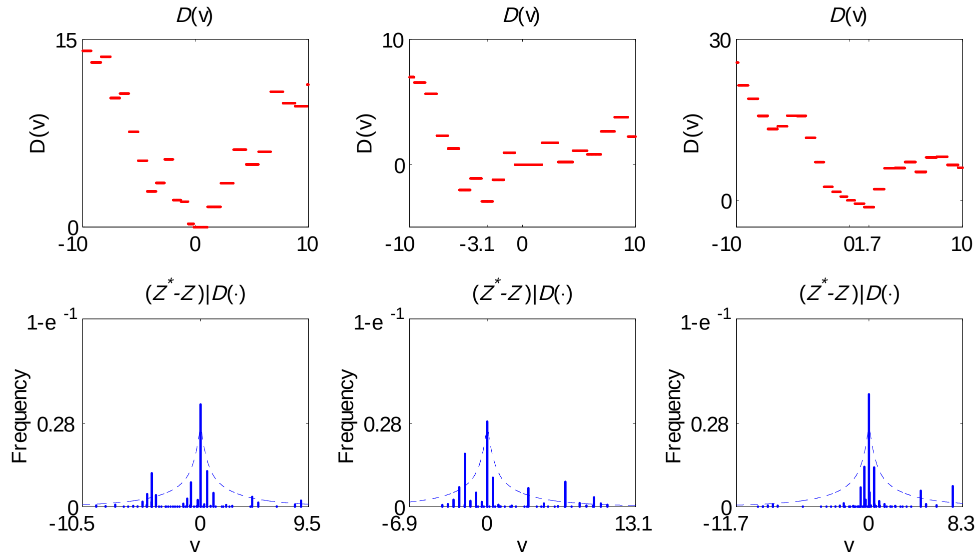

Figure 2: Dependence of the Component Distribution of $Z^* - Z$ on $D(\cdot)$

Now, for the LLSE,

$$P^*\left(n\left(\widehat{\gamma}^* - \widehat{\gamma}\right) \leq t | F_n\right) = \left(1 - \left(\widehat{F}_q\left(\widehat{\gamma}\right) - \widehat{F}_q\left(\widehat{\gamma} + \frac{t}{n}\right)\right)\right)^n \to e^t$$

for any $t < 0$ as long as $\widehat{f}_q(\cdot)$ is consistent in a neighborhood of γ_0 and $\widehat{\gamma}$ is a.s. consistent, where $\widehat{F}_q(\cdot)$ is the cdf associated with $\widehat{f}_q(\cdot)$. For the MLSE,

$$P^*\left(n\left(\widehat{\gamma}^* - \widehat{\gamma}\right) \leq t | F_n\right) = \int_0^\infty \left(1 - \left(\widehat{F}_q\left(\widehat{\gamma}\right) - \widehat{F}_q\left(\widehat{\gamma} + \frac{2t - s}{n}\right)\right)\right)^n d\left(1 - \left(1 - \left(\widehat{F}_q\left(\widehat{\gamma} + \frac{s}{n}\right) - \widehat{F}_q\left(\widehat{\gamma}\right)\right)\right)^n\right)$$

which converges to $e^{2t}/2$ for any $t < 0$. Similarly, for $t \geq 0$, $P^*\left(n\left(\widehat{\gamma}^* - \widehat{\gamma}\right) > t | F_n\right)$ converges to $e^{-2t}/2$.

2. Results Based on the MLSE for the Example in Section 3.2

This section contains the analysis about the example in Section 3.2 based on the MLSE. Figure 2 corresponds to Figure 1 in the main text. Comparing to Figure 1 there, the support of $\left(Z^* - Z\right)|D(\cdot)$ based on the MLSE is not completely contained in the set of middle points of the contiguous jumping locations of $D(v)$; the middle point of any two jumping locations can be included in the support of $\left(Z^* - Z\right)|D(\cdot)$. This is of course because $N^*_{i\pm}$ can be zero such that two or more jumps can be combined into one. Although there are more point masses in the distribution of $\left(Z^* - Z\right)|D(\cdot)$ based on the MLSE and we expect this conditional distribution would give a better approximation to the distribution of Z, the bootstrap nevertheless fails since the distribution of Z is continuous while the distribution of $\left(Z^* - Z\right)|D(\cdot)$ is discrete.

The first characteristic of the distribution of $(Z^* - Z)\,|D(\cdot)$ based on the LLSE still holds here. The largest point mass happens at zero. The second characteristic changes to that large point masses often happen *around* deeply negative jumps. Since the calculation based on the MLSE is too involved, we only provide some illustrations based on the LLSE here. Suppose $Min_- = 0$, $z_{11} > 0$, $z_{12} < 0$, $z_{1i} > 0$ for $i > 2$ and $z_{2i} \geq 0$ for all i. In this case, $P^*(Min_- = 1) = 0$.[3]

$$\begin{aligned}
&P^*(Min_- = 2) = P^*\left(N^*_{1-}z_{11} + N^*_{2-}z_{12} \leq 0, N^*_{3-} > 0\right) \\
&= \left(1 - e^{-1}\right)\sum\nolimits_{k=0}^{\infty} P^*\left(N^*_{1-}z_{11} + k \cdot z_{12} \leq 0\right) P^*\left(N^*_{2-} = k\right) \\
&= \left(1 - e^{-1}\right)\sum\nolimits_{k=0}^{\infty} \frac{e^{-1}}{k!} P^*\left(N^*_{1-} \leq -k\frac{z_{12}}{z_{11}}\right) \\
&= \left(1 - e^{-1}\right)\sum\nolimits_{k=0}^{\infty} \frac{e^{-1}}{k!} \sum\nolimits_{j=0}^{\left\lfloor -k\frac{z_{12}}{z_{11}}\right\rfloor} \frac{e^{-1}}{j!},
\end{aligned}$$

and

$$\begin{aligned}
&P^*(Min_- = 3) = P^*\left(N^*_{1-}z_{11} + N^*_{2-}z_{12} + N^*_{3-}z_{13} \leq 0, N^*_{4-} > 0\right) \\
&= P^*\left(N^*_{1-}z_{11} + N^*_{2-}z_{12} \leq 0, N^*_{3-} = 0, N^*_{4-} > 0\right) \\
&= e^{-1}\left(1 - e^{-1}\right)\sum\nolimits_{k=0}^{\infty} \frac{e^{-1}}{k!} \sum\nolimits_{j=0}^{\left\lfloor -k\frac{z_{12}}{z_{11}}\right\rfloor} \frac{e^{-1}}{j!} = e^{-1}P^*(Min_- = 2),
\end{aligned}$$

where $\lfloor x \rfloor$ is the greatest integer less than x. For $k > 3$, it can be similarly shown that $P^*(Min_- = k) = e^{-1}P^*(Min_- = k-1)$. For this simple sample path, the distribution on the left hand side of the deeply negative jump mimics the distribution in the simple threshold model without error term; see Figure 1 for intuitive impression. For more complicated sample paths of $D(v)$, the decaying rate may not be exactly e^{-1}. The insight for the LLSE should be extended to the MLSE. The third characteristic cannot apply here. The point masses in the right neighborhood of zero are not less than those in the left neighborhood. This is confirmed by checking the average bootstrap distribution $Z^* - Z$ in Figure 3 where the distribution of $Z^* - Z$ is symmetric.

	2.5% Quantile	97.5% Quantile	Coverage
Asymptotic	-12.18	12.18	95.00%
Min	-89.28	-0.03	24.44%
Max	0.06	88.97	99.99%
Average	-15.44	15.68	86.48%
Average Bootstrap	-23.01	24.23	99.17%

Table 5: Characterizations of Quantiles and Coverages In the Bootstrap

[3] There is a general result: if $z_{1k} < 0$, then $P^*(Min_- = k-1) = 0$; if $z_{2k} > 0$, then $P^*(Min_+ = k) = 0$, where Min_+ is the number of jumps before attaining the minimum of $D(v)$ on $v > 0$.

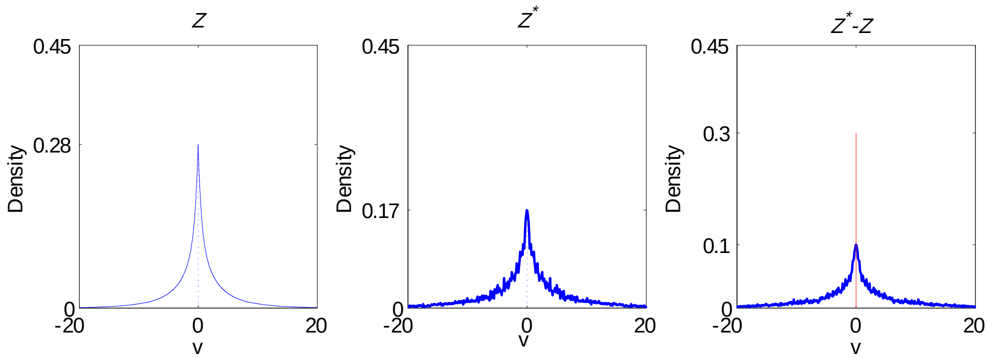

Figure 3: Comparison of Asymptotic Distributions under P_r

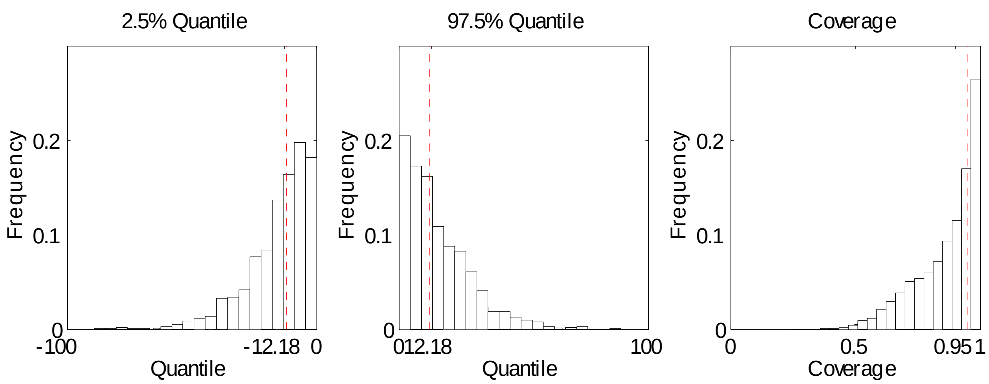

Figure 4: Distributions of 2.5% and 97.5% Quantiles and Coverage

Table 5 corresponds to Table 1 in the main text. The results here are quite similar as those based on the LLSE. For example, the average coverage is less than 95%, which means that there will be an undercover problem in practice when the bootstrap is based on the MLSE. Also, the absolute values of the quantiles of the average bootstrap are larger than those of the asymptotic counterparts, which indicates that $Z^* - Z$ in Figure 3 has a heavier tail than Z. This fact can also be confirmed by noticing that $Var(n(\widehat{\gamma}^* - \gamma_0)) = 121.45$ which is more than double of $Var(n(\widehat{\gamma} - \gamma_0)) = 31.06$. Theoretically, 2.5% quantiles and 97.5% quantiles should be symmetric; the little asymmetry in Table 5 is due to the simulation error. Figure 4 corresponds to Figure 3 in the main text. Different from the LLSE case, the distributions of the 2.5% quantile and 97.5% quantile are symmetric and there is no point mass at zero in the distribution of the 97.5% quantile.

