Supplement to "Adaptive Estimation of the Threshold Point in Threshold Regression": Technical Appendix

(*Journal of Econometrics*, Vol. 189, No. 1, November 2015, 83-100)

Ping Yu[1]

1. Extra Regularity Conditions

First, define some notations.

$$\theta_\ell \equiv \left(\beta_\ell', \sigma_\ell\right)', \ \theta_{\ell 0} \equiv \left(\beta_{\ell 0}', \sigma_{\ell 0}\right)', \ell = 1, 2.$$

$\mathcal{N}$ is an open neighborhood of θ_0, $\mathcal{N}_\ell$ is an open neighborhood of $\theta_{\ell 0}$, $\ell = 1, 2$, and $\mathcal{N}_\gamma$ is an open neighborhood of γ_0.

$$S\left(\mathsf{w}|\theta\right) = \begin{pmatrix} -\frac{\partial \ln f_{e|x,q}}{\partial e}\left(e|x,q\right)\frac{x'}{\sigma_1}\mathbf{1}\left(q \le \gamma\right) \\ -\frac{\partial \ln f_{e|x,q}}{\partial e}\left(e|x,q\right)\frac{x'}{\sigma_2}\mathbf{1}\left(q > \gamma\right) \\ -\frac{1}{\sigma_1}\left(1 + \frac{\partial \ln f_{e|x,q}}{\partial e}\left(e|x,q\right)e\right)\mathbf{1}\left(q \le \gamma\right) \\ -\frac{1}{\sigma_2}\left(1 + \frac{\partial \ln f_{e|x,q}}{\partial e}\left(e|x,q\right)e\right)\mathbf{1}\left(q > \gamma\right) \end{pmatrix} \equiv \begin{pmatrix} S_{\beta_1}\left(\theta\right) \\ S_{\beta_2}\left(\theta\right) \\ S_{\sigma_1}\left(\theta\right) \\ S_{\sigma_2}\left(\theta\right) \end{pmatrix}$$

is the score function of $\underline{\theta}$, and $\mathcal{I}\left(\theta\right) \equiv E\left[S\left(\mathsf{w}|\theta\right)S'\left(\mathsf{w}|\theta\right)\right]$ is the information matrix of regular parameters.

$$\begin{aligned}
\overline{z}_1\left(\mathsf{w}|\widetilde{\theta}_2, \theta_1\right) &= \ln \frac{\frac{\sigma_1}{\widetilde{\sigma}_2} f_{e|x,q}\left(\left.\frac{\sigma_1 e + x'\left(\beta_1 - \widetilde{\beta}_2\right)}{\widetilde{\sigma}_2}\right| x, q\right)}{f_{e|x,q}\left(e|x,q\right)} = \ln \frac{f_{e|x,q}\left(e|x,q;\widetilde{\theta}_2, \theta_1\right)}{f_{e|x,q}\left(e|x,q;\theta_1,\theta_1\right)}, \\
\overline{z}_2\left(\mathsf{w}|\widetilde{\theta}_1, \theta_2\right) &= \ln \frac{\frac{\sigma_2}{\widetilde{\sigma}_1} f_{e|x,q}\left(\left.\frac{\sigma_2 e + x'\left(\beta_2 - \widetilde{\beta}_1\right)}{\widetilde{\sigma}_1}\right| x, q\right)}{f_{e|x,q}\left(e|x,q\right)} = \ln \frac{f_{e|x,q}\left(e|x,q;\widetilde{\theta}_1, \theta_2\right)}{f_{e|x,q}\left(e|x,q;\theta_2,\theta_2\right)}, \\
\overline{z}_{1i} &= \overline{z}_1\left(\mathsf{w}_i|\theta_{20}, \theta_{10}\right) = \ln \frac{f_{e|x,q}\left(e|x,q;\theta_{20},\theta_{10}\right)}{f_{e|x,q}\left(e|x,q;\theta_{10}\right)}, \ z_{1i} = \overline{z}_{1i}\big|\left(q_i = \gamma_0 -\right), \\
\overline{z}_{2i} &= \overline{z}_2\left(\mathsf{w}_i|\theta_{10}, \theta_{20}\right) = \ln \frac{f_{e|x,q}\left(e|x,q;\theta_{10},\theta_{20}\right)}{f_{e|x,q}\left(e|x,q;\theta_{20}\right)}, \ z_{2i} = \overline{z}_{2i}\big|\left(q_i = \gamma_0 +\right),
\end{aligned}$$

where

$$f_{e|x,q}\left(e|x,q;\widetilde{\theta}_{\ell'}, \theta_\ell\right) = \frac{1}{\widetilde{\sigma}_{\ell'}} f_{e|x,q}\left(\left.\frac{\sigma_\ell e + x'\left(\beta_\ell - \widetilde{\beta}_{\ell'}\right)}{\widetilde{\sigma}_{\ell'}}\right| x, q\right), \ f_{e|x,q}\left(e|x,q;\theta_\ell\right) = f_{e|x,q}\left(e|x,q;\theta_\ell,\theta_{\ell 0}\right)$$

