GENERALIZED R-LEARNER FOR ESTIMATING POTENTIAL OUTCOME AND TREATMENT EFFECTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Estimating potential outcome and treatment effects from observation data is important for decision making. Orthogonal learners represent a growing area of research that develops target estimators insensitive to plug-in nuisance functions. Among existing orthogonal learners, the R-learner offers an approach that is free from numerical instability due to inverse propensities, yet it is derived based on partially linear outcome equations. We develop a *target-error* analysis for the R-learner to isolate an explicit *structural-drift* term that represents the gap between class-optimal plug-in target and the oracle target, revealing that the partially linear structure leads to a structural-drift *not reducible by the target estimator.* We accordingly present a generalized R-learner (GRL) that is able to accommodate this structural drift by utilizing a general nonlinear outcome equation in the target estimator, and provide an orthogonality regularized objective function that serves the dual purpose of reducing nuisance-induced structural drift in the target estimator while providing a nuisance-robust population risk bound. We establish conditions under which GRL achieves smaller target errors and when the target remains resolvable under nuisance estimation errors. On (semi-)synthetic and real datasets, we empirically demonstrated the strength of GRL in comparison to existing meta-learners in both potential outcome and treatment effect estimation.

1 INTRODUCTION

The ability to predict individual potential outcomes (PO) and treatment effects is important for answering *what-if* questions to inform treatment decisions (Hernán & Robins, 2020). Due to ethical and cost challenges with randomized controlled trials, learning treatment effects directly from observational data has received growing interests (Gianicolo et al., 2020; Bica et al., 2021). This setting presents a unique challenge where the treatment assignment mechanism is unknown and, if influenced by a patient's covariate, confounds the estimation of treatment effect on the outcome. *Model-agnostic meta-learners* (Künzel et al., 2019) represent a vibrant line of works in this space. It decomposes the target estimation into several supervised learning tasks where *nuisance* functions — such as propensity or outcome functions — are 1) first estimated by machine and deep learning (ML/DL), and 2) then used in a second-stage estimator to produce the *target* estimand. Depending on the second-stage target estimators, existing meta-learners fall into two broad categories.

Plug-in learners: This type of meta-learner focuses on approximating outcome functions and then directly plugging them into the defining expressions for the target estimand. Examples include the S-learner (a single outcome model (Johansson et al., 2016)), T-learner (one outcome model per treatment group (Künzel et al., 2019)), and a spectrum of designs in between with varying degrees of representation sharing (Shalit et al., 2017; Yao et al., 2018; 2019; Curth & Van der Schaar, 2021a; Chu et al., 2022) or disentanglement (Hassanpour & Greiner, 2019; Zhang et al., 2021; Curth & Van der Schaar, 2021b; Wu et al., 2022). Our work is readily complementary to these developments.

Target-oriented learners: In contrast, given the estimated nuisance functions, this type of learners construct a second-stage regression objective for the target estimand. Among existing approaches, a notable family shares the principle of *orthogonal learning:* the construction of *target estimators* that are locally insensitive to perturbations in *nuisance estimates*. This orthogonality permits the use of flexible ML/DL-based nuisance estimators, while ensuring their errors to have no first order effect on the *estimator* used to recover the target (Foster & Syrgkanis, 2023). Most existing orthogonal

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

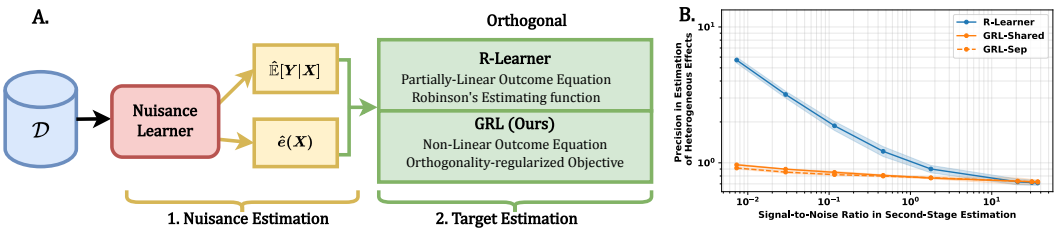


Figure 1: A: GRL *vs*. R-learner at a glance: GRL extends the partially-linear outcome to general nonlinear equations in the target estimator, with an orthogonality penalty to serve the dual purpose of reducing nuisance-induced structural drift while establishing a nuisance-robust risk bound. B: Empirical evidence for the improved robustness of GRL's CATE estimates to nuisance errors.

learners (Chernozhukov et al., 2018; Van der Laan et al., 2011; Ma et al., 2024), such as the DR-learner (Kennedy, 2023), build on the *augmented inverse probability weighting* (AIPW) estimating function that makes no structural assumption about the underlying data-generating process, although the calculation of inverse propensity can result in high variance and numerical instability. The R-learner (Nie & Wager, 2021) — built on Robinson's residual decomposition (Robinson, 1988) — offers a numerically-stable alternative that avoids inverse propensity, although derived based on the structural assumption of partially linear outcome equations.

Contribution I – Target-error analysis for the R-learner: Existing theory for the R-learner is mainly organized around the behavior of the *target estimator*: does a plug-in estimator using estimated nuisances behave nearly the same as its oracle counterpart using true nuisances (Robinson, 1988; Nie & Wager, 2021)? They however do not answer a closely-related yet different question: once nuisances are plugged in, how does the plug-in target shift away from the oracle target, and to what extent can this shift be addressed by the second-stage target estimation? A similar analysis was presented for the DR-learner, where the discrepancy induced by the estimated nuisances is characterized through product of nuisance errors, revealing its double-robustness property. Motivated by this, we perform a similar *target-error analysis* for the R-learner to isolate an explicit *structural-drift term* representing the gap between class-optimal plug-in target and oracle target. This reveals an effect largely invisible in *estimator-level* analyses: orthogonality may control nuisance sensitivity at the estimator level, yet the actual target error depends critically on whether the second-stage model class can absorb nuisance-induced structural distortions; more importantly, a partially-linear structure is inadequate for this purpose, leading to structural-drift *not reducible by the target estimator*.

Contribution II – A generalized R-learner: To mitigate this gap, we propose a generalized R-learner (GRL, Fig. 1A) to accommodate nuisance-induced structural drift in the plug-in class-optimal target by extending the partially-linear outcome to a general nonlinear outcome equation. We further propose an orthogonality regularized objective for this non-linear target to simultaneously reduce structural drift and retain a nuisance-robust population risk bound. We establish the conditions for GRL to achieve smaller target error than standard R-learner in both PO and CATE estimation, as well as the condition under which the second-stage target remains *resolvable* in the presence of first-stage nuisance estimation errors. We experimentally tested the resolvability conditions for GRL and R-learner, verifying the significantly improved robustness of GRL to nuisance estimation errors (Fig. 1B). We further compared GRL with a comprehensive suite of existing meta-learners on synthetic, semi-synthetic, and real datasets: while not a single approach stood out as the clear winner across all datasets, GRL was the only approach that consistently remained either top-performing or statistically comparable to top-performing models for both PO and CATE estimation.

2 Related Works

Plug-in learners & representation learning: In these approaches, the target (*e.g.*, average or conditional average treatment effect) is directly calculated via their defining expressions, using estimated nuisance outcome and/or propensity functions. At the two ends of the spectrum, we have S-learner that learns a single outcome model with T as an additional input (Johansson et al., 2016), and T-learner that learns a separate outcome model per treatment group (Künzel et al., 2019). With the advent of DL, a variety of multi-head architectures have emerged on this spectrum to model differ-

ent degrees of representation-sharing across outcome and propensity heads (Shalit et al., 2017; Shi et al., 2019), different degrees of disentanglement to control which features should be shared *vs*. private to each output (Hassanpour & Greiner, 2019; Curth & Van der Schaar, 2021a;b; Zhang et al., 2021), and invariant representations to align learned representations across treatment groups (Guo et al., 2021; Johansson et al., 2016; Shalit et al., 2017; Guo et al., 2023; Yao et al., 2018; 2019).

GRL is orthogonal to these developments. Indeed, various architectures may be leveraged as the nuisance learner in GRL. To demonstrate this, we instantiate GRL's nuisance learner with two simple variants: with and without representation sharing between propensity and outcome heads.

Target-oriented learners & orthogonal objectives: Alternatively, a second-stage learner for the target can be constructed using learned nuisance functions. One popular family involves the construction of *pseudo-outcomes*, represented by the PW-learner (Curth & Van der Schaar, 2021b), RA-learner (Curth & Van der Schaar, 2021b), and X-learner (Künzel et al., 2019). Another distinct family shares the goal of Neyman orthogonality, such that the target estimator is locally insensitive to nuisance estimation errors relative to its oracle version. Existing orthogonal learners are built on two types of estimating functions. One builds on the AIPW estimating function, such as double machine learning (Chernozhukov et al., 2018), targeted maximum likelihood estimation (Van der Laan et al., 2011), doubly robust (DR) learner (Kennedy, 2023), RieszNet (Chernozhukov et al., 2022), and DragonNet (Shi et al., 2019). These approaches do not impose structural assumptions on the outcome equations, although the sensitivity of inverse propensity to extreme propensities needs to be mitigated in practice by techniques such as propensity clipping and sample trimming (Crump et al., 2009), or automatic debiasing (Chernozhukov et al., 2022). Alternatively, Robinson's residual-on-residual transformation (Robinson, 1988) and its associated R-learner (Nie & Wager, 2021) are free from this sensitivity, yet they are derived based on a partially-linear outcome equation.

GRL examines the open question of nuisance-induced target distortion in the R-learner, reveals that the partially-linear structure leads to a *structural-drift* not resolvable by the target estimator, and present a solution to reduce this structural-drift while retaining robustness to nuisance estimates.

3 Preliminaries: Orthogonal PO & Treatment Effect Estimators

Consider covariate variable $X \in \mathbb{R}^d$, binary treatment variable $T \in \{0, 1\}$, and outcome variable $Y \in \mathbb{R}$. The target quantities of interest in the literature include: (i) ATE, $\tau = \mathbb{E}[Y^{(1)} - Y^{(0)}]$, as a low-dimensional parameter, (ii) CATE, $\tau(x) = \mathbb{E}[Y^{(1)} - Y^{(0)}|X = x]$, as a response surface, and (iii) potential outcome, $\mathbb{E}[Y|X = x, T = t]$, as a function that can be used to calculate CATE/ATE. We operate under the common assumptions of *SUTVA and consistency*, *ignorability,* and *positivity* for treatment-effect identifiability (Shalit et al., 2017; Yoon et al., 2018) elaborated in Appendix A. In Appendix B, we present a unified review of existing meta-learners organized under the principle of Neyman orthogonality, with those most relevant to GRL presented below.

Robinson's residual-on-residual transformation: Robinson's seminal *residual-on-residual* transformation (Robinson, 1988) is derived based on a partially-linear data generation process:

$$Y = g(X) + \tau * T + \xi_y \quad \mathbb{E}[\xi_y|X, T] = 0; \qquad T = e(X) + \xi_t, \quad \mathbb{E}[\xi_t|X] = 0 \tag{1}$$

where τ represents the ATE. By defining $\mathbb{E}[Y|X] = g(X) + \tau * e(X)$, we can re-write Equation 1:

$$Y - \mathbb{E}[Y|X] = \tau * (T - e(X)) + \xi_y \tag{2}$$

This transformation gives the foundation of Robinson's two-stage estimation process: $\eta = \{\mathbb{E}[Y|X], e\}$ are the *nuisance* parameters to be first estimated, and τ as the *target* is regressed from *residuals* of estimated $\{\hat{\mathbb{E}}[Y|X], \hat{e}\}$ via a *residual-on-residual* fit: $\hat{\tau} = \arg\min_\tau \frac{1}{n} \sum_i ||(Y_i - \hat{\mathbb{E}}[Y_i|X_i]) - \tau(T_i - \hat{e}(X_i))||_2^2$, yielding an estimation function Neyman orthogonal to $\{\mathbb{E}[Y|X], e\}$:

$$\psi_{Rob}(W; \tau, \eta) = [Y - \mathbb{E}[Y|X] - \tau * (T - e(X))](T - e(X)) \tag{3}$$

Note that, with binary T, any nonlinear outcome functions can be rearranged into Equation 1 with $g(X) = Y(X, 0)$ and $\tau = Y(X, 1) - Y(X, 0)$. The limitation of the partially-linear structure thus does not stem from its generality in representing oracle data-generation processes in Equation 1, but its use for deriving the second-stage estimator which we will elaborate in Section 4.

Quasi-oracle error bound: This residual-on-residual optimization as defined in Equation 3 is extended to estimate the response surface of CATE, $\tau(X)$, in the R-learner (Nie & Wager, 2021). To show that the estimated $\tau(X)$ is insensitive to the error introduced by estimated nuisance functions, Nie & Wager (2021) formalize a *quasi-oracle error bound*, defined through excess risk:

$$R(\tau^*(X); \eta_0) = L(\tau^*(X; \eta_0)) - L(\tau_0(X; \eta_0) \tag{4}$$

where L is the R-loss based on Equation 3, $\tau^*(X)$ is the optimal solution by minimizing R-loss under estimated nuisances $\hat{\eta}$, and $\tau_0(X)$ the oracle. Both terms are evaluated under oracle nuisances, η_0. They show that the feasible target, $\tau^*(X)$, attains the same risk bound as the target optimized with the true nuisance, where nuisance estimation errors only enter in second-order terms.

4 Target-Error and Structural-Drift Analyses

We begin by defining oracle nuisance functions $\eta_0 := (l_0(X) = \mathbb{E}[Y|X], e_0(X) := \mathbb{E}[T|X])$. With Robinson's transformation in Equation 2, we define the *oracle target outcome* function as:

$$Y(T, X) = \underline{l_0(X) + (T - e_0(X))\tau_0(X)} + \xi_y := f_0(X, T; \eta_0) \tag{5}$$

where $Y - l_0(X)$ is the *oracle outcome residual* to be resolved in the second-stage by the *oracle treatment residual* $T - e_0(X)$, with τ_0 being the oracle CATE function. Now let $\hat{\eta} = (\hat{l}(X), \hat{e}(X))$ denote estimated nuisance pair, and $\Delta l(X) = \hat{l}(X) - l_0(X)$ and $\Delta e(X) = \hat{e}(X) - e_0(X)$ denote their respective estimation errors. Under plug-in nuisances ($\hat{\eta}$), Equation 5 can be re-written as:

$$Y(T, X) \quad = \quad \underline{\hat{l}(X) + (T - \hat{e}(X)) * \tau_0(X)} + \Delta e(X)\tau_0(X) - \Delta l(X) + \xi_y \tag{6}$$

which gives the optimal *plug-in target outcome*. Contrasting the oracle (Equation 5) and plug-in target (Equation 6), two points become clear. First, with plug-in nuisances $\{\hat{l}(x), \hat{e}(x)\}$, the presence of nuisance estimation errors $\{\Delta e(X), \Delta l(X)\}$ $\{Y - l_0(X), T - e_0(X)\}$ induces a structural distortion to the optimal plug-in target outcome (Equation 6) in comparison to the oracle (Equation 5). Second, the use of partially-linear outcome equation, as the underlined portion in Equation 6, is not adequate to absorb this nuisance-induced distortion under plug-in nuisances.