3. Other Bootstrap Schemes

Parametric Bootstrap: Suppose the only unknown element in P is $\vartheta = (\theta', \sigma', \alpha')'$, where $\alpha \in \mathbb{R}^{d_\alpha}$ is the nuisance parameter affecting the shape of the conditional error distribution $f(\varepsilon|x, q; \alpha)$. Denote the dependence of P on θ by P_θ, then $P = P_{\theta_0}$, and $P_n = P_{\widehat{\theta}}$, where $\widehat{\theta}$ is the maximum likelihood estimator (MLE) or the Bayes estimator (BE) in Yu (2012). As mentioned in Section 3.4 of Yu (2012), the parametric bootstrap is valid.

It is noteworthy that in the parametric bootstrap, we must simulate from the joint distribution $f(y,x,q)$ to make the bootstrap procedure based on $\widehat{\theta}$ valid, where $f(y,x,q)$ covers both $f(\varepsilon|x,q;\widehat{\alpha})$ and $f(x,q)$. But even in a parametric model, $f(x,q)$ is seldom specified.[4] So a natural bootstrap scheme to avoid this difficulty is to condition on $\left\{(x_i',q_i)'\right\}_{i=1}^n$ and only utilize the randomness from $f(\varepsilon|x,q;\widehat{\alpha})$, which leads to the parametric wild bootstrap below.

Parametric Wild Bootstrap: P_n is constructed conditional on $\left\{(x_i',q_i)'\right\}_{i=1}^n$. For each $(x_i',q_i)'$, the conditional distribution of y_i is derived from

$$y_i = \begin{cases} x_i'\widehat{\beta}_1 + \widehat{\sigma}_1\varepsilon_i^*, & q_i \le \widehat{\gamma}; \\ x_i'\widehat{\beta}_2 + \widehat{\sigma}_2\varepsilon_i^*, & q_i > \widehat{\gamma}; \end{cases}$$

where $\varepsilon_i^* \sim f(\varepsilon|x,q;\widehat{\alpha})$. This bootstrap scheme is not consistent for γ. As shown in Yu (2012), the Bayesian credible set is an appropriate choice in parametric models since it does not rely on the specific form of $f(x,q)$ and has a good performance in finite samples.

Similarly, we have the wild bootstrap of Wu (1986) and Liu (1988).

Wild Bootstrap: P_n is similarly constructed as in the parametric wild bootstrap with the only difference being that $\varepsilon_i^* = d_i^*\widehat{\varepsilon}_i$ where $E^*\left[\varepsilon_i^*\right]=0$, $E^*\left[\varepsilon_i^{*2}\right]=\widehat{\varepsilon}_i^2$ and $E^*\left[\varepsilon_i^{*3}\right]=\widehat{\varepsilon}_i^3$. As noted in Liu (1988), matching the third moment can improve the rate of convergence of the bootstrap estimate for regular parameters like β. A popular choice of the d_i^* distribution is the two-point distribution:

$$P^*\left(d_i^* = \frac{1-\sqrt{5}}{2}\right) = \frac{1+\sqrt{5}}{2\sqrt{5}} \text{ and } P^*\left(d_i^* = \frac{1+\sqrt{5}}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}}. \tag{5}$$

For other choices of the d_i^* distribution, see Liu (1988) and Mammen (1993). Although the wild bootstrap is consistent and may have some finite-sample refinements for regular parameters (see Liu (1988) and Mammen (1993)), it is not consistent for γ. If ε_i is independent of $(x_i',q_i)'$, then we can extract more information from $\{\widehat{\varepsilon}_i\}_{i=1}^n$ by letting ε_i^* follow P_n^ε, where P_n^ε is the empirical distribution of $\{\widehat{\varepsilon}_i\}_{i=1}^n$. This procedure is often termed as *bootstrapping residuals*. To guarantee that the residuals have mean zero, the centered residuals, $\left\{\widehat{\varepsilon}_i - \overline{\widehat{\varepsilon}}\right\}_{i=1}^n$, are often used to substitute $\{\widehat{\varepsilon}_i\}_{i=1}^n$, where $\overline{\widehat{\varepsilon}} = n^{-1}\sum_{i=1}^n \widehat{\varepsilon}_i$. We still denote the corresponding empirical measure as P_n^ε.

In the wild bootstrap, we localize the objective function around $\left(\widehat{\beta}',\widehat{\gamma}\right)'$ since $\left(\widehat{\beta}',\widehat{\gamma}\right)'$ is the true value

[4] When q is independent of (x,ε), we can condition on $\{x_i\}_{i=1}^n$, and simulate only from $f(\varepsilon|x;\widehat{\alpha})$ and $f_q(q)$. But the problem remains since $f_q(q)$ is seldom known in reality and must be estimated nonparametrically. See more related discussions on the smoothed bootstrap in Section 4.1 below.

of $(\beta', \gamma)'$ under P_n, and uniformly for h in a compact set,

$$nP_n^*\left(m\left(\cdot\left|\widehat{\beta}+\frac{u}{\sqrt{n}},\widehat{\gamma}+\frac{v}{n}\right.\right)-m\left(\cdot\left|\widehat{\beta},\widehat{\gamma}\right.\right)\right) \tag{6}$$
$$= u_1'E\left[x_ix_i'\mathbf{1}\left(q_i\leq\gamma_0\right)\right]u_1+u_2'E\left[x_ix_i'\mathbf{1}\left(q_i>\gamma_0\right)\right]u_2-W_{wn}^*\left(u\right)+D_{wn}^*\left(v\right)+o_{P_r}\left(1\right).$$

where

$$W_{wn}^*\left(u\right)=W_{1wn}^*\left(u_1\right)+W_{2wn}^*\left(u_2\right),$$

and

$$D_{wn}^*\left(v\right)=\sum_{i=1}^n\overline{z}_{1i}^{w*}\mathbf{1}\left(\widehat{\gamma}+\frac{v}{n}<q_i\leq\widehat{\gamma}\right)+\sum_{i=1}^n\overline{z}_{2i}^{w*}\mathbf{1}\left(\widehat{\gamma}<q_i\leq\widehat{\gamma}+\frac{v}{n}\right),$$

with

$$W_{1wn}^*\left(u_1\right)=u_1'\left(\frac{2\sigma_{1,0}}{\sqrt{n}}\sum_{i=1}^n d_i^*x_i\varepsilon_i\mathbf{1}\left(q_i\leq\gamma_0\right)\right),\ W_{2wn}^*\left(u_2\right)=u_2'\left(\frac{2\sigma_{2,0}}{\sqrt{n}}\sum_{i=1}^n d_i^*x_i\varepsilon_i\mathbf{1}\left(q_i>\gamma_0\right)\right)$$
$$\overline{z}_{1i}^{w*}=2x_i'\left(\beta_{1,0}-\beta_{2,0}\right)\sigma_{1,0}d_i^*\varepsilon_i+\left(\beta_{1,0}-\beta_{2,0}\right)x_ix_i'\left(\beta_{1,0}-\beta_{2,0}\right),$$
$$\overline{z}_{2i}^{w*}=-2x_i'\left(\beta_{1,0}-\beta_{2,0}\right)\sigma_{2,0}d_i^*\varepsilon_i+\left(\beta_{1,0}-\beta_{2,0}\right)x_ix_i'\left(\beta_{1,0}-\beta_{2,0}\right).$$

and d_i^* following the distribution (5). The following Theorem 3 corresponds to Theorem 1 in the nonparametric bootstrap.

Theorem 1 *Under Assumption D with D5 changing to* $E[\varepsilon^4]<\infty$ *and* $E[\|x\|^{4+\delta}]<\infty$ *for some* $\delta>0$, *(6) holds and*

$$\left(W_{wn}^*\left(u\right),D_{wn}^*\left(v\right)\right)\rightsquigarrow\left(W^*\left(u\right),D_w^*\left(v\right)\right)$$

on any compact set, where $W^*\left(u\right)=2u_1'W_1^*+2u_2'W_2^*$ *is defined in Theorem 1, and*

$$D_w^*\left(v\right)=\begin{cases} z_{11}^{w*}+\sum\limits_{i=2}^{N_1(|v|)+1}z_{1i}^{w*},\ \text{if } v<0;\\ \sum\limits_{i=1}^{N_2(v)}z_{2i}^{w*},\ \text{if } v\geq 0;\end{cases}$$

with $z_{\ell i}^{w*}$ *following the conditional distribution of* $\overline{z}_{\ell i}^{w*}$ *given* $q_i=\gamma_0$. *Furthermore,* W_1^*, W_2^*, $\{z_{1i}^{w*},z_{2i}^{w*}\}_{i\geq1}$, $N_1(\cdot)$ *and* $N_2(\cdot)$ *are mutually independent of each other, and* $\{d_i^*\}_{i\geq1}$ *are i.i.d. sequences independent of the rest components of* $D_w^*\left(v\right)$.