Assumption E1: $P\left(q < \underline{\gamma}\right) > 0$ and $P\left(q > \overline{\gamma}\right) > 0$.
Assumption E2: $\mathcal{I}\left(\theta\right)$ is continuous, nonsingular and finite for $\theta \in \mathcal{N}$.
Assumption E3: For every (μ_1, σ_1) and (μ_2, σ_2) with σ_1 and σ_2 in a bounded set, there exists a slope function $m(\mathsf{w})$ such that $E\left[m(\mathsf{w})^4\right] < \infty$,

$$\left|\ln f_{e|x,q}\left(\mu_1 + \sigma_1 e|x,q\right) - \ln f_{e|x,q}\left(\mu_2 + \sigma_2 e|x,q\right)\right| \le m(\mathsf{w})\left(\left|\mu_1 - \mu_2\right| + \left|\sigma_1 - \sigma_2\right|\right).$$

[1] Department of Economics, 12 Grafton Road, University of Auckland, New Zealand. Email: p.yu@auckland.ac.nz.

Assumption E4: Uniformly for $\theta_1 \in \mathcal{N}_1$, $\theta_2 \in \mathcal{N}_2$ and $\gamma \in \mathcal{N}_\gamma$,

$$E\left[\sup_{\widetilde{\theta}_2 \in \mathcal{N}_2} \left|\overline{z}_1\left(\mathsf{w}|\widetilde{\theta}_2, \theta_1\right)\right| \middle| q = \gamma\right] < \infty, \text{ and } E\left[\sup_{\widetilde{\theta}_1 \in \mathcal{N}_1} \left|\overline{z}_2\left(\mathsf{w}|\widetilde{\theta}_1, \theta_2\right)\right| \middle| q = \gamma\right] < \infty.$$

Assumption E5: Both z_{1i} and z_{2i} have absolutely continuous distributions.
Assumption E6: $P\left(f_{e|x,q}\left(e|x,q;\theta_\ell\right) \neq f_{e|x,q}\left(e|x,q;\widetilde{\theta}_\ell\right)\middle| q\right) > 0$ for any $\theta_\ell, \widetilde{\theta}_\ell \in B_\ell \times \Omega_\ell$, $\theta_\ell \neq \widetilde{\theta}_\ell$, and $q \in \mathbb{Q}$, $\ell = 1, 2$.
Assumption E7: $E\left[\left|\ln f_{e|x,q}\left(e|x,q\right)\right|\right] < \infty$.
Assumption E8: Uniformly for $\theta_1 \in \mathcal{N}_1$, $\theta_2 \in \mathcal{N}_2$ and $\gamma \in \mathcal{N}_\gamma$,

$$E\left[\overline{z}_1\left(\mathsf{w}|\theta_2, \theta_1\right)\right| q = \gamma] < 0, \text{ and } E\left[\overline{z}_2\left(\mathsf{w}|\theta_1, \theta_2\right)\right| q = \gamma] < 0.$$

Assumption E9:

$$E\left[\frac{\partial}{\partial\widetilde{\theta}_1} \ln f_{e|x,q}\left(e|x,q;\widetilde{\theta}_1, \theta_1\right) \frac{\partial}{\partial\widetilde{\theta}_1'} \ln f_{e|x,q}\left(e|x,q;\widetilde{\theta}_1, \theta_1\right) \mathbf{1}\left(q \leq \gamma\right)\right]$$

and

$$E\left[\frac{\partial}{\partial\widetilde{\theta}_2} \ln f_{e|x,q}\left(e|x,q;\widetilde{\theta}_2, \theta_2\right) \frac{\partial}{\partial\widetilde{\theta}_2'} \ln f_{e|x,q}\left(e|x,q;\widetilde{\theta}_2, \theta_2\right) \mathbf{1}\left(q > \gamma\right)\right]$$

are nonsingular and finite for $\widetilde{\theta}_\ell$, θ_ℓ, $\ell = 1, 2$, and γ in an open neighborhood of the true value.
Assumption E10: For $\omega \in [0, 1]$,

