SUPPLEMENT TO "POLICY LEARNING WITH OBSERVATIONAL DATA"

(*Econometrica*, Vol. 89, No. 1, January 2021, 133–161)

SUSAN ATHEY
Stanford Graduate School of Business, Stanford University

STEFAN WAGER
Stanford Graduate School of Business, Stanford University

APPENDIX A: CHARACTERIZING THE VC DIMENSION

AS A PRELIMINARY TO OUR TECHNICAL ARGUMENT, we start by reviewing some practical characterizations of the VC dimension in terms of covering numbers in Hamming distance. For any discrete set of points $\{X_1, \ldots, X_m\}$ and any $\varepsilon > 0$, define the ε-Hamming covering number $N_H(\varepsilon, \Pi, \{X_1, \ldots, X_m\})$ as the smallest number of policies $\pi : \{X_1, \ldots, X_m\} \to \{0, 1\}$ (not necessarily contained in Π) required to ε-cover Π under Hamming distance,

$$H(\pi_1, \pi_2) = \frac{1}{m} \sum_{j=1}^{m} \mathbf{1}\big(\{\pi_1(X_j) \neq \pi_2(X_j)\}\big). \tag{48}$$

Then, define the ε-Hamming entropy of Π as $\log(N_H(\varepsilon, \Pi))$, where

$$N_H(\varepsilon, \Pi) = \sup\big\{N_H\big(\varepsilon, \Pi, \{X_1, \ldots, X_m\}\big) : X_1, \ldots, X_m \in \mathcal{X}; m \geq 1\big\} \tag{49}$$

is the number of functions needed to ε-cover Π under Hamming distance for any discrete set of points. We note that this notion of entropy is purely geometric, and does not depend on the distribution used to generate the X_i.

As argued in Pakes and Pollard (1989), a class Π has a finite VC dimension if and only if there is a constant κ for which

$$\log\big(N_H(\varepsilon, \Pi_n)\big) \leq \kappa \log\big(\varepsilon^{-1}\big) \quad \text{for all } 0 < \varepsilon < \frac{1}{2}. \tag{50}$$

Moreover, there are simple quantitative bounds for Hamming entropy in terms of the VC dimension: If Π is a VC class of dimension $\mathrm{VC}(\Pi)$, then (Haussler (1995))

$$\begin{aligned} \log\big(N_H(\varepsilon, \Pi)\big) &\leq \mathrm{VC}(\Pi)\big(\log\big(\varepsilon^{-1}\big) + \log(2) + 1\big) + \log\big(\mathrm{VC}(\Pi) + 1\big) + 1 \\ &\leq 5\,\mathrm{VC}(\Pi)\log\big(\varepsilon^{-1}\big) \quad \text{for all } 0 < \varepsilon < \frac{1}{2} \end{aligned} \tag{51}$$

whenever $\mathrm{VC}(\Pi) \geq 2$. Conversely, recall that if Π has VC-dimension d it can shatter a set of d points, and so we must have $N_H(1/d, \Pi) \geq 2^d$. Thus, the VC dimension d of any class whose Hamming entropy satisfies (50) must be bounded via the relationship

$$d \log(2) \leq \kappa \log(d). \tag{52}$$

Susan Athey: athey@stanford.edu
Stefan Wager: swager@stanford.edu

 https://doi.org/10.3982/ECTA15732

Whenever we invoke Assumption 3 in our proof, we actually work in terms of the covering number bound (51) and assume that $\mathrm{VC}(\Pi) \geq 2$ (the case with $\mathrm{VC}(\pi) = 1$, corresponding to nonpersonalized decision rules, is trivial).

APPENDIX B: ADDITIONAL SIMULATION EXPERIMENTS

We complement our experiments from Section 5 with another simulation example where, now, the treatment dose $W_i \in \mathbb{R}$ is continuous. As discussed in Section 2.1, we consider policies that infinitesimally nudge the treatment dose W_i for select samples; the value $V(\pi)$ of a policy π is then

$$\pi : \mathcal{X} \to \{0, 1\}, \qquad V(\pi) = \mathbb{E}\left[\pi(X_i)\left(\left[\frac{d}{d\nu} Y_i(W_i + \nu)\right]_{\nu=0} - C\right)\right], \tag{53}$$

where C is a cost of treatment. We assume W_i to be exogenous. As always, we learn our policy $\hat{\pi}$ via $\hat{\pi} = \mathrm{argmax}\{\frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i) - 1)(\widehat{\Gamma}_i - C) : \pi \in \Pi\}$, and the $\widehat{\Gamma}_i$ are appropriate cross-fit doubly robust scores (15),

$$\begin{aligned}\widehat{\Gamma}_i = &\left[\frac{d}{dw}\hat{m}^{(-i)}(X_i, w)\right]_{w=W_i} \\ &- \frac{d}{dw}\left[\log\left(\hat{f}^{(-i)}(w|X_i)\right)\right]_{w=W_i}\left(Y_i - \hat{m}^{(-i)}(X_i, W_i)\right),\end{aligned} \tag{54}$$

where $f(\cdot|x)$ denotes the conditional density of W_i given $X_i = x$, and $m(x, w) = \mathbb{E}[Y_i|X_i = x, W_i = w]$.

Unlike in our previous examples, the nonparametric regression problems underlying (54) have not received much attention in the statistical learning literature. First, (54) requires estimating derivatives of conditional response-functions; but many popular machine learning methods, such as random forests or boosted trees, do not have differentiable predictive surfaces. Second, the problem of estimating a conditional density function $f(\cdot|x)$ presents its own numerical challenges.

Here, we approach the problem as follows. In order to make sure that the derivatives of $\hat{m}(\cdot)$ and $\hat{f}(\cdot)$ are good estimates of $m(\cdot)$ and $f(\cdot)$, respectively, we use penalized series estimators throughout. We fit $\hat{m}(X_i, W_i)$ by penalized regression on third-order Hermite polynomials in (X_i, W_i). Meanwhile, we fit the conditional density function $f(\cdot|X_i)$ by adapting Lindsey's method, a technique for estimating distribution functions using software for generalized linear modeling (Efron and Tibshirani (1996), Lindsey (1974)). In the case without covariates, Lindsey's method involves first discretizing the support of W_i into a union of nonoverlapping equal-length intervals and, as with a histogram, counting the number of samples W_i that fall within each interval. Then these histogram counts are fit via Poisson regression using a series expansion of W_i. As shown in Efron (2011), the log-derivative of the estimated density function is well behaved as an estimate of the log-derivative of the true density. Now, in the case with covariates, we again discretize the support of W_i into K nonoverlapping intervals. However, instead of making a histogram, we now duplicate each sample K times: For each sample $i = 1, \ldots, n$ and interval $k = 1, \ldots, K$ we create a datapoint (X_i, w_k, L_{ik}), where w_k is the mid-point of the kth interval and L_{ik} is an indicator for whether the W_i is in the kth interval. Finally, we fit this model by penalized logistic regression on full interactions between third-order Hermite polynomials in X_i and an appropriate basis expansion $b(w)$ in w discussed further below.

In all cases, we fit penalized regression via `glmnet` (Friedman, Hastie, and Tibshirani (2010)), with the amount of penalization tuned via cross-validation.

We consider the following simulation designs, loosely motivated by a probit choice model in a pricing application (i.e., where W_i acts as a price and Y_i is a choice to purchase). In all cases, we generate independent samples as below, with $p = 6$:

$$\begin{aligned} X_i &\sim \mathcal{N}(0, \mathcal{I}_{p\times p}), \qquad & U_i &= 5/\left(1 + 3e^{-(X_{i1}+X_{i2})}\right) - 0.5, \\ W_i|X_i &\sim \mathcal{L}_w(X_i), \qquad & Y_i|U_i, W_i &\sim \text{Bernoulli}\left(\Phi(W_i - U_i)\right), \end{aligned} \tag{55}$$

where $\Phi(\cdot)$ is the standard Gaussian cumulative distribution function. We consider two choices for the conditional distribution $\mathcal{L}_w$ of W_i conditionally on X_i:

$$\text{Gaussian:} \quad W_i = 3/\left(1 + 3e^{-(X_{i1}+X_{i3})}\right) + \varepsilon_i, \qquad \varepsilon_i|X_i \sim \mathcal{N}(0,1), \quad \text{and} \tag{56}$$

$$\text{Non-Gaussian:} \quad W_i = 3/\left(1 + 3e^{-(X_{i1}+X_{i3}+\eta_i)}\right) + \varepsilon_i, \qquad (\varepsilon_i, \eta_i)|X_i \sim \mathcal{N}(0, \mathcal{I}_{2\times 2}). \tag{57}$$

In principle, the Gaussian case appears substantially easier than the non-Gaussian case, because the logistic regression problem underlying Lindsey's method as above is well specified with a quadratic expansion in w, that is, $b(w) = (1 w w^2)$. In the non-Gaussian case, no similar simplifications apply. In our experiments, we in fact set $b(w)$ to be the quadratic expansion in the Gaussian case; in the non-Gaussian case, we set $b(w)$ to a fifth-order natural spline basis.

