Econometric Theory, **30**, 2014, 676–714.
doi:10.1017/S0266466614000012

THE BOOTSTRAP IN THRESHOLD REGRESSION

PING YU
University of Auckland

This paper develops a general procedure to check the bootstrap validity in M-estimation. We apply the procedure in discontinuous threshold regression to show the inconsistency of the nonparametric bootstrap for inference on the threshold point. Especially, the conditional weak limit of the nonparametric bootstrap is shown not to exist. By comparing with two other boundaries in the literature, we show the fact that the threshold point is a boundary of the covariate that makes its bootstrap inference so different. The remedies to the bootstrap failure in the literature are summarized, and the nonparametric posterior interval is suggested by some simulation studies.

1. INTRODUCTION

Since the pioneering work by Tong (1978, 1983), threshold models get much popularity in current applied statistical and econometric practice. An encyclopedic survey on this kind of models is available in Tong (1990) and a selective review of the history of threshold models is given by Tong (2011); see also Lee and Seo (2008) and Hansen (2011) for a summary of applications especially in economics. However, it is also well understood that asymptotic distribution of threshold estimators is typically nonstandard and hard to simulate. A common antidote to inference when an estimator's distribution is nonstandard is to turn to the bootstrap. Validity of bootstrap inference for threshold regression is a long-standing open question in the literature despite much anticipation. Gonzalo and Wolf (2005) claim that "it is not known whether a bootstrap approach would work (for threshold regression)". This motivates them to use the subsampling in construction of confidence intervals (CIs) for the threshold point. Footnote 3 in Seo and Linton (2007) indicates that the bootstrap may be inconsistent in their simulations, but no theoretical analysis is conducted.

It is well known that the nonparametric bootstrap introduced by Efron (1979) can fail in many examples. Most of these examples are contributed by statisticians; see Andrews (2000, p. 400) for a list of them.[1] To our knowledge, there are also

I would like to thank K.S. Chan, Bruce Hansen, Peter Phillips, Ehsan Soofi, and seminar participants at ESWC 2010, Southern Methodist University, University of Iowa, University of Otago, UW-Milwaukee and UW-Madison for helpful comments. I also want to thank Peter Bickel for clarifying a key point of his paper in a personal communication and thank Botev Zdravko for providing his codes and giving detailed helps in coding. Special thanks are given to Jack Porter for a careful modification of the whole paper and to the co-editor and three referees for very insightful comments. Address correspondence and reprint requests to Ping Yu, Department of Economics, 12 Grafton Road, University of Auckland, Auckland, New Zealand. e-mail: p.yu@auckland.ac.nz.

 © Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

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).[2] This paper will contribute 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

$$y = \begin{cases} x'\beta_1 + \sigma_1\varepsilon, & q \le \gamma; \\ x'\beta_2 + \sigma_2\varepsilon, & q > \gamma; \end{cases} \tag{1}$$

where $E[\varepsilon|x, q] = 0$, q is the threshold variable used to split the sample and has a density $f_q(\cdot)$, $x \in \mathbb{R}^k$, $\beta \equiv (\beta_1', \beta_2')' \in \mathbb{R}^{2k}$, and $\sigma \equiv (\sigma_1, \sigma_2)'$ are threshold parameters on the mean and variance in the two regimes. We set $E[\varepsilon^2] = 1$ as a normalization of the error variance and allow for conditional heteroskedasticity. All the other variables have the same definitions as in the linear regression framework. Note that q can be an element of x, but to emphasize the special status of q in (1), we write the covariates explicitly as $(x', q)'$. Define $\theta = (\beta', \gamma)'$, which is the parameter of interest in most applications. Usually, the nonregular parameter γ is of main interest, since inference on the regular parameter β is standard. In this paper, we take the framework of discontinuous threshold regression where $(\beta_1', \sigma_1)' - (\beta_2', \sigma_2)'$ is a fixed constant. In this case, we show that the nonparametric bootstrap (and various wild bootstraps) for γ fails, while the smoothed bootstrap (when ε is independent of $(x', q)'$ which has a compact support) and the parametric bootstrap are still valid.

There are some related results in the structural change literature. For example, under the small-threshold-effect asymptotic structure of Hansen (2000), Antoch et al. (1995) show that the nonparametric 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. 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 nonparametric bootstrap has the correct convergence rate n in the structural change model but does not prove its validity before applying it to the famous Nile data. In this paper, we confirm that the nonparametric bootstrap has the right convergence rate, and also show that it is not valid for inference on the threshold point. In a class of parametric threshold models arising in survival analysis, Kosorok and Song (2007) suggest to use a form of specialized parametric bootstrap to conduct inference on γ. Their procedure tries to simulate the cadlag process in the asymptotic distribution of the threshold estimator, so is very burdensome and imprecise as argued in Section 4.1 of Yu (2008).[3] In a simple nonparametric threshold model where the only covariate is q and ε is independent of q, Gijbels et al. (2004) use a smoothed bootstrap procedure to construct CIs for γ, but their proof of its validity is not complete.

In an independent work by Seijo and Sen (2011), the authors consider similar problems as in this paper. Their discussion is restricted to a simplified threshold

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

model where $x = 1$, $\sigma_1 = \sigma_2$, and ε is independent of q. This paper goes beyond Seijo and Sen (2011) in two aspects. First, we emphasize that the threshold point is a "middle" boundary of q in discussing the bootstrap consistency. Inspired by the literature on the bootstrap inconsistency in boundary estimation, we give a *direct* proof that the conditional distribution of the nonparametric bootstrap estimator does not have a weak limit in probability, while Seijo and Sen (2011) only provide two *indirect* proofs of the nonparametric bootstrap failure. We also compare the threshold point with two other boundaries and find that it is the very fact that the threshold point is the boundary of a *covariate* that makes its bootstrap inference so different. Due to the same reason as in the nonparametric bootstrap, various wild bootstrap procedures are not valid either. Second, we prove the consistency of the smoothed bootstrap in a more practical setup than that in Seijo and Sen (2011). Also, we argue that the smoothed bootstrap is hard to apply in practice; even when it can be applied, our simulation studies show that it is outperformed by an alternative procedure proposed in Yu (2008). In summary, the results in Seijo and Sen (2011) and those given in this paper are more complements than substitutes.

The rest of this paper is organized as follows. In Section 2, we define our estimator, review the bootstrap schemes of interest, and provide a general procedure to check the bootstrap validity of M-estimators. In Section 3, we check the invalidity of the nonparametric bootstrap and compare the threshold point with two other boundaries. Section 4 summarizes the remedies of bootstrap failure in the literature and compares the smoothed bootstrap and the nonparametric posterior interval (NPI) in Yu (2008). Finally, Section 5 concludes. All proofs and lemmas are left to Appendix A and B, respectively. To save space, we put part of the paper in the supplementary materials.

Some notations are collected here for future reference. Any symbol with a superscript $*$ means an object under the bootstrap probability measure instead of under the outer measure as used in some other literature. For example, $E^*[\cdot]$ means the conditional expectation with respect to the bootstrap probability measure conditional on the original data; $\overset{*}{\rightsquigarrow}$ means weak convergence in probability. ω represents an original sample path as in Andrews (2000). $Qf = \int f dQ$ for any probability measure Q. $\|\cdot\|$ is the Euclidean metric, and $\|\cdot\|_\infty$ is the sup-norm for a function defined on $\mathbb{R}^d$, $1 \le d < \infty$. For any random vector x, $f_x(x)$ is the true density function of x. ℓ is always used for indicating the two regimes in (1), so is not written out explicitly as "$\ell = 1, 2$" throughout the paper.

2. THE SETUP AND OVERVIEW OF PROCEDURES

2.1. The Least Squares Estimator

With an i.i.d. (independent and identically distributed) sample $\{w_i\}_{i=1}^n$, where $w_i = \left(y_i, x_i', q_i\right)'$, the LSE of γ is usually defined by a profiled procedure:

$$\widehat{\gamma} = \arg\min_{\gamma} M_n(\gamma),$$

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

where

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

with

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

Usually, there is an interval of γ minimizing $M_n(\gamma)$. Most literature uses the left-endpoint LSE (LLSE), while Yu (2008, 2012) shows that the middle-point LSE (MLSE) is more efficient in most cases. Throughout this paper, we take the LLSE as our estimator of γ and the discussions on the MLSE are put in the supplementary materials.

The estimation of β is invariant to the two estimators of γ. To express the β estimator in matrix notation, define the $n \times 1$ vectors Y and ε by stacking the variables y_i and ε_i, and the $n \times k$ matrices X, $X_{\le\gamma}$, and $X_{>\gamma}$ by stacking the vectors x_i', $x_i'\mathbf{1}(q_i \le \gamma)$, and $x_i'\mathbf{1}(q_i > \gamma)$. Let

$$\begin{pmatrix} \widehat{\beta}_1(\gamma) \\ \widehat{\beta}_2(\gamma) \end{pmatrix} = \arg\min_{\beta_1,\beta_2} \frac{1}{n}\sum_{i=1}^{n} m(w_i|\theta) = \begin{pmatrix} \left(X_{\le\gamma}'X_{\le\gamma}\right)^{-1} X_{\le\gamma}'Y \\ \left(X_{>\gamma}'X_{>\gamma}\right)^{-1} X_{>\gamma}'Y \end{pmatrix},$$

then the LSE of β is defined as $\widehat{\beta} = \left(\widehat{\beta}_1'(\widehat{\gamma}), \widehat{\beta}_2'(\widehat{\gamma})\right)' \equiv \left(\widehat{\beta}_1', \widehat{\beta}_2'\right)'$. Given $\widehat{\theta} \equiv \left(\widehat{\beta}', \widehat{\gamma}\right)'$, we can estimate σ by

$$\widehat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i'\widehat{\beta}_1\right)^2 \mathbf{1}\left(q_i \le \widehat{\gamma}\right) \text{ and } \widehat{\sigma}_2^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i'\widehat{\beta}_2\right)^2 \mathbf{1}\left(q_i > \widehat{\gamma}\right).$$

The residuals

$$\widehat{\varepsilon}_i = \frac{y_i - x_i'\widehat{\beta}_1}{\widehat{\sigma}_1}\mathbf{1}\left(q_i \le \widehat{\gamma}\right) + \frac{y_i - x_i'\widehat{\beta}_2}{\widehat{\sigma}_2}\mathbf{1}\left(q_i > \widehat{\gamma}\right), \qquad i = 1, \ldots, n.$$

Under some regularity conditions specified in Section 3 below, $\widehat{\beta} - \beta_0$ and $\widehat{\sigma} - \sigma_0$ are both $O_P\left(n^{-1/2}\right)$, and $\widehat{\gamma} - \gamma_0$ is $O_P\left(n^{-1}\right)$.

2.2. The Bootstrap Schemes

Suppose we are interested in the distribution of some random variables $R_n \equiv R_n(\mathbf{w}_n, P)$, where $\mathbf{w}_n \equiv \{w_i\}_{i=1}^n$ is a random sample of size n from the probability measure P. Suppose the distribution of $R_n(\mathbf{w}_n, P)$ is continuous with respect to P, then if we can find some resampling measure P_n which converges to P under some metric, we expect the distribution of $R_n^* \equiv R_n\left(\mathbf{w}_n^*, P_n\right)$

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

is close to that of $R_n(\mathbf{w}_n, P)$, where $\mathbf{w}_n^* \equiv \{w_i^*\}_{i=1}^n$ is a random sample of size n from the probability measure P_n. This is the main idea of the bootstrap. In this paper, $R_n(\mathbf{w}_n, P) = \left(\sqrt{n}(\widehat{\beta} - \beta_0)', n(\widehat{\gamma} - \gamma_0)\right)'$ and $R_n(\mathbf{w}_n^*, P_n) = \left(\sqrt{n}(\widehat{\beta}^* - \beta_0(P_n))', n(\widehat{\gamma}^* - \gamma_0(P_n))\right)'$, where $(\beta_0(P_n)', \gamma_0(P_n))' = \arg\min_{\beta,\gamma} P_n m(w|\theta)$, $(\beta_0', \gamma_0)' = (\beta_0(P)', \gamma_0(P))'$ is the true value of θ under P, $(\widehat{\beta}', \widehat{\gamma})' = (\beta_0(F_n)', \gamma_0(F_n))'$ with F_n being the empirical measure of $\mathbf{w}_n = \{(y_i, x_i', q_i)'\}_{i=1}^n$ is the LSE defined in the last subsection, and $(\widehat{\beta}^{*\prime}, \widehat{\gamma}^*)'$ is the LSE using the bootstrap samples $\mathbf{w}_n^* = \{(y_i^*, x_i^{*\prime}, q_i^*)'\}_{i=1}^n$, that is,

$$\left(\widehat{\beta}^{*\prime}, \widehat{\gamma}^*\right)' = \arg\min_{\beta,\gamma} P_n^* m(w|\theta) \equiv \left(\beta_0(P_n^*)', \gamma_0(P_n^*)\right)' \tag{3}$$

with P_n^* being the empirical measure of $\mathbf{w}_n^*$. Different bootstrap schemes are distinguished by different resampling probability measures P_n from which the bootstrap samples $\mathbf{w}_n^*$ are drawn.

There are two sources of randomness in the bootstrap: one is the original data, and the other is the resampling. For the original sample, take w_i as from the probability space $(\Omega, \mathcal{A}, P)$. For the resampling, suppose the randomness is defined on a probability space $(\mathcal{T}, \mathcal{B}, P^*)$ which is independent of $(\Omega, \mathcal{A}, P)$. So the joint randomness is defined on the product probability space $(\Omega \times \mathcal{T}, \mathcal{A} \times \mathcal{B}, P_r) \equiv (\Omega, \mathcal{A}, P) \times (\mathcal{T}, \mathcal{B}, P^*)$, where $P_r \equiv P \times P^*$ denotes the whole randomness in the bootstrap. Denote the distribution of $R_n(\mathbf{w}_n, P)$ as $H_n(\cdot)$ (or just H_n), the conditional distribution of $R_n(\mathbf{w}_n^*, P_n)$ given $\mathbf{w}_n$ as $H_n^*(\cdot; \omega)$ (or just $H_n^*(\omega)$), and the distribution of $R_n(\mathbf{w}_n^*, P_n)$ under P_r as $\overline{H}_n^*(\cdot)$ (or just $\overline{H}_n^*$). Then the bootstrap method can be described in the following three steps.

Step 1: Construct an estimator P_n of P from the data $\mathbf{w}_n$.
Step 2: Generate random samples $\mathbf{w}_n^*$ from P_n, and construct R_n^*.
Step 3: Approximate $H_n(x)$ by $H_n^*(x; \omega) \equiv P^*(R_n^* \le x)$.

For notational simplicity, we suppress the dependence of H_n^* on ω unless necessary.

The consistency of the bootstrap is defined as follows. First note that a sequence of random vector X_n converges weakly to a random vector X if

$$\sup_{h \in BL_1} |E[h(X_n)] - E[h(X)]| \to 0,$$

where $BL_1 \equiv \{h : |h| \le 1, |h(x) - h(y)| \le \|x - y\| \text{ for all } x, y\}$ is the *bounded Lipschitz metric*. In the bootstrap case, $R_n(\mathbf{w}_n^*, P_n)$ is (weakly) *consistent* to $R_n(\mathbf{w}_n, P)$ if for any $\epsilon > 0$,

$$P\left(\sup_{h \in BL_1} \left|E^*\left[h(R_n(\mathbf{w}_n^*, P_n))\right] - E[h(R_n(\mathbf{w}_n, P))]\right| > \epsilon\right) \to 0. \tag{4}$$

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

If $R_n(\mathbf{w}_n, P)$ has a weak limit R which has the distribution H, then we denote (4) as $R_n^* \overset{*}{\rightsquigarrow} R$ or $H_n^* \overset{*}{\rightsquigarrow} H$. When the asymptotic distribution of R_n is continuous, consistency of the bootstrap can take a simpler form. For example, since the asymptotic distribution of $\widehat{\gamma}$ is continuous, consistency of the bootstrap is equivalent to

$$\left| P^* \left(n \left(\widehat{\gamma}^* - \widehat{\gamma} \right) \le \upsilon \right) - P \left(n \left(\widehat{\gamma} - \gamma_0 \right) \le \upsilon \right) \right| \xrightarrow{P} 0,$$

for any fixed value of $\upsilon \in \mathbb{R}$ by Pólya's theorem.

In this paper, we concentrate on two bootstrap schemes in threshold regression. Other schemes such as the parametric bootstrap and various wild bootstraps are described in the supplementary materials.

Nonparametric or Standard Bootstrap: $P_n = F_n$. This scheme is shown to be inconsistent for γ.

As will be shown in Section 3, the main reason for the nonparametric bootstrap failure is the discreteness generated from the bootstrap sampling on the local data around γ_0. To break such discreteness, we can smooth the data in the neighborhood of $q = \gamma_0$. Such a procedure is termed as the smoothed bootstrap by Efron (1979). In regular cases where the nonparametric bootstrap is valid, the smoothed bootstrap may have some refinements, e.g., in quantile estimation as shown in Silverman and Young (1987) and Hall et al. (1989).