Proof:. We first show (6). Note that $\widehat{\varepsilon}_i=\varepsilon_i+O_P\left(n^{-1/4}\right)$ uniformly over $i=1,\cdots,n$. From the modified Assumption D5, $\max\limits_{1\leq j\leq n}|x_j|<n^{1/(4+\delta)}$, and $\max\limits_{1\leq j\leq n}|\varepsilon_j|<n^{1/4}$ with probability one by the Borel-Cantelli

lemma. Also, we know that $\widehat{\gamma}-\gamma_0=O_P(n^{-1})$, $\widehat{\beta}-\beta_0=O_P(n^{-1/2})$ and $\widehat{\sigma}-\sigma_0=O_P(n^{-1/2})$. Now,

$$\begin{aligned}
\widehat{\varepsilon}_i-\varepsilon_i &= \frac{y_i-x_i'\widehat{\beta}_1}{\widehat{\sigma}_1}\mathbf{1}\left(q_i\leq\widehat{\gamma}\right)+\frac{y_i-x_i'\widehat{\beta}_2}{\widehat{\sigma}_2}\mathbf{1}\left(q_i>\widehat{\gamma}\right)-\varepsilon_i \\
&= \frac{x_i'\beta_{1,0}+\sigma_{1,0}\varepsilon_i-x_i'\widehat{\beta}_1}{\widehat{\sigma}_1}\mathbf{1}\left(q_i\leq\gamma_0\wedge\widehat{\gamma}\right)+\frac{x_i'\beta_{2,0}+\sigma_{2,0}\varepsilon_i-x_i'\widehat{\beta}_2}{\widehat{\sigma}_2}\mathbf{1}\left(q_i>\gamma_0\vee\widehat{\gamma}\right) \\
&+ \frac{x_i'\beta_{1,0}+\sigma_{1,0}\varepsilon_i-x_i'\widehat{\beta}_2}{\widehat{\sigma}_2}\mathbf{1}\left(\widehat{\gamma}<q_i\leq\gamma_0\right)+\frac{x_i'\beta_{2,0}+\sigma_{2,0}\varepsilon_i-x_i'\widehat{\beta}_1}{\widehat{\sigma}_1}\mathbf{1}\left(\gamma_0<q_i\leq\widehat{\gamma}\right)-\varepsilon_i \\
&= x_i'\frac{\beta_{1,0}-\widehat{\beta}_1}{\widehat{\sigma}_1}\mathbf{1}\left(q_i\leq\gamma_0\wedge\widehat{\gamma}\right)+x_i'\frac{\beta_{2,0}-\widehat{\beta}_2}{\widehat{\sigma}_2}\mathbf{1}\left(q_i>\gamma_0\vee\widehat{\gamma}\right) \\
&+ \left[\frac{\sigma_{1,0}}{\widehat{\sigma}_1}\mathbf{1}\left(q_i\leq\gamma_0\wedge\widehat{\gamma}\right)+\frac{\sigma_{2,0}}{\widehat{\sigma}_2}\mathbf{1}\left(q_i>\gamma_0\vee\widehat{\gamma}\right)-1\right]\varepsilon_i+O_P\left(n^{-3/4}\right) \\
&= O_P\left(n^{-1/4}\right) \text{ uniformly for } i.
\end{aligned}$$

Given these facts, uniformly for h in any compact set of $\mathbb{R}^{2k+1}$,

$$\begin{aligned}
& nP_n^*\left(m\left(\cdot\left|\widehat{\beta}+\frac{u}{\sqrt{n}},\widehat{\gamma}+\frac{v}{n}\right.\right)-m\left(\cdot\left|\widehat{\beta},\widehat{\gamma}\right.\right)\right) \\
&= \sum_{i=1}^n\left(u_1'\frac{x_ix_i'}{n}u_1-u_1'\frac{2\widehat{\sigma}_1}{\sqrt{n}}x_id_i^*\widehat{\varepsilon}_i\right)\mathbf{1}\left(q_i\leq\widehat{\gamma}\wedge\widehat{\gamma}+\frac{v}{n}\right)+\sum_{i=1}^n\left(u_2'\frac{x_ix_i'}{n}u_2-u_2'\frac{\widehat{\sigma}_2}{\sqrt{n}}x_id_i^*\widehat{\varepsilon}_i\right)\mathbf{1}(q_i>\widehat{\gamma}\vee\widehat{\gamma}+\frac{v}{n}) \\
&+ \sum_{i=1}^n\left[\left(\widehat{\beta}_1-\widehat{\beta}_2-\frac{u_2}{\sqrt{n}}\right)'x_ix_i'\left(\widehat{\beta}_1-\widehat{\beta}_2-\frac{u_2}{\sqrt{n}}\right)+2x_i'\left(\widehat{\beta}_1-\widehat{\beta}_2-\frac{u_2}{\sqrt{n}}\right)\widehat{\sigma}_1d_i^*\widehat{\varepsilon}_i\right]\mathbf{1}\left(\widehat{\gamma}+\frac{v}{n}<q_i\leq\widehat{\gamma}\right) \\
&+ \sum_{i=1}^n\left[\left(\widehat{\beta}_1+\frac{u_1}{\sqrt{n}}-\widehat{\beta}_2\right)'x_ix_i'\left(\widehat{\beta}_1+\frac{u_1}{\sqrt{n}}-\widehat{\beta}_2\right)-2x_i'\left(\widehat{\beta}_1+\frac{u_1}{\sqrt{n}}-\widehat{\beta}_2\right)\widehat{\sigma}_2d_i^*\widehat{\varepsilon}_i\right]\mathbf{1}\left(\widehat{\gamma}<q_i\leq\widehat{\gamma}+\frac{v}{n}\right) \\
&= \sum_{i=1}^n\left(u_1'\frac{x_ix_i'}{n}u_1-u_1'\frac{2\sigma_{1,0}}{\sqrt{n}}x_id_i^*\varepsilon_i\right)\mathbf{1}\left(q_i\leq\gamma_0\right)+\sum_{i=1}^n\left(u_2'\frac{x_ix_i'}{n}u_2-u_2'\frac{\sigma_{2,0}}{\sqrt{n}}x_id_i^*\varepsilon_i\right)\mathbf{1}(q_i>\gamma_0)+D_{wn}^*\left(v\right)+o_{P_r}(1),
\end{aligned}$$

where the simplification to $D_{wn}^*(v)$ in the last equality uses the fact that $\max\limits_{1\leq j\leq n}|x_j|<n^{1/(4+\delta)}$. Now, (6) follows directly.

The weak convergence of $W_{wn}^*(u)$ is straightforward, so we concentrate on $D_{wn}^*(v)$. As in the proof of Theorem 1, we need only calculate the limit of the characteristic function $P_r\left(\exp\left\{\sqrt{-1}\left(t_1\sum\limits_{i=1}^nT_{3i}^*+t_2\sum\limits_{i=1}^nT_{4i}^*\right)\right\}\right)$, where $T_{3i}^*=\overline{z}_{1i}^{w*}\mathbf{1}\left(\widehat{\gamma}+\frac{v}{n}<q_i\leq\widehat{\gamma}\right)$ and $T_{4i}^*=\overline{z}_{2i}^{w*}\mathbf{1}\left(\widehat{\gamma}<q_i\leq\widehat{\gamma}+\frac{v}{n}\right)$. The proof is almost the same as that of Theorem 1, but there are two points worth noting. First, since $\widehat{\gamma}-\gamma_0=O_P(n^{-1})$, the discussion can focus on $\widehat{\gamma}\in[\gamma_0-\frac{C}{n},\gamma_0+\frac{C}{n}]$ for a large enough C without affecting the asymptotic properties of $D_{wn}^*(v)$. So the conditioning on $q_i=\gamma_0$ in the calculation of Theorem 1 is still valid. Second, $\sum\limits_{i=1}^nT_{3i}^*=\sum\limits_{i=1}^n\overline{z}_{1i}^{w*}\mathbf{1}\left(q_i=\widehat{\gamma}\right)+\sum\limits_{i=1}^n\overline{z}_{1i}^{w*}\mathbf{1}\left(\widehat{\gamma}+\frac{v}{n}<q_i<\widehat{\gamma}\right)$ and the first term on the right hand side includes only one summand. ■

For $D_w^*(v)$, the randomness introduced by the wild bootstrap appears only in $\{d_i^*\}_{i\geq1}$, so the procedure (e.g., Step 2) in Section 2.3 can be similarly used as in the nonparametric bootstrap to show the invalidity of the wild bootstrap for γ. The reason for the wild bootstrap failure is the same as that for the nonparametric bootstrap failure: only the bootstrap sampling in the neighborhood of $\widehat{\gamma}$ (or γ_0) is informative to the inference on γ. Nevertheless, there are indeed some differences between these two schemes besides the multipliers in $D_w^*(\cdot)$ (d_i^*) and $D^*(\cdot)$ ($N_{i\pm}^*$). In $D_n^*(v)$ (and $D_n(v)$), we localize the objective function around γ_0, so there is a random interval around zero without any jump in $D^*(v)$ (and $D(v)$). In $D_{wn}^*(v)$, however, we localize the objective function around $\widehat{\gamma}$ and $\widehat{\gamma}$ equals some q_i, so there is a jump in $D_w^*(v)$ immediately as v gets negative. By the argmax continuous mapping theorem, $n(\widehat{\gamma}^* - \widehat{\gamma}) \rightsquigarrow Z_w^* \equiv \arg\min_v D_w^*(v)$. Given this special structure of $D_w^*(v)$, we expect that there will be a point mass at 0 in the component measure and also in Z_w^*.[5] The detailed analysis as in the example of Section 3.2 is omitted here to avoid replication. In summary, we first localize the objective function at γ_0 then recenter $\widehat{\gamma}^*$ at $\widehat{\gamma}$ by $(\widehat{\gamma}^* - \gamma_0) - (\widehat{\gamma} - \gamma_0)$ in the nonparametric bootstrap, while in the wild bootstrap, we localize the objective function directly at $\widehat{\gamma}$, so there is no need to recenter $\widehat{\gamma}^*$. In short, the order of the two operations, arg min and recentering at $\widehat{\gamma}$, are different in these two bootstrap schemes.