Before evaluating the accuracy of policy learning in this setting, we present some performance diagnostics on the associated doubly robust average derivative estimator $\hat{\theta}_{\text{DR}} = \sum_{i=1}^n \widehat{\Gamma}_i / n$ as, despite attracting a fair amount of interest in the literature on asymptotic estimation (including Chernozhukov, Escanciano, Ichimura, Newey, and Robins (2016), Chernozhukov, Newey, and Robins (2018), Hirshberg and Wager (2018)), we are not aware of existing Monte Carlo evaluations of this estimator in the literature.[1] We report bias and root-mean squared error for the doubly robust estimator $\hat{\theta}_{DR}$, the pure regression estimator $\hat{\theta}_{reg} = \sum_{i=1}^n d/dw \hat{m}^{(-i)}(X_i, W_i)/n$, and the pure weighting estimator $\hat{\theta}_{\text{weight}} = \sum_{i=1}^n d/dw \log \hat{f}^{(-i)}(X_i, W_i) Y_i / n$. We also report mean-squared standardized error $S = \mathbb{E}[(\hat{\theta}_{\text{DR}} - \theta)^2 / \hat{\sigma}^2]^{1/2}$ with $\hat{\sigma}^2 = \sum_{i=1}^n \widehat{\Gamma}_i / (n(n-1))$ which, under the conditions of Assumption 2, should converge as $\lim_{n\to\infty} S = 1$.

Table III shows results for both average derivative estimation as described above, and for policy learning with doubly robust scores. For policy learning, we use a cost of treatment parameter $C = 0.2$. First, encouragingly, we see that the doubly robust estimator of the average derivative, $\hat{\theta}_{\text{DR}}$, converges with sample size n, and that the value of our learned policies improves with n. Furthermore, we see that the doubly robust estimator outperforms the pure regression adjustment and weighting estimators here. However, even the doubly-robust estimator is still bias-dominated here, and the root-mean squared standardized error S is much bigger than 1 in all considered settings—especially the challenging ones with a non-Gaussian distribution of $W_i|X_i$. This suggests that the simulation problem considered here is a difficult nonparametric problem where semiparametric efficiency asymptotics kick in slowly at best. It is plausible that a more carefully tailored

[1]The closest experiments we are aware from are from Graham and Pinto (2018) and Hirshberg and Wager (2018), who report results results for doubly robust estimation in a closely related (but more restricted) model with a conditionally linear specification $\mathbb{E}[Y_i|X_i = x, W_i = w] = m(x) + w\tau(x)$.

TABLE III

SIMULATION RESULTS IN THE SETTING (55), WITH CONDITIONAL DISTRIBUTION OF $W_i|X_i$ AS IN (56) (SETUP 1) AND (57) (SETUP 2). WE REPORT BIAS AND ROOT-MEAN SQUARED ERROR FOR THE AVERAGE DERIVATE θ BASED ON THE REGRESSION ESTIMATOR $\hat{\theta}_{reg}$, THE WEIGHTED ESTIMATOR $\hat{\theta}_{weighted}$, AND THE DOUBLY ROBUST ESTIMATOR $\hat{\theta}_{DR}$. THE ROOT MEAN-SQUARED STANDARDIZED ERROR S CAPTURES THE ASYMPTOTIC BEHAVIOR OF STANDARD GAUSSIAN CONFIDENCE INTERVALS FOR θ BASED ON $\hat{\theta}_{DR}$. FINALLY, THE LAST COLUMN REPORTS POLICY VALUE OBTAINED BY LEARNING WITH DOUBLY ROBUST SCORES OVER THE CLASS Π OF DEPTH-2 TREES

		Regression		Weighted		Doubly Robust			
	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	S	Policy Value
Setup 1	600	−0.056	0.058	−0.132	0.133	−0.035	0.037	4.59	0.014
	1800	−0.035	0.036	−0.095	0.096	−0.017	0.019	2.97	0.024
	5400	−0.022	0.022	−0.081	0.081	−0.010	0.010	2.60	0.028
	16,200	−0.012	0.013	−0.073	0.073	−0.006	0.006	2.54	0.029
Setup 2	600	−0.069	0.072	−0.062	0.063	−0.049	0.050	8.01	0.018
	1800	−0.040	0.041	−0.052	0.053	−0.026	0.027	6.26	0.033
	5400	−0.023	0.024	−0.053	0.054	−0.014	0.014	5.17	0.035
	16,200	−0.015	0.015	−0.056	0.056	−0.009	0.009	5.25	0.037

estimator of the weighting function $d/dw \log f(x, w)$ following the lines of, for example, Chernozhukov, Newey, and Robins (2018) or Hirshberg and Wager (2018) could improve performance here.

APPENDIX C: PROOFS

C.1. *Proof of Lemma 2*

Our proof of this result follows the outline of the classical chaining argument of Dudley (1967), whereby we construct a sequence of approximating sets of increasing precision for $\widetilde{A}_n(\pi)$ with $\pi \in \Pi_n^\lambda$, and then use finite sample concentration inequalities to establish the behavior of $\widetilde{A}_n(\pi)$ over this approximation set. The improvements in our results relative to existing bounds described in the body of the text come from a careful construction of approximating sets targeted to the problem of doubly robust policy evaluation—for example, our use of chaining with respect to the random distance measure defined in (58)—and the use of sharp concentration inequalities.

Given these preliminaries, we start by defining the conditional 2-norm distance between two policies π_1, π_2 as

$$D_n^2(\pi_1, \pi_2) = \sum_{i=1}^n \Gamma_i^2 \big(\pi_1(X_i) - \pi_2(X_i)\big)^2 \Big/ \sum_{i=1}^n \Gamma_i^2, \tag{58}$$

and let $N_{D_n}(\varepsilon, \Pi_n^\lambda, \{X_i, \Gamma_i\})$ be the ε-covering number in this distance. To bound N_{D_n}, imagine creating another sample $\{X_j'\}_{j=1}^m$, with X_j' contained in the support of $\{X_i\}_{i=1}^n$, such that

$$\left| \big|\{j \in 1, \dots, m : X_j' = X_i\}\big| - m\Gamma_i^2 \Big/ \sum_{j=1}^n \Gamma_j^2 \right| \le 1.$$

We immediately see that, for any two policies π_1 and π_2,

$$\frac{1}{m}\sum_{j=1}^{m} 1\big(\{\pi_1(X_j') \neq \pi_2(X_j')\}\big) = D_n^2(\pi_1, \pi_2) + \mathcal{O}\left(\frac{1}{m}\right).$$

Moreover, recall that the Hamming covering number N_H as used in (50) does not depend on sample size, so we can without reservations make m arbitrarily large, and conclude that

$$N_{D_n}\big(\varepsilon, \Pi_n, \{X_i, \Gamma_i\}\big) \leq N_H\big(\varepsilon^2, \Pi_n\big). \tag{59}$$

In other words, we have found that we can bound the D_n-entropy of Π_n with respect to its distribution-independent Hamming entropy, which is controlled via Assumption 3.

Our proof strategy involves a chaining argument with respect to D_n. The lemma below describes the chaining that we use in our argument; we defer the proof of Lemma 6 to the end of this section.

LEMMA 6: *For any $J \geq 1$, there exists a chain of approximators $\Psi_j : \Pi_n^\lambda \to \Pi_n^\lambda$ for $j = 1, \ldots, J$, such that the following properties hold for all values of $j = 1, \ldots, J$ (we use the notational shorthand $\Psi_{J+1}(\pi) = \pi$):*

- *The approximation is accurate, that is, $D_n(\Psi_j(\pi), \Psi_{j+1}(\pi)) \leq 2^{-j}$ for all $\pi \in \Pi_n^\lambda$;*
- *There is no branching, such that $\Psi_j(\pi) = \Psi_j(\Psi_{j+1}(\pi))$ for all $\pi \in \Pi_n^\lambda$; and*
- *The set $\Pi_n^\lambda(j) := \{\Psi_j(\pi) : \pi \in \Pi_n^\lambda\}$ of jth order approximating policies has cardinality at most $N_{D_n}(2^{-(j+1)}, \Pi_n, \{X_i, \Gamma_i\})$.*

We now move to our main task, that is, bounding the Rademacher complexity $\mathbb{E}[\mathcal{R}_n(\Pi_n^\lambda)]$. In order to do so, we use a two-step strategy. We first prove the following weaker result below, with a bound that depends only on the worst-case variance S_n rather than the slice-adapted variance $S_n^\lambda \leq S_n$. We then use this bound to sharpen our argument and prove the desired bound (26).

