

The Bootstrap in Threshold Regression

Ping Yu*

University of Auckland

First Version: June 2008

This Version: January 2010

Abstract

This paper shows that the nonparametric bootstrap is inconsistent, and the parametric bootstrap is consistent for inference of the threshold point in discontinuous threshold regression. An interesting phenomenon is that the asymptotic nonparametric bootstrap distribution of the threshold point is discrete and depends on the sampling path of the original data. This is because the threshold point is essentially a boundary of the sample space, and only the bootstrap sampling on the data in the neighborhood of the threshold point is informative. The results are compared with Andrews (2000) where a parameter is on the boundary of the parameter space rather than the sample space. The method developed in this paper is generic in deriving the asymptotic bootstrap distribution. The remedies to the nonparametric bootstrap failure in the literature are also summarized.

KEYWORDS: bootstrap failure, threshold regression, boundary, average bootstrap, parametric wild bootstrap, truncated Poisson process, multiplier compound Poisson process

JEL-CLASSIFICATION: C13, C21.

*Email: p.yu@auckland.ac.nz. I would like to thank Bruce Hansen, Jack Porter, Peter Phillips and seminar participants at UW-Madison and Southern Methodist University for helpful comments.

1 Introduction

Since the introduction by Efron (1979), the bootstrap has become a popular alternative of the asymptotic inference, either when the asymptotic distribution is difficult to derive or the finite sample refinement can be achieved. Meanwhile, many examples of the bootstrap failure surface. Most of these examples are contributed by statisticians; see Andrews (2000, p400) for a list of them.¹ To my knowledge, there are also three examples contributed by econometricians: the case when a parameter is on the boundary of the parameter space in Andrews (2000), the maximum score estimator in Abrevaya and Huang (2005), and the matching estimator for the program evaluation in Abadie and Imbens (2008). This paper contributes another example where the nonparametric bootstrap fails: the least squares estimator (LSE) of the threshold point in discontinuous threshold regression.

The typical setup of threshold regression is

$$\begin{aligned} y &= \begin{cases} x' \beta_1 + \sigma_1 e, & q \leq \gamma; \\ x' \beta_2 + \sigma_2 e, & q > \gamma; \end{cases} \\ E[e|x, q] &= 0, \end{aligned} \tag{1}$$

where q is the threshold variable used to split the sample and has a density $f_q(\cdot)$, $x \in \mathbb{R}^k$, $\beta = (\beta'_1, \beta'_2)' \in \mathbb{R}^{2k}$ and $\sigma = (\sigma_1, \sigma_2)'$ are threshold parameters on mean and variance in the two regimes, $E[e^2] = 1$ is a normalization of the error variance and adopts conditional heteroskedasticity, and all the other variables have the same definitions as in the linear regression framework. The threshold regression model has many applications; see, e.g., Yu (2007) and Lee and Seo (2008) for the examples. Define $\theta = (\beta', \gamma)'$, which is the parameter of interest in applications. Most literature on threshold regression concentrates on the nonregular parameter γ since inference for the regular parameters β is standard.

In discontinuous threshold regression where $(\beta'_1, \sigma_1)' - (\beta'_2, \sigma_2)'$ is a fixed constant, Gonzalo and Wolf (2005) claim that "it is not known whether a bootstrap approach would work". This motivates them to use the subsampling for the construction of confidence intervals. Footnote 3 in Seo and Linton (2007) indicates that the bootstrap may be inconsistent in their simulations. There are some related results in the structural change literature. For example, in the framework with asymptotically diminishing threshold effect like that in Hansen (2000), Antoch et al (1995) show that the bootstrap is valid in the structural change model. See also Hušková and Kirch (2008) for the validity of the block bootstrap in the time series context. The extension of the arguments in these two papers to threshold regression is straightforward. In a similar framework as considered in this paper, Dümbgen (1991) finds that the bootstrap has the correct convergence rate in the structural change model without proving its validity. In this paper, we confirm that the bootstrap has the right convergence rate, and also show that it is not valid for inference of the threshold point.

Before presenting main results on the bootstrap inference, the LSE of θ and the basic probability structure in the bootstrap environment are defined. Suppose a random sample $\{w_i\}_{i=1}^n$ is observed, where $w_i = (y_i, x'_i, q_i)'$, the LSE of γ is usually defined by a profiled procedure.

$$\hat{\gamma} = \arg \min_{\gamma} Q_n(\gamma),$$

where

$$Q_n(\gamma) = \min_{\beta_1, \beta_2} \sum_{i=1}^n m(w_i|\theta),$$

¹See also, e.g., Horowitz (2001), Bickel et al (1997), and Beran (1997).

with

$$m(w|\theta) = (y - x'\beta_1\mathbf{1}(q \leq \gamma) - x'\beta_2\mathbf{1}(q > \gamma))^2. \quad (2)$$

Usually, there is an interval of γ minimizing this objective function. Following the literature, the left end point of the interval is taken as the minimizer. To express the model in matrix notation, define the $n \times 1$ vectors Y and \mathbf{e} by stacking the variables y_i and e_i , and the $n \times k$ matrices X , $X_{\leq \gamma}$ and $X_{> \gamma}$ by stacking the vectors x'_i , $x'_i\mathbf{1}(q_i \leq \gamma)$ and $x'_i\mathbf{1}(q_i > \gamma)$. Let

$$\begin{pmatrix} \hat{\beta}_1(\gamma) \\ \hat{\beta}_2(\gamma) \end{pmatrix} \equiv \arg \min_{\beta_1, \beta_2} \sum_{i=1}^n m(w_i|\theta) = \begin{pmatrix} (X'_{\leq \gamma} X_{\leq \gamma})^{-1} X'_{\leq \gamma} Y \\ (X'_{> \gamma} X_{> \gamma})^{-1} X'_{> \gamma} Y \end{pmatrix},$$

then the LSE of β is defined as $(\hat{\beta}'_1(\hat{\gamma}), \hat{\beta}'_2(\hat{\gamma}))' \equiv (\hat{\beta}'_1, \hat{\beta}'_2)'$. The bootstrap estimator $(\hat{\beta}^{*'}, \hat{\gamma}^*)'$ of $(\beta', \gamma)'$ is defined in the same way as above except using the bootstrap sample $\{w_i^*\}_{i=1}^n$.

The basic probability structure in the bootstrap environment is defined as follows. Let P_n and P_n^* be the empirical measure of the original data and the bootstrap sample w_1^*, \dots, w_n^* , respectively.² P_n is a random measure, and its randomness is defined by the data generating process (DGP) of the original data. P_n^* can be written as

$$P_n^* = \frac{1}{n} \sum_{i=1}^n \delta_{w_i^*} = \frac{1}{n} \sum_{i=1}^n M_{ni} \delta_{w_i},$$

where M_{ni} is the number of times that w_i is drawn from the original sample, and $\mathbf{M}_n = (M_{n1}, \dots, M_{nn})$ follows the multinomial distribution with parameters n and cell probabilities all equal to $\frac{1}{n}$ (and independent of the original data $\{w_i\}_{i=1}^n$). Suppose \mathbf{M}_n is defined on a probability space $(\mathcal{T}, \mathcal{B}, P^*)$ which describes a probability structure with an expanding support. For the original sample, take w_i as the i th coordinate projection from the probability space $(\mathcal{Z}^\infty, \mathcal{A}^\infty, P^\infty)$. For the joint randomness involving both \mathbf{M}_n and $\{w_i\}_{i=1}^n$, define the product probability space

$$(\mathcal{Z}^\infty \times \mathcal{T}, \mathcal{A}^\infty \times \mathcal{B}, P_r) \equiv (\mathcal{Z}^\infty, \mathcal{A}^\infty, P^\infty) \times (\mathcal{T}, \mathcal{B}, P^*),$$

where $P_r \equiv P^\infty \times P^*$ denotes the whole randomness in the nonparametric bootstrap.³

For the parametric bootstrap, a similar probability structure can be constructed as in the nonparametric case. The only difference is that the parametric distribution with the estimator as the true value is used as P^* .

This paper is organized as follows. In Section 2, an extreme case of threshold regression is used to illustrate why the nonparametric bootstrap is inconsistent, and the parametric bootstrap is consistent for the threshold point. Section 3 presents similar results for the general model. Section 4 uses numerical examples to show the intuition behind these results. Section 5 provides some remedies in the literature to the nonparametric bootstrap failure, and Section 6 concludes. All proofs and lemmas are left to Appendix A and B, respectively. A word on notation: any parameter with a subscript 0 means its true value, any symbol with a superscript $*$ means an object under P^* defined above instead of under the outer measure as used in some other literature, and \rightsquigarrow signifies weak convergence under the true parameter value.

²In other words, P_n is the population measure of P_n^* , and P , the original population measure, is the population measure of P_n .

³See Problem 3.6.1 in Van der Vaart and Wellner (1996) for a formal construction of this probability space.

2 When There is No Error Term: An Illustration

First, simplify (1) to the extreme case as follows:

$$y = \mathbf{1}(q \leq \gamma), \quad q \sim U[0, 1]. \quad (3)$$

This corresponds to $x = 1, \beta_{10} = 1, \beta_{20} = 0, \sigma_{10} = \sigma_{20} = 0$ in (1). 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 e in (3), so the observed y values can only be 0 or 1. In this case, the threshold point is essentially a "middle" boundary of q as shown in Yu (2007). $\hat{\gamma}$ 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 $\hat{\gamma}$ equals the q_i closest to $1/2$ from the left. Otherwise, $\hat{\gamma}$ 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 $(\frac{1}{2})^n$ which converges to zero, we can assume $\hat{\gamma} \leq \frac{1}{2}$ in the following discussion. For $t < 0$,

$$\begin{aligned} & P(n(\hat{\gamma} - \gamma_0) \leq t) \\ &= P(q_i \notin (\gamma_0 + \frac{t}{n}, \gamma_0] \text{ for all } i) \\ &= (1 + \frac{t}{n})^n \rightarrow e^t, \end{aligned} \quad (4)$$

so the asymptotic distribution of $n(\hat{\gamma} - \gamma_0)$ is a negative standard exponential, and there is no density on the positive axis. Note further that $\hat{\gamma}$ is a nondecreasing function of n , and there is no data point between $\hat{\gamma}$ and γ_0 .

2.1 Invalidity of the Nonparametric Bootstrap

The objective function of the least squares estimation is

$$\sum_{i=1}^n (y_i - \mathbf{1}(q_i \leq \gamma))^2.$$

In order to use the nonparametric bootstrap to approximate the distribution of $n(\hat{\gamma} - \gamma_0)$, we need to obtain the asymptotic distribution of $n(\hat{\gamma}^* - \hat{\gamma})$, where

$$\hat{\gamma}^* = \arg \min_{\gamma} \sum_{i=1}^n (y_i^* - \mathbf{1}(q_i^* \leq \gamma))^2,$$

and (y_i^*, q_i^*) follows the empirical distribution F_n .⁴ The asymptotic distribution of $n(\hat{\gamma}^* - \hat{\gamma})$ can be derived conditional on the data as n goes to infinity. Since γ is essentially a boundary, the following derivation is similar to that in Example 3 of Bickel et al (1997).

Suppose there are m y_i 's taking value 1, and the remaining $(n - m)$ y_i 's take value 0, then $\hat{\gamma} = q_{(m)}$, and $\hat{\gamma}$ converges to $1/2$ as n goes to infinity for any sample point ω in \mathcal{Z}^∞ according to Chan (1993), where $q_{(m)}$ is the m 'th order statistic of $\{q_i\}_{i=1}^n$. In the bootstrap sampling, as long as $(q_{(m)}, y_{(m)})$ is drawn, $\hat{\gamma}^* = q_{(m)}$. So $P_n^*(n(\hat{\gamma}^* - \hat{\gamma}) = 0 | F_n) = 1 - P_n^*((q_{(m)}, y_{(m)}) \text{ is not drawn}) = 1 - (1 - \frac{1}{n})^n \rightarrow 1 - e^{-1} > 0$, while $P(n(\hat{\gamma} - 1/2) = 0) \rightarrow 0$ since the asymptotic distribution of $n(\hat{\gamma} - 1/2)$ is continuous. Therefore, the bootstrap is not consistent. The following provides the whole asymptotic distribution of $n(\hat{\gamma}^* - \hat{\gamma}) | F_n$.

⁴In this paper, F_n is used exchangeably with P_n .

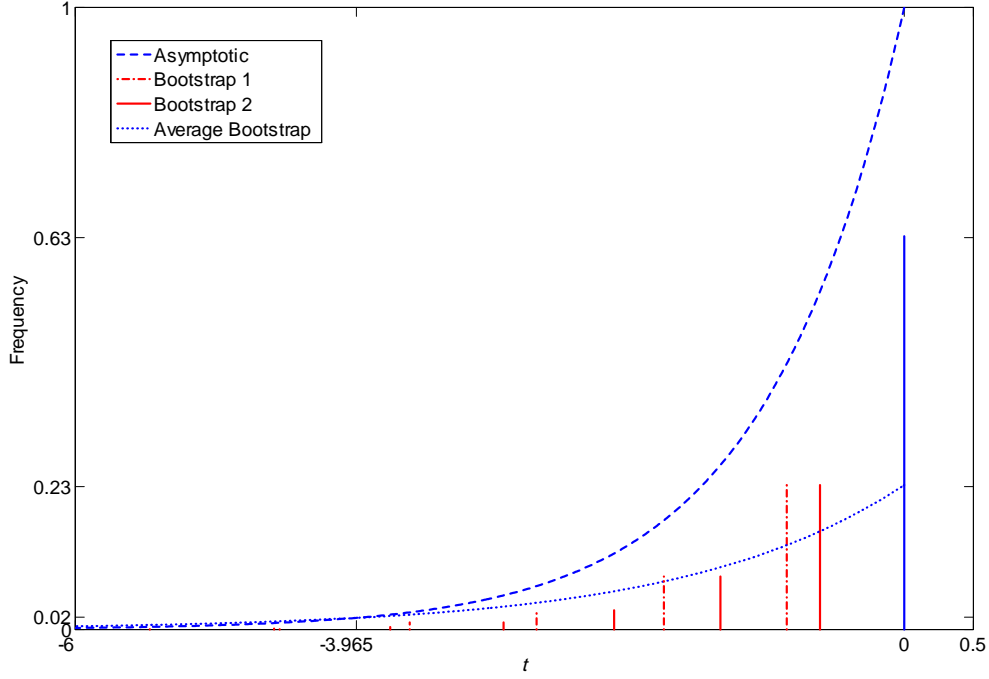


Figure 1: Comparison of the Asymptotic Distribution and Asymptotic Bootstrap Distribution in Threshold Regression without the Error Term

Note that for any $t \leq 0$,

$$\begin{aligned}
& P_n^*(n(\hat{\gamma}^* - \hat{\gamma}) < t | F_n) \\
&= P_n^*\left(\text{no } q_i^* \text{ is sampled from } \left[\hat{\gamma} + \frac{t}{n}, \hat{\gamma}\right]\right) \\
&= \left(1 - \frac{k}{n}\right)^n
\end{aligned}$$

where $k = \sum_{i=1}^n \mathbf{1}(\hat{\gamma} + \frac{t}{n} \leq q_i \leq \hat{\gamma})$. Conditional on $\hat{\gamma}$, $k \sim 1 + \text{Bin}\left(n-1, \frac{|t|/n}{1-(1/2-\hat{\gamma})}\right)$ converges weakly to $1 + N(|t|)$ for any $\hat{\gamma}$, where $N(\cdot)$ is a standard Poisson process, so $P_n^*(n(\hat{\gamma}^* - \hat{\gamma}) < t | F_n)$ converges to $e^{-(1+N(|t|))} |N(\cdot)|$ which is discrete. A new jump in $e^{-(1+N(|t|))} |N(\cdot)|$ happens as $|t|$ gets larger when the expanding interval $[\hat{\gamma} + \frac{t}{n}, \hat{\gamma}]$ covers a new q_i . Because $P_n^*(n(\hat{\gamma}^* - \hat{\gamma}) \leq 0 | F_n) \rightarrow e^{-(1+N(|t|))}|_{t=0} + (1 - e^{-1}) = 1$, there is no probability on the positive axis, which is similar as the asymptotic distribution.