Under some stronger assumptions on the conditional distribution of $(x_i', \varepsilon_i)'$ given $q_i = \gamma_0$, we can show the nonexistence of the conditional weak limit of $n(\widehat{\gamma}^* - \widehat{\gamma})$. To simplify notations, define the random variables with the conditional distribution of $(x_i', \varepsilon_i)'$ given $q_i = \gamma_0$ as $(\underline{x}_i', \underline{\varepsilon}_i)'$, and define the event

$$
\begin{aligned}
A = & \left\{\underline{x}_i'(\beta_{1,0} - \beta_{2,0}) \geq \epsilon,\, \underline{\varepsilon}_i \leq -(\beta_{1,0} - \beta_{2,0})\underline{x}_i\underline{x}_i'(\beta_{1,0} - \beta_{2,0}) \big/ \left[(1+\sqrt{5})\underline{x}_i'(\beta_{1,0} - \beta_{2,0})\sigma_{1,0}\right]\right\} \\
& \cup \left\{\underline{x}_i'(\beta_{1,0} - \beta_{2,0}) \leq -\epsilon,\, \underline{\varepsilon}_i \geq -(\beta_{1,0} - \beta_{2,0})\underline{x}_i\underline{x}_i'(\beta_{1,0} - \beta_{2,0}) \big/ \left[(1+\sqrt{5})\underline{x}_i'(\beta_{1,0} - \beta_{2,0})\sigma_{1,0}\right]\right\} \\
& \cup \left\{\underline{x}_i'(\beta_{1,0} - \beta_{2,0}) \geq \epsilon,\, \underline{\varepsilon}_i \geq -(\beta_{1,0} - \beta_{2,0})\underline{x}_i\underline{x}_i'(\beta_{1,0} - \beta_{2,0}) \big/ \left[(1-\sqrt{5})\underline{x}_i'(\beta_{1,0} - \beta_{2,0})\sigma_{1,0}\right]\right\} \\
& \cup \left\{\underline{x}_i'(\beta_{1,0} - \beta_{2,0}) \leq -\epsilon,\, \underline{\varepsilon}_i \leq -(\beta_{1,0} - \beta_{2,0})\underline{x}_i\underline{x}_i'(\beta_{1,0} - \beta_{2,0}) \big/ \left[(1-\sqrt{5})\underline{x}_i'(\beta_{1,0} - \beta_{2,0})\sigma_{1,0}\right]\right\} \\
& \equiv A_1 \cup A_2 \cup A_3 \cup A_4
\end{aligned}
$$

with ϵ being a positive constant. Now, it is not hard to see that for each $\omega \in A$, $P^*(\widehat{\gamma}^* \neq \widehat{\gamma})$ is at least $(\sqrt{5}-1)/(2\sqrt{5}) > 0$. This is because $z_{12}^{w*} \leq 0$ for $d_i^* = (1+\sqrt{5})/2$ on $A_1 \cup A_2$ and $d_i^* = (1-\sqrt{5})/2$ on $A_3 \cup A_4$. So if we assume $P(A) > 0$, then applying the proof idea in (iii) of Theorem 2, we can show the conditional weak limit of $n(\widehat{\gamma}^* - \widehat{\gamma})$ does not exist for ω's with a positive probability (which is at least as large as $P(A)$). The assumption that $P(A) > 0$ is not restrictive as it seems, e.g., if the conditional support of $\underline{\varepsilon}_i$ given $\underline{x}_i$ is $\mathbb{R}$, then $P(A) > 0$. For the specific example in Section 3.2, $\underline{x}_i' = 1$, $\beta_{1,0} - \beta_{2,0} = 1$, $\sigma_{1,0} = 1$, and $\underline{\varepsilon}_i = \varepsilon_i$. In this case, if we let $\epsilon = 1/2$, then $A = \{\varepsilon_i < -1/(1+\sqrt{5}) \text{ or } \varepsilon_i > -1/(1-\sqrt{5})\}$, whose probability is obviously positive given that $\varepsilon_i \sim N(0,1)$.

From Lemma 4.4 of Seijo and Sen (2011) (or Lemma 6 in the next section) and Theorem 3, the asymptotic results for bootstrapping residuals and the parametric wild bootstrap are much expected. In bootstrapping

[5] But Z_w^* should not have low densities in the right neighborhood of zero since the structure of $D_w^*(v)$ is somewhat like canceling the first waiting time on $v \leq 0$ in $D(v)$.

residuals, $D^*_{wn}(v)$ is changed to

$$D^*_{rn}(v) = \sum_{i=1}^{n} \overline{z}^{r*}_{1i} \mathbf{1}\left(\widehat{\gamma} + \frac{v}{n} < q_i \leq \widehat{\gamma}\right) + \sum_{i=1}^{n} \overline{z}^{r*}_{2i} \mathbf{1}\left(\widehat{\gamma} < q_i \leq \widehat{\gamma} + \frac{v}{n}\right),$$

where

$$\overline{z}^{r*}_{1i} = \left\{2x_i'(\beta_{1,0} - \beta_{2,0})\,\sigma_{1,0}\varepsilon_i^* + (\beta_{1,0} - \beta_{2,0})\,x_i x_i'(\beta_{1,0} - \beta_{2,0})\right\},$$
$$\overline{z}^{r*}_{2i} = \left\{-2x_i'(\beta_{1,0} - \beta_{2,0})\,\sigma_{2,0}\varepsilon_i^* + (\beta_{1,0} - \beta_{2,0})\,x_i x_i'(\beta_{1,0} - \beta_{2,0})\right\}.$$

The randomness introduced by bootstrapping residuals only appears in $\{\varepsilon_i^*\}_{i\geq 1}$, where ε_i^* follows the distribution $f_\varepsilon(\varepsilon_i^*)$ and is independent of all other components of $D^*_{rn}(v)$. In the parametric wild bootstrap, $D^*_{wn}(v)$ changes to

$$D^*_{pn}(v) = \sum_{i=1}^{n} \overline{z}^{p*}_{1i} \mathbf{1}\left(\widehat{\gamma} + \frac{v}{n} < q_i \leq \widehat{\gamma}\right) + \sum_{i=1}^{n} \overline{z}^{p*}_{2i} \mathbf{1}\left(\widehat{\gamma} < q_i \leq \widehat{\gamma} + \frac{v}{n}\right),$$

where

$$\overline{z}^{p*}_{1i} = \ln \frac{\frac{\sigma_{1,0}}{\sigma_{2,0}} f_{\varepsilon|x,q}\left(\left.\frac{\sigma_{1,0}\varepsilon_i^* + x_i'(\beta_{1,0} - \beta_{2,0})}{\sigma_{2,0}}\right| x_i, q_i; \alpha_0\right)}{f_{\varepsilon|x,q}\left(\varepsilon_i^* | x_i, q_i; \alpha_0\right)}, \overline{z}^{p*}_{2i} = \ln \frac{\frac{\sigma_{2,0}}{\sigma_{1,0}} f_{\varepsilon|x,q}\left(\left.\frac{\sigma_{2,0}\varepsilon_i^* + x_i'(\beta_{2,0} - \beta_{1,0})}{\sigma_{1,0}}\right| x_i, q_i; \alpha_0\right)}{f_{\varepsilon|x,q}\left(\varepsilon_i^* | x_i, q_i; \alpha_0\right)},$$

and $D_n(v)$ changes to

$$D_{pn}(v) = \sum_{i=1}^{n} \overline{z}^{p}_{1i} \mathbf{1}\left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0\right) + \sum_{i=1}^{n} \overline{z}^{p}_{2i} \mathbf{1}\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n}\right),$$

where

$$\overline{z}^{p}_{1i} = \ln \frac{\frac{\sigma_{1,0}}{\sigma_{2,0}} f_{\varepsilon|x,q}\left(\left.\frac{\sigma_{1,0}\varepsilon_i + x_i'(\beta_{1,0} - \beta_{2,0})}{\sigma_{2,0}}\right| x_i, q_i; \alpha_0\right)}{f_{\varepsilon|x,q}\left(\varepsilon_i | x_i, q_i; \alpha_0\right)}, \overline{z}^{p}_{2i} = \ln \frac{\frac{\sigma_{2,0}}{\sigma_{1,0}} f_{\varepsilon|x,q}\left(\left.\frac{\sigma_{2,0}\varepsilon_i + x_i'(\beta_{2,0} - \beta_{1,0})}{\sigma_{1,0}}\right| x_i, q_i; \alpha_0\right)}{f_{\varepsilon|x,q}\left(\varepsilon_i | x_i, q_i; \alpha_0\right)}.$$

The randomness introduced by the parametric wild bootstrap only appears in $\{\varepsilon_i^*\}_{i\geq 1}$, where ε_i^* follows the same conditional distribution $f_{\varepsilon|x,q}(\varepsilon_i^*|x_i, q_i; \alpha_0)$ as ε_i and is conditionally independent of ε_i given (x_i, q_i). So we expect $\arg\min_v D^*_{pn}(v)$ and $\arg\min_v D_{pn}(v)$ are highly correlated under P_r just as in the nonparametric bootstrap case. When ε is independent of $(x', q)'$, we simulate ε_i^* independently from f_ε without conditioning on $(x_i', q_i)'$. Accordingly, all conditional densities $f_{\varepsilon|x,q}$ in $D_{pn}(v)$ and $D^*_{pn}(v)$ change to f_ε, and ε_i^* is independent of all other components of $D^*_{pn}(v)$. The proofs combine the proof ideas in Yu (2012) and Theorem 4, but a formal development of these results is beyond the scope of this paper.

4. Proof of Theorem 3

To prove theorem 3, we need the following four lemmas. The first lemma extends Lemma 4.4 of Seijo and Sen (2011) to the case with nonconstant covariates. The other three correspond to Lemma 3, 4 and 1 in Appendix B, respectively. We use Q_n to denote the probability measure of w implied by the smoothed bootstrap procedure, and use Q_n^* for the empirical smoothed bootstrap measure. θ_n and θ_n^* are used for the true value and the bootstrap estimator of θ in the smoothed bootstrap; that is, $\theta_n = \theta_0(Q_n)$ and $\theta_n^* = \theta_0(Q_n^*)$.

Lemma 1 *Let F_ε and φ be, respectively, the distribution and characteristic function of ε. Then*

(i) *for any $\eta > 0$, $\sup_{|t|\leq\eta} \left|\int \exp\left\{\sqrt{-1}t\varepsilon\right\} dP_n^\varepsilon(\varepsilon) - \varphi(t)\right| \xrightarrow{P} 0$;*

(ii) $\|P_n^\varepsilon(\varepsilon) - F_\varepsilon(\varepsilon)\|_\infty \xrightarrow{P} 0$;

(iii) $\int \varepsilon^2 dP_n^\varepsilon(\varepsilon) \xrightarrow{P} 1$;

(iv) $\int |\varepsilon| \, dP_n^\varepsilon(\varepsilon) \xrightarrow{P} P\,|\varepsilon|$;

(v) $\int \varepsilon^4 dP_n^\varepsilon(\varepsilon) = O_P(1)$.

