A Conclusion

In this paper, we addressed the challenge of estimating treatment effects on rare, high-impact events by uniting tools from causal inference and extreme value theory. We introduced a new estimand that explicitly captures how an intervention shifts the tail average of the outcome distribution. Exploiting the spectral–magnitude decomposition inherent in multivariate regular variation, we obtained a simple, implementable identification formula (Proposition 3.3). Building on this, we constructed both inverse-propensity-weighted (IPW) and doubly-robust (DR) estimators, and we further established non-asymptotic error bounds under a Pareto-type tail assumption (Theorem 3.5, Corollary 3.6). In simulations and real-data experiments, our methods consistently outperformed naive estimators when targeting extreme outcomes, thereby validating our theoretical guarantees. We believe this work opens the door to more refined causal analyses in applications—such as disaster risk reduction and financial risk management—where understanding treatment effects in the tail is paramount. One limitation of this work is that we mainly use heuristics to estimate the scaling exponential α, which lacks a theoretical guarantee. For future direction, we would like to explore how to estimate the exponential α and develop a more adaptive method for choosing thresholds for our estimators.

B Proofs

B.1 Identification Formula

In this subsection, we derive the identification formula in Proposition 3.3.

Proof. We first prove (3.4). By Assumption 3.2,

$$\begin{aligned}\lim_{t\to\infty}\mathbb{E}\left[\frac{Y(1)-Y(0)}{t^\alpha}\mid \|U\|>t\right] &= \lim_{t\to\infty}\mathbb{E}\left[\frac{f(X,1,U)-f(X,0,U)}{t^\alpha}\mid \|U\|>t\right]\\ &= \lim_{t\to\infty}\mathbb{E}\left[\frac{f(X,1,U)-f(X,0,U)}{\|U\|^\alpha}\cdot\left(\frac{\|U\|}{t}\right)^\alpha\mid \|U\|>t\right]\\ &= \lim_{t\to\infty}\mathbb{E}\left[(g(X,1,U/\|U\|)-g(X,0,U/\|U\|)+2e(t))\cdot\left(\frac{\|U\|}{t}\right)^\alpha\mid \|U\|>t\right]\end{aligned}$$

We next argue that $\lim_{t\to\infty}\mathbb{E}[(\|U\|/t)^\alpha\mid\|U\|>t]=\alpha/(\beta-\alpha)$. We have

$$\begin{aligned}\mathbb{E}[(\|U\|/t)^\alpha\mid\|U\|>t] &= 1+\int_1^\infty P((\|U\|/t)^\alpha\geqslant r\mid\|U\|>t)\,\mathrm{d}r\\ &= 1+\frac{\int_1^\infty P(\|U\|/t>r^{1/\alpha})\,\mathrm{d}r}{P(\|U\|>t)}\\ &= 1+\frac{\int_1^\infty \alpha P(\|U\|>rt)r^{\alpha-1}\,\mathrm{d}r}{P(\|U\|>t)}\end{aligned}$$

Note that $\|U\|$ is also regularly varying. By Potter's theorem [Bingham et al., 1989, Theorem 1.56], for any $\epsilon>0$ and sufficiently large t, we have

$$\frac{P(\|U\|>rt)}{P(\|U\|>t)}\leqslant 2r^{-\beta+\epsilon}.$$

Take $\epsilon>0$ such that $\alpha-\beta+\epsilon-1<-1$, we have for sufficiently large t,

$$\frac{\int_1^\infty \alpha P(\|U\|>rt)r^{\alpha-1}\,\mathrm{d}r}{P(\|U\|>t)}\leqslant 2\int_1^\infty \alpha r^{\alpha-\beta+\epsilon-1}\,\mathrm{d}r<\infty.$$

Therefore, by the dominance convergence theorem,

$$\mathbb{E}[(\|U\|/t)^\alpha\mid\|U\|>t]\to 1+\int_1^\infty \alpha r^{-\beta}\cdot r^{\alpha-1}\,\mathrm{d}r=\beta/(\beta-\alpha),$$

which implies

$$\lim_{t\to\infty} e(t)\mathbb{E}[(\|U\|/t)^\alpha \mid \|U\| > t] = 0$$

and (3.4) holds. We then verify the uniform integrability of function

$$h(U) = \mathbb{E}_X[(g(X,1,U/\|U\|) - g(X,0,U/\|U\|))(\|U\|/t)^\alpha].$$

Note that by Assumption 3.2, g is a continuous function on a compact set and thus is bounded by some constant $C > 0$. We have

$$\mathbb{E}[|h(U)| \mid \|U\| > t] \leqslant 2C\mathbb{E}[(\|U\|/t)^\alpha \mid \|U\| > t].$$

