Effect-Invariant Mechanisms for Policy Generalization

Sorawit Saengkyongam[*]   SORAWIT.SAENGKYONGAM@STAT.MATH.ETHZ.CH
Seminar for Statistics
ETH Zürich
Zürich, Switzerland

Niklas Pfister   NP@MATH.KU.DK
Department of Mathematical Sciences
University of Copenhagen
Copenhagen, Denmark

Predrag Klasnja   KLASNJA@UMICH.EDU
School of Information
University of Michigan
Ann Arbor, MI, USA

Susan Murphy   SAMURPHY@G.HARVARD.EDU
Department of Statistics
Department of Computer Science
Harvard University
Cambridge, MA, USA

Jonas Peters[*]   JONAS.PETERS@STAT.MATH.ETHZ.CH
Seminar for Statistics
ETH Zürich
Zürich, Switzerland

Abstract

Policy learning is an important component of many real-world learning systems. A major challenge in policy learning is how to adapt efficiently to unseen environments or tasks. Recently, it has been suggested to exploit invariant conditional distributions to learn models that generalize better to unseen environments. However, assuming invariance of entire conditional distributions (which we call full invariance) may be too strong of an assumption in practice. In this paper, we introduce a relaxation of full invariance called effect-invariance (e-invariance for short) and prove that it is sufficient, under suitable assumptions, for zero-shot policy generalization. We also discuss an extension that exploits e-invariance when we have a small sample from the test environment, enabling few-shot policy generalization. Our work does not assume an underlying causal graph or that the data are generated by a structural causal model; instead, we develop testing procedures to test e-invariance directly from data. We present empirical results using simulated data and a mobile health intervention dataset to demonstrate the effectiveness of our approach.

Keywords: distribution generalization, policy learning, invariance, causality, domain adaptation

---

[*]. Part of this work was done while SS and JP were at the University of Copenhagen.

1. Introduction

When learning models from data, we often use these models in scenarios that are assumed to have similar or the same characteristics as the ones generating the training data. This holds for prediction tasks such as regression and classification but also for settings such as contextual bandits or dynamic treatment regimes. When we observe different regimes under training, we can hope to exploit this information to construct models that adapt better to an unseen environment (or task). Such problems are usually referred to as multi-task learning domain adaptation or domain generalization (Caruana, 1997; Crammer et al., 2008; Muandet et al., 2013; Wang et al., 2022); the nomenclature sometimes differs depending on whether one observes labeled and/or unlabeled data in the test domain. For prediction tasks, it has been suggested to learn invariant models by exploiting invariance of the conditional distributions. Under suitable assumptions, such models generalize better to unseen environments if the changes between the environments can be modeled by interventions (e.g., Rojas-Carulla et al., 2018; Magliacane et al., 2018; Christiansen et al., 2021). A similar approach has been applied in policy learning (Saengkyongam et al., 2023), where one searches for policies that yield an invariant reward distribution. We refer to the invariance of conditional distributions as 'full invariance'. More precisely, given covariates X_e and outcome Y_e from different environments $e \in \mathcal{E}$, the full invariance assumption posits the existence of a set of covariates X_e^S such that

$$\text{for all } e_1, e_2 \in \mathcal{E} : Y_{e_1}|X_{e_1}^S \text{ and } Y_{e_2}|X_{e_2}^S \text{ are identical.} \tag{1}$$
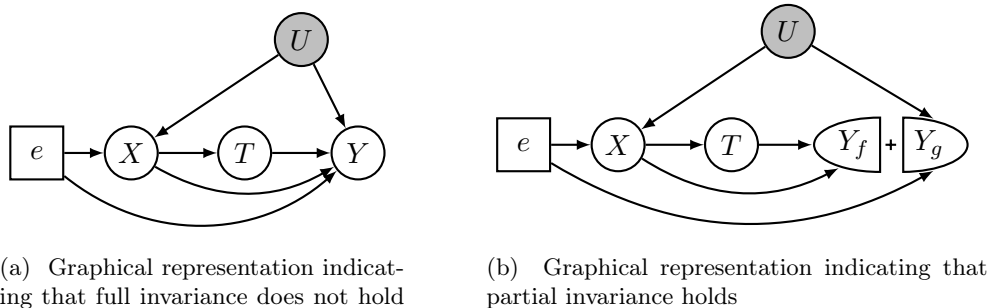
Full invariance, however, may be too strong of an assumption in practice. In prediction tasks, it has been suggested to relax the requirement of full invariance, such as vanishing empirical covariance, and instead use invariance as a form of regularization (e.g., Rothenhäusler et al., 2021; Jakobsen and Peters, 2022; Arjovsky et al., 2019). This approach comes with theoretical guarantees regarding generalization to bounded interventions, for example, but these results are often limited to restricted classes of models and interventions.

In this paper, we relax the full invariance assumption in a different direction and show how it can be applied to inferring optimal conditional treatments in policy learning. We illustrate our proposed relaxation based on an example. Consider the following class of structural causal models (SCMs, Pearl, 2009) indexed by environments $e \in \mathcal{E} := \{1, -1\}$, with the corresponding graph[1] shown in Figure 1,

$$\mathcal{S}(e): \quad \begin{cases} U := \epsilon_U \\ X := eU + \epsilon_X \\ T := \mathbb{1}(1 + X + \epsilon_T > 0) \\ Y := \underbrace{T(1 + X) + X}_{Y_f} + \underbrace{2e + U + \epsilon_Y}_{Y_g}, \end{cases} \tag{2}$$

where $(\epsilon_U, \epsilon_X, \epsilon_T, \epsilon_Y)$ are independent standard normal random variables. Here, $Y \in \mathbb{R}$ represents the outcome or reward, $T \in \{0, 1\}$ corresponds to the treatment or action, and $X \in \mathbb{R}$ and $U \in \mathbb{R}$ are observed and unobserved covariates, respectively. The mechanism

---

1. Visually, the graphical representation in Figure 1b is similar to SWIGs (Richardson and Robins, 2013), and we use 'tikz-swigs' LaTeX package for drawing the graph; the interpretation, however, is different.

(a) Graphical representation indicating that full invariance does not hold

(b) Graphical representation indicating that partial invariance holds

Hypothesis	P-value
Full-invariance, see (1)	0.01
Effect-invariance, see (5)	0.64

(c) Comparing invariance tests

Figure 1: Y is the outcome (or reward), X and U are observed and unobserved context variables, T is the treatment (or action), and e represents different environments. In example (2), the outcome mechanism is generally not invariant (as the environment enters Y directly), see (a). This paper introduces a type of partial invariance called e-invariance (Definition 3), which does hold here, see (b): when conditioning on X, the treatment effect is invariant across environments. The concepts of the paper are applicable even if the data generating process does not allow for a graphical representation. Instead, we propose testing procedures to test for e-invariance. (Bottom) Comparing the test result obtained from one of our proposed e-invariance tests, applied to a sample taken from (2), with the result of the full invariance test. While X does not satisfy the full invariance condition (as in (1)), it does satisfy the e-invariance condition (as in (5)).

for T can be considered as a fixed policy. Since the environment has a direct effect on the outcome, there is no subset satisfying the full invariance condition (1): regardless of whether we condition on $\varnothing$ or $\{X\}$, the outcome distribution is not independent of the environment. Consequently, methods that rely on the full invariance assumption such as the one proposed by Saengkyongam et al. (2023) would lead to a vacuous result.

However, the criterion of full invariance is not necessary when the goal is to learn an optimal policy. Instead it may suffice to find models that are *partially invariant*: In the above example, see (2), the outcome Y can be additively decomposed into two components: one being a function of U, e, and ϵ_Y, and another being a function of T and X. In this case, although the outcome mechanism is not entirely invariant, it contains an invariant component. When conditioning on X, the effect of the treatment is the same in all environments. More specifically, the conditional average treatment effect does not depend on e, that is,[2]

$$\forall x \in \mathcal{X} : \mathbb{E}^e[Y \mid X = x, T = 1] - \mathbb{E}^e[Y \mid X, T = 0] = 1 + x. \tag{3}$$

---

2. Using do-notation (Pearl, 2009), we have for all $x \in \mathcal{X}$ and $t \in \mathcal{T}$ that $\mathbb{E}^e[Y \mid X = x, \mathrm{do}(T = t)] = \mathbb{E}^e[Y \mid X = x, T = t]$.

We say that $\{X\}$ satisfies *effect-invariance (e-invariance)*. This condition suffices that, for an unseen test environment, we can still infer the optimal treatment among policies that only depend on X without having access to the outcome information in the test environment. In addition, if the environments are heterogeneous enough (see Assumption 3), such a policy is worst-case optimal. We refer to this setup as zero-shot generalization. We state the class of data generating processes and provide formal results in Section 2 and Section 3 below.

Moreover, if we can acquire a small sample – including observations of the outcome – from the test environment, we would want to optimize the policy using the data from the test environment. Ideally, this optimization also leverages information from training data from other environments to improve the finite sample performance of the learnt policy. We discuss that e-invariant information can be beneficial in such settings. We refer to this scenario as few-shot generalization and present it as an extension of the zero-shot methodology, in Section 5.

While SCMs provide a class of examples satisfying the assumptions of this work, we do not assume an underlying causal graph or SCM (but instead only require a sequential sampling procedure that ensures that the covariates X causally precede the outcome). In particular, e-invariance is not read off from a known graph but instead tested from data. Figure 1c illustrates the testing result obtained by applying one of the proposed e-invariance tests to a sample from (2), where we also include a comparison with the full-invariance test as proposed in (Peters et al., 2016, Method II).

The main contributions of this paper are four fold:

(1) Introducing e-invariance: In Section 2, we introduce the concept of e-invariance, which offers a relaxation of the full invariance assumption. An e-invariant set ensures that the conditional treatment effect function remains the same across different environments.

(2) Utilizing e-invariance for generalization: Section 3 discusses the use of e-invariance in learning policies that provably generalize well to unseen environments. We prove two generalization guarantees: The proposed method (i) outperforms an optimal context-free policy on new environments and (ii) outperforms any other policy in terms of worst-case performance.

(3) Methods for testing e-invariance: We propose hypothesis testing procedures, presented in Section 4, to test for e-invariance from data within both linear and nonlinear model classes.

(4) Semi-real-world case study: In Section 6, we demonstrate the effectiveness of our proposed policy learning methods in the semi-real-world case study of mobile health interventions. An optimal policy based on an e-invariance set is shown to generalize better to new environments than the policy that uses all the context information.

1.1 Further Related Work

Our work builds upon the existing research that leverages the invariance of conditional distributions (full invariance) for generalization to unseen environments (Schölkopf et al., 2012; Rojas-Carulla et al., 2018; Magliacane et al., 2018; Arjovsky et al., 2019; Christiansen et al., 2021; Saengkyongam et al., 2023). Several relaxations of the full invariance have been

suggested for the prediction tasks (Rothenhäusler et al., 2021; Jakobsen and Peters, 2022; Arjovsky et al., 2019; Guo and Bühlmann, 2022). In reinforcement learning, previous studies have suggested the use of invariance to achieve generalizable policies (Zhang et al., 2020; Sonar et al., 2021), however, they lack theoretical guarantees for generalization. Closely related to our work, Saengkyongam et al. (2023) has established the worst-case optimality of invariant policy learning based on the full invariance assumption, which may be too restrictive in practice.

Transportability in causal inference (e.g., Pearl and Bareinboim, 2011; Bareinboim and Pearl, 2014; Subbaswamy et al., 2019) addresses the task of identifying invariant distributions based on a known causal graph and structural differences between environments, which can be used to generalize causal findings. However, our approach differs in that we do not assume prior knowledge of the causal graph or structural differences between environments. Furthermore, our methods are applicable even if the data generating process does not allow for a graphical representation. Instead, we develop testing procedures to obtain invariant information from data. Additionally, methods based on causal graphs typically only capture full invariance information (through the Markov property), whereas our work relaxes the requirement of full invariance for policy learning.

2. Effect-invariance

2.1 Multi-environment policy learning

In this work, we consider the problem of multi-environment policy learning (or multi-environment contextual bandit) (see also Dawid, 2021; Saengkyongam et al., 2023). Given a fixed set of environments $\mathcal{E}$, we assume that for each environment $e \in \mathcal{E}$, there is a policy learning setup, where the distributions of covariates and outcome may differ between environments. Each of the setups is modelled by a three-step sequential sampling scheme: First, covariates (X, U) are sampled according to a fixed distribution depending on the environment, then X is revealed to an agent that uses it to select a treatment T (from a finite set $\mathcal{T}$) according to a policy π and, finally, an outcome Y is sampled conditionally on X, U and T. Formally, we assume the following setting throughout the paper.

Setting 1 (Multi-environment policy learning) *Let $\mathcal{E} \subset \mathbb{R}$ be a collection of environments, $Y \in \mathbb{R}$ an outcome variable, $X \in \mathcal{X} \subseteq \mathbb{R}^d$ observed covariates, $U \in \mathcal{U} \subseteq \mathbb{R}^p$ unobserved covariates and $T \in \mathcal{T} = \{1, \ldots, k\}$ a treatment. Let $\Delta(\mathcal{T})$ denote the probability simplex over the set of treatments $\mathcal{T}$ and let $\Pi := \{\pi \mid \pi : \mathcal{X} \to \Delta(\mathcal{T})\}$ denote the set of all policies. Moreover, for all $e \in \mathcal{E}$ let $\mathbb{P}^e_{X,U}$ be a distribution on $\mathcal{X} \times \mathcal{U}$ and for all $e \in \mathcal{E}$, $x \in \mathcal{X}$, $u \in \mathcal{U}$ and $t \in \mathcal{T}$ let $\mathbb{P}^e_{Y|X=x,U=u,T=t}$ be a distribution on $\mathbb{R}$. Given $e \in \mathcal{E}$ and $\pi \in \Pi$, this defines a random vector (Y, X, U, T) by $(X, U) \sim \mathbb{P}^e_{X,U}$, $T \sim \pi(X)$, and $Y \sim \mathbb{P}^e_{Y|X=X,U=U,T=T}$, see Figure 1(a) for an example. Correspondingly, n observations $(Y_i, X_i, T_i, e_i, \pi_i)_{i=1}^n$ from this model are generated by the following steps.*

(i) Select an environment $e_i \in \mathcal{E}$ and a policy $\pi_i \in \Pi$.

(ii) Sample covariates $(X_i, U_i) \sim \mathbb{P}^{e_i}_{X,U}$.

(iii) Sample the treatment $T_i \sim \pi_i(X_i)$.

(iv) Sample the outcome $Y_i \sim \mathbb{P}^{e_i}_{Y|X=X_i,U=U_i,T=T_i}$.[3]

The sampling in (ii)–(iv) is done independently for different i. In particular, we consider e_i and π_i to be deterministic; they are not random variables that depend on other variables. (Our results in Section 3 remain valid even if π_i depends on previous observations $\{j : j \leq i\}$, see Remark 2.)

Further denote by $\mathcal{E}^{\mathrm{tr}} \subseteq \mathcal{E}$ the set of observed environments within the n training observations and for each $e \in \mathcal{E}^{\mathrm{tr}}$ we denote by n_e the number of observations from environment e. We assume that there exists a product measure ν such that for all $e \in \mathcal{E}$ the joint distribution of (Y, X, U, T) in environment e, under policy π has density $p^{e,\pi}$ with respect to ν and that $\mathbb{P}^e_X$ has full support on $\mathcal{X}$. Next, we define $t_0 \in \mathcal{T}$ as a baseline treatment, which serves as the reference point for defining the conditional average treatment effect in (4). However, and importantly, our results hold for any choice of t_0. Finally, we assume that the policies generating the training observations are bounded, i.e., for all $i \in \{1,\ldots,n\}, t \in \mathcal{T}$, and $x \in \mathcal{X}$ it holds that $\pi_i(x)(t) > 0$.