Definition 1 (Target error)**.** *Let $f_m(\hat{\eta}) \in \mathcal{F}_m(\hat{\eta})$ where $\mathcal{F}_m(\hat{\eta})$ denote the second-stage model class with plug-in nuisances $\hat{\eta}$. Let population squared loss $\mathcal{L}(f) := \mathbb{E}[(Y - f(X, T; \eta))^2]$, under which the target error R_m of $f_m(\hat{\eta})$ relative to oracle target $f_0(\eta_0)$ is defined as:*

$$R_m := \mathcal{L}(f_m(\hat{\eta})) - \mathcal{L}(f_0(\eta_0)) = \|f_m(\hat{\eta}) - f_0(\eta_0)\|_2^2 \tag{7}$$

The proof of $R_m = \|f_m(\hat{\eta}) - f_0(\eta_0)\|_2^2$ is provided in Appendix C.1. Importantly, $f_0(\eta_0)$ (partially-linear as in Equation 5) and $f_m(\hat{\eta}) \in \mathcal{F}_m(\hat{\eta})$ do not need to be in the same model class.

Definition 2 (Structural drift)**.** *Given a model class $\mathcal{F}_m$, define class-specific population-optimal target $f_m^*(\hat{\eta}) \in \arg\min_{f \in \mathcal{F}_m(\hat{\eta})} \mathcal{L}(f)$. Define the function-level gap between the class-optimal plug-in and oracle target as: $d_m := f_m^*(\hat{\eta}) - f_0(\eta_0)$, and the risk-level structural drift S_m as:*

$$S_m := \mathcal{L}\big(f_m^*(\hat{\eta})\big) - \mathcal{L}(f_0(\eta_0)) = ||d_m||_2^2 \tag{8}$$

with $R_m = E_m + S_m$ and $E_m = \mathcal{L}\big(f_m(\hat{\eta})\big) - \mathcal{L}\big(f_m^*(\hat{\eta})\big)$ is the estimation error within $\mathcal{F}_m(\hat{\eta})$.

We now consider two second-stage model classes under plug-in nuisances:

$$\mathcal{F}_m(\hat{\eta}) = \begin{cases} \mathcal{F}_{\mathrm{PL}}(\hat{\eta}) = \{f(X, T) = \hat{l}(X) + (T - \hat{e}(X))\tau(X) : \tau \in \mathcal{T}\}, \\ \mathcal{F}_{\mathrm{NL}}(\hat{\eta}) = \{f(X, T) = h(X, \hat{l}(X), T - \hat{e}(X)) : h \in \mathcal{H}\} \end{cases} \tag{9}$$

where $\mathcal{F}_{\mathrm{PL}}$ represents the partially-linear model class used by R-learner, with $\tau \in \mathcal{T} : \mathcal{X} \to \mathbb{R}$, and $\mathcal{F}_{\mathrm{NL}}$ represents the nonlinear model class with $\mathcal{H}$ being the class of functions $h : \mathcal{X} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$.

Proposition 3 (Partially-linear structural drift)**.** Suppose that the partially-linear class is oracle-correct such that $\tau_{\mathrm{PL}}^*(\eta_0) = \tau_0$, then its structural drift is given by

$$S_{\mathrm{PL}} = \mathbb{E}\Big[\big(\Delta l(X) - \Delta e(X)\tau_0(X)\big)^2\Big] + O\big(\|\delta\eta\|^3\big). \tag{10}$$

Proposition 4 (Non-linear structural drift)**.** Suppose that the nonlinear class is oracle-correct such that $h^*(\eta_0) = f_0(\eta_0)$, its structural drift is given by

$$S_{\mathrm{NL}} = \|\nabla_\eta h^*(\eta_0)[\delta\eta]\|_2^2 + O\big(\|\delta\eta\|^3\big) \tag{11}$$

Proof. Refer to Appendix C.3 and C.4 for proofs of Proposition 3 and 4, respectively. □

With the partially-linear model class in Equation 10, the class-optimal plug-in target is associated with a quadratic structural drift induced by nuisance estimation errors: importantly, this term *is not reducible by the second-stage estimator*. In contrast, with the nonlinear model class, this structural drift depends on the sensitivity of $h(\eta)$ to η (Equation 11) and thus *is reducible in the second stage.*

4.1 Target error analysis for potential outcome

In general, comparing R_{PL} and R_{NL} using Propositions 3-4, we can establish that the nonlinear model class should be favored whenever there is a higher reduction of structural drift compared to the difference of estimation errors within each respective model class, *i.e.,*

$$\|\nabla_\eta h^*(\eta_0)[\delta\eta]\|_2^2 < \mathbb{E}\Big[\big(\Delta l(X) - \Delta e(X)\tau_0(X)\big)^2\Big] + (E_{\mathrm{PL}} - E_{\mathrm{NL}}) \tag{12}$$

Below we characterize the theoretical bounds for both error components.

For the structural drift, let $\|\Delta l\|_2 = O(r_l)$ and $\|\Delta e\|_2 = O(r_e)$, and assume $\|\tau_0\|_\infty \le M$. For partially linear class, structural drift is governed by $S_{\mathrm{PL}} = \|\Delta l - \tau_0 \Delta e\|_2^2$ and hence $S_{\mathrm{PL}} = O(r_l^2 + M^2 r_e^2)$. Under a mild non-cancellation condition, $\|\Delta l - \tau_0 \Delta e\|_2^2 \asymp r_l^2 + M^2 r_e^2$. More generally, if $r_l = O(n^{-\alpha_l})$ and $r_e = O(n^{-\alpha_e})$, then $S_{\mathrm{PL}} = O(n^{-2\alpha_l} + M^2 n^{-2\alpha_e})$ which is governed by $O(n^{-2\min(\alpha_l, \alpha_e)})$ when M is fixed. For non-linear class, if $\|\nabla_\eta h^*(\eta_0)\|_2 \le \kappa$, then the leading nonlinear structural-drift term satisfies $S_{\mathrm{NL}} \lesssim \kappa^2 \|\delta\eta\|_2^2$. Thus, S_{PL} is bottlenecked by the slowest nuisance rate, while S_{NL} is bounded by the controllable gradient of target function.

For the estimation error, let us consider MLP class such that estimation errors can be characterized by Rademacher complexity. Under bounded outcomes and predictions, standard generalization bounds give $E_m = O_p(\mathfrak{R}_n(\mathcal{F}_m))$ (Bartlett & Mendelson, 2002). Under matched depth and common layer-wise spectral- and $(2,1)$-norm constraints, norm-based neural-network bounds (Bartlett et al., 2017) give, up to logarithmic factors, $E_{\mathrm{PL}} = \widetilde{O}_p\left(RB_X/\sqrt{n}\right)$ and $E_{\mathrm{NL}} = \widetilde{O}_p\left((R\sqrt{B_X^2 + B_l^2 + 1})/\sqrt{n}\right)$ where $\|X\|_2 \le B_X$, $|\hat{l}(X)| \le B_l$, and R denotes a common upper bound on the norm-based network complexity. In other words, under matched complexity, both classes retain the same $n^{-1/2}$ estimation-error order, with the nonlinear bound increasing due to the augmented input $(X, \hat{l}(X), T - \hat{e}(X))$. For normalized inputs $|X_j| \le 1$ for $j = 1, \ldots, d$ and $|\hat{l}(X)| \le 1$, we have $\|X\|_2 \le \sqrt{d}$ and $\|(X, \hat{l}(X), T - \hat{e}(X))\|_2 \le \sqrt{d+2}$ and the two bounds scale as $\widetilde{O}_p(R\sqrt{d/n})$ and $\widetilde{O}_p(R\sqrt{(d+2)/n})$. Hence, with MLPs, the nonlinear model class reduces structural drift without changing the order of estimation complexity. Refer to Appendix C.8 for more details.

4.2 Target error analysis for CATE

Since the R-learner targets CATE directly at the second stage, below we present a counterpart target error analysis for CATE. For any potential outcome function f, define the treatment-contrast operator $(\mathcal{D}f)(x) := f(x,1) - f(x,0)$, and let $\tau_0 = \mathcal{D}f_0$, $\tau_m^*(\hat{\eta}) = \mathcal{D}f_m^*(\hat{\eta})$, and $\hat{\tau}_m = \mathcal{D}f_m(\hat{\eta})$.

Proposition 5 (CATE target error decomposition)**.** Let $S_m^\tau = \|\tau_m^*(\hat{\eta}) - \tau_0\|_2^2$, $E_m^\tau = \|\hat{\tau}_m - \tau_m^*(\hat{\eta})\|_2^2$, and $C_m^\tau = \langle \hat{\tau}_m - \tau_m^*(\hat{\eta}), \tau_m^*(\hat{\eta}) - \tau_0 \rangle$. The squared precision in estimation of heterogeneous effect (PEHE) admits decomposition $\mathrm{PEHE}_m^2 = \|\hat{\tau}_m - \tau_0\|_2^2 = S_m^\tau + E_m^\tau + 2C_m^\tau$. Because $|C_m^\tau| \le \sqrt{S_m^\tau E_m^\tau}$, we have $\left(\sqrt{S_m^\tau} - \sqrt{E_m^\tau}\right)^2 \le \mathrm{PEHE}_m^2 \le \left(\sqrt{S_m^\tau} + \sqrt{E_m^\tau}\right)^2$.

Proof. Refer to Appendix C.5. □

From Proposition 5, GRL has smaller PEHE when $S_{\mathrm{NL}}^\tau + E_{\mathrm{NL}}^\tau + 2C_{\mathrm{NL}}^\tau < S_{\mathrm{PL}}^\tau + E_{\mathrm{PL}}^\tau + 2C_{\mathrm{PL}}^\tau$. If the CATE-level estimation errors are orthogonal to their structural drifts, i.e., $C_{\mathrm{NL}}^\tau = C_{\mathrm{PL}}^\tau = 0$, this reduces to $S_{\mathrm{NL}}^\tau + E_{\mathrm{NL}}^\tau < S_{\mathrm{PL}}^\tau + E_{\mathrm{PL}}^\tau$. This is the CATE-level counterpart of Equation 12.

Under the Positivity assumption (Appendix A), PEHE is also bounded directly by the outcome target error R_m defined in Equation 7 and the rates established for S_m and E_m in Section 4.1 carry over.

Proposition 6 (PEHE bound)**.** If $\alpha \leq e_0(X) \leq 1-\alpha$ almost surely for some $\alpha > 0$, then $\text{PEHE}_m^2 \leq \frac{2}{\alpha}\|f_m(\hat{\eta}) - f_0(\eta_0)\|_2^2 = \frac{2}{\alpha}R_m = \frac{2}{\alpha}(S_m + E_m)$. *Proof.* Refer to Appendix C.6. □

5 GENERALIZED R-LEARNER (GRL)

We now introduce GRL with the nonlinear function class to reduce the structural drift term in Equation 11, while attempting to retain robustness to nuisances in the resulting nonlinear target estimator.

GRL as a two-stage meta-learner: Following the Robinson's decomposition and the standard practice of "cross-fitting", on a dataset $\mathcal{D}$ split into $\mathcal{D}_1$ and $\mathcal{D}_2$, we first learn nuisance functions $l(X) = \mathbb{E}[Y|X]$ and $e(X) = \mathbb{E}[T|X]$ on $\mathcal{D}_1$, and then learn $\mathbb{E}[Y|X,T] = h(X, l(X), T - e(X))$ on $\mathcal{D}_2$ with plug-in nuisances $\hat{l}(X)$ and $\hat{e}(X)$. To demonstrate that GRL is agnostic to the model choices for $l(X)$ and $e(X)$, we instantiate it in two architectural choices as illustrated in Fig.3: 1) GRL-Sep, where $l(X)$ and $e(X)$ are simply modeled and optimized separately (Fig.3a); and 2) GRL-Shared, where a latent representation $\phi = \Phi(X)$ is extracted from X to produce two outputs $l(\phi)$ and $e(\phi)$ (Fig.3b). For the latter, due to the unknown degree of confounding between X and T, optimization of $l(\phi)$ *vs.* $e(\phi)$ may introduce unexpected degrees of conflicts. To be robust to this, we first optimize $\Phi(X)$ via a reconstruction objective, using sum of dimension-wise mean squared error (MSE) for continuous covariates and binary cross entropy (BCE) for binary covariates. $e(\phi)$ head is optimized next with $\Phi(X)$ frozen. Finally, $l(\phi)$ is optimized with a MSE-loss, with each sample weighted by $\exp(-|T - \hat{e}(X)|/\gamma)$ such that propensity supported samples receive higher weights. Hyperparameter γ is between 0 and 1 and controls how quickly weight decays with $T - \hat{e}(X)$.

In the second-stage, to facilitate the residual-on-residual learning, we structure $h(X, l(X), T - e(X))$ to act as an identity function for $\hat{l}(X)$ via a skip connection. Unfortunately, directly fitting this nonlinear target induces an estimating function that no longer meets the property of Neyman orthogonality as described in Appendix B.1 and thus no longer robust to nuisance errors.

Lemma 1. (Nuisance Sensitivity). Consider raw loss, $\mathcal{L} = \mathbb{E}_n\left[(Y - h(X, l(X), T - e(X))\right]^2$ and its estimating function $\psi_{raw} = Y - h(X, l(X), T - e(X))$. Let $l_t = l_0 + t \cdot s$ and $e_t = e_0 + t \cdot r$ represent perturbations along arbitrary error directions $s(X)$ and $r(X)$. The pathwise derivatives of ψ_{raw} with respect to $l(X)$ and $e(X)$ do not vanish at $t = 0$ (*i.e.,* $l_0(X)$ and $e_0(X)$):

$$\partial_t \mathbb{E}[\psi(l_t)]\big|_{t=0} = \mathbb{E}\left[-\nabla_l h(\eta_0) \cdot s\right], \quad \partial_t \mathbb{E}[\psi(e_t)]\big|_{t=0} = \mathbb{E}\left[-\nabla_e h(\eta_0) \cdot r\right] \tag{13}$$

Coincidentally, the squared L2-norm of Equation 13 is equivalent to the quadratic structure-drift term for the nonlinear model class as defined in S_{NL} (Equation 11).

Orthogonality-regularized objective: To control for structural drift and re-enforce robustness to nuisance estimates, we introduce a regularized objective $\mathcal{L}_{\text{GRL-OP}}$:

$$\mathcal{L}_{\text{GRL-OP}} = \underbrace{\mathbb{E}_n\left[(Y - h(X, l(X), T - e(X)))^2\right]}_{\text{Raw Loss}} + \underbrace{\mathbb{E}_n\left[\|\nabla_l h\|^2 (Y - l(X))^2 + \|\nabla_e h\|^2 (T - e(X))^2\right]}_{\text{Orthogonality Penalty}} \tag{14}$$

We now establish the risk bound similar to Equation 4 in the R-learner but at the population level.