$$E\left[\frac{\partial}{\partial\widetilde{\theta}_1} \ln f_{e|x,q}\left(e|x,q;\omega\theta_1 + (1-\omega)\theta_2, \theta_1\right) \frac{\partial}{\partial\widetilde{\theta}_1'} \ln f_{e|x,q}\left(e|x,q;\omega\theta_1 + (1-\omega)\theta_2, \theta_1\right) \middle| q = \gamma\right]$$

and

$$E\left[\frac{\partial}{\partial\widetilde{\theta}_2} \ln f_{e|x,q}\left(e|x,q;\omega\theta_2 + (1-\omega)\theta_1, \theta_2\right) \frac{\partial}{\partial\widetilde{\theta}_2'} \ln f_{e|x,q}\left(e|x,q;\omega\theta_2 + (1-\omega)\theta_1, \theta_2\right) \middle| q = \gamma\right]$$

are nonsingular and finite for θ_1, θ_2 and γ in an open neighborhood of the true value.

2. Efficient Estimation of Regular Parameters

For the semiparametric efficient estimation of β, first note that the loss function is required to be additively separable, otherwise the estimation of γ may affect the estimation of β by a similar argument as in Section 2.2. When the loss function is additively separable, the semiparametric efficiency variance bound of β is the same as that when γ_0 is known. In other words, γ will not affect the efficiency of β. It follows from Chamberlain (1987) that the semiparametric efficiency variance bound of β is

$$\begin{pmatrix} E\left[\frac{xx'\mathbf{1}(q\leq\gamma_0)}{\sigma^2(x,q)}\right]^{-1} & \mathbf{0} \\ \mathbf{0} & E\left[\frac{xx'\mathbf{1}(q>\gamma_0)}{\sigma^2(x,q)}\right]^{-1} \end{pmatrix}.$$

Given this bound, there are basically two methods to find the efficient estimation. The first is summarized in Newey (1993). In short words, the GLSE is efficient. The second method is proposed in Kitamura et al.

(2004) in the empirical likelihood framework. In both methods, the unknown parameter γ is substituted by an n-consistent estimator of γ, which will not affect the asymptotic distribution of the efficient estimators of β.

A natural question is what is the asymptotic distribution of the SEBE of β. The joint SEBE of β and γ is defined by

$$\begin{aligned}\left(\widehat{\beta},\widehat{\gamma}\right) &= \arg\min_{s,t}\int_{B\times\Gamma} l_{1n}\left(s-\beta\right)l_{2n}(t-\gamma)\prod_{i=1}^{n}\left[\frac{1}{\widetilde{\sigma}_1}\widehat{f}\left(\frac{y_i-x_i'\beta_1}{\widetilde{\sigma}_1},x_i,q_i\right)\mathbf{1}(q_i\leq\gamma)\right.\\ &\qquad\left.+\frac{1}{\widetilde{\sigma}_2}\widehat{f}\left(\frac{y_i-x_i'\beta_2}{\widetilde{\sigma}_2},x_i,q_i\right)\mathbf{1}(q_i>\gamma)\right]\pi_1(\beta)\pi_2\left(\gamma\right)d\gamma,\end{aligned}$$

where $l_{1n}(s-\beta)=l_1\left(\sqrt{n}\left(s-\beta\right)\right)$ is the loss function for β, and $\pi_1\left(\beta\right)$ is the prior of β. The answer depends on which preliminary estimator $\widetilde{\beta}$ is used in Step G1. The basic result is that $\widehat{\beta}$ has the same asymptotic distribution as $\widetilde{\beta}$. For example, if $\widetilde{\beta}$ is the LSE, then the asymptotic variance of $\widehat{\beta}$ is that of the LSE rather than the GLSE.[2] In other words, there is no efficiency improvement for β in the SEB procedure. The proof involves an extension of Newey (1994), and will be pursued in a separate paper.

3. Algorithms in the Simulation Study

In the smoothed least squares estimation of Seo and Linton (2007), the objective function is

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\beta_{10}K\left(\frac{\gamma-q_i}{\sigma_n}\right)\right)^2.$$

As suggested by Seo and Linton (2007), $K(x)=\Phi(x)+x\phi(x)$ with Φ and ϕ being the standard normal cdf and pdf respectively, and the bandwidth $\sigma_n=\frac{\log n}{\sqrt{n}}$. The SLSE has a slower convergence rate $\sqrt{\frac{n}{\sigma_n}}$ rather than n.

The subsampling procedure in Gonzalo and Wolf (2005) is designed for time series. Here, we describe this method for the i.i.d. data with e independent of (x,q). Note that these algorithms are designed for the general model (1). For the simple setup in our simulations, the algorithms need to be adjusted correspondingly.