Notation When writing probabilities and expectations of the random variables Y, X, U and T or the corresponding observations, we use superscripts to make explicit any possible dependence on the environment and policy, e.g., $\mathbb{P}^{e,\pi}$ and $\mathbb{E}^{e,\pi}$. Moreover, by a slight abuse of notation, for a policy $\pi \in \Pi$ with a density, we let $\pi(x)$ denote the density rather than the distribution; we also use the commonly employed convention $\pi(t|x) := \pi(x)(t)$. Finally, for all $t \in \mathcal{T}$, we denote by $\pi_t \in \Pi$ the policy that always selects treatment t, that is, $\pi_t(\cdot|x) = \mathbb{1}(t = \cdot)$. If we assume the existence of an underlying SCM (see Section 2.3), we have for all $e \in \mathcal{E}$, $t \in \mathcal{T}$ and $x \in \mathcal{X}$ that $\mathbb{E}^{e,\pi_t}[Y \mid X = x] = \mathbb{E}^e[Y \mid X = x, \mathrm{do}(T = t)]$.

Remark 2 *Our results in Section 3 remain valid even if π_i in Setting 1 depends on previous observations. In this case, the sampling step (iii) is replaced by $T_i \sim \pi_i(X_i, H_i)$ with $H_i := \{(X_j, Y_j, T_j) : j < i\}$. Furthermore, in the Zero-shot setting in Section 3, we consider $D^{\mathrm{tr}} := (y_i^{\mathrm{tr}}, x_i^{\mathrm{tr}}, t_i^{\mathrm{tr}}, \pi_i^{\mathrm{tr}}(\cdot|\cdot, h_i), e_i^{\mathrm{tr}})_{i=1}^n$, where $(y_1^{\mathrm{tr}}, x_1^{\mathrm{tr}}, t_1^{\mathrm{tr}}), \ldots, (y_n^{\mathrm{tr}}, x_n^{\mathrm{tr}}, t_n^{\mathrm{tr}})$ are (jointly independent) realizations from $Q_1^{\mathrm{tr}} := \mathbb{P}^{e_1,\pi_1(\cdot|\cdot)}_{X,Y,T}, Q_2^{\mathrm{tr}} := \mathbb{P}^{e_2,\pi_2(\cdot|\cdot,h_2)}_{X,Y,T}, \ldots, Q_n^{\mathrm{tr}} := \mathbb{P}^{e_n,\pi_n(\cdot|\cdot,h_n)}_{X,Y,T}$ respectively, with $h_i := \{(y_j^{\mathrm{tr}}, x_j^{\mathrm{tr}}, t_j^{\mathrm{tr}}) : j < i\}$ for $i \geq 2$; in Appendix A.3, we replace π_i by $\pi_i(\cdot|\cdot, h_i)$.*

2.2 Invariant treatment effects

The concept of invariance has been connected to causality (Haavelmo, 1944; Pearl, 2009; Schölkopf et al., 2012) and it has been suggested to use it for causal discovery (Peters et al., 2016; Pfister et al., 2018; Heinze-Deml et al., 2018) or distribution generalization (Rojas-Carulla et al., 2018; Rothenhäusler et al., 2021; Magliacane et al., 2018). In our setting, the standard notion of invariance would correspond to the invariance in the outcome mechanism (Saengkyongam et al., 2023). In practice, this notion may be too strong. E.g., it does not hold if the environment directly influences the outcome (see Figure 1 for an example). In what follows, we introduce the notion of (treatment) effect-invariance, which relaxes the standard invariance condition.

---

3. Consequently, the distribution in (iv) is indeed the conditional distribution of Y, given X, U, and T, justifying the notation.

To this end, we recall the notion of the conditional average treatment effect (CATE) under different environments $e \in \mathcal{E}$. The CATE in environment $e \in \mathcal{E}$ for a subset of covariates $S \subseteq \{1, \ldots, d\}$ is defined for all $x \in \mathcal{X}^S$ and $t \in \mathcal{T}$ as

$$\tau_e^S(x, t) := \mathbb{E}^{e, \pi_t}[Y \mid X^S = x] - \mathbb{E}^{e, \pi_{t_0}}[Y \mid X^S = x]. \tag{4}$$

When $S = \{1, \ldots, d\}$, we simply denote τ_e^S by τ_e. In Setting 1, for some $S \subseteq \{1, \ldots, d\}$, the CATE functions, as defined in (4), may differ substantially from one environment to another. But there may exist a subset $S \subseteq \{1, \ldots, d\}$ such that the CATE functions do not change across environments. In this work, we exploit the existence of such sets, which we call *e-invariant* (for effect-invariant).[4]

Definition 3 (Effect-invariant sets) *Assume Setting 1. A subset $S \subseteq \{1, \ldots, d\}$ is said to be* effect-invariant *with respect to a set of environments $\mathcal{E}^\diamond \subseteq \mathcal{E}$ (e-invariant w.r.t. $\mathcal{E}^\diamond$ for short) if the following holds*

$$\forall e_1, e_2 \in \mathcal{E}^\diamond : \tau_{e_1}^S \equiv \tau_{e_2}^S. \tag{5}$$

For any $\mathcal{E}^\diamond \subseteq \mathcal{E}$, we denote by $\mathbf{S}_{\mathcal{E}^\diamond}^{\text{e-inv}}$ the collection of all e-invariant sets w.r.t. $\mathcal{E}^\diamond$.

The above definition does not depend on the choice of t_0 in Setting 1: if condition (5) holds for one choice of t_0, it holds for all other choices of $t_0 \in \mathcal{T}$. In this work, we focus on discrete treatments but, in principle, one could consider the continuous case by defining the CATE function as $(x, t) \mapsto \frac{\partial}{\partial t} \mathbb{E}^{e, \pi^t}[Y \mid X^s = x]$ and define the effect-invariance analogously to (5).

We now provide a characterization for e-invariance based on the outcome mechanism.

Proposition 4 *Assume Setting 1. A subset $S \subseteq \{1, \ldots, d\}$ is e-invariant w.r.t. $\mathcal{E}^\diamond$ if and only if there exists a pair of functions $\psi_S : \mathcal{X}^S \times \mathcal{T} \to \mathbb{R}$ and $\nu_S : \mathcal{X}^S \times \mathcal{E} \to \mathbb{R}$ such that*

$$\forall e \in \mathcal{E}^\diamond, \forall x \in \mathcal{X}^S, \forall t \in \mathcal{T} : \mathbb{E}^{e, \pi_t}[Y \mid X^S = x] = \psi_S(x, t) + \nu_S(x, e), \tag{6}$$

and $\psi_S(\cdot, t_0) \equiv 0$. In particular, we have for all $e \in \mathcal{E}^\diamond$ that $\psi_S \equiv \tau_e^S$.

Proof See Appendix A.1. ∎

The two equivalent conditions (5) and (6) provide two different viewpoints on e-invariant sets. The former shows that, when conditioning on an e-invariant set S, the CATE functions are invariant across environments, while the latter ensures that part of the conditional expected outcome $\mathbb{E}^{e, \pi_t}[Y \mid X^S]$ remains invariant across environments. In particular, the conditional expected outcome $\mathbb{E}^{e, \pi_t}[Y \mid X^S]$ can be additively decomposed into a fixed effect-modification term (ψ_S) that depends on the treatment and an environment-varying main-effect term (ν_S) that does not depend on the treatment. Here, the additivity stems from the definition of the CATE; different causal contrasts correspond to other forms of decomposition.

---

4. As an alternative to e-invariance, one could define *argmax-invariance* by requiring that $\forall e_1, e_2 \in \mathcal{E}^\diamond$: $\text{argmax}_{t \in \mathcal{T}} \tau_{e_1}^S(\cdot, t) \equiv \text{argmax}_{t \in \mathcal{T}} \tau_{e_2}^S(\cdot, t)$. A similar notion called 'invariant action prediction' has been introduced by Sonar et al. (2021). This condition would ensure that the optimal treatment is robust with respect to changes in the environment (even though the treatment effect may not be). E-invariance implies argmax-invariance but the latter condition is not sufficient to show generalization properties that we develop in Section 3.

In Section 3, we propose a method that utilizes e-invariant sets for zero-shot generalization. If there are multiple e-invariant sets, we choose the one that yields the highest expected outcome. Our results rely on the existence of an e-invariant set. We therefore make this assumption explicit.

Assumption 1 *In Setting 1, there exists a subset $S \subseteq \{1, \ldots, d\}$ such that S is e-invariant w.r.t. $\mathcal{E}$.*

The subsequent section connects Assumption 1 to a class of structural causal models (Pearl, 2009; Bongers et al., 2021; Dawid, 2021; Saengkyongam et al., 2023). For such models, proposition 6 below shows that Assumption 1 is satisfied if the outcome mechanism is of a specific form and an independence assumption holds. Furthermore, using a test for e-invariance, see Section 4, Assumption 1 is testable from data for the observed environments $\mathcal{E}^{\mathrm{tr}}$.

2.3 Effect-invariance in structural causal models

Assumption 1 is satisfied in a restricted class of structural causal models (SCMs). Formally, we consider the following class of SCMs inducing the sequential sampling steps (ii)–(iv) in Setting 1.

$$
\mathcal{S}(e, \pi): \quad
\begin{cases}
U := s_e(X, U, \epsilon_U) \\
X := h_e(X, U, \epsilon_X) \\
T := \ell_\pi(X, \epsilon_T) \\
Y := f(X^{\mathrm{PA}_{f,X}}, U^{\mathrm{PA}_{f,U}}, T) + g_e(X, U, \epsilon_Y),
\end{cases}
\tag{7}
$$

where $(U, X, T, Y) \in \mathcal{X} \times \mathcal{U} \times \mathcal{T} \times \mathbb{R}$, $(\epsilon_U, \epsilon_X, \epsilon_T, \epsilon_Y)$ are jointly independent noise variables, $(s_e, h_e, g_e)_{e \in \mathcal{E}}$, f and ℓ_π are measurable functions such that, for all $x \in \mathcal{X}$, $\ell_\pi(x, \epsilon_T)$ is a random variable on $\mathcal{T}$ with distribution $\pi(x)$, and $\mathrm{PA}_{f,X} \subseteq \{1, \ldots, d\}$ and $\mathrm{PA}_{f,U} \subseteq \{1, \ldots, p\}$. We call $\mathrm{PA}_{f,X}$ and $\mathrm{PA}_{f,U}$ the observed and unobserved policy-relevant parents, respectively.

To determine whether e-invariance holds, it is helpful to distinguish between the parents of Y that enter f (these are relevant to determine optimal policies) and those parents of Y that enter into g_e.[5] For building intuition, we therefore define a graphical representation, which splits Y into two nodes (as mentioned in Footnote 1, the graphical representation is similar to SWIGs (Richardson and Robins, 2013) but the interpretation is different).

Definition 5 (E-invariance graph) *We represent a class of SCMs of the form* (7) *by an e-invariance graph. This graph contains, as usually done when representing SCMs graphically, a directed edge from variables on the right-hand side of assignments to variables on the left-hand side, but with the exception that e is represented by a square node and the node Y is split into a part for Y_f and a part for Y_g; see Example 1 and also Figure 1.*

---

5. The set $\mathrm{PA}_{f,X}$ should be chosen as small as possible and, under this constraint, f and g_e should be chosen, such that the set of parents of Y entering g_e is as small as possible, too, see (see Bongers et al., 2021, Def. 2.6).

Example 1 *Consider the following SCMs*

$$
\mathcal{S}(e,\pi): \begin{cases}
U^1 := \epsilon_{U^1}, \quad U^2 := \epsilon_{U^2} \\
X^3 := \gamma_e^3 U^1 + \epsilon_{X^3} \\
X^2 := \gamma_e^2 U^2 + \epsilon_{X^2} \\
X^1 := X^2 + \gamma_e^1 U^1 + \epsilon_{X^1} \\
T := \ell_\pi(X^1, X^2, X^3, \epsilon_T) \\
Y := \underbrace{T(1 + 0.5X^2 + 0.5U^1) + X^2 + U^1}_{f} \\
\quad \underbrace{+ \mu_e + U^2 + X^3 + \epsilon_Y,}_{g_e}
\end{cases}
$$

where $\mathcal{T} = \{0,1\}$, $\epsilon_{U^1}, \epsilon_{U^2}, \epsilon_{X^1}, \epsilon_{X^2}, \epsilon_{X^3}, \epsilon_T, \epsilon_Y$ are jointly independent noise variables with mean zero, and $\gamma_e^1, \gamma_e^2, \gamma_e^3, \mu_e$ are environment-specific parameters. Here, $\mathrm{PA}_{f,X} = \{2\}$ and $\mathrm{PA}_{f,U} = \{1\}$ are the policy-relevant parents; the e-invariance graph is shown on the right. While in this example, the environment changes the coefficients $\gamma_e^1, \gamma_e^2, \gamma_e^3$ and μ_e, the generality of (7) allows for a change in the noise distributions, too.

Under the class of SCMs (7), the following proposition shows that an e-invariant set exists if the unobserved $U^{\mathrm{PA}_{f,U}}$ and the observed policy-relevant parents $X^{\mathrm{PA}_{f,X}}$ are independent, and the environments do not influence $U^{\mathrm{PA}_{f,U}}$.

Proposition 6 *Assume Setting 1 and that the sequential sampling steps (ii)–(iv) are induced by the SCMs in (7). If (i) for all $e \in \mathcal{E}^\diamond$, $U^{\mathrm{PA}_{f,U}} \perp\!\!\!\perp X^{\mathrm{PA}_{f,X}}$ in $\mathbb{P}_{X,U}^e$ and (ii) $\mathbb{P}_{U^{\mathrm{PA}_{f,U}}}^e$ are identical across $e \in \mathcal{E}^\diamond$, we have that*

$$
\mathrm{PA}_{f,X} \text{ is e-invariant w.r.t. } \mathcal{E}^\diamond. \tag{8}
$$

Proof See Appendix A.2. ∎

Example 7 (Example 1 continued) *Let $e \in \mathcal{E}$. In this example, it holds that $U^{\mathrm{PA}_{f,U}} \perp\!\!\!\perp X^{\mathrm{PA}_{f,X}}$ in $\mathbb{P}_{X,U}^e$, where $U^{\mathrm{PA}_{f,U}} = U^1$ and $X^{\mathrm{PA}_{f,X}} = X^2$. Therefore, $\mathrm{PA}_{f,X} = \{2\}$ satisfies the e-invariance condition (5) by Proposition 6. To illustrate this, consider the expected outcome conditioned on X^2,*

$$
\begin{aligned}
\mathbb{E}^{e,\pi_t}[Y \mid X^2] &= \mathbb{E}^{e,\pi_t}[T(1 + 0.5X^2 + 0.5U^1) + \mu_e + U^1 + U^2 + X^2 + X^3 \mid X^2] \\
&= \mathbb{1}(t = t_0)(1 + 0.5X^2 + 0.5\,\mathbb{E}^e[U^1 \mid X^2]) + \mu_e + X^2 + \mathbb{E}^e[U^1 + U^2 + X^3 \mid X^2] \\
&= \underbrace{\mathbb{1}(t = t_0)(1 + 0.5X^2) + X^2}_{\psi_{\{2\}}(X^2,t)} + \underbrace{\mu_e + \mathbb{E}^e[U^2 + X^3 \mid X^2]}_{\nu_{\{2\}}(X^2,e)} \quad \text{since } U^1 \perp\!\!\!\perp X^2.
\end{aligned}
$$

Thus, by Proposition 4, $\{2\}$ is e-invariant w.r.t. $\mathcal{E}$.

3. Zero-shot policy generalization through e-invariance

In this section, we consider zero-shot generalization (sometimes called unsupervised domain adaptation). We aim to find a policy that performs well (in terms of the expected outcome or reward) in a new test environment in which we have access to observations of the covariates but not the outcome. We formally lay out the setup and objective of zero-shot policy generalization and show that a policy that optimally uses information from e-invariant sets achieve desirable generalization properties.