Smoothed Bootstrap: P_n is the probability measure constructed as follows. First, estimate the joint density of $\mathsf{w} \equiv \left(\varepsilon, x', q\right)'$ from $\left\{\widehat{\mathsf{w}}_i\right\}_{i=1}^n \equiv \left\{\left(\widehat{\varepsilon}_i, x_i', q_i\right)'\right\}_{i=1}^n$ by some nonparametric techniques such as kernel smoothing; denote the estimate as $\widehat{f}(\mathsf{w})$. Second, the distribution of w_i^* is derived from

$$y_i^* = \begin{cases} x_i^{*\prime}\widehat{\beta}_1 + \widehat{\sigma}_1 \varepsilon_i^*, & q_i^* \le \widehat{\gamma}; \\ x_i^{*\prime}\widehat{\beta}_2 + \widehat{\sigma}_2 \varepsilon_i^*, & q_i^* > \widehat{\gamma}; \end{cases} \tag{5}$$

where $\mathsf{w}_i^* \equiv \left(\varepsilon_i^*, x_i^{*\prime}, q_i^*\right)'$ has the density $\widehat{f}(\mathsf{w})$. Here, we must use the joint distribution of w to maintain the correlation structure among ε, x, and q. When ε is independent of x and q, we can simulate $(x_i^{*\prime}, q_i^*)'$ from $\widehat{f}(x, q)$ and ε_i^* from P_n^{ε}, where P_n^{ε} is the empirical distribution of $\left\{\widehat{\varepsilon}_i\right\}_{i=1}^n$, then generate y_i^* by (5).

Other popular resampling schemes used in threshold regression include the m out of n (moon) bootstrap and the subsampling. The validity of the subsampling is proved in Gonzalo and Wolf (2005) where they even offer consistent inference on γ in the more general case when the (dis)continuity of the model is unknown; Seijo and Sen (2011) prove the validity of the moon bootstrap in a

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

simple threshold model, and their proof can be extended to the general model by some technical complication. We do not consider these two methods in this paper for two reasons. First, while their validity only requires $m/n \to 0$ and the continuity of the asymptotic distribution, selection of m and the associated bias problem in threshold models (see Section 4.2 of Yu, 2008) are quite tricky.[4] Second, they perform well in coverage in the simulations of Gonzalo and Wolf (2005), but do not in Yu (2008) and Seijo and Sen (2011), so their performance may depend on specific scenarios. Also, since m must be large enough ($m \to \infty$) to get a good approximation to the asymptotic distribution, n must be reasonably large to make the performance acceptable given that $m/n \to 0$.

2.3. Checking the Bootstrap Validity in M-Estimation

We now discuss how to check the validity of a bootstrap procedure in a general M-estimation based on the recent econometric and statistical literature. Let $\mathcal{M}$ be a collection of probability measures on $\mathbb{R}^d$ that includes all probability measures we need. Suppose under P_r, $\overline{H}_n^* \equiv \int_\Omega H_n^*(\omega) P(d\omega)$ converges weakly to a random measure $\overline{H}^* \equiv \int_\Omega H^*(\omega) P(d\omega)$, where $H^*(\omega) \in \mathcal{M}$ is the weak limit of $H_n^*(\omega)$ if exists.[5] Strong bootstrap consistency requires that $H^*(\omega) = H$ a.s. P, while weak consistency only requires a weaker version of $H^*(\omega) = H$ as defined in (4). So, roughly speaking, bootstrap consistency requires $H^*(\omega) = H$ in some sense, which implies that $\overline{H}^*$ is the point mass H in $\mathcal{M}$.

From Theorem 2.2 in Kosorok (2008), if the bootstrap is valid, $\overline{H}^*$ is an independent copy of H; that is, $\overline{H}^* \overset{d}{=} H$ and $\overline{H}^*$ and H are independent. Accordingly, we can check the validity of a bootstrap procedure in the following three steps.

Step 1: check whether $\overline{H}^*$ exists. If not, then the bootstrap cannot be valid.

Step 2: check whether $\overline{H}^* \overset{d}{=} H$. Usually, we need only check some of its implications. For example, suppose $R_n = a_n\left(\widehat{\theta} - \theta_0\right)$ and $R_n^* = a_n\left(\widehat{\theta}^* - \widehat{\theta}\right)$, where a_n is the convergence rate, $\widehat{\theta}^*$ is the bootstrap estimator, and $\widehat{\theta}$ is the original estimator. Under P_r, $a_n\left(\widehat{\theta}^* - \theta_0\right) = a_n\left(\widehat{\theta}^* - \widehat{\theta}\right) + a_n\left(\widehat{\theta} - \theta_0\right) \equiv \overline{H}_n^* + H_n \rightsquigarrow \overline{H}^* + H$, where $\rightsquigarrow$ signifies the weak convergence under P_r. If $\overline{H}^*$ and H are independent, $\overline{H}^* + H$ has double variance of H when $\overline{H}^* \overset{d}{=} H$. The bootstrap is obviously invalid if this condition fails. For most examples of bootstrap invalidity, this step is enough; see, e.g., the nonregular M-estimators with convex objective functions in Bose and Chatterjee (2001), the maximum score estimator in Abrevaya and Huang (2005), the Grenander estimator in Kosorok (2008) and Sen et al. (2010), and the LSE of the threshold point in threshold regression in Seijo and Sen (2011) and this paper.[6]

Step 3: check whether $\overline{H}^*$ and H are independent. For most bootstrap procedures, H_n^* is invariant under permutations of the original data; from the Hewitt–Savage zero–one law (see page 496 of Billingsley, 1995), $H_n^*(\omega)$ will be *invariant* of ω as $n \to \infty$ if it indeed converges. So if $\overline{H}^*$ and H are dependent, $H_n^*(\omega)$ cannot even have a weak limit. This procedure is first used in Sen et al. (2010)

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

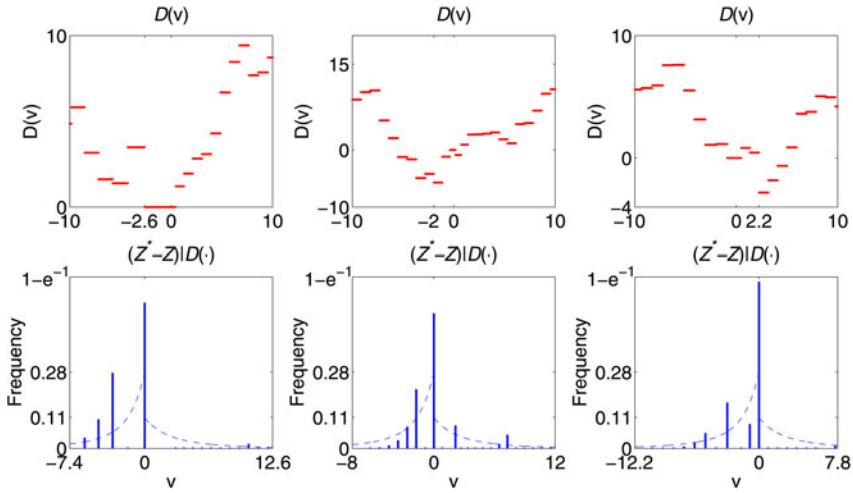


FIGURE 1. Dependence of the component measure $\left(Z^* - Z\right)|D(\cdot)$ on $D(\cdot)$.

to check that the bootstrap Grenander estimator has no weak limit in probability. Usually, we can draw the scatter plot of $\overline{H}^*$ vs. H, calculate their linear correlation, and test whether this correlation is zero; see, e.g., Figure 1 and calculations in Sen et al. (2010). Further, since $\overline{H}^*$ can often be expressed as a mixture of *component measures* $H^*(\omega)$[7], we can check whether $H^*(\omega)$ is invariant of ω to shed more light on the causes for this correlation. For this, some graphical checking may be helpful; see Figure 1 and 4 in Section 3.

The intuitive procedure in Step 3 cannot substitute rigorous mathematical proof. We divide such proof into three levels from the easier to the more advanced. For this purpose, suppose $R_n = a_n\left(\widehat{\theta} - \theta_0\right)$ and $R_n = a_n\left(\widehat{\theta}^* - \widehat{\theta}\right)$ as in step 2, and

$$a_n\left(\widehat{\theta} - \theta_0\right) = \arg\max_u G_n(u) + o_P(1), \qquad a_n\left(\widehat{\theta}^* - \theta_0\right) = \arg\max_u G_n^*(u) + o_{P_r}(1).$$

In level one, the localized objective function $G_n^*(\cdot)$ does not have the weak limit $G^*(\cdot;\omega)$, and $G_n(\cdot)$ does not have the weak limit $G(\cdot;\omega)$ conditional on ω. If not, then by the argmax continuous mapping theorem,

$$a_n\left(\widehat{\theta}^* - \widehat{\theta}\right) \overset{*}{\rightsquigarrow} \arg\max_u G^*(u;\omega) - \arg\max_u G(u;\omega) \equiv H^*(\omega).$$

Now, the component measure analysis will show that $H^*(\omega)$ depends on ω, which contradicts the Hewitt-Savage zero-one law. In level two, $G_n^*(\cdot)$ and $G_n(\cdot)$ do not have *any* weak limit conditional on ω. The analysis in this level can be very difficult; see, e.g., Sen et al. (2010) and Seijo and Sen (2011). In level three, $a_n\left(\widehat{\theta}^* - \widehat{\theta}\right)$ does not have any weak limit conditional on ω. Note that level two

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

does not imply level three. In this paper, we give nonexistence proofs in level three for two examples: the LSE of the threshold point in threshold regression in Section 3.1 and the boundary estimator of Andrews (2000) in Section 3.3.

To further check how $\overline{H}^*$ and H are correlated in Step 3, we first consider a benchmark (regular) model to see why they are not correlated. For regular parameters such as β in (1), the weak limit of the associated localized objective function in the (nonparametric) bootstrap usually takes the form

$$-\frac{1}{2}u'Vu+u'W+u'W^* \tag{6}$$

under P_r, where u is the local parameter, W and W^* are independent of each other and have the same normal distribution, and V is positive definite. The asymptotic distribution H is determined by the maximizer of $-\frac{1}{2}u'Vu+u'W$, i.e., $V^{-1}W$. $\overline{H}^*$ is determined by the difference between the maximizers of $-\frac{1}{2}u'Vu+u'W+u'W^*$ and $-\frac{1}{2}u'Vu+u'W$, i.e., $V^{-1}(W+W^*)-V^{-1}W=V^{-1}W^*$, which has the same distribution as and is independent of $V^{-1}W$. To achieve this desired property, (6) should have three key features. First, the centering process $-\frac{1}{2}u'Vu$ takes a quadratic form. Second, the randomness in the original data ($u'W$) and that in the bootstrap ($u'W^*$) are separable. Third, $u'W$ and $u'W^*$ take the linear form of u. Violation of any of the three features can invalidate the bootstrap.[8] For example, in the bootstrap of nonregular M-estimators with convex objective functions, the first feature is violated while the second and third hold; in bootstrapping the maximum score estimator, the third feature fails while the first and the second hold;[9] in bootstrapping the LSE of the threshold point in threshold regression, none of the three features holds.

If we pass all the three steps, it is more likely that the bootstrap is valid. Arcones and Giné (1992) show the bootstrap validity for regular *parametric* M-estimators. Recently, Cheng and Huang (2010) provide a general method to prove the bootstrap validity for regular *semiparametric* M-estimators. In both cases, the bootstrap validity essentially relies on representing the localized objective function as (6) such that the M-estimator is actually a Z-estimator; see Wellner and Zhan (1996) for further discussions. When usual bootstraps, such as the nonparametric bootstrap and the multiplier bootstrap, fail, common solutions to recover the bootstrap validity include the smoothed bootstrap and the moon bootstrap (or subsampling). For example, Sen et al. (2010) suggest to use the moon bootstrap, and Kosorok (2008) and Sen et al. (2010) suggest to use the smoothed bootstrap to get valid inference for the Grenander estimator; for the general cube root estimators, Delgado et al. (2001) suggest to use the subsampling (especially in hypothesis testing), Lee and Pun (2006) suggest to use the moon bootstrap, and Léger and MacGibbon (2006) suggest to use the smoothed bootstrap (especially for Chernoff's modal estimator when the unidimensional distribution is symmetric). For the nonregular M-estimators with convex objective functions, Bose and Chatterjee (2001) suggest to use the moon bootstrap. In

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

threshold regression, Seijo and Sen (2011) suggest to use the smoothed bootstrap, while we suggest to use the NPI as will be discussed in Section 4.

3. INVALIDITY OF THE NONPARAMETRIC BOOTSTRAP

Before a formal discussion of the bootstrap failure, we will first specify some regularity conditions and review the main asymptotic results in threshold regression. Throughout the following discussion, we maintain the assumption that the data are i.i.d. sampled and the mean independence assumption $E[\varepsilon|x,q]=0$ holds.

Assumption D.

1. $w_i \in \mathbb{W} \equiv \mathbb{R}\times\mathbb{X}\times\mathbb{Q} \subset \mathbb{R}^{k+2}$, $\beta_\ell \in B_\ell \subset \mathbb{R}^k$, $0<\sigma_\ell \in \Omega_\ell \subset \mathbb{R}$, Ω_ℓ is compact, $\gamma \in \Gamma = [\underline{\gamma}, \overline{\gamma}]$ which is compact, $\beta_{1,0}\neq\beta_{2,0}$, and $\sigma_{1,0}\neq\sigma_{2,0}$, where $\neq$ is an element by element operation.
2. $E\left[xx'\right] > E\left[xx'\mathbf{1}(q\le\gamma)\right] > 0$ for all $\gamma\in\Gamma$.
3. $E\left[xx'|q=\gamma\right] > 0$ for γ in an open neighborhood of γ_0.
4. $f_q(\gamma)$ is continuous, and $0<\underline{f}_q \le f_q(\gamma)\le \overline{f}_q<\infty$ for $\gamma\in\Gamma$.
5. $E[\varepsilon^4]<\infty$, and $E[\|x\|^4]<\infty$.
6. $E\left[\|x\|^2\middle| q=\gamma\right]<\infty$ and $E[\|x\varepsilon\||q=\gamma]<\infty$ for γ in an open neighborhood of γ_0.
7. The conditional distribution of $(x',\varepsilon)'$ given $q_i=\gamma$ is continuous in γ for γ in a neighborhood of γ_0.
8. z_{1i} and z_{2i} have absolutely continuous distributions, where $z_{\ell i}$ follows the conditional distribution of $\overline{z}_{\ell i}$ given $q_i=\gamma_0$, and

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

Assumption D is very standard and roughly a subset of Assumption 1 in Hansen (2000) except D7 and D8. Specifically, D1 does not require B_ℓ to be compact, and D2 excludes the possibility that γ_0 is on the boundary of q's support. D4 requires a positive and finite $f_q(\cdot)$ in the neighborhood of γ_0; see Yu and Zhao (2009) for relaxing this assumption. D7 states that the threshold effect is captured completely by the changes in mean and variance; the distribution of $(x',\varepsilon)'$ keeps stable in the neighborhood of $q_i=\gamma_0$. This assumption corresponds to that $D(\gamma)=E\left[x_ix_i'|q_i=\gamma\right]$ and $V(\gamma)=E\left[x_ix_i'\varepsilon_i^2|q_i=\gamma\right]$ are continuous at $\gamma=\gamma_0$ in Hansen (2000)'s Assumption 1. It implies that the conditional distribution of $\overline{z}_{\ell i}$ in D8 given $q_i=\gamma$ is continuous in γ for γ in a neighborhood of γ_0. $\overline{z}_{1i}$ represents the effect on the objective function when the threshold point is displaced on the left of γ_0, and $\overline{z}_{2i}$ represents the converse case. The absolute continuity of the distribution of $z_{\ell i}$ guarantees that the localized objective function has a unique minimizer and allows for some discrete covariates in x. See Section 3.1 of Hansen (2000) for more discussions on Assumption D.

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

From Chan (1993) or Yu (2012), under Assumption D,

$$\sqrt{n}\left(\widehat{\beta}_1-\beta_{1,0}\right)\xrightarrow{d} Z_{\beta_1}\sim E\left[xx'\mathbf{1}\left(q\le\gamma_0\right)\right]^{-1}\cdot N\left(\mathbf{0},E\left[xx'\sigma_{1,0}^2\varepsilon^2\mathbf{1}\left(q\le\gamma_0\right)\right]\right),$$
$$\sqrt{n}\left(\widehat{\beta}_2-\beta_{2,0}\right)\xrightarrow{d} Z_{\beta_2}\sim E\left[xx'\mathbf{1}\left(q>\gamma_0\right)\right]^{-1}\cdot N\left(\mathbf{0},E\left[xx'\sigma_{2,0}^2\varepsilon^2\mathbf{1}\left(q>\gamma_0\right)\right]\right),$$

and

$$n\left(\widehat{\gamma}-\gamma_0\right)\xrightarrow{d}\arg\min_{\upsilon} D\left(\upsilon\right)\equiv Z_\gamma, \tag{7}$$

with

$$D\left(\upsilon\right)=\begin{cases}\sum_{i=1}^{N_1(|\upsilon|)} z_{1i}, \text{ if } \upsilon\le 0;\\ \sum_{i=1}^{N_2(\upsilon)} z_{2i}, \text{ if } \upsilon>0.\end{cases}$$

Here, Z_{β_1}, Z_{β_2}, $\{z_{1i},z_{2i}\}_{i\ge1}$, $N_1(\cdot)$, and $N_2(\cdot)$ are independent of each other, and $N_\ell(\cdot)$ is a Poisson process with intensity $f_q(\gamma_0)$. Since $D(\upsilon)$ is the cadlag step version of a two-sided compound Poisson process with $D(0)=0$, there is a random interval $[M_-,M_+)$ minimizing $D(\upsilon)$. Since the left endpoint of the minimizing interval is taken as the LSE, $Z_\gamma=M_-$. The asymptotic distribution of $\widehat{\beta}$ is the same as that in the case when γ_0 is known, and the asymptotic distribution of $\widehat{\gamma}$ is the same as that in the case when β_0 is known. As shown in Yu (2008, 2012), this is mainly because γ is a "middle" boundary of q, so only the local information around γ_0 is useful in the estimation of γ_0. For example, $z_{\ell i}$ follows a distribution conditional on q in the neighborhood of γ_0, and the intensity of $N_\ell(\cdot)$ is $f_q(\gamma_0)$. The distribution of Z_γ is continuous, and the explicit form of its density is derived in Appendix D of Yu (2012).