LEMMA 7: *Under the conditions of Lemma 2 and for any λ,*

$$\limsup_{n\to\infty} \mathbb{E}\big[\mathcal{R}_n\big(\Pi_n^\lambda\big)\big] \Big/ \sqrt{\frac{S_n \operatorname{VC}(\Pi_n)}{n}} \leq 52. \tag{60}$$

PROOF: To start, it is helpful to decompose the random variable into several parts using the chaining established in Lemma 6. In doing so, the following thresholds play a key role:

$$J_0 := 1, \quad J(n) := \big\lfloor \log_2(n)(3-2\beta)/8 \big\rfloor, \quad \text{and} \quad J_+(n) := \big\lfloor \log_2(n)(1-\beta) \big\rfloor. \tag{61}$$

We then apply Lemma 6 to create a chain with $J = J_+(n)$ terms and note that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \xi_i \Gamma_i \big(2\pi(X_i) - 1\big) \\
&\quad = \frac{1}{n}\sum_{i=1}^{n} \xi_i \Gamma_i \big(2\Psi_{J_0}(\pi)(X_i) - 1\big)
\end{aligned}$$

$$+\sum_{j=J_0+1}^{J(n)}\frac{2}{n}\sum_{i=1}^{n}\xi_i\Gamma_i\big(\Psi_j(\pi)(X_i)-\Psi_{j-1}(\pi)(X_i)\big)$$

$$+\sum_{j=J(n)+1}^{J_+(n)}\frac{2}{n}\sum_{i=1}^{n}\xi_i\Gamma_i\big(\Psi_j(\pi)(X_i)-\Psi_{j-1}(\pi)(X_i)\big)$$

$$+\frac{2}{n}\sum_{i=1}^{n}\xi_i\Gamma_i\big(\pi(X_i)-\Psi_{J_+(n)}(\pi)(X_i)\big), \tag{62}$$

for any $\pi\in\Pi_n^\lambda$. Note that, for now, the first threshold $J_0=1$ is trivial; however, once we want to prove the stronger bound (26) instead of (60) we will need a more careful choice of J_0, so we already introduce this flexibility now for notational consistency.

We now proceed to successively control the $1/\sqrt{n}$-scale behavior of all four terms above, uniformly over all $\pi\in\Pi_n^\lambda$. The result will be that the first term can be characterized directly via Bernstein's inequality; the second term is controlled to $1/\sqrt{n}$-scale by chaining; the third term is shown to stochastically vanish at $1/\sqrt{n}$-scale by chaining; and the last term is shown to deterministically vanish at $1/\sqrt{n}$-scale.

Before embarking on this task, we recall Bernstein's inequality, which will be frequently used throughout the proof:

$$\mathbb{P}\left[\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}U_i\right|\geq t\right]\leq 2\exp\left[\frac{-t^2}{2}\Big/\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big[U_i^2\big]+\frac{Mt}{3\sqrt{n}}\right)\right], \tag{63}$$

for any independent, mean-zero variables U_i with $|U_i|\leq M$, and any constant $t>0$. To make use of this inequality, it is helpful to restrict ourselves to a study of $\mathcal{R}_n(\Pi_n^\lambda)$ on the event

$$\mathcal{B}_n=\left\{M_n\leq n^{\frac{1-2\beta}{16}}\text{ and }\widehat{\mathrm{Var}}\big[\big(2\pi(X_i)-1\big)\Gamma_i\big]\geq\frac{s^2}{2}\text{ for all }\pi\in\Pi_n^\lambda(J_0)\right\}, \tag{64}$$

where $M_n=\max_{i=1,\ldots,n}\{|\Gamma_i|\}$ and $0<\beta<1/2$ is the constant from Assumption 3. Recall that, by assumption, Γ_i is sub-Gaussian and $\mathrm{Var}[\Gamma_i|X_i]>s^2$, and so a simple calculation can be used to check that $\lim_{n\to\infty}\mathbb{P}[\mathcal{B}_n]=1$ and, furthermore,

$$\lim_{n\to\infty}\sqrt{n}\big(\mathbb{E}\big[\mathcal{R}_n\big(\Pi_n^\lambda\big)\big]-\mathbb{E}\big[\mathcal{R}_n\big(\Pi_n^\lambda\big)1(\mathcal{B}_n)\big]\big)=0. \tag{65}$$

Thus, for the rest of this proof, we will assume that the event $\mathcal{B}_n$ has occurred when convenient.

First Term. Because the chaining created in Lemma 6 has no branching, we see that

$$\sup\left\{\frac{1}{n}\sum_{i=1}^{n}\xi_i\Gamma_i\big(2\Psi_{J_0}(\pi)(X_i)-1\big):\pi\in\Pi_n^\lambda\right\}$$

$$=\sup\left\{\frac{1}{n}\sum_{i=1}^{n}\xi_i\Gamma_i\big(2\pi(X_i)-1\big):\pi\in\Pi_n^\lambda(J_0)\right\}. \tag{66}$$

Then, applying a union bound with Bernstein's inequality (63) on the event $\mathcal{B}_n$ in (64), we see that, for all large enough n and all $t \leq 2\widehat{S}^{0.5}\sqrt{\log(n) + \log(2|\Pi_n^\lambda(J_0)|)}$,

$$\begin{aligned}
1(\mathcal{B}_n)\mathbb{P}\left[\sqrt{n}\sup\left\{\frac{1}{n}\sum_{i=1}^n \xi_i\Gamma_i\big(2\pi(X_i)-1\big) : \pi \in \Pi_n^\lambda(J_0)\right\} \geq t \middle| \{X_i, \Gamma_i\}\right] \\
\leq 2\big|\Pi_n^\lambda(J_0)\big| \exp\left[-\frac{t^2}{2}/\big(\widehat{S} + tn^{-\frac{7+2\beta}{16}}/3\big)\right] \\
\leq 2\big|\Pi_n^\lambda(J_0)\big| \exp\left[-\frac{t^2}{4\widehat{S}}\right], \qquad (67)
\end{aligned}$$

where $\widehat{S} = \sum_{i=1}^n \Gamma_i^2/n$. Now, to bound expectations, we note the following fact: If a non-negative random variable satisfies $X \leq c_k$ with probability $1-2^{-k}$ for all $k = 1, 2, \ldots$, then $\mathbb{E}[X] \leq \sum_{k=1}^\infty 2^{-k}c_k$. Thus, applying the above bound for the choice

$$t_k = 2\widehat{S}^{0.5}\sqrt{\min\big\{k\log(2), \log(n)\big\} + \log\big(2\big|\Pi_n^\lambda(J_0)\big|\big)}, \quad k = 1, 2, \ldots, \lceil \log(n)/\log(2)\rceil$$

we then find that (the last term corresponds to a loose $\max|\Gamma_i|/n$ when all events fail)

$$\begin{aligned}
1(\mathcal{B}_n)\mathbb{E}\left[\sqrt{n}\sup\left\{\frac{1}{n}\sum_{i=1}^n \xi_i\Gamma_i\big(2\pi(X_i)-1\big) : \pi \in \Pi_n^\lambda(J_0)\right\} \middle| \{X_i, \Gamma_i\}\right] \\
\leq 2\widehat{S}^{0.5}\left(\sqrt{\log\big|\Pi_n^\lambda(J_0)\big|} + \sum_{k=1}^\infty 2^{-k}\sqrt{(k+1)\log(2)}\right) + n^{-\frac{7+2\beta}{16}} \\
\leq 2\widehat{S}^{0.5}\big(\sqrt{\log N_H(1/16, \Pi_n)} + 1.5\big) + n^{-\frac{7+2\beta}{16}} \\
\leq 2\widehat{S}^{0.5}\big(\sqrt{5\log(16)\,\mathrm{VC}(\Pi_n)} + 1.5\big) + n^{-\frac{7+2\beta}{16}} \\
\leq 11\sqrt{\widehat{S}\,\mathrm{VC}(\Pi_n)} + n^{-\frac{7+2\beta}{16}}, \qquad (68)
\end{aligned}$$

where for the third line we used Lemma 6 and (59) whereas for the last line we used Assumption 3 together with (51). Finally, noting that

$$\mathbb{E}[\sqrt{\widehat{S}}] \leq \sqrt{S_n} \qquad (69)$$

by concavity of the square-root function, we see that

$$\limsup_{n\to\infty} \mathbb{E}\left[1(\mathcal{B}_n)\sqrt{\frac{n}{S_n\,\mathrm{VC}(\Pi_n)}}\sup\left\{\frac{1}{n}\sum_{i=1}^n \xi_i\Gamma_i\big(2\pi(X_i)-1\big) : \pi \in \Pi_n^\lambda(J_0)\right\}\right] \leq 11. \qquad (70)$$

Second Term. First, we check that, for any choice of $\pi \in \Pi_n^\lambda$, $j = 1, \ldots, J$ and $t > 0$, we have

$$\mathbb{P}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Gamma_i\xi_i\big(\Psi_j(\pi)(X_i) - \Psi_{j+1}(\pi)(X_i)\big)\right| \geq t2^{-j}\sqrt{\widehat{\bar{S}}}\,|\{X_i, \Gamma_i\}\right] \leq 2\exp\left[\frac{-t^2}{2}\left(1 + \frac{1}{3}\frac{M_n t 2^j}{\sqrt{n\widehat{S}}}\right)^{-1}\right], \tag{71}$$

where $\widehat{S} = \sum_{i=1}^n \Gamma_i^2/n$, $M_n = \max\{|\Gamma_i| : 1 \leq i \leq n\}$. This can be verified using Bernstein's inequality (63), which establishes that, for any choice of $t > 0$, $\pi \in \Pi_n^\lambda$ and $j = 1, 2, \ldots, J$,