Proposition 7 (Nuisance-robust population risk bound)**.** Let $\Omega(h; \hat{\eta})$ represent the penalty terms of $\mathcal{L}_{\text{GRL-OP}}$ in Equation 14, so that $\mathcal{L}_{\text{GRL-OP}}(h; \hat{\eta}) = \mathcal{L}(h(\hat{\eta})) + \Omega(h; \hat{\eta})$, $h_0(\eta_0) \in \arg\min_{h \in \mathcal{H}} \mathcal{L}(h(\eta_0))$ and $h^*(\hat{\eta}) \in \arg\min_{h \in \mathcal{H}} \mathcal{L}_{\text{GRL-OP}}(h; \hat{\eta})$. Assume $h \in \mathcal{H}$ is twice continuously differentiable in its nuisance inputs with bounded derivatives, $\mathcal{H}$ and f_0 are uniformly bounded, and $\sigma_{\min}^2 \leq \text{Var}(Y \mid X, T) \leq \sigma_{\max}^2$ for some $0 < \sigma_{\min}^2 \leq \sigma_{\max}^2 < \infty$. Then there are constants $\kappa, C < \infty$, depending only on these bounds and positivity constant α, such that

$$\mathcal{L}(h^*(\eta_0)) - \mathcal{L}(h_0(\eta_0)) \;\leq\; \Omega(h_0; \hat{\eta}) + \sqrt{\kappa\, \Omega(h_0; \hat{\eta})}\, \|\delta\eta\| + C\, \|\delta\eta\|^2.$$

which approximates a second-order excessive risk via the minimization of $\Omega(h; \hat{\eta})$.

Proof. Refer to Appendix C.7. □

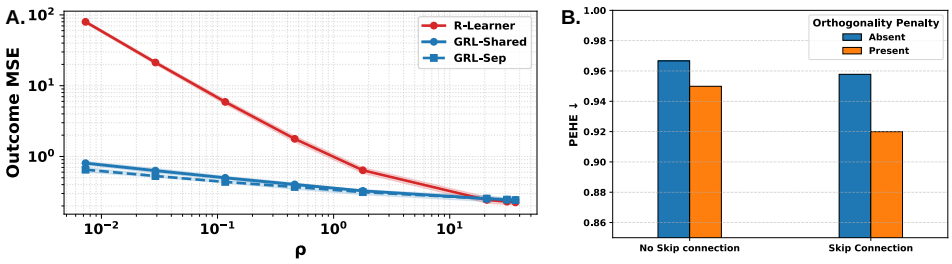


Figure 2: A. PO MSE of GRL *vs*. R-learner as the resolvability ratio ρ increased (PEHE in Fig. 1B). B. Ablation of the effect of orthogonal penalty and skip connection in the target estimator.

Proposition 7 shows that the population excess risk of h^* with respect to $\mathcal{L}$, *i.e.*, the gap between the optimal model h^* and the oracle model h_0, when evaluated at the true nuisances η_0, consists of second-order term $C\|\delta\eta\|^2$, and first order term $\sqrt{\kappa\,\Omega(h_0;\hat{\eta})}\,\|\delta\eta\|$, along with bias $\Omega(h_0;\hat{\eta})$. By minimizing $\Omega(h;\hat{\eta})$, the orthogonality penalty controls both the bias and first order effect, allowing the nuisance error to enter only through the second- or higher-order terms.

Therefore, the orthogonality penalty serves the dual role in reducing both the orthogonality gap identified in Lemma 1 and the quadratic structural-drift term in proposition 4.

Optimization: Because the raw loss in $\mathcal{L}_{\text{GRL-OP}}$ is much more difficult to optimize than the orthogonality penalty, instead of using a hyperparameter to combine the two objectives, we iteratively optimize them with two separate optimizers. We use a comparatively lower learning rate for the optimizer of the orthogonality penalty, to avoid it dominating the optimization with trivial solutions. $\nabla_l h$ and $\nabla_e h$ are approximated by PyTorch's autograd engine.

6 Experiments

Data: PO evaluation is challenging in real data that observes only factual outcomes. Following existing works, we adopted a combination of synthetic (ACIC2016), semi-synthetic (IHDP & Twins), and real-world (Jobs) datasets for evaluation. Details of each dataset are included in Appendix D.

Baselines: We evaluated GRL against a comprehensive suite of meta-learners. For plug-in learners, we included T-learner (Künzel et al., 2019), TARNet (Shalit et al., 2017), CFR-MMD (Shalit et al., 2017), SNet (Curth & Van der Schaar, 2021b), FlexTENet (Curth & Van der Schaar, 2021a), DiffPO (Ma et al., 2024), TEDVAE (Zhang et al., 2021) and MitNet (Guo et al., 2023). For target-oriented learners, we included *non-orthogonal* PW-learner (Curth & Van der Schaar, 2021b), RA-learner (Curth & Van der Schaar, 2021b), and X-learner (Künzel et al., 2019), as well as *orthogonal* DragonNet (Shi et al., 2019), R-learner (Nie & Wager, 2021), RieszNet (Chernozhukov et al., 2022), DR-learner (Kennedy, 2023) and its variant with propensity clipping (Crump et al., 2009) All meta-learners use identical MLP backbone and building blocks whenever applicable, with early-stopping strategy based on each model's validation loss, as detailed in Appendix E.

Metrics: For synthetic and semi-synthetic dataset,we considered 1) PEHE defined as root mean squared error of estimated CATE, and 2) MSE of estimated POs, separated by propensity-supported PO ($Y(x,0)$ for samples with $e_0(x) < 0.5$ and $Y(x,1)$ for samples $e_0(x) > 0.5$) and non-supported PO (the rest). For real-world Jobs dataset, we evaluated the *policy risk*, $\hat{R}_{\text{Pol}} = 1 - (\mathbb{E}[Y_1|\pi_f(x) = 1, T = 1] \cdot p(\pi_f = 1) + \mathbb{E}[Y_0|\pi_f(x) = 0, T = 0] \cdot p(\pi_f = 0))$ on a subset of randomized trial data following (Shalit et al., 2017), where the treatment policy $\pi_f(x) = 1$ if $\hat{\tau}(x) > 0$ and 0 otherwise.

6.1 Resolvability & Ablations

Resolvability test: The residual-on-residual fitting in stage 2 aims to address the oracle outcome and treatment residual from stage 1 that are *contaminated* by the nuisance estimation errors. We therefore quantify the *resolvability* of the oracle outcome residual by its *signal-to-noise* ratio to nuisance-induced distortions as: $\rho := (\mathbb{E}[\text{Var}((T - e_0(X))\tau_0(X)|X)])/(\mathbb{E}[(\Delta l(X) - \Delta e(X)\tau_0(X))^2])$. Note that ρ is inversely proportional to the structural drift S_{PL} defined in Equation 10.

Table 1: PEHE and MSE for propensity supported (SP) and non-supported (NSP) PO on synthetic data, the latter only calculated on baselines that provide PO as a target output. Best performance is in bold and second-best underlined; $*$ indicates statistical significance over the next best performance.

Model	$\lambda = 0.2$			$\lambda = 0.4$			$\lambda = 0.8$		
	PEHE ↓	MSE_{SP} ↓	MSE_{NSP} ↓	PEHE ↓	MSE_{SP} ↓	MSE_{NSP} ↓	PEHE ↓	MSE_{SP} ↓	MSE_{NSP} ↓
T-Learner	1.412 ± 0.005	3.306 ± 0.019	2.829 ± 0.027	1.412 ± 0.006	3.324 ± 0.024	2.831 ± 0.008	1.437 ± 0.027	3.304 ± 0.040	2.877 ± 0.014
TARNet	1.499 ± 0.460	0.474 ± 0.102	1.599 ± 1.020	2.410 ± 0.237	0.455 ± 0.085	4.062 ± 0.818	2.810 ± 0.220	0.379 ± 0.046	5.541 ± 0.840
CFR-MMD	1.235 ± 0.074	0.777 ± 0.091	0.883 ± 0.087	1.636 ± 0.261	1.082 ± 0.204	1.657 ± 0.559	1.813 ± 0.252	1.119 ± 0.081	2.098 ± 0.647
SNet	0.891 ± 0.036	0.361 ± 0.028	0.471 ± 0.033	1.643 ± 0.277	**0.335 ± 0.014**	1.874 ± 0.658	2.640 ± 0.029	**0.326 ± 0.037**	4.916 ± 0.096
FlexTENet	1.070 ± 0.007	0.978 ± 0.028	0.954 ± 0.015	1.072 ± 0.010	0.987 ± 0.021	0.970 ± 0.013	1.084 ± 0.012	0.984 ± 0.017	0.975 ± 0.009
DiffPO	1.032 ± 0.124	0.787 ± 0.111	0.958 ± 0.197	0.961 ± 0.081	0.568 ± 0.032	0.887 ± 0.153	1.371 ± 0.113	0.671 ± 0.071	1.904 ± 0.297
TEDVAE	0.890 ± 0.020	0.548 ± 0.054	0.487 ± 0.059	0.917 ± 0.022	0.552 ± 0.036	0.517 ± 0.052	0.939 ± 0.016	0.551 ± 0.036	0.553 ± 0.070
MitNet	1.289 ± 0.046	2.385 ± 0.028	2.092 ± 0.007	1.384 ± 0.006	2.460 ± 0.006	2.255 ± 0.010	1.387 ± 0.004	2.474 ± 0.019	2.302 ± 0.013
PW-Learner	1.897 ± 0.087	-	-	2.136 ± 0.143	-	-	2.194 ± 0.125	-	-
RA-Learner	1.343 ± 0.011	-	-	1.347 ± 0.007	-	-	1.359 ± 0.010	-	-
X-Learner	1.327 ± 0.009	-	-	1.325 ± 0.003	-	-	1.346 ± 0.019	-	-
DragonNet	1.567 ± 0.073	0.528 ± 0.105	1.603 ± 0.109	2.475 ± 0.058	0.471 ± 0.039	4.221 ± 0.190	2.829 ± 0.087	0.433 ± 0.035	5.564 ± 0.323
DR-Learner	1.470 ± 0.043	-	-	1.521 ± 0.055	-	-	1.494 ± 0.057	-	-
DR-Learner (clip)	1.459 ± 0.022	-	-	1.419 ± 0.017	-	-	1.433 ± 0.040	-	-
RieszNet	0.901 ± 0.021	0.437 ± 0.039	0.477 ± 0.035	0.918 ± 0.009	0.452 ± 0.028	0.528 ± 0.024	0.952 ± 0.026	0.434 ± 0.043	0.586 ± 0.052
R-Learner	0.909 ± 0.035	0.838 ± 0.025	0.982 ± 0.123	1.020 ± 0.048	0.958 ± 0.017	1.121 ± 0.151	1.213 ± 0.096	0.824 ± 0.171	1.569 ± 0.325
GRL-Shared	**0.741 ± 0.005**	0.349 ± 0.027	0.373 ± 0.007	**0.785 ± 0.031**	0.362 ± 0.007	0.414 ± 0.040	0.893 ± 0.058	0.403 ± 0.057	0.577 ± 0.015
GRL-Sep	0.758 ± 0.025	**0.305 ± 0.033**	**0.296 ± 0.024***	0.817 ± 0.012	0.339 ± 0.009	**0.359 ± 0.008***	**0.884 ± 0.034**	0.436 ± 0.039	**0.480 ± 0.040**

Experiments were conducted on a synthetic dataset that utilized real-world covariates $X \in \mathbb{R}^d$ from ACIC2016 competition (to ensure realistic feature correlations), while simulating treatment assignment and outcomes via known functions. To control nuisance estimation errors, we started with oracle nuisances and injected an increasing range of perturbations to the nuisance outcome before sending it in the second-stage target estimation, explicitly testing GRL *vs*. R-learner's vulnerability to nuisance-induced structural distortions. Fig. 1B and Fig. 2A show the resulting metrics as ρ changed. Consistent with our analyses in Section 4, the outcome MSE for R-learner increased at a linear rate with respect to S_{PL} (Proposition 3), and its PEHE increased at square-root rate (Proposition 5). In comparison, GRL was significantly less sensitive to ρ, confirming its effectiveness in controlling nuisance-induced structural drift at the second-stage target estimation.

GRL ablations: We first ablated the absence *vs*. presence of the orthogonal penalty in $\mathcal{L}_{\text{GRL-OP}}$, with its importance demonstrated in Fig. 2B especially in the presence of the skip connection for $\hat{l}$. With the orthogonality penalty present, we further examined the learning rates used for the two optimizers: results in Figure 4A showed that the lowest PEHE was generally obtained when the learning rate for optimizing the raw loss was approximately one order of magnitude larger than that for optimizing the penalty. Furthermore, to test whether the second-stage model may have degenerated into a naive outcome learner that ignores first-stage nuisance estimates, we removed $\hat{l}(X)$ from its input which led to deteriorated performance (4B). This confirmed that the second-stage model learned to use the first-stage nuisance outcome to focus on residual-on-residual learning as intended. Additional ablations on nuisance optimization strategies are provided in Appendix F.2.

6.2 Comparison with Existing Meta-Learners

Synthetic data: On synthetic data described in Section 6.1, we varied the scaling parameter λ in the propensity mechanism from 0 to 1 to test the robustness of GRL to extreme propensities (details in Appendix D.1). Table 1 lists PEHE and Outcome MSE metrics of GRL in comparison to all baseline meta-learners on representative values of λ, with complete results included in Appendix G.

In general, **plug-in learners** outperformed two-stage target-oriented learners at lower value of λ, but not necessarily at higher values of λ. Among **target-oriented meta-learners**, the orthogonal R-learner, RieszNet, and GRL consistently outperformed their non-orthogonal baselines in all settings. Compared to R-learner, GRL further achieved significant $20 - 35\%$ PEHE improvements, providing strong empirical evidence for the target-error analysis in Section 4. Furthermore, as λ increased, GRL showed a increasingly substantial improvement in outcome MSE on subgroups not supported by propensity compared to all baselines, contributing to a low disparity between supported and not-supported MSE; FlexTENet also showed an excellent disparity, but accompanied by lower accuracies. Between the two GRL variants, GRL-Sep offered more balance between accuracy and disparity, while GRL-Shared favored stronger accuracies at a moderate tradeoff of disparity.