3.1. The Nonparametric Bootstrap

In this bootstrap scheme, the empirical bootstrap measure in (3) is

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

where M_{ni} is the number of times that w_i is drawn from the original sample, and $\mathbf{M}_n\equiv(M_{n1},\ldots,M_{nn})$ follows the multinomial distribution with parameters n and cell probabilities all equal to $\frac{1}{n}$ (and independent of the original data $\{w_i\}_{i=1}^n$). Note that M_{ni}'s are dependent of each other, since the sum of M_{ni} must be n, but they become independent as n goes to infinity. In other words, Poissonization is possible,[10] i.e., $M_{n1},\ldots,M_{nn}$ can be approximated by $M_{N_n,1},\ldots,M_{N_n,n}$, which are i.i.d. Poisson variables with mean 1, where N_n is a Poisson random variable with mean n and independent of the original observations. Poissonization makes

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

the bootstrap sampling as if independent draws were conducted on each data point, which greatly simplifies the derivation of the asymptotic bootstrap distribution. From Lemma 1 in Appendix B, M_{ni} can indeed be substituted by $M_{N_n i}$ without affecting the asymptotic properties of P_n^*. Define $h = \left(u_1', u_2', v\right)' \equiv \left(u', v\right)'$ as the local parameter for θ, then uniformly for h in a compact set,

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

where

$$
\begin{aligned}
D_n(v) &= \sum_{i=1}^n \left[m\left(w_i | \gamma_0 + \frac{v}{n} \right) - m\left(w_i | \gamma_0 \right) \right] \\
&= \sum_{i=1}^n \overline{z}_{1i} \mathbf{1}\left(\gamma_0 + \frac{v}{n} < q_i \le \gamma_0 \right) + \sum_{i=1}^n \overline{z}_{2i} \mathbf{1}\left(\gamma_0 < q_i \le \gamma_0 + \frac{v}{n} \right), \\
D_n^*(v) &= \sum_{i=1}^n M_{N_n,i} \left[m\left(w_i | \gamma_0 + \frac{v}{n} \right) - m\left(w_i | \gamma_0 \right) \right] \\
&= \sum_{i=1}^n M_{N_n,i} \overline{z}_{1i} \mathbf{1}\left(\gamma_0 + \frac{v}{n} < q_i \le \gamma_0 \right) + \sum_{i=1}^n M_{N_n,i} \overline{z}_{2i} \mathbf{1}\left(\gamma_0 < q_i \le \gamma_0 + \frac{v}{n} \right), \\
W_n(u) &= W_{1n}(u_1) + W_{2n}(u_2),\ W_n^*(u) = W_{1n}^*(u_1) + W_{2n}^*(u_2),
\end{aligned}
$$

with

$$
\begin{aligned}
m(w|\gamma) &= \left(y - x'\beta_{1,0} \mathbf{1}(q \le \gamma) - x'\beta_{2,0} \mathbf{1}(q > \gamma) \right)^2, \\
W_{1n}(u_1) &= u_1' \left(\frac{2\sigma_{1,0}}{\sqrt{n}} \sum_{i=1}^n x_i \varepsilon_i \mathbf{1}\left(q_i \le \gamma_0 \right) \right), \\
W_{1n}^*(u_1) &= u_1' \left(\frac{2\sigma_{1,0}}{\sqrt{n}} \sum_{i=1}^n \left(M_{N_n,i} - 1 \right) x_i \varepsilon_i \mathbf{1}\left(q_i \le \gamma_0 \right) \right), \\
W_{2n}(u_2) &= u_2' \left(\frac{2\sigma_{2,0}}{\sqrt{n}} \sum_{i=1}^n x_i \varepsilon_i \mathbf{1}\left(q_i > \gamma_0 \right) \right), \\
W_{2n}^*(u_2) &= u_2' \left(\frac{2\sigma_{2,0}}{\sqrt{n}} \sum_{i=1}^n \left(M_{N_n,i} - 1 \right) x_i \varepsilon_i \mathbf{1}\left(q_i > \gamma_0 \right) \right).
\end{aligned}
$$

Corresponding to the local data usage in $D_n(v)$ (which is transferred to $D(v)$ asymptotically), essentially only the bootstrap sampling on finite data points

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

around γ_0 is relevant in $D_n^*(\upsilon)$. In contrast, $W_{1n}^*(u_1)$ and $W_{2n}^*(u_2)$ take an average form, which makes their asymptotic properties very different from those of $D_n^*(\upsilon)$. The following Theorem 1 gives the joint weak limit of $\left(W_n(u), W_n^*(u), D_n(\upsilon), D_n^*(\upsilon)\right)$ under P_r, which is Step 1 in Section 2.3 and is critical in checking the bootstrap validity in Theorem 2.

THEOREM 1. *Under Assumption D,*

$$\left(W_n(u), W_n^*(u), D_n(\upsilon), D_n^*(\upsilon)\right) \rightsquigarrow \left(W(u), W^*(u), D(\upsilon), D^*(\upsilon)\right)$$

on any compact set, where $\rightsquigarrow$ *signifies the weak convergence under* P_r*,*

$$W(u) = 2u_1' W_1 + 2u_2' W_2,\ W^*(u) = 2u_1' W_1^* + 2u_2' W_2^*,$$

with W_1 *and* W_1^* *following the same distribution* $N\left(0, E\left[x^2\sigma_{1,0}^2\varepsilon^2\mathbf{1}(q \le \gamma_0)\right]\right)$*,* W_2 *and* W_2^* *following the same distribution* $N\left(0, E\left[x^2\sigma_{2,0}^2\varepsilon^2\mathbf{1}(q > \gamma_0)\right]\right)$*, and* $(D(\upsilon), D^*(\upsilon))$ *being a bivariate vector of compound Poisson processes.* $D(\upsilon)$ *is the same as that in (7), and*

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

with $\left\{N_{i-}^*, N_{i+}^*\right\}_{i\ge1}$ *being independent standard Poisson variables.*[11] *Furthermore,* W_1*,* W_2*,* W_1^**,* W_2^**,* $\{z_{1i}, z_{2i}\}_{i\ge1}$*,* $\left\{N_{i-}^*, N_{i+}^*\right\}_{i\ge1}$*,* $N_1(\cdot)$*, and* $N_2(\cdot)$ *are independent of each other.*

$D^*(\upsilon)$ is called a *multiplier compound Poisson process*, which is highly correlated with $D(\upsilon)$, since $\{z_{1i}, z_{2i}\}_{i\ge1}$, $N_1(\cdot)$, and $N_2(\cdot)$ are the same as those in $D(\upsilon)$. The randomness in $\{z_{1i}, z_{2i}\}_{i\ge1}$, $N_1(\cdot)$, and $N_2(\cdot)$ is introduced by the original data, so the randomness introduced by the bootstrap appears only in $\left\{N_{i-}^*, N_{i+}^*\right\}_{i\ge1}$. Since $E\left[N_{i-}^*\right] = E\left[N_{i+}^*\right] = 1$ for any i, the average jump size of z_{1i}^* and z_{2i}^* in $D^*(\upsilon)$ is the same as z_{1i} and z_{2i} in $D(\upsilon)$. But the distribution of z_{1i}^* and z_{2i}^* instead of their mean determines the jumps in $D^*(\upsilon)$. The distribution of z_{1i}^* is

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

$$= \begin{cases} \sum_{k=1}^{\infty} \frac{e^{-1}}{k!}\Phi_1\left(\frac{x}{k}\right), & \text{if } x < 0; \\ e^{-1} + \sum_{k=1}^{\infty} \frac{e^{-1}}{k!}\Phi_1\left(\frac{x}{k}\right), & \text{if } x \ge 0; \end{cases}$$

where $\Phi_1(\cdot)$ is the cumulative distribution function 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

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

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 ith and $(i-1)$th jump on $v \le 0$ ($v > 0$) in $D(v)$ are combined into one jump. When N_{i-}^* (N_{i+}^*) is greater than 1, the ith jump on $v \le 0$ ($v > 0$) in $D(v)$ is magnified.

Although $D_n^*(v)$ converges weakly under P_r, Lemma 4.2 of Seijo and Sen (2011) shows that the conditional weak limit of $D_n^*(v)$ does not exist. Especially, the conditional weak limit of $D_n^*(v)$ cannot be $D^*(v)$ conditional on $D(v)$. Otherwise, by the argmax continuous mapping theorem, the conditional weak limit of $n\left(\widehat{\gamma}^* - \gamma_0\right)$ would be the conditional distribution of $\arg\min_v D^*(v) \equiv Z_\gamma^*$ given $D(v)$, which critically depends on $D(v)$. This, however, cannot happen by the Hewitt-Savage zero-one law. Such analyses correspond to level two and level one of the nonexistence proof in Section 2.3, respectively. The following theorem states the asymptotic bootstrap distributions of $\widehat{\beta}^*$ and $\widehat{\gamma}^*$. Particularly, we show that the conditional weak limit of $n\left(\widehat{\gamma}^* - \widehat{\gamma}\right)$ does not exist, which corresponds to level three of the nonexistence proof in Section 2.3.

THEOREM 2. *Under Assumption D,*

(i) $\sqrt{n}\left(\widehat{\beta}^* - \widehat{\beta}\right) \overset{*}{\rightsquigarrow} Z_\beta^*$ *as* $n \to \infty$, *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)

$$\left(n\left(\widehat{\gamma} - \gamma_0\right), n\left(\widehat{\gamma}^* - \gamma_0\right)\right) \rightsquigarrow \left(Z_\gamma, Z_\gamma^*\right) \text{ as } n \to \infty,$$

so

$$n\left(\widehat{\gamma}^* - \widehat{\gamma}\right) \rightsquigarrow Z_\gamma^* - Z_\gamma \text{ as } n \to \infty,$$

and $Z_\gamma^* - Z_\gamma$ *is independent of* Z_β^*.

(iii) $n\left(\widehat{\gamma}^* - \widehat{\gamma}\right)$ *does not converge in distribution in P probability as* $n \to \infty$.

When γ_0 is known, the validity of the nonparametric bootstrap for β is a standard result. When γ_0 is unknown, (i) shows that the nonparametric bootstrap for β is still valid and the asymptotic bootstrap distribution is the same as that in the case when γ_0 is known. As argued in Section 2.3, the validity of the bootstrap for β is due to the special form of the weak limit of $nP_n^*\left(m\left(\cdot\left|\beta_0 + \frac{u}{\sqrt{n}}, \gamma_0 + \frac{v}{n}\right.\right) - m\left(\cdot|\beta_0, \gamma_0\right)\right)$ as a function of u.

From (ii), $n\left(\widehat{\gamma}^* - \widehat{\gamma}\right) \xrightarrow{d} Z_\gamma^* - Z_\gamma$ as $n \to \infty$ in P_r probability. If we denote the component measure as $\left(Z_\gamma^* - Z_\gamma\right)|D(\cdot)$, then the distribution of $Z_\gamma^* - Z_\gamma$ is the average of all these component measures when $D(\cdot)$ is sampled according to the probability P. The distribution of Z_γ^* under P_r can be derived by the simulation method proposed in Appendix D of Yu (2012), with caution that two jumps in $D^*(v)$ can be combined into one when $N_{i\pm}^*$ equal zero. This distribution is expected to be continuous but more spreading than Z_γ as $N_{i\pm}^*$ can take values other

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

than 1. Since $D(\upsilon)$ and $D^*(\upsilon)$ are highly correlated, Z_γ^* and Z_γ are highly correlated under P_r. In fact, their correlation is so strong that there is a point mass at zero in the distribution of $Z_\gamma^* - Z_\gamma$, although Z_γ^* and Z_γ are both continuous; see the example in the next subsection for more details on $Z_\gamma^* - Z_\gamma$. Another result in (ii) is that $Z_\gamma^* - Z_\gamma$ is independent of Z_β^*, which is similar to the asymptotic distribution in (7). This is because the bootstrap samplings for inference on β and γ use information independently. The randomness introduced by the bootstrap in $\widehat{\beta}^*$ takes an average form; see, e.g., the term $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(M_{N_n,i}-1\right)x_i\varepsilon_i\mathbf{1}(q_i \le \gamma_0)$ 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 $\left\{M_{N_n,i}\right\}_{i=1}^n$ and makes the bootstrap distribution for β asymptotically normal. In contrast, only "local" sampling on the data around γ_0 is informative for inference on the threshold point as shown in D_n^* above. This is why $N_{i\pm}^*$ appear in $D^*(\upsilon)$, but not in $W^*(u)$. See Yu (2008) for more discussions on the independence between the local information and the global information.[12]

To understand (iii), more discussions on the component measure will be helpful. Since the randomness in $\left(Z_\gamma^* - Z_\gamma\right)|D(\cdot)$ comes only from $N_{i\pm}^*$, it is discrete and critically depends on $D(\cdot)$. The magnitude of the point masses depends on $\{z_{1i}, z_{2i}\}_{i\ge 1}$, and their locations depend on $N_\ell(\cdot)$. Since Z_γ is fixed conditional on $D(\cdot)$, the distribution of $\left(Z_\gamma^* - Z_\gamma\right)|D(\cdot)$ is just a location shift of $Z_\gamma^*|D(\cdot)$. If Z_γ^* equals Z_γ with a positive probability for a given $D(\cdot)$, then $\left(Z_\gamma^* - Z_\gamma\right)|D(\cdot)$ has a point mass at zero; see the example in the next subsection for more descriptions on this.

In the supplementary materials, we show that the wild bootstrap, bootstrapping residuals, and the parametric wild bootstrap are all invalid due to the same reason as in the nonparametric bootstrap, namely, only the bootstrap sampling in the neighborhood of γ_0 is informative to the inference on γ. We further construct a set A with $P(A) > 0$ such that for each $\omega \in A$, the conditional weak limit of $n\left(\widehat{\gamma}^* - \widehat{\gamma}\right)$ does not exist.

3.2. A Numerical Example

We pause here to provide more intuitions on the asymptotic bootstrap distributions in Theorem 2 using a concrete example. Suppose the population model is

$$y = \mathbf{1}(q \le \gamma) + \varepsilon, \quad q \sim U[0,1], \tag{8}$$

where $\varepsilon \sim N(0,1)$ is independent of q, and $\gamma_0 = 1/2$ is the only parameter of interest. From (7), the asymptotic distribution of the LSE is $\arg\min_{\upsilon} D(\upsilon)$, where in $D(\upsilon)$, $z_{1i} = 1 + 2\varepsilon_i^-$ and $z_{2i} = 1 - 2\varepsilon_i^+$ with ε_i^- and ε_i^+ following the same distribution as ε, and $N_1(\cdot)$ and $N_2(\cdot)$ are standard Poisson processes. The $D^*(\upsilon)$ in the asymptotic bootstrap distribution of the LSE is given in Theorem 1

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

with z_{1i} and z_{2i} being replaced by the specific form here. To simplify notations, Z^* and Z without subscripts are used for Z^*_γ and Z_γ in the following discussion.

We first check Step 2 in Section 2.3. By the algorithms in Appendix D of Yu (2012), we have $Var(Z^*) = 110.72$, which is more than double of $Var(Z) = 31.44$, so the nonparametric bootstrap cannot be consistent. A similar result can be found in Table 1 of Seijo and Sen (2011). From this result, we expect $Z^* - Z$ to have a thicker tail than Z. To study the distribution of $Z^* - Z$, we need to first study the component measure $(Z^* - Z)|D(\cdot)$. Such a measure can be thought as an approximation of the finite-sample distribution of $n\left(\widehat{\gamma}^* - \widehat{\gamma}\right)$ conditional on an original sample path, the weak limit of which does not exist by Theorem 2.

Figure 1 shows the dependence of $(Z^* - Z)|D(\cdot)$ on $D(\cdot)$. For comparison, the density of Z is also dotted in Figure 1, which is very different from $(Z^* - Z)|D(\cdot)$ in all cases. Our first impression of $(Z^* - Z)|D(\cdot)$ is that its distribution is discrete, which can also be seen from the finite distribution of $n\left(\widehat{\gamma}^* - \widehat{\gamma}\right)$ in Figure 1 of Seijo and Sen (2011). 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 $D(\cdot)$. Second, large point masses often happen at deeply negative jumps, and at the left of them there are decaying point masses. Third, there are not many point masses in the right neighborhood of 0. This phenomenon is due to the fact that the LLSE is taken as the estimator; when the MLSE is used, this will not happen. The distributions of Z, Z^*, and $Z^* - Z$ under P_r are shown in Figure 2. As expected, the distribution of Z^* has a thicker tail than Z. 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 at zero. Also, the distribution of $Z^* - Z$ shows that there is not much density in the right neighborhood of zero, which can be derived from the third point above.

Although Figure 1 and Figure 2 show that the nonparametric bootstrap is not valid, it is still useful in constructing CIs for γ if its quantiles are close to the asymptotic counterparts. In Table 1 and Figure 3, some simulation results are reported for this example based on $(Z^* - Z)|D(\cdot)$ and $Z^* - Z$, where an

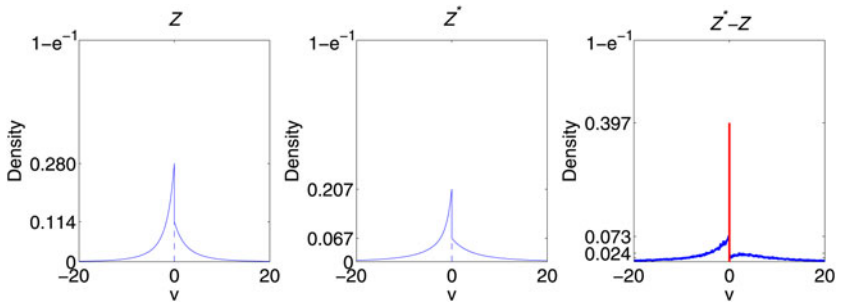


FIGURE 2. Comparison of asymptotic distributions under P_r.