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Gamma_i\xi_i\big(\Psi_j(\pi)(X_i) - \Psi_{j+1}(\pi)(X_i)\big)\right| \geq t2^{-j}\sqrt{\widehat{\bar{S}}}\,|\{X_i, \Gamma_i\}\right] \\
&\leq 2\exp\left[\frac{-t^2 4^{-j}\widehat{S}}{2}\Big/\left(\frac{1}{n}\sum_{i=1}^{n}\Gamma_i^2 1\big(\big\{\Psi_j(\pi)(X_i) \neq \Psi_{j+1}(\pi)(X_i)\big\}\big) + \frac{M_n t 2^{-j}\sqrt{\widehat{\bar{S}}}}{3\sqrt{n}}\right)\right] \\
&= 2\exp\left[\frac{-t^2}{2}4^{-j}\widehat{S}\Big/\left(D_n^2\big(\Psi_j(\pi), \Psi_{j+1}(\pi)\big)\widehat{S} + \frac{M_n t 2^{-j}\sqrt{\widehat{\bar{S}}}}{3\sqrt{n}}\right)\right].
\end{aligned}$$

Finally recall that, by Lemma 6, $D_n^2(\Psi_j(\pi), \Psi_{j+1}(\pi)) \leq 4^{-j}$; thus

$$4^{-j}\widehat{S}\Big/\left(D_n^2\big(\Psi_j(\pi), \Psi_{j+1}(\pi)\big)\widehat{S} + \frac{M_n t 2^{-j}\sqrt{\widehat{\bar{S}}}}{3\sqrt{n}}\right) \geq \left(1 + \frac{1}{3}\frac{M_n t 2^j}{\sqrt{n\widehat{S}}}\right)^{-1},$$

and so (71) follows.

Now, or every $j \geq J_0$ and $\delta > 1/(2n)$, define the event

$$\begin{aligned}
\mathcal{E}_{j,\delta} &:= \left\{\sup_{\pi\in\Pi_n^\lambda}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Gamma_i\xi_i\big(\Psi_j(\pi)(X_i) - \Psi_{j+1}(\pi)(X_i)\big)\right| \geq 2^{-j}t_{j,\delta}\sqrt{\widehat{\bar{S}}}\right\}, \\
t_{j,\delta} &:= 2\sqrt{7(j+2)\,\mathrm{VC}(\Pi_n) + \log\left(\frac{2j^2}{\delta}\right)}.
\end{aligned} \tag{72}$$

By (71), we immediately see that

$$\mathbb{P}\big[\mathcal{E}_{j,\delta}|\{X_i, \Gamma_i\}\big] \leq 2\big|\Pi_n^\lambda(j+1)\big|\exp\left[\frac{-t_{j,\delta}^2}{2}\left(1 + \frac{1}{3}\frac{M_n t_{j,\delta} 2^j}{\sqrt{n\widehat{S}}}\right)^{-1}\right]. \tag{73}$$

By invoking Assumption 3, Lemma 6 and (59) along with the fact that $5\log(4) < 7$, we see that

$$\log\big(\big|\Pi_n^\lambda(j+1)\big|\big) \leq \log\big(N_H\big(4^{-(j+2)}, \Pi_n\big)\big) \leq 7(j+2)\,\mathrm{VC}(\Pi_n). \tag{74}$$

Moreover, on the event $\mathcal{B}_n$ from (64) and recalling Assumption 3 along with the definition of $J(n)$, we see that

$$\begin{aligned}\frac{1}{3}\frac{M_n t_{j,\delta} 2^j}{\sqrt{n\widehat{S}}} &\leq \frac{2}{3}\frac{n^{\frac{1-2\beta}{16}}\sqrt{7\big(J(n)+2\big)\,\mathrm{VC}(\Pi_n)+\log\big(2nJ(n)^2\big)}2^{J(n)}}{\sqrt{ns^2/2}} \\ &= \exp\left[\log(n)\left(\frac{1-2\beta}{16}+\frac{\beta}{2}+\frac{3-2\beta}{8}-\frac{1}{2}\right)\right]\cdot \mathrm{polylog}(n) \\ &= n^{\frac{2\beta-1}{16}}\cdot \mathrm{polylog}(n) \leq 1\end{aligned}$$

for large enough values of n, simultaneously for all $j \leq J(n)$ and $\delta \geq 1/(2n)$, because $\beta < 1/2$. Thus, for large enough values of n, the bound (73) simplifies dramatically, and we get

$$1(\mathcal{B}_n)\mathbb{P}\big[\mathcal{E}_{j,n}|\{X_i,\Gamma_i\}\big] \leq \frac{\delta}{j^2}. \tag{75}$$

Applying this bound simultaneously to $j = J_0, \ldots, J(n)-1$:

$$1(\mathcal{B}_n)\mathbb{P}\left[\bigcup_{j=J_0}^{J(n)-1}\mathcal{E}_{j,n}\Big|\{X_i,\Gamma_i\}\right] \leq \sum_{j=J_0}^{J(n)-1}\frac{\delta}{j^2} \leq 2\delta. \tag{76}$$

Thus, for large enough n, we can directly verify that, with probability at least $1-2\delta$,

$$\begin{aligned}&\sqrt{n}1(\mathcal{B}_n)\sup_{\pi\in\Pi_n^\lambda}\left|\frac{2}{n}\sum_{i=1}^n\Gamma_i\xi_i\sum_{j=J_0}^{J(n)-1}\big(\Psi_{j+1}(\pi)-\Psi_j(\pi)\big)(X_i)\right| \\ &\leq 4\sqrt{\widehat{S}}\sum_{j=J_0}^{J(n)-1}2^{-j}\sqrt{7(j+2)\,\mathrm{VC}(\Pi_n)+\log\left(\frac{2j^2}{\delta}\right)} \\ &\leq 4\sqrt{\widehat{S}}\left(\sqrt{7\,\mathrm{VC}(\Pi_n)}\sum_{j=J_0}^{J(n)-1}2^{-j}\sqrt{j+2}+\sum_{j=J_0}^{J(n)-1}2^{-j}\sqrt{\log\big(2j^2\big)}+2^{1-J_0}\sqrt{\log\big(\delta^{-1}\big)}\right).\end{aligned}$$

Moreover, we can check by calculus that, for all $J_0 \geq 2$,

$$\begin{aligned}\sum_{j=J_0}^{J(n)-1}2^{-j}\sqrt{j+2} &\leq 2^{-J_0}\sum_{j=0}^{\infty}2^{-j}\left(\sqrt{J_0}+\frac{j+2}{2\sqrt{J_0}}\right) = 2\times 2^{-J_0}\sqrt{J_0}+3\times 2^{-J_0}, \\ \sum_{j=J_0}^{J(n)-1}2^{-j}\sqrt{\log\big(2j^2\big)} &\leq 2^{-J_0}\sum_{j=0}^{\infty}2^{-j}\left(\sqrt{\log\big(2J_0^2\big)}+\frac{2\log(J_0+j)-2\log(J_0)}{2\sqrt{\log\big(2J_0^2\big)}}\right) \\ &\leq 2^{-J_0}\sum_{j=0}^{\infty}2^{-j}\left(\sqrt{\log\big(2J_0^2\big)}+\frac{j}{J_0\sqrt{\log\big(2J_0^2\big)}}\right)\end{aligned}$$

$$= 2 \times 2^{-J_0}\left(\sqrt{\log\left(2J_0^2\right)} + \frac{1}{J_0\sqrt{\log\left(2J_0^2\right)}}\right)$$
$$\leq 4 \times 2^{-J_0}\sqrt{J_0};$$

moreover, the same final upper bounds can be verified directly for $J_0 = 1$. Thus the above expression can further be bounded by

$$\cdots \leq 4\sqrt{\widehat{S}}2^{-J_0}\left(\sqrt{7\,\mathrm{VC}(\Pi_n)}(2\sqrt{J_0}+3) + 4\sqrt{J_0} + 2\sqrt{\log\left(\delta^{-1}\right)}\right).$$