Table 2: Quantitative results on IHDP, Twins, and Jobs, with similar annotations used in Table 1

Model	IHDP (PEHE) ↓	Twins (PEHE) ↓	Jobs ($\mathcal{R}_{pol}$) ↓
T-Learner	1.7122 ± 0.3775	0.3303 ± 0.0052	0.2699 ± 0.0452
TARNet	5.0489 ± 1.5188	0.3185 ± 0.0052	0.2638 ± 0.0456
CFR-MMD	4.5710 ± 1.5332	0.3197 ± 0.0051	0.2622 ± 0.0402
SNet	6.2692 ± 2.2065	0.3186 ± 0.0054	0.2646 ± 0.0355
FlexTENet	1.3691 ± 0.3169	**0.3178 ± 0.0052**	0.2683 ± 0.0462
TEDVAE	1.2368 ± 0.2546	0.3217 ± 0.0054	0.2900 ± 0.0650
MitNet	2.9539 ± 1.2969	0.3188 ± 0.0063	0.2734 ± 0.0565
PW-Learner	4.1011 ± 1.9953	0.3551 ± 0.0072	0.2372 ± 0.0499
RA-Learner	1.4696 ± 0.3807	0.3301 ± 0.0058	0.2455 ± 0.0424
X-Learner	1.3700 ± 0.3697	0.3273 ± 0.0050	0.2322 ± 0.0475
DragonNet	4.9916 ± 1.5036	0.3185 ± 0.0052	0.2664 ± 0.0464
DR-Learner	2.3696 ± 0.9964	0.3454 ± 0.0065	0.2320 ± 0.0457
DR-Learner (clip)	2.1548 ± 0.7274	0.3483 ± 0.0045	0.2288 ± 0.0515
RieszNet	1.3251 ± 0.2612	0.3310 ± 0.0055	**0.2253 ± 0.0474**
R-Learner	2.8228 ± 1.1880	0.3433 ± 0.0078	0.2489 ± 0.0580
GRL-Shared	**1.1028 ± 0.1934**	0.3179 ± 0.0053	0.2347 ± 0.0469
GRL-Sep	1.1734 ± 0.2317	0.3179 ± 0.0052	0.2448 ± 0.0629

Semi-synthetic & real data: IHDP (Hill, 2011) is built upon real covariates from randomized control trial involving low-birth-weight infants. To mimic an observational study, we used the non-linear setup B in (Hill, 2011) where control units had exponential response surface and treated units had linear response surface. We used 100 realizations of this setup provided by Shalit et al. (2017). **Twins** (Louizos et al., 2017) is derived from US twin birth registry, and we simulated synthetic treatment assignment following Yoon et al. (2018). **Jobs** (LaLonde, 1986) was constructed by combining a randomized clinical trial with a large observational control group (Smith & Todd, 2005). Since true individual treatment effects were unavailable, we assessed the model's decision-making ability by the *policy risk* on the randomized subset. Details can be found in Appendix D.3.

As summarized in Table 2, **IHDP** induced significant performance differences among the methods tested, among which GRL delivered the strongest PEHE with statistically significant improvements over the next best baseline (TEDVAE). Furthermore, GRL showed a significant gain over R-learner, reducing PEHE by over 50% with significantly lower variances. On **Twins**, about half of the methods (including GRL) were able to obtan a PEHE≤ 0.32 with no statistically-significant differences among their performances. On **Jobs**, a small number of methods (including GRL) were able to achieve an average $\mathcal{R}_{pol} \leq 0.24$, although without statistical significance given the standard deviations observed in all methods. Interestingly, most baselines' performance fluctuated among the three datasets. In particular, **plug-in meta-learners** in general delivered stronger PEHE on Twins, but the opposite was observed on Jobs. Among the **orthogonal learners**, DragonNet showed strong performance on Twins yet among the poorest on Jobs and IHDP; in contrast, DR-learner was strong at Jobs yet suboptimal on Twins and IHDP. In comparison, GRL was the only method that was consistently among the top performers without significant deterioration at any particular dataset. These results provided strong evidence for the robustness of GRL to nuisance estimation errors.

7 Conclusions & Discussion

We present a target-error analysis for the R-learner that isolates the impact of nuisance-induced structural distortions and reveal the importance of the second-stage model class in addressing this distortion. We further present GRL that extends the R-learner with a nonlinear model class in stage two to reduce structural-drift while establishing a nuisance-robust population risk bound for the resulting nonlinear target. Experimental results demonstrated GRL's significant improvment in robustness to nuisance estimation error as well as accuracy in PO and CATE estimation. **Limitations.** We considered only two basic instantiations of GRL's nuisance learner. While this showed strong performance on synthetic data, gain of performance on real data was not as significant and could be enhanced by integration of more advanced nuisance learner designs. PO and CATE evaluation on real-data remain challenging due to the lack of ground-truth: the relative differences among models were moderate with limited statistical significance on the Jobs dataset, motivating the need of additional real-data benchmarks in this community. Finally, GRL considered only static treatment effect and needs further developments for temporal settings.

AI USE STATEMENT

In this work, we used generative AI tools for exploring different proof strategies. We have not used generative AI tools for formulating core hypothesis or designing experiments. Additionally, we used generative AI tools for polishing the writing, and generating code for some of the figures. We have reviewed all AI-assisted work. We reviewed all LLM-generated text and code for correctness. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work develops method for estimating causal effect from data. While our contributions are mostly theoretical, applying these tools in high stakes areas like healthcare or public policy may carry ethical risks. If the underlying data is flawed or unrepresentative, deploying these models can unintentionally lead to unfair decision choices. We strongly urge practitioners to rigorously audit their data for fairness and continuously monitor how these models impact different aspects of fairness, and safety before using them for real world interventions.

REPRODUCIBILITY STATEMENT

We provide details to ensure reproducibility in the appendices: mathematical proofs and theoretical claims are detailed in Appendix C, dataset details in Appendix D, and implementation details in Appendix E. Source code will be made publicly available upon publication.

REFERENCES

Peter Bartlett, Dylan J Foster, and Matus J Telgarsky. Spectrally-normalized margin bounds for neural networks. Advances in neural information processing systems, 30, 2017.

Peter L Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. Journal of machine learning research, 3(Nov):463–482, 2002.

Ioana Bica, Ahmed M Alaa, Craig Lambert, and Mihaela Van Der Schaar. From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges. Clinical Pharmacology & Therapeutics, 109(1):87–100, 2021.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters, 2018.

Victor Chernozhukov, Whitney Newey, Vıctor M Quintas-Martınez, and Vasilis Syrgkanis. Riesznet and forestriesz: Automatic debiased machine learning with neural nets and random forests. In International Conference on Machine Learning, pp. 3901–3914. PMLR, 2022.

Zhixuan Chu, Stephen L Rathbun, and Sheng Li. Learning infomax and domain-independent representations for causal effect inference with real-world data. In Proceedings of the 2022 SIAM International Conference on Data Mining (SDM), pp. 433–441. SIAM, 2022.

Richard K Crump, V Joseph Hotz, Guido W Imbens, and Oscar A Mitnik. Dealing with limited overlap in estimation of average treatment effects. Biometrika, 96(1):187–199, 2009.

Alicia Curth and Mihaela Van der Schaar. On inductive biases for heterogeneous treatment effect estimation. Advances in Neural Information Processing Systems, 34:15883–15894, 2021a.

Alicia Curth and Mihaela Van der Schaar. Nonparametric estimation of heterogeneous treatment effects: From theory to learning algorithms. In International Conference on Artificial Intelligence and Statistics, pp. 1810–1818. PMLR, 2021b.

Dylan J Foster and Vasilis Syrgkanis. Orthogonal statistical learning. The Annals of Statistics, 51 (3):879–908, 2023.

Emilio AL Gianicolo, Martin Eichler, Oliver Muensterer, Konstantin Strauch, and Maria Blettner. Methods for evaluating causality in observational studies. Deutsches Arzteblatt International, 116 (7):101–107, 2020.

Xingzhuo Guo, Yuchen Zhang, Jianmin Wang, and Mingsheng Long. Estimating heterogeneous treatment effects: Mutual information bounds and learning algorithms. In International Conference on Machine Learning, pp. 12108–12121. PMLR, 2023.

Zhenyu Guo, Shuai Zheng, Zhizhe Liu, Kun Yan, and Zhenfeng Zhu. Cetransformer: Casual effect estimation via transformer based representation learning. In Chinese Conference on Pattern Recognition and Computer Vision (PRCV), pp. 524–535. Springer, 2021.

Negar Hassanpour and Russell Greiner. Learning disentangled representations for counterfactual regression. In International Conference on Learning Representations, 2019.

Miguel A. Hernán and James M. Robins. Causal Inference: What If. Chapman & Hall/CRC, Boca Raton, FL, 2020.

Jennifer L Hill. Bayesian nonparametric modeling for causal inference. Journal of Computational and Graphical Statistics, 20(1):217–240, 2011.

Fredrik Johansson, Uri Shalit, and David Sontag. Learning representations for counterfactual inference. In International conference on machine learning, pp. 3020–3029. PMLR, 2016.

Edward H Kennedy. Towards optimal doubly robust estimation of heterogeneous causal effects. Electronic Journal of Statistics, 17(2):3008–3049, 2023.

Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the national academy of sciences, 116(10):4156–4165, 2019.

Robert J LaLonde. Evaluating the econometric evaluations of training programs with experimental data. The American economic review, pp. 604–620, 1986.

Christos Louizos, Uri Shalit, Joris M Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. Advances in neural information processing systems, 30, 2017.

Yuchen Ma, Valentyn Melnychuk, Jonas Schweisthal, and Stefan Feuerriegel. Diffpo: A causal diffusion model for learning distributions of potential outcomes. Advances in Neural Information Processing Systems, 37:43663–43692, 2024.

Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects. Biometrika, 108(2):299–319, 2021.

James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. Journal of the American statistical Association, 89 (427):846–866, 1994.

P. M. Robinson. Root-n-consistent semiparametric regression. Econometrica, 56(4):931–954, 1988. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1912705.

Andrea Rotnitzky, James M Robins, and Daniel O Scharfstein. Semiparametric regression for repeated outcomes with nonignorable nonresponse. Journal of the american statistical association, 93(444):1321–1339, 1998.

Uri Shalit, Fredrik D Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In International conference on machine learning, pp. 3076–3085. PMLR, 2017.

Claudia Shi, David Blei, and Victor Veitch. Adapting neural networks for the estimation of treatment effects. Advances in neural information processing systems, 32, 2019.

Jeffrey A Smith and Petra E Todd. Does matching overcome lalonde's critique of nonexperimental estimators? Journal of econometrics, 125(1-2):305–353, 2005.

Mark J Van der Laan, Sherri Rose, et al. Targeted learning: causal inference for observational and experimental data, volume 4. Springer, 2011.

Anpeng Wu, Junkun Yuan, Kun Kuang, Bo Li, Runze Wu, Qiang Zhu, Yueting Zhuang, and Fei Wu. Learning decomposed representations for treatment effect estimation. IEEE Transactions on Knowledge and Data Engineering, 35(5):4989–5001, 2022.

Liuyi Yao, Sheng Li, Yaliang Li, Mengdi Huai, Jing Gao, and Aidong Zhang. Representation learning for treatment effect estimation from observational data. Advances in neural information processing systems, 31, 2018.

Liuyi Yao, Sheng Li, Yaliang Li, Mengdi Huai, Jing Gao, and Aidong Zhang. Ace: Adaptively similarity-preserved representation learning for individual treatment effect estimation. In 2019 IEEE International Conference on Data Mining (ICDM), pp. 1432–1437. IEEE, 2019.

Jinsung Yoon, James Jordon, and Mihaela Van Der Schaar. Ganite: Estimation of individualized treatment effects using generative adversarial nets. In International conference on learning representations, 2018.

Weijia Zhang, Lin Liu, and Jiuyong Li. Treatment effect estimation with disentangled latent factors. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pp. 10923–10930, 2021.

A ASSUMPTIONS OF TREATMENT-EFFECT IDENTIFIABILITY

Assumption 8. Stable Unit Treatment Value Assumption (SUTVA) and Consistency: Let $Y_i^{(t)}$ denote the potential outcome for unit i under treatment $t \in \{0, 1\}$, and T_i the observed treatment. SUTVA implies no interference between units and no hidden versions of treatment, so potential outcomes are well-defined. Consistency requires $Y_i = Y_i^{(T_i)}$ meaning observed outcome under a treatment match their potential outcome.

Assumption 9. Ignorability (Unconfoundedness): Treatment assignment is independent of potential outcomes conditional on covariates i.e. $(Y^{(0)}, Y^{(1)}) \perp T \mid X$.

Assumption 10. Positivity (Overlap): Each treatment has non-zero probability of being assigned given covariates i.e. $0 < \alpha \leq \mathbb{P}(T = 1 \mid X) \leq 1 - \alpha < 1, \forall X$.

B A UNIFIED LENS OF ORTHOGONAL LEARNERS

We adopt the perspective of *orthogonal learners* to organize a range of existing approaches into a single conceptual framework, to reveal a clear and coherent structure underlying them. We anchor most existing orthogonal estimators of treatment effects on an estimating function known as the augmented inverse propensity weighting (AIPW) (Robins et al., 1994; Rotnitzky et al., 1998). More specifically, we show that DML (Chernozhukov et al., 2018) and TMLE (Van der Laan et al., 2011) represent two different approaches to arrive at this estimating function for ATE, while the DR-learner (Kennedy, 2023) and DragonNet (Shi et al., 2019) represent their respective extensions to CATE meta-learners. This estimating function has the advantage of not imposing structural assumptions on the underlying data generating equations. It however involves the calculation of inverse propensity, which can result in high variance and numerical instability at extreme propensities.

We then contrast the above approaches with a distinct orthogonal learner that overcomes this issue, found in the seminal work of *Robinson's residual-on-residual* ATE estimator (Robinson, 1988). Instead of treating outcome as the nuisance parameter, this approach defines a different nuisance function, $\mathbb{E}[Y|X]$, that absorbs the contribution of T that can be predicted by X. This important step partials out a residual treatment and residual outcome variable, resulting in the well-known *residual-on-residual* regression of ATE: importantly, this regression induces a Neyman-orthogonal estimating function that does not involve inverse propensity (Robinson, 1988). It was also extended to a CATE meta-learner, known as the R-learner (Nie & Wager, 2021). However, this approach and its estimating function are currently derived for a partially linear outcome equation. This restricts its ability to absorb first stage nuisance estimation errors..

In treatment effect estimation, the target quantity of interest can include: (i) ATE, $\tau = \mathbb{E}[Y^{(1)} - Y^{(0)}]$, representing the expected difference in potential control outcome $Y^{(0)}$ and treated outcome $Y^{(1)}$ across the entire population, and (ii) CATE, $\tau(x) = \mathbb{E}[Y^{(1)} - Y^{(0)}|X = x]$, representing heterogeneous effects conditioned on covariates and (iii) potential outcome, $\mathbb{E}[Y|X = x, T = t]$, as a function that can be used to calculate CATE/ATE. We adopt common assumptions for identifiability of ATE and CATE (Shalit et al., 2017; Yoon et al., 2018), including (i) *SUTVA and consistency*, meaning treatment and outcomes are well defined, and observed outcome under a treatment match their potential outcome; (ii) *ignorability,* meaning treatment assignment is independent of potential outcomes given covariates; and (iii) *positivity,* meaning each treatment has non-zero probability of being assigned given covariates as detailed in Appendix A.

B.1 NEYMAN ORTHOGONALITY

Consider the estimation of a target parameter τ in the presence of high-dimensional nuisance parameters η. Assume an estimating function $\psi(W; \tau, \eta)$ defined over observed data W, which satisfies the population moment condition at the true values of $\tau = \tau_0$ and $\eta = \eta_0$: $\mathbb{E}[\psi(W; \tau_0, \eta_0)] = 0$. The estimator $\hat{\tau}$ is obtained by solving the empirical counterpart $\mathbb{E}_n[\psi(W; \tau, \hat{\eta})] = 0$, where $\hat{\eta}$ is an estimate of the nuisance parameters and n the number of observed data samples.The estimating function ψ is said to satisfy Neyman orthogonality with respect to the nuisance parameter η at (τ_0, η_0), if the pathwise derivative of the population moment with respect to η vanishes at $\eta = \eta_0$:

$$\partial_\eta \mathbb{E}[\psi(W; \tau_0, \eta)]\Big|_{\eta=\eta_0} = 0$$

Intuitively, this condition implies that the moment used to estimate τ is locally insensitive to small errors in the estimation of nuisance parameters η. This allows the use of ML/DL models for the estimation of η, potentially at slower rates, without acting as bottleneck for the convergence of target estimator as if nuisance parameters were known.

