Supplemental Appendix for
"Robust Estimation for Average Treatment Effects"

Saraswata Chaudhuri* and Jonathan B. Hill†
Dept. of Economics, McGill University
Dept. of Economics, University of North Carolina

December 3, 2024

# Part I

# Omitted Theory and Proofs

Appendix C gives bias correction formulae under tail symmetry. We derive the first order mean-squared-error of the trim-by-$Z$ estimator in Appendix D. In Appendix E we prove scale estimator consistency Theorem 3.5. Appendix F presents a general background theory of the tail decay properties of the variable $Z = hY$ that point identifies the ATE. Finally, in Appendix G we study the trim-by-$X$ estimator and compare it with our estimator.

Assume without loss of generality that the ATE is:

$$\theta = 0.$$

Recall the assumptions. First, the data generating process.

**Assumption A1 (Unconfoundedness):** $Y_1, Y_0 \perp D|X$.

**Assumption A2 (Strict Overlap):** $0 < p_* \leq p(X) \equiv P(D = 1|X) \leq 1 - p_* < 1$ $a.s.$ for a constant $p_*$.

**Assumption A2′ (Limited Overlap):** $0 < p(X) \equiv P(D = 1|X) < 1$ $a.s.$

**Assumption A3 (Distribution Properties):**

$i$. All random variables lie in a complete probability measure space $(\Omega, \mathcal{F}, \mathcal{P})$. $(Y_i, D_i, X_i)'$ are iid.

*Dept. of Economics, McGill University, Quebec, Mongtreal, saraswata.chaudhuri@mcgill.ca.
†Corresponding author. Dept. of Economics, University of North Carolina at Chapel Hill, www.unc.edu/∼jbhill, jbhill@email.unc.edu.

We thank participants at the Cowles Foundation 2014 Conference on Econometrics at Yale University. In particular, we thank Xiaohong Chen and Shakeeb Khan for helpful comments. The authors are solely responsible for all errors.

*ii*. If $E[Z_i^2] = \infty$ then $Z_i$ has power law distribution tails:

$$P\left(Z_i - \theta \leq -c\right) \sim d_1 c^{-\kappa_1} \text{ and } P\left(Z_i - \theta \geq c\right) \sim d_2 c^{-\kappa_2}\text{, where } \kappa_i > 1 \text{ and } d_i \in (0, \infty). \tag{B1}$$

*iii*. Define $\xi \equiv [\gamma', \theta]' \in \mathbb{R}^{q+1}$ and $\mathcal{Z}_i(\xi) \equiv Z_i(\gamma) - \theta$, let $\xi_0$ be the true value of $\xi$, and let $\Xi$ be a compact subset of $\mathbb{R}^{q+1}$ containing $\xi_0$. Let $\{c_n(\xi)\}$ be any sequence of mappings $c_n : \Xi \to (0, \infty)$ that satisfy $P(|\mathcal{Z}_i(\xi)| > c_n(\xi)) = k_n/n$.

*a*. $\mathcal{Z}_i(\xi)$ has for each $\xi$ a continuous distribution with a continuous density function $f_{\mathcal{Z}(\xi)}$, and $E[\sup_{\xi \in \Xi} |\mathcal{Z}_i(\xi)|^\iota] < \infty$ for some $\iota > 0$.

*b*. $c_n(\xi)$ is continuously differentiable with $\inf_{\xi \in \Xi}\{c_n(\xi)\} \to \infty$, $\sup_{\xi \in \Xi}\{c_n(\xi)\} = O(n^{\varpi})$ for some $\varpi > 0$, and $(\partial/\partial\xi)c_n(\xi_0) = O(c_n \mathring{\mathcal{L}}_n)$ for some slowly varying function $\mathring{\mathcal{L}}_n \to (0, \infty]$.

*c*. There exists a continuously differentiable mapping $\mathcal{K} : \Xi \to (0, \infty)$ with $\inf_{\xi \in \Xi} \mathcal{K}(\xi) > 0$, $\sup_{\xi \in \Xi} \mathcal{K}(\xi) < \infty$ and $\sup_{\xi \in \Xi} ||(\partial/\partial\xi)\mathcal{K}(\xi)|| < \infty$, such that $\forall u \in \mathbb{R}$:

$$\lim_{n \to \infty} \sup_{\xi \in \Xi} \left| \frac{n}{k_n} c_n(\xi) \left\{ f_{\mathcal{Z}(\xi)} \left( -c_n(\xi) e^{u/k_n^{1/2}} \right) + f_{\mathcal{Z}(\xi)} \left( c_n(\xi) e^{u/k_n^{1/2}} \right) \right\} - \mathcal{K}(\xi) \right| = 0. \tag{B2}$$

**Assumption A3′ (Second Order Power Law):** A3(i) and A3(iii) hold. Further, $(ii)$ for some $d_i > 0$, $\eta_i > 0$, and $\kappa_i > 1$:

$$P\left(Z_i - \theta < -c\right) = d_1 c^{-\kappa_1} \left(1 + O(c^{-\eta_1})\right) \quad \text{and} \quad P\left(Z_i - \theta > c\right) = d_2 c^{-\kappa_2} \left(1 + O(c^{-\eta_2})\right). \tag{B3}$$

Further, $m_n \to \infty$, $m_n = o(n^{2\eta/(2\eta+\kappa)})$ and $m_n/k_n \to \infty$ where $\eta \equiv \min\{\eta_1, \eta_2\}$ and $\kappa \equiv \min\{\kappa_1, \kappa_2\}$.

**Assumption A4 (Trimming Rate):** $k_n \to \infty$ and $k_n = o(\ln(n))$.

**Assumption A5 (positive scale).** $\liminf_{n \to \infty} \mathcal{V}_n^2 > 0$.

Second, the parametric propensity score and plug-in estimator.

**Assumption B1 (parametric function):** Let $\mathbb{X} \subseteq \mathbb{R}^k$ denote the support of $X_i \in \mathbb{R}^k$, and let $\Gamma \subset \mathbb{R}^q$. There exists a known mapping $p : \mathbb{X} \times \Gamma \to (0,1)$ such that $p(x, \gamma_0) = P(D_i = 1|x)$ $\forall x \in \mathbb{X}$ for a unique interior point $\gamma_0 \in \Gamma$. $p(\cdot, \gamma)$ is Borel measurable for each $\gamma \in \Gamma$. $p(X_i, \gamma)$ is continuous and differentiable on $\Gamma$, $\sigma(X_i)$-*a.e.*

**Assumption B1′ (parametric function).** B1 holds, and $p(X_i, \gamma)$ is twice continuously differentiable, $\sigma(X_i)$-*a.e.*

**Assumption B2 (plug-in):** The plug-in $\hat{\gamma}_n$ satisfies $\sqrt{n}(\hat{\gamma}_n - \gamma_0) = 1/\sqrt{n} \sum_{i=1}^n w_i(1 + o_p(1))$ where $w_i \in \mathbb{R}^q$ is iid, $\sigma(X_i, D_i)$-measurable, it has a continuous distribution, $E[w_i] = 0$, $E[w_i^2] > 0$, and $E|w_i|^{2+\iota} < \infty$ for some $\iota > 0$.

**Assumption B3 (moment bounds):**

*i*. $\sup_{\gamma \in \Gamma}\{|h_i(\gamma) Z_i(\gamma)| \times ||(\partial/\partial\gamma) p_i(\gamma)||\}$ is $L_p$-bounded for some $p > 0$.

*ii.* $h_i(\gamma_0)(\partial/\partial\gamma)p(X_i,\gamma_0)$ is $L_{2+\iota}$-bounded for some $\iota > 0$.

Define

$$S_i(\gamma) \equiv h_i(\gamma)\frac{\partial}{\partial\gamma}p(X_i,\gamma).$$

Under B1$'$ $(\partial/\partial\gamma)S_i(\gamma)$ is well defined and satisfies

$$\begin{aligned}\frac{\partial}{\partial\gamma}S_i(\gamma) &= \frac{\partial}{\partial\gamma}h_i(\gamma)\frac{\partial}{\partial\gamma'}p(X_i,\gamma) + h_i(\gamma)\frac{\partial^2}{\partial\gamma\partial\gamma'}p(X_i,\gamma) \\ &= -h_i^2(\gamma)\frac{\partial}{\partial\gamma}p(X_i,\gamma)\frac{\partial}{\partial\gamma'}p(X_i,\gamma) + h_i(\gamma)\frac{\partial^2}{\partial\gamma\partial\gamma'}p(X_i,\gamma) \\ &= -S_i(\gamma)S_i(\gamma)' + h_i(\gamma)\frac{\partial^2}{\partial\gamma\partial\gamma'}p(X_i,\gamma).\end{aligned}$$

**Assumption B3$'$ (moment bounds):**

*i.* $\sup_{\gamma\in\Gamma}\{||S_i(\gamma)Z_i(\gamma)|\}$, $\sup_{\gamma\in\Gamma}||S_i(\gamma)S_i(\gamma)'Z_i(\gamma)||$ and $\sup_{\gamma\in\Gamma}||h_i(\gamma)(\partial^2/\partial\gamma\partial\gamma')p_i(\gamma) \times Z_i(\gamma)||$ are $L_p$-bounded for some $p > 0$.

*ii.* $\sup_{\gamma\in\Gamma}||S_i(\gamma)||$ is $L_4$-bounded, and $||h_i(\gamma)(\partial^2/\partial\gamma\partial\gamma')p_i(\gamma)||$ is $L_2$-bounded.

Recall

$$h_i(\gamma) \equiv h(X_i,\gamma) \equiv \frac{D_i}{p(X_i,\gamma)} - \frac{1-D_i}{1-p(X_i,\gamma)} \text{ with } h_i = h_i(\gamma_0), \text{ and } Z_i(\gamma) \equiv h_i(\gamma)Y_i \text{ with } Z_i \equiv Z_i(\gamma_0),$$

and

$$\hat{Z}_{n,i}(\gamma) \equiv Z_i(\gamma) - \frac{1}{n}\sum_{j=1}^n Z_j(\gamma), \;\; \hat{Z}^{(a)}_{n,i}(\gamma) \equiv \left|\hat{Z}_{n,i}(\gamma)\right| \;\text{ and }\; \hat{Z}^{(a)}_{n,(1)}(\gamma) \geq \hat{Z}^{(a)}_{n,(2)}(\gamma) \geq \cdots \geq \hat{Z}^{(a)}_{n,(n)}(\gamma),$$

and let $\{k_n\}$ be an *intermediate order* sequence: $k_n \in \{1,...,n\}$, $k_n \to \infty$ and $k_n/n \to 0$. Let $\hat{\gamma}_n$ be an estimator for $\gamma_0$. The tail-trimmed IPW estimator is

$$\hat{\theta}^{(tz)}_n(\hat{\gamma}_n) \equiv \frac{1}{n-k_n}\sum_{i=1}^n Z_i(\hat{\gamma}_n) I\left(\left|Z_i(\hat{\gamma}_n) - \frac{1}{n}\sum_{j=1}^n Z_j(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right).$$

Power law A3 implies

$$c_n = K\left(n/k_n\right)^{1/\kappa}. \tag{B4}$$

By Karamata's Theorem under A3(ii) (Resnick, 1987, Theorem 0.6):[1]

$$E\left[|Z_i|^\kappa I\left(|Z_i| \leq c_n\right)\right] \sim d\left\{\ln(n) - \ln(k_n)\right\} \sim d\ln(n) \tag{B5}$$

[1] Note that for any finite $a > 0$ and some $K(a) > 0$ we have $E[|Z_i|^\kappa I(|Z_i| \leq c_n)] = K(a) + \int_a^{c_n^\kappa} P(|Z_i| \geq u^{1/\kappa})du \sim K(a) + d\int_a^{c_n^\kappa} u^{-1}du = K(a) + d(\ln(c_n^\kappa) - \ln(a))$. Now use $c_n^\kappa = d(n/k_n)$ and $k_n = o(n)$ to deduce $E[|Z_i|^\kappa I(|Z_i| \leq c_n)] \sim d\{\ln(n) - \ln(k_n)\} \sim d\ln(n)$.

$$E\left[|Z_i|^p I\left(|Z_i| \leq c_n\right)\right] \sim \frac{p}{p-\kappa} c_n^p P\left(|Z_i| > c_n\right) \sim \frac{p}{p-\kappa} d^{p/\kappa} \left(\frac{n}{k_n}\right)^{p/\kappa - 1} \quad \forall p > \kappa.$$

## C Bias Correction under Tail Symmetry

Bias from tail-trimmed is defined as:

$$\mathcal{B}_n \equiv \frac{n}{n-k_n} E\left[(Z_i - \theta) I\left(|Z_i - \theta| \geq c_n\right)\right]$$

Recall tail specific versions of $\hat{Z}_{n,i}(\gamma) \equiv Z_i(\gamma) - 1/n \sum_{j=1}^n Z_j(\gamma)$, and their order statistics: $\hat{Z}_{n,i}^{(a)}(\gamma) \equiv |\hat{Z}_{n,i}(\gamma)|$ and

$$\hat{Z}_{n,i}^{(-)}(\gamma) \equiv -\hat{Z}_{n,i}(\gamma) I\left(\hat{Z}_{n,i}(\gamma) < 0\right) \text{ and } \hat{Z}_{n,i}^{(+)}(\gamma) \equiv \hat{Z}_{n,i}(\gamma) I(\hat{Z}_{n,i}(\gamma) > 0) \text{ with } \hat{Z}_{n,(j)}^{(\cdot)}(\gamma) \geq \hat{Z}_{n,(j+1)}^{(\cdot)}(\gamma).$$

Let $\{m_n\}$ be an intermediate order sequence: $m_n \in \{1, ..., n\}$, $m_n \to \infty$ and $m_n = o(n)$. Hill (1975)'s tail index estimators are

$$\hat{\kappa}_{m_n,1}^{-1}(\gamma) = \frac{1}{m_n - 1} \sum_{j=1}^{m_n - 1} \ln\left(\frac{\hat{Z}_{n,(j)}^{(-)}(\gamma)}{\hat{Z}_{n,(m_n)}^{(-)}(\gamma)}\right) \text{ and } \hat{\kappa}_{m_n,2}^{-1}(\gamma) = \frac{1}{m_n - 1} \sum_{j=1}^{m_n - 1} \ln\left(\frac{\hat{Z}_{n,(j)}^{(+)}(\gamma)}{\hat{Z}_{n,(m_n)}^{(+)}(\gamma)}\right).$$

and estimators of the scales are $(d_1, d_2)$:

$$\hat{d}_{m_n,1}(\gamma) \equiv \frac{m_n}{n} \left(\hat{Z}_{n,(m_n)}^{(-)}(\gamma)\right)^{\hat{\kappa}_{m_n,1}(\gamma)} \text{ and } \hat{d}_{m_n,2}(\gamma) \equiv \frac{m_n}{n} \left(\hat{Z}_{n,(m_n)}^{(+)}(\gamma)\right)^{\hat{\kappa}_{m_n,2}(\gamma)}.$$

The core bias estimator is:

$$\begin{aligned} \hat{\mathcal{B}}_n(\gamma) = \frac{n}{n-k_n} \Bigg\{ & \hat{d}_{m_n,2}^{1/\hat{\kappa}_{m_n,2}(\gamma)}(\gamma) \left(\frac{\hat{\kappa}_{m_n,2}(\gamma)}{\hat{\kappa}_{m_n,2}(\gamma) - 1}\right) \left(\frac{k_n}{n}\right)^{1 - 1/\hat{\kappa}_{m_n,2}(\gamma)} \\ & - \hat{d}_{m_n,1}^{1/\hat{\kappa}_{m_n,1}(\gamma)}(\gamma) \left(\frac{\hat{\kappa}_{m_n,1}(\gamma)}{\hat{\kappa}_{m_n,1}(\gamma) - 1}\right) \left(\frac{k_n}{n}\right)^{1 - 1/\hat{\kappa}_{m_n,1}(\gamma)} \Bigg\}. \end{aligned}$$

If the tail indices are known to be identical $\kappa_2 = \kappa_1 = \kappa$ then by Lemma 3.3

$$\mathcal{B}_n \sim \frac{n}{n-k_n} \left(\frac{\kappa}{\kappa - 1}\right) \left(\frac{k_n}{n}\right)^{1 - 1/\kappa} \left\{d_2^{1/\kappa} - d_1^{1/\kappa}\right\}.$$

This justifies the following bias estimator:

$$\hat{\mathcal{B}}_n(\gamma) = \frac{n}{n-k_n} \left(\frac{k_n}{n}\right)^{1 - 1/\hat{\kappa}_{m_n,0}(\gamma)} \left(\frac{\hat{\kappa}_{m_n,0}(\gamma)}{\hat{\kappa}_{m_n,0}(\gamma) - 1}\right) \left\{\hat{d}_{m_n,2}^{1/\hat{\kappa}_{m_n,0}(\gamma)}(\gamma) - \hat{d}_{m_n,1}^{1/\hat{\kappa}_{m_n,0}(\gamma)}(\gamma)\right\}$$

where

$$\hat{\kappa}_{m_n,0}^{-1}(\gamma) = \frac{1}{m_n - 1} \sum_{j=1}^{m_n - 1} \ln \left( \frac{\hat{Z}_{n,(j)}^{(a)}(\gamma)}{\hat{Z}_{n,(m_n)}^{(a)}(\gamma)} \right)$$

$$\hat{d}_{m_n,1}(\gamma) \equiv \frac{m_n}{n} \left( \hat{Z}_{n,(m_n)}^{(-)}(\gamma) \right)^{\hat{\kappa}_{m_n,0}(\gamma)} \text{ and } \hat{d}_{m_n,2}(\gamma) \equiv \frac{m_n}{n} \left( \hat{Z}_{n,(m_n)}^{(+)}(\gamma) \right)^{\hat{\kappa}_{m_n,0}(\gamma)}.$$

# D First Order Mean-Squared-Error of $\hat{\theta}_n^{(tz)}$

Recall from Section 3 of the main paper that the proper standardization for $\hat{\theta}_n^{(tz)}(\hat{\gamma}_n)$ requires:

$$\mathcal{D}_n \equiv -E\left[\frac{\partial}{\partial \gamma} p(X_i, \gamma_0) h_i Z_i I\left(|Z_i - \theta| < c_n\right)\right]$$

$$\vartheta_{n,i} \equiv \left(Z_i - \theta\right) I\left(|Z_i - \theta| < c_n\right) - E\left[\left(Z_i - \theta\right) I\left(|Z_i - \theta| < c_n\right)\right] + \mathcal{D}_n' w_i$$

$$\mathcal{B}_n \equiv \frac{n}{n - k_n} E\left[\left(Z_i - \theta\right) I\left(|Z_i - \theta| \geq c_n\right)\right]$$

and

$$\mathcal{V}_n^2 \equiv E\left[\vartheta_{n,i}^2\right]$$

$$\sigma_n^2 \equiv E\left[\left\{\left(Z_i - \theta\right) I\left(|Z_i - \theta| < c_n\right) - E\left[\left(Z_i - \theta\right) I\left(|Z_i - \theta| < c_n\right)\right]\right\}^2\right]$$

We know by Theorems 3.1 and 3.4 in the main paper that $\mathcal{V}_n^2$ gives the correct scale for $\hat{\theta}_n^{(tz:bc)}(\hat{\gamma}_n)$, while $\mathcal{V}_n^2 \sim K\sigma_n^2$ for some $K > 0$ ($K = 1$ if $E[Z_i^2] = \infty$). The asymptotic first order mean-squared-error of $\hat{\theta}_n^{(tz)}$ is therefore:

$$\mathcal{MSE}_n \equiv K\sigma_n^2/n + \mathcal{B}_n^2.$$

The following result characterizes $\mathcal{MSE}_n$ and the type of sequence $\{k_n\}$ that diminishes $\mathcal{MSE}_n$. Since characterizations $\sigma_n^2$ and $\mathcal{B}_n$ require the tail indices, we assume symmetry $\kappa_1 = \kappa_2 = \kappa$ to reduce notation.