We have proven that the Right Hand Side (RHS) converges to a constant as $t \to \infty$, which implies that $\mathbb{E}[|h(U)| \mid \|U\| > t]$ is uniformly bounded. We conclude that $h(U)$ is uniformly integrable. By the uniform-integrability convergence theorem,

$$\begin{aligned}\lim_{t\to\infty} \mathbb{E}\left[(g(X,1,U/\|U\|) - g(X,0,U/\|U\|)) \cdot \left(\frac{\|U\|}{t}\right)^\alpha \mid \|U\| > t\right] &= \mathbb{E}_{(r,\theta)\sim\mathcal{L}}[(g(X,1,\theta) - g(X,0,\theta))r^\alpha] \\ &= \mathbb{E}_{\theta\sim\mathcal{L}}[(g(X,1,\theta) - g(X,0,\theta))]\mathbb{E}_{r\sim\mathcal{L}}[r^\alpha].\end{aligned}$$

where $\mathcal{L}$ is the limiting distribution of $(\|U\|/t, U/\|U\|) \mid \|U\| > t$ and we use the asymptotic independent property of regularly varying distributions (Definition 2.3). □

B.2 Non-asymptotic Analysis

To obtain a convergence rate for the estimator $\widehat{\theta}_n^t$, we first analyze the rate of the two factors $\widehat{\gamma}_n$ and $\widehat{\eta}_{n,t}^{\mathrm{DR}}$.

Lemma B.1. Under Assumption 3.1, 3.4 with probability at least $1-\delta$, for sufficiently large n, we have

$$|\gamma - \widehat{\gamma}_n| \leqslant O\left(\left(\frac{\log(2/\delta)}{n}\right)^{1/(2+\beta)}\right),$$

where $\gamma = 1/\beta$ is the EVI of U.

Proof. We adopt the non-asymptotic analysis of the adaptive Hill estimator for EVI in [Boucheron and Thomas, 2015]. In the paper, the author adopts an adaptive estimator, choosing k to be

$$k = \max\left\{k \in \{l_n, \cdots, n\} \text{and } \forall i \in \{l_n, \cdots, n\}, |\widehat{\gamma}(i) - \widehat{\gamma}(k)| \leqslant \frac{\widehat{\gamma}(i)r_n(\delta)}{\sqrt{i}}\right\}$$

where $\widehat{\gamma}(i) = \frac{1}{i}\sum_{j=1}^i \log\frac{\|U_{(j)}\|}{\|U_{(i+1)}\|}$ and $r_n(\delta)$ scales like $\sqrt{\log((2/\delta)\log(n))}$. First, we verify the von Mises conditions in Boucheron and Thomas [2015] under Assumption 3.4. Let F be the CDF of $\|U\|$. By Assumption 3.4, we know that

$$g(z) = c\alpha^m\prod_{i=1}^m(1+z_i)^{-\alpha-1}, \|z\|_1 > \zeta k^{(1-2s)/\alpha}.$$

Note that

$$|c-1| = \left|\frac{g(z) - \alpha^m\prod_{i=1}^m(1+z_i)^{-\alpha-1}}{\alpha^m\prod_{i=1}^m(1+z_i)^{-\alpha-1}}\right| \leqslant \xi k^{-s} \leqslant \xi.$$

Let $\widetilde{Z}_1, \cdots, \widetilde{Z}_m \sim \alpha c^{1/m}(1+z)^{-\alpha-1}, z \geqslant c^{1/(m\alpha)} - 1$. Then, we verify the upper bound for the von Mises function, i.e., $\sup_{s\geqslant t}|\eta(s)| \leqslant O(t^\rho)$ for some $\rho < 0$, where η is the von Mises function.