Next, we bound expectations as in (68), and apply the above bound separately for the sequences $2\delta = \max\{2^{-k}, 1/n\}$ for $k = 1, 2, \ldots$ to show that, again for large enough n,

$$\begin{aligned}
&\sqrt{n}\mathbb{E}\left[1(\mathcal{B}_n)\sup_{\pi\in\Pi_n^\lambda}\left|\frac{2}{n}\sum_{i=1}^n \Gamma_i\xi_i\sum_{j=J_0}^{J(n)-1}\left(\Psi_{j+1}(\pi)-\Psi_j(\pi)\right)(X_i)\right|\right] \\
&\quad\leq 4\times 2^{-J_0}\left(\sqrt{7\,\mathrm{VC}(\Pi_n)}(2\sqrt{J_0}+3) + 4\sqrt{J_0} + 2\sum_{k=1}^{\infty}2^{-k}\sqrt{(k+1)\log(2)}\right)\mathbb{E}[\sqrt{\widehat{S}}] \\
&\quad\leq 2^{-J_0}\sqrt{\mathrm{VC}(\Pi_n)}(38\sqrt{J_0}+44)\mathbb{E}[\sqrt{\widehat{S}}] \\
&\quad\leq 2^{-J_0}\sqrt{S_n\,\mathrm{VC}(\Pi_n)}(38\sqrt{J_0}+44), \qquad (77)
\end{aligned}$$

where we note that the contribution of terms on the residual with-probability-$1/n$ scale as $M_n/n \ll 1/\sqrt{n}$ on $\mathcal{B}_n$ (64), and for the last inequality we also use (69). We thus conclude that

$$\limsup_{n\to\infty}\sqrt{\frac{n}{S_n\,\mathrm{VC}(\Pi_n)}}\mathbb{E}\left[1(\mathcal{B}_n)\sup_{\pi\in\Pi_n^\lambda}\left|\frac{1}{n}\sum_{i=1}^n \Gamma_i\xi_i\sum_{j=J_0}^{J(n)-1}\left(\Psi_{j+1}(\pi)-\Psi_j(\pi)\right)(X_i)\right|\right] \leq 41, \quad (78)$$

recalling our choice of $J_0 = 1$ from (61).

Third Term. We now verify that terms $\Psi_j(\pi)(X_i) - \Psi_{j+1}(\pi)(X_i)$ in (62) with $J(n) \leq j < J_+(n)$ are asymptotically negligible. To do so, we collapse all approximating policies with $J(n) \leq j < J_+(n)$, and directly compare $\Psi_{J(n)}(\pi)$ to $\Psi_{J_+(n)}(\pi)$. Because of our "no branching" construction, we know that $\Psi_{J(n)}(\pi) = \Psi_{J(n)}(\Psi_{J_+(n)}(\pi))$ for all policies $\pi \in \Pi_n^\lambda$, and so

$$\begin{aligned}
&\mathbb{P}\left[\sup\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \Gamma_i\xi_i\left(\Psi_{J(n)}(\pi)(X_i) - \Psi_{J_+(n)}(\pi)(X_i)\right)\right| : \pi\in\Pi_n^\lambda\right\} \geq 2\times t2^{-J(n)}\sqrt{\widehat{S}}\right] \\
&\quad= \mathbb{P}\left[\sup\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \Gamma_i\xi_i\left(\Psi_{J(n)}(\pi)(X_i) - \pi(X_i)\right)\right| : \pi\in\Pi_n^\lambda\left(J_+(n)\right)\right\} \geq 2\times t2^{-J(n)}\sqrt{\widehat{S}}\right] \\
&\quad\leq 2\left|\Pi_n^\lambda\left(J_+(n)\right)\right|\exp\left[\frac{-t^2}{2}\left(1+\frac{1}{6}\frac{M_n t 2^{J(n)}}{\sqrt{n\widehat{S}}}\right)^{-1}\right],
\end{aligned}$$

where the last inequality follows from Bernstein's inequality using exactly the same arguments as those used to establish (71). By Lemma 6, Assumption 3 and (51), we get

$$\begin{aligned}\log\bigl|\Pi_n^\lambda\bigl(J_+(n)\bigr)\bigr| &\le \log N_{D_n}\bigl(2^{-(J_+(n)+1)}, \Pi_n, \{X_i, \Gamma_i\}\bigr)\\ &\le \log N_H\bigl(4^{-(J_+(n)+1)}, \Pi_n\bigr)\\ &\le 5\log(4)\bigl(J_+(n)+1\bigr)n^\beta. \end{aligned} \tag{79}$$

The next step is to plug $t^2 = 4^{J(n)}n^{(2\beta-1)/4}/\widehat{S}$ into the previous bound. Given this choice along with Assumption 3 and (61) we see that, on event $\mathcal{B}_n$ from (64),

$$t2^{J(n)}/\sqrt{n} \ge \times 2^{2J(n)}n^{\frac{2\beta-5}{8}}n^{\frac{-1+2\beta}{16}} \ge n^{\frac{1-2\beta}{16}}/4$$

which grows with n, and so the bound simplifies on event $\mathcal{B}_n$ and for large enough n:

$$\begin{aligned}&\mathbb{P}\bigl[1(\mathcal{B}_n)\Delta_{\text{mid}}\bigl(\Pi_n^\lambda\bigr)\\ &\quad\ge 2n^{\frac{2\beta-1}{8}}\bigr] \le 1(\mathcal{B}_n)2\bigl|\Pi_n^\lambda\bigl(J_+(n)\bigr)\bigr|\exp\left[\frac{-(3/2)t\sqrt{n\widehat{S}}}{2^{J(n)}\max\{M_n,1\}}\right]\\ &\quad\le 2\exp\left[\sqrt{n}\left(5\log(4)\bigl(J_+(n)+1\bigr)n^{\beta-1/2} - \frac{3}{2}n^{\frac{6\beta-3}{16}}\right)\right], \quad \text{where}\\ &\Delta_{\text{mid}}\bigl(\Pi_n^\lambda\bigr) = \sup\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \Gamma_i\xi_i\bigl(\Psi_{J(n)}(\pi)(X_i) - \Psi_{J_+(n)}(X_i)\bigr)\right| : \pi\in\Pi_n^\lambda\right\}.\end{aligned}$$

Thus, noting that $\beta < 1/2$, we see that

$$\limsup_{n\to\infty} n^{\frac{5+6\beta}{16}}\log\bigl(\mathbb{P}\bigl[1(\mathcal{B}_n)\Delta_{\text{mid}}\bigl(\Pi_n^\lambda\bigr)\bigr] \ge 2n^{\frac{2\beta-1}{8}}\bigr) \le -\frac{3}{2}.$$

Meanwhile, we also know that $1(\mathcal{B}_n)\Delta_{\text{mid}}(\Pi_n^\lambda)/\sqrt{n} \le n^{(1-2\beta)/16}$, and so we conclude that

$$\lim_{n\to\infty}\mathbb{E}\left[\sup\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \Gamma_i\xi_i\bigl(\Psi_{J(n)}(\pi)(X_i) - \Psi_{J_+(n)}(X_i)\bigr)\right| : \pi\in\Pi_n^\lambda\right\}\right] = 0,$$

meaning that the third group of terms in the chaining (62) in fact do not contribute to the first-order behavior of the Rademacher complexity.

Fourth Term. Finally, the last term in (62) can be shown to vanish at $1/\sqrt{n}$-scale determinisitically. By Cauchy–Schwarz,

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^n \Gamma_i\xi_i\bigl(\pi(X_i) - \Psi_{J_+(n)}(\pi)(X_i)\bigr)\right| &\le \sqrt{\frac{1}{n}\sum_{i=1}^n \Gamma_i^2\bigl(\pi(X_i) - \Psi_{J_+(n)}(\pi)(X_i)\bigr)^2}\\ &= D_n\bigl(\pi, \Psi_{J_+(n)}(\pi)\bigr)\sqrt{\widehat{S}}\\ &\le 2^{-J_+(n)}\sqrt{\widehat{S}}.\end{aligned}$$

Furthermore, recalling the definition of $J_+(n)$ from (61) and on the event where M_n is controlled as in (64),

$$\lim_{n\to\infty} \sqrt{n} 2^{-J_+(n)} \sqrt{\widehat{S}} \leq 2\sqrt{n} n^{\beta-1} n^{\frac{1-2\beta}{16}} = n^{\frac{14\beta-7}{16}} = 0,$$

because $\beta < 1/2$ by Assumption 3.

Wrapping up Lemma 7. Combining (70) with (78) with our above results showing that the third and fourth terms in (62) are asymptotically negligible, we recover (60). *Q.E.D.*

We now turn to proving Lemma 2 itself, and specifically the bound (26). In doing so, we follow the proof of Lemma 7 closely, but with slightly stronger concentration bounds that are unlocked by the result we already have in Lemma 7. We also replace the choice $J_0 = 1$ in (61) with

$$J_0 := 9 + \left\lfloor \log_4\left(S_n / S_n^\lambda\right) \right\rfloor. \tag{80}$$

In the resulting new decomposition (62), we note that the third and fourth terms are still vanishing at the $1/\sqrt{n}$-scale, so we do not need to revisit those. Thus, our only task is to sharpen our bounds on the first and second terms.