**Lemma D.1.** *Under Assumption A3 with symmetric tail indices $\kappa_1 = \kappa_2 = \kappa$, it follows that:*

$$\kappa \in (1,2) : \mathcal{MSE}_n \sim \frac{1}{n}\left(\frac{n}{k_n}\right)^{2/\kappa - 1} + \left(\frac{n}{n - k_n}\right)^2 \left(\frac{k_n}{n}\right)^{2 - 2/\kappa} \left(\frac{\kappa}{\kappa - 1}\right)^2 \left\{d_2^{1/\kappa} - d_1^{1/\kappa}\right\}^2$$

$$\kappa = 2 : \mathcal{MSE}_n \sim \frac{d \ln(n)}{n} + \left(\frac{n}{n - k_n}\right)^2 \left(\frac{k_n}{n}\right)^{2 - 2/\kappa} \left(\frac{\kappa}{\kappa - 1}\right)^2 \left\{d_2^{1/\kappa} - d_1^{1/\kappa}\right\}^2$$

$$\kappa > 2 : \mathcal{MSE}_n \sim \left\{ \frac{E\left[(Z_i - \theta)^2\right]}{n} - d^{1/\kappa}\left(\frac{\kappa}{\kappa-2}\right)\frac{1}{n}\left(\frac{k_n}{n}\right)^{1-2/\kappa} \right\} + \left(\frac{n}{n-k_n}\right)^2 \left(\frac{k_n}{n}\right)^{2-2/\kappa} \left(\frac{\kappa}{\kappa-1}\right)^2 \left\{d_2^{1/\kappa} - d_1^{1/\kappa}\right\}^2.$$

*Let Assumption A4 hold. If $\kappa \neq 2$ then bias dominates and trimming less, and therefore using small $k_n$ and slow $k_n \to \infty$, diminishes $\mathcal{MSE}_n$ as $n$ increases. Conversely, if $\kappa = 2$ then the variance dominates and trimming more, and therefore using large $k_n$ and fast $k_n \to \infty$, diminishes $\mathcal{MSE}_n$ as $n$ increases.*

**Remark 1.** The proof reveals that the non-uniformity of the impact of $k_n$ on $\mathcal{MSE}_n$ arises from Assumption A4 property $k_n = o(\ln(n))$. If we were free to choose $k_n$ then $k_n/\ln(n) \to \infty$ would lead to bias dominating when $\kappa = 2$ and therefore a small $k_n$ and slow $k_n \to \infty$ leading to a smaller $\mathcal{MSE}_n$.

**Proof.** Observe that

$$\kappa \in (1,2) : \mathcal{MSE}_n \sim \frac{1}{n}\left(\frac{n}{k_n}\right)^{2/\kappa-1} + \left(\frac{n}{n-k_n}\right)^2 \left(\frac{k_n}{n}\right)^{2-2/\kappa} \left(\frac{\kappa}{\kappa-1}\right)^2 \left\{d_2^{1/\kappa} - d_1^{1/\kappa}\right\}^2$$

$$\kappa = 2 : \mathcal{MSE}_n \sim \frac{d\ln(n/k_n)}{n} + \left(\frac{n}{n-k_n}\right)^2 \frac{k_n}{n} 4\left\{d_2^{1/2} - d_1^{1/2}\right\}^2$$

$$\kappa > 2 : \mathcal{MSE}_n \sim \left\{ \frac{E\left[(Z_i - \theta)^2\right]}{n} - d^{1/\kappa}\left(\frac{\kappa}{\kappa-2}\right)\frac{1}{n}\left(\frac{k_n}{n}\right)^{1-2/\kappa} \right\} + \left(\frac{n}{n-k_n}\right)^2 \left(\frac{k_n}{n}\right)^{2-2/\kappa} \left(\frac{\kappa}{\kappa-1}\right)^2 \left\{d_2^{1/\kappa} - d_1^{1/\kappa}\right\}^2.$$

Cases $\kappa < 2$ and $\kappa = 2$ come from directly Lemmas 3.2 and 3.3 of the main paper. Case $\kappa > 2$ can be deduced similarly by using the arguments used to prove Lemma 3.3 in order to characterize the tail-trimmed variance $\sigma_n^2$. If $\kappa \in (1,2)$ then note $n^{-1}(n/k_n)^{2/\kappa-1} = o((n/(n-k_n))^2 (k_n/n)^{2-2/\kappa})$, hence bias dominates and $\mathcal{MSE}_n$ is smaller when trimming is less. If $\kappa > 2$ then $n^{-1}(k_n/n)^{1-2/\kappa} = o((n/(n-k_n))^2(k_n/n)^{2-2/\kappa})$ hence again trimming less reduces $\mathcal{MSE}_n$. Finally, if $\kappa = 2$ then use $k_n = o(\ln(n))$ under Assumption A4 to deduce $(n/(n-k_n))^2(k_n/n) = o(n^{-1}\ln(n/k_n))$, hence the variance term dominates. A large $k_n$ and fast $k_n \to \infty$ reduces variance and therefore $\mathcal{MSE}_n$. $\mathcal{QED}$.

# E Proof of Theorem 3.5

Recall $w_i$ appears in the Assumption B2 first order plug-in expansion

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n w_i(1 + o_p(1)).$$

In general $w_i$ is unobserved, so consider the MLE case:

$$w_i = (E[S_i S_i'])^{-1} S_i \quad \text{where} \quad S_i(\gamma) = h_i(\gamma)\frac{\partial}{\partial \gamma} p(X_i, \gamma).$$

Recall

$$\begin{aligned}
&p_i(\gamma) = p(X_i, \gamma) \\
&h_i(\gamma) \equiv \frac{D_i}{p_i(\gamma)} - \frac{1 - D_i}{1 - p_i(\gamma)} \\
&\sigma_n^2 \equiv E\left[Z_i^2 I\left(|Z_i| < c_n\right)\right] \\
&\mathcal{D}_n \equiv -E\left[\frac{\partial}{\partial \gamma} p_i h_i Z_i I\left(|Z_i - \theta| < c_n\right)\right] \\
&\vartheta_{n,i} \equiv (Z_i - \theta)\, I\left(|Z_i - \theta| < c_n\right) - E\left[(Z_i - \theta)\, I\left(|Z_i - \theta| < c_n\right)\right] + \mathcal{D}_n' w_i \\
&\mathcal{V}_n^2 \equiv E\left[\vartheta_{n,i}^2\right] = E\left[\left\{(Z_i - \theta)\, I\left(|Z_i - \theta| < c_n\right) - E\left[(Z_i - \theta)\, I\left(|Z_i - \theta| < c_n\right)\right] + \mathcal{D}_n' w_i\right\}^2\right]
\end{aligned}$$

and

$$\begin{aligned}
&\hat{Z}_{n,i}(\gamma) \equiv Z_i(\gamma) - \frac{1}{n}\sum_{j=1}^{n} Z_j(\gamma) \\
&\hat{w}_{n,i} \equiv \left(\frac{1}{n}\sum_{i=1}^{n} S_i(\hat{\gamma}_n) S_i(\hat{\gamma}_n)'\right)^{-1} \\
&\hat{\mathcal{D}}_n \equiv -\frac{1}{n}\sum_{i=1}^{n} S_i(\hat{\gamma}_n) Z_i(\hat{\gamma}_n) I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}_{n,(k_n)}^{(a)}(\hat{\gamma}_n)\right) \\
&\hat{\mathcal{V}}_n^2 \equiv \frac{1}{n - k_n}\sum_{i=1}^{n}\left\{\left(\hat{Z}_{n,i}(\hat{\gamma}_n) I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}_{n,(k_n)}^{(a)}(\hat{\gamma}_n)\right) + \left(\frac{n - k_n}{n}\right)\hat{\mathcal{B}}_n(\hat{\gamma}_n)\right) + \hat{\mathcal{D}}_n' \hat{w}_{n,i}\right\}^2.
\end{aligned}$$

Recall that by the definition of a derivative, any differentiable $f : \mathbb{R}^k \to \mathbb{R}$ satisfies

$$f(x_1) - f(x_0) = \frac{\partial}{\partial x'} f(x_1) \times (x_1 - x_0) + o\left(\|x_1 - x_0\|\right), \tag{E1}$$

where $o(||x_1 - x_0||) \to 0$ faster than $||x_1 - x_0|| \to 0$.

**Theorem 3.5.** *Under Assumptions A1, A2′, A3′, A4, A5, B1′, B2, and B3′* $\hat{\mathcal{V}}_n^2/\mathcal{V}_n^2 \xrightarrow{p} 1$.

**Proof.** In order to ease notation, assume:

$$\theta = 0.$$

Let $\iota > 0$ be a tiny number that may be different in different places. Write $\bar{w}_n \equiv 1/n \sum_{i=1}^{n} w_i$ and $p_i(\gamma) = p_i(\gamma)$. It suffices to prove $\tilde{\mathcal{V}}_n^2/\mathcal{V}_n^2 \xrightarrow{p} 1$ where $\tilde{\mathcal{V}}_n^2 = ((n - k_n)/n)\hat{\mathcal{V}}_n^2$.

Observe that:

$$
\begin{aligned}
\tilde{\mathcal{V}}_n^2 &= \frac{1}{n}\sum_{i=1}^{n}\left\{\left(\hat{Z}_{n,i}(\hat{\gamma}_n)I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) + \left(\frac{n-k_n}{n}\right)\hat{\mathcal{B}}_n(\hat{\gamma}_n)\right) + \mathcal{D}_n'w_i\right\}^2 \\
&\quad +2\left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right)'\frac{1}{n}\sum_{i=1}^{n} w_i\hat{Z}_{n,i}(\hat{\gamma}_n)I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) \\
&\quad +2\left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right)'\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i} - w_i\right)\left\{\hat{Z}_{n,i}(\hat{\gamma}_n)I\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right\} \\
&\quad +2\left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right)'\bar{w}_n\left(\frac{n-k_n}{n}\right)\hat{\mathcal{B}}_n(\hat{\gamma}_n) \\
&\quad +2\mathcal{D}_n'\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i} - w_i\right)\hat{Z}_{n,i}(\hat{\gamma}_n)I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) \\
&\quad +2\left(\frac{n-k_n}{n}\right)\hat{\mathcal{B}}_n(\hat{\gamma}_n)\mathcal{D}_n'\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i} - w_i\right) \\
&\quad +2\left(\frac{n-k_n}{n}\right)\hat{\mathcal{B}}_n(\hat{\gamma}_n)\left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right)'\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i} - w_i\right) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\hat{\mathcal{D}}_n'\hat{w}_{n,i}\right)^2 - \left(\mathcal{D}_n'w_i\right)^2\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{\left(\hat{Z}_{n,i}(\hat{\gamma}_n)I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) + \left(\frac{n-k_n}{n}\right)\hat{\mathcal{B}}_n(\hat{\gamma}_n)\right) + \mathcal{D}_n'w_i\right\}^2 + \mathcal{R}_n.
\end{aligned}
$$

By Theorems 3.1 and 3.4 in the main paper $\hat{\mathcal{B}}_n(\hat{\gamma}_n) = \mathcal{B}_n + o_p(\mathcal{V}_n/n^{1/2})$, and $\bar{w}_n = O_p(1/n^{1/2})$ since $w_i$ is iid and square integrable. Steps 1-4, below, imply $\mathcal{R}_n = o_p(\mathcal{V}_n^2)$. By Step 5:

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\hat{Z}_{n,i}(\hat{\gamma}_n)I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) + \left(\frac{n-k_n}{n}\right)\hat{\mathcal{B}}_n(\hat{\gamma}_n)\right) + \mathcal{D}_n'w_i\right\}^2 \\
&\qquad = \frac{1}{n}\sum_{i=1}^{n}\left\{\left(Z_iI\left(|Z_i| < c_n\right) + \left(\frac{n-k_n}{n}\right)\mathcal{B}_n\right) + \mathcal{D}_n'w_i\right\}^2 + o_p(\mathcal{V}_n^2).
\end{aligned} \tag{E2}
$$

Finally, by Step 6:

$$
\frac{1}{\mathcal{V}_n^2}\frac{1}{n}\sum_{i=1}^{n}\left\{\left(Z_iI\left(|Z_i| < c_n\right) + \left(\frac{n-k_n}{n}\right)\mathcal{B}_n\right) + \mathcal{D}_n'w_i\right\}^2 \xrightarrow{p} 1. \tag{E3}
$$

This proves $\tilde{\mathcal{V}}_n^2/\mathcal{V}_n^2 \xrightarrow{p} 1$ as required.

In the steps below we repeatedly use the A5 bound $\liminf_{n\to\infty}\mathcal{V}_n^2 > 0$, and the following three properties. First, by construction:

$$
\frac{1}{n}\sum_{i=1}^{n} I\left(|\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) = \frac{n-k_n}{n}.
$$

Second, by Lemma A.4.a in the main paper, for any $L_p$-bounded random variable $\zeta_i$, $p > 0$:

$$\frac{1}{\sigma_n n^{1/2}} \sum_{i=1}^{n} |\zeta_i| \times \left| I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) - I\left(|Z_i| < c_n\right)\right| = o_p(1). \tag{E4}$$

Third, by Theorem 3.1.b:

$$\mathcal{V}_n \sim K\sigma_n \text{ for some } K > 0. \tag{E5}$$

Fourth, independence and identical distributedness, $\theta = 0$ and power law tail property A3 imply for some slowly varying $\mathcal{L}_n$ (e.g. Ibragimov and Linnik, 1971):

$$\frac{1}{n}\sum_{i=1}^{n} Z_i = O_p\left(\frac{\mathcal{L}_n}{n^{1-1/\min\{\kappa,2\}}}\right). \tag{E6}$$

Note throughout that:

$$\frac{\partial}{\partial\gamma} h_i(\gamma) = -h_i(\gamma)^2 \frac{\partial}{\partial\gamma} p_i(\gamma)$$

**Step 1.** We want to show

$$\frac{1}{n}\sum_{i=1}^{n} w_i \hat{Z}_{n,i}(\hat{\gamma}_n) I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) = O_p(\mathcal{V}_n).$$

By the Cauchy-Schwartz inequality, square integrability of $w_i$, and (E5):

$$\left|E\left[w_i Z_i I\left(|Z_i| < c_n\right)\right]\right| \leq K\left(E\left[Z_i^2 I\left(|Z_i| < c_n\right)\right]\right)^2 = K\sigma_n \sim K\mathcal{V}_n.$$

Thus, it suffices to show

$$\frac{1}{n}\sum_{i=1}^{n} w_i \hat{Z}_{n,i}(\hat{\gamma}_n) I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) - E\left[w_i Z_i I\left(|Z_i| < c_n\right)\right] = o_p(\mathcal{V}_n).$$

Note that:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} w_i \hat{Z}_{n,i}(\hat{\gamma}_n) I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) - E\left[w_i Z_i I\left(|Z_i| < c_n\right)\right] \\
&\qquad = \frac{1}{n}\sum_{i=1}^{n} w_i Z_i I\left(|Z_i| < c_n\right) - E\left[w_i Z_i I\left(|Z_i| < c_n\right)\right] \\
&\qquad\quad + \frac{1}{n}\sum_{i=1}^{n} w_i Z_i \left\{I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) - I\left(|Z_i| < c_n\right)\right\} \\
&\qquad\quad + \frac{1}{n}\sum_{i=1}^{n} w_i \left\{Z_i(\hat{\gamma}_n) - Z_i\right\} I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right)
\end{aligned}$$

$$- \frac{1}{n}\sum_{i=1}^{n} Z_i \times \frac{1}{n}\sum_{i=1}^{n} w_i I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right)$$
$$- \frac{1}{n}\sum_{i=1}^{n} \{Z_i(\hat{\gamma}_n) - Z_i\} \times \frac{1}{n}\sum_{i=1}^{n} w_i I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right)$$
$$= \sum_{i=1}^{5} \mathcal{A}_{i,n}.$$

We will prove each $\mathcal{A}_{i,n} = o_p(\mathcal{V}_n)$.

$\mathcal{A}_{1,n} = o_p(\mathcal{V}_n)$. Define

$$\mathcal{W}_{n,i} \equiv \frac{w_i Z_i I\left(|Z_i| < c_n\right) - E\left[w_i Z_i I\left(|Z_i| < c_n\right)\right]}{\sigma_n}.$$

Under Assumption B2 $w_i$ is $L_{2+\iota}$-bounded. Hence, by Hôlder's inequality, for some $\delta > 0$ that satisfies $(1+\delta)(1+\delta/2) \leq 1+\iota$:

$$\begin{aligned} E\left|\mathcal{W}_{n,i}\right|^{1+\delta} &\leq K\frac{1}{\sigma_n^{1+\delta}} E\left[\left|w_i Z_i\right|^{1+\delta} I\left(|Z_i| < c_n\right)\right] \\ &\leq K\frac{1}{\sigma_n^{1+\delta}}\left(E\left|w_i\right|^{(1+\delta)(1+\delta/2)}\right)^{\frac{2}{2+\iota}}\left(E\left[\left|Z_i\right|^2 I\left(|Z_i| < c_n\right)\right]\right)^{\frac{1+\delta}{2}} \leq K\frac{1}{\sigma_n^{1+\delta}}\sigma_n^{1+\delta} = K. \end{aligned}$$

Therefore $\mathcal{W}_{n,i}$ is uniformly integrable. Since $\mathcal{W}_{n,i}$ is iid over $i \in \{1,...,n\}$, and uniformly integrable, it satisfies the conditions of Theorem 2 Andrews (1988), hence:

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{W}_{n,i} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{w_i Z_i I\left(|Z_i| < c_n\right) - E\left[w_i Z_i I\left(|Z_i| < c_n\right)\right]}{\sigma_n}\right\} \xrightarrow{p} 0. \tag{E7}$$

Therefore $\mathcal{A}_{1,n} = o_p(\sigma_n)$, hence $\mathcal{A}_{1,n} = o_p(\mathcal{V}_n)$ by (E5).

$\mathcal{A}_{2,n} = o_p(\mathcal{V}_n/n^{1/2})$. In view of $E[w_i^2] < \infty$, use (E4) to yield $\mathcal{A}_{2,n} = o_p(\mathcal{V}_n/n^{1/2})$.