Proof. Let F_n^ε and $\widehat{F}_n^\varepsilon$ be the empirical distribution of $\varepsilon_1, \cdots, \varepsilon_n$ and $\widehat{\varepsilon}_1, \cdots, \widehat{\varepsilon}_n$, respectively. Since

$$\int \exp\left\{\sqrt{-1}t\varepsilon\right\} dP_n^\varepsilon(\varepsilon) = \exp\left\{-\sqrt{-1}t\overline{\widehat{\varepsilon}}\right\} \widehat{F}_n^\varepsilon\left(\exp\left\{\sqrt{-1}t\varepsilon\right\}\right),$$

for any $t \in \mathbb{R}$ with $|t| \leq \eta$,

$$\begin{aligned} &\left|\int \exp\left\{\sqrt{-1}t\varepsilon\right\} dP_n^\varepsilon(\varepsilon) - \exp\left\{\sqrt{-1}t\overline{\widehat{\varepsilon}}\right\} F_n^\varepsilon\left(\exp\left\{\sqrt{-1}t\varepsilon\right\}\right)\right| \\ &= \left|\widehat{F}_n^\varepsilon\left(\exp\left\{\sqrt{-1}t\varepsilon\right\}\right) - F_n^\varepsilon\left(\exp\left\{\sqrt{-1}t\varepsilon\right\}\right)\right| \leq |\eta|\, F_n\left(|\widehat{\varepsilon} - \varepsilon|\right), \end{aligned}$$

where F_n is the empirical measure on the n data points and is understandable from the context. From the proof of Theorem 3, $F_n\left(|\widehat{\varepsilon} - \varepsilon|\right) = O_P(n^{-1/4})$. Thus

$$\sup_{|t|\leq\eta} \left|\int \exp\left\{\sqrt{-1}t\varepsilon\right\} dP_n^\varepsilon(\varepsilon) - \exp\left\{\sqrt{-1}t\overline{\widehat{\varepsilon}}\right\} F_n^\varepsilon\left(\exp\left\{\sqrt{-1}t\varepsilon\right\}\right)\right| \xrightarrow{P} 0$$

and (i) follows immediately because $\overline{\widehat{\varepsilon}} = \widehat{F}_n^\varepsilon\left(\varepsilon\right) \xrightarrow{P} 0$ and F_n^ε converges to F_ε in total variation distance with probability one. (ii) is implied by (i) since F_ε is assumed to be continuous and (i) implies that the characteristic functions of P_n^ε converges to the characteristic function of F_ε on the entire real line in probability.

For (iii), as $\int \varepsilon^2 dP_n^\varepsilon(\varepsilon) = \widehat{F}_n^\varepsilon\left(\varepsilon^2\right) - \overline{\widehat{\varepsilon}}^2$, we need only show $\widehat{F}_n^\varepsilon\left(\varepsilon^2\right) - F_n^\varepsilon\left(\varepsilon^2\right) \xrightarrow{P} 0$.

$$\begin{aligned}
\widehat{\varepsilon}_i &= \frac{y_i - x_i'\widehat{\beta}_1}{\widehat{\sigma}_1}\mathbf{1}\left(q_i \le \widehat{\gamma}\right) + \frac{y_i - x_i'\widehat{\beta}_2}{\widehat{\sigma}_2}\mathbf{1}\left(q_i > \widehat{\gamma}\right) \\
&= \frac{x_i'\beta_{1,0} + \sigma_{1,0}\varepsilon_i - x_i'\widehat{\beta}_1}{\widehat{\sigma}_1}\mathbf{1}\left(q_i \le \gamma_0 \wedge \widehat{\gamma}\right) + \frac{x_i'\beta_{2,0} + \sigma_{2,0}\varepsilon_i - x_i'\widehat{\beta}_2}{\widehat{\sigma}_2}\mathbf{1}\left(q_i > \gamma_0 \vee \widehat{\gamma}\right) \\
&+ \frac{x_i'\beta_{1,0} + \sigma_{1,0}\varepsilon_i - x_i'\widehat{\beta}_2}{\widehat{\sigma}_2}\mathbf{1}\left(\widehat{\gamma} < q_i \le \gamma_0\right) + \frac{x_i'\beta_{2,0} + \sigma_{2,0}\varepsilon_i - x_i'\widehat{\beta}_1}{\widehat{\sigma}_1}\mathbf{1}\left(\gamma_0 < q_i \le \widehat{\gamma}\right) \\
&= x_i'\frac{\left(\beta_{1,0} - \widehat{\beta}_1\right)\mathbf{1}\left(q_i \le \gamma_0 \wedge \widehat{\gamma}\right) + \left(\beta_{2,0} - \widehat{\beta}_1\right)\mathbf{1}\left(\gamma_0 < q_i \le \widehat{\gamma}\right)}{\widehat{\sigma}_1} \\
&+ x_i'\frac{\left(\beta_{2,0} - \widehat{\beta}_2\right)\mathbf{1}\left(q_i > \gamma_0 \vee \widehat{\gamma}\right) + \left(\beta_{1,0} - \widehat{\beta}_2\right)\mathbf{1}\left(\widehat{\gamma} < q_i \le \gamma_0\right)}{\widehat{\sigma}_2} \\
&+ \left[\frac{\sigma_{1,0}}{\widehat{\sigma}_2}\mathbf{1}\left(\widehat{\gamma} < q_i \le \gamma_0\right) + \frac{\sigma_{2,0}}{\widehat{\sigma}_1}\mathbf{1}\left(\gamma_0 < q_i \le \widehat{\gamma}\right)\right]\varepsilon_i \\
&+ \left[\frac{\sigma_{1,0}}{\widehat{\sigma}_1}\mathbf{1}\left(q_i \le \gamma_0 \wedge \widehat{\gamma}\right) + \frac{\sigma_{2,0}}{\widehat{\sigma}_2}\mathbf{1}\left(q_i > \gamma_0 \vee \widehat{\gamma}\right) - 1\right]\varepsilon_i + \varepsilon_i \\
&\equiv A + B + C + D + \varepsilon_i = O_P(n^{-1/2}) + O_P(n^{-1/2}) + O_P(n^{-1})\varepsilon_i + O_P(n^{-1/2})\varepsilon_i + \varepsilon_i \text{ uniformly for } i,
\end{aligned}$$

where A and B are $O_P(n^{-1/2})$ due to the boundedness of x_i. So $\widehat{\varepsilon}_i^2 - \varepsilon_i^2 = (A+B+C+D)^2 + 2(A+B+C+D)\varepsilon_i$ and $\widehat{F}_n^\varepsilon\left(\varepsilon^2\right) - F_n^\varepsilon\left(\varepsilon^2\right) = O_P(n^{-1/2}) + O_P(n^{-1/2})F_n^\varepsilon\left(\varepsilon^2\right) = O_P(n^{-1/2})$. For (iv), notice that $\left|\int |\varepsilon|\, dP_n^\varepsilon(\varepsilon) - \int |\varepsilon|\, dF_n^\varepsilon(\varepsilon)\right| = \left|F_n\left(\left|\widehat{\varepsilon} - \overline{\widehat{\varepsilon}}\right| - |\varepsilon|\right)\right| \le F_n\left(\left|\widehat{\varepsilon} - \varepsilon\right|\right) + \left|\overline{\widehat{\varepsilon}}\right| \xrightarrow{P} 0$, so $\int |\varepsilon|\, dP_n^\varepsilon(\varepsilon)$ has the same probability limit as $\int |\varepsilon|\, dF_n^\varepsilon(\varepsilon)$ which is $P|\varepsilon|$. For (v), notice that $\int \varepsilon^4 dP_n^\varepsilon(\varepsilon) \le \left|\overline{\widehat{\varepsilon}}\right|^4 + 4\left|\overline{\widehat{\varepsilon}}\right|^3 \widehat{F}_n^\varepsilon\left(|\varepsilon|\right) + 6\left|\overline{\widehat{\varepsilon}}\right|^2 \widehat{F}_n^\varepsilon\left(|\varepsilon|^2\right) + 4\left|\overline{\widehat{\varepsilon}}\right| \widehat{F}_n^\varepsilon\left(|\varepsilon|^3\right) + \widehat{F}_n^\varepsilon\left(|\varepsilon|^4\right)$, so we need only show $\widehat{F}_n^\varepsilon\left(|\varepsilon|^l\right) = O_P(1)$ in probability, $l = 1, 2, 3, 4$. Take $l = 4$ as an example. By expanding the expression of $\widehat{\varepsilon}_i$ above and using $E[\varepsilon^4] < \infty$, it is not hard to show this indeed holds. ■

Lemma 2 *Under Assumptions D and E,* $\theta_n^* - \widehat{\theta} \xrightarrow{Q_n} 0$.

Proof. First note that

$$\theta_n = \arg\min_\theta Q_n m(w|\theta),$$

where

$$
\begin{aligned}
Q_n m(w|\theta) &= Q_n \left(y - x'\beta_1 \mathbf{1}(q \le \gamma) - x'\beta_2 \mathbf{1}(q > \gamma)\right)^2 \\
&= Q_n \left(\left(x'\widehat{\beta}_1 + \widehat{\sigma}_1 \varepsilon\right) \mathbf{1}(q \le \widehat{\gamma}) + \left(x'\widehat{\beta}_2 + \widehat{\sigma}_2 \varepsilon\right) \mathbf{1}(q > \widehat{\gamma}) - x'\beta_1 \mathbf{1}(q \le \gamma) - x'\beta_2 \mathbf{1}(q > \gamma)\right)^2 \\
&= Q_n \left(\left[x'\left(\widehat{\beta}_1 - \beta_1\right)\right]^2 \mathbf{1}(q \le \widehat{\gamma} \wedge \gamma) + \left[x'\left(\widehat{\beta}_2 - \beta_2\right)\right]^2 \mathbf{1}(q > \widehat{\gamma} \vee \gamma)\right) \\
&+ Q_n \left(\left[x'\left(\widehat{\beta}_1 - \beta_2\right)\right]^2 \mathbf{1}(\gamma < q \le \widehat{\gamma}) + \left[x'\left(\widehat{\beta}_2 - \beta_1\right)\right]^2 \mathbf{1}(\widehat{\gamma} < q \le \gamma)\right) \\
&+ Q_n \left(\left[x'\left(\widehat{\beta}_1 - \beta_1\right)\widehat{\sigma}_1 \varepsilon\right] \mathbf{1}(q \le \widehat{\gamma} \wedge \gamma) + \left[x'\left(\widehat{\beta}_2 - \beta_2\right)\widehat{\sigma}_2 \varepsilon\right] \mathbf{1}(q > \widehat{\gamma} \vee \gamma)\right) \\
&+ Q_n \left(\left[x'\left(\widehat{\beta}_1 - \beta_2\right)\widehat{\sigma}_1 \varepsilon\right] \mathbf{1}(\gamma < q \le \widehat{\gamma}) + \left[x'\left(\widehat{\beta}_2 - \beta_1\right)\widehat{\sigma}_2 \varepsilon\right] \mathbf{1}(\widehat{\gamma} < q \le \gamma)\right) \\
&\qquad + Q_n \left(\widehat{\sigma}_1^2 \varepsilon^2 \mathbf{1}(q \le \widehat{\gamma}) + \widehat{\sigma}_2^2 \varepsilon^2 \mathbf{1}(q > \widehat{\gamma})\right) \equiv A + B + C + D + E.
\end{aligned}
$$

Since under Q_n, ε and $(x', q)'$ are independent and $P_n^{\varepsilon}(\varepsilon) = 0$, $C = D = 0$. E does not depend on θ and is finite as n gets large enough from Lemma 6. So $\arg\min_{\theta} Q_n m(w|\theta) = \arg\min_{\theta}(A + B)$. From Assumptions E1, E3 and E4 and the dominated convergence theorem, $A+B$ is finite and continuous in θ, so θ_n exists. Actually, $A + B \ge 0$ and $= 0$ when $\theta_n = \widehat{\theta}$, so without loss of generality, we can take $\theta_n = \widehat{\theta}$, and correspondingly, $\sigma_n = \widehat{\sigma} + o_P(1)$. In the general case when ε and $(x', q)'$ are not independent and $(x', q)'$ are not bounded, then C and D may not be zero and the dominated convergence theorem cannot be applied, so the analysis is more involved.