The above results are surprising in two aspects. First, while the asymptotic distribution is continuous, the asymptotic bootstrap distribution is discrete. This is different from conventional models where the asymptotic bootstrap distribution is normal. Essentially, this is because the asymptotic bootstrap distribution of the threshold point relies on the bootstrap sampling on the local data (i.e., $\hat{\gamma} + \frac{t}{n} \leq q_i \leq \hat{\gamma}$) rather than the sampling on the whole dataset in conventional models. Second, the asymptotic bootstrap distribution depends on the original data. Although the point mass at zero is always $1 - e^{-1}$, how to distribute the remaining e^{-1} probability depends on how the original data are sampled. When more data (in $1/n$ rate) are

sampled in the left neighborhood of $1/2$, the point masses are closer to zero. Figure 1 shows the asymptotic distribution and asymptotic bootstrap distributions for two different original sample paths. Clearly, more original data in bootstrap 2 lie in the left neighborhood of $1/2$ than in bootstrap 1.

One important similarity between the asymptotic bootstrap distribution and 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 the asymptotic bootstrap distribution depends on the local data around $\hat{\gamma}$ in the original sample. This is not difficult to understand considering 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 nonparametric bootstrap, the true distribution of q is the empirical distribution of $\{q_i\}_{i=1}^n$, and the true value of γ is $\hat{\gamma}$.

In summary, although this example is very simple, it shows one general feature of the nonparametric bootstrap of the threshold point: the local information around γ_0 (or $\hat{\gamma}$) is most important for the bootstrap inference. As a result, the asymptotic bootstrap distribution is discrete and depends on the original data. Therefore, the nonparametric bootstrap of the threshold point is invalid.

Since the asymptotic bootstrap distribution depends on the original data, a natural question is whether the average bootstrap is consistent, where the average bootstrap distribution is the distribution of $n(\hat{\gamma}^* - \hat{\gamma})$ under P_τ . From the frequentist point of view, infinitely many original sample paths can be potentially drawn according to P^∞ . So the average bootstrap distribution can be obtained by first running the nonparametric bootstrap for each sample path, and then averaging the bootstrap distribution for all sample paths.⁵ In this example, the question becomes whether $E[e^{-(1+N(|t|))}] = e^t$ for $t \leq 0$. The answer is no since there is a point mass at zero in the asymptotic bootstrap distribution for any sample path. The asymptotic density function resulted from the average bootstrap procedure is also shown in Figure 1.

As the last comment about the invalidity of the nonparametric bootstrap, note that γ is different from μ in Andrews (2000). In Andrews (2000), $\{X_i\}_{i=1}^n$ is a sequence of i.i.d. $N(\mu, 1)$ random variables, where $\mu \in \mathbb{R}^+ \equiv \{z : z \geq 0\}$. Note that 0 is on the boundary of the parameter space, but it is still in the interior of the sample space. In contrast, in threshold regression, γ is the boundary of the sample space, and is in the interior of the parameter space. So the reason for the invalidity of the nonparametric bootstrap is different. Checking the three sufficient conditions for the validity of the bootstrap provided on page 1209 in Bickel and Freedman (1981), the uniformity condition (6.1b) fails in the nonparametric bootstrap of γ , while in the nonparametric bootstrap of μ , the continuity condition (6.1c) fails. In Andrews (2000), even the parametric bootstrap fails at $\mu = 0$, but the parametric bootstrap works in threshold regression as seen in the next subsection.

2.2 Validity of the Parametric Bootstrap

In the parametric bootstrap sampling, the following DGP is used:

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

where $\hat{\gamma}$ is the MLE which is also the LSE in this simple case. For any bootstrap sample $\{w_i^*\}_{i=1}^n$ from this DGP, $\hat{\gamma}^*$ is the MLE using $\{w_i^*\}_{i=1}^n$. The question left is to derive the asymptotic distribution of $n(\hat{\gamma}^* - \hat{\gamma})$ conditioning on $\hat{\gamma}$.

For this simple setup, the exact distribution of $n(\hat{\gamma}^* - \hat{\gamma})$ conditioning on $\hat{\gamma}$ can be derived explicitly.

⁵Abadie and Imbens (2008) use a similar method as the average bootstrap to prove the invalidity of the bootstrap for inference of matching estimators. In conventional models, the average bootstrap is never considered since the asymptotic bootstrap distribution is the same for almost all original sample paths.

For any $t < 0$,

$$\begin{aligned}
& P_n^*(n(\hat{\gamma}^* - \hat{\gamma}) \leq t|\hat{\gamma}) \\
&= P_n^*\left(q_i^* \notin \left(\hat{\gamma} + \frac{t}{n}, \hat{\gamma}\right] \text{ for all } i|\hat{\gamma}\right) \\
&= \left(1 + \frac{t}{n}\right)^n \rightarrow e^t \text{ for any } \hat{\gamma},
\end{aligned}$$

so the parametric bootstrap for γ is consistent P^∞ almost surely. This is because the uniformity condition (6.1b) in Bickel and Freedman (1981) fails in the nonparametric bootstrap, but holds in the parametric bootstrap; see the arguments in their Counter-example 2 for more details. Note the similarity of this derivation with (4).

3 The Bootstrap in Threshold Regression: The General Model

Before the formal discussion of the bootstrap inference, we will first specify the regularity conditions and review the main asymptotic results in threshold regression.

Assumption D:

1. $w_i \in \mathbb{W} \equiv \mathbb{R} \times \mathbb{X} \times \mathbb{Q} \subset \mathbb{R}^{k+2}$, $\beta_1 \in B_1 \subset \mathbb{R}^k$, $\beta_2 \in B_2 \subset \mathbb{R}^k$, $0 < \sigma_1 \in \Omega_1 \subset \mathbb{R}$, $0 < \sigma_2 \in \Omega_2 \subset \mathbb{R}$, $\Omega_1 \times \Omega_2$ is compact, $\gamma \in \Gamma = [\underline{\gamma}, \bar{\gamma}]$, $\beta_{10} \neq \beta_{20}$, and $\sigma_{10} \neq \sigma_{20}$, where \neq is an element by element operation.
2. $E[xx'] > E[xx'\mathbf{1}(q \leq \gamma)] > 0$ for all $\gamma \in \Gamma$
3. $E[xx'|q = \gamma] > 0$ uniformly for γ in an open neighborhood of γ_0 .
4. $f_q(\cdot)$ is continuous, and $0 < \underline{f}_q \leq f_q(\gamma) \leq \bar{f}_q < \infty$ for $\gamma \in \Gamma$.
5. $E[\|xe\|^2] < \infty$, and $E[\|x\|^4] < \infty$.
6. $E[\|x\|^2|q = \gamma] < \infty$ and $E[\|xe\||q = \gamma] < \infty$ uniformly for γ in an open neighborhood of γ_0 .
7. Both z_{1i} and z_{2i} have absolutely continuous distributions, where the distribution of z_{1i} is the limiting conditional distribution of \bar{z}_{1i} given $\gamma_0 + \Delta < q_i \leq \gamma_0$, $\Delta < 0$ as $\Delta \uparrow 0$ with

$$\bar{z}_{1i} = \{2x'_i(\beta_{10} - \beta_{20})\sigma_{10}e_i + (\beta_{10} - \beta_{20})x_ix'_i(\beta_{10} - \beta_{20})\},$$

and the distribution of z_{2i} is the limiting conditional distribution of \bar{z}_{2i} given $\gamma_0 < q_i \leq \gamma_0 + \Delta$, $\Delta > 0$ as $\Delta \downarrow 0$ with

$$\bar{z}_{2i} = \{-2x'_i(\beta_{10} - \beta_{20})\sigma_{20}e_i + (\beta_{10} - \beta_{20})x_ix'_i(\beta_{10} - \beta_{20})\}.$$

Assumption D is roughly a subset of Assumption 1 in Hansen (2000). It is very standard. Assumption D1 does not require B_1 and B_2 to be compact, and Assumption D2 excludes the possibility that γ_0 is on the boundary of q 's support; see Section 3.1 of Hansen (2000) for more discussions. From Chan (1993) or Yu (2007), under Assumption D,

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_1 - \beta_{10}) &\xrightarrow{d} Z_{\beta_1} \sim E[xx'\mathbf{1}(q \leq \gamma_0)]^{-1} \cdot N(\mathbf{0}, E[xx'\sigma_{10}^2e^2\mathbf{1}(q \leq \gamma_0)]), \\
\sqrt{n}(\hat{\beta}_2 - \beta_{20}) &\xrightarrow{d} Z_{\beta_2} \sim E[xx'\mathbf{1}(q > \gamma_0)]^{-1} \cdot N(\mathbf{0}, E[xx'\sigma_{20}^2e^2\mathbf{1}(q > \gamma_0)]),
\end{aligned}$$

and

$$n(\hat{\gamma} - \gamma_0) \xrightarrow{d} \arg \min_v D(v) \equiv Z_\gamma, \quad (5)$$

with

$$D(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} z_{1i}, & \text{if } v \leq 0; \\ \sum_{i=1}^{N_2(v)} z_{2i}, & \text{if } v > 0. \end{cases}$$

Here, Z_{β_1} , Z_{β_2} , $\{z_{1i}, z_{2i}\}_{i \geq 1}$, $N_1(\cdot)$ and $N_2(\cdot)$ are independent of each other, and $N_\ell(\cdot)$ is a Poisson process with intensity $f_q(\gamma_0)$, $\ell = 1, 2$. Since $D(v)$ is the cadlag step version of a two-sided compound Poisson process with $D(0) = 0$, there is a random interval $[M_-, M_+)$ minimizing $D(v)$. Since the left end point of the minimizing interval is taken as the LSE, $Z_\gamma = M_-$. The asymptotic distribution of $\hat{\beta}$ is the same as that in the case when γ_0 is known, and the asymptotic distribution of $\hat{\gamma}$ is the same as that in the case when β_0 is known. The explicit form of the density of Z_γ is derived in Appendix D of Yu (2007).

3.1 Invalidity of the Nonparametric Bootstrap

Define N_n as a Poisson number with mean n and independent of the original observations, then $M_{N_n, 1}, \dots, M_{N_n, n}$ are i.i.d. Poisson variables with mean 1. From Lemma 1 in Appendix B, Poissonization is possible. See Kac (1949) and Klaassen and Wellner (1992) for an introduction of Poissonization. Define $h = (u', v) = (u'_1, u'_2, v)'$ as the local parameter for θ , then

$$\begin{aligned} & nP_n \left(m \left(\cdot \left| \beta_0 + \frac{u}{\sqrt{n}}, \gamma_0 + \frac{v}{n} \right. \right) - m(\cdot | \beta_0, \gamma_0) \right) \\ &= u'_1 E[x_i x'_i \mathbf{1}(q_i \leq \gamma_0)] u_1 + u'_2 E[x_i x'_i \mathbf{1}(q_i > \gamma_0)] u_2 - W_n(u) + D_n(v) + o_{P^\infty}(1), \end{aligned}$$

and

$$\begin{aligned} & nP_n^* \left(m \left(\cdot \left| \beta_0 + \frac{u}{\sqrt{n}}, \gamma_0 + \frac{v}{n} \right. \right) - m(\cdot | \beta_0, \gamma_0) \right) \\ &= u'_1 E[x_i x'_i \mathbf{1}(q_i \leq \gamma_0)] u_1 + u'_2 E[x_i x'_i \mathbf{1}(q_i > \gamma_0)] u_2 - W_n(u) - W_n^*(u) + D_n^*(v) + o_{P_r}(1). \end{aligned}$$

Here,

$$\begin{aligned} D_n(v) &= \sum_{i=1}^n \left[m \left(w_i \left| \gamma_0 + \frac{v}{n} \right. \right) - m(w_i | \gamma_0) \right] \\ &= \sum_{i=1}^n \bar{z}_{1i} \mathbf{1} \left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0 \right) + \sum_{i=1}^n \bar{z}_{2i} \mathbf{1} \left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n} \right), \end{aligned}$$

$$\begin{aligned} D_n^*(v) &= \sum_{i=1}^n M_{N_n, i} \left[m \left(w_i \left| \gamma_0 + \frac{v}{n} \right. \right) - m(w_i | \gamma_0) \right] \\ &= \sum_{i=1}^n M_{N_n, i} \bar{z}_{1i} \mathbf{1} \left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0 \right) + \sum_{i=1}^n M_{N_n, i} \bar{z}_{2i} \mathbf{1} \left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n} \right), \end{aligned}$$

with

$$m(w|\gamma) = (y - x' \beta_{10} \mathbf{1}(q \leq \gamma) - x' \beta_{20} \mathbf{1}(q > \gamma))^2,$$

and

$$\begin{aligned} W_n(u) &= W_{1n}(u_1) + W_{2n}(u_2), \\ W_n^*(u) &= W_{1n}^*(u_1) + W_{2n}^*(u_2), \end{aligned}$$

with

$$\begin{aligned} W_{1n}(u_1) &= u'_1 \left(\frac{2\sigma_{10}}{\sqrt{n}} \sum_{i=1}^n x_i e_i \mathbf{1}(q_i \leq \gamma_0) \right), \\ W_{2n}(u_2) &= u'_2 \left(\frac{2\sigma_{20}}{\sqrt{n}} \sum_{i=1}^n x_i e_i \mathbf{1}(q_i > \gamma_0) \right), \\ W_{1n}^*(u_1) &= u'_1 \left(\frac{2\sigma_{10}}{\sqrt{n}} \sum_{i=1}^n (M_{N_n, i} - 1) x_i e_i \mathbf{1}(q_i \leq \gamma_0) \right), \\ W_{2n}^*(u_2) &= u'_2 \left(\frac{2\sigma_{20}}{\sqrt{n}} \sum_{i=1}^n (M_{N_n, i} - 1) x_i e_i \mathbf{1}(q_i > \gamma_0) \right). \end{aligned}$$

From the form of $D_n^*(v)$, it is clear that the bootstrap sampling in the neighborhood of γ_0 is most important. So essentially only the sampling on the finite data points around γ_0 is relevant. In contrast, $W_{1n}^*(u_1)$ and $W_{2n}^*(u_2)$ take an average form, which makes their asymptotic properties very different from $D_n^*(v)$. The following Theorem 1 gives the joint weak limit of $(W_n(u), W_n^*(u), D_n(v), D_n^*(v))$, which is critical in deriving the asymptotic bootstrap distribution in Theorem 2.