$\mathcal{A}_{3,n} = O_p\left(\mathcal{V}_n/n^{1/2}\right)$. Write

$$\begin{aligned} \mathcal{A}_{3,n} &= \frac{1}{n}\sum_{i=1}^{n} w_i \{Z_i(\hat{\gamma}_n) - Z_i\} I\left(|Z_i| < c_n\right) \\ &+ \frac{1}{n}\sum_{i=1}^{n} w_i \{Z_i(\hat{\gamma}_n) - Z_i\}\left\{I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) - I\left(|Z_i| < c_n\right)\right\}. \end{aligned} \tag{E8}$$

Consider the first term. By derivative property (E1):

$$\frac{1}{n}\sum_{i=1}^{n} w_i \{Z_i(\hat{\gamma}_n) - Z_i\} I\left(|Z_i| < c_n\right) = -\frac{1}{n}\sum_{i=1}^{n} w_i S_i Z_i I\left(|Z_i| < c_n\right) \times (\hat{\gamma}_n - \gamma_0) + o_p(\|\hat{\gamma}_n - \gamma_0\|).$$

Under B2 $\hat{\gamma}_n - \gamma_0 = O_p(1/n^{1/2})$. Furthermore, by construction

$$w_i S_i Z_i I\left(|Z_i| < c_n\right) = \left(E\left[S_i S_i'\right]\right)^{-1} S_i S_i' \times Z_i I\left(|Z_i| < c_n\right),$$

and $S_i$ is $L_4$-bounded under B3′(ii). Hence by the Cauchy-Schwartz inequality and (E5):

$$\sup_{\lambda'\lambda} E\left[\left(\lambda' S_i\right)^2 \times Z_i I\left(|Z_i| < c_n\right)\right] \leq K\left(E\left[Z_i^2 I\left(|Z_i| < c_n\right)\right]\right)^{1/2} = K\sigma_n \sim K\mathcal{V}_n.$$

Now invoke Markov's inequality to yield:

$$\frac{1}{\mathcal{V}_n n}\sum_{i=1}^n w_i S_i Z_i I\left(|Z_i| < c_n\right) = O_p(1).$$

Therefore

$$\frac{1}{n}\sum_{i=1}^n w_i \left\{Z_i(\hat{\gamma}_n) - Z_i\right\} I\left(|Z_i| < c_n\right) = O_p\left(\mathcal{V}_n/n^{1/2}\right).$$

Turning to the second term in (E8), apply the mean-value-theorem to yield:

$$\left|\frac{1}{n}\sum_{i=1}^n w_i \left\{Z_i(\hat{\gamma}_n) - Z_i\right\}\left\{I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) - I\left(|Z_i| < c_n\right)\right\}\right|$$
$$\leq \left|\frac{1}{n}\sum_{i=1}^n |w_i| \sup_{\gamma\in\Gamma}\left\{\|S_i(\gamma)h_i(\gamma)\|\right\}\left\{I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) - I\left(|Z_i| < c_n\right)\right\}\right| \times \|\hat{\gamma}_n - \gamma_0\|.$$

Under B2 $||\hat{\gamma}_n - \gamma_0|| = o_p(1)$ and from B3′(i) $\sup_{\gamma\in\Gamma}\{||S_i(\gamma)h_i(\gamma)||\}$ is $L_p$-bounded for some $p > 0$, and $w_i$ is square integrable. The right hand side is therefore $o_p(\mathcal{V}_n/n^{1/2})$ by (E4).

$\mathcal{A}_{4,n} = o_p(\mathcal{V}_n/n^{1/2})$. Expansion (E4) implies:

$$\frac{1}{n}\sum_{i=1}^n w_i \left\{I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) - I\left(|Z_i| < c_n\right)\right\} = o_p(\mathcal{V}_n/n^{1/2}). \tag{E9}$$

Use the fact that $w_i$ is iid and square integrable, and $I(|Z_i| < c_n)$ is bounded, to deduce

$$\frac{1}{n}\sum_{i=1}^n w_i I\left(|Z_i| < c_n\right) = O_p(1/n^{1/2}). \tag{E10}$$

Combine the sample mean property (E6) with (E9) and (E10) to yield:

$$\mathcal{A}_{4,n} = O_p\left(\frac{\mathcal{L}_n}{n^{3/2-1/\min\{\kappa,2\}}}\right) = o_p\left(\sigma_n/n^{1/2}\right) = o_p\left(\mathcal{V}_n/n^{1/2}\right).$$

$\mathcal{A}_{5,n} = O_p(\mathcal{V}_n/n^{1/2})$. Derivative property (E1) implies:

$$\frac{1}{n}\sum_{i=1}^{n}\{Z_i(\hat{\gamma}_n) - Z_i\} = \frac{1}{n}\sum_{i=1}^{n} S_i Z_i I\left(|Z_i| < c_n\right) \times (\hat{\gamma}_n - \gamma_0) + o_p(\|\hat{\gamma}_n - \gamma_0\|).$$

Since $S_i$ is square integrable it follows

$$\sup_{\lambda'\lambda}\left|E\left[\lambda' S_i Z_i I\left(|Z_i| < c_n\right)\right]\right| \leq K\left(E\left[Z_i^2 I\left(|Z_i| < c_n\right)\right]\right)^{1/2} = K\sigma_n \sim K\mathcal{V}_n.$$

Now recall $\hat{\gamma}_n - \gamma_0 = O_p(1/\sqrt{n})$ to yield:

$$\frac{1}{n}\sum_{i=1}^{n}\{Z_i(\hat{\gamma}_n) - Z_i\} = O_p\left(\mathcal{V}_n/n^{1/2}\right). \tag{E11}$$

Next, by the arguments above:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} w_i I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) &= \frac{1}{n}\sum_{i=1}^{n} w_i I\left(|Z_i| < c_n\right) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} w_i \left\{I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) - I\left(|Z_i| < c_n\right)\right\} \\
&= O_p\left(1/n^{1/2}\right) + O_p\left(\mathcal{V}_n/n^{1/2}\right).
\end{aligned}$$

**Step 2.** We need

$$\left\|\hat{\mathcal{D}}_n - \mathcal{D}_n\right\| = o_p(\mathcal{V}_n) \text{ and } \|\mathcal{D}_n\| = O_p(\mathcal{V}_n).$$

The second equality $||\mathcal{D}_n|| = O_p(\mathcal{V}_n)$ is verified in the proof of Theorem 3.1.b.

Consider $||\hat{\mathcal{D}}_n - \mathcal{D}_n|| = o_p(\mathcal{V}_n)$. Observe that

$$|S_i(\hat{\gamma}_n)Z_i(\hat{\gamma}_n) - S_i Z_i| \leq \left\{\sup_{\gamma\in\Gamma}\left\|\frac{\partial}{\partial\gamma}S_i(\gamma) \times Z_i(\gamma)\right\| + \sup_{\gamma\in\Gamma}\left\|S_i(\gamma)S_i(\gamma)'Z_i(\gamma)\right\|\right\}\|\hat{\gamma}_n - \gamma_0\|.$$

Under B3′(i) $\sup_{\gamma\in\Gamma}||(\partial/\partial\gamma)S_i(\gamma) \times Z_i(\gamma)||$ and $\sup_{\gamma\in\Gamma}||S_i(\gamma)S_i(\gamma)'Z_i(\gamma)||$ are $L_p$-bounded for some $p > 0$. Now apply (E4) to deduce:

$$\hat{\mathcal{D}}_n = -\frac{1}{n}\sum_{i=1}^{n} S_i(\hat{\gamma}_n)Z_i(\hat{\gamma}_n)I\left(|Z_i| < c_n\right) + o_p(\mathcal{V}_n/n^{1/2}). \tag{E12}$$

Note that:

$$\frac{1}{n}\sum_{i=1}^{n} S_i(\hat{\gamma}_n)Z_i(\hat{\gamma}_n)I\left(|Z_i| < c_n\right) - E\left[S_i Z_i I\left(|Z_i| < c_n\right)\right]$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\left\{S_i Z_i I\left(|Z_i| < c_n\right) - E\left[S_i Z_i I\left(|Z_i| < c_n\right)\right]\right\} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{S_i(\hat{\gamma}_n) - S_i\right\} Z_i I\left(|Z_i| < c_n\right) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} S_i \left\{Z_i(\hat{\gamma}_n) - Z_i\right\} I\left(|Z_i| < c_n\right) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{S_i(\hat{\gamma}_n) - S_i\right\}\left\{Z_i(\hat{\gamma}_n) - Z_i\right\} I\left(|Z_i| < c_n\right) \\
&= \sum_{i=1}^{4} \mathcal{C}_{i,n}.
\end{aligned}
$$

It remains to show each $||\mathcal{C}_{i,n}|| = o_p(\mathcal{V}_n)$. In view of (E12), $||\hat{\mathcal{D}}_n - \mathcal{D}_n|| = o_p(\mathcal{V}_n)$ then follows.

$\mathcal{C}_{1,n}$. $||\mathcal{C}_{1,n}|| = o_p(\mathcal{V}_n)$ follows from (E7) and (E5).

$\mathcal{C}_{2,n}$. By the mean-value-theorem:

$$
\|\mathcal{C}_{2,n}\| \leq \frac{1}{n}\sum_{i=1}^{n}\sup_{\gamma\in\Gamma}\left\|\frac{\partial}{\partial\gamma}S_i(\gamma)\right\| |Z_i|\, I\left(|Z_i| < c_n\right) \times \|\hat{\gamma}_n - \gamma_0\|.
$$

Under B3′(ii) $\sup_{\gamma\in\Gamma} ||(\partial/\partial\gamma)S_i(\gamma)||$ is iid and integrable, hence by Markov's inequality:

$$
\frac{1}{n}\sum_{i=1}^{n}\sup_{\gamma\in\Gamma}\left\|\frac{\partial}{\partial\gamma}S_i(\gamma)\right\| |Z_i|\, I\left(|Z_i| < c_n\right) = O_p\left(c_n\right).
$$

Therefore, in view of B2: $||\mathcal{C}_{2,n}|| = O_p(c_n/n^{1/2})$. If $\kappa \geq 2$ then use the construction (B4) of $c_n$ to yield $c_n/n^{1/2} = K(n/k_n)^{1/\kappa}/n^{1/2} = o(1) = o(\sigma_n)$. If $\kappa < 2$ then by Karamata theory (B5) $c_n = (n/k_n)^{1/2}c_n/(n/k_n)^{1/2} \sim K\sigma_n/(n/k_n)^{1/2}$ hence $||\mathcal{C}_{2,n}|| = o_p(\sigma_n)$. Therefore $||\mathcal{C}_{2,n}|| = o_p(\mathcal{V}_n)$ by (E5).

$\mathcal{C}_{3,n}, \mathcal{C}_{4,n}$. $||\mathcal{C}_{3,n}|| = o_p(\mathcal{V}_n)$ follows from the argument following (E8). $||\mathcal{C}_{4,n}|| = o_p(\mathcal{V}_n)$ can be verified along the lines of $\mathcal{C}_{2,n}$ and $\mathcal{C}_{3,n}$.

**Step 3.** We will prove

$$
\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i} - w_i\right)\hat{Z}_{n,i}(\hat{\gamma}_n) I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right)\right\| = o_p\left(\mathcal{V}_n\right).
$$

An identical argument yields

$$
\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i} - w_i\right)\right\| = o_p\left(\mathcal{V}_n\right).
$$

Observe that

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i}-w_i\right)\hat{Z}_{n,i}(\hat{\gamma}_n)I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right|<\hat{Z}_{n,(k_n)}^{(a)}(\hat{\gamma}_n)\right)$$
$$=\left(E\left[S_iS_i'\right]\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left(S_i(\hat{\gamma}_n)-S_i\right)\hat{Z}_{n,i}(\hat{\gamma}_n)I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right|<\hat{Z}_{n,(k_n)}^{(a)}(\hat{\gamma}_n)\right)$$
$$+\left(\left(\frac{1}{n}\sum_{i=1}^{n}S_i(\hat{\gamma}_n)S_i(\hat{\gamma}_n)'\right)^{-1}-(E[S_iS_i'])^{-1}\right)\frac{1}{n}\sum_{i=1}^{n}S_i(\hat{\gamma}_n)\hat{Z}_{n,i}(\hat{\gamma}_n)I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right|<\hat{Z}_{n,(k_n)}^{(a)}(\hat{\gamma}_n)\right).$$

In Step 3.1 we prove $1/n\sum_{i=1}^{n}S_i(\hat{\gamma}_n)S_i(\hat{\gamma}_n)'-E[S_iS_i']=o_p(1)$. Further, by replicating arguments in Steps 1 and 2 it can be shown that:

$$\frac{1}{n}\sum_{i=1}^{n}\left(S_i(\hat{\gamma}_n)-S_i\right)\hat{Z}_{n,i}(\hat{\gamma}_n)\left\{I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right|<\hat{Z}_{n,(k_n)}^{(a)}(\hat{\gamma}_n)\right)-I\left(|Z_i|<c_n\right)\right\}=o_p\left(\sigma_n/n^{1/2}\right)$$
$$\frac{1}{n}\sum_{i=1}^{n}S_i(\hat{\gamma}_n)\hat{Z}_{n,i}(\hat{\gamma}_n)\left\{I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right|<\hat{Z}_{n,(k_n)}^{(a)}(\hat{\gamma}_n)\right)-I\left(|Z_i|<c_n\right)\right\}=o_p\left(\sigma_n/n^{1/2}\right).$$

Now, use $1/n\sum_{i=1}^{n}S_i(\hat{\gamma}_n)S_i(\hat{\gamma}_n)'-E[S_iS_i']=o_p(1)$ to deduce:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i}-w_i\right)\hat{Z}_{n,i}(\hat{\gamma}_n)I\left(|Z_i|<c_n\right)$$
$$=\left(E\left[S_iS_i'\right]\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left(S_i(\hat{\gamma}_n)-S_i\right)\hat{Z}_{n,i}(\hat{\gamma}_n)I\left(|Z_i|<c_n\right)+\frac{1}{n}\sum_{i=1}^{n}S_i(\hat{\gamma}_n)\hat{Z}_{n,i}(\hat{\gamma}_n)I\left(|Z_i|<c_n\right)\times o_p\left(1\right)$$
$$=\left(E\left[S_iS_i'\right]\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left(S_i(\hat{\gamma}_n)-S_i\right)Z_i(\hat{\gamma}_n)I\left(|Z_i|<c_n\right)-\frac{1}{n}\sum_{i=1}^{n}Z_i(\hat{\gamma}_n)\times\frac{1}{n}\sum_{i=1}^{n}S_i(\hat{\gamma}_n)I\left(|Z_i|<c_n\right)+o_p(\hat{\mathcal{D}}_n)$$
$$=\mathcal{E}_{1,n}+\mathcal{E}_{2,n}+\mathcal{E}_{3,n}.$$

Consider $\mathcal{E}_{1,n}$. By the sample mean  property (E6):

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(S_i(\hat{\gamma}_n)-S_i\right)\hat{Z}_{n,i}(\hat{\gamma}_n)I\left(|Z_i|<c_n\right) &= \frac{1}{n}\sum_{i=1}^{n}\left(S_i(\hat{\gamma}_n)-S_i\right)Z_iI\left(|Z_i|<c_n\right)\\ &+\frac{1}{n}\sum_{i=1}^{n}\left(S_i(\hat{\gamma}_n)-S_i\right)\left(Z_i(\hat{\gamma}_n)-Z_i\right)I\left(|Z_i|<c_n\right)\\ &-O_p\left(\frac{\mathcal{L}_n}{n^{1-1/\min\{\kappa,2\}}}\right)\times\frac{1}{n}\sum_{i=1}^{n}\left(S_i(\hat{\gamma}_n)-S_i\right)I\left(|Z_i|<c_n\right).\end{aligned}$$

Arguments in Step 2 prove each tern is $o_p(\mathcal{V}_n)$.

Next, for $\mathcal{E}_{2,n}$ write:

$$\begin{aligned}\mathcal{E}_{2,n} &= \frac{1}{n}\sum_{i=1}^{n} Z_i \times \frac{1}{n}\sum_{i=1}^{n} S_i I\left(|Z_i| < c_n\right) \\ &+\frac{1}{n}\sum_{i=1}^{n} Z_i \times \frac{1}{n}\sum_{i=1}^{n} \left(S_i(\hat{\gamma}_n) - S_i\right) I\left(|Z_i| < c_n\right) \\ &+\frac{1}{n}\sum_{i=1}^{n} \left(Z_i(\hat{\gamma}_n) - Z_i\right) \times \frac{1}{n}\sum_{i=1}^{n} S_i I\left(|Z_i| < c_n\right) \\ &+\frac{1}{n}\sum_{i=1}^{n} \left(Z_i(\hat{\gamma}_n) - Z_i\right) \times \frac{1}{n}\sum_{i=1}^{n} \left(S_i(\hat{\gamma}_n) - S_i\right) I\left(|Z_i| < c_n\right).\end{aligned}$$

Step 2 derivations and (E11) from Step 1 prove each term is $o_p(\mathcal{V}_n)$.

Finally, by Step 2 $||\hat{\mathcal{D}}_n|| = O_p(\mathcal{V}_n)$ hence

$$\|\mathcal{E}_{3,n}\| = o_p\left(\left\|\hat{\mathcal{D}}_n\right\|\right) = o_p\left(\mathcal{V}_n\right).$$

**Step 3.1** We need to show

$$\frac{1}{n}\sum_{i=1}^{n} S_i(\hat{\gamma}_n)S_i(\hat{\gamma}_n)' - E\left[S_iS_i'\right] = o_p\left(1\right). \tag{E13}$$

Add and subtract terms to yield:

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} S_i(\hat{\gamma}_n)S_i(\hat{\gamma}_n)' - E\left[S_iS_i'\right] \\ &\quad= \frac{1}{n}\sum_{i=1}^{n} S_iS_i' - E\left[S_iS_i'\right] + \frac{1}{n}\sum_{i=1}^{n} S_i\left\{S_i(\hat{\gamma}_n) - S_i\right\}' \\ &\qquad + \frac{1}{n}\sum_{i=1}^{n} \left\{S_i(\hat{\gamma}_n) - S_i\right\}S_i' + \frac{1}{n}\sum_{i=1}^{n} \left\{S_i(\hat{\gamma}_n) - S_i\right\}\left\{S_i(\hat{\gamma}_n) - S_i\right\}'.\end{aligned}$$

It suffices to prove each term is $o_p(1)$.