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

TABLE 1. Characterizations of quantiles and coverages in the bootstrap

	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%

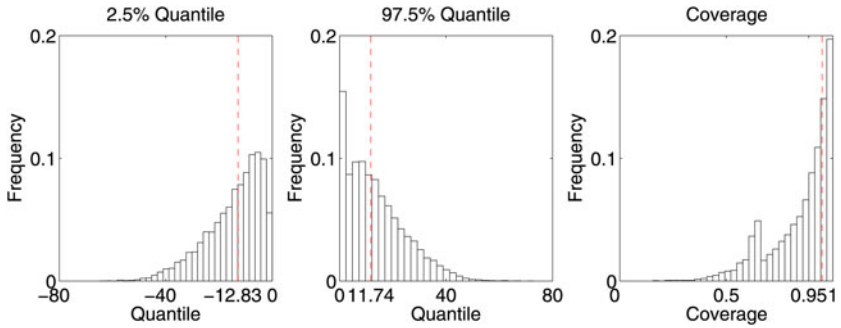


FIGURE 3. Distributions of 2.5% and 97.5% quantiles and coverage.

equal-tailed 95% coverage CI is used as the bootstrap CI. The numbers associated with "Asymptotic" are the benchmark for bootstrap inference. In Table 1, "Average" means the average of quantiles and coverage among all sample paths of $D(\cdot)$. The coverage 84.93% is less than the target 95%, which confirms the simulation results in Seo and Linton (2007) and Seijo and Sen (2011) that the bootstrap has an undercover problem. From Table 1 and Figure 3, the 2.5% and 97.5% quantiles of the component measure depend heavily on the original samples, and the same happens to coverage. In Table 1, the items associated with "Average" are different from those associated with "Average Bootstrap," because the two operations of taking a quantile and taking an average of distributions cannot be exchanged. By comparing the quantiles under "Asymptotic" and "Average Bootstrap", we conclude that $Z^* - Z$ has a thicker tail than Z as previously predicted. In Figure 3, there is a large point mass at zero in the distribution of the 97.5% quantile. This is not surprising according to Figure 1 where the probability in $(Z^* - Z)|D(\cdot)$ 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. In summary, Table 1 and Figure 3 suggest that we cannot use the bootstrap CI as the set estimator of the threshold point.

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

3.3. Comparison with Two Other Boundaries

We now compare the threshold point with two other boundaries in the literature. The first one is discussed in Andrews (2000), where $\{X_i\}_{i=1}^n$ is a sequence of i.i.d.$N(\mu,1)$ random variables with $\mu \in \mathbb{R}^+ \equiv \{z : z \geq 0\}$. A natural estimator of μ is $\widehat{\mu}_n = \max\left\{\overline{X}_n, 0\right\}$ which is also the MLE of μ, where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. The second one is the conventional boundary in the error term discussed in Hirano and Porter (2003) and Chernozhukov and Hong (2004): $y_i = g(x_i,\beta) + \varepsilon_i$, where the conditional density $f(\varepsilon|x;\alpha)$ has a jump at $\varepsilon = 0$, and the notations are similarly defined as in threshold regression. Only parametric models are considered here, so natural estimators of β are the MLE and the Bayes estimator.

The boundary estimator in Andrews (2000) is an example of nondifferentiable functions of the empirical; see, e.g., Example 6 of Bickel et al. (1997). We first check the bootstrap validity using the procedure in Section 2.3. When $\mu = 0$, $\sqrt{n}\left(\overline{X}_n^* - \overline{X}_n\right)$ and $\sqrt{n}\overline{X}_n$ are asymptotically independent $N(0,1)$ under P_r, where $\overline{X}_n^* = n^{-1}\sum_{i=1}^n X_i^*$ with X_i^* following the empirical measure F_n of $\{X_i\}_{i=1}^n$. Denote the limit variables as Z^* and Z, respectively, then the asymptotic distribution of the average bootstrap estimator of μ at 0 is as follows:

$$\begin{aligned}\sqrt{n}\left(\widehat{\mu}_n^* - \widehat{\mu}_n\right) &= \sqrt{n}\left(\max\left\{\overline{X}_n^*, 0\right\} - \max\left\{\overline{X}_n, 0\right\}\right)\\ &\rightsquigarrow \max\left\{Z + Z^*, 0\right\} - \max\{Z, 0\} \equiv Z_0^*. \qquad (9)\end{aligned}$$

While conditional on ω, $\sqrt{n}\left(\widehat{\mu}_n^* - \widehat{\mu}_n\right)$ has no weak limit just as in bootstrapping the threshold point. To understand this result, observe that $\sqrt{n}\overline{X}_n^*$ does not have a weak limit when conditional on ω, since $\sqrt{n}\overline{X}_n^* = \sqrt{n}\left(\overline{X}_n^* - \overline{X}_n\right) + \sqrt{n}\overline{X}_n$ and $\sqrt{n}\overline{X}_n$ has no limit, although $\sqrt{n}\left(\overline{X}_n^* - \overline{X}_n\right)$ has a weak limit by the multiplier central limit theorem (Lemma 2.9.5 of Van der Vaart and Wellner, 1996). Thus

$$\sqrt{n}\left(\widehat{\mu}_n^* - \widehat{\mu}_n\right) = \begin{cases} \sqrt{n}\left(\overline{X}_n^* - \overline{X}_n\right) - \sqrt{n}\overline{X}_n^* \mathbf{1}\left(\overline{X}_n^* < 0\right), & \overline{X}_n \geq 0,\\ \sqrt{n}\overline{X}_n^* \mathbf{1}\left(\overline{X}_n^* \geq 0\right), & \overline{X}_n < 0,\end{cases}$$

has no conditional weak limit; see Example 3.8 of Shao and Tu (1995) for related discussions.

As in bootstrapping the threshold point, the asymptotic average bootstrap distribution (9) can be treated as the average of the conditional distributions of Z_0^* given Z for all realizations of Z. Note that the component measure

$$Z_0^*|Z = \begin{cases} \max\{Z^*, -Z\}, & \text{if } Z \geq 0,\\ Z + \max\{Z^*, -Z\}, & \text{if } Z < 0,\end{cases}$$

also depends on the original sampling path, although only through an asymptotically sufficient statistic $\overline{X}_n$. This point is shown in Figure 4 where $Z_0^*|Z$ 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.

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

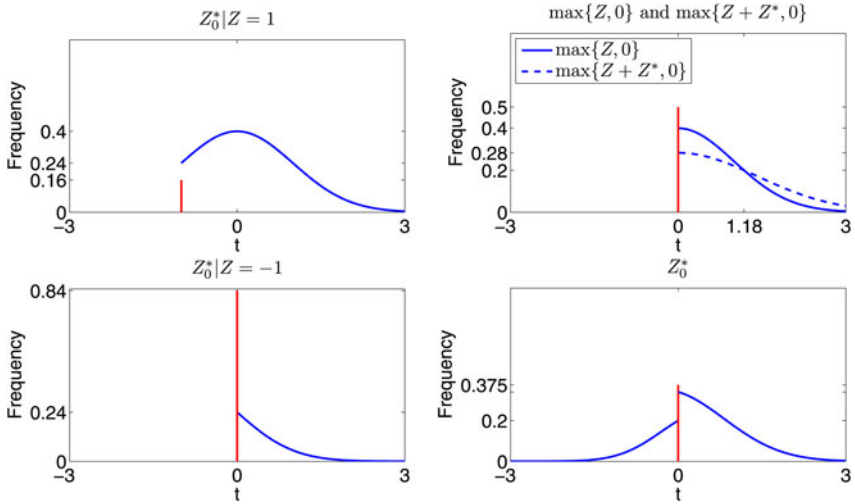


FIGURE 4. $Z_0^*|Z$, $\max\{Z,0\}$, $\max\left\{Z+Z^*,0\right\}$ and Z_0^* in Andrews (2000).

$Z_0^*|Z$ matches the asymptotic distribution only when $Z=0$. The asymptotic average bootstrap distribution is also shown in Figure 4, where there is also a point mass at zero as in the example of Section 3.2. When $\mu>0$, $\sqrt{n}\left(\overline{X}_n-\mu\right)\xrightarrow{d}Z$, so $\sqrt{n}\overline{X}_n\xrightarrow{d}\infty$. This corresponds to $Z=\infty$ in (9), so the bootstrap is valid if $\mu>0$. Because the bootstrap validity requires that the asymptotic bootstrap distribution matches the asymptotic distribution for all parameter values, the bootstrap fails in this case. The above argument provides an alternative way to show the bootstrap failure in Andrews (2000).

The reasons for the bootstrap failure in Andrews (2000) and threshold regression are different. In Andrews (2000), 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. Checking the three sufficient conditions for the bootstrap validity given on page 1209 in Bickel and Freedman (1981), the uniformity condition (6.1b) fails in the bootstrap of γ, while in the 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 mentioned in Section 3.4 of Yu (2012). In the former case, the continuity condition (6.1c) still fails, but in the later case, the uniformity condition (6.1b) holds; see the arguments in Counterexample 2 of Bickel and Freedman (1981) for more details.

In the conventional boundary estimation, Chernozhukov and Hong (2004) show that the parametric wild bootstrap is valid in their Remark 3.6, but as mentioned at the end of Section 3.1 it fails in threshold regression. 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

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

TABLE 2. Three boundaries, validity of the bootstraps, and reasons for bootstrap failure

	Andrews (2000)	Chernozhukov and Hong (2004)	Threshold regression
Parameter of interest	$\mu = 0$	β	γ
Boundary of what space?	Parameter space	Sample space (error term)	Sample space (covariate)
Parametric bootstrap	Invalid	Valid	Valid
(Parametric) wild bootstrap	-	Valid	Invalid
Nonparametric bootstrap	Invalid	-	Invalid
Reason for bootstrap failure	Continuity	-	Uniformity

Hong (2004), the boundary parameters appear in the conditional distribution of y given x, and there is no boundary parameter in the distribution of x, so simulating from $f(\varepsilon|x;\widehat{\alpha})$ in their setup is enough to mimic the original probability structure around the boundary. In threshold regression, however, γ is a boundary of q. To keep the probability structure around γ, we must simulate also from $f(x,q)$. The discussion above is summarized in Table 2. The subsampling and the moon bootstrap are valid in all cases, so are omitted from the table.

4. REMEDIES OF THE BOOTSTRAP FAILURE

From Section 3, the nonparametric bootstrap and various wild bootstrap schemes are inconsistent for inference on the threshold point in discontinuous threshold regression. The parametric bootstrap is valid, but not feasible in most cases due to the need to specify a complete likelihood. Additionally, inference based on the asymptotic distribution (7) is difficult. In this section, we briefly review alternative inferring approaches that have been proposed in the literature; see Yu (2008) for more detailed descriptions.

4.1. Smoothed Bootstrap and Other Remedies

Five remedies are potentially useful. The first one is suggested in Hansen (2000) who uses a framework with an asymptotically diminishing threshold effect in mean and proves that the CI constructed by inverting the likelihood ratio statistic is asymptotically valid.[13] The second one is the subsampling method of Gonzalo and Wolf (2005) or the moon bootstrap of Seijo and Sen (2011) as mentioned in Section 2.2. The third one 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. The fourth one is suggested by Yu (2008) who uses the nonparametric posterior to construct CIs for γ in the present framework. This method is described in the supplementary materials.

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

At last, Seijo and Sen (2011) suggest to use the smoothed bootstrap to cure the bootstrap failure. They show the bootstrap validity in a simple threshold model without nonconstant covariates in x. When there are nonconstant covariates in x, their proof can be extended to the case with q being independent of all other covariates (and ε). However, in practice, this extension is not applicable, since it is not very reasonable to assume that q is independent of all other components of the model.[14] When there is correlation between q and other covariates, the smoothed bootstrap is hard to apply even if ε is assumed to be independent of $(x', q)'$. There are two reasons. First, we must maintain the correlation structure between q and x to make the smoothed bootstrap valid, so we must simulate from the joint distribution of $(x', q)'$. When the dimension of $(x', q)'$ is large, we will suffer from the curse of dimensionality in estimating the joint density $f(x, q)$. Second, when there are complicated structures in the joint distribution of $(x', q)'$ such as discreteness or boundaries, a globally consistent estimator of $f(x, q)$ is hard to construct and simulate from.

The validity of the smoothed bootstrap in the simple threshold model is quite understandable. We know the bootstrap can work, because P_n can mimic P. From Assumption D4, $f_q(\cdot)$ is continuous in a neighborhood of γ_0, while in the empirical measure F_n, the distribution of q is discrete. This is the main reason why the nonparametric bootstrap fails. To mimic the continuity of $f_q(\cdot)$, we must provide a consistent estimator of $f_q(\cdot)$ at least in the neighborhood of γ_0, which is exactly what the smoothed bootstrap does in the simple threshold model. In fact, the smoothed bootstrap has already been used as a remedy of the bootstrap failure in the (covariate) boundary estimation; see, e.g., Section 2.4 of Bickel et al. (1997). In what follows, we prove the smoothed bootstrap validity in a more applicable setup than that in Seijo and Sen (2011). First, we specify some extra assumptions above Assumption D.

Assumption E.

1. The support of $(x', q)'$ is compact, and $f(x, q)$ is bounded on its support.
2. ε is independent of $(x', q)'$ and has a continuous distribution.
3. There is an estimator $\widehat{f}(x, q)$ such that $\left\| \widehat{f}(x, q) - f(x, q) \right\|_\infty \xrightarrow{a.s.} 0$.
4. The parameter space B_ℓ is compact.

Assumption E1 does not exclude the dependence between q and x. Also, we assume $(x', q)'$ is bounded and independent of ε. This assumption can avoid technical difficulties in our proof. E3 is a high-level assumption, which is a standard result in nonparametric estimation; see, e.g., Silverman (1978) or Hansen (2008) for the kernel smoothing estimator of $\widehat{f}(x, q)$.

THEOREM 3. *Under Assumptions D and E, the smoothed bootstrap is valid for both β and γ; that is, $\sqrt{n}\left(\widehat{\beta}^* - \widehat{\beta}\right) \overset{*}{\rightsquigarrow} Z_\beta^*$ as $n \to \infty$, where Z_β^* has the same distribution as Z_β, and $n\left(\widehat{\gamma}^* - \widehat{\gamma}\right) \overset{*}{\rightsquigarrow} Z_\gamma^*$ as $n \to \infty$, where Z_γ^* has the same distribution as Z_γ.*

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

As the conditional weak limit of $n\left(\widehat{\gamma}^{*}-\widehat{\gamma}\right)$ does not depend on the original sample path, the Hewitt-Savage zero-one law is not violated. Since $Z_{\gamma}^{*} \overset{d}{=} Z_{\gamma}$, the smoothed bootstrap is valid. The proof of Theorem 3, which is an extension of Theorem 1 and 2 to triangular arrays, is put in the supplementary materials.

When ε is independent of $(x', q)'$, the NPI does not suffer from the two drawbacks of the smoothed bootstrap mentioned above, since we need only estimate $f_{\varepsilon}(\cdot)$. In this case, the simulation studies in Yu (2008) show that the NPI performs the best among all the remedies aforementioned, including the smoothed bootstrap. Since Seijo and Sen (2011) suggest the smoothed bootstrap as the solution to bootstrap failure, we will concentrate on the NPI and the smoothed bootstrap in the simulation studies below, considering more general setups than those in Yu (2008) and Seijo and Sen (2011).

4.2. Simulation Results

We will consider four DGPs in our simulation. More specifically, in DGP1, $y = \beta_{1,0}\mathbf{1}(q \leq \gamma_0) + \beta_{2,0}\mathbf{1}(q > \gamma_0) + \varepsilon$, where $\left(\beta_{1,0}, \beta_{2,0}, \gamma_0\right) = (-1, 1, 0)$, q and ε are independent, $q \sim N(0, 2)$, and $\varepsilon \sim N(0, 1)$. This is the first DGP considered in Seijo and Sen (2011). DGP2 is the same as DGP1 except that $\varepsilon = \epsilon\sqrt{0.1 + 0.2|q| + 0.3q^2}$, where $\epsilon \sim N(0, 1)$ and is independent of q. Here, the correlation between q and ε is only through heteroskedasticity. In DGP3, $y = (1\ q)\beta_{1,0}\mathbf{1}(q \leq \gamma_0) + (1\ q)\beta_{2,0}\mathbf{1}(q > \gamma_0) + \varepsilon$, where $\beta_{1,0} = (-1\ -1)'$, $\beta_{2,0} = (1\ 1)'$, $\gamma_0 = 0$, ε and q are independent, $q \sim N(0, 2)$, and $\varepsilon \sim N(0, 1)$. In DGP4, $y = (1\ x)\beta_{1,0}\mathbf{1}(q \leq \gamma_0) + (1\ x)\beta_{2,0}\mathbf{1}(q > \gamma_0) + \varepsilon$, where $\beta_{1,0} = (-1\ -1)'$, $\beta_{2,0} = (1\ 1)'$, $\gamma_0 = 0$, ε and $(x, q)'$ are independent, $\varepsilon \sim N(0, 1)$, and $(x, q)' \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1.4 \\ 1.4 & 2 \end{pmatrix}\right)$.