$$\eta(t) = \frac{tU'(t)}{U(t)} - \frac{1}{\beta} \tag{B.1}$$

$$= \frac{1}{tU(t)f(U(t))} - \frac{1}{\beta}, \tag{B.2}$$

where $f(t)$ is the density function of $\sum_{i=1}^m a_i \widetilde{z}_i$. By [Nguyen, 2014, Theorem 2.1], we have that when $\|U\|_1 > \max_i\{a_i\}\zeta k^{(1-2s)/\alpha}$,

$$f(t) = C\beta t^{-\beta-1}(1 + D(1-1/\beta)t^{-1} + o(t^{-1})). \tag{B.3}$$

Then,

$$\bar{F}(t) = 1 - F(t) = Ct^{-\beta}(1 + Dt^{-1} + o(t^{-1}))$$

and

$$U(t) = C^{1/\beta}t^{1/\beta}(1 + DC^{-1/\beta}t^{-1/\beta}/\beta + o(t^{-1/\beta})). \tag{B.4}$$

Plug in (B.3) and (B.4) into (B.2), we get

$$\eta(t) = \frac{1}{\beta(1 - DC^{-1/\beta}t^{-1/\beta} + o(t^{1/\beta}))} - \frac{1}{\beta} = O(t^{-1/\beta}).$$

Therefore, the growth rate of the von Mises function is bounded. By [Boucheron and Thomas, 2015], with probability at least $1-\delta$, we have

$$|\gamma - \widehat{\gamma}_n| \leqslant O\left(\left(\frac{\log(2/\delta)}{n}\right)^{1/(2+\beta)}\right),$$

$\square$

Lemma B.2. Undet the assumption of Theorem 3.5, with probability at least $1-\delta$, we have

$$|\widehat{\eta}_{n,t}^{\mathrm{DR}} - \eta| \leqslant O(\sqrt{R_p(n/2,\delta)R_g(t^{-\beta}n,\delta)} + t^{\beta/2}n^{-1/2} + t^{-\min\{1,\beta\}} + t^{-\beta s/(1-2s)} + e(t) + \log(t)R_\alpha(n,\delta)).$$

Proof. Let

$$\eta^t = \mathbb{E}\left[\frac{f(X,1,U) - f(X,0,U)}{\|U\|^{\widehat{\alpha}_n}} \mid \|U\| > t\right],$$

We have the following decomposition

$$|\widehat{\eta}_{n,t}^{\mathrm{DR}} - \eta| \leqslant |\widehat{\eta}_{n,t}^{\mathrm{DR}} - \eta^t| + |\eta^t - \eta|.$$

The first term comes from the standard statistical error of DR estimator, while the second term is the bias term caused by the finite threshold. For the first term, by standard DML theory [Foster and Syrgkanis, 2023], we have

$$|\widehat{\eta}_{n,t}^{\mathrm{DR}} - \eta^t| \leqslant O\left(\sqrt{R_p(n/2,\delta)R_g(n_t,\delta)} + n_t^{-1/2}\right),$$

where $n_t = \sum_{i=1}^{n/2} I(\|U_i\| > t)$ is a random variable. By Bernstein's inequality, with probability at least $1-\delta$,

$$n_t - n\mathbb{P}(\|U\| > t)/2 \geqslant O\left(\log(1/\delta) + \sqrt{n\mathbb{P}(\|U\| > t)\log(1/\delta)}\right)$$

Therefore, with the same probability, when $n \geqslant \Theta(\log(1/\delta)t^{\beta})$.

$$n_t \geqslant \frac{1}{4}n\mathbb{P}(\|U\| > t) = \Theta(nt^{-\beta})$$

and we have

$$|\widehat{\eta}_{n,t}^{\mathrm{DR}} - \eta^t| \leqslant O\left(\sqrt{R_p(n/2,\delta)R_g(nt^{-\beta},\delta)} + t^{\beta/2}n^{-1/2}\right), \tag{B.5}$$

where we use the monotonicity of R_p, R_g.