Recall $S_i$ is iid and square integrable, hence:

$$\frac{1}{n}\sum_{i=1}^{n} S_iS_i' - E[S_iS_i'] = o_p(1).$$

The second term satisfies

$$\left\|\frac{1}{n}\sum_{i=1}^{n} S_i\left\{S_i(\hat{\gamma}_n) - S_i\right\}'\right\| \leq \frac{1}{n}\sum_{i=1}^{n}\sup_{\gamma\in\Gamma}\left\|\frac{\partial}{\partial\gamma}S_i(\gamma)S_i\right\| \times \|\hat{\gamma}_n - \gamma_0\| = o_p(1).$$

The $o_p(1)$ term follows by noting $||\hat{\gamma}_n - \gamma_0|| = o_p(1)$ by B2; and under B3′(ii) $\sup_{\gamma\in\Gamma} ||(\partial/\partial\gamma)S_i(\gamma)S_i||$ is integrable, hence

$$\frac{1}{n}\sum_{i=1}^{n}\sup_{\gamma\in\Gamma}\left\|\frac{\partial}{\partial\gamma}S_i(\gamma)S_i\right\| = O_p(1).$$

Similarly, $\sup_{\gamma\in\Gamma} ||(\partial/\partial\gamma)S_i(\gamma)(\partial/\partial\gamma)S_i(\gamma)||$ is integrable under B3′(ii). Hence, for the third term:

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\left\{S_i(\hat{\gamma}_n) - S_i\right\}\left\{S_i(\hat{\gamma}_n) - S_i\right\}'\right\| \leq \frac{1}{n}\sum_{i=1}^{n}\sup_{\gamma\in\Gamma}\left\|\frac{\partial}{\partial\gamma}S_i(\gamma)\frac{\partial}{\partial\gamma}S_i(\gamma)\right\| \times \left\|\hat{\gamma}_n - \gamma_0\right\|^2 = o_p(1).$$

**Step 4.** Next, we prove:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\hat{\mathcal{D}}_n'\hat{w}_{n,i}\right)^2 - \left(\mathcal{D}_n'w_i\right)^2\right\} = o_p(\mathcal{V}_n).$$

Expand:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\hat{\mathcal{D}}_n'\hat{w}_{n,i}\right)^2 - \left(\mathcal{D}_n'w_i\right)^2\right\} \\
&= \left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right)' \times \frac{1}{n}\sum_{i=1}^{n} w_i w_i' \times \left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right) \\
&\quad + 2\left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right)' \times \frac{1}{n}\sum_{i=1}^{n} w_i w_i' \times \mathcal{D}_n \\
&\quad + \mathcal{D}_n'\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i} - w_i\right)\left(\hat{w}_{n,i} - w_i\right)'\mathcal{D}_n \\
&\quad + 2\mathcal{D}_n'\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i} - w_i\right)\left(\hat{w}_{n,i} - w_i\right)'\left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right) \\
&\quad + \left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right)'\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i} - w_i\right)\left(\hat{w}_{n,i} - w_i\right)'\left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right) \\
&\quad + 2\mathcal{D}_n'\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i} - w_i\right)w_i'\mathcal{D}_n \\
&\quad + 4\mathcal{D}_n'\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i} - w_i\right)w_i'\left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right) \\
&\quad + 2\left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right)'\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i} - w_i\right)w_i'\left(\hat{\mathcal{D}}_n - \mathcal{D}_n\right).
\end{aligned}$$

Under B2 $w_i$ is iid and square integrable, hence $1/n\sum_{i=1}^{n} w_i w_i' \xrightarrow{p} E[w_i w_i']$. By Step 2 $\hat{\mathcal{D}}_n - \mathcal{D}_n = o_p(\mathcal{V}_n)$ and $\mathcal{D}_n = O_p(\mathcal{V}_n)$.

Further, use (E13) and $w_i = (E[S_i S_i'])^{-1} S_i$ to yield:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i}-w_i\right)w_i' &= 2\left(E\left[S_i S_i'\right]\right)^{-1}\times\frac{1}{n}\sum_{i=1}^{n}\left(S_i(\hat{\gamma}_n)-S_i\right)S_i'\times\left(E\left[S_i S_i'\right]\right)^{-1} \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}S_i S_i'\times\left(E\left[S_i S_i'\right]\right)^{-1}\times o_p\left(1\right) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\left\{S_i(\hat{\gamma}_n)-S_i\right\}S_i'\times\left(E\left[S_i S_i'\right]\right)^{-1}\times o_p\left(1\right).
\end{aligned}$$

Observe $1/n\sum_{i=1}^{n}S_i S_i' \xrightarrow{p} E[S_i S_i']$ in view of square integrability. The arguments in Step 3.1 imply $||1/n\sum_{i=1}^{n}\{S_i(\hat{\gamma}_n) - S_i\}S_i'|| = o_p(\mathcal{V}_n)$. Therefore:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i}-w_i\right)w_i' = o_p(\mathcal{V}_n).$$

The same argument implies:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\hat{w}_{n,i}-w_i\right)\left(\hat{w}_{n,i}-w_i\right)' \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\left\{\left(\frac{1}{n}\sum_{i=1}^{n}S_i(\hat{\gamma}_n)S_i(\hat{\gamma}_n)'\right)^{-1}S_i(\hat{\gamma}_n)-(E[S_i S_i'])^{-1}S_i\right\}\right. \\
&\qquad\left.\times\left\{\left(\frac{1}{n}\sum_{i=1}^{n}S_i(\hat{\gamma}_n)S_i(\hat{\gamma}_n)'\right)^{-1}S_i(\hat{\gamma}_n)-(E[S_i S_i'])^{-1}S_i\right\}'\right] \\
&= (E[S_i S_i'])^{-1}\times\frac{1}{n}\sum_{i=1}^{n}\left\{S_i(\hat{\gamma}_n)-S_i\right\}\left\{S_i(\hat{\gamma}_n)-S_i\right\}'\times(E[S_i S_i'])^{-1} \\
&\quad +(E[S_i S_i'])^{-1}\times\frac{1}{n}\sum_{i=1}^{n}\left\{S_i(\hat{\gamma}_n)-S_i\right\}\left(S_i(\hat{\gamma}_n)-S_i\right)'\times o_p\left(1\right) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\left(S_i(\hat{\gamma}_n)-S_i\right)\left\{S_i(\hat{\gamma}_n)-S_i\right\}'\times(E[S_i S_i'])^{-1}\times o_p\left(1\right) \\
&\quad +(E[S_i S_i'])^{-1}\times\frac{1}{n}\sum_{i=1}^{n}\left\{S_i(\hat{\gamma}_n)-S_i\right\}S_i'\times o_p\left(1\right) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}S_i\left\{S_i(\hat{\gamma}_n)-S_i\right\}'\times(E[S_i S_i'])^{-1}\times o_p\left(1\right) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}S_i(\hat{\gamma}_n)S_i(\hat{\gamma}_n)'\times o_p\left(1\right) \\
&= (E[S_i S_i'])^{-1}\times\frac{1}{n}\sum_{i=1}^{n}\left\{S_i(\hat{\gamma}_n)-S_i\right\}\left\{S_i(\hat{\gamma}_n)-S_i\right\}'\times(E[S_i S_i'])^{-1}
\end{aligned}$$

$$+ (E[S_iS_i'])^{-1} \times \frac{1}{n}\sum_{i=1}^n \{S_i(\hat\gamma_n) - S_i\}(S_i(\hat\gamma_n) - S_i)' \times o_p(1)$$

$$+ \frac{1}{n}\sum_{i=1}^n (S_i(\hat\gamma_n) - S_i)\{S_i(\hat\gamma_n) - S_i\}' \times (E[S_iS_i'])^{-1} \times o_p(1) \times o_p(1).$$

By the proof of Step 3.1:

$$\frac{1}{n}\sum_{i=1}^n \{S_i(\hat\gamma_n) - S_i\}\{S_i(\hat\gamma_n) - S_i\}' = o_p(\mathcal{V}_n),$$

hence

$$\frac{1}{n}\sum_{i=1}^n (\hat w_{n,i} - w_i)(\hat w_{n,i} - w_i)' = o_p(\mathcal{V}_n).$$

This proves the required result.

**Step 5.** We now prove (E2). Add and subtract terms, and expand:

$$\frac{1}{n}\sum_{i=1}^n \left\{\left(\hat Z_{n,i}(\hat\gamma_n) I\left(\left|\hat Z_{n,i}(\hat\gamma_n)\right| < \hat Z^{(a)}_{n,(k_n)}(\hat\gamma_n)\right) + \left(\frac{n-k_n}{n}\right)\hat{\mathcal{B}}_n(\hat\gamma_n)\right) + \mathcal{D}_n' w_i\right\}^2$$

$$= \frac{1}{n}\sum_{i=1}^n \left\{\left(Z_i I(|Z_i| < c_n) + \left(\frac{n-k_n}{n}\right)\mathcal{B}_n\right) + \mathcal{D}_n' w_i\right\}^2$$

$$+ \frac{1}{n}\sum_{i=1}^n (Z_i(\hat\gamma_n) - Z_i)^2 I\left(\left|\hat Z_{n,i}(\hat\gamma_n)\right| < \hat Z^{(a)}_{n,(k_n)}(\hat\gamma_n)\right)$$

$$+ \frac{1}{n}\sum_{i=1}^n Z_i^2 \left\{I\left(\left|\hat Z_{n,i}(\hat\gamma_n)\right| < \hat Z^{(a)}_{n,(k_n)}(\hat\gamma_n)\right) - I(|Z_i| < c_n)\right\}$$

$$+ 2\frac{1}{n}\sum_{i=1}^n \{Z_i(\hat\gamma_n) - Z_i\} Z_i I\left(\left|\hat Z_{n,i}(\hat\gamma_n)\right| < \hat Z^{(a)}_{n,(k_n)}(\hat\gamma_n)\right)\left\{I\left(\left|\hat Z_{n,i}(\hat\gamma_n)\right| < \hat Z^{(a)}_{n,(k_n)}(\hat\gamma_n)\right) - I(|Z_i| < c_n)\right\}$$

$$+ \frac{1}{n}\sum_{i=1}^n \left[\left\{\left(Z_i I(|Z_i| < c_n) + \left(\frac{n-k_n}{n}\right)\mathcal{B}_n\right) + \mathcal{D}_n' w_i\right\} \times \left\{\hat Z_{n,i}(\hat\gamma_n) I\left(\left|\hat Z_{n,i}(\hat\gamma_n)\right| < \hat Z^{(a)}_{n,(k_n)}(\hat\gamma_n)\right) - Z_i I(|Z_i| < c_n)\right\}\right]$$

$$+ \frac{n-k_n}{n} \times \left(\frac{1}{n}\sum_{i=1}^n Z_i\right)^2$$

$$- 2\frac{1}{n}\sum_{i=1}^n Z_i \times \frac{1}{n}\sum_{i=1}^n \{Z_i(\hat\gamma_n) - Z_i\} I\left(\left|\hat Z_{n,i}(\hat\gamma_n)\right| < \hat Z^{(a)}_{n,(k_n)}(\hat\gamma_n)\right)$$

$$- 2\frac{1}{n}\sum_{i=1}^n Z_i \times \frac{1}{n}\sum_{i=1}^n Z_i \left\{I\left(\left|\hat Z_{n,i}(\hat\gamma_n)\right| < \hat Z^{(a)}_{n,(k_n)}(\hat\gamma_n)\right) - I(|Z_i| < c_n)\right\} I\left(\left|\hat Z_{n,i}(\hat\gamma_n)\right| < \hat Z^{(a)}_{n,(k_n)}(\hat\gamma_n)\right)$$

$$+ 2\left(\frac{n-k_n}{n}\right)\left\{\hat{\mathcal{B}}_n(\hat\gamma_n) - \mathcal{B}_n\right\} \times \frac{1}{n}\sum_{i=1}^n \left\{Z_i(\hat\gamma_n) I\left(\left|\hat Z_{n,i}(\hat\gamma_n)\right| < \hat Z^{(a)}_{n,(k_n)}(\hat\gamma_n)\right) - Z_i I(|Z_i| < c_n)\right\}$$

$$- 2\left(\frac{n-k_n}{n}\right)\left\{\hat{\mathcal{B}}_n(\hat\gamma_n) - \mathcal{B}_n\right\} \times \frac{1}{n}\sum_{i=1}^n Z_i \times \frac{n-k_n}{n}$$

$$+\left(\frac{n-k_n}{n}\right)\left\{\hat{\mathcal{B}}_n(\hat{\gamma}_n)-\mathcal{B}_n\right\}\times\left\{\frac{1}{n}\sum_{i=1}^{n}\left(Z_iI\left(|Z_i|<c_n\right)+\left(\frac{n-k_n}{n}\right)\mathcal{B}_n\right)+\mathcal{D}_n'\frac{1}{n}\sum_{i=1}^{n}w_i\right\}$$
$$+\left(\frac{n-k_n}{n}\right)^2\left\{\hat{\mathcal{B}}_n(\hat{\gamma}_n)-\mathcal{B}_n\right\}^2$$
$$=\frac{1}{n}\sum_{i=1}^{n}\left\{\left(Z_iI\left(|Z_i|<c_n\right)+\left(\frac{n-k_n}{n}\right)\mathcal{B}_n\right)+\mathcal{D}_n'w_i\right\}^2+\sum_{i=1}^{11}\mathcal{G}_{i,n}.$$

The proof of Theorem 3.4 verifies $\hat{\mathcal{B}}_n(\hat{\gamma}_n)-\mathcal{B}_n=o_p(\mathcal{V}_n/n^{1/2})=o_p(1)$, and $1/n\sum_{i=1}^{n}Z_i=O_p(\mathcal{L}_n/n^{1-1/\min\{\kappa,2\}})$ $=o_p(\mathcal{V}_n)$ by (E6). We need only show each $\mathcal{G}_{i,n}=o_p(\mathcal{V}_n^2)$.

Consider $\mathcal{G}_{1,n}$. A first order expansion, $L_p$-boundedness of $\sup_{\gamma\in\Gamma}\{|Z_i(\gamma)h_i(\gamma)|\times||(\partial/\partial\gamma)p_i(\gamma)||\}^2$ under B3$'$(i), and approximation (E4) yield:

$$\left|\frac{1}{\sigma_n n^{1/2}}\sum_{i=1}^{n}\left(Z_i(\hat{\gamma}_n)-Z_i\right)^2\left\{I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right|<\hat{Z}_{n,(k_n)}^{(a)}(\hat{\gamma}_n)\right)-I\left(|Z_i|<c_n\right)\right\}\right| \tag{E14}$$
$$\leq\frac{1}{\sigma_n n^{1/2}}\sum_{i=1}^{n}\sup_{\gamma\in\Gamma}\left\{|Z_i(\gamma)h_i(\gamma)|\times\left\|\frac{\partial}{\partial\gamma}p_i(\gamma)\right\|\right\}^2\left|I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right|<\hat{Z}_{n,(k_n)}^{(a)}(\hat{\gamma}_n)\right)-I\left(|Z_i|<c_n\right)\right|\times\|\hat{\gamma}_n-\gamma_0\|^2$$
$$=o_p(1).$$

Hence we may work with $I(|Z_i|<c_n)$.

By the definition of a derivative (E1):

$$\frac{1}{n}\sum_{i=1}^{n}\left(Z_i(\hat{\gamma}_n)-Z_i\right)^2I\left(|Z_i|<c_n\right) \tag{E15}$$
$$=\frac{1}{n}\sum_{i=1}^{n}\left(Z_ih_i\frac{\partial}{\partial\gamma'}p_i\left(\hat{\gamma}_n-\gamma_0\right)\right)^2I\left(|Z_i|<c_n\right)+o_p\left(\|\hat{\gamma}_n-\gamma_0\|\right)$$
$$=\left(\hat{\gamma}_n-\gamma_0\right)'\frac{1}{n}\sum_{i=1}^{n}S_iS_i'Z_i^2I\left(|Z_i|<c_n\right)\times\left(\hat{\gamma}_n-\gamma_0\right)+o_p\left(\|\hat{\gamma}_n-\gamma_0\|\right).$$

Under B2 $||\hat{\gamma}_n-\gamma_0||=O_p(1/n^{1/2})$. Since under B3$'$(ii) $S_i$ is $L_4$-bounded, the Cauchy-Schwartz inequality implies:

$$\left\|E\left[S_iS_i'Z_i^2I\left(|Z_i|<c_n\right)\right]\right\|\leq K\left(E\left[Z_i^4I\left(|Z_i|<c_n\right)\right]\right)^{1/2}.$$

Now use Karamata theory (B5) to yield $E[Z_i^4I(|Z_i|<c_n)]=O(1)$ if $\kappa>4$; $E[Z_i^4I(|Z_i|<c_n)]\sim K\ln(n)$ if $\kappa=4$; and $E[Z_i^4I(|Z_i|<c_n)]\sim K(n/k_n)^{4/\kappa-1}$ if $\kappa<4$. Therefore

$$E\left[Z_i^4I\left(|Z_i|<c_n\right)\right]=O(1)\text{ if }\kappa>4$$
$$E\left[Z_i^4I\left(|Z_i|<c_n\right)\right]=O\left(\ln(n)\right)\text{ if }\kappa=4$$
$$E\left[Z_i^4I\left(|Z_i|<c_n\right)\right]=O\left((n/k_n)^{4/\kappa-1}\right)\text{ if }\kappa<4.$$

Further:

$$\begin{aligned}\sigma_n^2 &= E\left[Z_i^2 I\left(|Z_i| < c_n\right)\right] = O(1) \text{ if } \kappa > 2 \\ &\sim K \ln\left(n\right)) \text{ if } \kappa = 2 \\ &\sim K(n/k_n)^{2/\kappa - 1} \text{ if } \kappa < 2.\end{aligned}$$

This implies

$$(\hat{\gamma}_n - \gamma_0)' \frac{1}{n}\sum_{i=1}^{n} S_i S_i' Z_i^2 I\left(|Z_i| < c_n\right) \times (\hat{\gamma}_n - \gamma_0) \tag{E16}$$

$$= O_p\left(\frac{1}{n^2}\sum_{i=1}^{n} S_i S_i' Z_i^2 I\left(|Z_i| < c_n\right)\right)$$

$$= O_p\left(\frac{\sigma_n^2}{n^{1/2}}\right) \text{ if } \kappa > 4$$

$$= O_p\left(\frac{1}{n}\left(\ln\left(n\right)\right)^{1/2}\right) = O_p\left(\frac{\sigma_n^2}{n^{1/2}}\right) \text{ if } \kappa = 4$$

$$= O_p\left(\frac{1}{n}(n/k_n)^{2/\kappa - 1/2}\right) = O\left(\frac{1}{n^{1/2}}\right) = O_p\left(\frac{\sigma_n^2}{n^{1/2}}\right) \text{ if } \kappa \in [2, 4)$$

$$= O_p\left(\frac{1}{n}(n/k_n)^{2/\kappa - 1/2}\right) = O_p\left(\frac{(n/k_n)^{1/2}}{n}(n/k_n)^{2/\kappa - 1}\right) = o_p\left(\frac{\sigma_n^2}{n^{1/2}}\right) \text{ if } \kappa \in (1, 2)\,.$$

Combine (E14)-(E16) with (E5) to yield as required:

$$\frac{1}{n}\sum_{i=1}^{n}\left(Z_i(\hat{\gamma}_n) - Z_i\right)^2 I\left(|Z_i| < c_n\right) = o_p\left(\sigma_n^2\right) = o_p(\mathcal{V}_n^2).$$

The next terms $\mathcal{G}_{2,n}$, $\mathcal{G}_{3,n}$, and $\mathcal{G}_{4,n}$ satisfy $\mathcal{G}_{i,n} = o_p(\mathcal{V}_n/n^{1/2})$ by using approximation (E4) combined with arguments developed above. Further, $\mathcal{G}_{5,n} \sim (1/n\sum_{i=1}^{n} Z_i)^2 = o_p(1) = o_p(\mathcal{V}_n)$.

Next, $\mathcal{G}_{6,n} = o_p(\mathcal{V}_n/n^{1/2})$ follows from (E6), and

$$\frac{1}{n}\sum_{i=1}^{n}\left\{Z_i(\hat{\gamma}_n) - Z_i\right\} I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}_{n,(k_n)}^{(a)}(\hat{\gamma}_n)\right) = o_p(\mathcal{V}_n/n^{1/2})$$

by an argument identical to the proof of $\mathcal{A}_{n,3} = o_p(\mathcal{V}_n/n^{1/2})$ in Step 1.