In applying the smoothed bootstrap (SB) and the NPI, we need to estimate the densities of q_i (SB of DGP1 and DGP3), $\widehat{\varepsilon}_i$ (NPI of DGP1, DGP3, and DGP4), $(x_i, q_i)'$ (SB of DGP4), and $(q_i, \widehat{\varepsilon}_i)'$ (SB and NPI of DGP2). For this purpose, we use the Mablab functions kde.m and kde2d.m provided by Zdravko Botev. In these functions, the bandwidths are selected adaptively; see Botev et al. (2010). In the smoothed bootstrap, we need to simulate random samples from the kernel density estimators; such a procedure can be found in Section 6.4.1 of Silverman (1986). For each of these DGPs, we consider 1000 random samples of size $n = 50, 200, 500$. For each sample, we take 1000 SB replicates to approximate the bootstrap distribution. In the NPI, we use the slicesample procedure in Matlab to draw a Markov chain from the posterior; we use the MLSE as the starting value, and draw 1000 samples from the posterior after discarding the first 200 "burn-in" draws.

The simulation results of the SB and the NPI are summarized in Table 3. In DGP1, there are two rows associated with SB-LLSE (ET). The second row is copied from Seijo and Sen (2011) where the rule-of-thumb (ROT) bandwidth of Silverman (1986) is used. All other rows are based on the adaptive bandwidth

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

TABLE 3. Performance of SB and NPI: coverage and average length of the nominal 95% confidence intervals for γ (based on 1000 repetitions)

$n \rightarrow$	50		200		500	
CIs↓ Cov and Leng($\times 10^{-1}$) $\rightarrow$	Coverage	Length	Coverage	Length	Coverage	Length
DGP1						
SB-LLSE (ET)	0.956	7.321	0.950	1.609	0.953	0.625
	0.94	9.4	0.95	1.9	0.95	0.7
SB-LLSE (S)	0.961	7.753	0.950	1.700	0.957	0.661
SB-MLSE (ET)	0.955	6.862	0.953	1.510	0.954	0.586
SB-MLSE (S)	0.957	6.834	0.953	1.506	0.955	0.584
NPI	0.942	3.650	0.959	0.874	0.948	0.340
DGP2						
SB-LLSE (ET)	0.985	11.791	0.986	0.994	0.977	0.280
SB-LLSE (S)	0.989	14.772	0.975	1.170	0.968	0.445
SB-MLSE (ET)	0.989	11.264	0.980	0.841	0.965	0.243
SB-MLSE (S)	0.999	13.376	0.980	0.684	0.963	0.242
NPI	0.949	1.471	0.951	0.329	0.948	0.128
DGP3						
SB-LLSE (ET)	0.885	15.048	0.974	1.926	0.952	0.640
SB-LLSE (S)	0.966	18.795	0.980	1.962	0.956	0.678
SB-MLSE (ET)	0.897	14.006	0.975	1.788	0.953	0.603
SB-MLSE (S)	0.958	16.502	0.972	1.711	0.956	0.599
NPI	0.840	6.323	0.944	0.925	0.951	0.347
DGP4						
SB-LLSE (ET)	0.948	6.116	0.944	1.487	0.951	0.595
SB-LLSE (S)	0.955	6.647	0.956	1.591	0.948	0.633
SB-MLSE (ET)	0.945	5.627	0.947	1.381	0.947	0.557
SB-MLSE (S)	0.939	5.602	0.949	1.373	0.952	0.554
NPI	0.938	3.269	0.942	0.762	0.945	0.303

Note 1: In DGP1, the second row associated with SB-LLSE (ET) contains the simulation results in Seijo and Sen (2011) and the first row contains our results.
Note 2: ET for equal-tailed CIs and S for symmetric CIs.

of Botev et al. (2010). From Table 3, a few results of interest are summarized as follows. First, the NPI dominates the SB interval when the sample size is reasonably large.[15] In all cases except DGP3 with $n = 50$, the NPI coverage is close to the target 95%, and the length is about one half of the SB interval. In DGP1 and DGP4, the coverage of the SB interval is perfect, but for the equal-tailed intervals of DGP3 with $n = 50$, there is an undercover problem, and in DGP3 with $n = 200$ and DGP2, there is an obvious overcovering. Second, the correlation between q and ε provides information about γ_0, since most CIs in DGP2 are shorter

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

than those in DGP1. Comparing the CIs in DGP1 and DGP4 suggests that extra covariates also provide some information, although not as much as in DGP2. Also note that while the asymptotic distribution in DGP3 is the same as in DGP1, the finite-sample distribution seems less efficient. Third, the SB intervals based on the MLSE are better than those based on the LLSE: similar coverage but shorter length. This is due to the fact that the MLSE is more efficient than the LLSE in our setups. Also, since the distribution of the MLSE is more symmetric than that of the LLSE, the performance of the equal-tailed SB intervals based on the LLSE is better than that of the symmetric intervals, while for the SB intervals based on the MLSE, the converse statement is more suitable. Fourth, by comparing the two SB equal-tailed intervals based on the LLSE in DGP1, we can see that the adaptive selection of bandwidth can shorten the CIs without affecting their coverage. Fifth, except the case $n = 50$ in DGP2 and DGP3, the length of the CIs matches the prediction from the asymptotic arguments: the convergence rate of the γ estimators is n. For example, the length when $n = 50$ is about four times of when $n = 200$, which in turn is about 2.5 times of when $n = 500$. To avoid the curse of dimensionality, Section 7 of Seijo and Sen (2011) suggests to simulate only from the marginal density estimate of q in the smoothed bootstrap, but if q is not independent of ε or is correlated with other covariates, there is the misspecification problem. The simulation studies on DGP2 and DGP4 (where the misspecification problem is present) in the supplementary materials show that the SB procedure is not reliable under misspecification.

A rough picture from these simulation results is that the SB works well in coverage but not as well in length. The length of the CIs is very important in threshold regression, since it indicates how much data cannot be classified into the two regimes. In practice, some CIs are too long to be useful as most data are covered by the CIs; see the application in Yu (2008) for a concrete example. Based on the simulations, our suggestions are: (a) use the NPI to construct a CI for the threshold point and (b) if the smoothed bootstrap based on the LSE is used, the MLSE and adaptive bandwidths are preferable to the LLSE and the ROT bandwidths.

5. CONCLUSION

In this paper, we develop a general procedure to check the validity of the bootstrap in M-estimation. Based on this procedure, we show that the nonparametric bootstrap is inconsistent for inference on the threshold point in discontinuous threshold regression. It is found that its conditional weak limit does not even exist. By comparing with two other boundaries in the literature, we show that it is because the threshold point is a boundary of the covariate that its bootstrap inference is so different. We also summarize the remedies of the bootstrap failure in the literature. Especially, we show that the smoothed bootstrap is valid in a more practical environment than that in Seijo and Sen (2011). Moreover, our simulation

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

studies suggest that the smoothed bootstrap interval is outperformed by another confidence interval—the nonparametric posterior interval.

NOTES

1. See also, e.g., Horowitz (2001, Ch. 52), Bickel et al. (1997), and Beran (1997).

2. Another example is Romano and Shaikh (2008) which is intimately connected with Andrews (2000).

3. See also Li and Ling (2012) for an alternative simulator in threshold autoregressive (TAR) models.

4. See Chapter 9 of Politis et al. (1999) and Remark 1 in Section 4 of Lee and Pun (2006) for a summary of m selection. However, it seems that none of the existing methods are universally satisfactory.

5. See page 1961 of Sen et al. (2010) for a rigorous statement about the existence of $\overline{H}^*$. $\overline{H}^*$ can be treated as the asymptotic distribution in an *average bootstrap* procedure: first running the bootstrap for each sample path, and then averaging the bootstrap distribution for all sample paths.

6. Abadie and Imbens (2008) use a similar procedure to prove the invalidity of the bootstrap for inference of the matching estimator, although it is not an M-estimator. For regular parameters, such an average bootstrap procedure has never been considered, since $H^*(\omega)$ is the same for almost all ω and $\overline{H}^* \overset{d}{=} H$ automatically holds.

7. Note that $H^*(\omega)$ cannot be thought as the weak limit of $H_n^*(\omega)$ if this weak limit does not exist at all.

8. There are of course other features. For example, u is unrestricted or the true parameter is in the interior of the parameter space. In the example of Andrews (2000), although the localized process takes the form of (6), u is restricted to be positive. This restriction invalidates the bootstrap; see the analysis in Section 3.3 for more details.

9. Note that $\widehat{Z}^*$ in Figure 1 of Abrevaya and Huang (2005) is the unconditional distribution of $\widehat{Z}^*$ rather than the conditional distribution given ω. Actually, as mentioned at the end of Sen et al. (2010), their simulation studies show that $\widehat{Z}^*$ is highly correlated with Z^*, which indicates that $\widehat{Z}^*(\omega)$ may not exist. A corollary is that the corrected confidence intervals based on Table 1 of Abrevaya and Huang (2005) are not reliable. The problem in their proof is that when applying their Theorem 3 to prove the conditional version of $\widehat{Z}^*$ (their (16)), they have only checked conditions (ii) and (iii) without checking condition (i). However, as implied in the proof of Theorem 2 in Appendix A, condition (i) would fail in their case.

10. See Kac (1949) and Klaassen and Wellner (1992) for an introduction of Poissonization.

11. From the proof in Appendix B, $\left\{N_{i-}^*, N_{i+}^*\right\}_{i\geq 1}$ is just $\left\{M_{N_n,i}\right\}_{i\geq 1}$.

12. Such a difference in information usage implies another difference between the bootstrap for β and γ: for β, the exchangeably weighted bootstraps discussed in Praestgaard and Wellner (1993) are first-order equivalent, but they are not equivalent for γ. For example, if the Bayesian bootstrap of Rubin (1981) is used, then $\sqrt{n}\left(\widehat{\beta}^* - \widehat{\beta}\right)$ has the same conditional weak limit as in Theorem 2, but the distribution of $\left\{N_{i-}^*, N_{i+}^*\right\}_{i\geq 1}$ in $D^*(v)$ changes to a mixture Poisson distribution with the mixture measure following the standard exponential.

13. As mentioned in the introduction, the nonparametric bootstrap is also valid in this framework. But it is not a practical choice as there is no way to tell a real dataset following the framework of Hansen (2000) or that used in this paper. Note also that Theorem 3 in Hansen (2000) guarantees that the CI there is at least conservative in a dominating case of our framework.

14. The only exception is the structural change model in time series where q is the time index and follows the uniform distribution on $[0, 1]$. But we should definitely take into account the dependence structure in this case which complicates the bootstrap validity proof.

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

15. Any judgement must be based on some criterion. In CI construction, a natural loss function is

$$l_b(\gamma, a) = \mathbf{1}(\gamma \notin a) + b \cdot leb(a).$$

where a is a measurable set in $\mathbb{R}$, and $leb(a)$ is the Lebesgue measure of a; that is, the loss function gives some punishment on the uncovering of γ and also on the length of the CI. Unless one puts tremendous losses on uncovering, the NPI should be preferable to the SB interval.

REFERENCES

Abadie, A. & G.W. Imbens (2008) On the failure of the bootstrap for matching estimators. *Econometrica* 76, 1537–1557.

Abrevaya, J. & J. Huang (2005) On the bootstrap of the maximum score estimator. *Econometrica* 73, 1175–1204.

Aldous, D.J. (1978) Stopping times and tightness. *The Annals of Probability* 6, 335–340.

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.

Arcones, M.A. & E. Giné (1992) On the bootstrap of M-estimators and other statistical functionals. In R. LePage & L. Billard (eds.), *Exploring the Limits of Bootstrap*, pp. 13–47. Wiley.

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. & D.A. Freedman (1981) Some asymptotic theory for the bootstrap. *The Annals of Statistics* 9, 1196–1217.

Billingsley, P. (1995) *Probability and Measure*. Wiley.

Bose, A. & S. Chatterjee (2001) Generalised bootstrap in non-regular M-estimation problems. *Statistics and Probability Letters* 55, 319–328.

Botev, Z.I., et al. (2010) Kernel density estimation via diffusion. *The Annals of Statistics* 38, 2916–2957.

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.

Cheng, G. & J.Z. Huang (2010) Bootstrap consistency for general semiparametric M-estimation. *The Annals of Statistics* 38, 2884–2915.

Chernozhukov, V. & H. Hong (2004) Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica* 72, 1445–1480.

Delgado, M., et al. (2001) Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator. *Economics Letters* 73, 241–250.

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.

Gijbels, I., et al. (2004) Interval and band estimation for curves with jumps. *Journal of Applied Probability (Special Issue "Stochastic Methods and Their Applications", in honor of Chris Heyde)* 41A, 65–79.

Gonzalo, J. & M. Wolf (2005) Subsampling inference in threshold autoregressive models. *Journal of Econometrics* 127, 201–224.

Hall, P., et al. (1989) On smoothing and the bootstrap. *The Annals of Statistics* 17, 692–704.

Hansen, B.E. (2000) Sample splitting and threshold estimation. *Econometrica* 68, 575–603.

Hansen, B.E. (2008) Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726–748.

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

Hansen, B.E. (2011) Threshold autoregression in economics. *Statistics and Its Interface* 4, 123–127.
Hirano, K. & J.R. Porter (2003) Asymptotic efficiency in parametric structural models with parameter-dependent support. *Econometrica* 71, 1307–1338.
Horowitz, J.L. (2001) The bootstrap. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, pp. 3159–3228. Elsevier Science B.V.
Hušková, M. & C. Kirch (2008) Bootstrapping confidence intervals for the change-point of time series. *Journal of Time Series Analysis* 29, 947–972.
Huang, J. & 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. & D. Pollard (1990) Cube root asymptotics. *The Annals of Statistics* 18, 191–219.
Klaassen, C.A. & J.A. Wellner (1992) Kac empirical processes and the bootstrap. In M. Hahn & J. Kuelbs (eds.), *Proceedings of the Eighth International Conference on Probability in Banach Spaces*, pp. 411–429. Birkhäuser.
Kosorok, M.R. (2008) Bootstrapping in Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research; Festschrift in Honor of Professor Pranab K. Sen, IMS Collections*, vol. 1, pp. 282–292. IMS.
Kosorok, M.R. & R. Song (2007) Inference under right censoring for transformation models with a change-point based on a covariate threshold. *The Annals of Statistics* 35, 957–989.
Lee, S.M.S & M.C. Pun (2006) On m out of n bootstrapping for nonstandard M-estimation with nuisance parameters. *Journal of the American Statistical Association* 101, 1185–1197.
Lee, S. & 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.
Léger, C. & B. MacGibbon (2006) On the bootstrap in cube root asymptotics. *The Canadian Journal of Statistics* 34, 29–44.
Li, D. & S.Q. Ling (2012) On the least squares estimation of multiple-regime threshold autoregressive models. *Journal of Econometrics* 167, 240–253.
Pakes, A. & D. Pollard (1989) Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027–1057.
Praestgaard, J. & J.A. Wellner (1993) Exchangeably weighted bootstraps of the general empirical process. *The Annals of Probability* 21, 2053–2086.
Politis, D., et al. (1999) *Subsampling*. Springer-Verlag.
Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer-Verlag.
Romano, J.P. & A.M. Shaikh (2008) Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference* 138, 2786–2807.
Rubin, D.B. (1981) The Bayesian bootstrap. *The Annals of Statistics* 9, 130–134.
Seijo, E. & B. Sen (2011) Change-point in stochastic design regression and the bootstrap. *The Annals of Statistics* 39, 1580–1607.
Sen, B., et al. (2010) Inconsistency of bootstrap: The Grenander estimator. *The Annals of Statistics* 38, 1953–1977.
Seo, M.H. & O. Linton (2007) A smoothed least squares estimator for threshold regression models. *Journal of Econometrics* 141, 704–735.
Shao, J. & D. Tu (1995) *The Jackknife and Bootstrap*. Springer.
Silverman, B.W. (1978) Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics* 6, 177–184.
Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
Silverman, B.W. & G.A. Young (1987) The bootstrap: To smooth or not to smooth? *Biometrika* 74, 469–479.
Tong, H. (1978) On a threshold model. In C.H. Chen (ed.), *Pattern Recognition and Signal Processing*, pp. 575–586. Sijthoff & Noordhoff.
Tong, H. (1983) *Threshold Models in Nonlinear Time Series Analysis, Lecture Notes in Statistics*. Springer-Verlag.

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

Tong, H. (1990) *Non-linear Time Series: A Dynamical System Approach*. Clarendon Press.
Tong, H. (2011) Threshold models in time series analysis – 30 years on. *Statistics and Its Interface* 4, 107–118.
Van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.
Wellner, J.A. & Y. Zhan (1996) Bootstrapping Z-Estimators. Unpublished manuscript, Department of Statistics, University of Washington.
Yu, P. (2008) Adaptive Estimation of the Threshold Point in Threshold Regression. Unpublished manuscript, Department of Economics, University of Auckland.
Yu, P. (2012) Likelihood estimation and inference in threshold regression. *Journal of Econometrics* 167, 274–294.
Yu, P. & Y. Zhao (2009) Asymptotics for Threshold Regression Under General Conditions. Unpublished manuscript, Department of Economics, University of Auckland.