Algorithm S1 (Generating the Bootstrap Sample $\{y_i^*,x_i^*,q_i^*\}_{i=1}^n$)

1. Get the least squares estimation of θ, denoted as $(\widetilde{\beta}',\widetilde{\sigma}',\widetilde{\gamma})'$. The corresponding residuals are denoted as $\{\widetilde{e}_i\}_{i=1}^n$.
2. Generate a sequence $\{x_i^*,q_i^*\}_{i=1}^n$ by sampling with replacement from $\{x_i,q_i\}_{i=1}^n$.
3. Generate a sequence $\{e_i^*\}_{i=1}^n$ by sampling with replacement from $\{\widetilde{e}_i\}_{i=1}^n$. This step is independent of step 2.

[2] We can consider the linear regression for some intuition. Suppose $\{\widetilde{e}_i\}_{i=1}^n$ is the residual in the OLS estimation, and $\widehat{f}\left(e,x\right)$ is the kernel density estimator of $f\left(e,x\right)$ using $\{\widetilde{e}_i,x_i\}_{i=1}^n$, then $\widehat{E}[xe]\equiv\int xe\widehat{f}\left(e,x\right)dedx=\frac{1}{nh^{k+1}}\sum_{i=1}^{n}\int xeK\left(\frac{(e,x)-(\widetilde{e}_i,x_i)}{h}\right)dedx=\frac{1}{n}\sum_{i=1}^{n}\int(\widetilde{x}_i+uh)(\widetilde{e}_i+vh)K\left(v,u\right)dvdu=\frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i\widetilde{e}_i+O(h)\to 0$. But $\widehat{E}[e|x]\equiv\int e\widehat{f}\left(e|x\right)de$ will not converge to zero for every x.

4. Generate a sequence $\{y_i^*\}_{i=1}^n$ by

$$y_i^* = \begin{cases} x_i^{*\prime}\widetilde{\beta}_1 + \widetilde{\sigma}_1 e_i^*, & \text{if } q_i^* \le \widetilde{\gamma}, \\ x_i^{*\prime}\widetilde{\beta}_2 + \widetilde{\sigma}_2 e_i^*, & \text{if } q_i^* > \widetilde{\gamma}. \end{cases}$$

Algorithm S2 (Constructing the Subsampling CI for a Fixed Block Size m)

1. Generate $\{y_i^*, x_i^*, q_i^*\}_{i=1}^m$ by Algorithm S1. The only difference is to replace "replacement" by "non-replacement" and the sampling size "n" by "m".
2. Calculate $(\widetilde{\beta}^*, \widetilde{\gamma}^*)$ based on $\{y_i^*, x_i^*, q_i^*\}_{i=1}^m$ by least squares.
3. Repeat step 1 and 2 B times to get a sequence of estimates $\left\{\widetilde{\beta}_b^*, \widetilde{\gamma}_b^*\right\}_{b=1}^B$.
4. Find the $\frac{\tau}{2}$ and $1-\frac{\tau}{2}$ percentiles of $\{m\left(\widetilde{\gamma}_b^* - \widetilde{\gamma}\right)\}_{b=1}^B$, denoted as $c_{nm}\left(\frac{\tau}{2}\right)$ and $c_{nm}\left(1-\frac{\tau}{2}\right)$, then the equal-tailed subsampling CI for γ is $[\widetilde{\gamma} - n^{-1}c_{nm}\left(1-\frac{\tau}{2}\right), \widetilde{\gamma} - n^{-1}c_{nm}\left(\frac{\tau}{2}\right)]$. Similarly, the symmetric subsampling CI for γ is constructed by finding the $1-\tau$ percentile of $\{m\left|\widetilde{\gamma}_b^* - \widetilde{\gamma}\right|\}_{b=1}^B$, denoted as $c_{nm}(1-\tau)$, and constructing the CI as $[\widetilde{\gamma} - n^{-1}c_{nm}(1-\tau), \widetilde{\gamma} + n^{-1}c_{nm}(1-\tau)]$.

Algorithm S3 (Selecting the Block Size m)

1. Fix a selection of reasonable block size m between m_{low} and m_{up}.
2. Generate K pseudo sequences $\{y_{ki}^*, x_{ki}^*, q_{ki}^*\}_{i=1}^n$, $k = 1, \cdots, K$, by Algorithm S1. For each $k = 1, \cdots, K$ and for each m, compute a subsampling confidence interval $CI_{k,m}$ for γ by Algorithm S2.
3. Compute $\widehat{g}(m) = \#\left\{\widetilde{\gamma} \in CI_{k,m}\right\}/K$.
4. Find the value $\widetilde{m}$ that minimizes $|\widehat{g}(m) - (1-\tau)|$.