Apply (E4) and (E6) to yield:

$$|\mathcal{G}_{7,n}| \leq \left|\frac{1}{n}\sum_{i=1}^{n} Z_i\right| \times \frac{1}{n}\sum_{i=1}^{n}|Z_i|\left|I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}_{n,(k_n)}^{(a)}(\hat{\gamma}_n)\right) - I\left(|Z_i| < c_n\right)\right| = o_p(\mathcal{V}_n/n^{1/2}).$$

Consider $\mathcal{G}_{8,n}$, use (E4) and the argument for $\mathcal{G}_{6,n}$ to deduce:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{Z_i(\hat{\gamma}_n)I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) - Z_iI\left(|Z_i| < c_n\right)\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n} Z_i\left\{I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) - I\left(|Z_i| < c_n\right)\right\}$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\left\{Z_i(\hat{\gamma}_n) - Z_i\right\}I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{Z_i(\hat{\gamma}_n) - Z_i\right\}I\left(\left|\hat{Z}_{n,i}(\hat{\gamma}_n)\right| < \hat{Z}^{(a)}_{n,(k_n)}(\hat{\gamma}_n)\right) = o_p\left(\mathcal{V}_n\right).$$

Term $\mathcal{G}_{9,n} = o_p(1) = o_p(\mathcal{V}_n)$ given $\hat{\mathcal{B}}_n(\hat{\gamma}_n) - \mathcal{B}_n = o_p(\mathcal{V}_n/n^{1/2}) = o_p(1)$ and $1/n\sum_{i=1}^{n} Z_i = o_p(1)$.

Finally, for $\mathcal{G}_{10,n}$, by construction $Z_iI(|Z_i| < c_n) + ((n-k_n)/n)\mathcal{B}_n$ has a zero mean. Hence by independence and (E5):

$$\frac{1}{n}\sum_{i=1}^{n} Z_iI\left(|Z_i| < c_n\right) + \left(\frac{n-k_n}{n}\right)\mathcal{B}_n = O_p\left(\sigma_n/n^{1/2}\right) = O_p\left(\mathcal{V}_n/n^{1/2}\right).$$

Similarly, by assumption $w_i$ is iid, has a zero mean and is square integrable, hence $1/n\sum_{i=1}^{n} w_i = O_p(1/n^{1/2})$. By the proof of Theorem 3.1.b $\mathcal{D}_n = O(\mathcal{V}_n)$. Therefore

$$\frac{1}{n}\sum_{i=1}^{n}\left(Z_iI\left(|Z_i| < c_n\right) + \left(\frac{n-k_n}{n}\right)\mathcal{B}_n\right) + \mathcal{D}_n^{\prime}\frac{1}{n}\sum_{i=1}^{n} w_i = O_p(\mathcal{V}_n/n^{1/2}).$$

**Step 6.** In this last step we verify (E3). By definition

$$\begin{aligned}\mathcal{V}_n^2 &\equiv E[\vartheta_{n,i}^2] \\ &= E\left[\left\{\left(Z_iI\left(|Z_i| < c_n\right) + \left(\frac{n-k_n}{n}\right)\mathcal{B}_n\right) + \mathcal{D}_n^{\prime}w_i\right\}^2\right] \\ &= E\left[\left\{\left(Z_iI\left(|Z_i| < c_n\right) - Z_iI\left(|Z_i| < c_n\right)\right) + \mathcal{D}_n^{\prime}w_i\right\}^2\right].\end{aligned}$$

By construction $\mathcal{Y}_{n,i} \equiv \vartheta_{n,i}^2/\mathcal{V}_n^2$ is $L_1$-bounded uniformly in $n$, and iid over $1 \leq i \leq n$ for each $n$. Moreover, by Step 6.1 $\mathcal{Y}_{n,i}$ is uniformly integrable. The claim therefore follows from Theorem 2 in Andrews (1988).

**Step 6.1.** We will prove that for each $\varepsilon > 0$ there exists a $\mathcal{K}_\varepsilon > 0$ such that $\sup_{n\in\mathbb{N}} E[\mathcal{Y}_{n,i}I(\mathcal{Y}_{n,i} > \mathcal{K}_\varepsilon)] \leq \varepsilon$, which implies uniform integrability.

By sub-additivity and $|\mathcal{D}_n^{\prime}w_i| \leq ||\mathcal{D}_n|| \times ||w_i||$:

$$E\left[\mathcal{Y}_{n,i}I\left(\mathcal{Y}_{n,i} > \mathcal{K}_\varepsilon\right)\right] = \int_{\mathcal{K}_\varepsilon}^{\infty} P\left(\frac{\left\{\left(Z_iI\left(|Z_i| < c_n\right) - Z_iI\left(|Z_i| < c_n\right)\right) + \mathcal{D}_n^{\prime}w_i\right\}^2}{\mathcal{V}_n^2} > u\right)du$$

$$
\begin{aligned}
&\leq \int_{\mathcal{K}_\varepsilon}^{\infty} P\left(\frac{|Z_i| I(|Z_i| < c_n)}{\mathcal{V}_n} > u^{1/2}\right) du + \int_{\mathcal{K}_\varepsilon}^{\infty} I\left(\frac{|E[Z_i I(|Z_i| < c_n)]|}{\mathcal{V}_n} > u^{1/2}\right) du \\
&\quad + \int_{\mathcal{K}_\varepsilon}^{\infty} P\left(\frac{|\mathcal{D}_n' w_i|}{\mathcal{V}_n} > u^{1/2}\right) du \\
&\leq \int_{\mathcal{K}_\varepsilon}^{\infty} P\left(\frac{|Z_i| I(|Z_i| < c_n)}{\mathcal{V}_n} > u^{1/2}\right) du + \int_{\mathcal{K}_\varepsilon}^{\infty} I\left(E|Z_i| > \mathcal{V}_n u^{1/2}\right) du \qquad \text{(E17)} \\
&\quad + \int_{\mathcal{K}_\varepsilon}^{\infty} P\left(\frac{\|\mathcal{D}_n\|}{\mathcal{V}_n} \|w_i\| > u^{1/2}\right) du.
\end{aligned}
$$

We need only show each integral is bounded by $\varepsilon/3$ for some $\mathcal{K}_\varepsilon$.

Consider the second integral. Assumption A5 states $\liminf_{n\to\infty} \mathcal{V}_n^2 > 0$, hence for each $\varepsilon$ there exists a $\mathcal{K}_{1,\varepsilon}$ that implies:

$$
\int_{\mathcal{K}_{1,\varepsilon}}^{\infty} I\left(E|Z_i| > \mathcal{V}_n u^{1/2}\right) du \leq \int_{\mathcal{K}_{1,\varepsilon}}^{\infty} I\left(\frac{E|Z_i|}{K} > u^{1/2}\right) du = \max\left\{0, \left(\frac{E|Z_i|}{K}\right)^2 - \mathcal{K}_{1,\varepsilon}\right\} \leq \frac{\varepsilon}{3}. \qquad \text{(E18)}
$$

Further, for the third integral the proof of Theorem 3.1.b shows $||\mathcal{D}_n|| = O(\mathcal{V}_n)$ hence $\liminf_{n\to\infty} \mathcal{V}_n / ||\mathcal{D}_n|| > 0$. Therefore, for any $\mathcal{K}_\varepsilon > 0$:

$$
\int_{\mathcal{K}_\varepsilon}^{\infty} P\left(\frac{\|\mathcal{D}_n\|}{\mathcal{V}_n} \|w_i\| > u^{1/2}\right) du \leq \int_{\mathcal{K}_\varepsilon}^{\infty} P\left(\|w_i\| > K u^{1/2}\right) du.
$$

Since $||w_i||$ is iid and square integrable, it is uniformly square integrable. For each $\varepsilon$ there exists a $\mathcal{K}_{2,\varepsilon}$ such that:

$$
E\left[\left(\frac{\|w_i\|}{K}\right)^2 I\left(\left(\frac{\|w_i\|}{K}\right)^2 > \mathcal{K}_{2,\varepsilon}\right)\right] = \int_{\mathcal{K}_{2,\varepsilon}}^{\infty} P\left(\left(\frac{\|w_i\|}{K}\right)^2 > u\right) du \leq \frac{\varepsilon}{3},
$$

hence the third integral is bounded:

$$
\int_{\mathcal{K}_{2,\varepsilon}}^{\infty} P\left(\frac{\|\mathcal{D}_n\|}{\mathcal{V}_n} \|w_i\| > u^{1/2}\right) du \leq \frac{\varepsilon}{3}. \qquad \text{(E19)}
$$

Finally, for the first integral we have by Theorem 3.1.b $\mathcal{V}_n^2 \sim K\sigma_n^2$ for some $K > 0$, where $K = 1$ if $E[Z_i^2] = \infty$. If $E[Z_i^2] < \infty$ then by the iid property $Z_i$ is uniformly square integrable. Use $\mathcal{V}_n^2 \sim K\sigma_n^2 \to KE[Z_i^2]$ to deduce for each $\varepsilon$ there exists a $\mathcal{K}_{3,\varepsilon}^{(\kappa>2)}$ such that:

$$
\begin{aligned}
\int_{\mathcal{K}_{3,\varepsilon}^{(\kappa>2)}}^{\infty} P\left(\frac{|Z_i| I(|Z_i| < c_n)}{\mathcal{V}_n} > u^{1/2}\right) du &\leq \int_{\mathcal{K}_{3,\varepsilon}^{(\kappa>2)}}^{\infty} P\left(\frac{Z_i^2}{\mathcal{V}_n^2} > u\right) du \qquad \text{(E20)} \\
&\sim \int_{\mathcal{K}_{3,\varepsilon}^{(\kappa>2)}}^{\infty} P\left(Z_i^2 > E\left[Z_i^2\right] u\right) du
\end{aligned}
$$

$$= \frac{1}{E\left[Z_i^2\right]} \int_{\mathcal{K}_{3,\varepsilon}^{(\kappa>2)} E[Z_i^2]}^{\infty} P\left(Z_i^2 > u\right) du \leq \frac{\varepsilon}{3}.$$

If $E[Z_i^2] = \infty$ then use $\mathcal{V}_n^2 \sim \sigma_n^2$ and a change of variables to write for any $\mathcal{K}_\varepsilon > 0$:

$$\begin{aligned}
\int_{\mathcal{K}_\varepsilon}^{\infty} P\left(\frac{|Z_i| \, I\left(|Z_i| < c_n\right)}{\mathcal{V}_n} > u^{1/2}\right) du &\sim \int_{\mathcal{K}_\varepsilon}^{c_n^2/\sigma_n^2} P\left(\frac{Z_i^2}{\sigma_n^2} > u\right) du \\
&= K \int_{\mathcal{K}_\varepsilon}^{c_n^2/\sigma_n^2} P\left(Z_i^2 > \sigma_n^2 u\right) du \\
&= K \frac{1}{\sigma_n^2} \int_{\mathcal{K}_\varepsilon \sigma_n^2}^{c_n^2} P\left(Z_i^2 > v\right) dv \\
&\leq K \frac{\max\left\{0, c_n^2 - \mathcal{K}_\varepsilon \sigma_n^2\right\}}{\sigma_n^2} \\
&= K \frac{c_n^2}{\sigma_n^2} \left(\frac{\max\left\{0, 1 - \mathcal{K}_\varepsilon \sigma_n^2 / c_n^2\right\}}{\sigma_n^2}\right).
\end{aligned}$$

Recall $c_n = K(n/k_n)^{1/\kappa}$ by (B4). Now use the A3 power law property to deduce the following by Karamata theory (B5): if the tail index $\kappa < 2$ then $\sigma_n^2 \sim K c_n^2 P(|Z_i| > c_n) = K c_n^2 k_n/n$ hence $c_n^2/\sigma_n^2 \sim K n/k_n$; and if $\kappa = 2$ then $\sigma_n^2 \sim K \ln(n)$ hence $c_n^2/\sigma_n^2 \sim K(n/k_n)/\ln(n) = o((n/k_n))$. Finally, $P(|Z_i| > c_n) = k_n/n \to 0$ implies $c_n \to \infty$. Hence, for any $\mathcal{K}_{3,\varepsilon}^{(\kappa \leq 2)} > 0$

$$\int_{\mathcal{K}_{3,\varepsilon}^{(\kappa\leq 2)}}^{\infty} P\left(\frac{|Z_i| \, I\left(|Z_i| < c_n\right)}{\mathcal{V}_n} > u^{1/2}\right) du = O\left(\frac{n}{k_n}\right) \frac{1}{c_n^2 \left(k_n/n\right)} \left(1 + o\left(1\right)\right) = O\left(\frac{1}{c_n}\right) \to 0. \tag{E21}$$

Since $\mathcal{K}_{3,\varepsilon}^{(\kappa \leq 2)}$ is arbitrary, put $\mathcal{K}_{3,\varepsilon}^{(\kappa \leq 2)} = \mathcal{K}_{3,\varepsilon}^{(\kappa > 2)} = \mathcal{K}_{3,\varepsilon}$. Now set

$$\mathcal{K}_\varepsilon = \max\left\{\mathcal{K}_{1,\varepsilon}, \mathcal{K}_{2,\varepsilon}, \mathcal{K}_{3,\varepsilon}\right\} > 0. \tag{E22}$$

Together, (E18)-(E21) and monotonicity of probability measures imply for any $\varepsilon > 0$ and $\mathcal{K}_\varepsilon$ in (E22) that each integral in (E17) is bounded by $\varepsilon/3$. This completes the proof. $\mathcal{QED}$.

# F Probability Tail Decay - Threshold Crossing Latent Variable Model for Treatment Assignment

In this section, using a general environment we model tail decay for $Z$, the variable that point identifies the ATE. We work in a conventional latent variable threshold crossing framework with separable error and covariate for treatment assignment. In this setting we characterize the distribution tails of the variable that identifies the ATE. This framework is widely used (see Vytlacil (2002)) and hence is beneficial for appreciating why, where and how our estimator is robust to limited overlap.

The results here provide the required framework from which we derive the various examples in the main paper. See Sections F.1-F.4. Proofs are presented Section F.5. Assume without loss of generality that the ATE is:

$$\theta = 0.$$

Denote by $E_{Y_i}$ expectations with respect to the measure induced by $Y_i$. Let $\kappa \equiv \arg\sup_{\alpha>0}\{E|Z_i|^\alpha < \infty\}$. Recall $a \wedge b \equiv \min\{a, b\}$. We assume various distributions are smooth for the convenience of all subsequent derivations.

**D5.** The distributions of $DY/p(X)$ and $(1-D)Y/(1-p(X))$ are absolutely continuous on their support, and $p(X)|Y_1$ and $p(X)|Y_0$ have absolutely continuous distributions with Borel measurable density functions $f_{p(X)|Y_1}$ and $f_{p(X)|Y_0}$ for each $p(x) \in (0, 1)$ and $Y_1, Y_0$-*a.s.*

**Theorem F.1.** *Let $c > 1$ be arbitrary. Under A1, A2′ and D5:*

$$P(|Z| > c) = E_{Y_1}\left[\int_0^{\frac{|Y_1|}{c}\wedge 1} r f_{p(X)|Y_1}(r)dr\right] + E_{Y_0}\left[\int_{\left(1-\frac{|Y_0|}{c}\right)\vee 0}^1 (1-r) f_{p(X)|Y_0}(r)dr\right] \tag{F1}$$

$$\frac{\partial}{\partial c}P(|Z| > c) = -\frac{1}{c^3}E_{Y_1}\left[I(|Y_1| \leq c)Y_1^2 f_{p(X)|Y_1}\left(\frac{|Y_1|}{c}\right)\right] \tag{F2}$$
$$- \frac{1}{c^3}E_{Y_0}\left[I(|Y_0| \leq c)Y_0^2 f_{p(X)|Y_0}\left(1-\frac{|Y_0|}{c}\right)\right] = -\frac{1}{c^3}d(c).$$

If $Z$ had a Paretian tail, then $P(|Z| > c) \sim dc^{-\kappa}$ as $c \to \infty$ hence $(\partial/\partial c)P(|Z| > c) \sim -\kappa dc^{-\kappa-1}$. Property (F2) suggests that $Z$ has a tail structure similar to a power law with index $\kappa = 2$, but with a multiplicative scale $d(c)$ governed by the threshold $c$ and the distributions of $p(X), Y_0$ and $Y_1$.

We need the conditional density $f_{p(X)|Y_j}(r)$ in order to characterize $d(c)$. We therefore consider the popular latent variable threshold crossing framework for treatment assignment:

$$D = I(\alpha + \beta X - U \geq 0).$$

Obviously in practice $\beta$ and $V[U]$ cannot both be identified, hence either $\beta = 1$ or $Var(U) = 1$ are standard assumptions. Trivially the standardized form $D = I(X - u \geq 0)$ with $u = U/\beta$ is synonymous to $\beta = 1$ and $\{U, X\}$ having different tail thicknesses. We allow $\beta \gtreqless 1$ for ease, but everything that follows is synonymous to fixing $\beta = 1$ and inspecting the relative probability tails of $\{U, X\}$.

We assume for simplicity $U$ is independent of $X$, $Y_1$ and $Y_0$. Also, normalize $E[U] = 0$ and $Var(U) = 1$ for the rest of the paper. The assumption that the error $U$ is additively separable and independent of $X$ has implications on the treatment assignment (cf. Vytlacil (2002)). Generality is also lost due to the specific index structure $\alpha + \beta X$ , but these help to abstract from issues peripheral to the demonstration of the power law tail decay. Without loss of further generality, take $\beta > 0$.[2]

[2]Note that $\beta = 0$ implies $p(X) = F_U(\alpha) = p$ (constant) and as a result, under assumptions A1 and A2, $\theta = E[Y|D = 1] - E[Y|D = 0]$ meaning that there is no need for an IPW estimator. While its variance will increase with the proximity of $p$

**D6.** $U$ has an absolutely continuous distribution with density function $f_U$. $X$ has support $\mathcal{X}$. $X|Y_1$ and $X|Y_0$ have absolutely continuous distributions with Borel measurable density functions $f_{X|Y_1}(x)$ and $f_{X|Y_0}(x)$ for each $x \in \mathcal{X}$ and $a.s.$ with respect to $Y_1, Y_0$.