For the second term (the bias term),

$$\begin{aligned}
|\eta^t - \eta| &= \left|\mathbb{E}\left[\frac{f(X,1,U)-f(X,0,U)}{\|U\|^{\widehat{\alpha}_n}} \mid \|U\|>t\right] - \mathbb{E}[g(X,1,U/\|U\|)-g(X,0,U/\|U\|)]\right| \\
&\leqslant \left|\mathbb{E}\left[\frac{f(X,1,U)-f(X,0,U)}{\|U\|^{\alpha}} - g(X,1,U/\|U\|)-g(X,0,U/\|U\|) \mid \|U\|>t\right]\right| \\
&\quad + |\mathbb{E}[g(X,1,U/\|U\|)-g(X,0,U/\|U\|)] - \mathbb{E}[g(X,1,U/\|U\|)-g(X,0,U/\|U\|) \mid \|U\|>t]| \\
&\quad + \left|\mathbb{E}\left[\frac{f(X,1,U)-f(X,0,U)}{\|U\|^{\alpha}}\left(1-\|U\|^{\alpha-\widehat{\alpha}_n}\right)\right]\right| \\
&\leqslant 2e(t) + |\mathbb{E}[g(X,1,U/\|U\|)-g(X,0,U/\|U\|)] - \mathbb{E}[g(X,1,U/\|U\|)-g(X,0,U/\|U\|) \mid \|U\|>t]|, \\
&\quad + \left|\mathbb{E}\left[\frac{f(X,1,U)-f(X,0,U)}{\|U\|^{\alpha}}\left(1-\|U\|^{\alpha-\widehat{\alpha}_n}\right)\right] \mid \|U\|>t\right|
\end{aligned}$$

where we use Assumption 3.2 in the last equality. By the error rate assumption in Theorem 3.5,

$$\begin{aligned}
\left|\mathbb{E}\left[\frac{f(X,1,U)-f(X,0,U)}{\|U\|^{\alpha}}\left(1-\|U\|^{\alpha-\widehat{\alpha}_n}\right)\right] \mid \|U\|>t\right| &\leqslant C\mathbb{E}\left[\left|1-\|U\|^{\alpha-\widehat{\alpha}_n}\right| \mid \|U\|>t\right] \\
&= O(\log(t)R_\alpha(n,\delta)).
\end{aligned}$$

Since g is L-Lipschitz continuous, the second term is upper bounded by Wasserstein distance $LW_1(\mathcal{L}^t_{U/\|U\|}, \mathcal{L}_{U/\|U\|})$, where $\mathcal{L}^t_{U/\|U\|}$ is the distribution of $U/\|U\|$ conditioning on $\|U\|>t$ and $\mathcal{L}_{U/\|U\|}$ is its limiting distribution as $t\to\infty$. Therefore, we have

$$\begin{aligned}
|\eta^t - \eta| &\leqslant 2e(t) + LW_1(\mathcal{L}^t_{U/\|U\|}, \mathcal{L}_{U/\|U\|}) + O(\log(t)R_\alpha(n,\delta)) \\
&\leqslant 2e(t) + O(t^{-\min\{1,\beta\}} + t^{-\beta s/(1-2s)} + \log(t)R_\alpha(n,\delta)), \qquad \text{(B.6)}
\end{aligned}$$

where we use [Zhang et al., 2023, Proposition 3.1] in the last inequality to upper bound the bias term $W_1(\mathcal{L}^t_{U/\|U\|}, \mathcal{L}_{U/\|U\|})$. Combing (B.5) and (B.6), we get

$$|\widehat{\eta}^t_n - \eta| \leqslant O(\sqrt{R_p(n/2,\delta)R_g(t^{-\beta}n,\delta)} + t^{\beta/2}n^{-1/2} + t^{-\min\{1,\beta\}} + t^{-\beta s/(1-2s)} + e(t) + \log(t)R_\alpha(n,\delta)).$$

□

Lemma B.3. Under the assumption of Theorem 3.5, with probability at least $1-\delta$, we have

$$|\widehat{\eta}^{\text{IPW}}_{n,t} - \eta| \leqslant O(R_p(n/2,\delta) + t^{\beta/2}n^{-1/2} + t^{-\min\{1,\beta\}} + t^{-\beta s/(1-2s)}) + e(t).$$

Proof. Similar to Lemma B.2, we have the following decomposition.