Theorem 1 *Under Assumption D,*

$$(W_n(u), W_n^*(u), D_n(v), D_n^*(v)) \rightsquigarrow (W(u), W^*(u), D(v), D^*(v))$$

on any compact set, where

$$\begin{aligned} W(u) &= 2u'_1 W_1 + 2u'_2 W_2, \\ W^*(u) &= 2u'_1 W_1^* + 2u'_2 W_2^*, \end{aligned}$$

with W_1 and W_1^* following the same distribution $N(0, E[x^2 \sigma_{10}^2 e^2 \mathbf{1}(q \leq \gamma_0)])$, W_2 and W_2^* following the same distribution $N(0, E[x^2 \sigma_{20}^2 e^2 \mathbf{1}(q > \gamma_0)])$, and $(D(v), D^*(v))$ being a bivariate vector of compound Poisson process. $D(v)$ is the same as that in (5), and

$$D^*(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} N_{i-}^* z_{1i}, & \text{if } v \leq 0; \\ \sum_{i=1}^{N_2(v)} N_{i+}^* z_{2i}, & \text{if } v > 0; \end{cases} \equiv \begin{cases} \sum_{i=1}^{N_1(|v|)} z_{1i}^*, & \text{if } v \leq 0; \\ \sum_{i=1}^{N_2(v)} z_{2i}^*, & \text{if } v > 0; \end{cases}$$

with $\{N_{i-}^*, N_{i+}^*\}_{i \geq 1}$ being independent standard Poisson variables.⁶ Furthermore, $W_1, W_2, W_1^*, W_2^*, \{z_{1i}, z_{2i}\}_{i \geq 1}, \{N_{i-}^*, N_{i+}^*\}_{i \geq 1}, N_1(\cdot)$ and $N_2(\cdot)$ are independent of each other.

$D^*(v)$ is called a *multiplier compound Poisson process*, which is highly correlated with $D(v)$ since $\{z_{1i}, z_{2i}\}_{i \geq 1}, N_1(\cdot)$ and $N_2(\cdot)$ are the same as those in $D(v)$. The randomness in $\{z_{1i}, z_{2i}\}_{i \geq 1}, N_1(\cdot)$ and

⁶From the proof in Appendix B, $\{N_{i-}^*, N_{i+}^*\}_{i \geq 1}$ is just $\{M_{N_n, i}\}_{i \geq 1}$. The following is an intuitive derivation for the fact

$N_2(\cdot)$ is introduced by the original data, so the randomness introduced by the bootstrap appears only in $\{N_{i-}^*, N_{i+}^*\}_{i \geq 1}$. Since $E[N_{i-}^*] = E[N_{i+}^*] = 1$ for any i , the average jump size of z_{1i}^* and z_{2i}^* in $D^*(v)$ is the same as z_{1i} and z_{2i} in $D(v)$. But the distribution of z_{1i}^* and z_{2i}^* instead of their mean determines the jumps in $D^*(v)$. The distribution of z_{1i}^* is

$$\begin{aligned} & P_r(z_{1i}^* \leq x) \\ &= \begin{cases} \sum_{k=1}^{\infty} P_r(N_{i-}^* = k, z_{1i} \leq \frac{x}{k}) & \text{if } x < 0; \\ e^{-1} + \sum_{k=1}^{\infty} P_r(N_{i-}^* = k, z_{1i} \leq \frac{x}{k}) & \text{if } x \geq 0; \end{cases} \\ &= \begin{cases} \sum_{k=1}^{\infty} \frac{e^{-1}}{k!} \Phi_1\left(\frac{x}{k}\right), & \text{if } x < 0; \\ e^{-1} + \sum_{k=1}^{\infty} \frac{e^{-1}}{k!} \Phi_1\left(\frac{x}{k}\right), & \text{if } x \geq 0; \end{cases} \end{aligned}$$

where $\Phi_1(\cdot)$ is the cdf of z_{1i} . The distribution of z_{2i}^* can be similarly derived. Since there is a point mass e^{-1} at zero in the distribution of z_{1i}^* and z_{2i}^* , the sample path of $D^*(v)$ is very different from that of $D(v)$. When N_{i-}^* (N_{i+}^*) is equal to zero, the i th and $(i-1)$ th jump on $v \leq 0$ ($v > 0$) in $D(v)$ are combined into one jump. When N_{i-}^* (N_{i+}^*) is greater than 1, the i th jump on $v \leq 0$ ($v > 0$) in $D(v)$ is magnified.

Theorem 2 Under Assumption D,

(i) the bootstrap estimator $\hat{\beta}^*$ is consistent; that is, $\sqrt{n}(\hat{\beta}^* - \hat{\beta}) \xrightarrow{d} Z_{\beta}^*$ as $n \rightarrow \infty$ in P^{∞} probability, where Z_{β}^* has the same distribution as $Z_{\beta} \equiv \left(Z'_{\beta_1}, Z'_{\beta_2} \right)'$ and is independent of Z_{β} .

(ii)

$$(n(\hat{\gamma} - \gamma_0), n(\hat{\gamma}^* - \gamma_0)) \xrightarrow{d} (Z_{\gamma}, Z_{\gamma}^*) \text{ as } n \rightarrow \infty \text{ in } P_r \text{ probability,}$$

where

$$Z_{\gamma}^* = \arg \min_v D^*(v).$$

(iii)

$$n(\hat{\gamma}^* - \hat{\gamma}) \xrightarrow{d} Z_{\gamma}^* - Z_{\gamma} \text{ as } n \rightarrow \infty \text{ in } P^{\infty} \text{ probability.}$$

where $Z_{\gamma}^* - Z_{\gamma}$ and Z_{β}^* are independent conditional on the original data.

When γ_0 is known, the validity of the bootstrap for regular parameters β is a standard result. When γ_0 is unknown, (i) shows that the bootstrap is still valid and the asymptotic bootstrap distribution is the same as that for the case when γ_0 is known. The validity of the bootstrap for β is due to the separability of $W(u)$ and $W^*(u)$ in the weak limit of $nP_n^* \left(m \left(\cdot \mid \beta_0 + \frac{u}{\sqrt{n}}, \gamma_0 + \frac{v}{n} \right) - m(\cdot \mid \beta_0, \gamma_0) \right)$ and their linearity

that $\{M_{ni}\}_{i \geq 1}$ can be approximated by $\{N_{i-}^*, N_{i+}^*\}_{i \geq 1}$. The finite-dimensional marginal distribution of $\{N_{i+}^*, N_{i-}^*\}_{i=1}^{\infty}$ is

$$\begin{aligned} & P^* \left(N_{i_1+}^* = k_{i_1+}, \dots, N_{i_j+}^* = k_{i_j+}, N_{i_1-}^* = k_{i_1-}, \dots, N_{i_m-}^* = k_{i_m-} \right) \\ &= \lim_{n \rightarrow \infty} \frac{n!}{k_{i_1+}! \dots k_{i_j+}! k_{i_1-}! \dots k_{i_m-}! (n-k)!} \left(\frac{1}{n} \right)^k \left(1 - \frac{j+m}{n} \right)^{n-k} \\ &= \frac{e^{-(j+m)}}{k_{i_1+}! \dots k_{i_j+}! k_{i_1-}! \dots k_{i_m-}!} \\ &= \frac{e^{-1}}{k_{i_1+}!} \dots \frac{e^{-1}}{k_{i_j+}!} \cdot \frac{e^{-1}}{k_{i_1-}!} \dots \frac{e^{-1}}{k_{i_m-}!} \end{aligned}$$

where $k = k_{i_1+} + \dots + k_{i_j+} + k_{i_1-} + \dots + k_{i_m-}$. The independence is understandable. Note that $\text{Corr}(M_{ni}, M_{nj}) = -\frac{1}{n-1} < 0$, but $\text{Corr}(M_{ni}, M_{nj}) \rightarrow 0$ when $n \rightarrow \infty$. $\text{Corr}(M_{ni}, M_{nj}) = -\frac{1}{n-1}$ is generally true for exchangeable random variables with fixed sum n ; see, for example, Aldous (1985), page 8. The negativity of the correlation is understandable since for a fixed n , an increase in one component of a multinomial vector requires a decrease in another component.

with respect to u . This point is first noted in Abrevaya and Huang (2005). In their case, the weak limit is separable but not linear, so the bootstrap fails. In the bootstrap for γ , the randomness from the original data and that from the bootstrap sampling are neither separable nor linear. In consequence, the bootstrap is not valid for γ .

From (ii), $n(\hat{\gamma}^* - \hat{\gamma}) \xrightarrow{d} Z_\gamma^* - Z_\gamma$ as $n \rightarrow \infty$ in P_r probability by the continuous mapping theorem. $Z_\gamma^* - Z_\gamma$ is the asymptotic distribution of the average bootstrap of γ . The distribution of Z_γ^* under P_r can be derived by the simulation method proposed in Appendix D of Yu (2007), but take caution that two jumps in $D^*(v)$ can be combined into one when N_{i-}^* or N_{i+}^* equals zero. This distribution is expected to be continuous but more spreading than Z_γ as N_{i-}^* (N_{i+}^*) can take values other than 1. Since $D(v)$ and $D^*(v)$ are highly correlated, Z_γ^* and Z_γ are highly correlated under P_r . As in the case without the error term, there is expected to be a point mass at zero in the distribution of $Z_\gamma^* - Z_\gamma$ under P_r although Z_γ^* and Z_γ are both continuous.⁷

By (iii), the asymptotic bootstrap distributions for the nonregular parameter γ and for the regular parameters β are independent conditional on the original data, which is similar to the asymptotic distribution as shown in (5). This is because the bootstrap samplings for the inference of β and γ use information independently. Conditional on the original data, the randomness introduced by the bootstrap in $\hat{\beta}^*$ takes an average form; see, e.g., the term $\frac{1}{\sqrt{n}} \sum_{i=1}^n (M_{N_n, i} - 1) x_i e_i \mathbf{1}(q_i \leq \gamma_0)$ in $W_{1n}^*(u_1)$. Accordingly, the bootstrap sampling on a single data point does not contribute much to the asymptotic bootstrap distribution. This "global" sampling on the original data averages out the effect of $\{M_{N_n, i}\}_{i=1}^n$ and makes the bootstrap distribution for β converge to a normal distribution. In contrast, only "local" sampling on the data around γ_0 is informative to the inference of the threshold point as shown in D_n^* above. This is why $\{N_{i-}^*, N_{i+}^*\}_{i \geq 1}$ appears in $D^*(v)$, but not in $W^*(u)$.

The asymptotic distribution of $n(\hat{\gamma}^* - \hat{\gamma})|F_n$ is discrete and critically depends on $D(\cdot)$. The magnitude of the point masses depends on $\{z_{1i}, z_{2i}\}_{i \geq 1}$, and the location of the point masses depends on both $N_\ell(\cdot)$, $\ell = 1, 2$, and $\{z_{1i}, z_{2i}\}_{i \geq 1}$. Since Z_γ is fixed conditional on the original data, the nonparametric bootstrap distribution $(Z_\gamma^* - Z_\gamma)|D(\cdot)$ is just a location shift of $Z_\gamma^*|D(\cdot)$. See the numerical example in Section 4.2 below for more concrete description of $Z_\gamma^* - Z_\gamma$.

The method developed in Theorem 2 is very general in deriving the asymptotic bootstrap distribution. Take a revisit to Andrews (2000). A natural estimator of μ is $\hat{\mu}_n = \max\{\bar{X}_n, 0\}$ which is also the maximum likelihood estimator of μ , where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Under $\mu = 0$, $\sqrt{n}(\bar{X}_n^* - \bar{X}_n)$ and $\sqrt{n}\bar{X}_n$ are asymptotically independent $N(0, 1)$, where $\bar{X}_n^* = \frac{1}{n} \sum_{i=1}^n X_i^*$ with $X_i^* \sim F_n$, and the limit variables are denoted as Z^* and Z . Then the asymptotic distribution of the average bootstrap of μ at 0 is

$$\begin{aligned} \sqrt{n}(\hat{\mu}_n^* - \hat{\mu}_n) &= \sqrt{n} \left(\max\{\bar{X}_n^*, 0\} - \max\{\bar{X}_n, 0\} \right) \\ &\xrightarrow{d} \max\{Z + Z^*, 0\} - \max\{Z, 0\}. \end{aligned}$$

The bootstrap distribution converges weakly to the conditional distribution of $\max\{Z + Z^*, 0\} - \max\{Z, 0\}$ given Z :

$$\begin{cases} \max\{Z^*, -Z\}, & \text{if } Z \geq 0; \\ Z + \max\{Z^*, -Z\}, & \text{if } Z < 0; \end{cases}$$

⁷This point mass is less than $1 - e^{-1}$, because $\hat{\gamma}$ is not necessarily $\hat{\gamma}^*$ even if it is sampled in the bootstrap unlike in the case without error term. The numerical example in Section 4.2 below provides more description of $Z_\gamma^* - Z_\gamma$ under P_r .

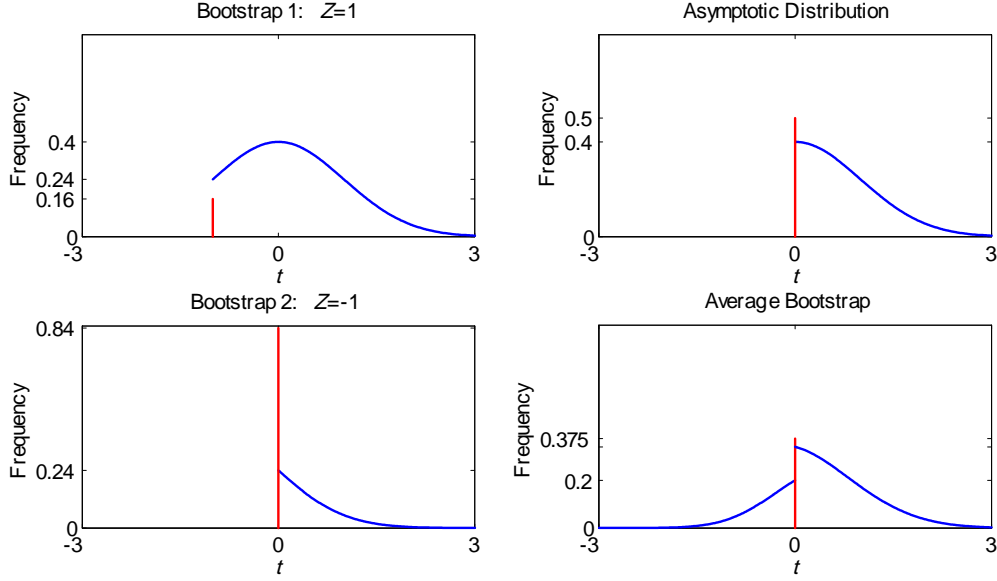


Figure 2: Comparison of the Asymptotic Distribution and Asymptotic Bootstrap Distribution in Andrews (2000)

so this asymptotic bootstrap distribution also depends on the original sampling path although only through an asymptotic sufficient statistic \bar{X}_n . This point is shown in Figure 2. In Figure 2, the asymptotic bootstrap distribution varies when Z changes. When $Z > 0$, there is a point mass at $-Z$, and when $Z < 0$, there is a point mass at 0. The point masses are increasing when Z gets smaller. The asymptotic bootstrap distribution matches the asymptotic distribution only when $Z = 0$.⁸

3.2 Validity of the Parametric Bootstrap

To prove the validity of the parametric bootstrap in the general model (1), Proposition 1.1 in Beran (1997) can be used. The critical step is to check the condition a) there, which is done in Yu (2007). To save space, the proof will not be repeated here.