B.2 AIPW ESTIMATING FUNCTION

Consider a general outcome generating process:

$$Y = m(T, X) + \xi_y, \quad \mathbb{E}[\xi_y|X, T] = 0; \quad T = e(X) + \xi_t, \quad \mathbb{E}[\xi_t|X] = 0 \tag{15}$$

With nuisance parameters $\eta = \{m(T,X), e(X)\}$ (outcome and propensity functions), a common Neyman-orthogonal estimating function for ATE τ is the AIPW function:

$$\psi_{AIPW}(W;\tau,\eta) = \underbrace{H(T,X)(Y - m(T,X))}_{\text{residual term (a)}} + \underbrace{m(1,X) - m(0,X) - \tau}_{\text{substitution term (b)}} \tag{16}$$

where $H(T,X) = \frac{T}{e(X)} - \frac{1-T}{1-e(X)}$.

ATE estimator For estimating ATE, DML as presented in (Chernozhukov et al., 2018) and TMLE (Van der Laan et al., 2011) can be viewed as two alternative approaches to leverage $\psi_{AIPW}(W;\tau,\hat{\eta})$ given the first-stage nuisance estimators $\hat{\eta}$. In DML, τ is directly estimated from Equation 16 by obtaining $\hat{\tau}$ that sets the empirical average of ψ_{AIPW} to zero given the nuisance estimators $\hat{\eta}$, *i.e.*,

$$\hat{\tau} \quad \text{solves} \quad \frac{1}{n}\sum\nolimits_i \psi_{AIPW}(W_i;\tau,\hat{\eta}) = 0 \tag{17}$$

In TMLE (Van der Laan et al., 2011), instead of targeting τ directly, the outcome estimator $\hat{m}$ is corrected to $\hat{m}^*$ in the form of a parametric submodel $\hat{m}^*(T,X) = \hat{m}(T,X) + \epsilon H(T,X)$. Substituting $\hat{m}^*$ in Equation 16, term (b) sums to zero while the residual term (a) does not. TMLE then optimizes ϵ to minimize:

$$\epsilon^* = \arg\min_{\epsilon} \frac{1}{n}\sum\nolimits_i ||Y_i - m_i^*||_2^2 \tag{18}$$

which induces an estimating function corresponding to the residual term (a) in Equation 16. The final estimate $\hat{\tau}_{TMLE}$ is then obtained from the updated outcome model.

CATE estimator For CATE, the DR-learner (Kennedy, 2023) constructs a pseudo-outcome $\tilde{Y}_{DR}$ by plugging estimates of the nuisance parameters ($\hat{\eta} = \{\hat{m}, \hat{e}\}$) as:

$$\tilde{Y}_{DR} = \hat{H}(T,X)(Y - \hat{m}(T,X)) + (\hat{m}(1,X) - \hat{m}(0,X))$$

CATE is then estimated by regressing $\tilde{Y}_{DR}$ on covariates as:

$$\hat{\tau}_{DR} = \arg\min_{\tau} \sum\nolimits_i (\tilde{Y}_{DR,i} - \tau(X_i))^2 \tag{19}$$

The DR learner can be viewed as extending the DML approach with the AIPW estimating function to the function-valued $\tau(x)$. Similarly, DragonNet (Shi et al., 2019) adapts the principle of TMLE to learn the nuisance outcome m as a neural network: it defines m^* as defined in TMLE and minimizes Equation 18 to jointly optimize ϵ and m. The corrected m^* is then used for ATE/CATE estimation.

Limitations While free of structural constraints, the AIPW involves inverse propensity calculation via the term $\hat{H}(T,X)$. When propensity scores are close to 0 or 1, it can lead to high variance and instability in the estimator. Although in practice this is mitigated by techniques such as propensity clipping or discarding samples whose propensity falls outside a predefined range (*e.g.*, [0.1, 0.9]) (Crump et al., 2009).

B.3 Robinson's Partialing-out Transformation

Robinson's seminal transformation (Robinson, 1988) provides an alternative path to orthogonality that avoids the instability of inverse probability calculation. Under the structural assumption of a partially linear model:

$$\begin{aligned} Y &= g(X) + \tau * T + \xi_y \quad \mathbb{E}[\xi_y|X,T] = 0 \\ T &= e(X) + \xi_t, \quad \mathbb{E}[\xi_t|X] = 0 \end{aligned} \tag{20}$$

we can write conditional mean of outcome as: $E[Y|X] = g(X) + \tau * e(X) = l(X)$, and re-write Equation 20 as:

$$Y - \mathbb{E}[Y|X] = \tau * (T - e(X))) + \xi_y \tag{21}$$

Equation 21 creates a different de-composition of the estimation process, with nuisance parameter $\eta = \{\mathbb{E}[Y|X], e\}$ and the target parameter τ estimated by an estimating function that is Neyman orthogonal to $\hat{\mathbb{E}}[Y|X]$ and $\hat{e}$:

$$\psi_{Rob}(W;\tau,\eta) = [Y - \mathbb{E}[Y|X] - \tau * (T - e(X))](T - e(X)) \tag{22}$$

ATE estimator Robinson's transformation is originally presented for ATE estimation via a residual-on-residual fit:

$$\hat{\tau} = \arg\min_{\tau} \frac{1}{n} \sum\nolimits_i ||(Y_i - \hat{\mathbb{E}}[Y_i|X_i]) - \tau(T_i - \hat{e}(X_i))||_2^2 \tag{23}$$

which induces $\psi_{Rob} = 0$ in Equation 22.

CATE estimator The R-learner (Nie & Wager, 2021) directly extends the residual-on-residual regression to CATE:

$$\hat{\tau}_R = \arg\min_{\tau} \frac{1}{n} \sum\nolimits_i \Big((Y_i - \hat{\mathbb{E}}[Y_i|X_i]) - \tau(X_i)(T_i - \hat{e}(X_i))\Big)^2 \tag{24}$$

Similar to the DR-learner, the R learner can be seen as extending the Robinson's estimating function from a low-dimensional parameter τ to the function-valued parameter $\tau(x)$.

Relation with DML Robinson's approach can be viewed as a specific approach to achieve orthogonal learners, anchored on a distinctive decomposition that *partials* out a residual effect from $\mathbb{E}[Y|X]$. One also has the option to approach the partially linear equation with a general DML approach discussed in Section B.2, considering g and e as the nuisance parameter followed by an orthogonal estimating function: $\psi_{DML_{PL}} = (Y - g(X) - \tau T)(T - e(X))$. As acknowledged in (Chernozhukov et al., 2018), this would require the construction of an estimator for $g(X) = E[Y - \tau * T|X]$ that either depends on τ or require specific assumptions. In comparison, $\mathbb{E}[Y|X]$ as a nuisance parameter can be directly estimated with standard DL/ML methods.

Limitations While numerically robust to extreme propensity scores, Robinson's partialling-out approach and its associated estimating function, $\psi_{Rob}(W; \tau, \eta)$, are derived under the structural assumption of a partially linear outcome equation.

C PROOFS

C.1 PROOF FOR $R_m = \|f_m(\hat{\eta}) - f_0\|_2^2$ IN DEFINITION 1

Proof. For outcome $Y = f_0(X,T) + \xi_y$, $\mathbb{E}[\xi_y \mid X,T] = 0$, and squared loss, we have

$$\begin{aligned} L(f) &= \mathbb{E}\Big[\big(f_0(X,T) + \xi_y - f(X,T)\big)^2\Big] \\ &= \mathbb{E}\Big[\big(\xi_y - (f(X,T) - f_0(X,T))\big)^2\Big] \\ &= \mathbb{E}[\xi_y^2] - 2\,\mathbb{E}\Big[\xi_y\,(f(X,T) - f_0(X,T))\Big] + \mathbb{E}\Big[(f(X,T) - f_0(X,T))^2\Big]. \end{aligned}$$

Since $f(X,T) - f_0(X,T)$ is measurable with respect to (X,T) and $\mathbb{E}[\xi_y \mid X,T] = 0$, the cross term vanishes:

$$\mathbb{E}\Big[\xi_y\,(f(X,T) - f_0(X,T))\Big] = \mathbb{E}\Big[\mathbb{E}[\xi_y \mid X,T]\,(f(X,T) - f_0(X,T))\Big] = 0.$$

Hence

$$L(f) = \mathbb{E}[\xi_y^2] + \mathbb{E}\Big[(f(X,T) - f_0(X,T))^2\Big].$$

Also,

$$L(f_0) = \mathbb{E}\Big[(Y - f_0(X,T))^2\Big] = \mathbb{E}[\xi_y^2].$$

Subtracting gives

$$L(f) - L(f_0) = \mathbb{E}\Big[(f(X,T) - f_0(X,T))^2\Big] = \|f - f_0\|_2^2.$$

Substituting f by $f_m(\hat{\eta})$, we get the result. □

C.2 STRUCTURAL DRIFT DECOMPOSITION

Expand the class-optimal plug-in target $f_m^*(\hat{\eta})$ around η_0,

$$d_m := f_m^*(\hat{\eta}) - f_0(\eta_0) = \underbrace{f_m^*(\eta_0) - f_0(\eta_0)}_{a_m} + \underbrace{\nabla_\eta f_m^*(\eta_0)\,[\delta\eta]}_{b_m(\delta\eta)} + r_m(\delta\eta),\, r_m(\delta\eta) = O\big(\|\delta\eta\|^2\big) \tag{25}$$

$$S_m = \|a_m + b_m(\delta\eta) + r_m(\delta\eta)\|_2^2 = \underbrace{\|a_m\|_2^2}_{\text{Oracle mismatch}} + \underbrace{2\langle a_m, b_m(\delta\eta)\rangle}_{\text{First order contribution}} + \underbrace{\|b_m(\delta\eta)\|_2^2}_{\text{Quadratic drift}} + O\big(\|\delta\eta\|^3\big). \tag{26}$$

where a_m captures whether the oracle target f_0 belongs to plug-in oracle class m, which we refer to as "oracle mismatch". $b_m(\delta\eta)$ represents the first-order effect of replacing oracle nuisance pair η_0 with plug-in nuisance pair $\hat{\eta}$ on d_m, which affects the risk-level structural drift S_m through both a first order term and a quadratic term, along with higher-order remainder terms.

C.3 PROOF OF PROPOSITION 3

Definition 11 (Oracle-correctness)**.** *A model class $\mathcal{F}_m$ is said to be oracle-correct if, under the true nuisance pair η_0, it contains the oracle target $f_0(\eta_0)$; that is, $f_0(\eta_0) \in \mathcal{F}_m(\eta_0)$. Equivalently, the class-optimal oracle target satisfies $f_m^*(\eta_0) = f_0(\eta_0)$.*

Partially-linear model class: $\mathcal{F}_{\mathrm{PL}}$ represents the class of functions used by R-learner. In the partially-linear class, the oracle mismatch takes the form $a_{\mathrm{PL}}(X,T) = \big(T - e_0(X)\big)\big(\tau_{\mathrm{PL},0}^*(X) - \tau_0(X)\big)$, so the class is oracle-correct when $\tau_{\mathrm{PL},0}^* = \tau_0$.

Proposition (Partially-linear structural drift)**.** Suppose that the partially-linear class is oracle-correct, so that $\tau_{\mathrm{PL}}^*(\eta_0) = \tau_0$, then structural drift is given by

$$S_{\mathrm{PL}} = \mathbb{E}\Big[\big(\Delta l(X) - \Delta e(X)\tau_0(X)\big)^2\Big] + O\big(\|\delta\eta\|^3\big). \tag{27}$$

Proof. There are two ways to derive the structural drift term.

First start with structural drift defined in equation 26. Under oracle correct case

$$S_{\mathrm{PL}} = \|b_{PL}(\delta\eta)\|_2^2 + O\big(\|\delta\eta\|^3\big). \tag{28}$$

where $b_{PL}(\delta\eta) = \nabla_\eta f^*_{PL}(\eta_0)[\delta\eta]$.

For the partially-linear class consider a function: $f_{\mathrm{PL}}(\eta)(X,T) = l(X) + \big(T - e(X)\big)\tau(X)$. Computing gradients we get,

$$\frac{\partial f_{PL}}{\partial l} = 1, \quad \frac{\partial f_{PL}}{\partial e} = -\tau(X). \tag{29}$$

Under oracle correct case and $\eta = \eta_0$, $\nabla_\eta f^*_{PL}(\eta_0) = (1, -\tau_0(X))$, hence

$$b_{PL}(\delta\eta) = \Delta l(X) - \Delta e(X)\tau_0(X). \tag{30}$$

Using this in equation equation 28, we get the desired result.