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 from line to line. P_n, P_n^*, P^*, and P_r are understood in the nonparametric bootstrap context.

$$\theta_\ell = \left(\beta_\ell', \sigma_\ell\right)'. \; m\left(w|\theta\right) = \left(y - x'\beta_1 \mathbf{1}(q \le \gamma) - x'\beta_2 \mathbf{1}(q > \gamma)\right)^2.$$

$$M_n\left(\theta\right) = P_n\left(m\left(\cdot|\theta\right)\right), \; M\left(\theta\right) = P\left(m\left(\cdot|\theta\right)\right), \; \mathbb{G}_n m = \sqrt{n}\left(M_n - M\right).$$

$$\begin{aligned}\overline{z}_1\left(w|\theta_2, \widetilde{\theta}_1\right) &= \left(\widetilde{\beta}_1 - \beta_2\right)' xx' \left(\widetilde{\beta}_1 - \beta_2\right) \\ &\quad + 2\widetilde{\sigma}_1 \left(\widetilde{\beta}_1 - \beta_2\right) x\varepsilon, \quad \text{so } \overline{z}_{1i} = \overline{z}_1\left(w_i|\theta_{2,0}, \theta_{1,0}\right).\end{aligned}$$

$$\begin{aligned}\overline{z}_2\left(w|\theta_1, \widetilde{\theta}_2\right) &= \left(\widetilde{\beta}_2 - \beta_1\right)' xx' \left(\widetilde{\beta}_2 - \beta_1\right) \\ &\quad + 2\widetilde{\sigma}_2 \left(\widetilde{\beta}_2 - \beta_1\right) x\varepsilon, \quad \text{so } \overline{z}_{2i} = \overline{z}_2\left(w_i|\theta_{1,0}, \theta_{2,0}\right).\end{aligned}$$

The following formulas are used repetitively in the following analysis:

$$\begin{aligned}m\left(w|\theta\right) &= \left(x'\left(\beta_{1,0} - \beta_1\right) + \sigma_{1,0}\varepsilon\right)^2 \mathbf{1}(q_i \le \gamma \wedge \gamma_0) + \left(x'\left(\beta_{2,0} - \beta_2\right) + \sigma_{2,0}\varepsilon\right)^2 \\ &\quad \times \mathbf{1}(q_i > \gamma \vee \gamma_0) + \left(x'\left(\beta_{1,0} - \beta_2\right) + \sigma_{1,0}\varepsilon\right)^2 \mathbf{1}(\gamma \wedge \gamma_0 < q_i \le \gamma_0) \\ &\quad + \left(x'\left(\beta_{2,0} - \beta_1\right) + \sigma_{2,0}\varepsilon\right)^2 \mathbf{1}(\gamma_0 < q \le \gamma \vee \gamma_0),\end{aligned}$$

so

$$\begin{aligned}m\left(w|\theta\right) - m\left(w|\theta_0\right) &= \left[\left(\beta_{1,0} - \beta_1\right)' xx' \left(\beta_{1,0} - \beta_1\right) + 2\sigma_{1,0}\left(\beta_{1,0} - \beta_1\right)' x\varepsilon\right] \mathbf{1}(q \le \gamma \wedge \gamma_0) \\ &\quad + \left[\left(\beta_{2,0} - \beta_2\right)' xx' \left(\beta_{2,0} - \beta_2\right) + 2\sigma_{2,0}\left(\beta_{2,0} - \beta_2\right)' x\varepsilon\right] \mathbf{1}(q > \gamma \vee \gamma_0) \\ &\quad + \overline{z}_1\left(w|\theta_2, \theta_{1,0}\right) \mathbf{1}(\gamma \wedge \gamma_0 < q \le \gamma_0) + \overline{z}_2\left(w|\theta_1, \theta_{2,0}\right) \mathbf{1}(\gamma_0 < q \le \gamma \vee \gamma_0) \\ &\equiv T\left(w|\theta_1, \theta_{1,0}\right) \mathbf{1}(q \le \gamma \wedge \gamma_0) + T\left(w|\theta_2, \theta_{2,0}\right) \mathbf{1}(q > \gamma \vee \gamma_0) \\ &\quad + \overline{z}_1\left(w|\theta_2, \theta_{1,0}\right) \mathbf{1}(\gamma \wedge \gamma_0 < q \le \gamma_0) + \overline{z}_2\left(w|\theta_1, \theta_{2,0}\right) \mathbf{1}(\gamma_0 < q \le \gamma \vee \gamma_0) \\ &\equiv A\left(w|\theta\right) + B\left(w|\theta\right) + C\left(w|\theta\right) + D\left(w|\theta\right).\end{aligned}$$

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

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

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

$$T_{1i} = \frac{1}{\sqrt{n}} x_i \varepsilon_i \mathbf{1}(q_i \le \gamma_0) \equiv \frac{1}{\sqrt{n}} S_{1i}, \qquad T_{2i} = \frac{1}{\sqrt{n}} x_i \varepsilon_i \mathbf{1}(q_i > \gamma_0) \equiv \frac{1}{\sqrt{n}} S_{2i},$$
$$T_{3i} = \bar{z}_{1i} \mathbf{1}\left(\gamma_0 + \frac{v_1}{n} < q_i \le \gamma_0\right), \qquad T_{4i} = \bar{z}_{2i} \mathbf{1}\left(\gamma_0 < q_i \le \gamma_0 + \frac{v_2}{n}\right),$$

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

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

it follows

$$\begin{aligned}
&P_r\left(\exp\left\{\sqrt{-1}\begin{bmatrix} s'_{11} T_{1i} + s'_{12} T_{2i} + s'_{21}\left(M_{N_n,i} - 1\right) T_{1i} + s'_{22}\left(M_{N_n,i} - 1\right) 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) \\
&\quad = P_r\left(\exp\left\{\sqrt{-1} s' T_i^R / \sqrt{n}\right\}\right) + \frac{v_1}{n} f_q(\gamma_0) P_r \\
&\qquad \times \left(\exp\left\{\sqrt{-1} s' T_i^R / \sqrt{n}\right\}\left[\exp\left\{\sqrt{-1} t_{11} \bar{z}_{1i}\right\} - 1\right] \middle| q_i = \gamma_0 - \right) \\
&\qquad + \frac{v_2}{n} f_q(\gamma_0) P_r\left(\exp\left\{\sqrt{-1} s' T_i^R / \sqrt{n}\right\}\left[\exp\left\{\sqrt{-1} t_{12} \bar{z}_{2i}\right\} - 1\right] \middle| q_i = \gamma_0 + \right) \\
&\qquad + \frac{v_1}{n} f_q(\gamma_0) P_r\left(\exp\left\{\sqrt{-1} s' T_i^R / \sqrt{n}\right\}\left[\exp\left\{\sqrt{-1} t_{21} M_{N_n,i} \bar{z}_{1i}\right\} - 1\right] \middle| \right. \\
&\qquad\qquad \left. \times q_i = \gamma_0 - \right) \\
&\qquad + \frac{v_2}{n} f_q(\gamma_0) P_r\left(\exp\left\{\sqrt{-1} s' T_i^R / \sqrt{n}\right\}\left[\exp\left\{\sqrt{-1} t_{22} M_{N_n,i} \bar{z}_{2i}\right\} - 1\right] \middle| \right. \\
&\qquad\qquad \left. \times q_i = \gamma_0 + \right) + o\left(\frac{1}{n}\right) \\
&\quad = 1 + \frac{1}{n}\left[-\frac{1}{2} s' \mathcal{J} s + f_q(\gamma_0) v_1 \left(P_r\left[\left\{\exp\left\{\sqrt{-1} t_1 \bar{z}_{1i}\right\}\right\} \middle| q_i = \gamma_0 - \right] - 1\right)\right. \\
&\qquad + f_q(\gamma_0) v_2 \left(P_r\left[\left\{\exp\left\{\sqrt{-1} t_2 \bar{z}_{2i}\right\}\right\} \middle| q_i = \gamma_0 + \right] - 1\right) \\
&\qquad + f_q(\gamma_0) v_1 \left(P_r\left[\left\{\exp\left\{\sqrt{-1} t_1 M_{N_n,i} \bar{z}_{1i}\right\}\right\} \middle| q_i = \gamma_0 - \right] - 1\right) \\
&\qquad \left. + f_q(\gamma_0) v_2 \left(P_r\left[\left\{\exp\left\{\sqrt{-1} t_2 M_{N_n,i} \bar{z}_{2i}\right\}\right\} \middle| q_i = \gamma_0 + \right] - 1\right)\right] + o\left(\frac{1}{n}\right),
\end{aligned}$$

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

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

$$\mathcal{J} = P_r\left(T_i^R T_i^{R\prime}\right) = \operatorname{diag}\left\{E\left[xx'\varepsilon^2 \mathbf{1}(q \le \gamma_0)\right], E\left[xx'\varepsilon^2 \mathbf{1}(q > \gamma_0)\right], \right.$$
$$\left. \times \, E\left[xx'\varepsilon^2 \mathbf{1}(q \le \gamma_0)\right], E\left[xx'\varepsilon^2 \mathbf{1}(q > \gamma_0)\right]\right\}.$$

So

$$P_r\left(\exp\left\{\sqrt{-1}\begin{bmatrix} s_{11}\sum_{i=1}^n T_{1i} + s_{12}\sum_{i=1}^n T_{2i} + s_{21}\sum_{i=1}^n \left(M_{N_n,i} - 1\right) T_{1i} \\ + s_{22}\sum_{i=1}^n \left(M_{N_n,i} - 1\right) T_{2i} \\ + t_{11}\sum_{i=1}^n T_{3i} + t_{12}\sum_{i=1}^n T_{4i} + t_{21}\sum_{i=1}^n M_{N_n,i} T_{3i} \\ + t_{22}\sum_{i=1}^n M_{N_n,i} T_{4i} \end{bmatrix}\right\}\right)$$
$$= \prod_{i=1}^n P_r\left(\exp\left\{\sqrt{-1}\left[s' T_i^R/\sqrt{n} + t_{11} T_{3i} + t_{12} T_{4i} + t_{21} M_{N_n,i} T_{3i} + t_{22} M_{N_n,i} T_{4i}\right]\right\}\right)$$
$$\to \exp\left\{-\frac{1}{2} s' \mathcal{J} s + f_q(\gamma_0) \upsilon_1 \left(P_r\left[\left\{\exp\left\{\sqrt{-1} t_{11} \bar{z}_{1i}\right\}\right\} \middle| q_i = \gamma_0 -\right] - 1\right)\right.$$
$$+ f_q(\gamma_0) \upsilon_2 \left(P_r\left[\left\{\exp\left\{\sqrt{-1} t_{12} \bar{z}_{2i}\right\}\right\} \middle| q_i = \gamma_0 +\right] - 1\right)$$
$$+ f_q(\gamma_0) \upsilon_1 \left(P_r\left[\left\{\exp\left\{\sqrt{-1} t_{21} M_{N_n,i} \bar{z}_{1i}\right\}\right\} \middle| q_i = \gamma_0 -\right] - 1\right)$$
$$\left. + f_q(\gamma_0) \upsilon_2 \left(P_r\left[\left\{\exp\left\{\sqrt{-1} t_{22} M_{N_n,i} \bar{z}_{2i}\right\}\right\} \middle| q_i = \gamma_0 +\right] - 1\right)\right\},$$

which matches the characteristic function of $\left(W(u), W^*(u), D(\upsilon), D^*(\upsilon)\right)$, and the result of interest follows. Here, note that $\left\{N_{i-}^*, N_{i+}^*\right\}_{i \ge 1}$ is just $\left\{M_{N_n,i}\right\}_{i \ge 1}$. Under Assumption D7, conditioning on $q_i = \gamma_0-$ and $q_i = \gamma_0+$ can be replaced by conditioning on $q_i = \gamma_0$. Also, the conditional distribution of $M_{N_n,i} \bar{z}_{\ell i}$ given $q_i = \gamma_0$ is the same as $M_{N_n,i} z_{\ell i}$, since $\left\{M_{N_n,i}\right\}_{i \ge 1}$ is independent of the original 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 $D_n(\upsilon)$ and $D_n^*(\upsilon)$. 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 prove the result only for $\upsilon > 0$. Suppose $0 < \upsilon_1 < \upsilon_2$ are stopping times in a compact set, then for any $\epsilon > 0$,

$$P\left(\sup_{|\upsilon_2 - \upsilon_1| < \delta} |D_n(\upsilon_2) - D_n(\upsilon_1)| > \epsilon\right)$$
$$\overset{(1)}{\le} P\left(\sum_{i=1}^n |\bar{z}_{2i}| \cdot \sup_{|\upsilon_2 - \upsilon_1| < \delta} \mathbf{1}\left(\gamma_0 + \frac{\upsilon_1}{n} < q_i \le \gamma_0 + \frac{\upsilon_2}{n}\right) > \epsilon\right)$$
$$\overset{(2)}{\le} \sum_{i=1}^n P\left[|\bar{z}_{2i}| \sup_{|\upsilon_2 - \upsilon_1| < \delta} \mathbf{1}\left(\gamma_0 + \frac{\upsilon_1}{n} < q_i \le \gamma_0 + \frac{\upsilon_2}{n}\right)\right] \Big/ \epsilon \overset{(3)}{\le} \frac{C\delta}{\epsilon},$$

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

where (1) is obvious, (2) is from Markov's inequality, and C in (3) can take $\overline{f}_q \sup_{\gamma_0<\gamma\le\gamma_0+\epsilon} E\left[|\overline{z}_{2i}||q_i=\gamma\right]<\infty$ from Assumptions D4 and D6. Similarly,

$$P_r\left(\sup_{|v_2-v_1|<\delta}\left|D_n^*(v_2)-D_n^*(v_1)\right|>\epsilon\right)$$
$$\le P_r\left(\sum_{i=1}^n\left|M_{N_n,i}\overline{z}_{2i}\right|\cdot\sup_{|v_2-v_1|<\delta}\mathbf{1}\left(\gamma_0+\frac{v_1}{n}<q_i\le\gamma_0+\frac{v_2}{n}\right)>\epsilon\right)$$
$$\le\sum_{i=1}^n P_r\left[\left|M_{N_n,i}\overline{z}_{2i}\right|\sup_{|v_2-v_1|<\delta}\mathbf{1}\left(\gamma_0+\frac{v_1}{n}<q_i\le\gamma_0+\frac{v_2}{n}\right)\right]/\epsilon\le\frac{C\delta}{\epsilon},$$

where C can take $\overline{f}_q \sup_{\gamma_0<\gamma\le\gamma_0+\epsilon} P_r\left[\left|M_{N_n,i}\overline{z}_{2i}\right||q_i=\gamma\right]=\overline{f}_q \sup_{\gamma_0<\gamma\le\gamma_0+\epsilon} E\left[|\overline{z}_{2i}||q_i=\gamma\right]<\infty$. ■

Proof of Theorem 2. First, we consider the weak convergence under P_r by applying Lemma 5. Here,

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

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

$$\begin{pmatrix}U_n(u)+D_n(v)\\U_n^*(u)+D_n^*(v)\end{pmatrix}\rightsquigarrow u_1'E\left[x_ix_i'\mathbf{1}(q_i\le\gamma_0)\right]u_1+u_2'E\left[x_ix_i'\mathbf{1}(q_i>\gamma_0)\right]$$
$$\times u_2-2u_1'W_1-2u_2'W_2+\begin{pmatrix}D(v)\\-2u_1'W_1^*-2u_2'W_2^*+D^*(v)\end{pmatrix}\equiv\begin{pmatrix}U(u)+D(v)\\U^*(u)+D^*(v)\end{pmatrix},$$

and from Lemma 4,

$$\begin{pmatrix}s_n\\t_n\end{pmatrix}=\begin{pmatrix}\sqrt{n}\begin{pmatrix}\widehat{\beta}_1-\beta_{1,0}\\\widehat{\beta}_2-\beta_{2,0}\end{pmatrix}\\n\left(\widehat{\gamma}-\gamma_0\right)\end{pmatrix}=O_P(1),\quad\begin{pmatrix}s_n^*\\t_n^*\end{pmatrix}=\begin{pmatrix}\sqrt{n}\begin{pmatrix}\widehat{\beta}_1^*-\beta_{1,0}\\\widehat{\beta}_2^*-\beta_{2,0}\end{pmatrix}\\n\left(\widehat{\gamma}^*-\gamma_0\right)\end{pmatrix}=O_{P_r}(1).$$

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

$$\begin{pmatrix}\left(s_n',t_n\right)'\\\left(s_n^{*\prime},t_n^*\right)'\end{pmatrix}\rightsquigarrow\begin{pmatrix}\left(s',t\right)'\\\left(s^{*\prime},t^*\right)'\end{pmatrix},$$

where

$$s=\arg\min_{u_1,u_2}U(u)=\begin{pmatrix}E\left[x_ix_i'\mathbf{1}(q_i\le\gamma_0)\right]^{-1}W_1\\E\left[x_ix_i'\mathbf{1}(q_i>\gamma_0)\right]^{-1}W_2\end{pmatrix},\ t=\arg\min_v D(v),$$
$$s^*=\arg\min_{u_1,u_2}U^*(u)=\begin{pmatrix}E\left[x_ix_i'\mathbf{1}(q_i\le\gamma_0)\right]^{-1}\left(W_1+W_1^*\right)\\E\left[x_ix_i'\mathbf{1}(q_i>\gamma_0)\right]^{-1}\left(W_2+W_2^*\right)\end{pmatrix},\ t^*=\arg\min_v D^*(v).$$

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

By the continuous mapping theorem,

$$\begin{aligned}
&\left(s_n^{*\prime}, t_n^*\right)' - \left(s_n', t_n\right)' \rightsquigarrow \left(s^{*\prime}, t^*\right)' - \left(s', t\right)' \\
&= \begin{pmatrix} E\left[x_i x_i' \mathbf{1}\left(q_i \le \gamma_0\right)\right]^{-1}\left(W_1 + W_1^*\right) - E\left[x_i x_i' \mathbf{1}\left(q_i \le \gamma_0\right)\right]^{-1} W_1 \\ E\left[x_i x_i' \mathbf{1}\left(q_i > \gamma_0\right)\right]^{-1}\left(W_2 + W_2^*\right) - E\left[x_i x_i' \mathbf{1}\left(q_i > \gamma_0\right)\right]^{-1} W_2 \\ \arg\min_{\upsilon} D^*(\upsilon) - \arg\min_{\upsilon} D(\upsilon) \end{pmatrix} \\
&= \begin{pmatrix} E\left[x_i x_i' \mathbf{1}\left(q_i \le \gamma_0\right)\right]^{-1} W_1^* \\ E\left[x_i x_i' \mathbf{1}\left(q_i > \gamma_0\right)\right]^{-1} W_2^* \\ \arg\min_{\upsilon} D^*(\upsilon) - \arg\min_{\upsilon} D(\upsilon) \end{pmatrix}.
\end{aligned}$$

Lemma 5 cannot be applied with respect to the probability measure P^* in P probability. (ii) and (iii) are satisfied, but (i) is not. This remarkable result is first observed by Seijo and Sen (2011) for $\left(t_n^*, t_n\right)'$ in a simple threshold model; see their Lemma 4.2. But as they mentioned on page 1592, (i), (ii), and (iii) are only sufficient conditions to find the conditional weak limit of the bootstrap estimator. Nonexistence of the conditional weak limit of the localized process does not imply the nonexistence of the conditional weak limit of the bootstrap estimator. Here, we first show that even for $\left(s_n^{*\prime}, s_n'\right)'$, condition (i) fails. So the intuitive derivation of the asymptotic bootstrap distribution for regular parameters in Section 2.3 or on page 1182 of Abrevaya and Huang (2005) is only valid under P_r but not under P^* in P probability. An appropriate adaptation of the arguments here can be used to show that the result (16) in Theorem 2 of Abrevaya and Huang (2005) may not hold, since condition (i) in their Theorem 3 does not hold.