$B = 1000$, and $m = n/4$ in our simulations. It is time-consuming to use Algorithm S3 to select the block size adaptively. A suggestion for the parameters in Algorithm S3 is: for $n = 100$, $m_{low} = 15$, $m_{up} = 40$; for $n = 400$, $m_{low} = 50$, $m_{up} = 150$. $K = 1000$.

The smoothed bootstrap procedure in Gijbels et al. (2004) and Seijo and Sen (2011) can only apply to the simple case in our simulation study where the error term is i.i.d. and there are not other covariates except q. We adapt their procedure to the following specific design:

$$y = \begin{cases} \beta_1 + \sigma_1 e, & q \le \gamma; \\ \beta_2 + \sigma_2 e, & q > \gamma. \end{cases}$$

Algorithm B (Smoothed Bootstrap)

1. Get the MLSE of θ, denoted as $(\widetilde{\beta}', \widetilde{\sigma}', \widetilde{\gamma})'$. The corresponding residuals are denoted as $\{\widetilde{e}_i\}_{i=1}^n$.
2. Generate a sequence $\{q_i^*\}_{i=1}^n$ from $\widehat{f}_q(q)$, where $\widehat{f}_q(q)$ is the kernel density estimator of $f_q(q)$. Such a procedure can be found in Section 6.4.1 of Silverman (1986).
3. Generate a sequence $\{e_i^*\}_{i=1}^n$ by sampling with replacement from $\{\widetilde{e}_i\}_{i=1}^n$ This step is independent of step 2.

4. Generate a sequence $\{y_i^*\}_{i=1}^n$ by

$$y_i^* = \begin{cases} \widetilde{\beta}_1 + \widetilde{\sigma}_1 e_i^*, & \text{if } q_i^* \leq \widetilde{\gamma}, \\ \widetilde{\beta}_2 + \widetilde{\sigma}_2 e_i^*, & \text{if } q_i^* > \widetilde{\gamma}. \end{cases}$$

5. Calculate $(\widetilde{\beta}^*, \widetilde{\gamma}^*)$ based on $\{y_i^*, x_i^*, q_i^*\}_{i=1}^n$ by the middle-point least squares.

6. Repeat step 1 to 5 B times to get a sequence of estimators $\left\{\widetilde{\beta}_b^*, \widetilde{\gamma}_b^*\right\}_{b=1}^B$.

7. Find the $\frac{\tau}{2}$ and $1-\frac{\tau}{2}$ percentiles of $\{\widetilde{\gamma}_b^* - \widetilde{\gamma}\}_{b=1}^B$, denoted as $c_n\left(\frac{\tau}{2}\right)$ and $c_n\left(1-\frac{\tau}{2}\right)$, then the equal-tailed bootstrap CI for γ is $[\widetilde{\gamma} - c_n\left(1-\frac{\tau}{2}\right), \widetilde{\gamma} - c_n\left(\frac{\tau}{2}\right)]$. Similarly, the symmetric bootstrap CI for γ is constructed by finding the $1-\tau$ percentile of $\{|\widetilde{\gamma}_b^* - \widetilde{\gamma}|\}_{b=1}^B$, denoted as $c_n(1-\tau)$, and constructing the CI as $[\widetilde{\gamma} - c_n(1-\tau), \widetilde{\gamma} + c_n(1-\tau)]$.

In our simulations, the standard normal kernel and the rule-of-thumb bandwidth of Silverman (1986) is used in Step 2.

4. Uniform Convergence of the Kernel Density Estimator

Lemma 1 *Define an infeasible estimator of the density function to be:*

$$\widetilde{f}(\mathsf{w}) = \frac{1}{n}\sum_{j=1}^n K_h(\mathsf{w}_j - \mathsf{w}),$$

where $K_h(\cdot) = \frac{1}{h^{k+2}} K\left(\frac{\cdot}{h}\right)$. *Then under Assumptions B, D, and K,*

$$n^{(1-\eta)/2} h^{(k+2)/2} \sup_{\mathsf{w}\in\mathbb{W}_n} \left|\widetilde{f}(\mathsf{w}) - f(\mathsf{w})\right| \xrightarrow{p} 0.$$