The main additional work we need to do is in bounding the first term. Starting from (66) we note that, because the ξ_i are all mean-zero,

$$\begin{aligned} &\mathbb{E}\left[\sup\left\{ \frac{1}{n}\sum_{i=1}^n \xi_i \Gamma_i \left(2\pi(X_i) - 1\right) : \pi \in \Pi_n^\lambda(J_0) \right\} \right] \\ &\quad = \mathbb{E}\left[\sup\left\{ \frac{1}{n}\sum_{i=1}^n \xi_i \left(\Gamma_i \left(2\pi(X_i) - 1\right) - A_n^*\right) : \pi \in \Pi_n^\lambda(J_0) \right\} \right], \end{aligned} \tag{81}$$

where $A_n^* = \sup\{A_n(\pi) : \pi \in \Pi_n^\lambda\}$. Then, applying Bernstein's inequality as in (67), we get that for all large enough n and all $t \leq 2\widehat{S}_{\max}^{0.5}\sqrt{\log(n) + \log(2|\Pi_n^\lambda(J_0)|)}$,

$$\begin{aligned} &1(\mathcal{B}_n)\mathbb{P}\left[\sqrt{n} \sup\left\{ \frac{1}{n}\sum_{i=1}^n \xi_i \Gamma_i \left(2\pi(X_i) - 1\right) : \pi \in \Pi_n^\lambda(J_0) \right\} \geq t \middle| \{X_i, \Gamma_i\} \right] \\ &\quad \leq 2\left|\Pi_n^\lambda(J_0)\right| \exp\left[-\frac{t^2}{4\widehat{S}_{\max}} \right], \\ &\widehat{S}_{\max} := \sup\left\{ \frac{1}{n}\sum_{i=1}^n \left(\Gamma_i \left(2\pi(X_i) - 1\right) - A_n^*\right)^2 : \pi \in \Pi_n^\lambda(J_0) \right\}. \end{aligned} \tag{82}$$

Then, following (68), we get that

$$\begin{aligned} &1(\mathcal{B}_n)\mathbb{E}\left[\sqrt{n} \sup\left\{ \frac{1}{n}\sum_{i=1}^n \xi_i \Gamma_i \left(2\pi(X_i) - 1\right) : \pi \in \Pi_n^\lambda(J_0) \right\} \middle| \{X_i, \Gamma_i\} \right] \\ &\quad \leq 2\widehat{S}_{\max}^{0.5}\left(\sqrt{\log N_H\left(4^{-(J_0+1)}, \Pi_n\right)} + 1.5\right) + n^{-\frac{7+2\beta}{16}} \end{aligned}$$

$$\leq 2\widehat{S}_{\max}^{0.5}\left(\sqrt{5\log(4)\,\mathrm{VC}(\Pi_n)(J_0+1)}+1.5\right)+n^{-\frac{7+2\beta}{16}}$$
$$\leq 6\sqrt{\widehat{S}_{\max}\,\mathrm{VC}(\Pi_n)\left(10+\left\lfloor\log_4\left(S_n/S_n^\lambda\right)\right\rfloor\right)}+n^{-\frac{7+2\beta}{16}}. \tag{83}$$

Now, combining the bound we already have from Lemma 7 with the proof of Lemma 4, we see that under the conditions of Lemma 2 and provided that $S_n\,\mathrm{VC}(\Pi_n)/n \to 0$, we have that

$$\limsup_n \mathbb{E}[\sqrt{\widehat{S}_{\max}}]/\sqrt{S_n^\lambda+4\lambda^2}\leq 1;$$

to check this, we also used the fact that, by (24),

$$\sup\left\{\mathbb{E}\left[\left(2\left(\pi(X_i)-1\right)\Gamma_i-A_n^*\right)^2\right]:\pi\in\Pi_n^\lambda\right\}$$
$$=\sup\left\{\mathrm{Var}\left[2\left(\pi(X_i)-1\right)\Gamma_i\right]+\left(A_n(\pi)-A_n^*\right)^2:\pi\in\Pi_n^\lambda\right\}\leq S_n^\lambda+4\lambda^2. \tag{84}$$

Thus, we conclude that

$$\limsup_{n\to\infty}\mathbb{E}\left[1(\mathcal{B}_n)\sqrt{\frac{n}{\left(S_n^\lambda+4\lambda^2\right)\mathrm{VC}(\Pi_n)}}\sup\left\{\frac{1}{n}\sum_{i=1}^n\xi_i\Gamma_i\left(2\pi(X_i)-1\right):\pi\in\Pi_n^\lambda(J_0)\right\}\right]$$
$$\Big/\left(1+18\sqrt{1+\left\lfloor\log_4\left(S_n/S_n^\lambda\right)\right\rfloor/9}\right)\leq 1. \tag{85}$$

Meanwhile, for the second term, we proceed exactly as before up to (77). Here, however, we invoke the new (larger) choice of J_0 and, noting that

$$2^{-8}\left(44+38\sqrt{9+\left\lfloor\log_4\left(\frac{S_n}{S_n^\lambda}\right)\right\rfloor}\right)\leq\sqrt{1+\left\lfloor\log_4\left(\frac{S_n}{S_n^\lambda}\right)\right\rfloor/9},$$

we get

$$\limsup_{n\to\infty}\sqrt{\frac{n}{S_n^\lambda\,\mathrm{VC}(\Pi_n)}}\mathbb{E}\left[1(\mathcal{B}_n)\sup_{\pi\in\Pi_n^\lambda}\left|\frac{1}{n}\sum_{i=1}^n\Gamma_i\xi_i\sum_{j=J_0}^{J(n)-1}\left(\Psi_{j+1}(\pi)-\Psi_j(\pi)\right)(X_i)\right|\right]$$
$$\Big/\sqrt{1+\left\lfloor\log_4\left(\frac{S_n}{S_n^\lambda}\right)\right\rfloor/9}\leq 1. \tag{86}$$

Finally, we establish (26) by combining this bound with (85), and the fact that clipping as in (64) has an asymptotically negligible effect.

PROOF OF LEMMA 6: We construct the chaining by backwards recursion, as follows. First, for the largest index J under consideration, we do the following:

1. Let $\Psi_J':\Pi_n\to\{\mathcal{X}\to\{0,1\}\}$ be an optimal $2^{-(J+1)}$ covering of Π_n, such that the cardinality of the set $\{\Psi_J'(\pi):\pi\in\Pi_n\}$ is at most $N_{D_n}(2^{-(J+1)},\Pi_n,\{X_i,\Gamma_i\})$.
2. For every approximating policy $\pi'\in\{\Psi_J'(\pi):\pi\in\Pi_n\}$, construct a function neighbor$(\cdot)$ such that neighbor$(\pi')\in\{\pi\in\Pi_n^\lambda:D_n(\pi,\pi')\leq 2^{-(J+1)}\}$ if this set is nonempty, and neighbor$(\pi')=\emptyset$ else.
3. Define $\Psi_J:\Pi_n^\lambda\to\Pi_n^\lambda$ via $\Psi_J(\pi)=\text{neighbor}(\Psi_J'(\pi))$.

We can see by construction that $\Psi_J(\pi) \in \Pi_n^\lambda$ for all $\pi \in \Pi_n^\lambda$ (because no element in Π_n^λ can be mapped by Ψ_J to an element π' with neighbor$(\pi') = \emptyset$), and that the cardinality of the set $\Pi_n^\lambda(J) = \{\Psi_J(\pi) : \pi \in \Pi_n^\lambda\}$ is at most $N_{D_n}(2^{-(J+1)}, \Pi_n, \{X_i, \Gamma_i\})$. Furthermore, by the triangle inequality, $D_n(\Psi_J(\pi), \pi) \le 2^{-J}$ for all $\pi \in \Pi_n^\lambda$.

Next, for every $1 \le j < J$, we first define the mapping Ψ_j as a 2^{-j}-approximation of $\Pi_n^\lambda(j+1)$ using exactly the same construction as above. Thus, $\Psi_j : \Pi_n^\lambda(j+1) \to \Pi_n^\lambda(j+1)$, $\Pi_n^\lambda(j) = \{\Psi_j(\pi) : \pi \in \Pi_n^\lambda(j+1)\}$ has cardinality at most $N_{D_n}(2^{-(j+1)}, \Pi_n, \{X_i, \Gamma_i\})$, and $D_n(\Psi_j(\pi), \pi) \le 2^{-j}$ for all $\pi \in \Pi_n^\lambda(j+1)$. Finally, we extend the mappings Ψ_j to the whole domain Π_n^λ via the relationship $\Psi_j(\pi) = \Psi_j(\Psi_{j+1}(\pi))$ for all $\pi \in \Pi_n^\lambda$. Note that this extension does not grow the size of the set $\Pi_n^\lambda(j)$, and that the mapping Ψ_j has no branching by construction. *Q.E.D.*

C.2. *Proof of Corollary 3*

First, as argued by Bartlett and Mendelson (2002) in the proof of their Theorem 8,

$$\mathbb{E}\big[\sup\big\{\big|\widetilde{A}_n(\pi) - A_n(\pi)\big| : \pi \in \Pi_n^\lambda\big\}\big] \le 2\mathbb{E}\big[\mathcal{R}_n\big(\Pi_n^\lambda\big)\big], \tag{87}$$