It is worth pointing out that the parametric wild bootstrap is not valid. In the parametric wild bootstrap, we condition on $\{x_i, q_i\}_{i=1}^n$, and only utilize the randomness from $f(e|x, q; \hat{\alpha})$, where $\alpha \in \mathbb{R}^{d_\alpha}$ is some nuisance parameter affecting the shape of the error distribution, and $\hat{\alpha}$ is its MLE. The invalidity of the parametric wild bootstrap can be seen from the illustrative example in Section 2. 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 and does not include any randomness when conditioning on the original data. In consequence, the bootstrap confidence interval only includes the MLE itself and

⁸It can be shown that the asymptotic bootstrap distribution is
$$\begin{cases} \begin{cases} \phi(t), & \text{when } t > -Z; \\ \Phi(-Z), & \text{when } t = -Z; \\ \phi(t-Z), & \text{when } t > 0; \\ \Phi(-Z), & \text{when } t = 0; \end{cases} & \text{if } Z > 0; \\ \begin{cases} \phi(t), & \text{when } t > 0; \\ \Phi(-Z), & \text{when } t = 0; \end{cases} & \text{if } Z \leq 0. \end{cases} \quad \text{When } Z \rightarrow \infty,$$
 this distribution converges to the standard normal. When $Z \rightarrow -\infty$, this distribution converges to a point mass at zero. The asymptotic average bootstrap distribution is
$$\begin{cases} 2\phi(t)\Phi(t), & \text{when } t < 0; \\ \int_{-\infty}^0 \Phi(-z)\phi(z)dz, & \text{when } t = 0; \\ \frac{1}{2}\phi(t) + \int_{-\infty}^0 \phi(t-z)\phi(z)dz, & \text{when } t > 0; \end{cases} \quad \text{where } \phi(\cdot) \text{ is the standard normal density, and } \Phi(\cdot) \text{ is the standard normal cdf, which has a point mass only at zero.}$$

does not cover γ_0 for almost all original sample paths. On the contrary, Remark 3.6 in Chernozhukov and Hong (2004) shows that the parametric wild bootstrap is valid in constructing confidence intervals for the boundary parameters they considered. This difference can be explained as follows. We know the parametric bootstrap is valid because it maintains the probability structure around the boundary. In Chernozhukov and Hong (2004), the boundary parameters appear in the conditional distribution of y on x , and there is no boundary parameter in the distribution of x , so simulating from $f(\epsilon|x; \hat{\alpha})$ in their setup is enough to mimic the original probability structure around the boundary. In threshold regression, however, γ is a boundary of q , so we must simulate from the joint distribution $f(y, x, q)$, which covers both $f(\epsilon|x, q; \hat{\alpha})$ and $f(x, q)$, to keep the probability structure around γ . See Algorithm 1 and 2 in Section 4.2 of Yu (2007) for a concrete description of such a procedure. In practice, even in parametric models, $f(x, q)$ is seldom specified, so the parametric bootstrap was never used in applications.⁹ As shown in Yu (2007), the Bayesian credible set is a good choice in parametric models since it does not rely on the specific form of $f(x, q)$.

4 Numerical Examples

It is appropriate to pause here to provide more intuition behind Theorems 1 and 2 by considering the following two numerical examples. We first apply the general results in Section 3 to the simple example in Section 2, then consider a more practical example where the error term is present. To simplify notations, Z^* and Z without subscripts are used for Z_γ^* and Z_γ in the following discussion.

4.1 When There is No Error Term: A Revisit

In the case without the error term,

$$D(v) = \begin{cases} N_1(|v|), & \text{if } v \leq 0; \\ N_2(v), & \text{if } v > 0; \end{cases}$$

and

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

Now,

$$P_n^*(n(\hat{\gamma}^* - \hat{\gamma}) = 0 | F_n) \longrightarrow P^*(N_{1-}^* > 0) = 1 - e^{-1},$$

and for $t \leq 0$,

$$\begin{aligned} & P_n^*(n(\hat{\gamma}^* - \hat{\gamma}) < t | F_n) \\ \longrightarrow & \begin{cases} P^*(N_{1-}^* = 0) = e^{-1}, & \text{if } N(|t|) = 0, \\ P^*(N_{1-}^* = 0, N_{2-}^* = 0) = e^{-2}, & \text{if } N(|t|) = 1, \\ \vdots & \vdots \\ P^*(\text{Poisson}(k+1) = 0) = e^{-(k+1)}, & \text{if } N(|t|) = k, \\ \vdots & \vdots \end{cases} \\ = & e^{-(1+N(|t|))} \end{aligned}$$

⁹When q is independent of (x, ϵ) , we can condition on $\{x_i\}_{i=1}^n$, and simulate only from $f(\epsilon|x_i, q; \hat{\alpha})$ and $f(q)$. But the problem remains since $f(q)$ is seldom known in reality.

where $N(|t|)$ is a *truncated Poisson process* starting from $t_0 \equiv \sup \{t : N_1(|t|) = 0\}$. This $N(|t|)$ is the same $N(|t|)$ as in Section 2.1. Because $n(\hat{\gamma}^* - \gamma_0) | F_n$ is a location shift of $n(\hat{\gamma}^* - \hat{\gamma}) | F_n$, it can be shown that for $t \leq 0$,

$$P_n^*(n(\hat{\gamma}^* - \gamma_0) < t | F_n) \longrightarrow e^{-N_1(|t|)}.^{10}$$

The point mass of $Z^* | N_1(\cdot)$ at the k th jump of $D(v)$ on $v \leq 0$ is

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

which is exponentially decaying. Under P_r , for $t \leq 0$,

$$P_r(n(\hat{\gamma}^* - \gamma_0) < t) \rightarrow E[e^{-N_1(|t|)}] = \exp\{t - t/e\},$$

which is continuous. Note that Z has a thinner tail than Z^* and $Z^* - Z$. For comparison, their densities on $t < 0$ are listed below:

$$\begin{aligned} f_Z(t) &= e^t; \\ f_{Z^*}(t) &= \exp\{t - t/e\} (1 - e^{-1}); \\ f_{Z^* - Z}(t) &= \exp\{t - t/e - 1\} (1 - e^{-1}). \end{aligned}$$

4.2 When There is An Error Term

Suppose the population model is

$$y = \mathbf{1}(q \leq \gamma) + e, \quad q \sim U[0, 1], \quad (6)$$

where $e \sim N(0, 1)$ is independent of q . This setup is the same as (3) except that an error term e is added in. γ is the only parameter of interest.

From (5), the asymptotic distribution of the LSE is as follows:

$$n(\hat{\gamma} - \gamma_0) \xrightarrow{d} \arg \min_v D(v)$$

where

$$D(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} z_{1i} = \sum_{i=1}^{N_1(|v|)} (1 + 2e_i^-), & \text{if } v \leq 0; \\ \sum_{i=1}^{N_2(v)} z_{2i} = \sum_{i=1}^{N_2(v)} (1 - 2e_i^+), & \text{if } v > 0; \end{cases}$$

$\{e_i^-, e_i^+, i = 1, \dots, N_1(\cdot), N_2(\cdot)\}$ are independent of each other, e_i^- and e_i^+ follow the same distribution as e , and $N_1(\cdot)$ and $N_2(\cdot)$ are standard Poisson processes.

From Theorem 2,

$$n(\hat{\gamma}^* - \hat{\gamma}) \xrightarrow{d} \arg \min_v D^*(v) - \arg \min_v D(v) \text{ in } P^\infty \text{ probability,}$$

where $D(v)$ is defined above and

$$D^*(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} z_{1i}^* = \sum_{i=1}^{N_1(|v|)} N_{i-}^* z_{1i}, & \text{if } v \leq 0; \\ \sum_{i=1}^{N_2(v)} z_{2i}^* = \sum_{i=1}^{N_2(v)} N_{i+}^* z_{2i}, & \text{if } v > 0. \end{cases}$$

¹⁰The asymptotic distribution of $n(\hat{\gamma}^* - \gamma_0) | F_n$ is discrete but has no point at 0. This is easy to understand since the support of $Z^* | N_1(\cdot)$ is the jumping locations of $N_1(\cdot)$, and $N_1(\cdot)$ does not have a jump at zero.

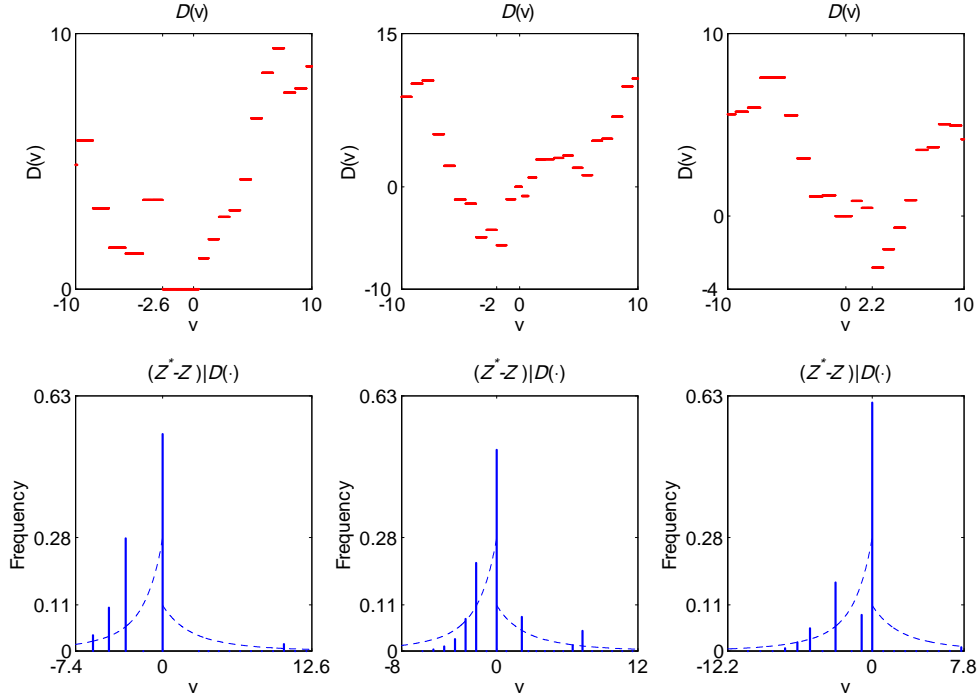


Figure 3: Dependence of the Distribution of $Z^* - Z$ on $D(\cdot)$

Figure 3 shows the dependence of the distribution of $Z^* - Z$ on $D(\cdot)$. For comparison, the density of Z is also dotted in Figure 3, which is very different from the conditional distribution of $(Z^* - Z) | D(\cdot)$ in all cases. The support of $(Z^* - Z) | D(\cdot)$ is a subset of the jumping locations of $D(\cdot)$. For the three sample paths of $D(\cdot)$ in Figure 3, $\arg \min_v D(v)$ is obtained at the 0th, 2nd on $v \leq 0$, and 3rd on $v > 0$ jump, respectively. Compared to Figure 1, there are some differences in the distribution of $(Z^* - Z) | D(\cdot)$ with and without the error term. First, there is positive probability on the positive axis. Second, not every jumping location of $D(\cdot)$ on $v \leq 0$ 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 $1 - e^{-1}$.

The distribution of $(Z^* - Z) | D(\cdot)$ has three main characteristics. First, the largest point mass is at zero in all cases and depends on $D(\cdot)$.¹¹ For example, when $Min_- = 0$, where Min_- is the number of jumps before attaining the minimum of $D(v)$ on $v \leq 0$,

$$\begin{aligned}
& P_n^*(\hat{\gamma}^* = \hat{\gamma} | F_n) \\
\rightarrow & P^*(Z_1^* > 0, Z_2^* \geq 0 | D(\cdot)) \\
= & (1 - F_1^*(0))(1 - F_2^*(0-))
\end{aligned}$$

where $F_1^*(\cdot)$ is the conditional cdf of $\min \left\{ \sum_{i=1}^k z_{1i}^*, k = 1, 2, \dots \right\}$, and $F_2^*(\cdot)$ is similarly defined.¹² z_{1i}^* or

¹¹For comparison, Z has the largest density also at zero.

¹²Note that on such sample paths, $Z_1 > 0$ and $Z_2 \geq 0$ with $Z_1 = \min \left\{ \sum_{i=1}^k z_{1i}, k = 1, 2, \dots \right\}$ and $Z_2 =$

z_{2i}^* follows a discrete conditional distribution, so $F_2^*(0-) \neq F_2^*(0)$ in general.¹³ This is different from what happens in Appendix D of Yu (2007), where $F_2(0-) = F_2(0)$ since z_{2i} follows an absolute continuous distribution. Furthermore, the conditional distribution of z_{1i}^* (z_{2i}^*) depends on i . So

$$\begin{aligned}
& P^*(Z_1^* > 0) \\
&= P^*(N_{1-}^* z_{11} > 0, N_{1-}^* z_{11} + N_{2-}^* z_{12} > 0, \dots) \\
&= P^*(N_{1-}^* > 0, N_{1-}^* z_{11} + N_{2-}^* z_{12} > 0, N_{1-}^* z_{11} + N_{2-}^* z_{12} + N_{3-}^* z_{13} > 0, \dots) \\
&\leq 1 - e^{-1}.
\end{aligned}$$

This probability depends on the realized value of $\{z_{1i}\}_{i=1}^\infty$. If $z_{1i} \geq 0$ for all $i \geq 2$, then $P^*(Z_1^* > 0) = 1 - e^{-1}$.¹⁴ For a general realization of $\{z_{1i}\}_{i=1}^\infty$, the probability $P^*(Z_1^* > 0)$ is not tractable. Similar arguments can apply to $P^*(Z_2^* \geq 0)$. When $Min_- \neq 0$, the calculation is more complicated.

Second, large point masses often happen at deeply negative jumps, and at the left of them there are decaying point masses. This will be illustrated by the following calculations. 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, $P^*(Min_- = 1) = 0$.¹⁵

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

and

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

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 more complicated sample paths of $D(v)$, the decaying rate may not be exactly e^{-1} .

Third, there is no point mass in the right neighborhood of 0. For example, when $Min_- = 0$, there are no point masses on $0 < v < |v_{0-}| + v_{1+}$, where $v_{k-} = \sup\{v : N_1(|v|) = k\}$ and $v_{k+} = \sup\{v : N_2(v) = k\}$ for $k \geq 0$. This phenomenon is due to the fact that the left endpoint of the minimizing interval is taken as the estimator.

$\min \left\{ \sum_{i=1}^k z_{2i}, k = 1, 2, \dots \right\}$, which implies $z_{11} > 0$.

¹³For example, suppose $z_{2i} > 0$ for all $i \geq 1$, then $F_2^*(0-) = 0$, but $F_2^*(0) = 1 - P^*(Z_2^* > 0) = 1 - (1 - e^{-1}) = e^{-1}$.

¹⁴Such sample paths have P^∞ probability 0.

¹⁵There 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 $v > 0$.