By independence of $U$ and $X$:

$$p(X) = P(D = 1|X) = P(\alpha + \beta X \geq U) = F_U(\alpha + \beta X)$$

hence under D6 it follows for $j = 0, 1$:

$$f_{p(X)|Y_j}(r) = f_{X|Y_j}\left(\frac{F_U^{-1}(r) - \alpha}{\beta}\right)\frac{1}{\beta}\frac{1}{f_U\left(F_U^{-1}(r)\right)} \text{ where } r \in (0,1).$$

The result in (F2) can therefore be written as

$$\frac{\partial}{\partial c}P(|Z| > c) = -\frac{1}{c^3}\frac{1}{\beta}\mathcal{F}(\alpha,\beta,c) \;\text{ where }\; \mathcal{F}(\alpha,\beta,c) \equiv \mathcal{F}_1(\alpha,\beta,c) + \mathcal{F}_0(\alpha,\beta,c), \tag{F3}$$

and

$$\begin{aligned}\mathcal{F}_1(\alpha,\beta,c) &\equiv E_{Y_1}\left[Y_1^2 I(|Y_1| \leq c)\, f_{X|Y_1}\left(\frac{F_U^{-1}\left(\frac{|Y_1|}{c}\right) - \alpha}{\beta}\right)\frac{1}{f_U\left(F_U^{-1}\left(\frac{|Y_1|}{c}\right)\right)}\right]\\ \mathcal{F}_0(\alpha,\beta,c) &\equiv E_{Y_0}\left[Y_0^2 I(|Y_0| \leq c)\, f_{X|Y_0}\left(\frac{F_U^{-1}\left(1 - \frac{|Y_0|}{c}\right) - \alpha}{\beta}\right)\frac{1}{f_U\left(F_U^{-1}\left(1 - \frac{|Y_0|}{c}\right)\right)}\right].\end{aligned}$$

It remains to deduce power law properties as a consequence of the behavior of $\mathcal{F}(\alpha,\beta,c)$ as $c \to \infty$. The behavior of the ratios $f_{X|Y_j}((q_1 - \alpha)/\beta)/f_U(q_j)$ and therefore the relative tail decay of $X|Y_j$ and $U$ plays a key roll, where for $j = 0, 1$ the $q_j's$ are quantiles

$$q_0 \equiv F_U^{-1}(1 - |Y_0|/c) \;\text{ and }\; q_1 \equiv F_U^{-1}(|Y_1|/c) \text{ for } |Y_j|/c \leq 1. \tag{F4}$$

We demonstrate below by example how these two ratios influence the tail behavior of $Z$. Given the simplicity of the setup and a similar setting in Busso, DiNardo, and McCrary (2009) and Khan and Tamer (2010a), we focus on the cases where $Y_j \perp X, U$, and either $\{U, X\}$ are identically distributed, or normally or Laplace distributed. Further, in order to avoid notational clutter we simply assume $\alpha = 0$:

$$D = I(\beta X - U \geq 0). \tag{F5}$$

to 0 or 1, the IPW estimator does not, however, suffer from the limited overlap problem asymptotically as long as the constant $p \in (0, 1)$.

In the settings discussed below, $\alpha = 0$ implies $Z$ has a symmetric distribution about the ATE $\theta$. Allowing $\alpha \neq 0$ merely generalizes to tail asymmetry. In practice, a more general setting will clearly be desired. The following derivations serve as a basic groundwork for showing under limited overlap why heavy tails arise, and how sensitive they are to $\beta$.

### F.1 Example: iid Error and Covariate

A brief example sheds some light on how the covariate slope $\beta$ and the relative tail behavior of $X$ and $U$ affects the tail behavior of $Z$ . In Khan and Tamer (2010a), following Lewbel (1997), the latent variable case treated is the standardization $\beta = 1$.

Then

$$\frac{f_{X|Y_j}((q_1 - \alpha)/\beta)}{f_U(q_j)} = \frac{f_{X|Y_j}(q_j)}{f_U(q_j)},$$

and since $Y_j \perp X$ this further reduces to $f_X(q_j)/f_U(q_j)$. Thus, if $X$ and $U$ have the same densities, then

$$\mathcal{F}_j(0,1,c) \equiv E_{Y_j}\left[Y_j^2 I\left(|Y_j| \leq c\right)\right],$$

and if $Y_j$ has a finite variance then by dominated convergence $\lim_{c\to\infty} \mathcal{F}_j(0,1,c) = E[Y_j^2]$. This implies by (F3) that

$$\frac{\partial}{\partial c} P\left(|Z| > c\right) \sim -c^{-3}\left(E\left[Y_0^2\right] + E\left[Y_1^2\right]\right),$$

hence $Z$ has a Paretian tail with index 2. This proves the following.

**Theorem F.2.** *Let the treatment assignment be (F5) with $\beta = 1$, let $Y_j \perp X, U$, and let $\{U, X\}$ be iid. Then $P(|Z| > c) = dc^{-2}(1 + o(1))$ where $d = (1/2)(E[Y_0^2] + E[Y_1^2])$.*

**Remark 2.** By dominated convergence the same conclusion follows when $f_X(r)/f_U(r) \to (0,\infty)$ as $|r| \to \infty$. Hence, the tail index is identically 2 when $X$ and $U$ have the same rate of distribution tail decay.

Two simple lessons are $(i)$ when $Y_j \perp X, U$, and $X$ and $U$ have the same impact on $\mathcal{F}_j(\alpha,\beta,c)$ for $j = 0,1$, then $Z$ is heavy tailed with a hairline infinite variance; and $(ii)$ lighter or heavier tails are driven by tail differences in $X$ and $U$, and $\beta \gtreqless 1$, an issue largely ignored in the literature on IPW estimators for the ATE. Notice $(i)$ explains Khan and Tamer (2010a, Section 4.1)'s finding that their tail-trimmed ATE estimator has a $o(n^{1/2})$ rate of convergence when $X$ and $U$ are identically logit distributed: $Z$ has an infinite variance hence negligible trimming results in sub-$n^{1/2}$ convergence (Csörgo, Horváth, and Mason, 1986; Hahn, Weiner, and Mason, 1991; Hill, 2015).

### F.2 Example: Laplace Error and Covariate

Let $(Y_1, Y_0, X, U)$ be independently distributed Laplace with mean 0 and variance 1. The cdf is

$$F(r) = \frac{1}{2}e^{\sqrt{2}r} \text{ if } r \leq 0 \text{ and } F(r) = 1 - \frac{1}{2}e^{-\sqrt{2}r} \text{ if } r > 0, \tag{F6}$$

hence $f(r) = (1/\sqrt{2})e^{-\sqrt{2}|r|}$.

**Theorem F.3.** *Let the treatment assignment be (F5), let $Y_j \perp X, U$, and let $\{U, X\}$ be iid with cdf (F6). $Z$ is symmetrically distributed about zero, and $P(|Z| > c) = d(\beta)c^{-(1+1/\beta)}(1 + O(e^{-c/4}))$ where $d(\beta) \equiv \beta^{-1}2^{1/(2\beta)} \int_0^\infty \exp\{-y\} y^{1+1/\beta} dy \in (0, \infty)$ for all $\beta \in (0, \infty)$.*

**Remark 3.** The distribution is symmetric due to the treatment assignment location $\alpha = 0$, independence $Y_j \perp X, U$, and symmetry about zero for the distributions of all variables $(Y_1, Y_0, X, U)$. The tail index $1 + 1/\beta > 1$, so the ATE always exists.[3] As $\beta$ increases the signal $\beta X$ is stronger, ceteris parabus, hence the probability tails of $Z$ become monotonically heavier.

**Remark 4.** The second order term $O(e^{-c/4})$ is $O(c^{-\eta})$ for any $\eta > 0$. This implies power law assumption A3′ holds, and since $\eta > 0$ is arbitrary then any fractile $m_n \to \infty$ and $m_n = o(n)$ can in theory be used for tail exponent estimation in the bias-corrected ATE estimator.

### F.3 Example: Normal Error and Covariate

Repeat the setup above, except assume $(Y_1, Y_0, X, U)$ are independently distributed $N(0, 1)$. Then

$$\begin{aligned}
\mathcal{F}_j(0, \beta, c) &= E_{Y_j}\left[Y_j^2 I\left(|Y_j| \leq c\right) \frac{f_{X|Y_j}\left(q_j/\beta\right)}{f_U\left(q_j\right)}\right] = E_{Y_j}\left[Y_j^2 I\left(|Y_j| \leq c\right) \frac{f_X\left(q_j/\beta\right)}{f_U\left(q_j\right)}\right] \\
&= \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^c y^2 \exp\left\{-\frac{y^2}{2}\right\} \exp\left\{-\frac{1-\beta^2}{2\beta^2} q_j^2\right\} dy.
\end{aligned}$$

Let $\Phi(z)$ denote the standard normal cdf. In this setting, it follows by a change of variables $z = y/c$ that, e.g.,

$$\mathcal{F}_1(0, \beta, c) = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^1 c^3 z^2 \exp\left\{-\frac{c^2 z^2}{2}\right\} \exp\left\{-\frac{1-\beta^2}{2\beta^2} (\Phi^{-1}(z))^2\right\} dz. \tag{F7}$$

**Theorem F.4.** *Let the treatment assignment be (F5), let $Y_j \perp X, U$, and let $\{U, X\}$ be iid $N(0, 1)$. $Z$ is symmetrically distributed about zero, and $P(|Z| > c) = d(\beta)c^{-(1+1/\beta^2))}(1 + o(e^{-c/2}))$ where $d(\beta) \equiv \beta^{-1}(2\pi)^{-K(1-1/\beta^2)} \int_0^\infty u^2 \exp\left\{-u^2/2\right\} u^{-K(1-1/\beta^2)}$ for some $K > 0$.*

**Remark 5.** The higher order term $o(e^{-c/2})$ is $O(c^{-\eta})$ for any $\eta > 0$, hence again A3′ holds and any $m_n \to \infty$ and $m_n = o(n)$ is valid.

**Remark 6.** Although exponential tails in general will lead to results similar to the Laplace case, there are concrete differences worth noting. In particular, in the present case $Z$ has a Paretian tail with index $\kappa = 1 + 1/\beta^2$ which is more sensitive to changes in $\beta$ than the Laplace index $1 + 1/\beta$ is when $\beta \in (0, 2)$.

### F.4 Example: Non-iid Error and Covariate

The preceding examples exclude, for simplicity, the case where the errors and covariates have different distributions. Consider $Y_j \perp X, U$, as above, hence

$$\mathcal{F}_j(0, \beta, c) = E_{Y_j}\left[Y_j^2 I\left(|Y_j| \leq c\right) \frac{f_X(q_j/\beta)}{f_U(q_j)}\right],$$

[3] This is trivial: $\theta = E[Y_1 - Y_0]$ exists because the $Y_j' s$ are iid Laplace, and therefore integrable, hence the tail index must be greater than one.

where $q_j$ are defined by (F4). Then, for a given $\beta > 0$, a relatively heavier (thinner) tailed error $U$ is associated with thinner (heavier) tails in $Z$. For example, if $(Y_0, Y_1, X)$ are Laplace and $U$ is normal then $Z$ is heavier tailed than if all $(Y_0, Y_1, X, U)$ are Laplace, and additionally in this case if $\beta = 1$ then $\kappa < 2$. Conversely, if $(Y_0, Y_1, X)$ are normal and $U$ is Laplace or has a power law distribution tail, then $Z$ is thinner tailed than if all $(Y_0, Y_1, X, U)$ are normal. A similar scenario arises if $(Y_0, Y_1, X)$ and $U$ belong to the same distribution class but have different variances.

As a final brief example, consider Khan and Tamer (2010a, Section 4.1)'s example with $\beta = 1$, $Y_j \perp X, U$, logistic $X$ and normal $U$. Since logistic tails are heavier the normal tails, $f_X(q_j/\beta)/f_U(q_j) \to \infty$ such that $\kappa < 2$. This explains their derived sub-$(n/\ln(n))^{1/2}$ rate of convergence for $\hat{\theta}_n^{(tx)}$ with minimum mse thresholds. However, if $U$ is logistic and $X$ is normal then $f_X(q_j/\beta)/f_U(q_j) \to 0$ and $Z$ has a power law with tail index $\kappa > 2$, hence identification is "regular". Our simulation experiments with Laplace and normal $\{U, X\}$ clearly demonstrate these opposite cases.

## F.5 Proofs

**Proof of Theorem F.1.** By mutual exclusivity of the events $D = 1$ and $D = 0$ it follows:

$$P(|hY| > c) = P\left(\left|\frac{DY_1}{p(X)} - \frac{(1-D)Y_0}{1-p(X)}\right| > c\right) = P\left(\left|\frac{DY_1}{p(X)}\right| > c\right) + P\left(\left|\frac{(1-D)Y_0}{1-p(X)}\right| > c\right). \tag{F8}$$

Observe that:

$$\begin{aligned} P\left(\frac{DY_1}{p(X)} > c\right) &= E_{Y_1}\left[I\left(\frac{Y_1}{p(X)} > c\right)p(X)|Y_1\right] \\ &= E_{Y_1}\left(E\left[p(X)I\left(p(X) < \frac{Y_1}{c} \wedge 1\right)|Y_1\right]\right) = E_{Y_1}\left[\int_0^{\frac{Y_1}{c}\wedge 1} r f_{p(X)|Y_1}(r|y)\, dr\right] \end{aligned}$$

and

$$P\left(\frac{DY_1}{p(X)} < -c\right) = E_{Y_1}\left[\int_0^{\frac{-Y_1}{c}\wedge 1} r f_{p(X)|Y_1}(r|y)\, dr\right],$$

hence

$$P\left(\left|\frac{DY_1}{p(X)}\right| > c\right) = E_{Y_1}\left[\int_0^{\frac{|Y_1|}{c}\wedge 1} r f_{p(X)|y}(r|y)\, dr\right]. \tag{F9}$$

By the same argument

$$P\left(\left|\frac{(1-D)Y_0}{1-p(X)}\right| > c\right) = E_{Y_0}\left[\int_{\left(1-\frac{|Y_0|}{c}\right)\vee 0}^{1} (1-r) f_{p(X)|Y_0}(r) dr\right]. \tag{F10}$$

Differentiate both sides of (F9) and (F10) with respect to $c$ to deduce:

$$\frac{\partial}{\partial c} P\left(\left|\frac{DY_1}{p(X)}\right| > c\right) \tag{F11}$$

$$= \frac{\partial}{\partial c}\int_{|Y_1|>c}\left\{\int_0^1 r f_{p(X)|Y_1}(r)dr\right\} f_{Y_1}(y)dy$$
$$+ \frac{\partial}{\partial c}\int_{|Y_1|\le c}\left\{\int_0^{\frac{|Y_1|}{c}} r f_{p(X)|Y_1}(r)dr\right\} f_{Y_1}(y)dy$$
$$= \frac{\partial}{\partial c}\int_{-\infty}^{-c}\left\{\int_0^1 r f_{p(X)|Y_1}(r)dr\right\} f_{Y_1}(y)dy + \frac{\partial}{\partial c}\int_{-c}^{\infty}\left\{\int_0^1 r f_{p(X)|Y_1}(r)dr\right\} f_{Y_1}(y)dy$$
$$+ \frac{\partial}{\partial c}\int_{-c}^{c}\left\{\int_0^{\frac{|Y_1|}{c}} r f_{p(X)|Y_1}(r)dr\right\} f_{Y_1}(y)dy$$
$$= -E[p(X)|Y_1=-c]f_{Y_1}(-c) - E[p(X)|Y_1=c]f_{Y_1}(c) + E[p(X)|Y_1=-c]f_{Y_1}(-c)$$
$$+ E[p(X)|Y_1=c]f_{Y_1}(c) + \int_{-c}^{c}\frac{\partial}{\partial c}\left\{\frac{|Y_1|}{c}\right\}\frac{|Y_1|}{c} f_{p(X)|Y_1}\left(\frac{|Y_1|}{c}\right) f_{Y_1}(y)dy$$
$$= -\frac{1}{c^3}E_{Y_1}\left[Y_1^2 f_{p(X)|Y_1}\left(\frac{|Y_1|}{c}\right)1(|Y_1|\le c)\right],$$

and
$$\frac{\partial}{\partial c}P\left(\left|\frac{(1-D)Y_0}{1-p(X)}\right|>c\right) = -\frac{1}{c^3}E_Y\left[Y_0^2 f_{p(X)|Y_0}\left(1-\frac{|Y_0|}{c}\right)1(|Y_0|\le c)\right]. \tag{F12}$$

Now combine (F8)-(F12) to prove the claims. $\mathcal{QED}$ .

**Proof of Theorem F.3.** We only characterize $\mathcal{F}_1(\alpha,\beta,c)$ in (F3) since $\mathcal{F}_0(\alpha,\beta,c)$ is similar. Define $q_1 \equiv F_U^{-1}(|Y_1|/c)$. By the Laplace definition it follows

$$q_1 = \frac{1}{\sqrt{2}}\left\{\ln 2 + \ln\left(\frac{|Y_1|}{c}\right)\right\} < 0 \text{ if } \frac{|Y_1|}{c} \le 1/2 \text{ and } q_1 = -\frac{1}{\sqrt{2}}\left\{\ln 2 + \ln\left(1-\frac{|Y_1|}{c}\right)\right\} > 0 \text{ if } \frac{|Y_1|}{c} > 1/2.$$

Use $Y_j \perp X, U$ and substitute $y = |Y_1|$ to deduce

$$\mathcal{F}_1(\alpha,\beta,c) \tag{F13}$$
$$= E_{Y_1}\left[Y_1^2 I\left(|Y_1|\le c\right)\frac{f_X\left(q_1/\beta\right)}{f_U\left(q_1\right)}\right]$$
$$= \sqrt{2}\int_0^{c/2} y^2\exp\left\{-\sqrt{2}y\right\}\times\exp\left\{\left(\ln 2+\ln\left(y/c\right)\right)\left(1/\beta-1\right)\right\}dy$$
$$+ \sqrt{2}\int_{c/2}^{c} y^2\exp\left\{-\sqrt{2}y\right\}\times\exp\left\{\left(\ln 2+\ln\left(1-y/c\right)\right)\left(1/\beta-1\right)\right\}dy$$
$$= 2^{\frac{2-\beta}{2\beta}}\int_0^{c/2} y^2\exp\left\{-\sqrt{2}y\right\}\times\left(y/c\right)^{1/\beta-1}dy + 2^{\frac{2-\beta}{2\beta}}\int_{c/2}^{c} y^2\exp\left\{-\sqrt{2}y\right\}\times\left(1-y/c\right)^{1/\beta-1}dy$$

$$= 2^{\frac{1-2\beta}{2\beta}} c^{-(1/\beta-1)} \left[ \int_0^{c/\sqrt{2}} \exp\{-y\} \times y^{1+1/\beta} dy + \int_{c/\sqrt{2}}^{\sqrt{2}c} y^2 \exp\{-y\} \times \left(\sqrt{2}c - y\right)^{1/\beta-1} dy \right]$$

$$= 2^{1/(2\beta)-1} c^{-(1/\beta-1)} \left(\mathcal{I}_1(c) + \mathcal{I}_2(c)\right).$$

It suffices to show each $\mathcal{I}_i(c) = K + O(e^{-c/4})$ and at least one $\lim_{c\to\infty} \mathcal{I}_i(c) > 0$. It then follows by (F3) that $(\partial/\partial c)P(|Z| > c) = -Kc^{-2-1/\beta}(1 + O(e^{-c/4}))$, hence by dominated convergence $P(|Z| > c) = d(\beta)(1 + O(e^{-c/4}))$. Use (F3) and (F13) to deduce $d(\beta) = \beta^{-1}2^{1/(2\beta)} \int_0^{\infty} \exp\{-y\} y^{1+1/\beta} dy \in (0,\infty)$ for all $\beta \in (0,\infty)$.