Proof. Since $\mathbb{W}_n$ is compact, it can be covered by b_n open set W_{in}, $i = 1, \cdots, b_n$, such that the diameter of each set is less than some δ_n which will be determined later. $b_n = O\left(\frac{n^{\nu(k+2)}}{\delta_n^{k+2}}\right)$. By the triangle inequality,

$$\begin{aligned} & \sup_{\mathsf{w}\in\mathbb{W}_n} \left|\widetilde{f}(\mathsf{w}) - E\left[\widetilde{f}(\mathsf{w})\right]\right| \\ \leq\ & \max_{1\leq i\leq b_n}\left\{\sup_{\mathsf{w}\in W_{in}}\left|\widetilde{f}(\mathsf{w}) - \widetilde{f}(\mathsf{w}_i)\right| + \sup_{\mathsf{w}\in W_{in}}\left|E\left[\widetilde{f}(\mathsf{w}) - \widetilde{f}(\mathsf{w}_i)\right]\right| + \left|\widetilde{f}(\mathsf{w}_i) - E\left[\widetilde{f}(\mathsf{w}_i)\right]\right|\right\}, \end{aligned}$$

where $\mathsf{w}_i \in W_{in}$. We will analyze each part of the right hand side in turn. For each i,

$$\sup_{\mathsf{w}\in W_{in}}\left|\widetilde{f}(\mathsf{w}) - \widetilde{f}(\mathsf{w}_i)\right| \overset{(1)}{\leq} \frac{1}{n}\sum_{j=1}^n \left|K_h'(\xi_j)\right| \sup_{\mathsf{w}\in W_{in}}|\mathsf{w} - \mathsf{w}_i| \overset{(2)}{\leq} \frac{C}{h^{k+3}}\delta_n.$$

Here, (1) is from the mean value theorem, ξ_j is some point between $\mathsf{w}_j - \mathsf{w}$ and $\mathsf{w}_j - \mathsf{w}_i$, K_h' is the total derivative with respect to the arguments of K_h, and (2) is from the boundedness of $K'(\cdot)$. Let $\delta_n = \frac{h^{k+3}}{3C}\epsilon_n$ with ϵ_n determined later, then

$$\sup_{\mathsf{w}\in W_{in}}\left|\widetilde{f}(\mathsf{w}) - \widetilde{f}(\mathsf{w}_i)\right| \leq \frac{\epsilon_n}{3}.$$

Similarly, $\sup_{\mathsf{w}\in W_{in}}\left|E\left[\widetilde{f}\left(\mathsf{w}\right)-\widetilde{f}(\mathsf{w}_i)\right]\right|\leq\frac{\epsilon_n}{3}$. As for the third part,

$$\begin{aligned}
&P\left(\max_{1\leq i\leq b_n}\left|\widetilde{f}(\mathsf{w}_i)-E\left[\widetilde{f}(\mathsf{w}_i)\right]\right|>\frac{\epsilon_n}{3}\right)\overset{(1)}{\leq}b_nP\left(\left|h^{k+2}\widetilde{f}(\mathsf{w}_i)-E\left[h^{k+2}\widetilde{f}(\mathsf{w}_i)\right]\right|>\frac{\epsilon_n h^{k+2}}{3}\right)\\
&\overset{(2)}{\leq}2b_n\exp\left\{-3n\left(\frac{\epsilon_n h^{k+2}}{3}\right)^2\bigg/\left(6\sigma_i^2+2C\left(\frac{\epsilon_n h^{k+2}}{3}\right)\right)\right\}\\
&\overset{(3)}{\leq}2b_n\exp\left\{-3n\left(\frac{\epsilon_n h^{k+2}}{3}\right)^2\bigg/\left(6Ch^{k+2}+2C\frac{\epsilon_n h^{k+2}}{3}\right)\right\}\\
&=O\left(b_n\exp\left(-Cnh^{k+2}\epsilon_n^2\right)\right).
\end{aligned}$$

(1) is from the basic probability property. (2) is from the Bernstein's inequality, where

$$\sigma_i^2=var\left(K\left(\frac{\mathsf{w}_i-\mathsf{w}}{h}\right)\right).$$

(3) is because $\sigma_i^2\leq Ch^{k+2}$. Let $\epsilon_n=O\left(\left(n^{(1-\eta)/2}h^{(k+2)/2}\right)^{-1}\right)$, then

$$P\left(\max_{1\leq i\leq b_n}\left|\widetilde{f}(\mathsf{w}_i)-E\left[\widetilde{f}(\mathsf{w}_i)\right]\right|>\frac{\epsilon_n}{3}\right)\xrightarrow{p}0.$$

Combining the arguments above,

$$P\left(\sup_{\mathsf{w}\in\mathbb{W}_n}\left|\widetilde{f}\left(\mathsf{w}\right)-E\left[\widetilde{f}\left(\mathsf{w}\right)\right]\right|>\epsilon_n\right)\xrightarrow{p}0.$$