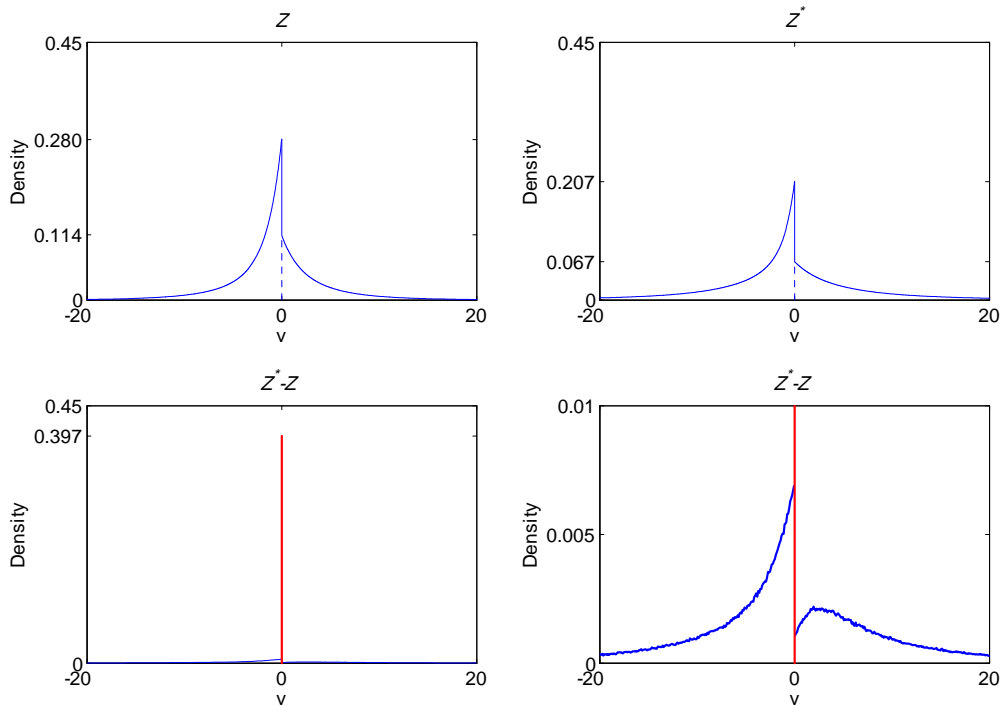


Figure 4: Comparison of the Asymptotic Distributions under P_r

The distributions of Z , Z^* and $Z^* - Z$ under P_r are shown in Figure 4. The distribution of Z^* has a thicker tail than Z as expected. The left thick tail is due to the fact that N_{i-}^* can take value 0 on positive z_{1i} 's and values greater than 1 on negative z_{1i} 's, and the right thick tail is due to the fact that N_{i+}^* can take values greater than 1 on negative z_{2i} 's. The distribution of $Z^* - Z$ is approximated by 1 million simulated draws. The striking feature of this distribution is that there is a point mass (less than $1 - e^{-1}$) at zero. Also, the magnified version of the distribution of $Z^* - Z$ at bottom right of Figure 3 shows that there is no much density in the right neighborhood of zero, which can be derived from the third point above.

Although Figure 3 and 4 show that the asymptotic bootstrap distribution is very different from the asymptotic distribution, the bootstrap is still useful in practice if the quantiles are close to each other. In Table 1 and Figure 5, some simulation results are reported for the example above, where an equal-tailed 95% coverage confidence interval is used as the bootstrap confidence interval. In Table 1, "Average" means the average of the quantiles and coverage among all sample paths of $D(\cdot)$. The numbers under "Asymptotic" are the benchmark for the bootstrap inference. From Table 1 and Figure 5, the 2.5% and 97.5% quantiles of the asymptotic bootstrap distribution depends heavily on the original samples, and the same happens to the coverage. In Table 1, the items under "Average" are different from those under "Average Bootstrap" because the two operations of taking a quantile and taking an average of distributions can not be exchanged. By comparing the quantiles under "Asymptotic" and "Average Bootstrap", we conclude that $Z^* - Z$ has a thicker tail than Z as in the case without the error term. In Figure 5, there is a large point mass at zero in the distribution of the 97.5% quantile. This is not surprising according to Figure 3 where the asymptotic bootstrap probability on the positive axis is small for the three representative sample paths of $D(\cdot)$. In the distribution of coverage, an interesting phenomenon is that there is a hump between 0.6 and 0.7. Table 1

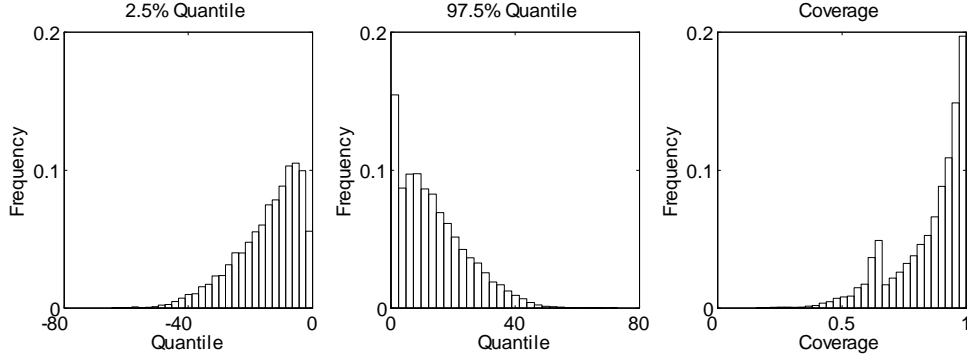


Figure 5: Distributions of 2.5% and 97.5% Quantiles and the Coverage

and Figure 5 suggest that it is risky to use the bootstrap confidence interval as the set estimation of the threshold point.

	2.5% Quantile	97.5% Quantile	Coverage
Asymptotic	-12.83	11.74	95.00%
Min	-64.68	0	15.49%
Max	-0.02	73.00	99.99%
Average	-14.55	13.62	84.93%
Average Bootstrap	-21.86	21.44	99.01%

Table 1: Quantiles and Coverage under the Asymptotic Bootstrap Distribution

5 Remedies in the Semiparametric Case

As shown in Section 3, the nonparametric bootstrap is inconsistent for inference of the threshold point in discontinuous threshold regression, so some remedies are needed to the nonparametric bootstrap failure. Another motivation of the remedies comes from the difficulty of statistical inference based on the asymptotic distribution (5).

Four remedies are potentially useful. The first one is suggested in Hansen (2000). Hansen (2000) uses a framework with an asymptotically diminishing threshold effect in mean and proves that the confidence interval constructed by inverting the likelihood ratio statistic is asymptotically valid. As mentioned in the introduction, the nonparametric bootstrap is also valid in this framework. But the nonparametric bootstrap is not a practical choice as there is no way to distinguish a real dataset following the framework of Hansen (2000) or that used in this paper.¹⁶

The second remedy in the literature is the subsampling method in discontinuous threshold regression by Gonzalo and Wolf (2005). The subsampling method is asymptotically valid as long as there is an asymptotic distribution which is continuous. Gonzalo and Wolf (2005) do not prove the continuity of the asymptotic distribution (5), and Appendix D of Yu (2007) fills this gap.

The third remedy is proposed in Seo and Linton (2007), which is based on the smoothed least squares estimation in our framework. The drawbacks of this method are that the convergence rate is less than n and a smoothing parameter needs to be specified in practice.

¹⁶Note that Theorem 3 in Hansen (2000) guarantees that the confidence interval there is at least conservative in a dominating case of Chan (1993)'s framework.

The fourth remedy is suggested by Yu (2008). Yu (2008) uses the nonparametric posterior to construct confidence intervals for γ in the present framework. The simulation study there indicates that this method performs better than other methods under some standard setups.

6 Conclusion

In this paper, we show that the nonparametric bootstrap is inconsistent and the parametric bootstrap is consistent for inference of the threshold point in discontinuous threshold regression. It is found that the asymptotic nonparametric bootstrap distribution depends on the sample path of the original data. Such a phenomenon also appears in Andrews (2000). The method developed in this paper is generic in deriving the asymptotic bootstrap distribution. The remedies to the nonparametric bootstrap failure in the semiparametric case are summarized.

References

- Abadie, A. and G.W. Imbens, 2008, On the Failure of the Bootstrap for Matching Estimators, *Econometrica*, 1537-1557.
- Abrevaya, J. and J. Huang, 2005, On the Bootstrap of the Maximum Score Estimator, *Econometrica*, 73, 1175-1204.
- Aldous, D., 1978, Stopping Times and Tightness, *The Annals of Probability*, 6, 335-340.
- Aldous, D.J., 1985, Exchangeability and Related Topics, *Lecture Notes in Math.* 1117, 1-198, Springer, New York.
- Andrews, D.W.K., 2000, Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space, *Econometrica*, 68, 399-405.
- Antoch, J. et al, 1995, Change-point Problem and Bootstrap, *Journal of Nonparametric Statistics*, 5, 123-144.
- Beran, R., 1997, Diagnosing Bootstrap Success, *Annals of the Institute of Statistical Mathematics*, 49, 1-24.
- Bickel, P.J. et al, 1997, Resampling Fewer than n Observations: Gains, Losses, and Remedies for Losses, *Statistica Sinica*, 7, 1-31.
- Bickel, P.J. and D.A. Freedman, 1981, Some Asymptotic Theory for the Bootstrap, *The Annals of Statistics*, 9, 1196-1217.
- Chan, K.S., 1993, Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model, *The Annals of Statistics*, 21, 520-533.
- Chernozhukov, V. and H. Hong, 2004, Likelihood Estimation and Inference in a Class of Nonregular Econometric Models, *Econometrica*, 72, 1445-1480.
- Dümbgen, L., 1991, The Asymptotic Behavior of Some Nonparametric Change-point Estimators, *The Annals of Statistics*, 19, 1471-1495.
- Efron, B., 1979, Bootstrap Methods: Another Look at the jackknife, *The Annals of Statistics*, 7, 1-26.

- Gonzalo, J. and M. Wolf, 2005, Subsampling inference in Threshold Autoregressive Models, *Journal of Econometrics*, 127, 201-224.
- Hansen, B.E., 2000, Sample Splitting and Threshold Estimation, *Econometrica*, 575-603.
- Horowitz, J.L., 2001, The Bootstrap, *Handbook of Econometrics*, Vol. 5, J.J. Heckman and E.E. Leamer, eds., Elsevier Science B.V., Ch. 52, 3159-3228.
- Hušková, M. and C. Kirch, 2008, Bootstrapping Confidence Intervals for the Change-point of Time Series, *Journal of Time Series Analysis*, 29, 947-972.
- Huang, J. and J.A. Wellner, 1995, Estimation of a Monotone Density or Monotone Hazard Under Random Censoring, *Scandinavian Journal of Statistics*, 22, 3-33.
- Kac, M., 1949, On Deviations between Theoretical and Empirical Distributions, *Proceedings of the National Academy of Sciences USA* 35, 252-257.
- Kim, J. and D. Pollard, 1990, Cube Root Asymptotics, *The Annals of Statistics*, 18, 191-219.
- Klaassen, C. A. and Wellner, J. A., 1992, Kac Empirical Processes and the Bootstrap, in M. Hahn and J. Kuelbs, eds., *Proceedings of the Eighth International Conference on Probability in Banach Spaces*, 411-429, Birkhäuser, New York.
- Lee, S. and M.H. Seo, 2008, Semiparametric Estimation of a Binary Response Model with a Change-point Due to a Covariate Threshold, *Journal of Econometrics*, 144, 492-499.
- Newey, W.K. and D.L. McFadden, 1994, Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics*, Vol. 4, R.F. Egle and D.L. McFadden, eds., Elsevier Science B.V., Ch. 36, 2113-2245.
- Pakes, A., and D. Pollard, 1989, Simulation and the Asymptotics of Optimization Estimators, *Econometrica*, 57, 1027-1057.
- Pollard, D., 1984, *Convergence of Stochastic Processes*, New York : Springer-Verlag.
- Seo, M.H. and O. Linton, 2007, A Smoothed Least Squares Estimator for Threshold Regression Models, *Journal of Econometrics*, 141, 704-735.
- Van der Vaart, A.W. and J.A. Wellner, 1996, *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York : Springer.
- Wellner, J.A. and Y. Zhan, 1996, Bootstrapping Z-Estimators, unpublished manuscript, Department of Statistics, University of Washington.
- Yu, P., 2007, Likelihood-based Estimation and Inference in Threshold Regression, unpublished manuscript, Department of Economics, University of Wisconsin at Madison.
- Yu, P., 2008, Semiparametric Efficient Estimation in Threshold Regression, unpublished manuscript, Department of Economics, University of Wisconsin at Madison.

Appendix A: Proofs

First, some notations are collected for reference in all lemmas and proofs. The letter C is used as a generic positive constant, which need not be the same in each occurrence.

$$\theta_\ell = (\beta'_\ell, \sigma_\ell)', \ell = 1, 2$$

$$\begin{aligned} m(w|\theta) &= (y - x'\beta_1\mathbf{1}(q \leq \gamma) - x'\beta_2\mathbf{1}(q > \gamma))^2, \\ M_n(\theta) &= P_n(m(\cdot|\theta)), \\ M(\theta) &= P(m(\cdot|\theta)), \\ \mathbb{G}_n m &= \sqrt{n}(M_n - M). \end{aligned}$$

$$\begin{aligned} \bar{z}_1(w|\theta_2, \tilde{\theta}_1) &= (\tilde{\beta}_1 - \beta_2)' xx' (\tilde{\beta}_1 - \beta_2) + 2\tilde{\sigma}_1 (\tilde{\beta}_1 - \beta_2) xe, \text{ so } \bar{z}_{1i} = \bar{z}_1(w_i|\theta_{20}, \theta_{10}), \\ \bar{z}_2(w|\theta_1, \tilde{\theta}_2) &= (\tilde{\beta}_2 - \beta_1)' xx' (\tilde{\beta}_2 - \beta_1) + 2\tilde{\sigma}_2 (\tilde{\beta}_2 - \beta_1) xe, \text{ so } \bar{z}_{2i} = \bar{z}_2(w_i|\theta_{10}, \theta_{20}), \end{aligned}$$

The following formulas are used repetitively in the following analysis:

$$\begin{aligned} m(w|\theta) &= (x'(\beta_{10} - \beta_1) + \sigma_{10}e)^2 \mathbf{1}(q_i \leq \gamma \wedge \gamma_0) + (x'(\beta_{20} - \beta_2) + \sigma_{20}e)^2 \mathbf{1}(q_i > \gamma \vee \gamma_0) \\ &+ (x'(\beta_{10} - \beta_2) + \sigma_{10}e)^2 \mathbf{1}(\gamma \wedge \gamma_0 < q_i \leq \gamma_0) + (x'(\beta_{20} - \beta_1) + \sigma_{20}e)^2 \mathbf{1}(\gamma_0 < q \leq \gamma \vee \gamma_0), \end{aligned}$$

so

$$\begin{aligned} &m(w|\theta) - m(w|\theta_0) \\ &= [(\beta_{10} - \beta_1)' xx' (\beta_{10} - \beta_1) + 2\sigma_{10}(\beta_{10} - \beta_1) xe] \mathbf{1}(q \leq \gamma \wedge \gamma_0) \\ &\quad + [(\beta_{20} - \beta_2)' xx' (\beta_{20} - \beta_2) + 2\sigma_{20}(\beta_{20} - \beta_2) xe] \mathbf{1}(q > \gamma \vee \gamma_0) \\ &\quad + \bar{z}_1(w|\theta_2, \theta_{10}) \mathbf{1}(\gamma \wedge \gamma_0 < q \leq \gamma_0) + \bar{z}_2(w|\theta_1, \theta_{20}) \mathbf{1}(\gamma_0 < q \leq \gamma \vee \gamma_0) \\ &\equiv T(w|\theta_1, \theta_{10}) \mathbf{1}(q \leq \gamma \wedge \gamma_0) + T(w|\theta_2, \theta_{20}) \mathbf{1}(q > \gamma \vee \gamma_0) \\ &\quad + \bar{z}_1(w|\theta_2, \theta_{10}) \mathbf{1}(\gamma \wedge \gamma_0 < q \leq \gamma_0) + \bar{z}_2(w|\theta_1, \theta_{20}) \mathbf{1}(\gamma_0 < q \leq \gamma \vee \gamma_0) \\ &\equiv A(w|\theta) + B(w|\theta) + C(w|\theta) + D(w|\theta). \end{aligned}$$

Proof of Theorem 1. This proof includes two parts: (i) the finite-dimensional limit distributions of $(W_n(u), W_n^*(u), D_n(v), D_n^*(v))$ are the same as specified in the theorem; (ii) the process $(W_n(u), W_n^*(u), D_n(v), D_n^*(v))$ is stochastically equicontinuous.