Setting Zero-shot *Assume Setting 1 and that we are given $n \in \mathbb{N}$ training observations $D^{\mathrm{tr}} := (Y_i^{\mathrm{tr}}, X_i^{\mathrm{tr}}, T_i^{\mathrm{tr}}, \pi_i^{\mathrm{tr}}, e_i^{\mathrm{tr}})_{i=1}^n$ from the observed environments $e_i^{\mathrm{tr}} \in \mathcal{E}^{\mathrm{tr}}$. During test time, we are given $m \in \mathbb{N}$ observations $D_X^{\mathrm{tst}} := (X_i^{\mathrm{tst}})_{i=1}^m$ from a single test environment $e^{\mathrm{tst}} \in \mathcal{E}$. We denote by $Q^{\mathrm{tr}} := Q_1^{\mathrm{tr}} \otimes \ldots \otimes Q_n^{\mathrm{tr}}$, where $Q_i^{\mathrm{tr}} := \mathbb{P}_{X,Y,T}^{e_i,\pi_i}$ and $Q_X^{\mathrm{tst}} := \mathbb{P}_X^{e^{\mathrm{tst}}}$ the distributions of D^{tr} and D_X^{tst}, respectively.*

We seek to find a policy that generalizes well to the test environment e^{tst}. As we only have access to the observed covariate distribution $\mathbb{P}_X^{e^{\mathrm{tst}}}$ and since there may be multiple potential test environments $e \in \mathcal{E}$ with $\mathbb{P}_X^e = \mathbb{P}_X^{e^{\mathrm{tst}}}$, we propose to evaluate the performance of a policy π based on its expected outcome (relative to a fixed baseline policy π_{t_0} that always chooses t_0) in the worst-case scenario across all environments with covariate distribution equal to $\mathbb{P}_X^{e^{\mathrm{tst}}}$. Formally, let $[e^{\mathrm{tst}}] := \{e \in \mathcal{E} \mid \mathbb{P}_X^e = Q_X^{\mathrm{tst}}\}$ be an equivalence class of environments under which the covariate distribution $\mathbb{P}_X^e$ is the same as Q_X^{tst}. We then consider the following worst-case objective

$$V^{[e^{\mathrm{tst}}]}(\pi) := \inf_{e \in [e^{\mathrm{tst}}]} \big(\mathbb{E}^{e,\pi}[Y] - \mathbb{E}^{e,\pi_{t_0}}[Y] \big). \tag{9}$$

The goal of (population) zero-shot generalization applied to our setting is then to find a policy that (i) is identifiable from Q_i^{tr} (for an arbitrary $1 \leq i \leq n$) and Q_X^{tst} and (ii) maximizes the worst-case performance defined in (9).

We now introduce a policy $\pi^{\mathrm{e\text{-}inv}}$ that optimally uses information from e-invariant sets and show that $\pi^{\mathrm{e\text{-}inv}}$ achieves the aforementioned goal under suitable assumptions. To this end, for all $S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}$ (see Definition 3), we denote the set of all policies that depend only on X^S by $\Pi^S := \{\pi \in \Pi \mid \exists \bar{\pi} : \mathcal{X}^S \to \Delta(\mathcal{T}) \text{ s.t. } \forall x \in \mathcal{X}, \pi(\cdot|x) = \bar{\pi}(\cdot|x^S)\} \subseteq \Pi$. Next, for all $S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}$, we define $\Pi_{\mathrm{opt}}^S \subseteq \Pi^S$ to be a set of policies such that each $\pi^S \in \Pi_{\mathrm{opt}}^S$ satisfies for all $x \in \mathcal{X}$ and $t \in \mathcal{T}$ that

$$\pi^S(t|x) > 0 \implies t \in \operatorname*{argmax}_{t' \in \mathcal{T}} \frac{1}{|\mathcal{E}^{\mathrm{tr}}|} \sum_{e \in \mathcal{E}^{\mathrm{tr}}} \tau_e^S(x^S, t'). \tag{10}$$

That is, all the mass of $\pi^S(\cdot|x)$ is distributed on treatments that maximize the treatment effect conditioned on X^S. Since $\mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}$ contains only e-invariant sets w.r.t. $\mathcal{E}^{\mathrm{tr}}$, we also have that $\frac{1}{|\mathcal{E}^{\mathrm{tr}}|} \sum_{e \in \mathcal{E}^{\mathrm{tr}}} \tau_e^S \equiv \tau_f^S$ for any fixed $f \in \mathcal{E}^{\mathrm{tr}}$ (but for finite samples, we approximate the former). Finally, we denote by

$$\Pi_{\mathrm{opt}}^{\mathrm{e\text{-}inv}} := \{\pi \in \Pi \mid \exists S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}} \text{ s.t. } \pi \in \Pi_{\mathrm{opt}}^S\}$$

the collection of all such policies.

We now propose to use a policy from the collection of policies that are optimal among $\Pi_{\mathrm{opt}}^{\text{e-inv}}$, i.e.,

$$\underset{\pi \in \Pi_{\mathrm{opt}}^{\text{e-inv}}}{\operatorname{argmax}} \mathbb{E}^{e^{\mathrm{tst}}, \pi}[Y]. \tag{11}$$

Although the set (11) depends on the expected value of Y in the test environment, in Proposition 8 we show that we can construct a policy, denoted by $\pi^{\text{e-inv}}$, that satisfies the argmax property (11) and is identifiable from the data available during training (i.e., i.i.d. observations from Q^{tr} and Q_X^{tst}).

In Theorem 9, we then prove generalization properties of an optimal e-invariant policy $\pi^{\text{e-inv}}$. This generalization result requires the following two assumptions.

Assumption 2 (Generalizing environments) *It holds for all $S \subseteq \{1, \ldots, d\}$ that*

$$S \text{ is e-invariant w.r.t. } \mathcal{E}^{\mathrm{tr}} \implies S \text{ is e-invariant w.r.t. } \mathcal{E}^{\mathrm{tr}} \cup [e^{\mathrm{tst}}]. \tag{12}$$

Assumption 2 imposes some commonalities between environments which allows a transfer of e-invariance from the observed to the test environments. Similar assumptions are used when proving guarantees of other invariance-based learning methods (e.g., Rojas-Carulla et al. (2018); Magliacane et al. (2018); Christiansen et al. (2021); Pfister et al. (2021); Saengkyongam et al. (2023)).

Assumption 3 (Adversarial environment) *There exist $e \in [e^{\mathrm{tst}}]$ and $S \in \mathbf{S}_{\mathcal{E}}^{\text{e-inv}}$ such that for all $x \in \mathcal{X}$ it holds that*

$$\max_{t \in \mathcal{T}} \tau_e(x, t) = \max_{t \in \mathcal{T}} \tau_e^S(x^S, t). \tag{13}$$

Assumption 3 ensures that there exists at least one environment that does not benefit from non-e-invariant covariates and facilitates the worst-case optimality result of our proposed optimal e-invariant policy $\pi^{\text{e-inv}}$. Without Assumption 3, relying only on e-invariant covariates can become suboptimal if other (non-e-invariant) covariates are beneficial across all environments. For example, consider Example 1 and assume that the coefficients γ_e^1 and γ_e^3 in different environments are relatively close, e.g., $\forall e \in \mathcal{E} : \gamma_e^1, \gamma_e^3 \in (0.9, 1)$. In this scenario, $\{X^1, X^3\}$ is not e-invariant. Still, it is preferable to use these variables for policy learning as they provide valuable information for predicting U^1, which modifies the treatment effect. In the above setting, Assumption 3 does not hold; it would be satisfied if there is at least one additional environment $e \in [e^{\mathrm{tst}}]$ where $\gamma_e^1 = \gamma_e^3 = 0$. The reason is that in such an environment the variables X^1 and X^3 do not offer any relevant information for predicting U^1. A similar assumption, known as confounding-removing interventions, is introduced in (Christiansen et al., 2021) in the prediction setting.

Proposition 8 (Identifiability) *Assume Setting Zero-shot and Assumptions 1 and 2. Let $e \in \mathcal{E}^{\mathrm{tr}}$ be an arbitrary training environment, for all $S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\text{e-inv}}$ let $\pi^S \in \Pi_{\mathrm{opt}}^S$, that is, a policy that satisfies (10), and let S^* be a subset such that*

$$S^* \in \mathbf{S}^{\mathrm{opt}} := \underset{S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\text{e-inv}}}{\operatorname{argmax}} \mathbb{E}^{e^{\mathrm{tst}}} \left[\sum_{t \in \mathcal{T}} \tau_e^S(X^S, t) \pi^S(t \mid X) \right]. \tag{14}$$

Define $\pi^{\text{e-inv}} := \pi^{S^*}$. Then, the following holds: (i) the set $\mathbf{S}^{\text{opt}}$ is identifiable from the distributions Q_i^{tr} (for an arbitrary $1 \leq i \leq n$) and Q_X^{tst} (which makes it possible to choose S^* and π^{S^*} during test time) and (ii) $\pi^{\text{e-inv}}$ is an element in (11).

Proof See Appendix A.3. ∎

Theorem 9 (Generalizability) *Assume Setting Zero-shot and Assumptions 1 and 2. Let $\pi^{\text{e-inv}}$ be as defined in Proposition 8. Then, the two following statements hold.*

(i) *Let π_t, as defined in Section 2.1, be the policy that always chooses treatment $t \in \mathcal{T}$. We have that*

$$\max_{t \in \mathcal{T}} \mathbb{E}^{e^{\text{tst}}, \pi_t}[Y] \leq \mathbb{E}^{e^{\text{tst}}, \pi^{\text{e-inv}}}[Y]. \tag{15}$$

(ii) *Given Assumption 3, we have that*

$$\forall \pi \in \Pi : V^{[e^{\text{tst}}]}(\pi^{\text{e-inv}}) \geq V^{[e^{\text{tst}}]}(\pi). \tag{16}$$

Proof See Appendix A.4. ∎

Theorem 9 provides two generalization properties of the policy $\pi^{\text{e-inv}}$. First, Theorem 9(i) shows that $\pi^{\text{e-inv}}$ guarantees to outperform, in any (unseen) test environment, an optimal policy that does not use covariates X. In other words, it is always beneficial to utilize the information from e-invariant sets when generalizing treatment regimes, compared to ignoring the covariates. Second, Theorem 9(ii) shows that $\pi^{\text{e-inv}}$ maximizes the worst-case performance defined in (9), that is, it outperforms all other policies when evaluating each policy in the respective worst case environment if Assumption 3 holds true.

3.1 Estimation of $\pi^{\text{e-inv}}$

As shown in Proposition 8, the policy $\pi^{\text{e-inv}}$ is identifiable from Q^{tr} and Q_X^{tst}. We now turn to the problem of estimating $\pi^{\text{e-inv}}$ given data D^{tr} and D_X^{tst} of Q^{tr} and Q_X^{tst}, respectively. For now, assume we are given the collection $\mathbf{S}_{\mathcal{E}^{\text{tr}}}^{\text{e-inv}}$ of all e-invariant sets w.r.t. $\mathcal{E}^{\text{tr}}$. We discuss how to estimate $\mathbf{S}_{\mathcal{E}^{\text{tr}}}^{\text{e-inv}}$ in Section 4.

Proposition 8 suggests a plug-in estimator of $\pi^{\text{e-inv}}$ based on (14). Specifically, the estimate can be obtained as follows.

(i) For all $S \in \mathbf{S}_{\mathcal{E}^{\text{tr}}}^{\text{e-inv}}$, compute an estimate $\hat{\tau}^S$ for τ_e^S, $e \in \mathcal{E}^{\text{tr}}$, by pooling the data from the training environments (as the τ_e^S's are equal across environments by effect-invariance). There is a rich literature on estimating CATE from observational data (see Zhang et al. (2021b) for a survey), one can choose an estimator that is appropriate to a given dataset. In particular, the choice of the CATE estimator can differ from the one employed in the testing step outlined in Section 4. Finally, once an estimate $\hat{\tau}^S$ is obtained, we then plug $\hat{\tau}^S$ into (10) to construct an estimate $\hat{\pi}^S$ for π^S, that is, $\hat{\pi}^S$ satisfies for all $x \in \mathcal{X}$ and $t \in \mathcal{T}$ that

$$\hat{\pi}^S(t|x) > 0 \implies t \in \operatorname*{argmax}_{t' \in \mathcal{T}} \hat{\tau}^S(x^S, t'). \tag{17}$$

We distribute the probabilities equally if there are more than one t satisfying (17).

(ii) Find an optimal subset S^* among $\mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}$, see (14):

$$\hat{S}^* \in \operatorname*{argmax}_{S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}} \frac{1}{m} \sum_{i=1}^{m} \left[\sum_{t \in \mathcal{T}} \hat{\tau}^S((X_i^{\mathrm{tst}})^S, t) \hat{\pi}^S(t \mid X_i^{\mathrm{tst}}) \right]. \tag{18}$$

If there are multiple S satisfying (18), we randomly choose $\hat{S}^*$ among one of them.

(iii) Return $\hat{\pi}^{\hat{S}^*}$ – which was already computed in step (i) – as the estimate of $\pi^{\mathrm{e\text{-}inv}}$.

4. Inferring e-invariant sets

We now turn to the problem of testing the e-invariance condition (5) based on training observations $D^{\mathrm{tr}} := (Y_i, X_i, T_i, \pi_i, e_i)_{i=1}^{n}$ from the observed environments $e_i \in \mathcal{E}^{\mathrm{tr}}$.

Throughout this section, we assume a fixed initial (or training) policy π^{tr}, i.e., $\forall i \in \{1, \ldots, n\} : \pi^{\mathrm{tr}} = \pi_i$. The initial policy π^{tr} can either be given or estimated from the available data (see, e.g., Algorithm 2). Our proposed testing methods remain valid even if the initial policies $(\pi_i)_{i=1}^{n}$ are different as long as they are both known and independent of all observed quantities[6]. Furthermore, we consider discrete environments, $\mathcal{E}^{\mathrm{tr}} = \{1, \ldots, \ell\}$, and consider a binary treatment variable, $\mathcal{T} = \{0, 1\}$. One can generalize to a multi-level treatment variable by repeating the proposed procedures for each level $1, \ldots, k$ with the baseline treatment $t_0 = 0$ and combining the test results with a multiple testing correction method.

To begin with, we define for all $S \subseteq \{1, \ldots, d\}$ the e-invariance null hypothesis

$$H_{0,S}^{\mathrm{tr}} : S \text{ is e-invariant w.r.t. } \mathcal{E}^{\mathrm{tr}}, \tag{19}$$

see Definition 3.

In Section 4.1, we propose a testing procedure under the assumption that, for all $S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}$, the functions $(\tau_e^S)_{e \in \mathcal{E}^{\mathrm{tr}}}$ can be modelled by linear functions and provide its statistical guarantees. In Section 4.2, we relax the linearity assumption by using a doubly robust pseudo-outcome learner (see, e.g., Kennedy, 2020).

4.1 Linear CATE functions

One way of creating e-invariance tests is to assume a parametric form of the CATEs. In this section, we rely on the following linearity assumption.

Assumption 4 (Linear CATEs) *For all $S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}$, there exist coefficients $(\gamma_t^S)_{t \in \mathcal{T}} \in \mathbb{R}^{k \times |S|}$ and intercepts $(\mu_t^S)_{t \in \mathcal{T}} \in \mathbb{R}^k$ such that*

$$\forall e \in \mathcal{E}, \forall t \in \mathcal{T}, \forall x^S \in \mathcal{X}^S : \tau_e^S(x^S, t) = \mu_t^S + \gamma_t^S x^S. \tag{20}$$

Under Assumption 4, we now present a testing method for the e-invariance hypothesis $H_{0,S}^{\mathrm{tr}}$ for a fixed set $S \subseteq \{1, \ldots, d\}$. Let $u_e \in \{0,1\}^{1 \times \ell}$ and $v_t \in \{0,1\}$ be the one-hot encodings of the environment $e \in \mathcal{E}^{\mathrm{tr}}$ and the treatment $t \in \mathcal{T}$, respectively, and let $\alpha \in \mathbb{R}^{1 \times (1+d)}$,

---

6. Specifically, we do not allow for data collected with adaptive algorithms, which we leave for future work, see Section 7.

$A \in \mathbb{R}^{\ell \times (1+d)}$, $\beta \in \mathbb{R}^{1 \times (1+|S|)}$ and $B \in \mathbb{R}^{\ell \times (1+|S|)}$ be model parameters. For notational convenience, we define $\tilde{X}_i := \begin{bmatrix} 1 & X_i \end{bmatrix}^\top \in \mathbb{R}^{(1+d) \times 1}$ and $\tilde{X}_i^S := \begin{bmatrix} 1 & X_i^S \end{bmatrix}^\top \in \mathbb{R}^{(1+|S|) \times 1}$. We consider the following (potentially misspecified) response model under treatment $t \in \mathcal{T}$ and environment $e \in \mathcal{E}^{\mathrm{tr}}$

$$\underbrace{\alpha \tilde{X} + (u_e A) \tilde{X}}_{\text{main effect}} + \underbrace{(v_t \beta) \tilde{X}^S}_{\text{treatment effect}} + \underbrace{(v_t u_e B) \tilde{X}^S}_{\text{environment} \times \text{treatment effect}} . \tag{21}$$

In this model, we have that the CATE functions τ_e^S are identical across environments $e \in \mathcal{E}^{\mathrm{tr}}$ if and only if $B = 0$. Thus, testing (19) is equivalent to testing the null hypothesis $H_0 : B = 0$.