$$|\widehat{\eta}^{\text{IPW}}_{n,t} - \eta| \leqslant |\widehat{\eta}^{\text{IPW}}_{n,t} - \eta^t| + |\eta^t - \eta|.$$

Term $|\eta^t - \eta|$ can be bounded in the same way as in the proof of Lemma B.2. By [Su et al., 2023, Theorem 1], we have

$$|\widehat{\eta}^{\text{IPW}}_{n,t} - \eta^t| \leqslant O(R_p(n/2,\delta) + n_t^{-1/2}) = O(R_p(n/2,\delta) + t^{\beta/2}n^{-1/2} + \log(t)R_\alpha(n,\delta)).$$

The rest of the proof is similar. □

Now we are ready to prove Theorem 3.5.

Proof of Theorem 3.5. Note that by the asymptotic independence property of regularly varying distribution,

$$\begin{aligned}
|\widehat{\theta}^{\text{DR}}_{n,t} - \theta^{\text{NETE}}| &= |\widehat{\eta}^{\text{DR}}_{n,t}\cdot\widehat{\mu}_n - \eta\cdot\mu| \\
&\leqslant |\mu|\cdot|\widehat{\eta}^{\text{DR}}_{n,t} - \eta| + |\widehat{\eta}^{\text{DR}}_{n,t}|\cdot|\widehat{\mu}_n - \mu|. \qquad \text{(B.7)}
\end{aligned}$$

By Lemma B.1 and B.2, with high probability, $\widehat{\gamma}_n$ and $|\widehat{\eta}_{n,t}^{\mathrm{DR}}|$ is bounded. Note that by Lemma B.1,

$$|\widehat{\mu}_n - \mu| = \left|\frac{1}{1-\widehat{\alpha}_n\widehat{\gamma}_n} - \frac{1}{1-\alpha\gamma}\right| = O(|\widehat{\alpha}_n - \alpha| + |\widehat{\gamma}_n - \gamma|) = O(\log(1/\delta)n^{-1/(2+\beta)} + R_\alpha(n,\delta))$$

Therefore, by Lemma B.2 and (B.7),

$$\begin{aligned}|\widehat{\theta}_{n,t}^{\mathrm{DR}} - \theta^{\mathrm{NETE}}| \leqslant O(&\sqrt{R_p(n/2,\delta)R_g(nt^{-\beta},\delta)} + t^{\beta/2}n^{-1/2} + \log(1/\delta)n^{-1/(2+\beta)} \\ &+ t^{-\min\{1,\beta\}} + t^{-\beta s/(1-2s)} + \log(t)R_\alpha(n,\delta) + e(t)).\end{aligned}$$

The bound for $\widehat{\theta}_{n,t}^{\mathrm{IPW}}$ can be proven similarly. $\square$

Corollary B.4 (Convergence rate for IPW)**.** Under the assumptions of Theorem 3.5, further suppose that

$$R_p(n,\delta) = \Theta(\log(1/\delta)n^{-1/2}), R_g(n,\delta) = \Theta(\log(1/\delta)n^{-1/2}), R_\alpha(n,\delta) = \Theta(\log(1/\delta)n^{-c_\alpha}),$$

for some $c_\alpha > 0$, the following conclusions hold.

1. If $s \in (0, 1/(2+\max\{1,\beta\}))$, takes $t_n = \Theta(n^{(1-2s)\widehat{\gamma}_n})$, with probability at least $1-\delta$, we have
$$|\widehat{\theta}_{n,t}^{\mathrm{IPW}} - \theta^{NETE}| = O(e(t_n) + n^{-s}\log(1/\delta) + n^{-c_\alpha}\log(n)\log(1/\delta)).$$
2. If $s \in [1/(2+\max\{1,\beta\}), 1/2)$, takes $t = \Theta(n^{(\widehat{\gamma}_n/(1+2\min\{1,\widehat{\gamma}_n\})})$, with probability at least $1-\delta$, we have
$$|\widehat{\theta}_{n,t}^{\mathrm{IPW}} - \theta^{NETE}| = O(e(t_n) + n^{-1/(2+\max\{\beta,1\})}\log(1/\delta) + n^{-c_\alpha}\log(n)\log(1/\delta)).$$