We now prove $\gamma_n^* - \gamma_n \xrightarrow{Q_n} 0$. The proof below mimics the proof of Lemma 3. Suppose $\gamma \ge \gamma_n$. Note that $Y^* = X^*\beta_{2n} + \sigma_{2n}\varepsilon^* + X^*_{\le\gamma_n}\delta_{\beta n} + \delta_{\sigma n}\varepsilon^*_{\le\gamma_n}$, and X^* lies in the space spanned by $P^*_\gamma = \overline{X}^*_\gamma \left(\overline{X}^{*\prime}_\gamma \overline{X}^*_\gamma\right)^{-1} \overline{X}^{*\prime}_\gamma$, where Y^*, X^*, $X^*_{\le\gamma_n}$, $\overline{X}^*_\gamma$, ε^* and $\varepsilon^*_{\le\gamma_n}$ are the counterparts of Y, X, $X_{\le\gamma_0}$, $\overline{X}_\gamma$, ε and $\varepsilon_{\le\gamma_0}$ in the smoothed bootstrap environment, $\delta_{\beta n} = \beta_{1n} - \beta_{2n}$, and $\delta_{\sigma n} = \sigma_{1n} - \sigma_{2n}$. So as the counterpart of $M_n(\gamma)$ in Lemma 3,

$$
\begin{aligned}
M_n^*(\gamma) &= \frac{1}{n} Y^{*\prime}\left(I - P^*_\gamma\right) Y^* \\
&= \frac{1}{n}\left[2\sigma_{2n}\delta'_{\beta n} X^{*\prime}_{\le\gamma_n}\left(I - P^*_\gamma\right)\varepsilon^* + 2\delta_{\sigma n}\delta'_{\beta n} X^{*\prime}_{\le\gamma_n}\left(I - P^*_\gamma\right)\varepsilon^*_{\le\gamma_n} + 2\sigma_{2n}\delta_{\sigma n}\varepsilon^{*\prime}\left(I - P^*_\gamma\right)\varepsilon^*_{\le\gamma_n}\right. \\
&\left. + \delta'_{\beta n} X^{*\prime}_{\le\gamma_n}\left(I - P^*_\gamma\right) X^*_{\le\gamma_n}\delta_{\beta n} + \sigma_{2n}^2 \varepsilon^{*\prime}\left(I - P^*_\gamma\right)\varepsilon^* + \delta_{\sigma n}^2 \varepsilon^{*\prime}_{\le\gamma_n}\left(I - P^*_\gamma\right)\varepsilon^*_{\le\gamma_n}\right].
\end{aligned}
$$

By a Glivenko-Cantelli theorem, see, e.g., Theorem 8.3 of Pollard (1990),

$$
\sup_{\gamma\in\Gamma:\gamma\ge\gamma_n} \left|M_n^*(\gamma) - M^*(\gamma)\right| \xrightarrow{Q_n} 0,
$$

where

$$
M^*(\gamma) = \delta'_{\beta n}\left(M(\gamma_n) - M(\gamma_n) M(\gamma)^{-1} M(\gamma_n)\right)\delta_{\beta n} + Q_n\left[\left(\sigma_{2n}\varepsilon + \delta_{\sigma n}\varepsilon \mathbf{1}(q \le \gamma_n)\right)^2\right]
$$

with $M(\gamma) = Q_n[xx'\mathbf{1}(q \leq \gamma)]$. The second term is finite as n gets large and does not depend on γ, so $\arg\min_{\gamma:\gamma\geq\gamma_n} M^*(\gamma) = \arg\min_{\gamma:\gamma\geq\gamma_n} \delta'_{\beta n}\left(M(\gamma_n) - M(\gamma_n)M(\gamma)^{-1}M(\gamma_n)\right)\delta_{\beta n}$. Since $\delta_{\beta n} \neq 0$ as n gets large, it is enough to show $M(\gamma_n) - M(\gamma_n)M(\gamma)^{-1}M(\gamma_n) > M(\gamma_n) - M(\gamma_n)M(\gamma_n)^{-1}M(\gamma_n) = 0$. $M(\gamma_n)$ is positive definite for n large enough by the dominated convergence theorem and Assumption D2, so we need only show $M(\gamma) > M(\gamma_n)$. But $M(\gamma) - M(\gamma_n) = Q_n[xx'\mathbf{1}(\gamma_n < q \leq \gamma)] > 0$ for n large enough by the dominated convergence theorem and Assumptions D3 and D4. Symmetrically, we can show the case with $\gamma < \gamma_n$. An extended version of Theorem 2.1 of Newey and McFadden (1994) can be applied to show $\gamma_n^* - \gamma_n \xrightarrow{Q_n} 0$. With the consistency of γ_n^* in hand, it is easy to show the consistency of $\beta_n^*(\gamma_n^*)$, the counterpart of $\widehat{\beta}(\widehat{\gamma})$, by a dominance argument. ■

Lemma 3 *Under Assumptions D and E,* $\gamma_n^* = \widehat{\gamma} + O_{Q_n}\left(\frac{1}{n}\right)$*, and* $\beta_n^* = \widehat{\beta} + O_{Q_n}\left(\frac{1}{\sqrt{n}}\right)$*.*

Proof. We apply the proof idea of Theorem 3.2.5 in Van der Vaart and Wellner (1996) in this proof.

Define $d\left(\theta, \widehat{\theta}\right) = \left\|\beta - \widehat{\beta}\right\| + \sqrt{|\gamma - \widehat{\gamma}|}$. We first bound $Q_n(m(w|\theta) - m(w|\widehat{\theta}))$ for θ in a neighborhood of $\widehat{\theta}$.

$$
\begin{aligned}
& Q_n(m(w|\theta) - m(w|\widehat{\theta})) \\
& = \left(\widehat{\beta}_1 - \beta_1\right)' Q_n[xx'\mathbf{1}(q \leq \gamma \wedge \widehat{\gamma})]\left(\widehat{\beta}_1 - \beta_1\right) + \left(\widehat{\beta}_2 - \beta_2\right)' Q_n[xx'\mathbf{1}(q > \gamma \vee \widehat{\gamma})]\left(\widehat{\beta}_2 - \beta_2\right) \\
& + \left(\widehat{\beta}_1 - \beta_2\right)' Q_n[xx'\mathbf{1}(\gamma \wedge \widehat{\gamma} < q \leq \widehat{\gamma})]\left(\widehat{\beta}_1 - \beta_2\right) + \left(\widehat{\beta}_2 - \beta_1\right)' Q_n[xx'\mathbf{1}(\widehat{\gamma} < q \leq \gamma \vee \widehat{\gamma})]\left(\widehat{\beta}_2 - \beta_1\right) \\
& \geq C\left(\left\|\widehat{\beta}_1 - \beta_1\right\|^2 + \left\|\widehat{\beta}_2 - \beta_2\right\|^2 + |\gamma - \widehat{\gamma}|\right) = Cd^2\left(\theta, \widehat{\theta}\right),
\end{aligned}
$$

where the last inequality need some explanations. By the dominated convergence theorem, all Q_n in the equality can be substituted by P and $\widehat{\gamma}$ substituted by γ_0 for n large enough, so by the arguments in Lemma 4, we can find some C such that the inequality holds for the given θ.

For each n, the parameter space (minus the point $\widehat{\theta}$) can be partitioned into the "shells" $S_{j,n} = \left\{\theta : 2^{j-1} < r_n d\left(\theta, \widehat{\theta}\right) \leq 2^j\right\}$ with $r_n = \sqrt{n}$ and j ranging over the integers. Given an integer J,

$$
Q_n\left(r_n d\left(\theta_n^*, \widehat{\theta}\right) > 2^J\right) \leq \sum_{\substack{j \geq J \\ 2^j \leq \eta r_n}} Q_n\left(\inf_{\theta \in S_{j,n}}\left(Q_n^* m(w|\theta) - Q_n^* m\left(w|\widehat{\theta}\right)\right) \leq 0\right) + Q_n\left(2d\left(\theta_n^*, \widehat{\theta}\right) \geq \eta\right).
$$

The second term converges to zero as $n \to \infty$ for every $\eta > 0$ by the consistency of θ_n^*, so we can concentrate on the first term. Note that

$$
Q_n^* m(w|\theta) - Q_n^* m\left(w|\widehat{\theta}\right) = (Q_n^* - Q_n)\left(m(w|\theta) - m\left(w|\widehat{\theta}\right)\right) + Q_n\left(m(w|\theta) - m\left(w|\widehat{\theta}\right)\right).
$$

By the bound of $Q_n(m(w|\theta) - m(w|\widehat{\theta}))$ above and the maximal inequality, we have

$$\begin{aligned}
&Q_n\left(\inf_{\theta\in S_{j,n}}\left(Q_n^* m\left(w|\theta\right) - Q_n^* m\left(w|\widehat{\theta}\right)\right)\leq 0\right)\\
&\leq Q_n\left(\sup_{\theta\in S_{j,n}}\left|\left(Q_n^* - Q_n\right)\left(m\left(w|\theta\right) - m\left(w|\widehat{\theta}\right)\right)\right|\geq C\frac{2^{2j-2}}{r_n^2}\right)\\
&\leq C\ \frac{2^j}{\sqrt{n}r_n}\Big/ C\frac{2^{2j-2}}{r_n^2} = C\frac{r_n}{\sqrt{n}2^j}.
\end{aligned}$$

Consequently,

$$Q_n\left(r_n d\left(\theta_n^*, \widehat{\theta}\right) > 2^M\right)\leq C\sum_{j\geq J}\frac{r_n}{\sqrt{n}2^j},$$

which can be made arbitrarily small by letting J large enough since by definition $r_n = \sqrt{n}$. ■