Note that

$$E\left[\widetilde{f}\left(\mathsf{w}\right)\right]=\int K(x)f(\mathsf{w}+xh)dx=\int K(x)dxf(\mathsf{w})+O\left(h^2\right)=f(\mathsf{w})+O\left(h^2\right)$$

by the Taylor expansion and Assumption D1 and K. From Assumption B, $h^2=o\left(n^{-(1-\eta)/2}h^{-(k+2)/2}\right)$, so

$$n^{(1-\eta)/2}h^{(k+2)/2}\sup_{\mathsf{w}\in\mathbb{W}_n}\left|\widetilde{f}\left(\mathsf{w}\right)-f(\mathsf{w})\right|\xrightarrow{p}0.$$

■

5. Three Criteria of Efficiency

In this section, we will review three criteria of efficiency in parametric models, and discuss how to extend these efficiency concepts to semiparametric cases. The point here is that the usual concepts of semiparametric efficiency bound cannot be applied in threshold regression. Our discussion is based on asymptotic approximation; see Chamberlain (2007) for an excellent summary on decision theory in finite samples.

The classical efficiency theory is based on the average risk (AR). Suppose the loss function is $L\left(\theta,d\right)$, where $\theta\in\Theta\subset\mathbb{R}^k$ is the parameter of interest, $d:\mathcal{Z}\to\mathcal{A}$ is the decision rule, $\mathcal{Z}$ is the sample space, and $\mathcal{A}$ is the action space. For our estimation problem, $\mathcal{A}$ is the same as the parameter space Θ. Suppose the observation z is drawn from the distribution P_θ which has a density p_θ with respect to some dominating measure, then the risk

$$R(\theta,d)=E_\theta\left[L(\theta,d)\right]\equiv\int_{\mathcal{Z}}L(\theta,d(z))\,dP_\theta(z)=\int_{\mathcal{Z}}L(\theta,d(z))\,p_\theta(z)dz,$$

and the average risk with respect to the prior π is

$$R^*(\pi,d)=\int_\Theta R(\theta,d)d\pi(\theta).$$

The optimal (or efficient) estimator is the decision rule that minimizes $R^*(\pi,d)$ for a given π. Because

$$R^*(\pi,d)=\int_\Theta\left[\int_{\mathcal{Z}}L(\theta,d(z))\,p_\theta(z)dz\right]d\pi(\theta)=\int_{\mathcal{Z}}\left[\int_\Theta L(\theta,d(z))\,p_\theta(z)d\pi(\theta)\right]dz$$

by Fubini's theorem, the Bayes estimator (BE) which minimizes the posterior expected loss $\int_\Theta L(\theta,d(z))\,p_\theta(z)d\pi(\theta)$ is optimal. If an asymptotic argument is taken into account, then we should use the asymptotic average risk (AAR)

$$\lim_{K\uparrow\mathbb{R}^k}\overline{\lim_{n\to\infty}}\frac{1}{\lambda(K)}\int_K R(h,d)dh,$$

where $R(h,d)=\int_{\mathcal{Z}}L(\theta_0+\varphi_nh,d(z))\,p_{\theta_0+\varphi_nh}(z)dz$, $h=\varphi_n^{-1}(\theta-\theta_0)$ is the local parameter, φ_n is the appropriate normalization rate for θ, $z=(z_1,\cdots,z_n)$ is the data observed, λ is the Lebesgue measure, and K is an increasing sequence of cubes centered at the origin and converging to $\mathbb{R}^k$. Lemma 3.1 of Chernozhukov and Hong (2004) show that the BE is optimal based on the AAR. In the semiparametric case, the parameter space Θ also includes an infinite-dimensional component. The literature concentrates on simulating the prior using Dirichlet processes, and few efficiency results are available. As mentioned in Section 3.2, the semiparametric Bayes method is under development even in regular models.

Minimaxity is another efficiency criterion. In finite samples, we try to choose d to minimize the sup-risk

$$\sup_{\theta\in\Theta}R(\theta,d).$$

The corresponding asymptotic version is the local asymptotic minimaxity (LAM): the estimator d^* is optimal if it satisfies

$$\lim_{\delta\to0}\underline{\lim_{n\to\infty}}\left[\inf_d\sup_{\|\theta-\theta_0\|<\delta}E_\theta\left[L\left(\varphi_n^{-1}(d-\theta)\right)\right]\right]=\lim_{\delta\to0}\underline{\lim_{n\to\infty}}\left[\sup_{\|\theta-\theta_0\|<\delta}E_\theta\left[L\left(\varphi_n^{-1}(d^*-\theta)\right)\right]\right],\tag{1}$$