The model proposed in (21) is more restrictive than Assumption 4 as it additionally requires the main effect to be linear. To avoid this requirement, we propose using a testing methodology that explicitly allows for the misspecification in the main effect, where we employ the centered and weighted estimation method proposed by Boruvka et al. (2018), which uses a Neyman orthogonal score (Neyman, 1959, 1979). (A standard approach of weighted least-squares using weights $1/\pi^{\mathrm{tr}}(T_i|X_i)$ may not yield a test with the correct asymptotic level for the null hypothesis H_0.) More precisely, we consider the following steps:

(i) Treatment centering: We center the treatment indicators v_{T_i} by an arbitrary fixed policy $\tilde{\pi}$ that depends only on X^S (i.e., $\tilde{\pi} \in \Pi^S$). More precisely, we replace v_{T_i} with $v_{T_i} - \tilde{\pi}(1|X_i^S)$. As an example, one could consider a fixed random policy $\tilde{\pi}(t|x) := q^t(1-q)^{(1-t)}$ for some $q \in [0, 1]$.

(ii) Weighted least squares: We estimate the model parameters via a weighted least-squares approach. The weights are defined by $W_i := \tilde{\pi}(T_i|X_i^S)/\pi^{\mathrm{tr}}(T_i|X_i)$, where $\tilde{\pi}$ is the policy chosen in step (i) and π^{tr} is the initial policy.

The use of the above steps ensures that the estimator for treatment effects remains consistent even if the main effect is misspecified (Boruvka et al., 2018) and allows us to obtain a test with pointwise asymptotic level, see Proposition 10.

Formally, we employ a generalized method of moments estimator. Define $\zeta_i(\alpha, A, \beta, B) :=$ $\alpha \tilde{X}_i + (u_{e_i} A) \tilde{X}_i + (v_{T_i} - \tilde{\pi}(1|X_i^S))(\beta \tilde{X}_i^S + (u_{e_i} B) \tilde{X}_i^S)$ and $\nabla \zeta_i := \begin{bmatrix} \frac{\partial \zeta_i}{\partial \alpha} & \frac{\partial \zeta_i}{\partial A} & \frac{\partial \zeta_i}{\partial \beta} & \frac{\partial \zeta_i}{\partial B} \end{bmatrix}^\top$. We then estimate $\hat{\alpha}, \hat{A}, \hat{\beta}, \hat{B}$ as the solutions to the estimating equations

$$\sum_{i=1}^{n} G_i(\alpha, A, \beta, B) = 0, \tag{22}$$

where $G_i(\alpha, A, \beta, B) := W_i[Y_i - \zeta_i(\alpha, A, \beta, B)]\nabla \zeta_i$.

Under additional regularity conditions (see Appendix A.5), we have, for a vectorized B, that $\sqrt{n}(\hat{B} - B) \xrightarrow{d} \mathcal{N}(0, \mathbb{V}[B])$.

This allows us to construct a hypothesis test for $H_0 : B = 0$. To this end, we estimate $\mathbb{V}[B]$ as follows. First, for all $i \in \{1, \ldots, n\}$ define

$$\widehat{G}_i := G_i(\hat{\alpha}, \hat{A}, \hat{\beta}, \hat{B}) \in \mathbb{R}^{s+q} \quad \text{and} \quad \widehat{J}_i := J_i(\hat{\alpha}, \hat{A}, \hat{\beta}, \hat{B}) \in \mathbb{R}^{(s+q) \times (s+q)},$$

where J_i is the Jacobian of G_i and $s := (1 + d + |S|) + \ell(1 + d)$ and $q := \ell(1 + |S|)$. Then, the covariance matrix $\mathbb{V}[B]$ can be consistently estimated by the lower block diagonal $q \times q$ entry of the matrix, i.e,

$$\hat{\mathbb{V}} := \left(\frac{1}{n} \sum_{i=1}^n \hat{J}_i \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \hat{G}_i \hat{G}_i^\top \right) \left(\frac{1}{n} \sum_{i=1}^n \hat{J}_i \right)^{-\top}, \tag{23}$$

(Boruvka et al., 2018, Proposition 3.1).

We can then use the Wald test to test the null hypothesis $H_0 : B = \mathbf{0}$ using the consistent estimator $\hat{\mathbb{V}}$ of $\mathbb{V}[B]$ (see, e.g., Boos et al. (2013)). When both $\tilde{\pi}$ and π^{tr} are given, the covariance estimate can be obtained using standard implementations (e.g., Huber-White covariance estimator (Huber, 1967; White, 1980)). However, when either $\tilde{\pi}$ or π^{tr} are estimated, one needs to adjust the covariance estimator to incorporate the additional estimation error (see Supplement C in Boruvka et al. (2018)). The full testing procedure is given in Algorithm 1.

Algorithm 1 (Wald e-invariance test) *Given: a training sample D^{tr} of size n, a subset $S \subseteq \{1, \ldots, d\}$ and a significance level $\alpha \in (0, 1)$.*

(i) *Solve the estimating equation (22) and compute the covariance estimator (23) and the test statistic $T_n := n\hat{B}^\top \hat{\mathbb{V}} \hat{B}$.*

(ii) *Return $\psi_n^{Wd}(D^{\mathrm{tr}}, S, \alpha) := \mathbb{1}(T_n > q_\alpha)$, where q_α is the $(1 - \alpha)$-quantile of a chi-squared distribution with $\ell(1 + |S|)$-degrees of freedom.*

Proposition 10 shows that the above results carry over to our setting in that the proposed procedure achieves pointwise asymptotic level for testing the e-invariance hypothesis $H_{0,S}^{\mathrm{tr}}$.

Proposition 10 *Assume Setting 1 and Assumption 4. Let $S \subseteq \{1, \ldots, d\}$ be a subset of interest, $\alpha \in (0, 1)$ be a significance level, and $\psi_n^{Wd}(D^{\mathrm{tr}}, S, \alpha)$ be the Wald invariance test detailed in Algorithm 1. Under some regularity conditions (see Appendix A.5), it holds that $\psi_n^{Wd}(D^{\mathrm{tr}}, S, \alpha)$ has pointwise asymptotic level for testing the e-invariance hypothesis $H_{0,S}^{\mathrm{tr}}$, that is,*

$$\sup_{\mathbb{P} \in H_{0,S}^{\mathrm{tr}}} \limsup_{n \to \infty} \mathbb{P}(\psi_n^{Wd}(D^{\mathrm{tr}}, S, \alpha) = 1) \leq \alpha. \tag{24}$$

Proof The proof follows directly from (Boruvka et al., 2018, Proposition 3.1), see Appendix A.5. ∎

4.2 Non-linear CATE functions

This section relaxes the assumption of linear CATEs (Assumption 4) and proposes a non-parametric approach for testing the e-invariance hypothesis $H_{0,S}^{\mathrm{tr}}$. The key idea is to employ a pseudo-outcome approach to estimate non-linear CATE functions (see (4)) and apply a conditional mean independence test based on the pseudo-outcome. In particular, we consider the Doubly Robust (DR) learner due to Kennedy (2020).

For all $e \in \mathcal{E}^{\mathrm{tr}}$, let $\bar{\mu}^e : \mathcal{X} \times \mathcal{T} \to \mathbb{R}$ denote a model of the conditional expected outcome $\mathbb{E}^e[Y \mid X = \cdot, T = \cdot]$ and $\bar{\pi}$ denote a model of the initial policy π^{tr}. Assume $t_0 = 0$. We consider, for all $e \in \mathcal{E}^{\mathrm{tr}}, x \in \mathcal{X}, t \in \mathcal{T}$ and $y \in \mathcal{Y}$, the function

$$O^e(x,t,y) = \bar{\mu}^e(x,1) - \bar{\mu}^e(x,0) + \frac{\mathbb{1}(t=1)(y-\bar{\mu}^e(x,1))}{\bar{\pi}(1|x)} - \frac{\mathbb{1}(t=0)(y-\bar{\mu}^e(x,0))}{1-\bar{\pi}(1|x)}, \quad (25)$$

and generate pseudo-outcomes by plugging in the observed data. The motivation for constructing the above pseudo-outcome is that, under Setting 1, the conditional mean of $O^e(X,T,Y)$ given X^S is equal to the CATE function τ_e^S if at least one of the models $\bar{\mu}^e$ or $\bar{\pi}$ is correct. Formally, we have the following result.

Proposition 11 *Assume Setting 1. Let $S \subseteq \{1,\ldots,d\}, e \in \mathcal{E}^{\mathrm{tr}}, \pi \in \Pi$ and $O^e(\cdot)$ be the pseudo-outcome defined in (25). Assume $t_0 = 0$. If for all $x \in \mathcal{X}$ and $t \in \mathcal{T}$ $\bar{\mu}^e(x,t) = \mathbb{E}^{e,\pi^{\mathrm{tr}}}[Y \mid X = x, T = t]$ or for all $x \in \mathcal{X}$ and $t \in \mathcal{T}$ $\bar{\pi}(t \mid x) = \pi^{\mathrm{tr}}(t \mid x)$, we have for all $x \in \mathcal{X}^S$ that*

$$\mathbb{E}^{e,\pi^{\mathrm{tr}}}[O^e(X,T,Y) \mid X^S = x] = \tau_e^S(x,1). \quad (26)$$

Proof See Appendix A.6. ∎

Under the assumptions of Proposition 11, it holds for all $S \subseteq \{1,\ldots,p\}$ that the null hypothesis $H_{0,S}^{\mathrm{tr}}$ is equal to

$$\forall e_1, e_2 \in \mathcal{E}^{\mathrm{tr}}, \forall x \in \mathcal{X}^S : \mathbb{E}^{e_1,\pi^{\mathrm{tr}}}[O^{e_1}(X,T,Y) \mid X^S = x] = \mathbb{E}^{e_2,\pi^{\mathrm{tr}}}[O^{e_2}(X,T,Y) \mid X^S = x]. \quad (27)$$

We can thus test for e-invariance by using an appropriate conditional mean independence test that has a correct level under the null hypothesis (27). For example, one can use the generalised covariance measure[7] (Shah and Peters, 2020; Scheidegger et al., 2022) or the projected covariance measure (Lundborg et al., 2022).

We therefore propose the following steps to construct a non-parametric test for the e-invariance hypothesis $H_{0,S}^{\mathrm{tr}}$.

Algorithm 2 (DR-learner e-invariance test) *Given a training sample D^{tr} of size n, subset of interest S, significance level α and conditional mean independence test ϕ. Let $D_1 \subset D^{\mathrm{tr}}$ denote a random sample of D^{tr}, and $D_2 := D^{\mathrm{tr}} \setminus D_1$.*

(i) Fit models $\bar{\mu}^e$ and $\bar{\pi}$ from the data D_1.

(ii) Construct the pseudo-outcomes

$$O_i^{e_i}(X_i,T_i,Y_i) = \bar{\mu}_1^{e_i}(X_i) - \bar{\mu}_0^{e_i}(X_i) + T_i \frac{Y_i - \bar{\mu}_1^{e_1}(X_i)}{\bar{\pi}(1 \mid X_i)} - (1-T_i)\frac{Y_i - \bar{\mu}_0^{e_i}(X_i)}{1 - \bar{\pi}(1 \mid X_i)}.$$

for each observation $(X_i, T_i, Y_i) \in D_2$ in D_2.

---

7. The generalised covariance measure (GCM) does not directly test for the conditional mean independence. However, it preserves the level guarantees under the conditional mean independence null hypothesis. Specifically, consider a random vector (A,B,C). It holds that $\mathbb{E}[A \mid B, C] = \mathbb{E}[A \mid B] \implies \mathrm{Cov}(A,B \mid C) = 0 \implies \mathbb{E}[\mathrm{Cov}(A,B \mid C)] = 0$, where the first equality is the conditional mean independence hypothesis and the last equality is the null hypothesis of the GCM test.

(iii) Apply the test ϕ on $O_i^{e_i}(X_i, T_i, Y_i)$ and observations in D_2 with a significance level α and return the test result.

In practice, when using Algorithm 1 or Algorithm 2 to search for all e-invariant set $\mathbf{S}_{\mathcal{E}^{\text{tr}}}^{\text{e-inv}}$, one needs to iterate over all subsets $S \subseteq \{1, \ldots, d\}$, which can be computationally challenging when d is large. To mitigate this issue, as suggested in e.g., Peters et al. (2016); Rojas-Carulla et al. (2018); Saengkyongam et al. (2023), we can employ a variable screening method, such as Lasso regression (Tibshirani, 1994), to filter out variables that are not relevant in estimating the CATE function and employ Algorithm 1 or Algorithm 2 on the resulting set.

5. Extension: Few-shot policy generalization through e-invariance

In the Zero-shot setting, the outcome is not observed in the test environment and, as shown in Theorem 9, relying on e-invariant covariates is optimal under certain assumptions. This is no longer true if, in the test environment, we have access to observations not only of the covariates but also of the corresponding outcomes obtained after using a test policy in the test environment. We may then want to adapt to the test environment while exploiting the e-invariance information gathered in the training environments. In this section, we illustrate how our method could be extended to such a setup (called few-shot generalization), where we observe a large number of training observations from the training environments and a small number of test observations (including the outcome) from the test environment.

Setting Few-shot *Assume Setting 1 and that we are given $n \in \mathbb{N}$ training observations $D^{\text{tr}} := (Y_i^{\text{tr}}, X_i^{\text{tr}}, T_i^{\text{tr}}, \pi_i^{\text{tr}}, e_i^{\text{tr}})_{i=1}^n$ from the observed environments $e_i^{\text{tr}} \in \mathcal{E}^{\text{tr}}$ and $m \in \mathbb{N}$ test observations $D^{\text{tst}} := (Y_i^{\text{tst}}, X_i^{\text{tst}}, T_i^{\text{tst}}, \pi_i^{\text{tst}})_{i=1}^m$ from a test environment $e^{\text{tst}} \in \mathcal{E}$ and assume that $m \ll n$.*

The goal of few-shot policy generalization is to find a policy $\pi \in \Pi$ that maximizes the expected outcome in the test environment e^{tst} by exploiting the common information shared between the training and test environments. We consider using Assumption 2 as the commonalities shared between the environments. In what follows, we propose a constrained optimization approach to learn a policy that aims to maximize the expected outcome in the test environment while exploiting the e-invariance condition.

An optimal policy π_{tst}^* in the test environment e^{tst} distributes all its mass on treatments which maximize the CATE in the test environment – conditioned on the covariates X. That is, an optimal policy π_{tst}^* satisfies for all $x \in \mathcal{X}$ and $t \in \mathcal{T}$ that

$$\pi_{\text{tst}}^*(t|x) > 0 \implies t \in \operatorname*{argmax}_{t' \in \mathcal{T}} \tau_{e^{\text{tst}}}(x, t'). \tag{28}$$

Therefore, learning an optimal policy π_{tst}^* can be reduced to learning the CATE function $\tau_{e^{\text{tst}}}$ in the test environment.

As mentioned in Section 3.1, the problem of learning $\tau_{e^{\text{tst}}}$ from observational data is a well-studied problem. Here, we abstract away from a specific method and assume that we are given a function class $\mathcal{H} \subseteq \{\tau \mid \tau : \mathcal{X} \times \mathcal{T} \to \mathbb{R}\}$ and a loss function $\ell : \mathcal{Y} \times \mathcal{X} \times \mathcal{T} \times \Pi \times \mathcal{H} \to \mathbb{R}$ such that $\hat{\tau} \in \operatorname{argmin}_{\tau \in \mathcal{H}} \sum_{i=1}^m \ell(Y_i^{\text{tst}}, X_i^{\text{tst}}, T_i^{\text{tst}}, \pi_i^{\text{tst}}, \tau)$ is a consistent estimator of $\tau_{e^{\text{tst}}}$ as

$m \to \infty$. Now, we propose to leverage Assumption 2 when estimating $\tau_{e^{\mathrm{tst}}}$ in the test environment. In particular, by Assumption 2 we have for all $S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}$ and for any fixed $e \in \mathcal{E}^{\mathrm{tr}}$ that

$$\forall x \in \mathcal{X}^S, \forall t \in \mathcal{T} : \tau_e^S(x, t) = \mathbb{E}^{e^{\mathrm{tst}}}[\tau_{e^{\mathrm{tst}}}(X, t) \mid X^S = x]. \tag{29}$$

Let $S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}$ and for all $x \in \mathcal{X}^S$, $t \in \mathcal{T}$ and $\tau \in \mathcal{H}$, we define $h^S(\tau, x, t) := \mathbb{E}^{e^{\mathrm{tst}}}[\tau(X, t) \mid X^S = x]$ and $\tau_{\mathrm{tr}}^S(x, t) := \tau_e^S(x, t)$ (for an arbitrary $e \in \mathcal{E}^{\mathrm{tr}}$).