We apply Lemma 3.2 of Sen et al. (2010)[16] to show that condition (i) of Lemma 5 fails for $\left(s_n^{*\prime}, s_n'\right)'$ under P^* in P probability. It is enough to show that $U_n^*(u)$ does not have a conditional weak limit for a fixed u. In Sen et al. (2010)'s notations, $Y_n = W_n(u)$, and $Z_n = W_n^*(u)$. Let K_n and L_n be the conditional distribution functions of $Y_n + Z_n$ and Z_n given the original data, respectively. From the multiplier central limit theorem, $d(L_n, L) \xrightarrow{P} 0$, where L follows the same distribution as $W(u)$ so is nondegenerate. If $d(K_n, K) \xrightarrow{P} 0$ for some distribution function K, then Lemma 3.2 of Sen et al. (2010) claims that there must be a random variable Y for which $Y_n \xrightarrow{P} Y$. Since Y_n is invariant to permutations of the original data, from the Hewitt–Savage zero–one law, Y must be a constant P almost surely. But $Y_n \xrightarrow{d} W(u)$ which is not a constant. This contradiction implies that $d(K_n, K) \xrightarrow{P} 0$ cannot hold; in other words, $U_n^*(u)$ does not have a conditional weak limit. This argument can also serve as a proof that $\sqrt{n}\overline{X}_n^*$ in Section 3.3 does not have a conditional weak limit.

Although Lemma 5 cannot be applied, we can still show that the bootstrap estimator $\widehat{\beta}^*$ is consistent by a straightforward argument. From Lemma 1 and Lemma 2,

$$\sqrt{n}\left(\widehat{\beta}^* - \widehat{\beta}\right) = s_n^* - s_n = \begin{pmatrix} E\left[x_i x_i' \mathbf{1}\left(q_i \le \gamma_0\right)\right]^{-1} W_{1n}^* \\ E\left[x_i x_i' \mathbf{1}\left(q_i > \gamma_0\right)\right]^{-1} W_{2n}^* \end{pmatrix} + \Delta_n$$

where $\Delta_n = o_{P^*}(1)$ in P probability. By the multiplier central limit theorem, $\sqrt{n}\left(\widehat{\beta}^* - \widehat{\beta}\right) \overset{*}{\rightsquigarrow} Z_\beta^*$. Since the distribution of Z_β^* is invariant to the original data, Z_β^* is independent of Z_β.

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

Next, we give a direct proof that $n(\widehat{\gamma}^* - \widehat{\gamma})$ does not have a weak limit in P probability. The proof idea is inspired by the counter-example 2 (page 1210) of Bickel and Freedman (1981). Suppose $n(\widehat{\gamma}^* - \widehat{\gamma})$ has a conditional weak limit, then this conditional weak limit must be the same as $\arg\min_{\upsilon} D_n^*(\upsilon) - \arg\min_{\upsilon} D_n(\upsilon)$ by Lemma 1. To simplify notations, we still use t_n^* and t_n for the minimizers of $D_n^*(\upsilon)$ and $D_n(\upsilon)$, and $\widehat{\gamma}^*$ and $\widehat{\gamma}$ for the minimizers of $P_n^* m(w|\gamma)$ and $P_n m(w|\gamma)$; that is, $t_n^* - t_n = n\left(\widehat{\gamma}^* - \widehat{\gamma}\right)$. Conditional on ω, $\{q_i, \overline{z}_{1i}, \overline{z}_{2i}\}_{i=1}^n$ in $D_n(\cdot)$ and $D_n^*(\cdot)$ are fixed. So $\widehat{\gamma}^*$ must equal $q_{(j)}$ for some j and $\widehat{\gamma}$ must equal $q_{(k)}$ for some k, where $q_{(i)}$ is the ith order statistics of $\{q_i\}_{i=1}^n$, $i = 1, \cdots, n$, and the j, k in $\widehat{\gamma}^*$ and $\widehat{\gamma}$ may depend on ω. A key observation is that for almost every ω, $\lim_{n\to\infty} P^*\left(\widehat{\gamma}^* \neq \widehat{\gamma}\right) > 0$. This can be seen from the following arguments. If $\widehat{\gamma} = q_{(k)}$, then for whatever values the other $M_{N_n,i}$'s may take, $\widehat{\gamma}^* \neq \widehat{\gamma}$ as long as $M_{N_n,k-1} = 0$, since we are using the LLSE. As a result, $\lim_{n\to\infty} P^*\left(\widehat{\gamma}^* \neq \widehat{\gamma}\right) \geq P^*\left(M_{N_n,k-1} = 0\right) = e^{-1} > 0$. Given ω, if the limit conditional distribution of $n(\widehat{\gamma}^* - \widehat{\gamma})$ exists, then there must exist points masses $\left\{p_1^n, \cdots, p_{k-1}^n, p_{k+1}^n, \cdots, p_n^n\right\}$ such that $\lim_{n\to\infty} \sum_{j=1, j\neq k}^n p_j^n \mathbf{1}\left(n\left|q_{(j)} - q_{(k)}\right| \geq x\right)$ exists for almost every $x > 0$, where p_j^n, $j = 1, \cdots, k-1, k+1, \cdots, n$, is the probability of obtaining $\widehat{\gamma}^* = q_{(j)}$. Since $\overline{\lim}_{n\to\infty} n\left|q_{(j)} - q_{(k)}\right| = \infty$ and $\underline{\lim}_{n\to\infty} n\left|q_{(j)} - q_{(k)}\right| = 0$ P almost surely for any $j \neq k$, $\overline{\lim}_{n\to\infty} \sum_{j=1, j\neq k}^n p_j^n \mathbf{1}\left(n\left|q_{(j)} - q_{(k)}\right| \geq x\right) = \overline{\lim}_{n\to\infty} \sum_{j=1, j\neq k}^n p_j^n \geq e^{-1} > 0$ and $\underline{\lim}_{n\to\infty} \sum_{j=1, j\neq k}^n p_j^n \mathbf{1}\left(n\left|q_{(j)} - q_{(k)}\right| \geq x\right) = 0$ P almost surely. As a result, $\lim_{n\to\infty} \sum_{j=1, j\neq k}^n p_j^n \mathbf{1}\left(n\left|q_{(j)} - q_{(k)}\right| \geq x\right)$ does not converge P almost surely. This implies that $n(\widehat{\gamma}^* - \widehat{\gamma})$ cannot have a conditional weak limit. ■

APPENDIX B: Lemmas

LEMMA 1. *Under Assumption D, uniformly for h in any compact set of* $\mathbb{R}^{2k+1}$,

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

and

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

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

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

Proof. This proof is inspired by Klaassen and Wellner (1992) and Chapter 3.6 of Van der Vaart and Wellner (1996). First,

$$
\begin{aligned}
nP_n\left(m\left(\cdot\middle|\beta_0+\frac{u}{\sqrt{n}},\gamma_0+\frac{v}{n}\right)-m\left(\cdot|\beta_0,\gamma_0\right)\right) &= \sum_{i=1}^{n}\left(u_1'\frac{x_ix_i'}{n}u_1-u_1'\frac{2\sigma_{1,0}}{\sqrt{n}}x_i\varepsilon_i\right)\\
&\times\mathbf{1}\left(q_i\le\gamma_0\wedge\gamma_0+\frac{v}{n}\right)+\sum_{i=1}^{n}\left(u_2'\frac{x_ix_i'}{n}u_2-u_2'\frac{\sigma_{2,0}}{\sqrt{n}}x_i\varepsilon_i\right)\mathbf{1}(q_i>\gamma_0\vee\gamma_0+\frac{v}{n})\\
&+\sum_{i=1}^{n}\left[\left(\beta_{1,0}-\beta_{2,0}-\frac{u_2}{\sqrt{n}}\right)'x_ix_i'\left(\beta_{1,0}-\beta_{2,0}-\frac{u_2}{\sqrt{n}}\right)\right.\\
&\left.+2x_i'\left(\beta_{1,0}-\beta_{2,0}-\frac{u_2}{\sqrt{n}}\right)\sigma_{1,0}\varepsilon_i\right]\mathbf{1}\left(\gamma_0+\frac{v}{n}<q_i\le\gamma_0\right)\\
&+\sum_{i=1}^{n}\left[\left(\beta_{1,0}+\frac{u_1}{\sqrt{n}}-\beta_{2,0}\right)'x_ix_i'\left(\beta_{1,0}+\frac{u_1}{\sqrt{n}}-\beta_{2,0}\right)\right.\\
&\left.-2x_i'\left(\beta_{1,0}+\frac{u_1}{\sqrt{n}}-\beta_{2,0}\right)\sigma_{2,0}\varepsilon_i\right]\mathbf{1}\left(\gamma_0<q_i\le\gamma_0+\frac{v}{n}\right)\\
&= u_1'E\left[x_ix_i'\mathbf{1}\left(q_i\le\gamma_0\right)\right]u_1+u_2'E\left[x_ix_i'\mathbf{1}\left(q_i>\gamma_0\right)\right]u_2-W_n\left(u\right)+D_n\left(v\right)+o_P\left(1\right),
\end{aligned}
$$

where $o_P(1)$ is from Assumptions D4 and D5. Second,

$$
\begin{aligned}
nP_n^*\left(m\left(\cdot\middle|\beta_0+\frac{u}{\sqrt{n}},\gamma_0+\frac{v}{n}\right)-m\left(\cdot|\beta_0,\gamma_0\right)\right)&\\
&=\sum_{i=1}^{n}M_{ni}\left(u_1'\frac{x_ix_i'}{n}u_1-u_1'\frac{2\sigma_{1,0}}{\sqrt{n}}x_i\varepsilon_i\right)\mathbf{1}\left(q_i\le\gamma_0\wedge\gamma_0+\frac{v}{n}\right)\\
&+\sum_{i=1}^{n}M_{ni}\left(u_2'\frac{x_ix_i'}{n}u_2-u_2'\frac{\sigma_{2,0}}{\sqrt{n}}x_i\varepsilon_i\right)\mathbf{1}\left(q_i>\gamma_0\vee\gamma_0+\frac{v}{n}\right)\\
&+\sum_{i=1}^{n}M_{ni}\left[\left(\beta_{1,0}-\beta_{2,0}-\frac{u_2}{\sqrt{n}}\right)'x_ix_i'\left(\beta_{1,0}-\beta_{2,0}-\frac{u_2}{\sqrt{n}}\right)\right.\\
&\left.+2x_i'\left(\beta_{1,0}-\beta_{2,0}-\frac{u_2}{\sqrt{n}}\right)\sigma_{1,0}\varepsilon_i\right]\mathbf{1}\left(\gamma_0+\frac{v}{n}<q_i\le\gamma_0\right)\\
&+\sum_{i=1}^{n}M_{ni}\left[\left(\beta_{1,0}+\frac{u_1}{\sqrt{n}}-\beta_{2,0}\right)'x_ix_i'\left(\beta_{1,0}+\frac{u_1}{\sqrt{n}}-\beta_{2,0}\right)\right.\\
&\left.-2x_i'\left(\beta_{1,0}+\frac{u_1}{\sqrt{n}}-\beta_{2,0}\right)\sigma_{2,0}\varepsilon_i\right]\mathbf{1}\left(\gamma_0<q_i\le\gamma_0+\frac{v}{n}\right)\\
&\equiv T1+T2+T3+T4=\sum_{i=1}^{n}\left(u_1'\frac{x_ix_i'}{n}u_1-u_1'\frac{2\sigma_{1,0}}{\sqrt{n}}x_i\varepsilon_i\right)\mathbf{1}\left(q_i\le\gamma_0\right)\\
&+\sum_{i=1}^{n}\left(u_2'\frac{x_ix_i'}{n}u_2-u_2'\frac{\sigma_{2,0}}{\sqrt{n}}x_i\varepsilon_i\right)\mathbf{1}(q_i>\gamma_0)-W_n^*\left(u\right)+D_n^*\left(v\right)+o_{P_r}\left(1\right)\\
&=u_1'E\left[x_ix_i'\mathbf{1}\left(q_i\le\gamma_0\right)\right]u_1+u_2'E\left[x_ix_i'\mathbf{1}\left(q_i>\gamma_0\right)\right]u_2-W_n\left(u\right)-W_n^*\left(u\right)\\
&+D_n^*\left(v\right)+o_{P_r}\left(1\right),
\end{aligned}
$$

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

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

Note that $\left(M_{N_n,1},\cdots,M_{N_n,n}\right)=\sum_{j=1}^{N_n}\mathbf{1}_j\equiv\sum_{j=1}^{N_n}\left(1_{1j},\cdots,1_{nj}\right)$, where $\left(1_{1j},\cdots,1_{nj}\right)$ follows the multinomial distribution with parameters 1 and cell probabilities all equal to $\frac{1}{n}$. So conditional on $N_n=k$, $\left\{\left|M_{N_n,i}-M_{ni}\right|\right\}_{i=1}^n\overset{d}{=}\left\{\left|M_{|N_n-n|,i}\right|\right\}_{i=1}^n$. By the analysis in Theorem 3 of Klaassen and Wellner (1992) or Theorem 3.6.2 in Van der Vaart and Wellner (1996),

$$P^*\left(\max_{1\le i\le n}\left|M_{N_n,i}-M_{ni}\right|>2\right)=O\left(n^{-1/2}\right). \tag{10}$$

Define $I_n^j=\left\{i:\left|M_{N_n,i}-M_{ni}\right|\ge j\right\}$, $j=1,2,\cdots$, then $\#I_n^j\le|N_n-n|=O_{P^*}\left(\sqrt{n}\right)$. Thus,

$$\begin{aligned}
&\sum_{i=1}^n\left(M_{N_n,i}-M_{ni}\right)\left(u_1'\frac{x_ix_i'}{n}u_1-u_1'\frac{2\sigma_{1,0}}{\sqrt{n}}x_i\varepsilon_i\right)\mathbf{1}\left(q_i\le\gamma_0\wedge\gamma_0+\frac{\upsilon}{n}\right)\\
&\quad=\operatorname{sign}\left(N_n-n\right)\sum_{j=1}^2\left[\frac{\#I_n^j}{n}u_1'\left(\frac{1}{\#I_n^j}\sum_{i\in I_n^j}x_ix_i'\mathbf{1}\left(q_i\le\gamma_0\wedge\gamma_0+\frac{\upsilon}{n}\right)\right)u_1\right.\\
&\qquad\left.-\frac{\#I_n^j}{\sqrt{n}}2\sigma_{1,0}u_1'\frac{1}{\#I_n^j}\sum_{i\in I_n^j}x_i\varepsilon_i\mathbf{1}\left(q_i\le\gamma_0\wedge\gamma_0+\frac{\upsilon}{n}\right)\right]+o_{P_r}(1)\\
&\quad=o_{P_r}(1),
\end{aligned}$$

where the $o_{P_r}(1)$ in the first equality is from (10), and the last equality is from $\#I_n^j=O_{P^*}\left(\sqrt{n}\right)$ and a multiplier Glivenko-Cantelli theorem (see, e.g., Lemma 3.6.16 of Van der Vaart and Wellner, 1996). Now,

$$\sum_{i=1}^n M_{N_n,i}\left(u_1'\frac{x_ix_i'}{n}u_1-u_1'\frac{2\sigma_{1,0}}{\sqrt{n}}x_i\varepsilon_i\right)\mathbf{1}\left(\gamma_0\wedge\gamma_0+\frac{\upsilon}{n}<q_i\le\gamma_0\right)=o_{P_r}(1)$$

from Assumption D4. So

$$\begin{aligned}
T1&=\sum_{i=1}^n M_{N_n,i}\left(u_1'\frac{x_ix_i'}{n}u_1-u_1'\frac{2\sigma_{1,0}}{\sqrt{n}}x_i\varepsilon_i\right)\mathbf{1}\left(q_i\le\gamma_0\right)+o_{P_r}(1)\\
&=u_1'E\left[x_ix_i'\mathbf{1}\left(q_i\le\gamma_0\right)\right]u_1-u_1'\left(\frac{2\sigma_{1,0}}{\sqrt{n}}\sum_{i=1}^n M_{N_n,i}x_i\varepsilon_i\mathbf{1}\left(q_i\le\gamma_0\right)\right)+o_{P_r}(1).
\end{aligned}$$

Similarly, we could prove the result for $T2$.