where $L(\theta,d)$ depends on θ and d only through their difference $d-\theta$, or it satisfies a little weaker criterion

$$\inf_d\sup_J\underline{\lim_{n\to\infty}}\sup_{h\in J}R(\theta_0+\varphi_nh,d)=\sup_J\underline{\lim_{n\to\infty}}\sup_{h\in J}R(\theta_0+\varphi_nh,d^*),\tag{2}$$

where the outer supremum is over all finite subsets J of $\mathbb{R}^k$. From Ibragimov and Has'minskii (1981), the BE is also efficient under this criterion, which is used as a base in Hirano and Porter (2003) and Yu (2012). In asymptotic shift experiments, the arguments can be somewhat simplified. We can first search among equivariant estimators in the limit experiments and prove that the minimum risk equivariant (MRE) estimator is optimal among all estimators by the Hunt-Stein theorem, then apply the asymptotic representation theorem to find the finite-sample analog. Actually, Hirano and Porter (2003) show that the bias-corrected MLE is also efficient in such experiments. Extending this criterion to the semiparametric case is hard unless the model is locally asymptotically normal (LAN). Chamberlain (1987, 1992) essentially uses the semiparametric version of (1); see also Hirano and Porter (2009) for searching optimal treatment rules using the semiparametric version of (2). In threshold regression, it is hard to find an estimator satisfying the semiparametric version of (1) or (2). In fact, from Ritov and Bickel (1990), the usual semiparametric variance bound can be misleading even in regular models, let alone the present nonregular model. If we want to show our SEBE is optimal in the sense of (1), we must show our estimator is adaptive to the estimation of $f(\mathsf{w})$ *uniformly* in a neighborhood of the true density under some metric. But such a proof seems difficult. Our estimator is actually semiparametric efficient in a weaker sense: it is "pointwise" (instead of "uniform") efficient; that is, for any fixed (unknown) $f(\mathsf{w})$, we can estimate γ as if $f(\mathsf{w})$ were known. This efficiency criterion seems enough for practical purposes.

The last efficiency criterion is optimality among regular estimators. $\widehat{\theta}$ is a regular estimator of θ if

$$\varphi_n^{-1}\left(\widehat{\theta}-(\theta+\varphi_n h)\right) \overset{\theta+\varphi_n h}{\rightsquigarrow} D_\theta,$$

where $\overset{\theta+\varphi_n h}{\rightsquigarrow}$ denotes convergence in distribution under $P_{\theta+\varphi_n h}$, and the limit distribution D_θ may depend on θ but not on h. This kind of estimators correspond to the equivariant estimator in the limit experiment. In asymptotic shift experiments, we still search for the MRE estimator, but show it is optimal in this sense by the convolution theorem, which states that any equivariant estimator can be expressed as the MRE estimator plus an independent error term; see Hirano and Porter (2003) for more discussions. In the semiparametric case, most literature using this criterion concentrates on LAN models; Newey (1990) and Bickel et al. (1998) are the most-cited references. In threshold regression, it is not hard to check that the BE is regular, but it is hard to show that it is optimal among all regular estimators. This is because we do not know what the limit experiment is given that the sufficient statistic of γ is infinite-dimensional; see Section 3.3 of Yu (2012) on this point. Thus, we cannot use this criterion in the semiparametric environment.

In summary, LAM is the only possible criterion for semiparametric efficiency in threshold regression, and our estimator is efficient in a weaker form of this criterion.

Additional References

Chamberlain, G., 1992, Efficiency Bounds for Semiparametric Regression, *Econometrica*, 60, 567-596.

Chamberlain, G., 2007, Decision Theory Applied to An Instrumental Variables Model, *Econometrica*, 75, 609-652.

Hirano, K. and J.R. Porter, 2003, Efficiency in Asymptotic Shift Experiments, unpublished manuscript, University of Wisconsin-Madison.

Hirano, K. and J.R. Porter, 2009, Asymptotics for Statistical Treatment Rules, *Econometrica*, 77, 1683–1701.

Kitamura, Y. et al., 2004, Empirical Likelihood-Based Inference in Conditional Moment Restriction Models, *Econometrica*, 6, 1667-1714.

Newey, W.K., 1993, Efficient Estimation of Models with Conditional Moment Restrictions, in G.S. Maddala, C.R. Rao, and H.D. Vinod, eds., *Handbook of Statistics*, Volume 11: Econometrics, Amsterdam: North-Holland.

Newey, W.K., 1994, The Asymptotic Variance of Semiparametric Estimators, *Econometrica*, 1349-1382.