We then consider the following constrained optimization

$$\hat{\tau}^S \in \operatorname*{argmin}_{\tau} \sum_{i=1}^m \ell(Y_i^{\mathrm{tst}}, X_i^{\mathrm{tst}}, T_i^{\mathrm{tst}}, \pi_i^{\mathrm{tst}}, \tau) \tag{30}$$
$$\text{s.t. } \tau \in \mathcal{H} \text{ and } \tau_{\mathrm{tr}}^S(\cdot, \cdot) \equiv h^S(\tau, \cdot, \cdot).$$

If there are multiple $S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}$ satisfying e-invariance, that is, $|\mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}| > 1$, one may choose an optimal set S^* as in (14).

We now impose the following separability assumption on the CATE function $\tau_{e^{\mathrm{tst}}}$, which allows us to find a solution to the optimization problem (30).

Assumption 5 (Separability of CATEs) *Let $e^{\mathrm{tst}} \in \mathcal{E}$ be a test environment, $S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}$ and $N := \{1, \ldots, d\} \setminus S$. There exist function classes $\mathcal{F} \subseteq \{\mathcal{X}^S \times \mathcal{T} \to \mathbb{R}\}$ and $\mathcal{G} \subseteq \{\mathcal{X}^N \times \mathcal{T} \to \mathbb{R}\}$ and a pair of functions $f \in \mathcal{F}$ and $g \in \mathcal{G}$ such that*

$$\forall x \in \mathcal{X}, \forall t \in \mathcal{T} : \tau_{e^{\mathrm{tst}}}(x, t) = f(x^S, t) + g(x^N, t). \tag{31}$$

Under Assumptions 2 and 5, there exists $f \in \mathcal{F}$ and $g \in \mathcal{G}$ such that for all $x \in \mathcal{X}^S$ and $t \in \mathcal{T}$

$$\begin{aligned}
\tau_{\mathrm{tr}}^S(x, t) &= \mathbb{E}^{e^{\mathrm{tst}}}[\tau_{e^{\mathrm{tst}}}(X, t) \mid X^S = x] \\
&= \mathbb{E}^{e^{\mathrm{tst}}}[f(X^S, t) + g(X^N, t) \mid X^S = x] \\
&= f(x, t) + \mathbb{E}^{e^{\mathrm{tst}}}[g(X^N, t) \mid X^S = x],
\end{aligned}$$

which is equivalent to

$$f(x, t) = \tau_{\mathrm{tr}}^S(x, t) - \mathbb{E}^{e^{\mathrm{tst}}}[g(X^N, t) \mid X^S = x]. \tag{32}$$

Combining (32) and (31), we then have for all $x \in \mathcal{X}$ and $t \in \mathcal{T}$ that

$$\tau_{e^{\mathrm{tst}}}(x, t) = \tau_{\mathrm{tr}}^S(x^S, t) - \mathbb{E}^{e^{\mathrm{tst}}}[g(X^N, t) \mid X^S = x^S] + g(x^N, t). \tag{33}$$

Instead of optimizing over the function class $\mathcal{H}$, we now optimize over the function class $\mathcal{G}$ by replacing τ in (30) with $\tau_g^S : (x, t) \mapsto \tau_{\mathrm{tr}}^S(x^S, t) - \mathbb{E}^{e^{\mathrm{tst}}}[g(X^N, t) \mid X^S = x^S] + g(x^N, t)$. More specifically, we consider the unconstrained optimization

$$\hat{g} \in \operatorname*{argmin}_{g \in \mathcal{G}} \sum_{i=1}^m \ell(Y_i^{\mathrm{tst}}, X_i^{\mathrm{tst}}, T_i^{\mathrm{tst}}, \pi_i^{\mathrm{tst}}, \tau_g^S). \tag{34}$$

Then, $\tau_{\hat{g}}^S$ is a solution to the constraint optimization (30).

In practice, we estimate the conditional expectation $\mathbb{E}^{e^{\text{tst}}}[g(X^N, t) \mid X^S = \cdot]$ by an estimator $\hat{q}_{g,t}$. Intuitively, if the function class $\mathcal{G}$ (see Assumption 5) has a lower complexity compared to $\mathcal{H}$, and $\hat{q}_{g,t}$ has good finite-sample properties, one may expect an improvement (e.g., $\tau_{\hat{g}}^S$ has a lower variance) using this approach over an estimator that does not take into account the training sample. Without additional assumptions on $\mathcal{G}$, the optimization problem (34) requires the computation of $\hat{q}_{g,t}$ at each iteration (since $\hat{q}_{g,t}$ depends on g). In Appendix B, we present an example to demonstrate that the optimization can simplify when imposing an additional assumption, such as linearity.

6. Experiments

This section presents the empirical experiments conducted on both simulated and real-world datasets. Firstly, we demonstrate through simulations that the testing methods introduced in Section 4 provide level guarantees that hold empirically in finite samples. Secondly, we demonstrate the effectiveness of our e-invariance approach in a semi-real-world case study of mobile health interventions, where it outperforms the baselines in terms of generalization to a new environment. The code for all experiments is available at `https://github.com/sorawitj/effect-invariance`.

6.1 Testing for e-invariance (simulated data)

We now conduct simulated experiments to validate the e-invariance tests proposed in Section 4. We generate datasets of size $n \in \{1000, 2000, 4000, 8000\}$ according to the SCM in Example 1 with two training environments $\mathcal{E}^{\text{tr}} = \{0, 1\}$. Each of the noise variables $(\epsilon_{U^1}, \epsilon_{U^2}, \epsilon_{X^1}, \epsilon_{X^2}, \epsilon_{X^3}, \epsilon_T, \epsilon_Y)$ is independently drawn from a standard Gaussian distribution. The environment-specific parameters $(\gamma_e^1, \gamma_e^2, \gamma_e^3, \mu_e)$ are drawn independently from a uniform distribution on $[-3, 3]$. As for the initial policy, we consider a policy that depends on the full covariate set $\{X_1, X_2, X_3\}$. More precisely, for all $x \in \mathcal{X}$, the initial policy π^{tr} selects a treatment according to $\pi^{\text{tr}}(T = 1 \mid X = x) = 1/(1 + e^{-(0.5 + x^1 - 0.5x^2 + 0.3x^3)})$. Moreover, we explore a scenario where the assumption of linear main effects in Equation (21) is violated. Specifically, we modify the structural assignment of Y in Example 1 as $Y := T(1 + 0.5X^2 + 0.5U^1) + U^1 + X^2 + \mu_e + U^2 - 0.5X^2X^3 + X^3 + \epsilon_Y$. Lastly, we also consider a setting where the treatment effect itself is nonlinear. In this case, the structural assignment for Y is defined as $Y := T(1 + 0.5(X^2)^2 + 0.5(X^2)^3 + 0.5U^1) + U^1 + X^2 + \mu_e + U^2 - 0.5X^2X^3 + X^3 + \epsilon_Y$.

We then conduct the Wald and DR-learner e-invariance tests (Wald test and DR test for short, respectively) for all candidate subsets according to Algorithm 1 and Algorithm 2, where we assume that the initial policy π^{tr} is given. For the DR test, we estimate the conditional mean function ($\bar{\mu}^e$) with a random forest (Breiman, 2001) and use the weighted generalised covariance measure (Shah and Peters, 2020; Scheidegger et al., 2022) as the final test ϕ in Algorithm 2.

Figure 2 reports the rejection rates at the 5% significance level for each candidate set under various settings. Recall that in Example 1, $\{X^2\}$ is the only e-invariant set. The results indicate that, for finite sample sizes, both of the proposed methods hold the correct level at 5% in all settings (the rejection rates for the e-invariant set $\{X^2\}$ are approximately 5% in all settings) except in the bottom left setting: here, the linear CATEs assumption
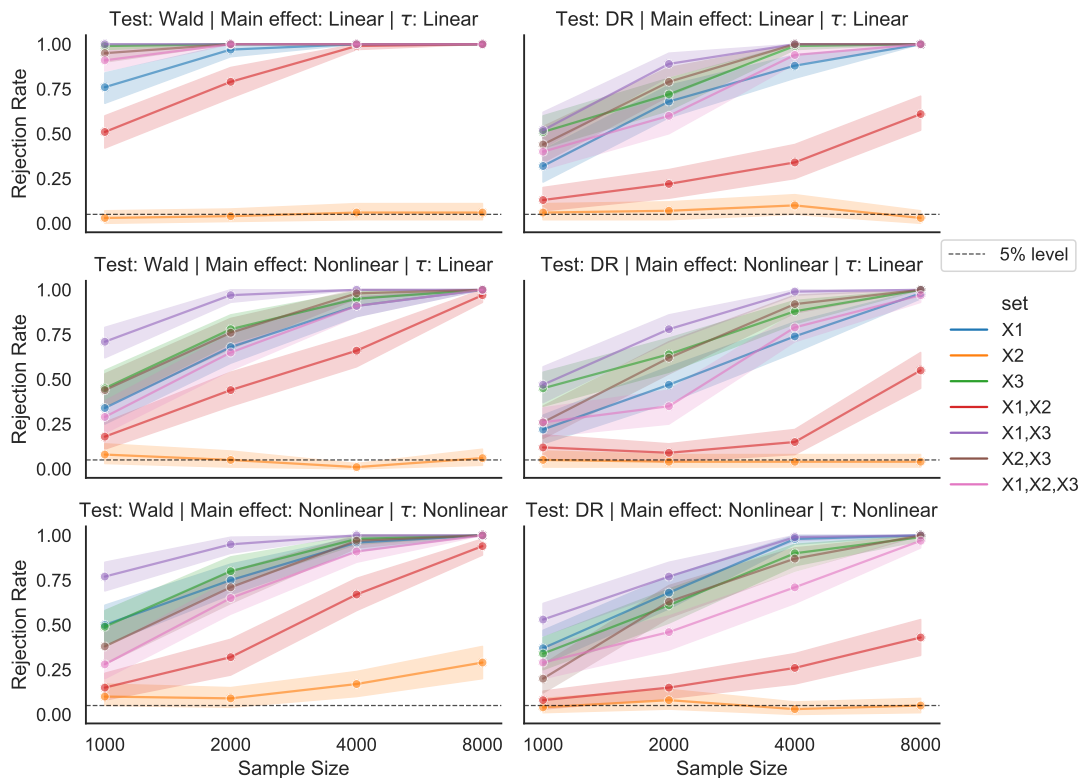
Figure 2: Rejection rates (at the 5% significance level) of the proposed e-invariance tests from Section 4 for varying sample sizes. (Top) the main effect and treatment effect – see (21) – are linear, (Middle) the main effect is nonlinear while the treatment effect is linear, (Bottom) both the main effect and treatment effect are nonlinear. In all settings, the DR-learner test achieves the correct level, i.e., the e-invariant set $\{X^2\}$ has a 5% rejection rate. Similarly, the Wald test correctly rejects the set $\{X^2\}$ at 5% level in all scenarios except in the bottom scenario due to the violation of Assumption 4. As sample size increases, all other sets are rejected with increasing empirical probability.

(Assumption 4) is violated and the Wald test fails to maintain the correct level. When the linear main effect and treatment effect assumptions in (21) are specified correctly (top row), the Wald test shows superior performance compared to the DR test (that is, the Wald test rejects the non-e-invariant sets more often). When the linear main effect assumption is violated (middle row), the Wald test remains valid but the power of the test drops significantly. The Wald test, nonetheless, slightly performs better than the DR test in terms of test power in this setting.

6.2 A case study using HeartSteps V1 dataset

We apply our proposed approach to the study of a mobile health intervention for promoting physical activity called HeartSteps V1 (Klasnja et al., 2019). HeartSteps V1 was a 42-day micro-randomized trial with 37 adults that aimed to optimize the effectiveness of

two intervention components for promoting physical activity. One of the interventions was contextual-aware activity suggestions, delivered as push notifications, which aimed to encourage short bouts of walking throughout the day. Each participant was equipped with a wearable tracker that linked to the mobile application, which gathered sensor data and contextual information about the user. This information was used to tailor the content of activity suggestions that users received and to determine whether the user was available to receive an activity suggestion (e.g., if the sensor data indicated that the user was currently walking, they would not be sent a suggestion). The application randomized the delivery of activity suggestion up to five times a day at user-selected times spaced approximately 2.5 hours apart. If the contextual information indicated that the person was unavailable for the intervention, no suggestion was sent.

In this paper, we consider users as environments. We filter out users that had zero interactions with the application, resulting in a total of 27 users. For each user $u \in \{1, \ldots, 27\}$, we have the user's trajectory $(X_{u,1}, T_{u,1}, Y_{u,1}), \ldots, (X_{u,\ell_u}, T_{u,\ell_u}, Y_{u,\ell_u})$ of size ℓ_u (on average ℓ_u is 160), where the covariates $X_{u,i}$ are the contextual information about the user at time step i, the treatment $T_{u,i} \in \{0, 1\}$ is whether to deliver an activity suggestion, and the outcome $Y_{u,i}$ is the log transformation of the 30-minute step count after the decision time. In this analysis, we make Assumption 4 and consider the following approximation for the conditional mean of $Y_{u,i}$:

$$\alpha_u^\top g(X_{u,i}) + \beta_u^\top f(X_{u,i}) T_{u,i}, \tag{35}$$

where $g(X_{u,i})$ is a (known) baseline feature vector and $f(X_{u,i})$ is a (known) feature vector for the treatment effect. We allow the main effect to be misspecified. As for the feature vectors, we consider the same features (with minor modifications[8]) as in Liao et al. (2020); the vector $f(X_{u,i})$ contains Decision Bucket (DB) (bucketized decision time), Application Engagement (AE) (indicating how frequently users interact with the application), Location (LC) (indicating whether users are at home, at work or somewhere else) and Variation Indicator (VI) (the variation level of step count 60 minutes around the current time slot in past 7 days). The baseline vector $g(X_{u,i})$ contains $f(X_{u,i})$ along with the prior 30-minute step count, the previous day's total step count and the current temperature.

Since, for a given user, the outcome model (35) does not change over time, we can combine all users' trajectories and obtain the combined observations under multiple environments; that is, we have the dataset $D := (X_i, T_i, Y_i, e_i)_{i=1}^n$, with $n = \sum_{u \in \{1, \ldots, 27\}} \ell_u$, collected from multiple environments (users) $\mathcal{E}$, where $e_i \in \mathcal{E}$ for all $i \in \{1, \ldots, n\}$. In particular, we do not account for potential temporal dependencies that are not captured by the bucketized decision times. In practice, one may allow for dependence across time in the observations $(X_{u,i}, T_{u,i}, Y_{u,i})_{i=1}^{\ell_u}$ within each user, which we leave for future work, see Section 7.

6.3 Inferring e-invariant sets (HeartSteps V1)

We begin our analysis on the HeartSteps V1 data by conducting the Wald e-invariance test detailed in Algorithm 1 to find subsets of the treatment effect feature vector $f(X)$ that satisfy the e-invariant condition (5). As a comparison, we also apply the invariance test

---

8. We replace the dosage variable with the bucketized decision time variable to account for potential non-linear time dependency.