As to $T3$ and $T4$, note that

$$\begin{aligned}
&\sum_{i=1}^n\left(M_{N_n,i}-M_{ni}\right)\left[\left(\beta_{1,0}-\beta_{2,0}-\frac{u_2}{\sqrt{n}}\right)'x_ix_i'\left(\beta_{1,0}-\beta_{2,0}-\frac{u_2}{\sqrt{n}}\right)\right.\\
&\qquad\left.+2x_i'\left(\beta_{1,0}-\beta_{2,0}-\frac{u_2}{\sqrt{n}}\right)\sigma_{1,0}\varepsilon_i\right]\mathbf{1}\left(\gamma_0+\frac{\upsilon}{n}<q_i\le\gamma_0\right)\\
&\quad=\operatorname{sign}\left(N_n-n\right)\sum_{j=1}^2\frac{1}{\#I_n^j}\sum_{i\in I_n^j}\left[\left(\beta_{1,0}-\beta_{2,0}-\frac{u_2}{\sqrt{n}}\right)'x_ix_i'\left(\beta_{1,0}-\beta_{2,0}-\frac{u_2}{\sqrt{n}}\right)\right.\\
&\qquad\left.+2x_i'\varepsilon_i\left(\beta_{1,0}-\beta_{2,0}-\frac{u_2}{\sqrt{n}}\right)\sigma_{1,0}\right]\left(\#I_n^j\cdot\mathbf{1}\left(\gamma_0+\frac{\upsilon}{n}<q_i\le\gamma_0\right)\right)=o_{P_r}(1),
\end{aligned}$$

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

where the last equality is from the fact that $\#I_n^j \cdot \mathbf{1}\left(\gamma_0 + \frac{v}{n} < q_i \le \gamma_0\right) = o_{P_r}(1)$ uniformly for $i \in I_n^j$. So it is easy to see that

$$\begin{aligned} T3+T4 &= \sum_{i=1}^{n} M_{N_n,i}\left[\left(\beta_{1,0}-\beta_{2,0}\right)' x_i x_i' \left(\beta_{1,0}-\beta_{2,0}\right) + 2x_i'\left(\beta_{1,0}-\beta_{2,0}\right)\sigma_{1,0}\varepsilon_i\right] \\ &\quad \times \mathbf{1}\left(\gamma_0 + \frac{v}{n} < q_i \le \gamma_0\right) + \sum_{i=1}^{n} M_{N_n,i}\left[\left(\beta_{1,0}-\beta_{2,0}\right)' x_i x_i' \left(\beta_{1,0}-\beta_{2,0}\right)\right. \\ &\quad \left. -2x_i'\left(\beta_{1,0}-\beta_{2,0}\right)\sigma_{2,0}\varepsilon_i\right]\mathbf{1}\left(\gamma_0 < q_i \le \gamma_0 + \frac{v}{n}\right) + o_{P_r}(1) \\ &= D_n^*(v) + o_{P_r}(1). \end{aligned}$$

■

The following lemma appears in Wellner and Zhan (1996) and is used in Abrevaya and Huang (2005); see also Lemma 3 of Cheng and Huang (2010). It states relationships among the probability measures P, P^*, and P_r in terms of $o_p(\cdot)$ and $O_p(\cdot)$. Given a real-valued function Δ_n defined on the product probability space $(\Omega \times \mathcal{T}, \mathcal{A} \times \mathcal{B}, P_r)$, we say that Δ_n is of an order $o_{P^*}(1)$ in P probability if for any $\epsilon, \delta > 0$,

$$P\left(P^*\left(|\Delta_n| > \epsilon\right) > \delta\right) \to 0 \text{ as } n \to \infty,$$

and that Δ_n is of an order $O_{P^*}(1)$ in P probability if for any $\delta > 0$, there exists a $0 < M < \infty$ such that

$$P\left(P^*\left(|\Delta_n| \ge M\right) > \delta\right) \to 0 \text{ as } n \to \infty.$$

LEMMA 2. *(i) If Δ_n is defined only on the probability space $(\Omega, \mathcal{A}, P)$ and $\Delta_n = o_P(1)$ $(O_P(1))$, then $\Delta_n = o_{P_r}(1)$ $\left(O_{P_r}(1)\right)$; (ii) If $\Delta_n = o_{P_r}(1)$ $\left(O_{P_r}(1)\right)$, then $\Delta_n = o_{P^*}(1)$ $(O_{P^*}(1))$ in P probability.*

LEMMA 3. *Under Assumption D, both $\widehat{\theta}$ and $\widehat{\theta}^*$ are consistent in P_r probability.*

Proof. The proof idea follows from Lemma A.5 of Hansen (2000), so omitted here. ■

LEMMA 4. *Under Assumption D, $\widehat{\gamma} = \gamma_0 + O_P\left(\frac{1}{n}\right)$, $\widehat{\gamma}^* = \gamma_0 + O_{P_r}\left(\frac{1}{n}\right)$, $\widehat{\beta} = \beta_0 + O_P\left(\frac{1}{\sqrt{n}}\right)$, and $\widehat{\beta}^* = \beta_0 + O_{P_r}\left(\frac{1}{\sqrt{n}}\right)$.*

Proof. We will first prove $\widehat{\gamma} = \gamma_0 + O_P\left(\frac{1}{n}\right)$. In this process, Corollary 3.2.6 of Van der Vaart and Wellner (1996) is used. We then prove $\widehat{\gamma}^* = \gamma_0 + O_{P_r}\left(\frac{1}{n}\right)$ by applying the proof idea of Theorem 3.2.5 in Van der Vaart and Wellner (1996), which is also used in Lemma 3 of Abrevaya and Huang (2005).

First, $M(\theta) - M(\theta_0) \ge Cd^2(\theta, \theta_0)$ with $d(\theta, \theta_0) = \|\beta - \beta_0\| + \sqrt{|\gamma - \gamma_0|}$ for θ in a neighborhood of θ_0.

$$\begin{aligned} &M(\theta) - M(\theta_0) \\ &\quad = E\left[T\left(w|\theta_1, \theta_{1,0}\right)\mathbf{1}(q \le \gamma \wedge \gamma_0)\right] + E\left[T\left(w|\theta_2, \theta_{2,0}\right)\mathbf{1}(q > \gamma \vee \gamma_0)\right] \\ &\qquad + E\left[\bar{z}_1\left(w|\theta_2, \theta_{1,0}\right)\mathbf{1}(\gamma \wedge \gamma_0 < q \le \gamma_0)\right] + E\left[\bar{z}_2\left(w|\theta_1, \theta_{2,0}\right)\mathbf{1}(\gamma_0 < q \le \gamma \vee \gamma_0)\right] \end{aligned}$$

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

$$
\begin{aligned}
&= \left(\beta_{1,0}-\beta_1\right)' E\left[xx'\mathbf{1}(q \le \gamma \wedge \gamma_0)\right]\left(\beta_{1,0}-\beta_1\right) + \left(\beta_{2,0}-\beta_2\right)' \\
&\quad \times E\left[xx'\mathbf{1}(q > \gamma \vee \gamma_0)\right]\left(\beta_{2,0}-\beta_2\right) + \left(\beta_{1,0}-\beta_2\right)' E\left[xx'\mathbf{1}(\gamma \wedge \gamma_0 < q \le \gamma_0)\right] \\
&\quad \times \left(\beta_{1,0}-\beta_2\right) + \left(\beta_{2,0}-\beta_1\right)' E\left[xx'\mathbf{1}(\gamma_0 < q \le \gamma \vee \gamma_0)\right]\left(\beta_{2,0}-\beta_1\right) \\
&\ge C\left(\left\|\beta_{1,0}-\beta_1\right\|^2 + \left\|\beta_{2,0}-\beta_2\right\|^2 + |\gamma-\gamma_0|\right),
\end{aligned}
$$

where the last inequality is from Assumptions D1–D4.

Second, $E\left[\sup_{d(\theta,\theta_0)<\delta} |\mathbb{G}_n\left(m\left(w|\theta\right)-m\left(w|\theta_0\right)\right)|\right] \le C\delta$ for any sufficiently small δ. Since $\left\{T\left(w|\theta_1,\theta_{1,0}\right) : d\left(\theta,\theta_0\right) < \delta\right\}$ is a finite-dimensional vector space of functions and $\{\mathbf{1}(q \le \gamma \wedge \gamma_0) : d\left(\theta,\theta_0\right) < \delta\}$ is a VC subgraph class of functions by Lemma 2.4 of Pakes and Pollard (1989), $\{A\left(w|\theta\right) : d\left(\theta,\theta_0\right) < \delta\}$ is VC subgraph by Lemma 2.14 (ii) of Pakes and Pollard (1989). Similarly, $\{B\left(w|\theta\right) : d\left(\theta,\theta_0\right) < \delta\}$, $\{C\left(w|\theta\right) : d\left(\theta,\theta_0\right) < \delta\}$, and $\{D\left(w|\theta\right) : d\left(\theta,\theta_0\right) < \delta\}$ are VC subgraph. From Theorem 2.14.2 of Van der Vaart and Wellner (1996),

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

where F is the envelope of $\{m\left(w|\theta\right)-m\left(w|\theta_0\right) : d\left(\theta,\theta_0\right) < \delta\}$ and can take the form of, e.g.,

$$
\begin{aligned}
F = &\left(\delta^2\|x\|^2 + 2\sigma_{1,0}\delta\|x\varepsilon\|\right)\sup_{d(\theta,\theta_0)<\delta} 1\left(q \le \gamma \wedge \gamma_0\right) + \left(\delta^2\|x\|^2 + 2\sigma_{2,0}\delta\|x\varepsilon\|\right) \\
&\times \sup_{d(\theta,\theta_0)<\delta} 1\left(q > \gamma \vee \gamma_0\right) \\
&+ \sup_{d(\theta,\theta_0)<\delta}\left(\left\|\beta_{1,0}-\beta_2\right\|^2\|x\|^2 + 2\sigma_{1,0}\left\|\beta_{1,0}-\beta_2\right\|\|x\varepsilon\|\right)1\left(\gamma \wedge \gamma_0 < q \le \gamma_0\right) \\
&+ \sup_{d(\theta,\theta_0)<\delta}\left(\left\|\beta_{2,0}-\beta_1\right\|^2\|x\|^2 + 2\sigma_{2,0}\left\|\beta_{2,0}-\beta_1\right\|\|x\varepsilon\|\right)1\left(\gamma_0 < q \le \gamma \vee \gamma_0\right).
\end{aligned}
$$

By Assumptions D4 and D5, $\sqrt{PF^2} \le C\delta$ for $\delta < 1$. So $\phi\left(\delta\right) = \delta$ in Corollary 3.2.6 of Van der Vaart and Wellner (1996) and δ/δ^α is decreasing for all $1 < \alpha < 2$. Since $r_n^2\phi\left(\frac{1}{r_n}\right) = r_n$, $\sqrt{n}d\left(\widehat{\theta}-\theta_0\right) = O_P\left(1\right)$. By the definition of d, the result follows.

Now, we prove $\widehat{\gamma}^* = \gamma_0 + O_{P_r}\left(\frac{1}{n}\right)$. For each n, the parameter space (minus the point θ_0) can be partitioned into the "shells" $S_{j,n} = \left\{\theta : 2^{j-1} < r_n d\left(\theta,\theta_0\right) \le 2^j\right\}$ with $r_n = \sqrt{n}$ and j ranging over the integers. Given an integer J,

$$
\begin{aligned}
P_r\left(r_n d\left(\widehat{\theta}^*,\theta_0\right) > 2^J\right) \le &\sum_{j\ge J, 2^j \le \eta r_n} P_r\left(\inf_{\theta\in S_{j,n}}\left(P_n^* m\left(w|\theta\right) - P_n^* m\left(w|\theta_0\right)\right) \le 0\right) \\
&+ P_r\left(2d\left(\widehat{\theta}^*,\theta_0\right) \ge \eta\right).
\end{aligned}
$$

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

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

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

From the first step in the first part of our proof, $P\left(m\left(w|\theta\right)-m\left(w|\theta_0\right)\right)\geq Cd^2\left(\theta,\theta_0\right)$ for θ in a neighborhood of θ_0. By choosing η small enough, this result holds for $d\left(\theta,\theta_0\right)\leq\eta$, so

$$\begin{aligned}
&P_r\left(\inf_{\theta\in S_{j,n}}\left(P_n^*m\left(w|\theta\right)-P_n^*m\left(w|\theta_0\right)\right)\leq 0\right)\\
&\quad\leq P_r\left(\sup_{\theta\in S_{j,n}}\left|\left(P_n^*-P_n\right)\left(m\left(w|\theta\right)-m\left(w|\theta_0\right)\right)\right|\geq\frac{C}{2}\frac{2^{2j-2}}{r_n^2}\right)\\
&\qquad+P_r\left(\sup_{\theta\in S_{j,n}}\left|\left(P_n-P\right)\left(m\left(w|\theta\right)-m\left(w|\theta_0\right)\right)\right|\geq\frac{C}{2}\frac{2^{2j-2}}{r_n^2}\right).
\end{aligned}$$

From the second step in the first part of our proof, where $\delta<\eta$, and Markov's inequality,

$$\begin{aligned}
P_r\left(\sup_{\theta\in S_{j,n}}\left|\left(P_n-P\right)\left(m\left(w|\theta\right)-m\left(w|\theta_0\right)\right)\right|\geq\frac{C}{2}\frac{2^{2j-2}}{r_n^2}\right)&\leq C\frac{2^j}{\sqrt{n}r_n}\Bigg/\frac{C}{2}\frac{2^{2j-2}}{r_n^2}\\
&=C\frac{r_n}{\sqrt{n}2^j}.
\end{aligned}$$

Conditioning on $\{w_i\}_{i=1}^n$,

$$P^*\sup_{d(\theta,\theta_0)<\delta}\left|\sqrt{n}\left(P_n^*-P_n\right)\left(m\left(w|\theta\right)-m\left(w|\theta_0\right)\right)\right|\leq C\sqrt{P_nF^2}$$

by the maximal inequality, where C can be chosen uniformly for any sample path of the original data, since the class of function $\{m\left(w|\theta\right)-m\left(w|\theta_0\right):d\left(\theta,\theta_0\right)<\delta\}$ is VC subgraph; see Theorem 3 of Léger and MacGibbon (2006). So

$$\begin{aligned}
&P_r\sup_{d(\theta,\theta_0)<\delta}\left|\sqrt{n}\left(P_n^*-P_n\right)\left(m\left(w|\theta\right)-m\left(w|\theta_0\right)\right)\right|\\
&\quad=PP^*\sup_{d(\theta,\theta_0)<\delta}\left|\sqrt{n}\left(P_n^*-P_n\right)\left(m\left(w|\theta\right)-m\left(w|\theta_0\right)\right)\right|\\
&\quad\leq CP\sqrt{P_nF^2}\leq C\sqrt{PP_nF^2}=C\sqrt{PF^2}\leq C\delta,
\end{aligned}$$

where the second to last inequality is from Jensen's inequality. Consequently,

$$P_r\left(\sup_{\theta\in S_{j,n}}\left|\left(P_n^*-P_n\right)\left(m\left(w|\theta\right)-m\left(w|\theta_0\right)\right)\right|\geq\frac{C}{2}\frac{2^{2j-2}}{r_n^2}\right)\leq C\frac{r_n}{\sqrt{n}2^j}.$$

In summary, $P_r\left(r_nd\left(\widehat{\theta}^*,\theta_0\right)>2^J\right)$ is bounded by $C\sum_{j\geq J}\frac{r_n}{\sqrt{n}2^j}$, which can be made arbitrarily small by letting J large enough, since by definition $r_n=\sqrt{n}$. ■

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 $\mathbf{B}_{\text{loc}}\left(\mathbb{R}^d\right)$ as the space of all locally bounded real functions on $\mathbb{R}^d$, endowed with the uniform metric on compacta. The space $\mathbf{C}_{\min}\left(\mathbb{R}^d\right)$ is defined as the subset of continuous functions $x\left(\cdot\right)\in\mathbf{B}_{\text{loc}}\left(\mathbb{R}^d\right)$ for which (i) $x\left(t\right)\to\infty$ as $\|t\|\to\infty$ and (ii) $x\left(t\right)$

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press

achieves its minimum at a unique point in $\mathbb{R}^d$. $\mathbf{B}'_{\text{loc}}(\mathbb{R})$ is the same as $\mathbf{B}_{\text{loc}}(\mathbb{R})$ except being endowed with the Skorohod metric on compacta. The space $\mathbf{D}_{\min}(\mathbb{R})$ is defined as the subset of cadlag functions $x(\cdot) \in \mathbf{B}'_{\text{loc}}(\mathbb{R})$ for which the conditions (i) and (ii) in the definition of $\mathbf{C}_{\min}(\mathbb{R})$ are satisfied.

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

(i) $\left(U_n + D_n, U_n^* + D_n^*\right) \rightsquigarrow \left(U + D, U^* + D^*\right)$, *where* $\rightsquigarrow$ *signifies the weak convergence under some probability measure* P,

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

and

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

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

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

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

https://doi.org/10.1017/S0266466614000012 Published online by Cambridge University Press