Alternatively,

For the partially-linear class, $f_{\mathrm{PL}}(\hat\eta)(X,T) = \hat l(X) + \big(T - \hat e(X)\big)\tau(X)$.
Subtracting the oracle target, $f_0(X,T) = l_0(X) + \big(T - e_0(X)\big)\tau_0(X)$ gives

$$f_{\mathrm{PL}}(\hat\eta)(X,T) - f_0(X,T) = \Delta l(X) + \big(T - e_0(X)\big)\big(\tau(X) - \tau_0(X)\big) - \Delta e(X)\tau(X). \tag{31}$$

Expanding the class-optimal coefficient around η_0:

$$\tau^*_{\mathrm{PL}}(\hat\eta)(X) = \tau^*_{\mathrm{PL}}(\eta_0)(X) + \nabla_\eta \tau^*_{\mathrm{PL}}(\eta_0)(X)[\delta\eta] + r_{\mathrm{PL}}(\delta\eta), \qquad r_{\mathrm{PL}}(\delta\eta) = O\big(|\delta\eta|^2\big).$$

Evaluating equation 31 at $\tau = \tau^*_{\mathrm{PL}}(\hat\eta)$ yields

$$\begin{aligned} d_{\mathrm{PL}}(\hat\eta) = {} & \underbrace{\big(T - e_0(X)\big)\big(\tau^*_{\mathrm{PL}}(\eta_0)(X) - \tau_0(X)\big)}_{a_{\mathrm{PL}}} \\ & + \underbrace{\Big[\Delta l(X) + \big(T - e_0(X)\big)\nabla_\eta \tau^*_{\mathrm{PL}}(\eta_0)(X)[\delta\eta] - \Delta e(X)\,\tau^*_{\mathrm{PL}}(\eta_0)(X)\Big]}_{b_{\mathrm{PL}}(\delta\eta)} \\ & + \underbrace{\Big[\big(T - e_0(X)\big) r_{\mathrm{PL}}(X) - \Delta e(X)\,\nabla_\eta \tau^*_{\mathrm{PL}}(\eta_0)(X)[\delta\eta] - \Delta e(X)\, r_{\mathrm{PL}}(X)\Big]}_{\text{Higher order}}. \end{aligned}$$

Let $U = T - e_0(X)$, $\sigma_T^2(X) = \mathrm{Var}(T \mid X)$, and define, for fixed X,

$$J_X(\tau) := \mathbb{E}\Big[\big(\Delta l(X) + U(\tau(X) - \tau_0(X)) - \Delta e(X)\,\tau(X)\big)^2 \,\Big|\, X\Big].$$

Using $\mathbb{E}[U \mid X] = 0$ and $\mathbb{E}[U^2 \mid X] = \sigma_T^2(X)$,

$$J_X(\tau) = (\Delta l(X) - \Delta e(X)\,\tau(X))^2 + \sigma_T^2(X)(\tau(X) - \tau_0(X))^2.$$

Differentiating gives

$$-2\Delta e(X)\big(\Delta l(X) - \Delta e(X)\tau(X)\big) + 2\sigma_T^2(X)\big(\tau(X) - \tau_0(X)\big) = 0,$$

Hence,

$$\tau^*_{\mathrm{PL}}(\hat\eta)(X) = \frac{\Delta e(X)\Delta l(X) + \sigma_T^2(X)\tau_0(X)}{\Delta e(X)^2 + \sigma_T^2(X)}.$$

$$\tau^*_{\mathrm{PL}}(\hat\eta)(X) - \tau_0(X) = \frac{\Delta e(X)\big(\Delta l(X) - \Delta e(X)\tau_0(X)\big)}{\Delta e(X)^2 + \sigma_T^2(X)} = O\big(\|\delta\eta\|^2\big).$$

Since it is second order, the first-order term $\nabla_\eta \tau^*_{\mathrm{PL}}(\eta_0)(X)[\delta\eta] = 0$.
Substituting $\tau^*_{\mathrm{PL}}(X) = \tau_0(X)$ for oracle-correct and $\nabla_\eta \tau^*_{\mathrm{PL}}(\eta_0)(X)[\delta\eta] = 0$ into $b_{\mathrm{PL}}(\delta\eta)$ gives

$$b_{\mathrm{PL}}(\delta\eta) = \Delta l(X) - \Delta e(X)\,\tau_0(X).$$

Hence

$$\|b_{\mathrm{PL}}(\delta\eta)\|_2^2 = \mathbb{E}\Big[\big(\Delta l(X) - \Delta e(X)\tau_0(X)\big)^2\Big].$$

$$S_{\mathrm{PL}} = \|b_{\mathrm{PL}}(\delta\eta)\|_2^2 + O\big(\|\delta\eta\|^3\big),$$
$$S_{\mathrm{PL}} = \mathbb{E}\Big[\big(\Delta l(X) - \Delta e(X)\tau_0(X)\big)^2\Big] + O\big(\|\delta\eta\|^3\big).$$

□

C.4 Proof of Proposition 4

Proposition (Nonlinear structural drift)**.** Suppose the nonlinear class is oracle-correct, so that $f^*_{\mathrm{NL}}(\eta_0) = f_0$, then structural drift is given by

$$S_{\mathrm{NL}} = \|\nabla_\eta h^*(\eta_0)[\delta\eta]\|_2^2 + O\big(\|\delta\eta\|^3\big).$$

Proof. For the nonlinear class, $f_{\mathrm{NL}}(\hat\eta)(X,T) = h\big(X, \hat l(X), T - \hat e(X)\big)$.
Subtracting the oracle target $f_0(X,T) = l_0(X) + \big(T - e_0(X)\big)\tau_0(X)$ gives,

$$h\big(X, \hat l(X), T - \hat e(X)\big) - f_0(X,T) = h\big(X, \hat l(X), T - \hat e(X)\big) - \Big(l_0(X) + \big(T - e_0(X)\big)\tau_0(X)\Big) \quad (32)$$

Expanding the class-optimal nonlinear target around η_0:

$$h^*(\hat\eta)(X,T) = h^*(\eta_0)(X,T) + \nabla_\eta h^*(\eta_0)(X,T)[\delta\eta] + r_{\mathrm{NL}}(\delta\eta), \qquad r_{\mathrm{NL}}(\delta\eta) = O\big(\|\delta\eta\|^2\big).$$

Subtracting $f_0(X,T)$ yields

$$\begin{aligned} d_{\mathrm{NL}}(\hat\eta) &= h^*(\hat\eta) - f_0 \\ &= \underbrace{\Big[h^*(\eta_0)(X,T) - \big(l_0(X) + (T - e_0(X))\tau_0(X)\big)\Big]}_{a_{\mathrm{NL}}} \\ &\quad + \underbrace{\nabla_\eta h^*(\eta_0)(X,T)[\hat\eta - \eta_0]}_{b_{\mathrm{NL}}(\delta\eta)} \\ &\quad + \underbrace{r_{\mathrm{NL}}(\delta\eta)}_{\text{Higher order}}. \end{aligned}$$

For oracle correct case, $h^*(\eta_0)(X,T) = f_0(X,T) = l_0(X) + \big(T - e_0(X)\big)\tau_0(X)$
Hence

$$S_{\mathrm{NL}} = \|b_{\mathrm{NL}}(\delta\eta)\|_2^2 + O\big(\|\delta\eta\|^3\big) = \|\nabla_\eta h^*(\eta_0)[\delta\eta]\|_2^2 + O\big(\|\delta\eta\|^3\big).$$

□

C.5 Proof of Proposition 5

Proposition (CATE target error decomposition)**.** Let $S^\tau_m = \|\tau^*_m(\hat\eta) - \tau_0\|_2^2, E^\tau_m = \|\hat\tau_m - \tau^*_m(\hat\eta)\|_2^2$, and $C^\tau_m = \langle \hat\tau_m - \tau^*_m(\hat\eta), \tau^*_m(\hat\eta) - \tau_0\rangle$. The squared precision in estimation of heterogeneous effect (PEHE) admits decomposition $\mathrm{PEHE}^2_m = \|\hat\tau_m - \tau_0\|_2^2 = S^\tau_m + E^\tau_m + 2C^\tau_m$. Because $|C^\tau_m| \le \sqrt{S^\tau_m E^\tau_m}$, we have $\big(\sqrt{S^\tau_m} - \sqrt{E^\tau_m}\big)^2 \le \mathrm{PEHE}^2_m \le \big(\sqrt{S^\tau_m} + \sqrt{E^\tau_m}\big)^2$.

Proof. We have, $\hat\tau_m - \tau_0 = (\hat\tau_m - \tau^*_m(\hat\eta)) + (\tau^*_m(\hat\eta) - \tau_0)$. Expanding the squared L_2 norm gives $\|\hat\tau_m - \tau_0\|_2^2 = \Big\|\hat\tau_m - \tau^(_m\hat\eta)\Big\|_2^2 + \|\tau^*_m(\hat\eta) - \tau_0\|_2^2 + 2\,\langle \hat\tau_m - \tau^*_m(hat\eta), \tau^*_m(\hat\eta) - \tau_0\rangle$.

The interaction bound follows from Cauchy–Schwarz:

$$\begin{aligned} |C^\tau_m| &\le \|\hat\tau_m - \tau^*_m(\hat\eta)\|_2 \, \|\tau^*_m(\hat\eta) - \tau_0\|_2 \\ &= \sqrt{E^\tau_m S^\tau_m} \end{aligned}$$

□

C.6 Proof of proposition 6

Proposition (PEHE bound)**.** If $\alpha \le e_0(X) \le 1-\alpha$ almost surely for some $\alpha > 0$, then $\mathrm{PEHE}_m^2 \le \frac{2}{\alpha}\|f_m(\hat{\eta}) - f_0\|_2^2 = \frac{2}{\alpha}R_m = \frac{2}{\alpha}(S_m + E_m)$.

Proof. Let $g_m = f_m(\hat{\eta}) - f_0(\eta_0)$. Since $(a-b)^2 \le 2a^2 + 2b^2$,

$$\mathrm{PEHE}_m^2 \le 2\mathbb{E}[g_m(X,1)^2 + g_m(X,0)^2].$$

Moreover,

$$\begin{aligned}\mathbb{E}\left[g_m(X,T)^2 \mid X\right] &= e_0(X)g_m(X,1)^2 + (1-e_0(X))\,g_m(X,0)^2 \\ &\ge \alpha\left(g_m(X,1)^2 + g_m(X,0)^2\right).\end{aligned}$$

Therefore,

$$\mathbb{E}\left[g_m(X,1)^2 + g_m(X,0)^2\right] \le \frac{1}{\alpha}\mathbb{E}\left[g_m(X,T)^2\right].$$

Combining the two inequalities gives

$$\mathrm{PEHE}_m^2 \le \frac{2}{\alpha}\mathbb{E}\left[g_m(X,T)^2\right] = \frac{2}{\alpha}\left\|f_m(\hat{\eta}) - f_0(\eta_0)\right\|_2^2.$$

□

C.7 Proof of proposition 7

Proposition (Nuisance-robust population risk bound)**.** Let $\Omega(h;\hat{\eta})$ represent the penalty terms of $\mathcal{L}_{\text{GRL-OP}}$ in Equation 14, so that $\mathcal{L}_{\text{GRL-OP}}(h;\hat{\eta}) = \mathcal{L}(h(\hat{\eta})) + \Omega(h;\hat{\eta})$, $h_0(\eta_0) \in \arg\min_{h\in\mathcal{H}} \mathcal{L}(h(\eta_0))$ and $h^*(\hat{\eta}) \in \arg\min_{h\in\mathcal{H}} \mathcal{L}_{\text{GRL-OP}}(h;\hat{\eta})$. Assume $h \in \mathcal{H}$ is twice continuously differentiable in its nuisance inputs with bounded derivatives, $\mathcal{H}$ and f_0 are uniformly bounded, and $\sigma_{\min}^2 \le \mathrm{Var}(Y \mid X,T) \le \sigma_{\max}^2$ for some $0 < \sigma_{\min}^2 \le \sigma_{\max}^2 < \infty$. Then there are constants $\kappa, C < \infty$, depending only on these bounds and positivity constant α, such that

$$\mathcal{L}(h^*(\eta_0)) - \mathcal{L}(h_0(\eta_0)) \;\le\; \Omega(h_0;\hat{\eta}) + \sqrt{\kappa\,\Omega(h_0;\hat{\eta})}\,\|\delta\eta\| + C\,\|\delta\eta\|^2.$$

which approximates a second-order excessive risk via the minimization of $\Omega(h;\hat{\eta})$.

Proof. Let M bound $|h|$, $|f_0|$ and the derivatives of h in its nuisance inputs, and recall $\Omega(h;\hat{\eta}) = \mathbb{E}[\|\nabla_l h\|^2(Y-\hat{l})^2] + \mathbb{E}[\|\nabla_e h\|^2(T-\hat{e})^2]$.

Fix $h \in \mathcal{H}$ and write $\psi_{\text{raw}} = Y - h(\hat{\eta})$. Since $\eta_0 = \hat{\eta} - \delta\eta$, expanding h around $\hat{\eta}$ inside the loss,

$$\mathcal{L}(h(\eta_0)) = \mathbb{E}\Big[\big(\psi_{\text{raw}} + \nabla_\eta h(\hat{\eta})[\delta\eta] - r_h(\delta\eta)\big)^2\Big] = \mathcal{L}(h(\hat{\eta})) + 2\,\mathbb{E}\big[\psi_{\text{raw}}\,\nabla_\eta h(\hat{\eta})[\delta\eta]\big] + \mathrm{Rem},$$

where $\nabla_\eta h(\hat{\eta})[\delta\eta] = \nabla_l h\,\Delta l + \nabla_e h\,\Delta e$, $r_h(\delta\eta) = O(\|\delta\eta\|^2)$ uniformly over $\mathcal{H}$ by the bounded second derivatives, and Rem collects the squared first-order term and all terms containing r_h. Each of these is bounded by Cauchy–Schwarz, M, and $\mathrm{Var}(Y \mid X,T) \le \sigma_{\max}^2$, so $|\mathrm{Rem}| \le 2C_0\|\delta\eta\|^2$ with C_0 depending only on M and $\sigma_{\max}^2$.

For the first-order term, Cauchy–Schwarz gives

$$\big|2\,\mathbb{E}[\psi_{\text{raw}}\,\nabla_\eta h(\hat{\eta})[\delta\eta]]\big| \le 2\Big(\mathbb{E}\big[\psi_{\text{raw}}^2\big(\|\nabla_l h\|^2 + \|\nabla_e h\|^2\big)\big]\Big)^{1/2}\|\delta\eta\|.$$

Conditionally on (X,T), $\psi_{\text{raw}} = (f_0 - h(\hat{\eta})) + \xi_y$ with $|f_0 - h(\hat{\eta})| \le 2M$, so $\mathbb{E}[\psi_{\text{raw}}^2 \mid X,T] \le 4M^2 + \sigma_{\max}^2$, while $\mathbb{E}[(Y-\hat{l})^2 \mid X,T] \ge \mathrm{Var}(Y \mid X,T) \ge \sigma_{\min}^2$ and $(T-\hat{e})^2 \ge \alpha^2$ by positivity. Hence, with $\kappa_0 = (4M^2 + \sigma_{\max}^2)/\min\{\sigma_{\min}^2, \alpha^2\}$,

$$\mathbb{E}[\psi_{\text{raw}}^2 \mid X,T] \le \kappa_0\,\mathbb{E}[(Y-\hat{l})^2 \mid X,T], \qquad \mathbb{E}[\psi_{\text{raw}}^2 \mid X,T] \le \kappa_0\,(T-\hat{e})^2.$$

Because $\nabla_l h$ and $\nabla_e h$ are functions of (X,T), conditioning on (X,T) yields $\mathbb{E}[\psi_{\text{raw}}^2\|\nabla_l h\|^2] \le \kappa_0\,\mathbb{E}[\|\nabla_l h\|^2(Y-\hat{l})^2]$ and likewise for the e-term, so the first-order term is at most $2\sqrt{\kappa_0\,\Omega(h;\hat{\eta})}\,\|\delta\eta\|$. Therefore, for every $h \in \mathcal{H}$,

$$\big|\mathcal{L}(h(\eta_0)) - \mathcal{L}(h(\hat{\eta}))\big| \le 2\sqrt{\kappa_0\,\Omega(h;\hat{\eta})}\,\|\delta\eta\| + 2C_0\|\delta\eta\|^2. \tag{33}$$

Apply equation 33 to h^* and use $2\sqrt{ab} \le a + b$ with $a = \Omega(h^*; \hat{\eta})$, $b = \kappa_0 \|\delta\eta\|^2$:

$$\mathcal{L}(h^*(\eta_0)) \le \mathcal{L}(h^*(\hat{\eta})) + \Omega(h^*; \hat{\eta}) + (\kappa_0 + 2C_0)\|\delta\eta\|^2 = \mathcal{L}_{\text{GRL-OP}}(h^*; \hat{\eta}) + (\kappa_0 + 2C_0)\|\delta\eta\|^2.$$

By minimality of h^*, $\mathcal{L}_{\text{GRL-OP}}(h^*; \hat{\eta}) \le \mathcal{L}_{\text{GRL-OP}}(h_0; \hat{\eta}) = \mathcal{L}(h_0(\hat{\eta})) + \Omega(h_0; \hat{\eta})$, and by equation 33 applied to h_0, $\mathcal{L}(h_0(\hat{\eta})) \le \mathcal{L}(h_0(\eta_0)) + 2\sqrt{\kappa_0\, \Omega(h_0; \hat{\eta})}\, \|\delta\eta\| + 2C_0\|\delta\eta\|^2$. Combining the three inequalities,

$$\mathcal{L}(h^*(\eta_0)) - \mathcal{L}(h_0(\eta_0)) \le \Omega(h_0; \hat{\eta}) + \sqrt{4\kappa_0\, \Omega(h_0; \hat{\eta})}\, \|\delta\eta\| + (\kappa_0 + 4C_0)\|\delta\eta\|^2,$$

which gives the claim with $\kappa = 4\kappa_0$ and $C = \kappa_0 + 4C_0$. □

C.8 Estimation Complexity of Partially-Linear and Non-Linear Classes

We provide details for the estimation-error comparison. Recall the partially linear and nonlinear second-stage classes $\mathcal{F}_{\mathrm{PL}}$ and $\mathcal{F}_{\mathrm{NL}}$ defined in equation 9. Standard Rademacher generalization bounds imply $E_m = O_p(\mathfrak{R}_n(\mathcal{F}_m))$ under bounded outcomes and predictions (Bartlett & Mendelson, 2002).