Then, to check concentration, we need to bound $\sup_{\pi\in\Pi_n} |\widetilde{A}_n(\pi) - A_n(\pi)|$ in terms of its expectation. Recall that $\widetilde{A}_n(\pi) = n^{-1}\sum \Gamma_i(2\pi(X_i) - 1)$, and that the Γ_i are uniformly sub-Gaussian. Because the Γ_i are not bounded, it is convenient to define truncated statistics

$$\widetilde{A}_n^{(-)}(\pi) = \frac{1}{n}\sum_{i=1}^n \Gamma_i^{(-)}\big(2\pi(X_i) - 1\big),$$
$$\Gamma_i^{(-)} = \Gamma_i \mathbf{1}\big(\big\{|\Gamma_i| \le \log(n)\big\}\big).$$

Here, we of course have that $|\Gamma_i^{(-)}| \le \log(n)$, and so we can apply Talagrand's inequality as described in Bousquet (2002) to these truncated statistics. We see that, for any $\delta > 0$, with probability at least $1 - \delta$,

$$\begin{aligned}
&\sup_{\pi\in\Pi_n^\lambda} \big|\widetilde{A}_n^{(-)}(\pi) - A_n^{(-)}(\pi)\big| \\
&\quad\le \mathbb{E}\Big[\sup_{\pi\in\Pi_n^\lambda} \big|\widetilde{A}_n^{(-)}(\pi) - A_n^{(-)}(\pi)\big|\Big] + \frac{\log(n)\log(\delta)}{3n} \\
&\qquad + \sqrt{2\log\big(\delta^{-1}\big)\Big(\sup_{\pi\in\Pi_n^\lambda} \mathrm{Var}\big[\widetilde{A}_n(\pi)\big] + \frac{2\log(n)}{n}\mathbb{E}\Big[\sup_{\pi\in\Pi_n^\lambda} \big|\widetilde{A}_n^{(-)}(\pi) - A_n^{(-)}(\pi)\big|\Big]\Big)},
\end{aligned}$$

where we used the shorthand $A_n^{(-)}(\pi) = \mathbb{E}[\widetilde{A}_n^{(-)}(\pi)]$. Moreover, because the Γ_i are uniformly sub-Gaussian, we can immediately verify that

$$\mathbb{E}\Big[\Big|\sup_{\pi\in\Pi_n^\lambda} \big|\widetilde{A}_n^{(-)}(\pi) - A_n^{(-)}(\pi)\big| - \sup_{\pi\in\Pi_n} \big|\widetilde{A}_n(\pi) - A_n(\pi)\big|\Big|\Big]$$

decays exponentially fast in n; similarly, $n\sup_{\pi\in\Pi_n^\lambda} \mathrm{Var}[\widetilde{A}_n(\pi)] - S_n^\lambda$ also decays exponentially fast. Using (87) and noting that, by Lemma 2 and Assumption 3, $\mathbb{E}[\mathcal{R}_n(\Pi_n^\lambda)]$ decays

polynomially in n, we conclude that with probability at least $1-\delta$,

$$\begin{aligned}&\sup\bigl\{\bigl|\widetilde{A}_n(\pi)-A_n(\pi)\bigr|:\pi\in\Pi_n^\lambda\bigr\}\\&\quad\le\bigl(1+o(1)\bigr)\biggl(\mathbb{E}\bigl[\sup\bigl\{\bigl|\widetilde{A}_n(\pi)-A_n(\pi)\bigr|:\pi\in\Pi_n^\lambda\bigr\}\bigr]+\sqrt{\frac{2S_n^\lambda\log\bigl(\delta^{-1}\bigr)}{n}}\biggr),\end{aligned}\tag{88}$$

thus establishing our second claim.

C.3. *Proof of Lemma 4*

In the argument below, we omit all n-subscripts for readability, for example, we write $\widehat{A}(\pi)$ instead of $\widehat{A}_n(\pi)$. For any fixed policy π, we begin by expanding out the difference of interest as

$$\begin{aligned}&\widehat{A}(\pi)-\widetilde{A}(\pi)\\&\quad=\frac{1}{n}\sum_{i=1}^n\bigl(2\pi(X_i)-1\bigr)\bigl(Y_i-m(X_i,W_i)\bigr)\bigl(\hat{g}^{(-k(i))}(X_i,Z_i)-g(X_i,Z_i)\bigr)\\&\qquad+\frac{1}{n}\sum_{i=1}^n\bigl(2\pi(X_i)-1\bigr)\bigl(\tau_{\hat{m}^{(-k)}}(X_iW_i)-\tau_m(X_i,W_i)\\&\qquad-g(X_i,Z_i)\bigl(\hat{m}^{(-k(i))}(X_i,W_i)-m(X_i,W_i)\bigr)\bigr)\\&\qquad-\frac{1}{n}\sum_{i=1}^n\bigl(2\pi(X_i)-1\bigr)\bigl(\hat{m}^{(-k(i))}(X_i,W_i)-m(X_i,W_i)\bigr)\\&\qquad\times\bigl(\hat{g}^{(-k(i))}(X_i,Z_i)-g(X_i,Z_i)\bigr).\end{aligned}$$

Denote these three summands by $D_1(\pi)$, $D_2(\pi)$, and $D_3(\pi)$. We will bound all 3 summands separately.

To bound the first term, it is helpful separate out the contributions of the K different folds:

$$D_1^{(k)}(\pi)=\frac{1}{n}\sum_{\{i:k(i)=k\}}\bigl(2\pi(X_i)-1\bigr)\bigl(Y_i-m(X_i,W_i)\bigr)\bigl(\hat{g}^{(-k(i))}(X_i,Z_i)-g(X_i,Z_i)\bigr).\tag{89}$$

Now, because $\hat{g}^{(-k)}(\cdot)$ was only computed using data from the $K-1$ folds, we can condition on the value of this function estimate to make the individual terms in the above sum independent. Moreover, by exogeneity of the instrument and the exclusion restriction, we see that $\mathbb{E}[Y_i-m(X_i,W_i)|X_i,Z_i,\hat{g}^{(-k(i))}(\cdot)]=0$, and so the expected second moment of $D_1^{(k)}(\pi)$ reduces to the sum of the variances of its constituent terms.

Next, by Assumption 2, we know that

$$\sup_{x\in\mathcal{X}}\bigl|\bigl(\hat{g}^{(-k)}(x,z)-g(x,z)\bigr)\bigr|\le 1$$

with probability tending to 1, and so the individual summands in (89) are all ν-sub-Gaussian with probability tending to 1. Then, writing

$$V_n(k)=\mathbb{E}\bigl[\bigl(\hat{g}^{(-k)}(X_i,Z_i)-g(X_i,Z_i)\bigr)^2\operatorname{Var}\bigl[Y_i-m(X_i,W_i)|X_i,Z_i\bigr]|\hat{g}^{(-k)}(\cdot)\bigr]$$

for the variance of $D_1^{(k)}(\pi)$ conditionally on the model $\hat{g}^{(-k)}(\cdot)$ fit on the other $K-1$ folds, we can apply Corollary 3 to establish that

$$\frac{n}{n_k}\mathbb{E}\Big[\sup_{\pi\in\Pi}\big|D_1^{(k)}(\pi)\big|\,\big|\hat{g}^{(-k)}(\cdot)\Big] = \mathcal{O}\left(\sqrt{\mathrm{VC}(\Pi_n)\frac{V_n(k)}{n_k}}\right), \tag{90}$$

where $n_k = |\{i : k(i) = k\}|$ denotes the number of observations in the kth fold. Since we compute our doubly robust scores using a finite number of evenly-sized folds, $n_k/n \to 1/K$, we can use our risk bounds in Assumption 2 to check that

$$\begin{aligned}\mathbb{E}\big[V_n(k)\big] &\le \mathbb{E}\Big[\nu^2\mathbb{E}\Big[\big(\hat{g}^{(-k)}(X_i, Z_i) - g(X_i, Z_i)\big)^2\,\big|\hat{g}^{(-k)}(\cdot)\Big]\Big] \\ &= \mathcal{O}\left(a\left(\frac{K-1}{K}n\right)n^{-\zeta_g}\right).\end{aligned} \tag{91}$$

Then, applying (90) separately to all K folds and using Jensen's inequality, we find that

$$\mathbb{E}\Big[\sup_{\pi\in\Pi}\big|D_1(\pi)\big|\Big] = \mathcal{O}\left(\nu\sqrt{\mathrm{VC}(\Pi_n)\frac{a\big(\big(1-K^{-1}\big)n\big)}{n^{1+\zeta_g}}}\right), \tag{92}$$

thus bounding the first term.

Meanwhile, recall that by the properties of our weighting function (10), we know that $\mathbb{E}[\tau_{\tilde{m}}(X_i, W_i) - g(X_i, Z_i)\tilde{m}(X_i, W_i)|X_i] = 0$ for any conditional response function $\tilde{m}(\cdot)$, which in particular means that, by cross-fitting,

$$\begin{aligned}&\mathbb{E}\Big[\tau_{\hat{m}^{(-k)}}(X_i, W_i) - \tau_m(X_i, W_i) \\ &\quad - g(X_i, Z_i)\big(\hat{m}^{(-k(i))}(X_i, W_i) - m(X_i, W_i)\big)|X_i, \hat{m}^{-k(i)}(\cdot)\Big] = 0.\end{aligned}$$

Thus, by a similar argument as before, we find that

$$\mathbb{E}\Big[\sup_{\pi\in\Pi}\big|D_2(\pi)\big|\Big] = \mathcal{O}\left(\frac{1}{\eta}\sqrt{\mathrm{VC}(\Pi_n)\frac{a\big(\big(1-K^{-1}\big)n\big)}{n^{1+\zeta_m}}}\right), \tag{93}$$

where η is the uniform "overlap" bound on the weighting function $g(\cdot)$.