Proof of Corollary 3.6. By Theorem 3.5 and the error rate assumption in Corollary 3.6, we have

$$\begin{aligned}\left|\widehat{\theta}_{n,t}^{\mathrm{DR}} - \theta^{NETE}\right| &\leqslant O(\log(1/\delta)t^{\beta/4}n^{-1/2} + t^{\beta/2}n^{-1/2} + \log(1/\delta)n^{-1/(2+\beta)} \\ &\qquad + t^{-\min\{1,\beta\}} + t^{-\beta s/(1-2s)} + \log(t)\log(1/\delta)n^{-c_\alpha} + e(t)) \\ &\leqslant O(\log(1/\delta)t^{\beta/2}n^{-1/2} + \log(1/\delta)n^{-1/(2+\beta)} \\ &\qquad + t^{-\min\{1,\beta\}} + t^{-\beta s/(1-2s)} + \log(t)\log(1/\delta)n^{-c_\alpha} + e(t)).\end{aligned}$$

If $s \in (0, 1/(2+\max\{1,\beta\}))$, we have

$$\begin{aligned}\left|\widehat{\theta}_{n,t}^{\mathrm{DR}} - \theta^{NETE}\right| \leqslant O(&\log(1/\delta)t^{\beta/2}n^{-1/2} + \log(1/\delta)n^{-1/(2+\beta)} \\ &+ t^{-\beta s/(1-2s)} + \log(t)\log(1/\delta)n^{-c_\alpha} + e(t)).\end{aligned}$$

Takes $t_n = \Theta(n^{(1-2s)\widehat{\gamma}_n})$, we get

$$\begin{aligned}\left|\widehat{\theta}_{n,t_n}^{\mathrm{DR}} - \theta^{NETE}\right| \leqslant O(&\log(1/\delta)n^{(1-2s)\widehat{\gamma}_n\beta/2-1/2} + \log(1/\delta)n^{-1/(2+\beta)} \\ &+ n^{-\widehat{\gamma}_n\beta s} + \log(t)\log(1/\delta)n^{-c_\alpha} + e(t_n)).\end{aligned} \tag{B.8}$$

By Lemma B.1, we have

$$|\widehat{\gamma}_n - \gamma| = |\widehat{\gamma}_n - 1/\beta| \leqslant O(\log(1/\delta)n^{-1/(2+\beta)}).$$

Therefore, $n^{\widehat{\gamma}_n\beta} = 1 + O(n^{-1/(2+\beta)})$. Plug this bound into (B.8) and we can get the results. Similarly, if $s \in [1/(2+\max\{1,\beta\}), 1/2)$, we have

$$\begin{aligned}\left|\widehat{\theta}_{n,t}^{\mathrm{DR}} - \theta^{NETE}\right| \leqslant O(&\log(1/\delta)t^{\beta/2}n^{-1/2} + \log(1/\delta)n^{-1/(2+\beta)} \\ &+ t^{-\min\{1,\beta\}} + \log(t)\log(1/\delta)n^{-c_\alpha} + e(t)).\end{aligned}$$

Take $t_n = \Theta(\widehat{\gamma}_n/(1+2\min\{1,\widehat{\gamma}_n\}))$, we get the results. $\square$

C Experiment Details

In this section, we introduce some details in our experiments. The two baseline estimators we consider are naive-IPW:

$$\widehat{\theta}_{n,t}^{\text{Naive-IPW}} = \frac{1}{t^{\alpha} n_t} \sum_{i>n/2:\|U_i\|\geqslant t} Y_i\Big(\frac{D_i}{\widehat{p}(X_i)} - \frac{1-D_i}{1-\widehat{p}(X_i)}\Big).$$

and naive-DR:

$$\widehat{\eta}_{n,t}^{\mathrm{Naive-DR}} = \frac{1}{t^{\alpha} n_t} \sum_{i>n/2:\|U_i\|\geqslant t} \Big[\widehat{g}(X_i,1,U_i) - \widehat{g}(X_i,0,U_i) + \frac{D_i - \widehat{p}(X_i)}{\widehat{p}(X_i)(1-\widehat{p}(X_i))}\big(Y_i - \widehat{g}(X_i,D_i,U_i)\big)\Big],$$