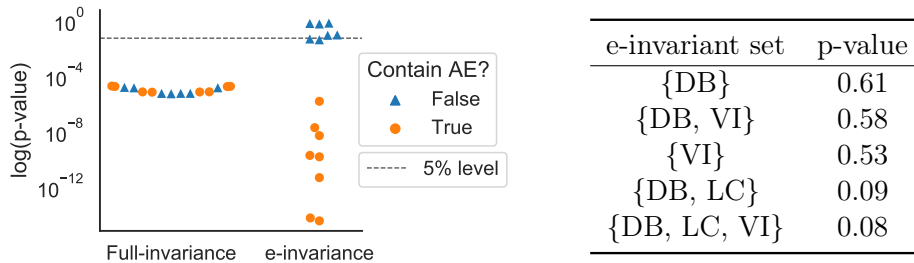| e-invariant set	p-value
{DB}	0.61
{DB, VI}	0.58
{VI}	0.53
{DB, LC}	0.09
{DB, LC, VI}	0.08

Figure 3: (Left) P-values for all subsets, considering the full- and e-invariance hypotheses. (Right) all five subsets for which we do not reject the e-invariance hypothesis.

proposed in Peters et al. (2016, Method II), which tests for a full-invariance instead of our proposed e-invariance (see Figure 1). Figure 3(Left) reports the p-values of all subsets for the full-invariance and e-invariance tests. The p-values for the full-invariance are all below the 5% level and hence there is no subset that satisfies the full-invariance hypothesis. However, we find several subsets that satisfy the e-invariance condition (those with p-values of the e-invariance hypothesis greater than the 5% level). Interestingly, all subsets that contain Application Engagement (AE) have p-values close to zero, suggesting that AE is a variable that renders the conditional treatment effect unstable between environments if included in the model. We report all the subsets for which we accept the e-invariance hypothesis at the 5% significance level in Figure 3(Right).

The above finding demonstrates that the relaxed notion of invariance that we propose can be beneficial in practice. The full-invariance condition may be too strict in that there is no full-invariant set. But if our goal is to learn a generalizable policy, it may suffice to test for the weaker notion of e-invariance, which the following section investigates using semi-real data.

6.4 Zero-shot generalization (augmented HeartSteps V1)

As the HeartSteps V1 study has been completed, it is not possible to implement and test a proposed policy on a new subject. In this section, we instead conduct a simulation study using HeartSteps V1 data to illustrate the use of e-invariance for zero-shot generalization, see Section 3. To evaluate the performance of a policy, we consider 'leave-one-environment-out' cross validation. Specifically, we first choose $e \in \mathcal{E}$ as a test environment (user) and split the dataset D into the test set $D^{\text{tst}} \coloneqq \{(X_i^{\text{tst}}, T_i^{\text{tst}}, Y_i^{\text{tst}}, e_i^{\text{tst}})\}_{i=1}^{n_{\text{tst}}}$ and the training set $D^{\text{tr}} \coloneqq \{(X_i^{\text{tr}}, T_i^{\text{tr}}, Y_i^{\text{tr}}, e_i^{\text{tr}})\}_{i=1}^{n_{\text{tr}}}$, where $e_i^{\text{tst}} = e$ and $e_i^{\text{tr}} \in \mathcal{E}^{\text{tr}} \coloneqq \mathcal{E} \setminus \{e\}$ for all i. We then conduct the training and testing procedure as follows.

Training phase: Using the training data D^{tr}, we find all sets that are not rejected by the Wald e-invariance test detailed in Algorithm 1. Using the inferred e-invariant sets, we then compute an estimate of $\pi^{\text{e-inv}}$ as discussed in Section 3.1, where we use the R-learner due to Nie and Wager (2021) as the CATE estimator – based on the implementation of the `econml` Python package (Battocchi et al., 2019). As a baseline, we include an optimal policy which utilizes all variables in $f(X)$ (denoted as 'full-set'). This baseline is computed by pooling all data from the training environments and fitting the R-learner CATE estimator

on the complete covariate set. Additionally, we include a uniformly random policy denoted as 'random' as another baseline for comparison.

To illustrate this procedure for the test user $e = 1$, consider the set of training users $\mathcal{E}^{\text{tr}} = \{2, \ldots, 27\}$. Using the observations from $\mathcal{E}^{\text{tr}}$, we apply Algorithm 1 to obtain the inferred e-invariant sets $\mathbf{S}^{\text{e-inv}}_{\mathcal{E}^{\text{tr}}} = \{\{\text{DB}\}, \{\text{DB}, \text{VI}\}, \{\text{VI}\}, \{\text{DB}, \text{LC}, \text{VI}\}\}$. For each $S \in \mathbf{S}^{\text{e-inv}}_{\mathcal{E}^{\text{tr}}}$, we then train a policy $\hat{\pi}^S$ as in (17) using the R-learner as the CATE estimator and choose an optimal $\hat{S}^*$ as in (18). We then use $\pi^{\hat{S}^*}$ as the final estimate of $\pi^{\text{e-inv}}$.

Testing phase: To perform policy evaluation, we create a semi-real test environment. To do so, we follow Liao et al. (2020). Given a test dataset D^{tst}, the value $V(\pi)$ of a policy $\pi \in \Pi$ is computed by the following procedure.

(1) Fit a regression model (35) on D^{tst}

$$Y_i^{\text{tst}} = \alpha_{\text{tst}}^\top g(X_i^{\text{tst}}) + \beta_{\text{tst}}^\top f(X_i^{\text{tst}}) T_i^{\text{tst}} + \epsilon_i, \tag{36}$$

and obtain pairs of covariates and residuals $\{(X_i^{\text{tst}}, \hat{\epsilon}_i)\}_{i=1}^{n_{\text{tst}}}$ and parameters $\hat{\alpha}_{\text{tst}}$ and $\hat{\beta}_{\text{tst}}$.

(2) Generate more pairs to obtain a total of 1000 observations $\{(\bar{X}_i^{\text{tst}}, \bar{\epsilon}_i)\}_{i=1}^{1000}$ by uniformly sampling with replacement from the original pairs.

(3) For each i, the treatment $\bar{T}_i^{\text{tst}}$ is selected based on the covariates $\bar{X}_i^{\text{tst}}$ according to π.

(4) For each i, the reward $\bar{Y}_i^{\text{tst}}$ is defined by

$$\bar{Y}_i^{\text{tst}} = \hat{\alpha}_{\text{tst}}^\top g(\bar{X}_i^{\text{tst}}) + \hat{\beta}_{\text{tst}}^\top f(\bar{X}_i^{\text{tst}}) \bar{T}_i + \bar{\epsilon}_i, \tag{37}$$

where the coefficients $\hat{\alpha}_{\text{tst}}$ and $\hat{\beta}_{\text{tst}}$ are obtained from the regression model fitted in step (1). The value is then given as the average reward: $\hat{V}(\pi) = \frac{1}{1000} \sum_{i=1}^{1000} \bar{Y}_i^{\text{tst}}$.

The performance of a policy π is then computed as $\hat{V}(\pi) - \hat{V}(\pi_0)$, where π_0 is the policy that always selects to not deliver a suggestion. This corresponds to an empirical version of the expected relative reward as in (9).

Figure 4(Left) shows the performance of different policies trained on the data available during training. Our proposed approach (e-inv) shows a slight improvement over the baseline approaches in terms of the mean and median performances over all users. Furthermore, as presented in Figure 4(Right), the e-invariance policy $\pi^{\text{e-inv}}$ yields higher relative reward comparing to the policy that uses all the variables in $f(X)$ in the majority of users (17 out of 27 users). We use the Wilcoxon signed-rank test (Wilcoxon, 1945) to compare the performance of the proposed e-inv policy with that of the full-set policy. It shows a p-value of 0.008, indicating that the improvement is statistically significant.

7. Conclusion and future work

This work addresses the challenge of adjusting for distribution shifts between environments in the context of policy learning. We propose an approach that leverages e-invariance, which is a relaxation of the full invariance assumption commonly used in causal inference
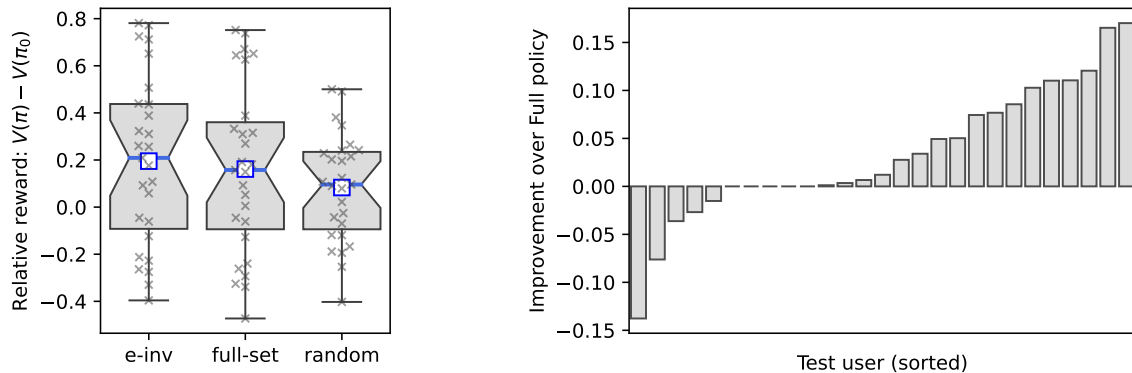
Figure 4: (Left) The out-of-environment (here: out-of-user) performance of different policies including the proposed e-invariance policy (e-inv), an optimal policy that uses all variables in $f(X)$ (full-set), and a uniformly random policy (random). (Right) Comparing the out-of-environment performance between the e-inv policy and the full-set policy for each test user. For the majority of test users, the e-inv policy outperforms the baseline (p-value 0.008).

literature. We show that despite being a weaker assumption, e-invariance is sufficient for building policies that generalize better to unseen environments compared to other policies. That is, under suitable assumptions, an optimal e-invariance policy is worst-case optimal. Additionally, we present a method for leveraging e-invariance information in the few-shot generalization setting, when a sample from the test environment is available.

To enable the practical use of e-invariance, we propose two testing procedures; one to test for e-invariance in linear and one in nonlinear model classes. Moreover, we validate the effectiveness of our policy learning methods through a semi-real-world case study in the domain of mobile health interventions. Our experiments show that an optimal policy based on an e-invariant set outperforms policies that rely on the complete context information when it comes to generalizing to new environments.

There are several promising directions for future research. It might be worthwhile to develop e-invariance testing procedures that can handle more complex temporal dependencies, especially when the data is collected from adaptive algorithms such as contextual bandit algorithms. Existing works have proposed inference methods to handle such scenarios (e.g., Zhang et al., 2021a; Hadad et al., 2021), but how to incorporate these methods effectively into our framework remains an open question.

Another interesting area of future work is how best to use the e-invariant set S^* (see (14)) in order to warm-start a contextual bandit algorithm. In the digital health field, one frequently conducts a series of optimization trials (each on a set of different users) in the process of optimizing a full digital health intervention. The data from each trial is used to inform the design of the subsequent trial. In the case of HeartSteps, 3 trials (V1, V2 and V3) were conducted beginning with HeartSteps V1. HeartSteps V2 & V3 deployed a Bayesian Thompson-Sampling algorithm (Russo et al., 2018; Liao et al., 2020) which uses a prior distribution on the parameters to warm-start the algorithm. Clearly the knowledge of an optimal e-invariant set S^* should guide the formation of the prior. Determining the most effective approach to achieve this is still an open question.

Lastly, our work also contributes to the field of causal inference by introducing a relaxation of the full invariance assumption. We believe that there are other scenarios where the full invariance assumption is too restrictive, and a relaxation of the assumption may be sufficient to address the task at hand. Further investigating the potential for relaxation in different causal inference settings would be a promising future research direction.

Acknowledgments

References

M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz. Invariant risk minimization. *ArXiv e-prints (1907.02893)*, 2019.

Elias Bareinboim and Judea Pearl. Transportability from multiple environments with limited experiments: Completeness results. *Advances in neural information processing systems*, 27, 2014.

Keith Battocchi, Eleanor Dillon, Maggie Hei, Greg Lewis, Paul Oka, Miruna Oprescu, and Vasilis Syrgkanis. EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation. https://github.com/microsoft/EconML, 2019. Version 0.x.

S. Bongers, P. Forre, J. Peters, and J. M. Mooij. Foundations of structural causal models with cycles and latent variables. *Annals of Statistics*, 49(5):2885–2915, 2021.

Dennis D Boos, Leonard A Stefanski, et al. *Essential statistical inference*. Springer, 2013.

Audrey Boruvka, Daniel Almirall, Katie Witkiewitz, and Susan A Murphy. Assessing time-varying causal effect moderation in mobile health. *Journal of the American Statistical Association*, 113(523):1112–1121, 2018.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Rich Caruana. Multitask learning. *Machine learning*, 28(1):41–75, 1997.

Rune Christiansen, Niklas Pfister, Martin Emil Jakobsen, Nicola Gnecco, and Jonas Peters. A causal framework for distribution generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2021.

Koby Crammer, Michael Kearns, and Jennifer Wortman. Learning from multiple sources. *Journal of Machine Learning Research*, 9(8), 2008.

Philip Dawid. Decision-theoretic foundations for statistical causality. *Journal of Causal Inference*, 9(1):39–77, 2021.

Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

Zijian Guo and Peter Bühlmann. Two stage curvature identification with machine learning: Causal inference with possibly invalid instrumental variables. *arXiv preprint arXiv:2203.12808*, 2022.

T. Haavelmo. The probability approach in econometrics. *Econometrica*, 12:S1–S115 (supplement), 1944.

Vitor Hadad, David A Hirshberg, Ruohan Zhan, Stefan Wager, and Susan Athey. Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the national academy of sciences*, 118(15):e2014602118, 2021.

C. Heinze-Deml, J. Peters, and N. Meinshausen. Invariant causal prediction for nonlinear models. *Journal of Causal Inference*, 6(2):1–35, 2018.

Peter J Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Weather Modification; University of California Press: Berkeley, CA, USA*, page 221, 1967.

M. Jakobsen and J. Peters. Distributional robustness of K-class estimators and the PULSE. *The Econometrics Journal*, 25(2):404–432, 2022.

Edward H Kennedy. Towards optimal doubly robust estimation of heterogeneous causal effects. *arXiv preprint arXiv:2004.14497*, 2020.

Predrag Klasnja, Shawna Smith, Nicholas J Seewald, Andy Lee, Kelly Hall, Brook Luers, Eric B Hekler, and Susan A Murphy. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of heartsteps. *Annals of Behavioral Medicine*, 53(6):573–582, 2019.

Peng Liao, Kristjan Greenewald, Predrag Klasnja, and Susan Murphy. Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–22, 2020.

Anton Rask Lundborg, Ilmun Kim, Rajen D Shah, and Richard J Samworth. The projected covariance measure for assumption-lean variable significance testing. *arXiv preprint arXiv:2211.02039*, 2022.

S. Magliacane, T. van Ommen, T. Claassen, S. Bongers, P. Versteeg, and J. M. Mooij. Domain adaptation by using causal inference to predict invariant conditional distributions. In *Advances in Neural Information Processing Systems 31 (NeurIPS)*, pages 10846–10856. Curran Associates, Inc., 2018.

Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *Proceedings of the 30th International Conference on Machine Learning*, pages 10–18. PMLR, 2013.

Jerzy Neyman. Optimal asymptotic tests of composite statistical hypotheses. *Probability and statsitics*, page 416–444, 1959.

Jerzy Neyman. C (α) tests and their use. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 1–21, 1979.

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

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, USA, 2nd edition, 2009.

Judea Pearl and Elias Bareinboim. Transportability of causal and statistical relations: A formal approach. In *Twenty-fifth AAAI conference on artificial intelligence*, 2011.

J. Peters, P. Bühlmann, and N. Meinshausen. Causal inference using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (with discussion)*, 78(5):947–1012, 2016.

N. Pfister, P. Bühlmann, and J. Peters. Invariant causal prediction for sequential data. *Journal of the American Statistical Association*, 114(527):1264–1276, 2018.

Niklas Pfister, Evan G Williams, Jonas Peters, Ruedi Aebersold, and Peter Bühlmann. Stabilizing variable selection and regression. *The Annals of Applied Statistics*, 15(3): 1220–1246, 2021.

Thomas S Richardson and James M Robins. Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128 (30):2013, 2013.

Mateo Rojas-Carulla, Bernhard Schölkopf, Richard Turner, and Jonas Peters. Invariant models for causal transfer learning. *The Journal of Machine Learning Research*, 19(1): 1309–1342, 2018.

Dominik Rothenhäusler, Nicolai Meinshausen, Peter Bühlmann, and Jonas Peters. Anchor regression: Heterogeneous data meet causality. *Journal of the Royal Statistical Society: Series B*, 83(2):215–246, 2021.