For the partially linear class,

$$\hat{\mathfrak{R}}_S(\mathcal{F}_{\mathrm{PL}}) = \mathbb{E}_\sigma\left[\sup_{\tau\in\mathcal{T}} \frac{1}{n}\sum_{i=1}^{n} \sigma_i \left(\hat{l}(X_i) + (T_i - \hat{e}(X_i))\,\tau(X_i)\right)\right] \le \hat{\mathfrak{R}}_{S_X}(\mathcal{T}),$$

where $\sigma_1, \ldots, \sigma_n$ are independent Rademacher random variables. The fixed term $\hat{l}(X_i)$ vanishes under the Rademacher expectation, and the inequality follows from the contraction principle since $|T_i - \hat{e}(X_i)| \le 1$. Hence, the complexity of F_{PL} is controlled by that of the CATE class $\mathcal{T}$.

We now specialize to the feed-forward MLP classes used in our implementation. For a depth-D network with weight matrices $A_1, \ldots, A_D$, Bartlett et al. (2017) characterize its norm-based spectral complexity by

$$R_A = \left(\prod_{j=1}^{D} \rho_j \|A_j\|_\sigma\right)\left[\sum_{j=1}^{D}\left(\frac{\|A_j^\top - M_j^\top\|_{2,1}}{\|A_j\|_\sigma}\right)^{2/3}\right]^{3/2},$$

where ρ_j is the Lipschitz constant of the activation, $\|\cdot\|_\sigma$ denotes the spectral norm, $\|\cdot\|_{2,1}$ denotes the $(2,1)$ matrix norm, and M_j is a fixed reference matrix. Their covering-number bound, combined with Dudley's entropy integral, yields, up to logarithmic factors,

$$\mathfrak{R}_n(\mathcal{F}_{\mathrm{PL}}) = \widetilde{O}\left(\frac{R_{\mathrm{PL}} B_X}{\sqrt{n}}\right),$$

when $\|X\|_2 \le B_X$, where R_{PL} bounds the corresponding norm-based complexity of networks in $\mathcal{T}$.

For the nonlinear class, the network input is $(X, \hat{l}(X), T - \hat{e}(X))$. If $\|X\|_2 \le B_X$ and $|\hat{l}(X)| \le B_l$, then

$$\left\|\left(X, \hat{l}(X), T - \hat{e}(X)\right)\right\|_2^2 = \|X\|_2^2 + |\hat{l}(X)|^2 + |T - \hat{e}(X)|^2 \le B_X^2 + B_l^2 + 1.$$

Therefore,

$$\mathfrak{R}_n(\mathcal{F}_{\mathrm{NL}}) = \widetilde{O}\left(\frac{R_{\mathrm{NL}}\sqrt{B_X^2 + B_l^2 + 1}}{\sqrt{n}}\right),$$

where R_{NL} bounds the norm-based complexity of networks in $\mathcal{H}$.

To compare the two classes, consider MLPs with matched depth and common layerwise spectral- and $(2,1)$-norm constraints. Let R denote the resulting common upper bound on their norm-based network complexities. Then

$$E_{\mathrm{PL}} = \widetilde{O}_p\left(\frac{R B_X}{\sqrt{n}}\right), \qquad E_{\mathrm{NL}} = \widetilde{O}_p\left(\frac{R\sqrt{B_X^2 + B_l^2 + 1}}{\sqrt{n}}\right).$$

Thus, under a matched norm-complexity, both classes retain the same $n^{-1/2}$ estimation-error order, while the nonlinear upper bound increases through the augmented input radius.

As a concrete normalized setting, let $X \in \mathbb{R}^d$, where d is the covariate dimension, and suppose $|X_j| \le 1$ for $j = 1, \ldots, d$ and $|\hat{l}(X)| \le 1$. Then $\|X\|_2 \le \sqrt{d}$, while $|T - \hat{e}(X)| \le 1$ gives

$$\left\|\left(X, \hat{l}(X), T - \hat{e}(X)\right)\right\|_2 \le \sqrt{d+2}.$$

$$\mathfrak{R}_n(\mathcal{F}_{\mathrm{PL}}) = \widetilde{O}\left(R\sqrt{\frac{d}{n}}\right), \qquad \mathfrak{R}_n(\mathcal{F}_{\mathrm{NL}}) = \widetilde{O}\left(R\sqrt{\frac{d+2}{n}}\right),$$

Therefore,

$$E_{\mathrm{PL}} = \widetilde{O}_p\left(R\sqrt{\frac{d}{n}}\right), \qquad E_{\mathrm{NL}} = \widetilde{O}_p\left(R\sqrt{\frac{d+2}{n}}\right).$$

The ratio of these derived upper-bound scales is therefore $\sqrt{(d+2)/d}$. Thus, under the stated matched-complexity conditions, moving from the partially linear to the nonlinear MLP class preserves the order of the second-stage estimation error in n, while incurring only the additional complexity associated with the augmented nuisance-dependent inputs.

D Dataset Details

D.1 Synthetic data

For evaluation with controlled ground truth, we utilize real-world covariates $\mathbf{x} \in \mathbb{R}^d$ from the ACIC2016 competition to preserve realistic feature distributions and correlations, while simulating treatment assignment and outcomes via known functions.

We generate treatment using a mixed linear and nonlinear propensity model. Specifically, we sample weight matrices $\mathbf{W}_p \in \mathbb{R}^{d\times 1}$ and $\mathbf{W}_i \in \mathbb{R}^{d\times 5}$ with entries drawn from $\mathcal{N}(0,1)$, and define the logit as $L(\mathbf{x}) = \mathbf{x}^\top \mathbf{W}_p + \sum_{k=1}^{5} \tanh(\mathbf{x}^\top \mathbf{W}_{i,k})$. The propensity score is then $e(\mathbf{x}) = \sigma(\lambda \cdot L(\mathbf{x}))$, where $\sigma(\cdot)$ is the logistic function and λ controls the sharpness of treatment assignment. Treatment is sampled as $T \sim \text{Bernoulli}(e(\mathbf{x}))$. We construct nonlinear outcome surfaces using both raw covariates and pairwise feature interactions. Let $\mathbf{X}_{\text{int}}$ denote all pairwise interactions of $\mathbf{x}$ and define the extended feature vector $\mathbf{x}_{\text{ext}} = [\mathbf{x}, \mathbf{X}_{\text{int}}]$. We sample weight matrices $\mathbf{W}_0 \in \mathbb{R}^{d\times 5}$ and $\mathbf{W}_1^{\text{ext}} \in \mathbb{R}^{(d+d_{\text{int}})\times 5}$ from $\mathcal{N}(0,1)$, where d_{int} is the number of interaction features. The base components are defined as $\ell_0(\mathbf{x}) = \log(1 + \exp(\mathbf{x}^\top \mathbf{W}_0))$ and $\ell_1(\mathbf{x}) = \mathbf{x}_{\text{ext}}^\top \mathbf{W}_1^{\text{ext}}$, and we define aggregated signals $g_0(\mathbf{x}) = \sum_{k=1}^{5} \ell_{0,k}(\mathbf{x})$ and $g_1(\mathbf{x}) = \sum_{k=1}^{5} \ell_{1,k}(\mathbf{x})$. The control outcome is standardized using training data statistics as $\mu_0(\mathbf{x}) = (g_0(\mathbf{x}) - \mathbb{E}_{\text{train}}[g_0])/(\text{Std}_{\text{train}}(g_0))$. We define the raw treatment effect as $\tau_{\text{raw}}(\mathbf{x}) = g_1(\mathbf{x}) - \mu_0(\mathbf{x})$, which induces heterogeneous and nonlinear effects. To control scale, we apply an affine transformation $\tau(\mathbf{x}) = s \cdot \tau_{\text{raw}}(\mathbf{x}) + b$, where s and b are chosen so that $\tau(\mathbf{x})$ achieves a target mean and standard deviation on the training set. The treated outcome is then defined as $\mu_1(\mathbf{x}) = \mu_0(\mathbf{x}) + \tau(\mathbf{x})$. Finally, observed outcomes are generated as $Y = T \cdot \mu_1(\mathbf{x}) + (1-T) \cdot \mu_0(\mathbf{x}) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ with small variance.

D.2 Semi-synthetic data

IHDP: IHDP (Hill, 2011) is a widely utilized benchmark dataset built upon real covariates from randomized control trial involving low-birth-weight infants. To mimic an observational study, bias was introduced by removing a non-random subset of treated individuals (individuals with non-white mothers), resulting in imbalanced dataset with incomplete overlap. We used the non-linear setup described as setup B in Hill (2011) where control units had exponential response surface and treated units had linear response surface. We used 100 realizations of this simulation setup provided by Shalit et al. (2017). Since the magnitude of CATE varied drastically across simulation runs, we applied the rescaling procedure similar to Curth & Van der Schaar (2021b), where for any run that had variance of CATE greater than one, the original outcome noise were extracted, followed by scaling of potential outcomes with standard deviation of CATE and adding back original noise.

Twins: Twins (Louizos et al., 2017) is a benchmark dataset derived from US twin birth registry for evaluating treatment effect estimation models. The heavier twin was considered as the treated unit and lighter twin as the control. The outcome was binary representing 1-year mortality. Since the outcomes for both twins were recorded, the unobserved counterfactual for any unit was approximated using observed outcome of their sibling. To simiulate observation study with selection bias, we generated synthetic treatment assignment based on the covariates similar to Yoon et al. (2018), where we sampled a weight vector $\mathbf{w}$ from a uniform distribution, $\mathbf{w} \sim \mathcal{U}(-0.1, 0.1)$. The treatment probability was then computed via the logistic function $e(\mathbf{x}) = \sigma(\mathbf{x}^T \mathbf{w})$, and the final treatment assignment was sampled as $T \sim \text{Bern}(e(\mathbf{x}))$.

D.3 Real data

Jobs: Jobs (LaLonde, 1986) is popularly used real-world benchmark dataset for causal inference, where the goal is to estimate the causal effect of job training on future employment status. The dataset was constructed by combining a randomized clinical trial from the National Supported Work (NSW) program with a large observational control group from the Panel Study of Income Dynamics (PSID) (Smith & Todd, 2005). Unlike synthetic and semi-synthetic benchmarks, individual ground truth treatment effects were unavailable for this real-world data. We thus assessed the model's decision-making ability by the policy risk metric on the randomized NSW subset.

E IMPLEMENTATION DETAILS

Fig. 3 illustrates the two variants we instantiated for GRL, without (a) and with (b) shared representations for the treatment outcome and propensity heads.

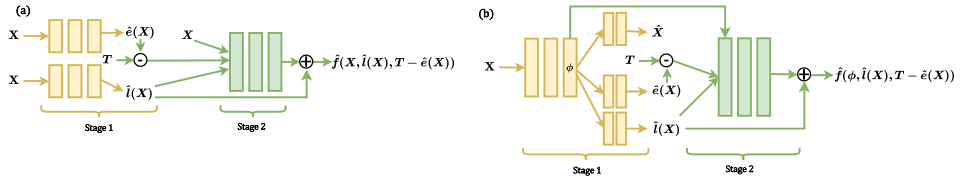


Figure 3: Instantiations of GRL without (a) and with (b) shared representations.

Similarly, Table 3 shows architectural details of GRL-Shared and GRL-Sep models.

Table 3: Architecture of GRL-Shared and GRL-Sep models. BN = BatchNorm, DO = Dropout.

Model	Stage	Component	Input	Layers	Units	Activation	BN/DO	Output
GRL-Shared	Stage 1	Shared representation	X	2	100	ELU	BN ✓ / DO ✓ (0.2)	$\Phi(X)$
		Reconstruction head	$\Phi(X)$	2	50	ELU	BN ✓ / DO ✓ (0.2)	$\hat{X}$
		Outcome head	$\Phi(X)$	1	50	ELU	BN ✓ / DO ✓ (0.2)	$\hat{l}(X)$
		Treatment head	$\Phi(X)$	1	50	ELU + Sigmoid	BN ✓ / DO ✓ (0.2)	$\hat{e}(X)$
	Stage 2	Target model	$[\Phi(X), \hat{l}(X), T - \hat{e}(X)]$	2	50	ELU	BN ✓ / DO ✓ (0.2)	$\hat{y}$ (or $\hat{y} + \hat{l}(X)$ if skip)
GRL-Sep	Stage 1	Outcome head	X	2	100	ELU	BN ✓ / DO ✓ (0.2)	$\hat{l}(X)$
		Treatment head	X	2	100	ELU + Sigmoid	BN ✓ / DO ✓ (0.2)	$\hat{e}(X)$
	Stage 2	Target model	$[X, \hat{l}(X), T - \hat{e}(X)]$	2	100	ELU	BN ✓ / DO ✓ (0.2)	$\hat{y}$ (or $\hat{y} + \hat{l}(X)$ if skip)

We utilized publicly available code, CATENets (Curth & Van der Schaar, 2021a;b) as the underlying backbone for our experiments. We used the PyTorch implementation provided for different baselines with same building blocks. Our proposed method was implemented within this same framework and use same blocks to ensure fair comparison. For baselines with specialized architecture that are not the part of this codebase, we brought models like TEDVAE, and MitNet into the same codebase using authors' official public code release, while for DiffPO we used their publicly released implementation. All models use the same early-stopping strategy based on each model's validation loss and were trained and evaluated on Nvidia GeForce RTX 4060 Ti 16 GB GPU.

F Additional Ablations Results

F.1 Ablations on Second-Stage Optimization Strategies

Figure 4A summarizes ablation results of the learning rates used for optimizing the raw loss and the orthogonal penalty in the second-stage optimization. As shown, the lowest PEHE was generally obtained when the learning rate for optimizing the raw loss was approximately one order of magnitude larger than that for optimizing the penalty.

Figure 4B summarizes the results when removing the nuisance outcome $\hat{l}(X)$ from the inputs to the second-stage outcome function. The deterioration of in both PEHE and outcome MSE metrics provided evidence that the second-stage model learned to use the first-stage nuisance outcome to focus on residual-on-residual learning as intended, instead of degenerating into a naive outcome learner that ignores first-stage nuisance estimates.