It now remains to bound the final term, $D_3(\pi)$. Here, we can use the Cauchy–Schwarz inequality to verify that

$$\begin{aligned}\big|D_3(\pi)\big| = \Bigg|\frac{1}{n}\sum_{i=1}^{n}\big(2\pi(X_i) - 1\big)\big(\hat{m}^{(-k(i))}(X_i, W_i) - m(X_i, W_i)\big) \\ \times\big(\hat{g}^{(-k(i))}(X_i, Z_i) - g(X_i, Z_i)\big)\Bigg|\end{aligned}$$

$$\leq \sqrt{\frac{1}{n}\sum_{i=1}^{n}\big(\hat{m}^{(-k(i))}(X_i, W_i) - m(X_i, W_i)\big)^2} \times \sqrt{\frac{1}{n}\sum_{i=1}^{n}\big(\hat{g}^{(-k(i))}(X_i, Z_i) - g(X_i, Z_i)\big)^2}.$$

This bound is deterministic and does not depend on π; thus, it also holds as a bound for the supremum of $|D_3(\pi)|$ over all π. Then, applying Cauchy–Schwarz again to the above product, we see that

$$\begin{aligned}\mathbb{E}\left[\frac{n\sup_{\pi\in\Pi}|D_3(\pi)|}{|\{i : W_i = 1\}|}\right] &\leq \sqrt{\mathbb{E}\big[\big(\hat{m}^{(-k(i))}(X_i, W_i) - m(X_i, W_i)\big)^2\big]} \\ &\quad\times \sqrt{\mathbb{E}\big[\big(\hat{g}^{(-k(i))}(X_i, Z_i) - g(X_i, Z_i)\big)^2\big]} \\ &\leq a\left(\left\lfloor\frac{K-1}{K}n\right\rfloor\right)\Big/\sqrt{\left\lfloor\frac{K-1}{K}n\right\rfloor},\end{aligned}$$

The desired conclusion now follows from combining these three bounds.

C.4. *Proof of Theorem 5*

Writing $\mathrm{VC}(\Pi) = d$, we know that there exists a collection of d nonoverlapping sets $\mathcal{A}_j$ for $j = 1, \ldots, d$ such that Π shatters this collection of sets, that is, for any vector $v \in \{0, 1\}^d$, there exist a policy $\pi_v \in \Pi$ such that $\pi_v(x) = v_j$ for all $x \in \mathcal{A}_j$. Our proof starts with such a collection of sets $\{\mathcal{A}_j\}_{j=1}^d$ and a distribution $\mathcal{P}$ over $\mathcal{X}_s$ such that

$$\mathbb{E}_{\mathcal{P}}\left[1\big(\{X_i \in \mathcal{A}_j\}\big)\frac{\sigma^2(X_i)}{e(X_i)\big(1 - e(X_i)\big)}\right] = \frac{S_{\mathcal{P}}}{d} \quad \text{for } j = 1, \ldots, d, \tag{94}$$

where $S_{\mathcal{P}}$ is as defined in (36). We will establish our result by studying learning over Π with features drawn from this distribution $\mathcal{P}$.

Now, to lower-bound the minimax risk for policy learning for unknown bounded treatment effect functions $\tau(\cdot)$, it is sufficient to bound minimax risk over a smaller class of policies T, as minimax risk increases with the complexity of the class T. Noting this fact, we restrict our analysis to treatment functions T such that

$$\tau(x) = \frac{\sigma^2(x)c_j}{e(x)\big(1 - e(x)\big)}\Big/\mathbb{E}\left[\frac{\sigma^2(x)1\big(\{X_i \in \mathcal{A}_j\}\big)}{e(X_i)\big(1 - e(X_i)\big)}\right]$$

for all $x \in \mathcal{A}_j$, where $c_j \in \mathbb{R}$ is an unknown coefficient for each $j = 1, \ldots, d$. If we knew the values of c_j for $j = 1, 2, \ldots, d$, the optimal policy $\pi^* \in \Pi$ would be treat only those j-groups with a positive c_j, that is, $\pi^*(x) = 1(\{c_j > 0\})$ for all $x \in \mathcal{A}_j$.

Now, following the argument of Hirano and Porter (2009) (we omit details for brevity), the minimax policy learner is of the form $\hat{\pi}^*(x) = 1(\{\hat{c}_j^* > 0\})$ for all $x \in \mathcal{A}_j$, where $\hat{c}_j^*$ is an efficient estimator for c_j. Moreover, in this example, we can use (94) to verify that the

semiparametric efficient variance for estimating c_j is $S_{\mathcal{P}}/d$. Thus, the efficient estimator $\hat{c}_j^*$ will incorrectly estimate the sign of c_j with probability tending to $\Phi(-c_j\sqrt{d/S_{\mathcal{P}}})$, where $\Phi(\cdot)$ denotes the standard Gaussian cumulative distribution function. (Recall that, in our sampling model (35), the signal also decays as $1/\sqrt{n}$.)

By construction, we suffer an expected utility loss of $2|c_j|$ from failing to accurately estimate the sign of c_j. Thus, by the above argument, given fixed values of c_j, the policy learner will suffer an asymptotic regret

$$\lim_{n\to\infty} \sqrt{n}\mathbb{E}[R_n] = \sum_{j=1}^{d} 2|c_j|\Phi\big(-|c_j|\sqrt{d/S_{\mathcal{P}}}\big),$$

using an efficient estimator $\hat{c}_j^*$. Setting $|c_j| = 0.75\sqrt{S_{\mathcal{P}}/d}$, this limit becomes

$$\lim_{n\to\infty} \sqrt{n}\mathbb{E}[R_n] = 1.5\Phi(-0.75)\sqrt{dS_{\mathcal{P}}},$$

which, noting that $1.5\Phi(-0.75) \geq 0.33$, concludes the proof.

REFERENCES

BARTLETT, P. L., AND S. MENDELSON (2002): "Rademacher and Gaussian Complexities: Risk Bounds and Structural Results," *Journal of Machine Learning Research*, 3, 463–482. [14]

BOUSQUET, O. (2002): "A Bennett Concentration Inequality and Its Application to Suprema of Empirical Processes," *Comptes Rendus Mathematique*, 334 (6), 495–500. [14]

CHERNOZHUKOV, V., J. C. ESCANCIANO, H. ICHIMURA, W. K. NEWEY, AND J. M. ROBINS (2016): "Locally Robust Semiparametric Estimation," arXiv preprint arXiv:1608.00033. [3]

CHERNOZHUKOV, V., W. NEWEY, AND J. ROBINS (2018): "Double/de-Biased Machine Learning Using Regularized Riesz Representers," arXiv preprint arXiv:1802.08667. [3,4]

DUDLEY, R. M. (1967): "The Sizes of Compact Subsets of Hilbert Space and Continuity of Gaussian Processes," *Journal of Functional Analysis*, 1 (3), 290–330. [4]

EFRON, B. (2011): "Tweedie's Formula and Selection Bias," *Journal of the American Statistical Association*, 106 (496), 1602–1614. [2]

EFRON, B., AND R. TIBSHIRANI (1996): "Using Specially Designed Exponential Families for Density Estimation," *The Annals of Statistics*, 24 (6), 2431–2461. [2]

FRIEDMAN, J., T. HASTIE, AND R. TIBSHIRANI (2010): "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, 33 (1), 1–22. [3]

GRAHAM, B. S., AND C. C. D. X. PINTO (2018): "Semiparametrically Efficient Estimation of the Average Linear Regression Function," Technical report, National Bureau of Economic Research. [3]

HAUSSLER, D. (1995): "Sphere Packing Numbers for Subsets of the Boolean n-Cube With Bounded Vapnik-Chervonenkis Dimension," *Journal of Combinatorial Theory, Series A*, 69 (2), 217–232. [1]

HIRANO, K., AND J. R. PORTER (2009): "Asymptotics for Statistical Treatment Rules," *Econometrica*, 77 (5), 1683–1701. [17]

HIRSHBERG, D. A., AND S. WAGER (2018): "Augmented Minimax Linear Estimation," arXiv preprint arXiv:1712.00038. [3,4]

LINDSEY, J. (1974): "Comparison of Probability Distributions," *Journal of the Royal Statistical Society: Series B (Methodological)*, 36 (1), 38–47. [2]

PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57 (5), 1027–1057. [1]

Co-editor Ulrich K. Müller handled this manuscript.

Manuscript received 2 October, 2017; final version accepted 3 September, 2020; available online 4 September, 2020.