Lemma 4 *Under Assumptions D and E, uniformly for h in any compact set of $\mathbb{R}^{2k+1}$,*

$$\begin{aligned}
&nQ_n^*\left(m\left(\cdot\left|\widehat{\beta} + \frac{u}{\sqrt{n}}, \widehat{\gamma} + \frac{v}{n}\right.\right) - m\left(\cdot\left|\widehat{\beta}, \widehat{\gamma}\right.\right)\right)\\
&= u_1' P\left[x_i x_i' \mathbf{1}\left(q_i\leq\gamma_0\right)\right] u_1 + u_2' P\left[x_i x_i' \mathbf{1}\left(q_i > \gamma_0\right)\right] u_2 - W_n\left(u\right) + D_{sn}^*\left(v\right) + o_{Q_n}\left(1\right).
\end{aligned}$$

where $W_n(u)$ is specified in Section 3.1, and

$$D_{sn}^*\left(v\right) = \sum_{i=1}^n \overline{z}_{1i}\mathbf{1}\left(\widehat{\gamma} + \frac{v}{n} < q_i\leq\widehat{\gamma}\right) + \sum_{i=1}^n \overline{z}_{2i}\mathbf{1}\left(\widehat{\gamma} < q_i\leq\widehat{\gamma} + \frac{v}{n}\right)$$

with w *in $W_n(u)$ and $D_{sn}^*(v)$ following Q_n rather than P.*

Proof. Note that

$$\begin{aligned}
&nQ_n^*\left(m\left(\cdot\left|\widehat{\beta} + \frac{u}{\sqrt{n}}, \widehat{\gamma} + \frac{v}{n}\right.\right) - m\left(\cdot\left|\widehat{\beta}, \widehat{\gamma}\right.\right)\right)\\
&= \sum_{i=1}^n\left(u_1'\frac{x_i x_i'}{n}u_1 - u_1'\frac{2\widehat{\sigma}_1}{\sqrt{n}}x_i\varepsilon_i\right)\mathbf{1}\left(q_i\leq\widehat{\gamma}\wedge\widehat{\gamma} + \frac{v}{n}\right) + \sum_{i=1}^n\left(u_2'\frac{x_i x_i'}{n}u_2 - u_2'\frac{\widehat{\sigma}_2}{\sqrt{n}}x_i\varepsilon_i\right)\mathbf{1}(q_i > \widehat{\gamma}\vee\widehat{\gamma} + \frac{v}{n})\\
&+\sum_{i=1}^n\left[\left(\widehat{\beta}_1 - \widehat{\beta}_2 - \frac{u_2}{\sqrt{n}}\right)' x_i x_i'\left(\widehat{\beta}_1 - \widehat{\beta}_2 - \frac{u_2}{\sqrt{n}}\right) + 2x_i'\left(\widehat{\beta}_1 - \widehat{\beta}_2 - \frac{u_2}{\sqrt{n}}\right)\widehat{\sigma}_1\varepsilon_i\right]\mathbf{1}\left(\widehat{\gamma} + \frac{v}{n} < q_i\leq\widehat{\gamma}\right)\\
&+\sum_{i=1}^n\left[\left(\widehat{\beta}_1 + \frac{u_1}{\sqrt{n}} - \widehat{\beta}_2\right)' x_i x_i'\left(\widehat{\beta}_1 + \frac{u_1}{\sqrt{n}} - \widehat{\beta}_2\right) - 2x_i'\left(\widehat{\beta}_1 + \frac{u_1}{\sqrt{n}} - \widehat{\beta}_2\right)\widehat{\sigma}_2\varepsilon_i\right]\mathbf{1}\left(\widehat{\gamma} < q_i\leq\widehat{\gamma} + \frac{v}{n}\right).
\end{aligned}$$

By a uniform law of large number, $\left|\sum_{i=1}^n u_1'\frac{x_i x_i'}{n}u_1\mathbf{1}\left(q_i\leq\widehat{\gamma}\wedge\widehat{\gamma} + \frac{v}{n}\right) - u_1' Q_n\left[x_i x_i'\mathbf{1}\left(q_i\leq\widehat{\gamma}\wedge\widehat{\gamma} + \frac{v}{n}\right)\right]u_1\right|$ converges to zero in probability for h in any compact set. By the dominated convergence theorem and the a.s. convergence of $\widehat{\gamma}$ to γ_0, $Q_n\left[x_i x_i'\mathbf{1}\left(q_i\leq\widehat{\gamma}\wedge\widehat{\gamma} + \frac{v}{n}\right)\right]\to P\left[x_i x_i'\mathbf{1}\left(q_i\leq\gamma_0\right)\right]$ as n goes to infinity. Similarly, $\sum_{i=1}^n u_1'\frac{x_i x_i'}{n}u_1\mathbf{1}\left(q_i > \widehat{\gamma}\vee\widehat{\gamma} + \frac{v}{n}\right)$ converges to $u_2' P\left[x_i x_i'\mathbf{1}\left(q_i > \gamma_0\right)\right]u_2$ for h in any compact set.

Next, we show that the difference between $u_1' \frac{2\widehat{\sigma}_1}{\sqrt{n}} \sum_{i=1}^{n} x_i \varepsilon_i \mathbf{1}\left(q_i \leq \widehat{\gamma} \wedge \widehat{\gamma} + \frac{v}{n}\right)$ and $W_{1n}(u_1)$ is $o_{Q_n}(1)$. Since $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \varepsilon_i \mathbf{1}\left(q_i \leq \widehat{\gamma} \wedge \widehat{\gamma} + \frac{v}{n}\right) = O_{Q_n}(1)$, e.g., by calculating its second moment, and $\widehat{\sigma}_1$ converges to $\sigma_{1,0}$ almost surely, we need only show that the difference between $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \varepsilon_i \mathbf{1}\left(q_i \leq \widehat{\gamma} \wedge \widehat{\gamma} + \frac{v}{n}\right)$ and $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \varepsilon_i \mathbf{1}(q_i \leq \gamma_0)$ is $o_{Q_n}(1)$. Since $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \varepsilon_i \mathbf{1}\left(q_i \leq \widehat{\gamma} \wedge \widehat{\gamma} + \frac{v}{n}\right) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \varepsilon_i \mathbf{1}(q_i \leq \gamma_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \varepsilon_i \mathbf{1}(\gamma_0 < q_i \leq \widehat{\gamma} \wedge \widehat{\gamma} + \frac{v}{n})$ and $\widehat{\gamma} = \gamma_0 + O_P(n^{-1})$, by Assumption E3, D4 and the dominated convergence theorem, we can show the second moment of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \varepsilon_i \mathbf{1}\left(\gamma_0 < q_i \leq \widehat{\gamma} \wedge \widehat{\gamma} + \frac{v}{n}\right)$ converges to zero. By tedious but quite similar arguments, we can show the difference between the remaining terms and the targets we want is $o_{Q_n}(1)$. ■

Proof of Theorem 3. The proof completely parallels the proof of Theorem 1 and Theorem 2. We first derive the finite-dimensional limit distributions of $(W_n(u), D_{sn}^*(v))$, then check its stochastic equicontinuity, and finally apply the argmax continuous mapping theorem to show that the asymptotic distribution of θ_n^* is as specified in the theorem.

First, define T_{1i} and T_{2i} as in the proof of Theorem 1, $T_{3i} = \overline{z}_{1i} \mathbf{1}\left(\widehat{\gamma} + \frac{v}{n} < q_i \leq \widehat{\gamma}\right)$, and $T_{4i} = \overline{z}_{2i} \mathbf{1}\left(\widehat{\gamma} < q_i \leq \widehat{\gamma} + \frac{v}{n}\right)$, then

$$
\begin{aligned}
& Q_n\left(\exp\left\{\sqrt{-1}\left[s_1' T_{1i} + s_2' T_{2i} + + t_1 T_{3i} + t_2 T_{4i}\right]\right\}\right) \\
&= Q_n\left(\exp\left\{\sqrt{-1} s' T_i / \sqrt{n}\right\}\right) + \frac{v_1}{n} \widehat{f}_q(\widehat{\gamma}) Q_n\left(\exp\left\{\sqrt{-1} s' T_i / \sqrt{n}\right\} \left[\exp\left\{\sqrt{-1} t_1 \overline{z}_{1i}\right\} - 1\right] \middle| q_i = \widehat{\gamma}\right) \\
&+ \frac{v_2}{n} \widehat{f}_q(\widehat{\gamma}) Q_n\left(\exp\left\{\sqrt{-1} s' T_i / \sqrt{n}\right\} \left[\exp\left\{\sqrt{-1} t_2 \overline{z}_{2i}\right\} - 1\right] \middle| q_i = \widehat{\gamma}\right) + o\left(\frac{1}{n}\right),
\end{aligned}
$$

where $s = (s_1' \ s_2')'$, $T_i = \begin{pmatrix} S_{1i}' & S_{2i}' \end{pmatrix}'$, $\widehat{f}_q(q) = \int \widehat{f}(x, q) dx$ is a uniformly consistent estimator of $f_q(q)$, and $o(1)$ is a quantity going to zero uniformly over $i = 1, \cdots, n$ from Assumption E3. By the Taylor expansion of $\exp\left\{\sqrt{-1} s' T_i / \sqrt{n}\right\}$, the dominated convergence theorem and Lemma 6,

$$
Q_n\left(\exp\left\{\sqrt{-1} s' T_i / \sqrt{n}\right\}\right) = 1 - \frac{1}{2n} s' \mathcal{J} s + o\left(\frac{1}{n}\right),
$$

where

$$
\mathcal{J} = \lim_{n \to \infty} Q_n(T_i T_i') = \text{diag}\left\{P\left[x x' \varepsilon^2 \mathbf{1}(q \leq \gamma_0)\right], P\left[x x' \varepsilon^2 \mathbf{1}(q > \gamma_0)\right]\right\} = \text{diag}\left\{P\left[x x' \mathbf{1}(q \leq \gamma_0)\right], P\left[x x' \mathbf{1}(q > \gamma_0)\right]\right\}.
$$