Daniel J Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, Zheng Wen, et al. A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning*, 11(1): 1–96, 2018.

Sorawit Saengkyongam, Nikolaj Thams, Jonas Peters, and Niklas Pfister. Invariant policy learning: A causal perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

Cyrill Scheidegger, Julia Hörrmann, and Peter Bühlmann. The weighted generalised covariance measure. *Journal of Machine Learning Research*, 23(273):1–68, 2022.

B. Schölkopf, D. Janzing, J. Peters, E. Sgouritsa, K. Zhang, and J. M. Mooij. On causal and anticausal learning. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*. Omnipress, 2012.

R. Shah and J. Peters. The hardness of conditional independence testing and the generalised covariance measure. *Annals of Statistics*, 48(3):1514–1538, 2020.

Anoopkumar Sonar, Vincent Pacelli, and Anirudha Majumdar. Invariant policy optimization: Towards stronger generalization in reinforcement learning. In *Learning for Dynamics and Control*, pages 21–33. PMLR, 2021.

Adarsh Subbaswamy, Peter Schulam, and Suchi Saria. Preventing failures due to dataset shift: Learning predictive models that transport. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3118–3127. PMLR, 2019.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.

Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and Philip Yu. Generalizing to unseen domains: A survey on domain generalization. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

Halbert White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, pages 817–838, 1980.

Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6): 80–83, 1945.

Amy Zhang, Clare Lyle, Shagun Sodhani, Angelos Filos, Marta Kwiatkowska, Joelle Pineau, Yarin Gal, and Doina Precup. Invariant causal prediction for block MDPs. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 11214–11224. PMLR, 2020.

Kelly Zhang, Lucas Janson, and Susan Murphy. Statistical inference with m-estimators on adaptively collected data. *Advances in neural information processing systems*, 34:7460–7471, 2021a.

Weijia Zhang, Jiuyong Li, and Lin Liu. A unified survey of treatment effect heterogeneity modelling and uplift modelling. *ACM Computing Surveys (CSUR)*, 54(8):1–36, 2021b.

Appendix A. Proofs

A.1 Proof of Proposition 4

Proof We split the proof into three parts. First (Part 1), we show that the expected outcome function can be decomposed into an effect-modification term that depends on the

treatment and a main-effect term that does not depend on the treatment. We then proceed and prove the 'only if' part of the main result in Part 2 and the 'if' part in Part 3.

Part 1: We show the following lemma.

Lemma 12 *Assume Setting 1. Let $S \subseteq \{1, \ldots, d\}$ be an arbitrary subset and $t_0 \in \mathcal{T}$ be the baseline treatment. Then, there exists a pair of functions $\kappa_S : \mathcal{X}^S \times \mathcal{T} \times \mathcal{E} \to \mathbb{R}$ and $\nu_S : \mathcal{X}^S \times \mathcal{E} \to \mathbb{R}$ such that*

$$\forall e \in \mathcal{E}, \forall x \in \mathcal{X}^S, \forall t \in \mathcal{T} : \mathbb{E}^{e,\pi_t}[Y|X^S = x] = \mathbb{1}(t \neq t_0)\kappa_S(x, t, e) + \nu_S(x, e). \tag{38}$$

Proof Fix $e \in \mathcal{E}$ and $t \in \mathcal{T}$, and define $\delta_S(\cdot, t, e) \coloneqq \mathbb{E}^{e,\pi_t}[Y \mid X^S = \cdot]$ and $\nu_S(\cdot, e) \coloneqq \mathbb{E}^{e,\pi_{t_0}}[Y \mid X^S = \cdot]$. It then holds for all $x \in \mathcal{X}^S$ that

$$\mathbb{E}^{e,\pi_t}[Y \mid X^S = x] = \mathbb{1}(t \neq t_0)(\delta_S(x, t, e) - \nu_S(x, e)) + \nu_S(x, e).$$

We then define $\kappa_S(\cdot, t, e) \coloneqq \delta_S(\cdot, t, e) - \nu_S(\cdot, e)$, which concludes the proof. ■

Part 2: Assume a subset $S \subseteq \{1, \ldots, d\}$ is e-invariant w.r.t. $\mathcal{E}^\diamond$. Fix $e_0 \in \mathcal{E}^\diamond$ as a reference environment. By Lemma 12, there exists a pair of functions $\kappa_S : \mathcal{X}^S \times \mathcal{T} \times \mathcal{E} \to \mathbb{R}$ and $\nu_S : \mathcal{X}^S \times \mathcal{E} \to \mathbb{R}$ such that for all $e \in \mathcal{E}^\diamond$, $x \in \mathcal{X}^S$ and $t \in \mathcal{T}$

$$\tau_e^S(x, t) = (\mathbb{1}(t \neq t_0)\kappa_S(x, t, e) + \nu_S(x, e)) - \nu_S(x, e) = \mathbb{1}(t \neq t_0)\kappa_S(x, t, e).$$

Next, we define the function $\tilde{\psi}_S : \mathcal{X}^S \times \mathcal{T} \to \mathbb{R}$ for all $x \in \mathcal{X}^S$ and $t \in \mathcal{T}$ by

$$\tilde{\psi}_S(x, t) \coloneqq \kappa_S(x, t, e_0).$$

Now, since S is e-invariant w.r.t. $\mathcal{E}^\diamond$ it holds for all $\forall e \in \mathcal{E}^\diamond$, $x \in \mathcal{X}^S$ and $t \in \mathcal{T}$ that

$$\mathbb{1}(t \neq t_0)\tilde{\psi}_S(x, t) = \mathbb{1}(t \neq t_0)\kappa_S(x, t, e). \tag{39}$$

Then, combining (38) and (39) implies that (6) is true.

Part 3: Assume (6) holds for a subset $S \subseteq \{1, \ldots, d\}$. It then holds for all $e, h \in \mathcal{E}^\diamond$, $x \in \mathcal{X}^S$ and $t \in \mathcal{T}$ that

$$\tau_e^S(x, t) - \tau_h^S(x, t)$$
$$= (\psi_S(x, t) + \nu_S(x, e) - \psi_S(x, t_0) - \nu_S(x, e)) - (\psi_S(x, t) + \nu_S(x, h) - \psi_S(x, t_0) - \nu_S(x, h))$$
$$= 0,$$

which proves that S is e-invariant w.r.t. $\mathcal{E}^\diamond$. ■

A.2 Proof of Proposition 6

From the SCM (7), we have for all $x \in \mathcal{X}^{\mathrm{PA}_{f,X}}$ that

$$\mathbb{E}^{e,\pi_t}[Y \mid X^{\mathrm{PA}_{f,X}} = x] = \mathbb{E}^e[f(X^{\mathrm{PA}_{f,X}}, U^{\mathrm{PA}_{f,U}}, t) \mid X^{\mathrm{PA}_{f,X}} = x] + \mathbb{E}^e[g_e(X, U, \epsilon_Y) \mid X^{\mathrm{PA}_{f,X}} = x]$$

Using the assumption (i) that $U^{\mathrm{PA}_{f,U}} \perp\!\!\!\perp X^{\mathrm{PA}_{f,X}}$ in $\mathbb{P}^e_{X,U}$ for all $e \in \mathcal{E}^\diamond$, we have

$$\mathbb{E}^{e,\pi_t}[Y \mid X^{\mathrm{PA}_{f,X}} = x] = \mathbb{E}^e[f(x, U^{\mathrm{PA}_{f,U}}, t)] + \mathbb{E}^e[g_e(X, U, \epsilon_Y) \mid X^{\mathrm{PA}_{f,X}} = x], \tag{40}$$

where a formal proof for the equality (40) is given, for example, in (Durrett, 2019, Example 4.1.7). Next, using the assumption (ii) that $\mathbb{P}^e_{U^{\mathrm{PA}_{f,U}}}$ are identical across $e \in \mathcal{E}^\diamond$, we can drop the dependency on e from the component $\mathbb{E}^e[f(x, U^{\mathrm{PA}_{f,U}}, t)]$ in (40) and have that

$$\begin{aligned}
\mathbb{E}^{e,\pi_t}[Y \mid X^{\mathrm{PA}_{f,X}} = x] &= \mathbb{E}[f(x, U^{\mathrm{PA}_{f,U}}, t)] + \mathbb{E}^e[g_e(X, U, \epsilon_Y) \mid X^{\mathrm{PA}_{f,X}} = x] \\
&= \mathbb{1}(t \neq t_0)\big(\mathbb{E}[f(x, U^{\mathrm{PA}_{f,U}}, t)] - \mathbb{E}[f(x, U^{\mathrm{PA}_{f,U}}, t_0)]\big) \\
&\quad + \mathbb{E}[f(x, U^{\mathrm{PA}_{f,U}}, t_0)] + \mathbb{E}^e[g_e(X, U, \epsilon_Y) \mid X^{\mathrm{PA}_{f,X}} = x].
\end{aligned}$$

Defining $\psi_S : (x, t) \mapsto \mathbb{1}(t \neq t_0)\big(\mathbb{E}[f(x, U^{\mathrm{PA}_{f,U}}, t)] - \mathbb{E}[f(x, U^{\mathrm{PA}_{f,U}}, t_0)]\big)$ and $\nu_S : (x, e) \mapsto \mathbb{E}[f(x, U^{\mathrm{PA}_{f,U}}, t_0)] + \mathbb{E}^e[g_e(X, U, \epsilon_Y) \mid X^{\mathrm{PA}_{f,X}} = x]$, we have that $\psi_S(\cdot, t_0) \equiv 0$ and

$$\mathbb{E}^{e,\pi_t}[Y \mid X^{\mathrm{PA}_{f,X}} = x] = \psi_S(x, t) + \nu_S(x, e).$$

Thus, by Proposition 4, $\mathrm{PA}_{f,X}$ is e-invariant w.r.t. $\mathcal{E}^\diamond$.

A.3 Proof of Proposition 8

Proof We begin with the proof of the first statement (i) of Proposition 8. First, we show that the collections of policies $(\Pi^S_{\mathrm{opt}})_{S \in \mathbf{S}^{\text{e-inv}}_{\mathcal{E}^{\mathrm{tr}}}}$ are identifiable from Q^{tr}_i (for an arbitrary $1 \leq i \leq n$). Fix (an arbitrary) $f \in \mathcal{E}^{\mathrm{tr}}$ and $S \in \mathbf{S}^{\text{e-inv}}_{\mathcal{E}^{\mathrm{tr}}}$. Let $\pi^S \in \Pi^S_{\mathrm{opt}}$. Then, π^S satisfies

$$\pi^S(t|x) > 0 \implies t \in \operatorname*{argmax}_{t' \in \mathcal{T}} \frac{1}{|\mathcal{E}^{\mathrm{tr}}|} \sum_{e \in \mathcal{E}^{\mathrm{tr}}} \tau^S_e(x^S, t') = \operatorname*{argmax}_{t' \in \mathcal{T}} \tau^S_f(x^S, t'). \tag{41}$$

Thus, the identifiability of Π^S_{opt} depends on the identifiability of τ^S_f.

Fix $i \in \{1, \ldots, n\}$. Recall that in Setting 1 we assume $\forall x \in \mathcal{X}, t \in \mathcal{T} : \pi_i(t \mid x) > 0$. It then holds for all $x \in \mathcal{X}^S$ and $t \in \mathcal{T}$ that

$$\begin{aligned}
\mathbb{E}^{e_i,\pi_t}[Y_i \mid X^S_i = x] &= \mathbb{E}^{e_i}\big[\mathbb{E}^{e_i,\pi_t}[Y_i \mid X_i] \mid X^S_i = x\big] \\
&\overset{(*)}{=} \mathbb{E}^{e_i}\big[\mathbb{E}^{e_i,\pi_i}\big[\tfrac{\mathbb{1}(T_i=t)}{\pi_i(T_i|X_i)} Y_i \mid X_i\big] \mid X^S_i = x\big] \\
&= \mathbb{E}^{e_i,\pi_i}\big[\tfrac{\mathbb{1}(T_i=t)}{\pi_i(t|X_i)} Y_i \mid X^S_i = x\big], \tag{42}
\end{aligned}$$

where the equality $(*)$ holds by definition of π_t. Since the right-hand side of (42) is the expectation w.r.t. Q^{tr}_i, the quantity $\mathbb{E}^{e_i,\pi_t}[Y_i \mid X^S_i = x]$ is identifiable from Q^{tr}_i.

Next, we have

$$\begin{aligned}
\mathbb{E}^{e_i,\pi_t}[Y_i \mid X^S_i = x] - \mathbb{E}^{e_i,\pi_0}[Y_i \mid X^S_i = x] &= \mathbb{E}^{e_i,\pi_t}[Y \mid X^S = x] - \mathbb{E}^{e_i,\pi_0}[Y \mid X^S = x] \\
&= \tau^S_{e_i}(x, t) \\
&= \tau^S_f(x, t) \qquad \text{since } S \in \mathbf{S}^{\text{e-inv}}_{\mathcal{E}^{\mathrm{tr}}}. \tag{43}
\end{aligned}$$

From (43), we then have that τ_f^S is identifiable from Q_i^{tr} and therefore Π_{opt}^S is identifiable from Q_i^{tr}. Consequently, the collection

$$A := \operatorname*{argmax}_{S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}} \mathbb{E}^{e^{\mathrm{tst}}} \left[\sum_{t \in \mathcal{T}} \tau_e^S(X^S, t) \pi^S(t \mid X) \right]$$

is identifiable from Q_i^{tr} and Q^{tst}.

Next, we show the proof of the second statement (ii) of Proposition 8. Fix $e^{\mathrm{tst}} \in \mathcal{E}$, (an arbitrary) $f \in \mathcal{E}^{\mathrm{tr}}$ and let $S^* \subseteq \{1, \ldots, d\}$ be a subset that satisfies

$$S^* \in \operatorname*{argmax}_{S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}} \mathbb{E}^{e^{\mathrm{tst}}} \left[\sum_{t \in \mathcal{T}} \tau_f^S(X^S, t) \pi^S(t \mid X^S) \right].$$

Next, we recall the definition $\Pi_{\mathrm{opt}}^{\mathrm{e\text{-}inv}} = \{\pi \in \Pi \mid \exists S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}} \text{ s.t. } \pi \in \Pi_{\mathrm{opt}}^S\}$ and let $\pi_{e^{\mathrm{tst}}}^{\diamond} \in \operatorname{argmax}_{\pi \in \Pi_{\mathrm{opt}}^{\mathrm{e\text{-}inv}}} \mathbb{E}^{e^{\mathrm{tst}}, \pi}[Y]$. Then, using that $\pi_{e^{\mathrm{tst}}}^{\diamond} \in \Pi_{\mathrm{opt}}^{\mathrm{e\text{-}inv}}$, choose $S^{\diamond} \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}$ such that $\pi_{e^{\mathrm{tst}}}^{\diamond} \in \Pi_{\mathrm{opt}}^{S^{\diamond}}$. We have

$$
\begin{aligned}
\mathbb{E}^{e^{\mathrm{tst}}, \pi_{e^{\mathrm{tst}}}^{\diamond}}[Y] - \mathbb{E}^{e^{\mathrm{tst}}, \pi_{t_0}}[Y] &= \mathbb{E}^{e^{\mathrm{tst}}} \Big[\sum_{t \in \mathcal{T}} (\mathbb{E}^{e^{\mathrm{tst}}, \pi_t}[Y \mid X^{S^{\diamond}}] - \mathbb{E}^{e^{\mathrm{tst}}, \pi_{t_0}}[Y \mid X^{S^{\diamond}}]) \pi_{e^{\mathrm{tst}}}^{\diamond}(t \mid X^{S^{\diamond}}) \Big] \\
&= \mathbb{E}^{e^{\mathrm{tst}}} \Big[\sum_{t \in \mathcal{T}} \tau_{e^{\mathrm{tst}}}^{S^{\diamond}}(X^{S^{\diamond}}, t) \pi_{e^{\mathrm{tst}}}^{\diamond}(t \mid X^{S^{\diamond}}) \Big] \\
&= \mathbb{E}^{e^{\mathrm{tst}}} \Big[\sum_{t \in \mathcal{T}} \tau_f^{S^{\diamond}}(X^{S^{\diamond}}, t) \pi_{e^{\mathrm{tst}}}^{\diamond}(t \mid X^{S^{\diamond}}) \Big] && \text{by Assumption 2} \\
&= \mathbb{E}^{e^{\mathrm{tst}}} \Big[\sum_{t \in \mathcal{T}} \tau_f^{S^{\diamond}}(X^{S^{\diamond}}, t) \pi^{S^{\diamond}}(t \mid X^{S^{\diamond}}) \Big] && \text{since } \pi_{e^{\mathrm{tst}}}^{\diamond} \in \Pi_{\mathrm{opt}}^{S^{\diamond}} \\
&\leq \max_{S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}} \mathbb{E}^{e^{\mathrm{tst}}} \Big[\sum_{t \in \mathcal{T}} \tau_f^S(X^S, t) \pi^S(t \mid X^S) \Big] \\
&= \mathbb{E}^{e^{\mathrm{tst}}} \Big[\sum_{t \in \mathcal{T}} \tau_f^{S^*}(X^{S^*}, t) \pi^{S^*}(t \mid X^{S^*}) \Big] \\
&= \mathbb{E}^{e^{\mathrm{tst}}} \Big[\sum_{t \in \mathcal{T}} \tau_{e^{\mathrm{tst}}}^{S^*}(X^{S^*}, t) \pi^{S^*}(t \mid X^{S^*}) \Big] && \text{by Assumption 2} \\
&= \mathbb{E}^{e^{\mathrm{tst}}} \Big[\sum_{t \in \mathcal{T}} (\mathbb{E}^{e^{\mathrm{tst}}, \pi_t}[Y \mid X^{S^*}] - \mathbb{E}^{e^{\mathrm{tst}}, \pi_{t_0}}[Y \mid X^{S^*}]) \pi^{S^*}(t \mid X^{S^*}) \Big] \\
&= \mathbb{E}^{e^{\mathrm{tst}}, \pi^{S^*}}[Y] - \mathbb{E}^{e^{\mathrm{tst}}, \pi_{t_0}}[Y].
\end{aligned}
$$

We therefore have that

$$\mathbb{E}^{e^{\mathrm{tst}}, \pi_{e^{\mathrm{tst}}}^{\diamond}}[Y] \leq \mathbb{E}^{e^{\mathrm{tst}}, \pi^{S^*}}[Y],$$

which concludes the proof of the second statement (ii) of Proposition 8.