where $n_t = \sum_{i=\lfloor n/2\rfloor+1}^{n} I(\|U_i\| \geqslant t)$ and $\widehat{p}(X)$ and $\widehat{g}(\cdot)$ are the estimated propensity function and the outcome function respectively. The nuisance estimation of $\widehat{g}$ is obtained by running a regression $Y \sim (X, D, U)$. We clip the propensity to $[10^{-4}, 1-10^{-4}]$ to ensure the overlap assumption (Assumption 2.2). $\epsilon \sim \text{Unif}([-1,1])$ in the data generation in synthetic experiments. We use sample splitting in our experiment, using the first half for nuisance estimation. In the experiment, we use the same threshold t for all estimators, which is given by Corollary 3.6. To choose the threshold, we first use the adaptive Hill estimator Boucheron and Thomas [2015] to get an estimation of EVI $\widehat{\gamma}_n$ and then set the threshold to be $t = 0.25 n^{(\widehat{\gamma}_n/(1+2\min\{1,\widehat{\gamma}_n\})}$ as in Theorem 3.5. The approximate exponential $\widehat{\alpha}_n$ is coefficient of $\log(\|U\|)$ in linear regression $\log(|Y|) \sim \log(\|U\|)$. For the adaptive Hill estimator Boucheron and Thomas [2015], we follow authors' choice for hyperparameters and choose $l_n = 30, r(\delta) = \sqrt{\log\log(n)}$ and

$$k = \min\left\{k \in \{l_n, \cdots, n\} \text{ and } \exists\, i \in \{l_n, \cdots, n\}, |\widehat{\gamma}(i) - \widehat{\gamma}(k)| > \frac{\widehat{\gamma}(i) r_n(\delta)}{\sqrt{i}}\right\} - 1,$$

where $\widehat{\gamma}(i) = \frac{1}{i}\sum_{j=1}^{i} \log \frac{\|U_{(j)}\|}{\|U_{(i+1)}\|}$.

We run logistic regression to estimate the propensity function and use random forest to model the outcome.

For the semi-synthetic experiment, we apply the same hyperparameter as above to estimate NETE. We shift the data to make it positive and normalize each dimension by its 10 % quantile. The Fig. 3 shows the rough distribution of the wavesurge data after these transformations.

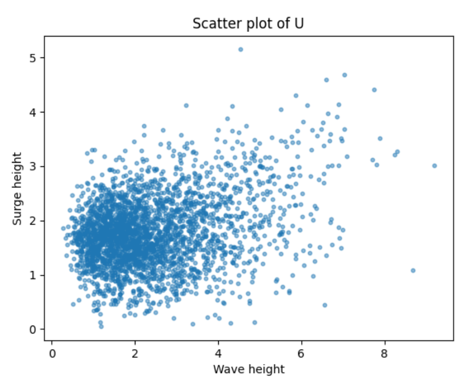

Figure 3: The scatter plot of wavesurge data.

We now describe how we calculate test-set estimation in our experiments. By the data generation process,

$$\begin{aligned}\theta^{\text{NETE}} &= \lim_{t\to\infty} \mathbb{E}[\frac{W^{\alpha_1} S^{\alpha_2}}{t^{\alpha_1+\alpha_2}} \mid \|U\| > t] \\ &= \lim_{t\to\infty} \mathbb{E}[\frac{W^{\alpha_1} S^{\alpha_2}}{\|U\|^{\alpha_1+\alpha_2}} \mid \|U\| > t] \cdot \frac{1}{1-(\alpha_1+\alpha_2)\gamma},\end{aligned}$$

where we use Proposition 3.3 in the second equality. We know the ground-truth α_1, α_2 and we can estimate the EVI γ using the test set. Suppose the estimated EVI is $\widehat{\gamma}$, we set the threshold to $t_n = 0.25 n^{(\widehat{\gamma}/(1+2\min\{1,\widehat{\gamma}\})}$ and get estimation

$$\widehat{\theta}_{\text{test}}^{\text{NETE}} = \mathbb{E}_n\big[\frac{W^{\alpha_1} S^{\alpha_2}}{\|U\|^{\alpha_1+\alpha_2}} \mid \|U\| > t_n\big] \cdot \frac{1}{1-(\alpha_1+\alpha_2)\widehat{\gamma}}.$$