Part (i): We only prove the result for a fixed h , or the Cramér-Wold device can be used. Define

$$\begin{aligned} T_{1i} &= \frac{1}{\sqrt{n}} x_i e_i \mathbf{1}(q_i \leq \gamma_0) \equiv \frac{1}{\sqrt{n}} S_{1i}, \\ T_{2i} &= \frac{1}{\sqrt{n}} x_i e_i \mathbf{1}(q_i > \gamma_0) \equiv \frac{1}{\sqrt{n}} S_{2i}, \\ T_{3i} &= \bar{z}_{1i} \mathbf{1}\left(\gamma_0 + \frac{v_1}{n} < q_i \leq \gamma_0\right), \\ T_{4i} &= \bar{z}_{2i} \mathbf{1}\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v_2}{n}\right), \end{aligned}$$

where $v_1 \leq 0$ and $v_2 > 0$. Since

$$\begin{aligned}\exp\{\sqrt{-1}t_{11}T_{3i}\} &= 1 + \mathbf{1}\left(\gamma_0 + \frac{v_1}{n} < q_i \leq \gamma_0\right) [\exp\{\sqrt{-1}t_{11}\bar{z}_{1i}\} - 1], \\ \exp\{\sqrt{-1}t_{12}T_{4i}\} &= 1 + \mathbf{1}\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v_2}{n}\right) [\exp\{\sqrt{-1}t_{12}\bar{z}_{2i}\} - 1], \\ \exp\{\sqrt{-1}t_{21}M_{N_n,i}T_{3i}\} &= 1 + \mathbf{1}\left(\gamma_0 + \frac{v_1}{n} < q_i \leq \gamma_0\right) [\exp\{\sqrt{-1}t_{21}M_{N_n,i}\bar{z}_{1i}\} - 1], \\ \exp\{\sqrt{-1}t_{22}M_{N_n,i}T_{4i}\} &= 1 + \mathbf{1}\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v_2}{n}\right) [\exp\{\sqrt{-1}t_{22}M_{N_n,i}\bar{z}_{2i}\} - 1],\end{aligned}$$

it follows

$$\begin{aligned}& P_r \left(\exp \left\{ \sqrt{-1} \begin{bmatrix} s'_{11}T_{1i} + s'_{12}T_{2i} + s'_{21}(M_{N_n,i} - 1)T_{1i} + s'_{22}(M_{N_n,i} - 1)T_{2i} \\ + t_{11}T_{3i} + t_{12}T_{4i} + t_{21}M_{N_n,i}T_{3i} + t_{22}M_{N_n,i}T_{4i} \end{bmatrix} \right\} \right) \\ &= P_r \left(\exp \left\{ \sqrt{-1}s'T_i^R/\sqrt{n} \right\} + \frac{v_1}{n}f_q(\gamma_0)P_r \left(\exp \left\{ \sqrt{-1}s'T_i^R/\sqrt{n} \right\} [\exp\{\sqrt{-1}t_{11}\bar{z}_{1i}\} - 1] \mid q_i = \gamma_0- \right) \right. \\ &\quad + \frac{v_2}{n}f_q(\gamma_0)P_r \left(\exp \left\{ \sqrt{-1}s'T_i^R/\sqrt{n} \right\} [\exp\{\sqrt{-1}t_{12}\bar{z}_{2i}\} - 1] \mid q_i = \gamma_0+ \right) \\ &\quad + \frac{v_1}{n}f_q(\gamma_0)P_r \left(\exp \left\{ \sqrt{-1}s'T_i^R/\sqrt{n} \right\} [\exp\{\sqrt{-1}t_{21}M_{N_n,i}\bar{z}_{1i}\} - 1] \mid q_i = \gamma_0- \right) \\ &\quad \left. + \frac{v_2}{n}f_q(\gamma_0)P_r \left(\exp \left\{ \sqrt{-1}s'T_i^R/\sqrt{n} \right\} [\exp\{\sqrt{-1}t_{22}M_{N_n,i}\bar{z}_{2i}\} - 1] \mid q_i = \gamma_0+ \right) + o\left(\frac{1}{n}\right) \right) \\ &= 1 + \frac{1}{n} \left[-\frac{1}{2}s'\mathcal{J}s + f_q(\gamma_0)v_1 (E[\{\exp\{\sqrt{-1}t_{11}\bar{z}_{1i}\}\} \mid q_i = \gamma_0-] - 1) \right. \\ &\quad + f_q(\gamma_0)v_2 (E[\{\exp\{\sqrt{-1}t_{12}\bar{z}_{2i}\}\} \mid q_i = \gamma_0+] - 1) + f_q(\gamma_0)v_1 (E[\{\exp\{\sqrt{-1}t_{21}M_{N_n,i}\bar{z}_{1i}\}\} \mid q_i = \gamma_0-] - 1) \\ &\quad \left. + f_q(\gamma_0)v_2 (E[\{\exp\{\sqrt{-1}t_{22}M_{N_n,i}\bar{z}_{2i}\}\} \mid q_i = \gamma_0+] - 1) \right] + o\left(\frac{1}{n}\right),\end{aligned}$$

where $s = (s'_{11} \ s'_{12} \ s'_{21} \ s'_{22})'$, $T_i^R = \begin{pmatrix} S'_{1i} & S'_{2i} & (M_{N_n,i} - 1)S'_{1i} & (M_{N_n,i} - 1)S'_{2i} \end{pmatrix}'$, $o(1)$ in the first equality is a quantity going to zero uniformly over $i = 1, \dots, n$ from Assumption D4, the last equality is from the Taylor expansion of $\exp\{\sqrt{-1}s'T_i^R/\sqrt{n}\}$, and

$$\begin{aligned}\mathcal{J} &= P_r(T_i^R T_i^{R'}) \\ &= \begin{pmatrix} E[xx'e^2\mathbf{1}(q \leq \gamma_0)] & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & E[xx'e^2\mathbf{1}(q > \gamma_0)] & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E[xx'e^2\mathbf{1}(q \leq \gamma_0)] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & E[xx'e^2\mathbf{1}(q > \gamma_0)] \end{pmatrix}.\end{aligned}$$

So

$$\begin{aligned}
& P_r \left(\exp \left\{ \sqrt{-1} \left[\begin{array}{l} s_{11} \sum_{i=1}^n T_{1i} + s_{12} \sum_{i=1}^n T_{2i} + s_{21} \sum_{i=1}^n (M_{N_n,i} - 1) T_{1i} + s_{22} \sum_{i=1}^n (M_{N_n,i} - 1) T_{2i} \\ + t_{11} \sum_{i=1}^n T_{3i} + t_{12} \sum_{i=1}^n T_{4i} + t_{21} \sum_{i=1}^n M_{N_n,i} T_{3i} + t_{22} \sum_{i=1}^n M_{N_n,i} T_{4i} \end{array} \right] \right\} \right) \\
&= \prod_{i=1}^n P_r \left(\exp \left\{ \sqrt{-1} \left[s' T_i^R / \sqrt{n} + t_{11} T_{3i} + t_{12} T_{4i} + t_{21} M_{N_n,i} T_{3i} + t_{22} M_{N_n,i} T_{4i} \right] \right\} \right) \\
&\rightarrow \exp \left\{ -\frac{1}{2} s' \mathcal{J} s + f_q(\gamma_0) v_1 (E [\{\exp \{\sqrt{-1} t_{11} \bar{z}_{1i}\}\} | q_i = \gamma_0^-] - 1) \right. \\
&\quad + f_q(\gamma_0) v_2 (E [\{\exp \{\sqrt{-1} t_{12} \bar{z}_{2i}\}\} | q_i = \gamma_0^+] - 1) + f_q(\gamma_0) v_1 (E [\{\exp \{\sqrt{-1} t_{21} M_{N_n,i} \bar{z}_{1i}\}\} | q_i = \gamma_0^-] - 1) \\
&\quad \left. + f_q(\gamma_0) v_2 (E [\{\exp \{\sqrt{-1} t_{22} M_{N_n,i} \bar{z}_{2i}\}\} | q_i = \gamma_0^+] - 1) \right\},
\end{aligned}$$

and the result of interest follows. Note that $\{N_{i-}^*, N_{i+}^*\}_{i \geq 1}$ is just $\{M_{N_n,i}\}_{i \geq 1}$, and $E[M_{N_n,i} \bar{z}_{\ell i} | q_i = \gamma_0^+] = M_{N_n,i} z_{\ell i}$, $\ell = 1, 2$, since $\{M_{N_n,i}\}_{i \geq 1}$ is independent of the data.

Part (ii): The stochastic equicontinuity of $W_n(u)$ and $W_n^*(u)$ can be trivially proved since they are linear functions of u . Now, we concentrate on $D_n(v)$ and $D_n^*(v)$. For this purpose, a condition called Aldous's (1978) condition is sufficient; see Theorem 16 on Page 134 of Pollard (1984). Without loss of generality, we only prove the result for $v > 0$. Suppose $0 < v_1 < v_2$ are stopping times in a compact set K , then for any $\varepsilon > 0$,

$$\begin{aligned}
& P^\infty \left(\sup_{|v_2 - v_1| < \delta} |D_n(v_2) - D_n(v_1)| > \varepsilon \right) \\
&\stackrel{(1)}{\leq} P \left(\sum_{i=1}^n |\bar{z}_{2i}| \cdot \sup_{|v_2 - v_1| < \delta} \mathbf{1} \left(\gamma_0 + \frac{v_1}{n} < q_i \leq \gamma_0 + \frac{v_2}{n} \right) > \varepsilon \right) \\
&\stackrel{(2)}{\leq} \sum_{i=1}^n E \left[|\bar{z}_{2i}| \sup_{|v_2 - v_1| < \delta} \mathbf{1} \left(\gamma_0 + \frac{v_1}{n} < q_i \leq \gamma_0 + \frac{v_2}{n} \right) \right] / \varepsilon \\
&\stackrel{(3)}{\leq} \frac{C\delta}{\varepsilon},
\end{aligned}$$

where (1) is obvious, (2) is from Markov's inequality, and C in (3) can take $\bar{f}_q \sup_{\gamma_0 < \gamma \leq \gamma_0 + \varepsilon} E[|\bar{z}_{2i}| | q_i = \gamma] < \infty$ for some $\varepsilon > 0$ from Assumptions D4 and D6. Similarly,

$$\begin{aligned}
& P_r \left(\sup_{|v_2 - v_1| < \delta} |D_n^*(v_2) - D_n^*(v_1)| > \varepsilon \right) \\
&\leq P \left(\sum_{i=1}^n |M_{N_n,i} \bar{z}_{2i}| \cdot \sup_{|v_2 - v_1| < \delta} \mathbf{1} \left(\gamma_0 + \frac{v_1}{n} < q_i \leq \gamma_0 + \frac{v_2}{n} \right) > \varepsilon \right) \\
&\leq \sum_{i=1}^n E \left[|M_{N_n,i} \bar{z}_{2i}| \sup_{|v_2 - v_1| < \delta} \mathbf{1} \left(\gamma_0 + \frac{v_1}{n} < q_i \leq \gamma_0 + \frac{v_2}{n} \right) \right] / \varepsilon \\
&\leq \frac{C\delta}{\varepsilon},
\end{aligned}$$

where C can take $\bar{f}_q \sup_{\gamma_0 < \gamma \leq \gamma_0 + \varepsilon} E[|M_{N_n,i} \bar{z}_{2i}| | q_i = \gamma] = \bar{f}_q \sup_{\gamma_0 < \gamma \leq \gamma_0 + \varepsilon} E[|\bar{z}_{2i}| | q_i = \gamma] < \infty$. ■

Proof of Theorem 2. This proof uses Lemma 5. In this model,

$$\begin{aligned} U_n + D_n &= nP_n \left(m \left(\cdot \left| \beta_0 + \frac{u}{\sqrt{n}}, \gamma_0 + \frac{v}{n} \right) - m(\cdot | \beta_0, \gamma_0) \right), \\ U_n^* + D_n^* &= nP_n^* \left(m \left(\cdot \left| \beta_0 + \frac{u}{\sqrt{n}}, \gamma_0 + \frac{v}{n} \right) - m(\cdot | \beta_0, \gamma_0) \right), \end{aligned}$$

and $d = 2k$. From Theorem 1,

$$\begin{aligned} \begin{pmatrix} U_n(u) + D_n(v) \\ U_n^*(u) + D_n^*(v) \end{pmatrix} &\rightsquigarrow u'_1 E[x_i x'_i \mathbf{1}(q_i \leq \gamma_0)] u_1 + u'_2 E[x_i x'_i \mathbf{1}(q_i > \gamma_0)] u_2 - 2u'_1 W_1 - 2u'_2 W_2 \\ &+ \begin{pmatrix} D(v) \\ -2u'_1 W_1^* - 2u'_2 W_2^* + D^*(v) \end{pmatrix} \equiv \begin{pmatrix} U(u) + D(v) \\ U^*(u) + D^*(v) \end{pmatrix}, \end{aligned}$$

and from Lemma 4,

$$\begin{aligned} \begin{pmatrix} s_n \\ t_n \end{pmatrix} &= \begin{pmatrix} \sqrt{n} \begin{pmatrix} \widehat{\beta}_1 - \beta_{10} \\ \widehat{\beta}_2 - \beta_{20} \end{pmatrix} \\ n(\widehat{\gamma} - \gamma_0) \end{pmatrix} = O_{P_\infty}(1), \\ \begin{pmatrix} s_n^* \\ t_n^* \end{pmatrix} &= \begin{pmatrix} \sqrt{n} \begin{pmatrix} \widehat{\beta}_1^* - \beta_{10} \\ \widehat{\beta}_2^* - \beta_{20} \end{pmatrix} \\ n(\widehat{\gamma}^* - \gamma_0) \end{pmatrix} = O_{P_r}(1). \end{aligned}$$