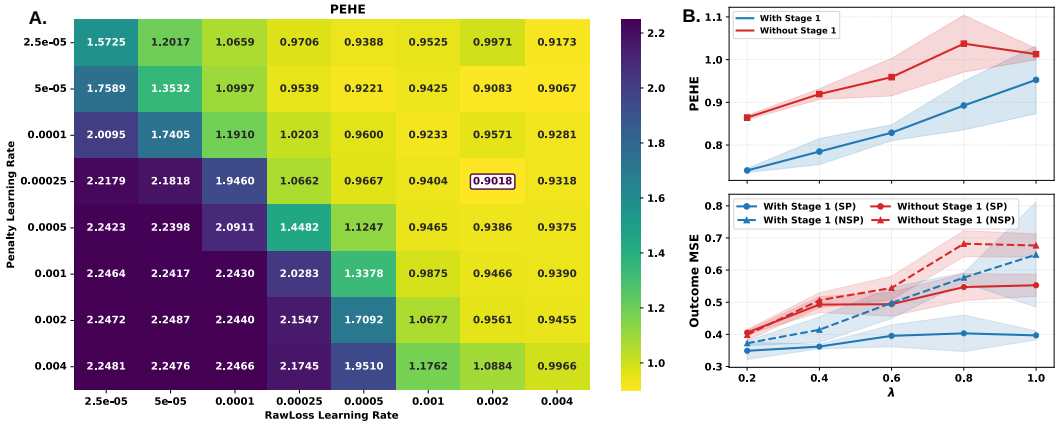


Figure 4: A. Heatmap showing the effect of Raw Loss and Orthogonaity Penalty learning rates on PEHE with box highlighting the lowest value. B. GRL with and without using first-stage nuisance prediction $\hat{l}(x)$ as input in the second stage. (N)SP: propensity-(non)-supported.

F.2 Ablations on First-Stage Optimization Strategies

Table 4 shows ablations on different optimization strategies used in first stage for nuisance learners with sequential indicating optimizing the $l(X)$ and $e(X)$ heads one after another and simultaneous indicating optimizing them together. The weighting column ablates the benefit of including the weighting when learning the $l(X)$ head as described in Section 5. The model column ablates the importance of optimizing the reconstruction objective prior to optimizing $l(X)$ and $e(X)$ heads.

Table 4: Ablation on first stage nuisance models with similar annotations used in Table 1

Model	Optimization	Weighting	PEHE ↓	MSE ↓
Reconstruction	Sequential	Yes	**0.8288 ± 0.0186***	**0.4467 ± 0.0414**
		No	0.9303 ± 0.0160	0.5082 ± 0.0211
	Simultaneous	Yes	1.1152 ± 0.0474	0.7538 ± 0.1412
		No	1.0935 ± 0.0442	0.7125 ± 0.1510
No-reconstruction	Simultaneous	Yes	0.9409 ± 0.0436	0.5368 ± 0.0575
		No	0.9387 ± 0.0337	0.5277 ± 0.0714
	Sequential	Yes	0.9105 ± 0.0484	0.4882 ± 0.0493
		No	0.9077 ± 0.0541	0.4883 ± 0.0477

G COMPLETE RESULTS ON SYNTHETIC DATASETS

Tables 5-7 list the complete results for PEHE, MSE of estimated outcome, and supported and non-supported MSE of estimated outcome, on the synthetic dataset across all values of λ.

Table 5: Performance comparison based on PEHE with similar annotations used in Table 1

Model	$\lambda = 0$	$\lambda = 0.2$	$\lambda = 0.4$	$\lambda = 0.6$	$\lambda = 0.8$	$\lambda = 1$
T-Learner	1.3957 ± 0.0139	1.4122 ± 0.0050	1.4119 ± 0.0061	1.423 ± 0.0075	1.4366 ± 0.0265	1.4525 ± 0.0195
TARNet	0.8047 ± 0.0538	1.4985 ± 0.4603	2.4057 ± 0.2372	2.6477 ± 0.2844	2.8091 ± 0.2195	2.7449 ± 0.189
CFR-MMD	0.9523 ± 0.1113	1.2350 ± 0.0738	1.6362 ± 0.2607	1.6983 ± 0.2351	1.8128 ± 0.2515	1.6955 ± 0.2117
SNet	0.7234 ± 0.0353	0.8905 ± 0.0356	1.6427 ± 0.2770	2.2525 ± 0.0878	2.6396 ± 0.0286	2.5200 ± 0.1215
FlexTENet	1.0478 ± 0.0057	1.0704 ± 0.0065	1.0724 ± 0.0095	1.0732 ± 0.0046	1.0838 ± 0.0120	1.0867 ± 0.0202
DiffPO	0.8749 ± 0.0924	1.0319 ± 0.1241	0.9613 ± 0.0806	1.5869 ± 0.064	1.3709 ± 0.1125	1.0879 ± 0.1135
TEDVAE	0.8457 ± 0.0094	0.8901 ± 0.018	0.9169 ± 0.0216	0.9348 ± 0.0179	0.9387 ± 0.0162	**0.9428 ± 0.0168**
MitNet	1.1989 ± 0.0280	1.2885 ± 0.0460	1.3839 ± 0.0056	1.3766 ± 0.0063	1.3867 ± 0.0041	1.4353 ± 0.0074
PW-Learner	1.6527 ± 0.0731	1.8970 ± 0.0874	2.1363 ± 0.1431	2.1529 ± 0.0232	2.194 ± 0.1246	2.1448 ± 0.0864
RA-Learner	1.3491 ± 0.0149	1.3426 ± 0.0110	1.3468 ± 0.0071	1.3474 ± 0.0085	1.3588 ± 0.0099	1.3673 ± 0.0072
X-Learner	1.3104 ± 0.0074	1.3273 ± 0.0086	1.3245 ± 0.0031	1.3245 ± 0.0015	1.3462 ± 0.0191	1.3563 ± 0.0059
DragonNet	0.8899 ± 0.0701	1.5666 ± 0.0733	2.4747 ± 0.0578	2.7284 ± 0.0347	2.8287 ± 0.0870	2.7931 ± 0.0754
DR-Learner	1.4451 ± 0.0220	1.4704 ± 0.0426	1.5207 ± 0.0548	1.4760 ± 0.0195	1.4937 ± 0.0573	1.4749 ± 0.0003
DR-Learner (clip)	1.4451 ± 0.0220	1.4589 ± 0.0223	1.4189 ± 0.0165	1.4428 ± 0.0086	1.4328 ± 0.0395	1.4822 ± 0.0597
RieszNet	0.8733 ± 0.0205	0.9011 ± 0.0211	0.9183 ± 0.0091	0.936 ± 0.0263	0.9518 ± 0.0264	0.9440 ± 0.0084
R-Learner	0.8465 ± 0.0217	0.9094 ± 0.0354	1.0195 ± 0.0477	1.1917 ± 0.1017	1.2129 ± 0.0955	1.283 ± 0.0583
GRL-Shared	**0.7035 ± 0.0310**	**0.7405 ± 0.0048**	**0.7847 ± 0.0306**	**0.8288 ± 0.0186**	0.8925 ± 0.0571	0.9526 ± 0.0794
GRL-Sep	0.7059 ± 0.0117	0.7577 ± 0.025	0.8168 ± 0.0115	0.8574 ± 0.0431	**0.8838 ± 0.0337**	0.9516 ± 0.0459

Table 6: Performance comparison based on outcome MSE with similar annotations used in Table 1

Model	$\lambda = 0.2$	$\lambda = 0.4$	$\lambda = 0.6$	$\lambda = 0.8$	$\lambda = 1.0$
T-Learner	3.0679 ± 0.0058	3.0778 ± 0.016	3.0674 ± 0.0288	3.0903 ± 0.0178	3.1215 ± 0.0267
TARNet	1.0366 ± 0.5543	2.2581 ± 0.426	2.6904 ± 0.5925	2.9601 ± 0.3986	2.8387 ± 0.4012
CFR-MMD	0.8299 ± 0.0752	1.3694 ± 0.374	1.4766 ± 0.3427	1.6083 ± 0.3636	1.4439 ± 0.3272
SNet	0.4158 ± 0.0153	1.1045 ± 0.3318	1.9784 ± 0.1303	2.6211 ± 0.0499	2.3969 ± 0.2416
FlexTENet	0.9656 ± 0.0208	0.9781 ± 0.0148	0.9758 ± 0.0111	0.9793 ± 0.0074	0.9751 ± 0.0071
DiffPO	0.8723 ± 0.1541	0.7277 ± 0.0929	1.6971 ± 0.1061	1.2873 ± 0.1838	0.9193 ± 0.1676
TEDVAE	0.5175 ± 0.0564	0.5343 ± 0.044	0.5494 ± 0.0485	0.5521 ± 0.0529	0.5564 ± 0.0572
MitNet	2.2385 ± 0.0170	2.3574 ± 0.0079	2.3061 ± 0.0031	2.3879 ± 0.0156	2.2919 ± 0.0005
DragonNet	1.0656 ± 0.1063	2.3455 ± 0.0779	2.8314 ± 0.1088	2.9986 ± 0.1782	2.9146 ± 0.1329
RieszNet	0.4572 ± 0.0367	0.4900 ± 0.0252	0.5027 ± 0.0451	0.5101 ± 0.0469	**0.5055 ± 0.0236**
R-Learner	0.9103 ± 0.0683	1.0392 ± 0.0795	1.0882 ± 0.0975	1.1965 ± 0.231	1.0732 ± 0.0953
GRL-Shared	0.3606 ± 0.0171	0.3882 ± 0.0166	0.4467 ± 0.0414	0.4898 ± 0.0344	0.5223 ± 0.0884
GRL-Sep	**0.3004 ± 0.0279***	**0.3488 ± 0.0009***	**0.3983 ± 0.0213**	**0.4578 ± 0.0377**	0.5057 ± 0.0594

Table 7: Outcome MSE breakdown based on propensity-supported and non-supported PO with similar annotations used in Table 1.

Model	$\lambda = 0.2$		$\lambda = 0.4$		$\lambda = 0.6$		$\lambda = 0.8$		$\lambda = 1.0$	
	$\text{MSE}_{\text{SP}} \downarrow$	$\text{MSE}_{\text{NSP}} \downarrow$	$\text{MSE}_{\text{SP}} \downarrow$	$\text{MSE}_{\text{NSP}} \downarrow$	$\text{MSE}_{\text{SP}} \downarrow$	$\text{MSE}_{\text{NSP}} \downarrow$	$\text{MSE}_{\text{SP}} \downarrow$	$\text{MSE}_{\text{NSP}} \downarrow$	$\text{MSE}_{\text{SP}} \downarrow$	$\text{MSE}_{\text{NSP}} \downarrow$
T-learner	3.3064 ± 0.0183	2.8293 ± 0.0265	3.3242 ± 0.0242	2.8314 ± 0.0080	3.2881 ± 0.0616	2.8468 ± 0.0051	3.3038 ± 0.0398	2.8768 ± 0.0136	3.3383 ± 0.0463	2.9048 ± 0.0164
TARNet	0.4739 ± 0.1016	1.5993 ± 1.0203	0.4545 ± 0.0852	4.0617 ± 0.8175	0.3772 ± 0.0424	5.0035 ± 1.1636	0.3792 ± 0.0462	5.5411 ± 0.8397	0.3661 ± 0.0434	5.3113 ± 0.8454
CFR-MMD	0.7765 ± 0.0911	0.8833 ± 0.0873	1.0817 ± 0.2041	1.6570 ± 0.5586	1.0939 ± 0.1886	1.8593 ± 0.4971	1.1186 ± 0.0808	2.0981 ± 0.6467	1.0152 ± 0.1437	1.8726 ± 0.5120
SNet	0.3611 ± 0.0279	0.4705 ± 0.0330	**0.3353 ± 0.0138**	1.8738 ± 0.6576	**0.3343 ± 0.0390**	3.6225 ± 0.2860	**0.3259 ± 0.0372**	4.9164 ± 0.0960	**0.3222 ± 0.0206**	4.4716 ± 0.4712
FlexTENet	0.9776 ± 0.0276	0.9535 ± 0.0146	0.9867 ± 0.0208	0.9696 ± 0.0130	0.9770 ± 0.0129	0.9746 ± 0.0132	0.9839 ± 0.0174	0.9748 ± 0.0088	0.9699 ± 0.0166	0.9803 ± 0.0088
DiffPO	0.7870 ± 0.1111	0.9577 ± 0.1971	0.5680 ± 0.0324	0.8874 ± 0.1534	0.9235 ± 0.0316	2.4707 ± 0.1806	0.6705 ± 0.0709	1.9041 ± 0.2967	0.5827 ± 0.0310	1.2559 ± 0.3042
TEDVAE	0.5481 ± 0.0543	0.4869 ± 0.0590	0.5516 ± 0.0361	0.5169 ± 0.0523	0.5541 ± 0.0331	0.5447 ± 0.0641	0.5511 ± 0.0358	0.5531 ± 0.0700	0.5502 ± 0.0377	0.5626 ± 0.0768
MitNet	2.3854 ± 0.0275	2.0917 ± 0.0065	2.4601 ± 0.0063	2.2548 ± 0.0095	2.3784 ± 0.0006	2.2337 ± 0.0064	2.4736 ± 0.0187	2.3022 ± 0.0131	2.2951 ± 0.0034	2.2886 ± 0.0043
DragonNet	0.5279 ± 0.1054	1.6033 ± 0.1089	0.4706 ± 0.0393	4.2205 ± 0.1902	0.4116 ± 0.0752	5.2513 ± 0.1430	0.4327 ± 0.0350	5.5644 ± 0.3229	0.3934 ± 0.0376	5.4357 ± 0.2286
RieszNet	0.4373 ± 0.0389	0.4771 ± 0.0346	0.4517 ± 0.0278	0.5283 ± 0.0241	0.4433 ± 0.0426	0.5621 ± 0.0476	0.4339 ± 0.0425	0.5863 ± 0.0516	0.4255 ± 0.0201	0.5854 ± 0.0273
R-Learner	0.8384 ± 0.0251	0.9823 ± 0.1227	0.9575 ± 0.0167	1.1210 ± 0.1507	0.8341 ± 0.1099	1.3424 ± 0.0854	0.8237 ± 0.1714	1.5694 ± 0.3248	0.8263 ± 0.1039	1.3201 ± 0.0918
GRL-Shared	0.3488 ± 0.0269	0.3725 ± 0.0074	0.3620 ± 0.0071	0.4144 ± 0.0398	0.3954 ± 0.0341	0.4980 ± 0.0487	0.4032 ± 0.0568	0.5765 ± 0.0151	0.3969 ± 0.0139	0.6478 ± 0.1636
GRL-Sep	**0.3046 ± 0.0330**	**0.2961 ± 0.0240***	0.3387 ± 0.0087	**0.3588 ± 0.0076***	0.3894 ± 0.0269	**0.4072 ± 0.0158***	0.4355 ± 0.0394	**0.4801 ± 0.0403**	0.4536 ± 0.0396	**0.5579 ± 0.0794**