If $\beta = 1$ then $\mathcal{I}_1(c) + \mathcal{I}_2(c) = \int_0^{\sqrt{2}c} y^2 \exp\{-y\} dy = 2 + o(e^{-c/4})$, and if $\beta \neq 1$ then $\lim_{c\to\infty} \mathcal{I}_1(c) \in (0,\infty)$ in view of the exponential term $\exp\{-y\}$. It remains to bound $\mathcal{I}_2(c)$. If $\beta < 1$ then

$$\begin{aligned}
\mathcal{I}_2(c) &= \int_{c/\sqrt{2}}^{\sqrt{2}c} y^2 \exp\{-y\} \times \left(\sqrt{2}c - y\right)^{1/\beta-1} dy \leq 2^{(1-\beta)/2\beta} c^{1/\beta-1} \int_{c/\sqrt{2}}^{\sqrt{2}c} y^2 \exp\{-y\} dy \\
&\leq 2^{(1+\beta)/2\beta} \frac{c^{1/\beta+1}}{\exp\left\{\sqrt{2}c\right\}} = o\left(e^{-c/4}\right).
\end{aligned}$$

Finally, if $\beta > 1$ then $e^{c/4} y^2 \exp\{-y\} \times (\sqrt{(2)}c - y)^{1/\beta-1} dy \leq K y^{-(1+\delta)}$ for all $y \in [c/\sqrt{2}, \sqrt{2}c - \iota]$, tiny $\iota > 0$, some tiny $\delta > 0$ and some large $K > 0$. Therefore $\int_{c/\sqrt{2}}^{\sqrt{2}c-\iota} y^2 \exp\{-y\} \times (\sqrt{2}c - y)^{1/\beta-1} dy = o(e^{-c/4})$ for any tiny $\iota > 0$, hence $\mathcal{I}_2(c) = K + O(e^{-c/4})$. $\mathcal{QED}$.

**Proof of Theorem F.4.** Symmetry follows from $\alpha = 0$, independence $Y_j \perp X, U$, and distribution symmetry for all $(Y_0, Y_1, X, U)$.

We only compute $\mathcal{F}_1(0,\beta,c)$ since $\mathcal{F}_0(0,\beta,c)$ is similar. Now let $\Phi(w)$ and $\phi(w)$ be the normal cdf and pdf. In order to characterize the standard normal quantile $\Phi^{-1}(u/c)$ for $u \in [0,c]$, we use the expansion $1 - \Phi(w) = (1 + O(1/w^2)) \times \phi(w)/w$ to solve $u/c = \phi(w(c))/w(c)$ for some $w(c)$ as $c \to \infty$ hence as $w(c) \to \infty$. See Gray and Wang (1991), cf. Lew (1981) and Hawkes (1982). Rudimentary algebra reveals $w(c)$ satisfies

$$w(c) = 2^{1/2} (\ln(c))^{1/2} \left(1 - \frac{\ln(u)}{\ln(c)} - \frac{\ln(2\pi)}{\ln(c)}\right)^{1/2} (1 + O(1/\ln(c))).$$

Since $|\Phi^{-1}(u/c)| = w(c)$ use formula (F7) to deduce $\mathcal{F}_1(0,\beta,c)$ is identically:

$$\begin{aligned}
&\left(\frac{2}{\pi}\right)^{1/2} \int_0^c \frac{u^2}{\exp\{u^2/2\}} \exp\left\{\frac{\beta^2-1}{\beta^2} \ln(c) \left(1 - \frac{\ln(u)}{\ln(c)} - \frac{\ln(2\pi)}{\ln(c)}\right)(1 + O(1/\ln(c)))\right\} du \\
&= \left(\frac{2}{\pi}\right)^{1/2} \int_0^c u^2 \exp\left\{-u^2/2\right\} c^{(1-1/\beta^2)(1-\ln(u)/\ln(c)-\ln(2\pi)/\ln(c))(1+O(1/\ln(c)))} du \\
&= \left(\frac{2}{\pi}\right)^{1/2} c^{(1-1/\beta^2)(1+O(1/\ln(c)))} \int_0^c u^2 \exp\left\{-u^2/2\right\} c^{-(1-1/\beta^2)O(1/\ln(c))(\ln(u)+\ln(2\pi))} du.
\end{aligned}$$

If $\beta = 1$ then $\lim_{c\to\infty} \mathcal{F}_1(0,1,c) = 1$, in particular $\mathcal{F}_1(0,1,c) = 1 + o(e^{-c/2})$ is easily verified given the normal density.

Now assume $\beta \neq 1$ and let $d(\beta)$ be a positive finite function of $\beta$ that may change from line to line. Observe

$$\ln\left(c^{-(1-1/\beta^2)O(1/\ln(c))(\ln(u)+\ln(2\pi))}\right) = -\left(1-1/\beta^2\right) \times O\left(1\right) \times \ln(2\pi u),$$

hence by the monotonicity of the natural log $c^{-(1-1/\beta^2)O(1/\ln(c))(\ln(u)+\ln(2\pi))} = (2\pi u)^{-(1-1/\beta^2)\times O(1)}$. Similarly $c^{(1-1/\beta^2)O(1/\ln(c))} = O(1) \times c^{(1-1/\beta^2)}$. Therefore:

$$\begin{aligned}
\mathcal{F}_1(0,\beta,c) &= (2\pi)^{-(1-1/\beta^2)\times O(1)} \times O\left(1\right) \times c^{(1-1/\beta^2)} \int_0^c u^2 \exp\left\{-u^2/2\right\} u^{-(1-1/\beta^2)\times O(1)} du \\
&= \tilde{d}(\beta) \times \left(1 + o(e^{-c/2})\right) \times c^{(1-1/\beta^2)},
\end{aligned}$$

say. Use the formula for $(\partial/\partial c)P(|Z| > c)$ in (F3) to deduce

$$\frac{\partial}{\partial c}P\left(|Z| > c\right) = -\frac{1}{\beta}\tilde{d}(\beta)c^{-3}c^{(1-1/\beta^2)}\left(1 + o(e^{-c/2})\right) = -\frac{1}{\beta}\tilde{d}(\beta)c^{-2-\beta^{-2}}\left(1 + o(e^{-c/2})\right),$$

hence by dominated convergence $P(|Z| > c) = d(\beta)c^{-(1+1/\beta^2)}(1 + o(e^{-c/2}))$ where $d(\beta) \equiv \beta^{-1}\tilde{d}(\beta) = \beta^{-1}(2\pi)^{-K(1-1/\beta^2)} \int_0^\infty u^2 \exp\left\{-u^2/2\right\} u^{-K(1-1/\beta^2)}$. $\mathcal{QED}$.

# G Other Tail-Trimmed Estimators

We study the properties of the trim-by-$X$ estimator for scalar $X_i$:

$$\theta_n^{(tx)} = \frac{1}{n}\sum_{i=1}^{n} Z_i I\left(|X_i| \leq \nu_n\right),$$

where $\{\nu_n\}$ is a sequence of positive numbers, $\nu_n \to \infty$. We therefore ignore sampling error associated with propensity score estimation in order to focus on the merit of trimming by $X_i$. In any event, under a threshold crossing treatment assignment model $D_i = I(\beta X_i - U_i \geq 0)$ with independent $U_i$ and $X_i$, trimming symmetrically by $p(X_i)$ or $X_i$ are equivalent when $U_i$ has a continuous symmetric distribution about zero.[4]

We further simplify derivations by working in the following latent variable framework with Laplace distributed variables, in which case the ATE $\theta = 0$. See Chaudhuri and Hill (2014, Part I) for a broad treatment of the latent variable model and a demonstration that limited overlap in this environment yields power law tails in $Z_i$.

**Assumption A6 (treatment assignment):** The treatment assignment satisfies $D = I\left(\alpha + \beta X - U \geq 0\right)$; $X \perp U$; $Y_j \perp X, U$; and $(Y_0, Y_1, X, U)$ are iid Laplace distributed with cdf (F6).

Under Assumption A6, $\theta_n^{(tx)}$ is unbiased since by independence $E[Z_i I(|X_i| \leq \nu_n)] = E[\{D_i Y_{1,i} + (1 - D_i)Y_{0,i}\}h_i I(|X_i| \leq \nu_n)] = 0 = \theta$. We abstract from the possibility of bias in order to focus on the convergence rate. Note that Khan and Tamer's (2010a: Section 4.1) characterization of bias for $\theta_n^{(tx)}$ is presumably under the assumption $\alpha \neq 0$ (see their footnote 7). Define the variance

$$\mathcal{S}_n^2 \equiv E\left[\left\{Z_i I\left(|X_i| \leq \nu_n\right) - E\left[Z_i I\left(|X_i| \leq \nu_n\right)\right]\right\}^2\right] = E\left[Z_i^2 I\left(|X_i| \leq \nu_n\right)\right] - \theta^2 \times (1 + o(1)), \tag{G14}$$

where the second equality follows from $\nu_n \to \infty$ and dominated convergence.

Khan and Tamer (2010b,a) study the convergence rate of $\theta_n^{(tx)}$ under A6 with $\beta = 1$, and with other distributions. Under A6 we characterize the limit distribution and rate of convergence $n^{1/2}/\mathcal{S}_n$ of $\theta_n^{(tx)}$, and compare $\theta_n^{(tx)}$ to the trim-by-$Z$ estimator $\hat{\theta}_n^{(tz)}$. We again use $\beta \gtreqless 1$ to mimic the setting where $\beta = 1$ and $\{U, X\}$ may have different distribution tails. We then reveal the weak correspondence between extremes in $X_i$ and in $Z_i$. This sheds light on the inability of $\theta_n^{(tx)}$ to control for heavy tails in small samples, based on our simulation study, unless the sample portion of trimmed $Z_i$ is very large. Finally, we present an improved version of $\theta_n^{(tx)}$ that uses a stochastic threshold and discuss how to set the trimming fractile such that it compares closely with $\hat{\theta}_n^{(tz)}$.

[4] Observe that $p(X_i) = F_U(\alpha + \beta' X_i)$, where $F_U(c) \equiv P(U_i \leq c)$, hence $\tilde{\nu}_{n,1} \leq p(X_i) \leq \tilde{\nu}_{n,2}$ *if and only if* $F_U^{-1}(\tilde{\nu}_{n,1}) \leq \alpha + \beta' X_i \leq F_U^{-1}(\tilde{\nu}_{n,2})$. If $\alpha = 0$, $U$ has a symmetric distribution about zero, $X_i$ is scalar, and the cutpoints are symmetric $\tilde{\nu}_n = \tilde{\nu}_{n,2} = 1 - \tilde{\nu}_{n,1}$, then trimming symmetrically by $p(X_i)$ or $X_i$ are arithmetically identical when $\nu_n \equiv F_U^{-1}(\tilde{\nu}_n)/\beta$ since $1 - \tilde{\nu}_n \leq p(X_i) \leq \tilde{\nu}_n$ *if and only if* $|X_i| \leq F_U^{-1}(\tilde{\nu}_n)/\beta = \nu_n$. If $\alpha \neq 0$ and/or $U_i$ has an asymmetric distribution, and $X_i$ is a scalar, or $D_i = I(\gamma_0' X_i - U_i \geq 0)$ for vector-valued $X_i$ that may contain a constant term, then trimming by $p(X_i)$ or $X_i$ are similar, but not identical.

## G.1 Properties of $\theta_n^{(tx)}$

We first characterize $\mathcal{S}_n^2$. Under A6, $E[Y_j^2] = 1$ and hence by independence $E[Z_i^2 I(|X_i| \leq \nu_n)] = E[h_i^2 I(|X_i| \leq \nu_n)]$. Now apply dominated convergence and $\theta = 0$ to deduce $\mathcal{S}_n^2 \sim E[h_i^2 I(|X_i| \leq \nu_n)]$, while

$$E\left[h_i^2 I\left(|X_i| \leq \nu_n\right)\right] = E\left[\frac{1}{F_u(\beta X_i)} I\left(|X_i| \leq \nu_n\right)\right] + E\left[\frac{1}{1 - F_u(\beta X_i)} I\left(|X_i| \leq \nu_n\right)\right]. \tag{G15}$$

By the Laplace assumption, the first term in (G15) is (the second term has a similar expression):

$$\begin{aligned} E\left[\frac{1}{F_u(\beta X_i)} I\left(|X_i| \leq \nu_n\right)\right] &= \int_{-\nu_n}^{\nu_n} \frac{1}{F(\beta x)} \frac{\partial}{\partial x} F(x)\, dx \qquad \text{(G16)} \\ &= \sqrt{2}\left[\int_{-\nu_n}^{0} e^{\sqrt{2}x(\beta-1)} dx + \int_0^{\nu_n} \frac{e^{-\sqrt{2}x}}{2 - e^{-\sqrt{2}\beta x}} dx\right] \\ &= \int_0^{\sqrt{2}\nu_n} e^{x(\beta-1)} dx + \int_0^{\sqrt{2}\nu_n} \frac{e^{(\beta-1)x}}{2e^{\beta x} - 1} dx = \int_0^{\sqrt{2}\nu_n} e^{x(\beta-1)} dx \times (1 + o(1)). \end{aligned}$$

By Theorem F.3, $Z_i$ has a tail

$$P\left(|Z_i - \theta| \geq c\right) = dc^{-\kappa}(1 + o(1)) \text{ with } \kappa = 1 + 1/\beta.$$

If $\beta < 1$ then $\kappa > 2$ and $\int_0^{\sqrt{2}\nu_n} e^{x(\beta-1)} dx = O(1)$, hence $E[h_i^2 I(|X_i| \leq \nu_n)] \sim 2\int_0^\infty e^{-x(1-\beta)} dx = 2/(1 - \beta) = E[h_i^2]$. The case studied in Khan and Tamer (2010a) is $\beta = 1$ which aligns with a tail index $\kappa = 2$, and $E[h_i^2 I(|X_i| \leq \nu_n)] \sim \sqrt{2}\nu_n \to \infty$ by (G16). Finally, if $\beta > 1$ then $\kappa < 2$ , and $\int_0^{\sqrt{2}\nu_n} e^{x(\beta-1)} dx = (\beta - 1)^{-1}\left(\exp\{\sqrt{2}\nu_n(\beta - 1)\} - 1\right)$ hence $E[h_i^2 I(|X_i| \leq \nu_n)] \sim 2(\beta - 1)^{-1}\left(\exp\{\sqrt{2}\nu_n(\beta - 1)\} - 1\right) \to \infty$.

This proves $\mathcal{S}_n^2$ is finite for each $n$ and any $\beta$, and monotonically increasing in $\beta$ when $\beta \geq 1$. Khan and Tamer (2010b, Theorem 4.1) assume the Lindeberg condition holds in order to prove asymptotic normality in a general environment. Using arguments in Khan and Tamer (2010b, Section 3), however, the condition is straightforward to verify here, hence we omit a proof.

**Theorem G.1.** *Under Assumption A6* $n^{1/2}\mathcal{S}_n^{-1}(\theta_n^{(tx)} - \theta) \xrightarrow{d} N(0, 1)$. *In particular: if* $\beta < 1$ *then* $n^{1/2}(\theta_n^{(tx)} - \theta) \xrightarrow{d} N(0, 2/(1 - \beta))$; *if* $\beta = 1$ *then* $(n^{1/2}/\nu_n)(\theta_n^{(tx)} - \theta) \xrightarrow{d} N(0, 2)$; *and if* $\beta > 1$ *then* $(n^{1/2}/e^{\sqrt{2}\nu_n(\beta-1)})(\theta_n^{(tx)} - \theta) \xrightarrow{d} N(0, 2/(\beta - 1))$.

**Remark 7.** There are substantial differences in estimator behavior for the full range of $\beta > 0$. Small $\beta \in (0, 1)$ implies $Z_i$ has a finite variance hence $\theta_n^{(tx)}$ is $n^{1/2}$-convergent with asymptotic variance $2/(1 - \beta)$, identical to the untrimmed $1/n \sum_{i=1}^n Z_i$. Unity $\beta = 1$ aligns with a hairline infinite variance, and convergence rate $n^{1/2}/\nu_n = o(n^{1/2})$, with an asymptotic variance that depends on $\nu_n$. Greater than unity $\beta > 1$ aligns with a power law tail with index $1 + 1/\beta < 2$, and for a chosen sequence $\{\nu_n\}$ the rate of convergence is exponentially slower. For example, if we use $\nu_n = \lambda(\ln(n))^\delta$ for $\delta \in (0, 1]$ as do Khan and Tamer (2010b) when the errors and regressors have exponential tails, then $\theta_n^{(tx)}$ has a convergence rate $n^{1/2}/(\ln(n))^\delta$ when $\beta = 1$ but only $n^{1/2}/e^{\sqrt{2}(\beta-1)\lambda(\ln(n))^\delta}$ when $\beta > 1$. Consider that if $\nu_n = \lambda \ln(n)$ and $\beta > 1$ then the rate is just $n^{1/2-\sqrt{2}\lambda(\beta-1)}$. We therefore require information on $\beta$ in order to set $\lambda$ small enough just to ensure $n^{1/2-\sqrt{2}\lambda(\beta-1)} \to \infty$. The choice of $\nu_n = \lambda \ln(\ln(n))$, however, is always valid since $n^{1/2}/e^{\nu_n(\beta-1)} = n^{1/2}/(\ln(n))^{\sqrt{2}\lambda(\beta-1)} \to \infty$.

## G.2 Comparison of Estimators

We now compare $\theta_n^{(tx)}$ and $\hat{\theta}_n^{(tz)}$ based on their rates of convergence and ability to remove extreme observations of $Z_i$. Recall we assume the propensity score $p(\cdot)$ is known.

### G.2.1 Rates of Convergence

We first derive the limit distribution of $\hat{\theta}_n^{(tz)}$ with its case-dependent asymptotic variance. Combine $E[Z_i^2]$ derived above for the case $\beta < 1$, with Theorem F.3 for the power law property with index $1 + 1/\beta$, and Lemma 3.2 in the main paper for rates of convergence, to deduce the following.

**Theorem G.2.** *Let A6 hold. If $\beta < 1$ then $n^{1/2}(\hat{\theta}_n^{(tz)} - \theta) \xrightarrow{d} N(0, 2/(1-\beta))$, if $\beta = 1$ then $(n/\ln(n))^{1/2}(\hat{\theta}_n^{(tz)} - \theta) \xrightarrow{d} N(0, d)$, and if $\beta > 1$ then $n^{1/2}/((n/k_n)^{\beta/(\beta+1)-1/2})(\hat{\theta}_n^{(tz)} - \theta) \xrightarrow{d} N(0, d^{2\beta/(\beta+1)} \times (\beta+1)/(\beta-1))$.*

A comparison of the convergence rates when $\beta \geq 1$ is complicated by the presence of the threshold $\nu_n$ in $\theta_n^{(tx)}$ and fractile $k_n$ (with associated threshold $c_n$) in $\hat{\theta}_n^{(tz)}$. Khan and Tamer (2010a) suggest $\nu_n = \lambda \ln(n)$ for some $\lambda > 0$ for the logit case with $\beta = 1$. Since Laplace and logit distributions will lead to the same essential results, consider $\nu_n = \lambda \ln(n)$. Then $\theta_n^{(tx)}$ and $\hat{\theta}_n^{(tz)}$ have the same rates of convergence when $\beta \leq 1$ by Theorems G.1 and G.2.