Next,

$$
\begin{aligned}
& Q_n\left(\exp\left\{\sqrt{-1} s' T_i / \sqrt{n}\right\} \left[\exp\left\{\sqrt{-1} t_1 \overline{z}_{1i}\right\} - 1\right] \middle| q_i = \widehat{\gamma}\right) = Q_n\left(\left[\exp\left\{\sqrt{-1} t_1 \overline{z}_{1i}\right\} - 1\right] \middle| q_i = \widehat{\gamma}\right) + o(1) \\
&= \int \left[\exp\left\{\sqrt{-1} t_1 \overline{z}_{1i}\right\} - 1\right] \frac{\widehat{f}(x, \widehat{\gamma})}{\widehat{f}_q(\widehat{\gamma})} dx dP_n^{\varepsilon}(\varepsilon) + o(1) \to P\left[\left\{\exp\left\{\sqrt{-1} t_1 \overline{z}_{1i}\right\}\right\} \middle| q_i = \gamma_0\right] - 1,
\end{aligned}
$$

where the first equality is from Taylor expanding $\exp\left\{\sqrt{-1}s'T_i/\sqrt{n}\right\}$, the second equality is from the definition of Q_n, and the convergence is from the dominated convergence theorem and Lemma 6. Similarly, $Q_n\left(\exp\left\{\sqrt{-1}s'T_i/\sqrt{n}\right\}\left[\exp\left\{\sqrt{-1}t_2\overline{z}_{2i}\right\}-1\right]\middle|\, q_i=\widehat{\gamma}\right)\to P\left[\left\{\exp\left\{\sqrt{-1}t_2\overline{z}_{2i}\right\}\right\}\middle|\, q_i=\gamma_0\right]-1$. In summary,

$$\begin{aligned}&Q_n\left(\exp\left\{\sqrt{-1}\left[s_1'T_{1i}+s_2'T_{2i}++t_1T_{3i}+t_2T_{4i}\right]\right\}\right)\\&=1+\frac{1}{n}\left[-\frac{1}{2}s'\mathcal{J}s+f_q\left(\gamma_0\right)v_1\left(P\left[\left\{\exp\left\{\sqrt{-1}t_1\overline{z}_{1i}\right\}\right\}\middle|\, q_i=\gamma_0\right]-1\right)\right.\\&\left.+f_q\left(\gamma_0\right)v_2\left(P\left[\left\{\exp\left\{\sqrt{-1}t_2\overline{z}_{2i}\right\}\right\}\middle|\, q_i=\gamma_0\right]-1\right)\right]+o\left(\frac{1}{n}\right).\end{aligned}$$

So

$$\begin{aligned}&Q_n\left(\exp\left\{\sqrt{-1}\left[s_1\sum_{i=1}^n T_{1i}+s_2\sum_{i=1}^n T_{2i}++t_1\sum_{i=1}^n T_{3i}+t_2\sum_{i=1}^n T_{4i}\right]\right\}\right)\\&=\prod_{i=1}^n Q_n\left(\exp\left\{\sqrt{-1}\left[s'T_i/\sqrt{n}+t_1T_{3i}+t_2T_{4i}\right]\right\}\right)\\&\to\exp\left\{-\frac{1}{2}s'\mathcal{J}s+f_q\left(\gamma_0\right)v_1\left(P\left[\left\{\exp\left\{\sqrt{-1}t_1\overline{z}_{1i}\right\}\right\}\middle|\, q_i=\gamma_0\right]-1\right)\right.\\&\left.+f_q\left(\gamma_0\right)v_2\left(P\left[\left\{\exp\left\{\sqrt{-1}t_2\overline{z}_{2i}\right\}\right\}\middle|\, q_i=\gamma_0\right]-1\right)\right\},\end{aligned}$$

which matches the characteristic function of $(W(u), D(v))$, and the result of interest follows.

Second, for the stochastic equicontinuity, note that for any $\epsilon>0$ and $0<v_1<v_2$ which are stopping times in a compact set,

$$\begin{aligned}&Q_n\left(\sup_{|v_2-v_1|<\delta}|D_{sn}^*(v_2)-D_{sn}^*(v_1)|>\epsilon\right)\le Q_n\left(\sum_{i=1}^n|\overline{z}_{2i}|\cdot\sup_{|v_2-v_1|<\delta}\mathbf{1}\left(\widehat{\gamma}+\frac{v_1}{n}<q_i\le\widehat{\gamma}+\frac{v_2}{n}\right)>\epsilon\right)\\&\le\sum_{i=1}^n Q_n\left[|\overline{z}_{2i}|\sup_{|v_2-v_1|<\delta}\mathbf{1}\left(\widehat{\gamma}+\frac{v_1}{n}<q_i\le\widehat{\gamma}+\frac{v_2}{n}\right)\right]\Big/\epsilon\le\frac{C\delta}{\epsilon}\text{ for }n\text{ large enough,}\end{aligned}$$

where C in the last inequality can take $\left(\overline{f}_q+\epsilon\right)\sup_{\gamma_0-\epsilon\le\gamma\le\gamma_0+\epsilon}Q_n\left[|\overline{z}_{2i}|\,|q_i=\gamma\right]<\infty$ from Assumption D4, D6, E3 and Lemma 6. ■

5. Simulation Results for Pseudo-SB

To avoid the curse of dimensionality in the SB of DGP2 and DGP4, we can simulate only from the marginal density estimate of q instead of the joint density estimate of $(q,\varepsilon)'$ and $(x,q)'$, but then there is a misspecification problem. We label such a SB procedure as the pseudo-SB; see Section 7 of Seijo and Sen (2011) for a detailed description. Similarly, we label the NPI that is based on the marginal density of ε in DGP2 as the pseudo-NPI. The results for pseudo-SB and pseudo-NPI are summarized in Table 4. The results in the table show that the pseudo-SB intervals are not reliable. In DGP 2, their coverage is almost 1, which induces long intervals, while in DGP4, their performance is close to that of the correct SB procedures. It is hard to filter out which cases are suitable to use pseudo-SB intervals.

$n \to$	50		200		500	
CIs↓ Cov and Leng($\times 10^{-1}$)→	Coverage	Length	Coverage	Length	Coverage	Length
DGP2						
Pseudo-SB-LLSE (ET)	0.988	7.461	0.997	1.539	0.998	0.593
Pseudo-SB-LLSE (S)	0.992	8.141	0.992	1.643	0.994	0.636
Pseudo-SB-MLSE (ET)	0.989	7.089	1	1.425	1	0.549
Pseudo-SB-MLSE (S)	0.995	7.305	1	1.419	1	0.546
Pseudo-NPI	0.976	2.565	0.991	0.575	0.987	0.222
DGP4						
Pseudo-SB-LLSE (ET)	0.961	6.170	0.936	1.423	0.949	0.560
Pseudo-SB-LLSE (S)	0.958	6.688	0.948	1.534	0.938	0.603
Pseudo-SB-MLSE (ET)	0.945	5.650	0.943	1.308	0.941	0.516
Pseudo-SB-MLSE (S)	0.943	5.616	0.942	1.301	0.943	0.513

Table 4: Performance of Pseudo-SB and Pseudo-NPI: Coverage and Average Length of the Nominal 95% Confidence Intervals for γ (Based on 1000 Repetitions)

6. Construction of the Nonparametric Posterior Interval

The following algorithm is given in Yu (2008).

Step 1: Get the LSE $\left(\widehat{\gamma}, \widehat{\beta}', \widehat{\sigma}'\right)'$ and the corresponding residuals $\{\widehat{\varepsilon}_i\}_{i=1}^n$.

Step 2: Estimate the joint density of w by kernel smoothing,

$$\widehat{f}\left(\mathsf{w}\right) = \frac{1}{nh^{k+2}} \sum_{i=1}^{n} K\left(\frac{\widehat{\mathsf{w}}_i - \mathsf{w}}{h}\right)$$

where h is the bandwidth, and $K(\cdot) : \mathbb{R}^{k+2} \to \mathbb{R}$ is a kernel density.

Step 3: Construct the estimated likelihood function as

$$\begin{aligned}
\widehat{\mathcal{L}}_n(\gamma) &= \prod_{i=1}^{n} \left[\frac{1}{\widehat{\sigma}_1}\widehat{f}\left(\frac{y_i - x_i'\widehat{\beta}_1}{\widehat{\sigma}_1}, x_i, q_i\right)\mathbf{1}(q_i \le \gamma) + \frac{1}{\widehat{\sigma}_2}\widehat{f}\left(\frac{y_i - x_i'\widehat{\beta}_2}{\widehat{\sigma}_2}, x_i, q_i\right)\mathbf{1}(q_i > \gamma)\right] \\
&= \exp\left\{\sum_{i=1}^{n}\mathbf{1}(q_i \le \gamma)\ln\left(\frac{1}{\widehat{\sigma}_1}\widehat{f}\left(\frac{y_i - x_i'\widehat{\beta}_1}{\widehat{\sigma}_1}, x_i, q_i\right)\right) + \sum_{i=1}^{n}\mathbf{1}(q_i > \gamma)\ln\left(\frac{1}{\widehat{\sigma}_2}\widehat{f}\left(\frac{y_i - x_i'\widehat{\beta}_2}{\widehat{\sigma}_2}, x_i, q_i\right)\right)\right\} \\
&\equiv \exp\left\{\widehat{L}_n(\gamma)\right\},
\end{aligned}$$

and the posterior distribution as

$$\widehat{p}_n(\gamma) = \frac{\exp\left\{\widehat{L}_n(\gamma)\right\}\pi_2(\gamma)}{\int_\Gamma \exp\left\{\widehat{L}_n(\widetilde{\gamma})\right\}\pi_2(\widetilde{\gamma})\,d\widetilde{\gamma}},$$

where $\pi_2(\gamma)$ is the prior of γ, e.g., $\pi_2(\gamma)$ can be the uniform distribution on $(q_{\min}, q_{\max})$ with $q_{\min}$ ($q_{\max}$) being the minimum (maximum) of $\{q_i\}_{i=1}^n$.

Step 4: Based on a MCMC algorithm, draw a Markov chain

$$S = \left(\gamma^{(1)}, \cdots, \gamma^{(B)}\right)$$

whose marginal density is approximately $\widehat{p}_n(\gamma)$. Then the $(1-\tau)$ NPI is constructed by picking out the $\tau/2$ and $1-\tau/2$ quantiles of S.

When ε is independent of $(x', q)'$, $\widehat{f}(\mathsf{w})$ in Step 2 and 3 is substituted by $\widehat{f}_\varepsilon(\varepsilon) = \frac{1}{nh}\sum_{i=1}^n K(\widehat{\varepsilon}_i - \varepsilon)$, where h is the bandwidth, and $K(\cdot): \mathbb{R} \to \mathbb{R}$ is a kernel density.

Additional References

Liu, R.Y., 1988, Bootstrap Procedures under Some Non i.i.d. Models, *The Annals of Statistics*, 16, 1696-1708.

Mammen, E., 1993, Bootstrap and Wild Bootstrap for High Dimensional Linear Models, *The Annals of Statistics*, 21, 255-285.

Newey, W.K. and D.L. McFadden, 1994, Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics*, Vol. 4, R.F. Eagle and D.L. McFadden, eds., Elsevier Science B.V., Ch. 36, 2113-2245.

Pollard, D., 1990, *Empirical Processes: Theory and Applications*, Hayward : Institute of Mathematical Statistics.

Wu, C.F.J., 1986, Jackknife, Bootstrap and other Resampling Methods in Regression Analysis, *The Annals of Statistics*, 14, 1261-1295.