$\alpha_n = \alpha_n^* = 0$ in this model. So from Lemma 5,

$$\begin{pmatrix} (s'_n, t'_n)' \\ (s'^*_n, t'^*_n)' \end{pmatrix} \xrightarrow{d} \begin{pmatrix} (s', t)' \\ (s'^*, t^*)' \end{pmatrix},$$

where

$$\begin{aligned} s &= \arg \min_{u_1, u_2} u'_1 E[x_i x'_i \mathbf{1}(q_i \leq \gamma_0)] u_1 + u'_2 E[x_i x'_i \mathbf{1}(q_i > \gamma_0)] u_2 - 2u'_1 W_1 - 2u'_2 W_2 \\ &= \begin{pmatrix} E[x_i x'_i \mathbf{1}(q_i \leq \gamma_0)]^{-1} W_1 \\ E[x_i x'_i \mathbf{1}(q_i > \gamma_0)]^{-1} W_2 \end{pmatrix}, \\ t &= \arg \min_v D(v), \\ s^* &= \arg \min_{u_1, u_2} u'_1 E[x_i x'_i \mathbf{1}(q_i \leq \gamma_0)] u_1 + u'_2 E[x_i x'_i \mathbf{1}(q_i > \gamma_0)] u_2 - 2u'_1 W_1 - 2u'_2 W_2 - 2u'_1 W_1^* - 2u'_2 W_2^* \\ &= \begin{pmatrix} E[x_i x'_i \mathbf{1}(q_i \leq \gamma_0)]^{-1} (W_1 + W_1^*) \\ E[x_i x'_i \mathbf{1}(q_i > \gamma_0)]^{-1} (W_2 + W_2^*) \end{pmatrix}, \\ t^* &= \arg \min_v D^*(v). \end{aligned}$$

By the continuous mapping theorem,

$$\begin{aligned}
& (s_n^{*'}, t_n^{*'})' - (s_n', t_n')' \xrightarrow{d} (s^{*'}, t^{*'})' - (s', t)' \\
&= \begin{pmatrix} E[x_i x_i' \mathbf{1}(q_i \leq \gamma_0)]^{-1} (W_1 + W_1^*) - E[x_i x_i' \mathbf{1}(q_i \leq \gamma_0)]^{-1} W_1 \\ E[x_i x_i' \mathbf{1}(q_i > \gamma_0)]^{-1} (W_2 + W_2^*) - E[x_i x_i' \mathbf{1}(q_i > \gamma_0)]^{-1} W_2 \\ \arg \min_v D^*(v) - \arg \min_v D(v) \end{pmatrix} \\
&= \begin{pmatrix} E[x_i x_i' \mathbf{1}(q_i \leq \gamma_0)]^{-1} W_1^* \\ E[x_i x_i' \mathbf{1}(q_i > \gamma_0)]^{-1} W_2^* \\ \arg \min_v D^*(v) - \arg \min_v D(v) \end{pmatrix}.
\end{aligned}$$

The above discussion is under the P_r probability. From Lemma 2, $(s_n', t_n)'$ and $(s_n^{*'}, t_n^{*'})'$ are both $O_{P^*}(1)$ in P^∞ probability. Therefore, Lemma 5 can also be applied with respect to the probability measure P^* in P^∞ probability, yielding

$$\begin{pmatrix} (s_n', t_n)' \\ (s_n^{*'}, t_n^{*'})' \end{pmatrix} \xrightarrow{d} \begin{pmatrix} (s', t)' \\ (s^{*'}, t^{*'})' \end{pmatrix} \text{ in } P^\infty \text{ probability,}$$

whence by the continuous mapping theorem,

$$\begin{aligned}
& (s_n^{*'}, t_n^{*'})' - (s_n', t_n)' \xrightarrow{d} (s^{*'}, t^{*'})' - (s', t)' \\
&= \begin{pmatrix} E[x_i x_i' \mathbf{1}(q_i \leq \gamma_0)]^{-1} W_1^* \\ E[x_i x_i' \mathbf{1}(q_i > \gamma_0)]^{-1} W_2^* \\ \arg \min_v D^*(v) - \arg \min_v D(v) \end{pmatrix} \text{ in } P^\infty \text{ probability.}
\end{aligned}$$

■

Appendix B: Lemmas

Lemma 1 *Under Assumptions D4 and D5, for any $h \in \mathbb{R}^{2k+1}$,*

$$\begin{aligned}
& nP_n \left(m \left(\cdot \left| \beta_0 + \frac{u}{\sqrt{n}}, \gamma_0 + \frac{v}{n} \right. \right) - m(\cdot | \beta_0, \gamma_0) \right) \\
&= u_1' E[x_i x_i' \mathbf{1}(q_i \leq \gamma_0)] u_1 + u_2' E[x_i x_i' \mathbf{1}(q_i > \gamma_0)] u_2 - W_n(u) + D_n(v) + o_{P^\infty}(1),
\end{aligned}$$

and

$$\begin{aligned}
& nP_n^* \left(m \left(\cdot \left| \beta_0 + \frac{u}{\sqrt{n}}, \gamma_0 + \frac{v}{n} \right. \right) - m(\cdot | \beta_0, \gamma_0) \right) \\
&= u_1' E[x_i x_i' \mathbf{1}(q_i \leq \gamma_0)] u_1 + u_2' E[x_i x_i' \mathbf{1}(q_i > \gamma_0)] u_2 - W_n(u) - W_n^*(u) + D_n^*(v) + o_{P_r}(1),
\end{aligned}$$

where $W_n(u)$, $W_n^*(u)$, $D_n(v)$ and $D_n^*(v)$ are specified in Section 3.1.

Proof. First,

$$\begin{aligned}
& nP_n \left(m \left(\cdot \mid \beta_0 + \frac{u}{\sqrt{n}}, \gamma_0 + \frac{v}{n} \right) - m \left(\cdot \mid \beta_0, \gamma_0 \right) \right) \\
&= \sum_{i=1}^n \left(u_1' \frac{x_i x_i'}{n} u_1 - u_1' \frac{2\sigma_{10}}{\sqrt{n}} x_i e_i \right) \mathbf{1} \left(q_i \leq \gamma_0 \wedge \gamma_0 + \frac{v}{n} \right) + \sum_{i=1}^n \left(u_2' \frac{x_i x_i'}{n} u_2 - u_2' \frac{\sigma_{20}}{\sqrt{n}} x_i e_i \right) \mathbf{1} \left(q_i > \gamma_0 \vee \gamma_0 + \frac{v}{n} \right) \\
&+ \sum_{i=1}^n \left[\left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right)' x_i x_i' \left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right) + 2x_i' \left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right) \sigma_{10} e_i \right] \mathbf{1} \left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0 \right) \\
&+ \sum_{i=1}^n \left[\left(\beta_{10} + \frac{u_1}{\sqrt{n}} - \beta_{20} \right)' x_i x_i' \left(\beta_{10} + \frac{u_1}{\sqrt{n}} - \beta_{20} \right) - 2x_i' \left(\beta_{10} + \frac{u_1}{\sqrt{n}} - \beta_{20} \right) \sigma_{20} e_i \right] \mathbf{1} \left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n} \right) \\
&= u_1' E [x_i x_i' \mathbf{1} (q_i \leq \gamma_0)] u_1 + u_2' E [x_i x_i' \mathbf{1} (q_i > \gamma_0)] u_2 - W_n(u) + D_n(v) + o_{P^\infty}(1),
\end{aligned}$$

where $o_{P^\infty}(1)$ is from Assumption D4 and D5.

Second,

$$\begin{aligned}
& nP_n^* \left(m \left(\cdot \mid \beta_0 + \frac{u}{\sqrt{n}}, \gamma_0 + \frac{v}{n} \right) - m \left(\cdot \mid \beta_0, \gamma_0 \right) \right) \\
&= \sum_{i=1}^n M_{ni} \left(u_1' \frac{x_i x_i'}{n} u_1 - u_1' \frac{2\sigma_{10}}{\sqrt{n}} x_i e_i \right) \mathbf{1} \left(q_i \leq \gamma_0 \wedge \gamma_0 + \frac{v}{n} \right) + \sum_{i=1}^n M_{ni} \left(u_2' \frac{x_i x_i'}{n} u_2 - u_2' \frac{\sigma_{20}}{\sqrt{n}} x_i e_i \right) \mathbf{1} \left(q_i > \gamma_0 \vee \gamma_0 + \frac{v}{n} \right) \\
&+ \sum_{i=1}^n M_{ni} \left[\left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right)' x_i x_i' \left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right) + 2x_i' \left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right) \sigma_{10} e_i \right] \mathbf{1} \left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0 \right) \\
&+ \sum_{i=1}^n M_{ni} \left[\left(\beta_{10} + \frac{u_1}{\sqrt{n}} - \beta_{20} \right)' x_i x_i' \left(\beta_{10} + \frac{u_1}{\sqrt{n}} - \beta_{20} \right) - 2x_i' \left(\beta_{10} + \frac{u_1}{\sqrt{n}} - \beta_{20} \right) \sigma_{20} e_i \right] \mathbf{1} \left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n} \right) \\
&= \sum_{i=1}^n \left(u_1' \frac{x_i x_i'}{n} u_1 - u_1' \frac{2\sigma_{10}}{\sqrt{n}} x_i e_i \right) \mathbf{1} (q_i \leq \gamma_0) + \sum_{i=1}^n \left(u_2' \frac{x_i x_i'}{n} u_2 - u_2' \frac{\sigma_{20}}{\sqrt{n}} x_i e_i \right) \mathbf{1} (q_i > \gamma_0) - W_n^*(u) + D_n^*(v) + o_{P_r}(1) \\
&= u_1' E [x_i x_i' \mathbf{1} (q_i \leq \gamma_0)] u_1 + u_2' E [x_i x_i' \mathbf{1} (q_i > \gamma_0)] u_2 - W_n(u) - W_n^*(u) + D_n^*(v) + o_{P_r}(1),
\end{aligned}$$

and $o_{P_r}(1)$ here need some explanation. Poissonization is key in the following discussion.

Note that $M_{N_n,1}, \dots, M_{N_n,n}$ are i.i.d. Poisson variables with mean 1. By the analysis in Theorem 3 of Klaassen and Wellner (1992),

$$P \left(\max_{1 \leq i \leq n} |M_{N_n,i} - M_{ni}| > 2 \right) = O \left(n^{-1/2} \right).$$

Define $I_n^j = \{i : |M_{N_n,i} - M_{ni}| \geq j\}$, then $\#I_n^j = O_p(\sqrt{n})$.

With probability approaching 1,

$$\begin{aligned}
& \sum_{i=1}^n (M_{N_n,i} - M_{ni}) \left(u_1' \frac{x_i x_i'}{n} u_1 - u_1' \frac{2\sigma_{10}}{\sqrt{n}} x_i e_i \right) \mathbf{1} \left(q_i \leq \gamma_0 \wedge \gamma_0 + \frac{v}{n} \right) \\
&= \text{sign}(N_n - n) \sum_{j=1}^2 \left(\frac{\#I_n^j}{n} u_1' \left(\frac{1}{\#I_n^j} \sum_{i \in I_n^j} x_i x_i' \mathbf{1} \left(q_i \leq \gamma_0 \wedge \gamma_0 + \frac{v}{n} \right) \right) u_1 \right. \\
&\quad \left. - \frac{\#I_n^j}{\sqrt{n}} 2\sigma_{10} u_1' \frac{1}{\#I_n^j} \sum_{i \in I_n^j} x_i e_i \mathbf{1} \left(q_i \leq \gamma_0 \wedge \gamma_0 + \frac{v}{n} \right) \right) \\
&= o_{P_r}(1),
\end{aligned}$$

where the last equality is from a multiplier Glivenko-Cantelli theorem; see, e.g., Lemma 3.6.16 of Van der Vaart and Wellner (1996). Now,

$$\sum_{i=1}^n M_{N_n,i} \left(u_1' \frac{x_i x_i'}{n} u_1 - u_1' \frac{2\sigma_{10}}{\sqrt{n}} x_i e_i \right) \mathbf{1} \left(\gamma_0 \wedge \gamma_0 + \frac{v}{n} < q_i \leq \gamma_0 \right) = o_{P_r}(1)$$

from Assumption D4. So

$$\begin{aligned}
& \sum_{i=1}^n M_{ni} \left(u_1' \frac{x_i x_i'}{n} u_1 - u_1' \frac{2\sigma_{10}}{\sqrt{n}} x_i e_i \right) \mathbf{1} \left(q_i \leq \gamma_0 \wedge \gamma_0 + \frac{v}{n} \right) \\
&= \sum_{i=1}^n M_{N_n, i} \left(u_1' \frac{x_i x_i'}{n} u_1 - u_1' \frac{2\sigma_{10}}{\sqrt{n}} x_i e_i \right) \mathbf{1} (q_i \leq \gamma_0) + o_{P_r}(1) \\
&= u_1' E[x_i x_i' \mathbf{1}(q_i \leq \gamma_0)] u_1 - u_1' \left(\frac{2\sigma_{10}}{\sqrt{n}} \sum_{i=1}^n M_{N_n, i} x_i e_i \mathbf{1}(q_i \leq \gamma_0) \right) + o_{P_r}(1).
\end{aligned}$$

Similarly, we could prove the result for $\sum_{i=1}^n M_{ni} \left(u_2' \frac{x_i x_i'}{n} u_2 - u_2' \frac{\sigma_{20}}{\sqrt{n}} x_i e_i \right) \mathbf{1}(q_i > \gamma_0 \vee \gamma_0 + \frac{v}{n})$.

With probability approaching 1,

$$\begin{aligned}
& \sum_{i=1}^n (M_{N_n, i} - M_{ni}) \left[\left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right)' x_i x_i' \left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right) \right. \\
& \quad \left. + 2x_i' \left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right) \sigma_{10} e_i \right] \mathbf{1} \left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0 \right) \\
&= \text{sign}(N_n - n) \sum_{j=1}^2 \frac{1}{\#I_n^j} \sum_{i \in I_n^j} \left[\left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right)' x_i x_i' \left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right) \right. \\
& \quad \left. + 2x_i' e_i \left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right) \sigma_{10} \right] \left(\#I_n^j \cdot \mathbf{1} \left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0 \right) \right) \\
&= o_{P_r}(1),
\end{aligned}$$

since $\#I_n^j \cdot \mathbf{1} \left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0 \right) = o_{P_r}(1)$ uniformly for $i \in I_n^j$ by Assumption D4. So

$$\begin{aligned}
& \sum_{i=1}^n M_{ni} \left[\left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right)' x_i x_i' \left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right) + 2x_i' \left(\beta_{10} - \beta_{20} - \frac{u_2}{\sqrt{n}} \right) \sigma_{10} e_i \right] \mathbf{1} \left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0 \right) \\
& \quad + \sum_{i=1}^n M_{ni} \left[\left(\beta_{10} + \frac{u_1}{\sqrt{n}} - \beta_{20} \right)' x_i x_i' \left(\beta_{10} + \frac{u_1}{\sqrt{n}} - \beta_{20} \right) - 2x_i' \left(\beta_{10} + \frac{u_1}{\sqrt{n}} - \beta_{20} \right) \sigma_{20} e_i \right] \mathbf{1} \left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n} \right) \\
&= \sum_{i=1}^n M_{N_n, i} \left[(\beta_{10} - \beta_{20})' x_i x_i' (\beta_{10} - \beta_{20}) + 2x_i' (\beta_{10} - \beta_{20}) \sigma_{10} e_i \right] \mathbf{1} \left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0 \right) \\
& \quad + \sum_{i=1}^n M_{N_n, i} \left[(\beta_{10} - \beta_{20})' x_i x_i' (\beta_{10} - \beta_{20}) - 2x_i' (\beta_{10} - \beta_{20}) \sigma_{20} e_i \right] \mathbf{1} \left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n} \right) + o_{P_r}(1) \\
&= D_n^*(v) + o_{P_r}(1).
\end{aligned}$$

■

The following lemma appears in Wellner and Zhan (1996) and states relationships among the probability measures P^∞ , P^* and P_r .