However, if $\beta > 1$ then $e^{\nu_n(\beta-1)} = n^{\sqrt{2}\lambda(\beta-1)}$ hence $\theta_n^{(tx)}$ has rate $n^{1/2-\sqrt{2}\lambda(\beta-1)} \to \infty$ *only provided* $\beta < 1 + 1/(2^{3/2}\lambda)$. Conversely, $\hat{\theta}_n^{(tz)}$ has a rate $n^{1/2}/(n/k_n)^{\beta/(\beta+1)-1/2} \to \infty$ for any value $\beta > 1$. Now, Paretian tail decay and the threshold construction imply $c_n \sim d^{1/(1+1/\beta)}(n/k_n)^{1/(1+1/\beta)}$ . If the fractile $k_n$ implies the thresholds of $\hat{\theta}_n^{(tz)}$ satisfy $c_n \sim \lambda \ln(n)$, similar to $\nu_n$, then we must have a number of trimmed $Z_i's$ equal to $k_n \sim Kn/(\ln(n))^{1+1/\beta}$. In this case the rate of convergence for $\hat{\theta}_n^{(tz)}$ is $n^{1/2}/(\ln(n))^{1-(\beta+1)/(2\beta)}$ which is faster than the rate $n^{1/2-\sqrt{2}\lambda(\beta-1)}$ for $\theta_n^{(tx)}$ with threshold $\nu_n = \lambda \ln(n)$.

This suggests that the trim-by-$Z$ estimator $\hat{\theta}_n^{(tz)}$ has a faster rate of convergence than the trim-by-$X$ estimator $\theta_n^{(tx)}$ in the heavy tail case $\beta > 1$ when the same type of thresholds are used. Although we only treat the Laplace case here, in general this follows from the fact that limited overlap and therefore heavy tails imply potentially many large values of $Z_i$ are present, while this slows down the convergence rate. The estimator $\hat{\theta}_n^{(tz)}$ removes extreme $Z_i's$ by construction, and as we show next, for a given threshold sequence $\theta_n^{(tx)}$ is more likely to leave extremes present, which leads to its comparatively slower rate.

### G.2.2 Ability to Remove Extreme Observations

By construction $\theta_n^{(tx)}$ removes $Z_i$ only when $X_i$ is large. We now demonstrate the correspondence between extreme values of $X_i$ and $Z_i$ can be weak by simulating $P(|Z_i| > c_z \mid |X_i| > c_x)$, the conditional probability that $Z_i$ is large when $X_i$ is large, for various thresholds $\{c_x, c_z\}$.

We use a latent variable model for treatment assignment $D = I(\beta X - U \geq 0)$, for choices $\beta \in \{.25, 1, 2\}$. Each $(Y_0, Y_1, X, U)$ is iid standard normal, or Laplace with cdf (F6), hence $\beta \in \{.25, 1, 2\}$ aligns with finite, hairline infinite, and infinite variance cases. We draw $R = 10,000$ samples $\{Z_i\}_{i=1}^n$ of size $n = 1,000,000$, and compute

$$P_{n,r} = P_{n,r}(c_z, c_x) \equiv \frac{1/n \sum_{i=1}^n I\left(|Z_i| > c_z\right) I\left(|x_i| > c_x\right)}{1/n \sum_{i=1}^n I\left(|x_i| > c_x\right)} \tag{G17}$$

for each $r^{th}$ sample and $\{c_x, c_z\} \in [1, 10]$ with increments of 1. By the law of large numbers and independence, $P_{n,r}$ will be very close to $P(|Z_i| > c_z \mid |X_i| > c_x)$ with high probability.

Plots of $1/R \sum_{r=1}^R P_{n,r}$ are contained in Figure G.1. In all cases $P_{n,r} \leq .6$, and $P_{n,r} \leq .05$ when both $c_x, c_z \geq 4$. The event $|X_i| > c_x$ for large $c_x$ is a very weak predictor of $|Z_i| > c_z$ for large $c_z$. Furthermore, the probability is smaller when tails are heavier: when $\beta = 2$, hence $\kappa < 2$, we have $P_{n,r} \leq .3$ and .4 for Laplace and Normal cases, respectively. However, $P_{n,r}$ is monotonically *higher* for each $c_z$ and *small* $c_x$.

This is precisely what we find in our simulation experiments below: we must use small $c_x$ to ensure as close a correspondence between $X_i$ and $Z_i$ sample extremes as possible. Specifically, we must trim a large number of observations to ensure an adaptive version of $\theta_n^{(tx)}$ is close to normally distributed, and has small bias when $Z$ is symmetrically distributed. If we let $c_x$ be large, and therefore trim few observations, then $P(|Z_i| > c_z||X_i| > c_x)$ is small and in any given sample $\theta_n^{(tx)}$ tends not to remove enough, or any, extremes: $\theta_n^{(tx)}$ performs roughly on par with the untrimmed estimator $1/n \sum_{i=1}^n Z_i$.

### G.3 Adaptive Trim-by-X and Trim-by-p(X) Estimators

A chosen $\nu_n$ may result in no trimming at all in some samples, or very few observations trimmed that do not sufficiently align with sample extremes of $Z_i$, and therefore estimator instability may still exist. A simple improvement for $\theta_n^{(tx)}$ bases trimming on an order statistic of $X_i$.

Under the assumption that there is only one covariate $X$ that matters for trimming, define $X_i^{(a)} \equiv |X_i|$, denote the order statistics $X_{(1)}^{(a)} \geq X_{(2)}^{(a)} \cdots$, and let $\{k_n^{(x)}\}$ be an intermediate order sequence: $k_n^{(x)} \to \infty$ as $k_n^{(x)}/n \to 0$. Then an adaptive version of $\theta_n^{(tx)}$ is $\hat{\theta}_n^{(tx)} \equiv 1/n \sum_{i=1}^n Z_i I(|X_i| \leq X_{(k_n^{(x)})}^{(a)})$, in which case $\nu_n$ satisfies $P(|X_i| > \nu_n) \sim k_n^{(x)}/n$. Lemma A.4 in the main paper can be extended to $\hat{\theta}_n^{(tx)}$ to verify $(n^{1/2}/\mathcal{S}_n)(\hat{\theta}_n^{(tx)} - \theta_n^{(tx)}) \xrightarrow{p} 0$. Coupled with Theorem G.1, this proves the next claim.

**Theorem G.3.** *Under A6 $\hat{\theta}_n^{(tx)}$ satisfies Theorem G.1.*

**Remark 8.** The result can be extended to other distributions, evidently case-by-case since the Lindeberg condition must be verified. Thus, another advantage of the trim-by-$Z$ estimator $\hat{\theta}_n^{(tz)}$ is we do not need to make any assumptions on the distributions of $U$ and $X$ since, by theory and references presented in Hill (2015), the Lindeberg condition holds under very general conditions.

We have thus far assumed the propensity score is known in order to reduce notation. In the same manner as Theorem 3.1 in the main paper, if a parametric plug-in $p_i(\hat{\gamma}_n)$ is used, and Assumptions B1-B3 hold, then

$$\theta_n^{(tx)}(\hat{\gamma}_n) = \frac{1}{n}\sum_{i=1}^n Z_i(\hat{\gamma}_n) I\left(|X_i| \leq \nu_n\right) \text{ and } \hat{\theta}_n^{(tx)}(\hat{\gamma}_n) \equiv \frac{1}{n}\sum_{i=1}^n Z_i(\hat{\gamma}_n) I\left(|X_i| \leq X_{(k_n^{(x)})}^{(a)}\right) \tag{G18}$$

satisfy, e.g., $n^{1/2}\mathcal{S}_n^{-1}(\hat{\theta}_n^{(tx)}(\hat{\gamma}_n) - \theta) \xrightarrow{d} N(0,1)$ in the heavy tail case $E[Z_i^2] = \infty$, and $n^{1/2}\mathcal{S}_n^{-1}(\hat{\theta}_n^{(tx)}(\hat{\gamma}_n) -$

$\theta) \xrightarrow{d} N(0, K)$ for some $K \in (0, \infty)$ that depends on $p_i(\gamma_0)$.

A fully adaptive trim-by-$p(X_i)$ estimator operates similarly. Define order statistics $p_{(1)}(\gamma) \geq \cdots \geq p_{(n)}(\gamma)$, and an intermediate order sequence $\{k_n^{(p)}\}$. The estimator is

$$\hat{\theta}_n^{(tp)}(\hat{\gamma}_n) \equiv \frac{1}{n}\sum_{i=1}^{n} Z_i(\hat{\gamma}_n) I\left(p_{(n-k_n^{(p)}+1)}(\hat{\gamma}_n) \leq p_i(\hat{\gamma}_n) \leq p_{(k_n^{(p)})}(\hat{\gamma}_n)\right). \tag{G19}$$

Since $k_n^{(p)} \to \infty$ and $k_n^{(p)}/n \to 0$ it follows $p_{(n-k_n^{(p)}+1)} \xrightarrow{p} 0$ and $p_{(n-k_n^{(p)}+1)} \xrightarrow{p} 1$, hence trimming is negligible. In the threshold crossing model $D_i = I(\beta X_i - U_i \geq 0)$ where $U_i$ and $X_i$ are independent, and $U_i$ has a symmetric distribution about zero, then it can be shown that $(n^{1/2}/\tilde{\mathcal{S}}_n)(\hat{\theta}_n^{(tp)}(\hat{\gamma}_n) - \theta) \xrightarrow{d} N(0, 1)$ for some sequence of positive constants $\{\tilde{\mathcal{S}}_n\}$, where $\tilde{\mathcal{S}}_n \to \infty$ if $E[Z_i^2] = \infty$.

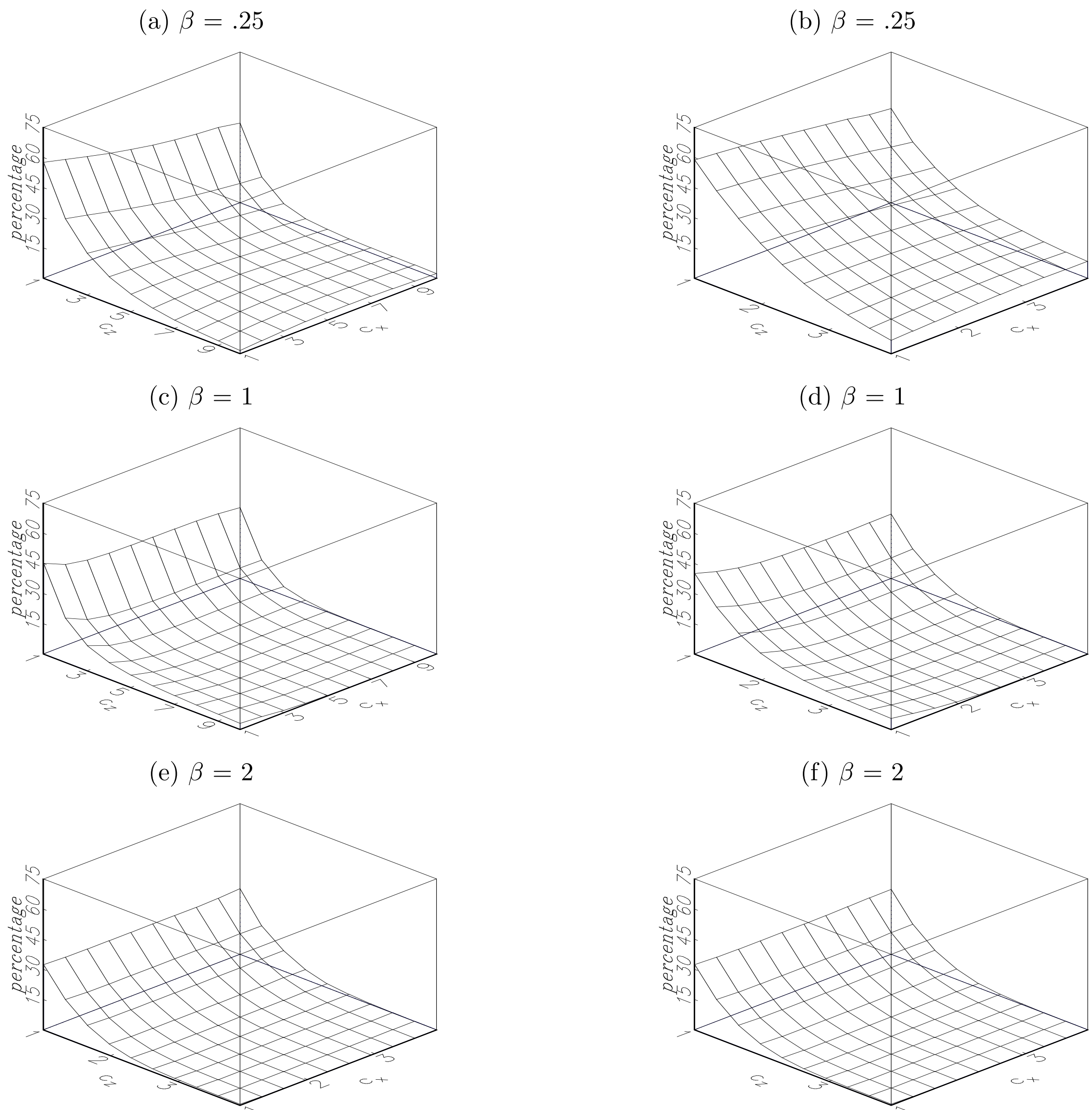


Figure G.1. $P(|Z| > c_z \mid |X| > c_x)$: $(Y_1, Y_2, U, X)$ are iid Laplace (left panels) or Normal (right panels).

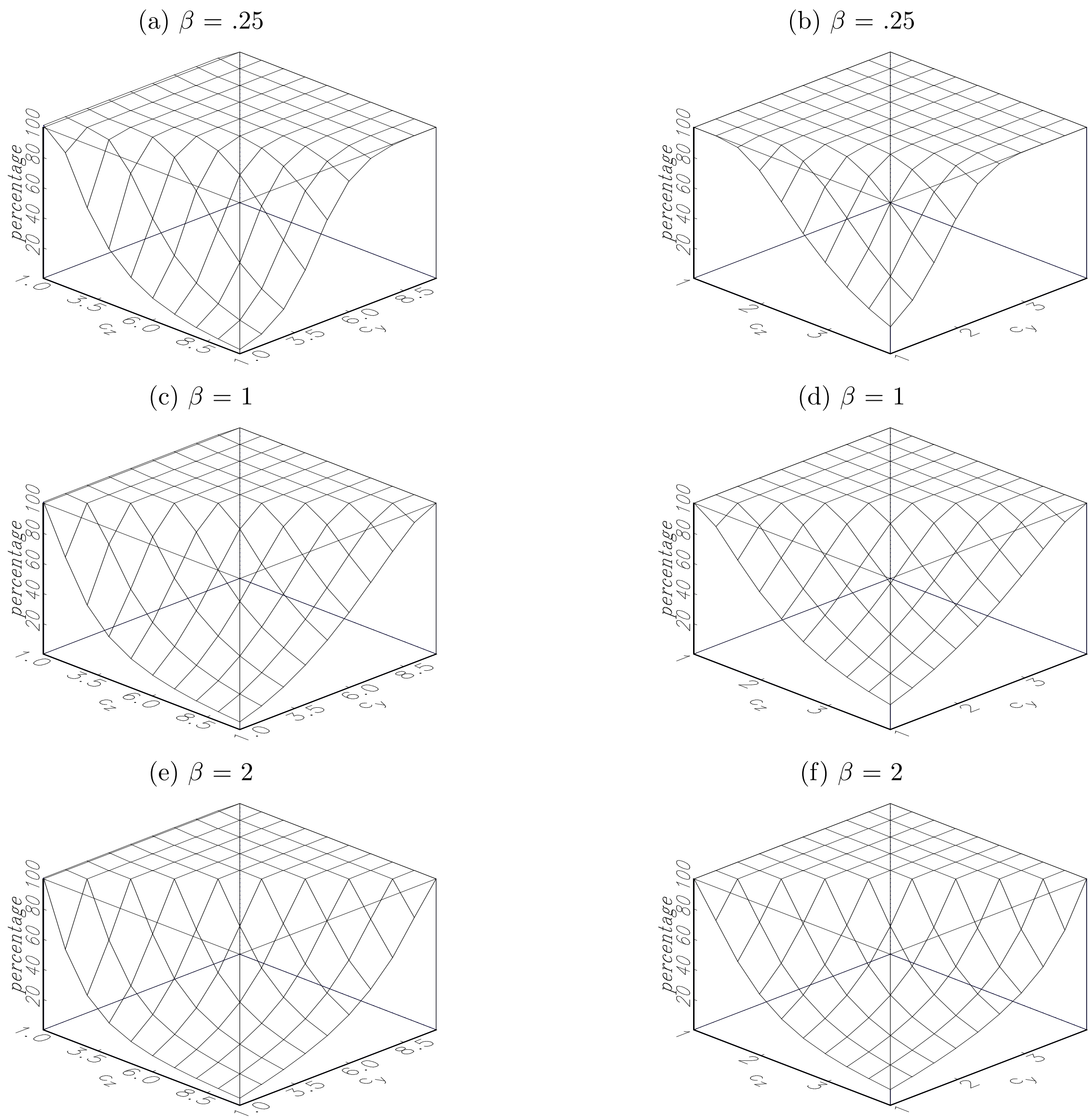


Figure G.2. $P(|Z| > c_z \mid |Y| > c_y)$: $(Y_1, Y_2, U, X)$ are iid Laplace (left panels) or Normal (right panels).

# References


Andrews, D. (1988): "Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables," *Econometric Theory*, 4, 458–467.

Busso, M., J. DiNardo, and J. McCrary (2009): "Finite Sample Properties of Semiparametric Estimators of Average Treatment Effects," Discussion paper, University of Michigan.

Chaudhuri, S., and J. Hill (2014): "Supplemental Appendices I and II for Robust Estimation for Average Treatment Effects," mimeo.

Csörgo, S., L. Horváth, and D. Mason (1986): "What Portion of the Sample Makes a Partial Sum Asymptotically Stable or Normal?," *Probability Theory and Related Fields*, 72, 1–16.

Gray, H., and S. Wang (1991): "A General Method for Approximating Tail Probabilities," *Journal of American Statistical Association*, 86, 159–166.

Hahn, M., D. Weiner, and D. Mason (1991): *Sums, Trimmed Sums and Extremes*. Birkhäuser: Berlin.

Hawkes, A. (1982): "Approximating the Normal Tail," *Journal of the Royal Statistical Societ Series D*, 31, 231–236.

Hill, B. M. (1975): "A Simple General Approach to Inference about the Tail of a Distribution," *Annals of Statistics*, 3(5), 1163–1174.

Hill, J. B. (2015): "Robust Expected Shorfall Estimation for Infinite Variance Time Series," *Journal of Financial Econometrics*, 13, 1–44.

Ibragimov, I., and I. Linnik (1971): *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff.

Khan, S., and E. Tamer (2010a): "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 78, 2021–2042.

——— (2010b): "Irregular Identification, Support Conditions, and Inverse Weight Estimation," Discussion paper, Duke University.

Lew, R. (1981): "An Approximation to the Cumulative Normal Distribution with Simple Coefficients,," *Journal of the Royal Statistical Societ Series C*, 30, 299–301.

Lewbel, A. (1997): "Semiparametric Estimation of Location and Other Discrete Choice Moments," *Econometric Theory*, 13, 32–51.

Resnick, S. (1987): *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag: New York.

Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341.