Lemma 2 (i) If Δ_n is defined only on the probability $(\mathcal{Z}^\infty, \mathcal{A}^\infty, P^\infty)$ and $\Delta_n = o_{P^\infty}(1)$ ($O_{P^\infty}(1)$), then $\Delta_n = o_{P_r}(1)$ ($O_{P_r}(1)$); (ii) If $\Delta_n = o_{P_r}(1)$ ($O_{P_r}(1)$), then $\Delta_n = o_{P^*}(1)$ ($O_{P^*}(1)$) in P^∞ probability.

Lemma 3 Under Assumptions D1-D5, both $\hat{\theta}$ and $\hat{\theta}^*$ are consistent in P_r probability.

Proof. First, we will prove $\hat{\gamma}$ is consistent. The idea of proof follows from Lemma A.5 of Hansen (2000).

Suppose $\gamma \geq \gamma_0$. Note that $Y = X\beta_{20} + \sigma_{20}\mathbf{e} + X_{\leq\gamma_0}\delta_{\beta 0} + \delta_{\sigma 0}\mathbf{e}_{\leq\gamma_0}$, and X lies in the space spanned by $P_\gamma = \bar{X}_\gamma \left(\bar{X}'_\gamma \bar{X}_\gamma \right)^{-1} \bar{X}'_\gamma$, where $\mathbf{e}_{\leq\gamma}$ is defined by stacking the variables $e_i \mathbf{1}(q_i \leq \gamma)$, $\delta_\beta = \beta_1 - \beta_2$, $\delta_\sigma = \sigma_1 - \sigma_2$, and $\bar{X}_\gamma = [X \ X_{\leq\gamma}]$. So

$$\begin{aligned} \frac{1}{n}Q_n(\gamma) &= \frac{1}{n}Y'(I - P_\gamma)Y \\ &= \frac{1}{n} \left[2\sigma_{20}\delta'_{\beta 0}X'_{\leq\gamma_0}(I - P_\gamma)\mathbf{e} + 2\delta_{\sigma 0}\delta'_{\beta 0}X'_{\leq\gamma_0}(I - P_\gamma)\mathbf{e}_{\leq\gamma_0} + 2\sigma_{20}\delta_{\sigma 0}\mathbf{e}'(I - P_\gamma)\mathbf{e}_{\leq\gamma_0} \right. \\ &\quad \left. + \delta'_{\beta 0}X'_{\leq\gamma_0}(I - P_\gamma)X_{\leq\gamma_0}\delta_{\beta 0} + \sigma_{20}^2\mathbf{e}'(I - P_\gamma)\mathbf{e} + \delta_{\sigma 0}^2\mathbf{e}'_{\leq\gamma_0}(I - P_\gamma)\mathbf{e}_{\leq\gamma_0} \right] \\ &\xrightarrow{p} \delta'_{\beta 0} \left(M(\gamma_0) - M(\gamma_0)M(\gamma)^{-1}M(\gamma_0) \right) \delta_{\beta 0} + E \left[(\sigma_{20}e + \delta_{\sigma 0}e\mathbf{1}(q \leq \gamma_0))^2 \right] \end{aligned}$$

uniformly for $\gamma \in \Gamma$ by a Glivenko-Cantelli theorem, where $M(\gamma) = E[xx'\mathbf{1}(q \leq \gamma)]$. The second term of this limit does not depend on γ , so we concentrate on the first term. We need only prove that $M(\gamma_0) - M(\gamma_0)M(\gamma)^{-1}M(\gamma_0) > M(\gamma_0) - M(\gamma_0)M(\gamma_0)^{-1}M(\gamma_0) = 0$ for any $\gamma > \gamma_0$, which reduces to prove $M(\gamma) > M(\gamma_0)$. Since $M(\gamma_2) > M(\gamma_1)$ if $\gamma_2 > \gamma_1$, we need only prove that for any $\epsilon > 0$, $M(\gamma_0 + \epsilon) - M(\gamma_0) > 0$. But from Assumptions D3 and D4, $M(\gamma_0 + \epsilon) - M(\gamma_0) = E[xx'\mathbf{1}(\gamma_0 < q \leq \gamma_0 + \epsilon)] \geq \left(\inf_{\gamma_0 < \gamma \leq \gamma_0 + \epsilon} E[xx'|q = \gamma] \right) (F(\gamma_0 + \epsilon) - F(\gamma_0)) > 0$.

Symmetrically, we can show that $M(\gamma_0 - \epsilon) - M(\gamma_0) > 0$ for any $\epsilon > 0$. Theorem 2.1 of Newey and McFadden (1994) can be applied to show $\hat{\gamma}$ is consistent. With the consistency of $\hat{\gamma}$ in hand, it is easy to show $\hat{\beta}_1(\hat{\gamma})$ and $\hat{\beta}_2(\hat{\gamma})$ are consistent by a dominance argument. Similar arguments show $\hat{\theta}^*$ is consistent in P_r probability, but now a multiplier Glivenko-Cantelli theorem is used for the uniform convergence. ■

Lemma 4 *Under Assumptions D1-D5, $\hat{\gamma} = \gamma_0 + O_{P^\infty}(\frac{1}{n})$, $\hat{\gamma}^* = \gamma_0 + O_{P_r}(\frac{1}{n})$, $\hat{\beta} = \beta_0 + O_{P^\infty}(\frac{1}{\sqrt{n}})$, and $\hat{\beta}^* = \beta_0 + O_{P_r}(\frac{1}{\sqrt{n}})$.*

Proof. We will first prove $\hat{\gamma} = \gamma_0 + O_{P^\infty}(\frac{1}{n})$. Corollary 3.2.6 of Van der Vaart and Wellner (1996) is used in this proof.

First, $M(\theta) - M(\theta_0) \geq Cd^2(\theta, \theta_0)$ with $d(\theta, \theta_0) = \|\underline{\theta} - \underline{\theta}_0\| + \sqrt{|\gamma - \gamma_0|}$ for θ in a neighborhood of θ_0 .

$$\begin{aligned} &M(\theta) - M(\theta_0) \\ &= E[T(w|\theta_1, \theta_{10})\mathbf{1}(q \leq \gamma \wedge \gamma_0)] + E[T(w|\theta_2, \theta_{20})\mathbf{1}(q > \gamma \vee \gamma_0)] \\ &\quad + E[\bar{z}_1(w|\theta_2, \theta_{10})\mathbf{1}(\gamma \wedge \gamma_0 < q \leq \gamma_0)] + E[\bar{z}_2(w|\theta_1, \theta_{20})\mathbf{1}(\gamma_0 < q \leq \gamma \vee \gamma_0)] \\ &= (\beta_{10} - \beta_1)' E[xx'\mathbf{1}(q \leq \gamma \wedge \gamma_0)](\beta_{10} - \beta_1) + (\beta_{20} - \beta_2)' E[xx'\mathbf{1}(q > \gamma \vee \gamma_0)](\beta_{20} - \beta_2) \\ &\quad + (\beta_{10} - \beta_2)' E[xx'\mathbf{1}(\gamma \wedge \gamma_0 < q \leq \gamma_0)](\beta_{10} - \beta_2) + (\beta_{20} - \beta_1)' E[xx'\mathbf{1}(\gamma_0 < q \leq \gamma \vee \gamma_0)](\beta_{20} - \beta_1) \\ &\geq C \left(\|\beta_{10} - \beta_1\|^2 + \|\beta_{20} - \beta_2\|^2 + |\gamma - \gamma_0| \right), \end{aligned}$$

where the last inequality is from Assumptions D1-D4.

Second, $E \left[\sup_{d(\theta, \theta_0) < \delta} |\mathbb{G}_n(m(w|\theta) - m(w|\theta_0))| \right] \leq C\delta$. Since $\{T(w|\theta_1, \theta_{10}) : d(\theta, \theta_0) < \delta\}$ is a finite-dimensional vector space of functions and $\{\mathbf{1}(q \leq \gamma \wedge \gamma_0) : d(\theta, \theta_0) < \delta\}$ is a VC subgraph class of functions by Lemma 2.4 of Pakes and Pollard (1989), $\{A(w|\theta) : d(\theta, \theta_0) < \delta\}$ is VC subgraph by Lemma 2.14 (ii) of Pakes and Pollard (1989). Similarly, $\{B(w|\theta) : d(\theta, \theta_0) < \delta\}$, $\{C(w|\theta) : d(\theta, \theta_0) < \delta\}$, and

$\{D(w|\theta) : d(\theta, \theta_0) < \delta\}$ are VC subgraph. From Theorem 2.14.2 of Van der Vaart and Wellner (1996),

$$E \left[\sup_{d(\theta, \theta_0) < \delta} |\mathbb{G}_n(m(w|\theta) - m(w|\theta_0))| \right] \leq C\sqrt{PF^2},$$

where F is the envelope of $\{m(w|\theta) - m(w|\theta_0) : d(\theta, \theta_0) < \delta\}$. But from the function form of $m(w|\theta) - m(w|\theta_0)$, $\sqrt{PF^2} \leq C\delta$ by Assumption D5. So $\phi(\delta) = \delta$ in Corollary 3.2.6 of Van der Vaart and Wellner (1996) and $\frac{\delta}{\delta^\alpha}$ is decreasing for all $1 < \alpha < 2$. Since $r_n^2 \phi\left(\frac{1}{r_n}\right) = r_n$, $\sqrt{nd}(\hat{\theta} - \theta_0) = O_{P^\infty}(1)$. By the definition of d , the result follows.

Now, we prove $\hat{\gamma}^* = \gamma_0 + O_{P_r}\left(\frac{1}{n}\right)$. Since $\hat{\theta}^*$ is consistent, we need only concentrate on a neighborhood \mathcal{N} of θ_0 . By the definition of $\hat{\theta}^*$,

$$P_n^* m(w|\hat{\theta}^*) \leq P_n^* m(w|\theta_0).$$

Thus

$$\begin{aligned} 0 &\geq P_n^* \left(m(w|\hat{\theta}^*) - m(w|\theta_0) \right) \\ &= (P_n^* - P_n) \left(m(w|\hat{\theta}^*) - m(w|\theta_0) \right) + (P_n - P) \left(m(w|\hat{\theta}^*) - m(w|\theta_0) \right) + P \left(m(w|\hat{\theta}^*) - m(w|\theta_0) \right). \end{aligned}$$

From the proof above, $P(m(w|\theta) - m(w|\theta_0)) \geq Cd^2(\theta, \theta_0)$, $(P_n - P)(m(w|\theta) - m(w|\theta_0)) = O_{P^\infty}(n^{-1/2}d(\theta, \theta_0))$ for $\theta \in \mathcal{N}$. If we could prove $(P_n^* - P_n)(m(w|\theta) - m(w|\theta_0)) = O_{P_r}(n^{-1/2}d(\theta, \theta_0))$ for $\theta \in \mathcal{N}$, the result follows. Conditioning on $\{w_i\}_{i=1}^n$, a similar argument as above shows that

$$\begin{aligned} &P_n \sup_{d(\theta, \theta_0) < \delta} |\sqrt{n}(P_n^* - P_n)(m(w|\theta) - m(w|\theta_0))| \\ &\leq C\sqrt{P_n F^2} \leq C\delta \end{aligned}$$

for n large enough by the strong law of large numbers. ■

The following Lemma is an extension of Theorem 6.1 of Huang and Wellner (1995) which is an extension of the argmax continuous mapping theorem. As in Kim and Pollard (1990), define $\mathbf{B}_{loc}(\mathbb{R}^d)$ as the space of all locally bounded real functions on \mathbb{R}^d , endowed with the uniform metric on compacta. The space $\mathbf{C}_{min}(\mathbb{R}^d)$ is defined as the subset of continuous functions $x(\cdot) \in \mathbf{B}_{loc}(\mathbb{R}^d)$ for which (i) $x(t) \rightarrow \infty$ as $|t| \rightarrow \infty$ and (ii) $x(t)$ achieves its minimum at a unique point in \mathbb{R}^d . $\mathbf{B}'_{loc}(\mathbb{R})$ is the same as $\mathbf{B}_{loc}(\mathbb{R})$ except being endowed with the Skorohod metric on compacta. The space $\mathbf{D}_{min}(\mathbb{R})$ is defined as the subset of cadlag functions $x(\cdot) \in \mathbf{B}'_{loc}(\mathbb{R})$ for which the conditions (i) and (ii) in the definition of $\mathbf{C}_{min}(\mathbb{R})$ are satisfied.

Lemma 5 Suppose (U_n, U_n^*) are random maps onto $\mathbf{B}_{loc}(\mathbb{R}^d) \times \mathbf{B}_{loc}(\mathbb{R}^d)$, and (D_n, D_n^*) are random maps onto $\mathbf{B}'_{loc}(\mathbb{R}) \times \mathbf{B}'_{loc}(\mathbb{R})$. Let (s_n, t_n) and (s_n^*, t_n^*) be random maps onto $\mathbb{R}^{d+1} \times \mathbb{R}^{d+1}$ such that:

(i) $(U_n + D_n, U_n^* + D_n^*) \rightsquigarrow (U + D, U^* + D^*)$, where

$$P((U, U^*) \in \mathbf{C}_{min}(\mathbb{R}^d) \times \mathbf{C}_{min}(\mathbb{R}^d)) = 1$$

and

$$P((D, D^*) \in \mathbf{D}_{min}(\mathbb{R}) \times \mathbf{D}_{min}(\mathbb{R})) = 1;$$

(ii) (s_n, t_n) and (s_n^*, t_n^*) are uniquely defined and both $O_p(1)$;

(iii) $U_n(s_n) + D_n(t_n) \leq \inf_u U_n(u) + \inf_v D_n(v) + \alpha_n$ and $U_n^*(s_n^*) + D_n^*(t_n^*) \leq \inf_u U_n^*(u) + \inf_v D_n^*(v) + \alpha_n^*$, where α_n and α_n^* are both $o_p(1)$.

Then $((s_n, t_n), (s_n^*, t_n^*)) \xrightarrow{d} \left(\arg \min_{u,v} U(u) + D(v), \arg \min_{u,v} U^*(u) + D^*(v) \right)$.

Proof. The proof follows from Theorem 6.1 of Huang and Wellner (1995) by using Dudley's representation theorem. The only difference is that the metric for $\mathbf{D}_{\min}(\mathbb{R})$ is substituted by the Skorohod metric on compacta and the product metric is used on $\mathbf{C}_{\min}(\mathbb{R}^d) \times \mathbf{D}_{\min}(\mathbb{R})$. ■