$\blacksquare$

A.4 Proof of Theorem 9

Proof Let $e^{\text{tst}} \in \mathcal{E}$ be a test environment, $\pi^{\text{e-inv}}$ be a policy satisfying (11) and, for all $S \subseteq \{1, \ldots, d\}$, Π_{opt}^S be the set policies satisfying (10).

We now prove the first statement, see Theorem 9(i). By definition, there exists $S^* \in \mathbf{S}_{\mathcal{E}^{\text{tr}}}^{\text{e-inv}}$ such that $\pi^{\text{e-inv}} \in \Pi_{\text{opt}}^{S^*}$. It then holds that

$$\mathbb{E}^{e^{\text{tst}}, \pi^{\text{e-inv}}}[Y] - \mathbb{E}^{e^{\text{tst}}, \pi_{t_0}}[Y] = \mathbb{E}^{e^{\text{tst}}}\Big[\sum_{t \in \mathcal{T}}(\mathbb{E}^{e^{\text{tst}}, \pi_t}[Y \mid X^{S^*}] - \mathbb{E}^{e^{\text{tst}}, \pi_{t_0}}[Y \mid X^{S^*}])\pi^{\text{e-inv}}(t \mid X^{S^*})\Big]$$

$$= \mathbb{E}^{e^{\text{tst}}}\Big[\sum_{t \in \mathcal{T}} \tau_{e^{\text{tst}}}^{S^*}(X^{S^*}, t)\pi^{\text{e-inv}}(t \mid X^{S^*})\Big].$$

Fix $e^{\text{tr}} \in \mathcal{E}^{\text{tr}}$. We have

$$\mathbb{E}^{e^{\text{tst}}, \pi^{\text{e-inv}}}[Y] - \mathbb{E}^{e^{\text{tst}}, \pi_{t_0}}[Y] = \mathbb{E}^{e^{\text{tst}}}\Big[\sum_{t \in \mathcal{T}} \tau_{e^{\text{tst}}}^{S^*}(X^{S^*}, t)\pi^{\text{e-inv}}(t \mid X^{S^*})\Big]$$

$$= \mathbb{E}^{e^{\text{tst}}}\Big[\sum_{t \in \mathcal{T}} \tau_{e^{\text{tr}}}^{S^*}(X^{S^*}, t)\pi^{\text{e-inv}}(t \mid X^{S^*})\Big] \quad \text{by Assumption 2}$$

$$= \mathbb{E}^{e^{\text{tst}}}\Big[\max_{t \in \mathcal{T}} \tau_{e^{\text{tr}}}^{S^*}(X^{S^*}, t)\Big] \quad \text{by the definition of } \Pi_{\text{opt}}^{S^*}$$

$$= \mathbb{E}^{e^{\text{tst}}}\Big[\max_{t \in \mathcal{T}} \tau_{e^{\text{tst}}}^{S^*}(X^{S^*}, t)\Big] \quad \text{by Assumption 2}$$

$$\geq \max_{t \in \mathcal{T}}(\mathbb{E}^{e^{\text{tst}}}[\tau_{e^{\text{tst}}}^{S^*}(X^{S^*}, t)])$$

$$= \max_{t \in \mathcal{T}}(\mathbb{E}^{e^{\text{tst}}, \pi_t}[Y] - \mathbb{E}^{e^{\text{tst}}, \pi_{t_0}}[Y]) \quad \text{by the tower property}$$

This implies,

$$\mathbb{E}^{e^{\text{tst}}, \pi^{\text{e-inv}}}[Y] \geq \max_{t \in \mathcal{T}} \mathbb{E}^{e^{\text{tst}}, \pi_t}[Y], \tag{44}$$

which concludes the proof of Theorem 9(i). Next, we prove the second statement, see Theorem 9(ii). Recall that $[e^{\text{tst}}] := \{e \in \mathcal{E} \mid \mathbb{P}_X^e = Q_X^{\text{tst}}\}$. From Assumption 3, there exists an environment $f \in [e^{\text{tst}}]$ and $S^\diamond \in \mathbf{S}_{\mathcal{E}}^{\text{e-inv}}$ such that

$$\forall x \in \mathcal{X} : \max_{t \in \mathcal{T}} \tau_f(x, t) = \max_{t \in \mathcal{T}} \tau_f^{S^\diamond}(x^{S^\diamond}, t). \tag{45}$$

We have for all $S \in \mathbf{S}_{\mathcal{E}^{\text{tr}}}^{\text{e-inv}}$ that

$$\mathbb{E}^f[\max_{t \in \mathcal{T}} \tau_f^{S^\diamond}(X^{S^\diamond}, t)] = \mathbb{E}^f[\max_{t \in \mathcal{T}} \tau_f(X, t)] \quad \text{from (45)}$$

$$= \mathbb{E}^f[\mathbb{E}^f[\max_{t \in \mathcal{T}} \tau_f(X, t) \mid X^S]]$$

$$\geq \mathbb{E}^f[\max_{t \in \mathcal{T}} \mathbb{E}^f[\tau_f(X, t) \mid X^S]]$$

$$= \mathbb{E}^f[\max_{t \in \mathcal{T}} \tau_f^S(X^S, t)]. \tag{46}$$

Now, we have for all $e \in [e^{\text{tst}}]$ and for all $S \in \mathbf{S}_{\mathcal{E}^{\text{tr}}}^{\text{e-inv}}$

$$
\begin{aligned}
\mathbb{E}^e[\max_{t \in \mathcal{T}} \tau_e^{S^\diamond}(X^{S^\diamond}, t)] &= \mathbb{E}^f[\max_{t \in \mathcal{T}} \tau_f^{S^\diamond}(X^{S^\diamond}, t)] && \text{by Assumption 2 and } f \in [e^{\text{tst}}] \\
&\geq \mathbb{E}^f[\max_{t \in \mathcal{T}} \tau_f^S(X^S, t)] && \text{from (46)} \\
&= \mathbb{E}^e[\max_{t \in \mathcal{T}} \tau_e^S(X^S, t)] && \text{by Assumption 2 and } f \in [e^{\text{tst}}]. && (47)
\end{aligned}
$$

Next, we recall the definition $\Pi_{\text{opt}}^{\text{e-inv}} = \{\pi \in \Pi \mid \exists S \in \mathbf{S}_{\mathcal{E}^{\text{tr}}}^{\text{e-inv}} \text{ s.t. } \pi \in \Pi_{\text{opt}}^S\}$ and that $\pi^{\text{e-inv}} \in \text{argmax}_{\pi \in \Pi_{\text{opt}}^{\text{e-inv}}} \mathbb{E}^{e^{\text{tst}}, \pi}[Y]$. Then there exists $S^* \in \mathbf{S}_{\mathcal{E}^{\text{tr}}}^{\text{e-inv}}$ such that $\pi^{\text{e-inv}} \in \Pi_{\text{opt}}^{S^*}$. We therefore have for all $e \in [e^{\text{tst}}]$ that

$$
\begin{aligned}
\mathbb{E}^{e, \pi^{\text{e-inv}}}[Y] - \mathbb{E}^{e, \pi_{t_0}}[Y] &= \mathbb{E}^e[\sum_{t \in \mathcal{T}} (\mathbb{E}^{e, \pi_t}[Y \mid X^{S^*}] - \mathbb{E}^{e, \pi_{t_0}}[Y \mid X^{S^*}]) \pi^{\text{e-inv}}(t \mid X^{S^*})] \\
&= \mathbb{E}^e[\sum_{t \in \mathcal{T}} \tau_e^{S^*}(X^{S^*}, t) \pi^{\text{e-inv}}(t \mid X^{S^*})]. && (48)
\end{aligned}
$$

Fix $e^{\text{tr}} \in \mathcal{E}^{\text{tr}}$, we have for all $e \in [e^{\text{tst}}]$

$$
\begin{aligned}
\mathbb{E}^{e, \pi^{\text{e-inv}}}[Y] - \mathbb{E}^{e, \pi_{t_0}}[Y] &= \mathbb{E}^e[\sum_{t \in \mathcal{T}} \tau_{e^{\text{tr}}}^{S^*}(X^{S^*}, t) \pi^{\text{e-inv}}(t \mid X^{S^*})] && \text{by (48) and Assumption 2} \\
&= \mathbb{E}^e[\max_{t \in \mathcal{T}} \tau_{e^{\text{tr}}}^{S^*}(X^{S^*}, t)] && \text{by the definition of } \Pi_{\text{opt}}^{S^*} \\
&= \mathbb{E}^e[\max_{t \in \mathcal{T}} \tau_e^{S^*}(X^{S^*}, t)] && \text{by Assumption 2.} && (49)
\end{aligned}
$$

Let $\pi^{S^\diamond} \in \Pi_{\text{opt}}^{S^\diamond}$, we then have that

$$
\begin{aligned}
\mathbb{E}^{e^{\text{tst}}}[\max_{t \in \mathcal{T}} \tau_{e^{\text{tst}}}^{S^*}(X^{S^*}, t)] &= \mathbb{E}^{e^{\text{tst}}, \pi^{\text{e-inv}}}[Y] - \mathbb{E}^{e^{\text{tst}}, \pi_{t_0}}[Y] && \text{from (49)} \\
&\geq \mathbb{E}^{e^{\text{tst}}, \pi^{S^\diamond}}[Y] - \mathbb{E}^{e^{\text{tst}}, \pi_{t_0}}[Y] \\
&= \mathbb{E}^{e^{\text{tst}}}[\sum_{t \in \mathcal{T}} (\mathbb{E}^{e^{\text{tst}}, \pi_t}[Y \mid X^{S^\diamond}] - \mathbb{E}^{e^{\text{tst}}, \pi_{t_0}}[Y \mid X^{S^\diamond}]) \pi^{S^\diamond}(t \mid X^{S^\diamond})] \\
&= \mathbb{E}^{e^{\text{tst}}}[\sum_{t \in \mathcal{T}} \tau_{e^{\text{tst}}}^{S^\diamond}(X^{S^\diamond}, t) \pi^{S^\diamond}(t \mid X^{S^\diamond})] \\
&= \mathbb{E}^{e^{\text{tst}}}[\sum_{t \in \mathcal{T}} \tau_{e^{\text{tr}}}^{S^\diamond}(X^{S^\diamond}, t) \pi^{S^\diamond}(t \mid X^{S^\diamond})] && \text{by Assumption 2} \\
&= \mathbb{E}^{e^{\text{tst}}}[\max_{t \in \mathcal{T}} \tau_{e^{\text{tr}}}^{S^\diamond}(X^{S^\diamond}, t)] && \text{by the definition of } \Pi_{\text{opt}}^{S^\diamond} \\
&= \mathbb{E}^{e^{\text{tst}}}[\max_{t \in \mathcal{T}} \tau_{e^{\text{tst}}}^{S^\diamond}(X^{S^\diamond}, t)], && \text{by Assumption 2} && (50)
\end{aligned}
$$

where the above inequality holds because

$$
\pi^{\text{e-inv}} \in \underset{\pi \in \Pi_{\text{opt}}^{\text{e-inv}}}{\text{argmax}} \mathbb{E}^{e^{\text{tst}}, \pi}[Y] = \underset{\pi \in \Pi_{\text{opt}}^{\text{e-inv}}}{\text{argmax}} \mathbb{E}^{e^{\text{tst}}, \pi}[Y] - \mathbb{E}^{e^{\text{tst}}, \pi_{t_0}}[Y].
$$

Combining the two inequalities (50) and (47), we then have that

$$
\mathbb{E}^{e^{\text{tst}}}[\max_{t \in \mathcal{T}} \tau_{e^{\text{tst}}}^{S^*}(X^{S^*}, t)] = \mathbb{E}^{e^{\text{tst}}}[\max_{t \in \mathcal{T}} \tau_{e^{\text{tst}}}^{S^\diamond}(X^{S^\diamond}, t)]. \tag{51}
$$

Then, from (51) and since $e^{\text{tst}} \in [e^{\text{tst}}]$ and $S^\diamond$ is e-invariant w.r.t. $\mathcal{E}^{\text{tr}} \cup [e^{\text{tst}}]$ (by Assumption 2), we have for all $e \in [e^{\text{tst}}]$

$$\mathbb{E}^e[\max_{t\in\mathcal{T}} \tau_e^{S^*}(X^{S^*}, t)] = \mathbb{E}^e[\max_{t\in\mathcal{T}} \tau_e^{S^\diamond}(X^{S^\diamond}, t)]. \tag{52}$$

We are now ready to prove the main statement of Theorem 9(ii).

$$\begin{aligned}
V^{[e^{\text{tst}}]}(\pi^{\text{e-inv}}) &= \inf_{e\in[e^{\text{tst}}]} (\mathbb{E}^{e,\pi^{\text{e-inv}}}[Y] - \mathbb{E}^{e,\pi_{t_0}}[Y]) \\
&= \inf_{e\in[e^{\text{tst}}]} \mathbb{E}^e[\max_{t\in\mathcal{T}} \tau_e^{S^*}(X^{S^*}, t)] && \text{from (49).}
\end{aligned}$$

By the definition of $[e^{\text{tst}}]$ and Assumption 2, we then have

$$\begin{aligned}
V^{[e^{\text{tst}}]}(\pi^{\text{e-inv}}) &= \mathbb{E}^f[\max_{t\in\mathcal{T}} \tau_f^{S^*}(X^{S^*}, t)] \\
&= \mathbb{E}^{e^f}[\max_{t\in\mathcal{T}} \tau_{ef}^{S^\diamond}(X^{S^\diamond}, t)] && \text{from (52).} \tag{53}
\end{aligned}$$

Finally, we show that for all $\pi \in \Pi$, $V^{[e^{\text{tst}}]}(\pi)$ is bounded above by $V^{[e^{\text{tst}}]}(\pi^{\text{e-inv}})$.

$$\begin{aligned}
V^{[e^{\text{tst}}]}(\pi) &= \inf_{e\in[e^{\text{tst}}]} (\mathbb{E}^{e^{\text{tst}},\pi}[Y] - \mathbb{E}^{e^{\text{tst}},\pi_{t_0}}[Y]) \\
&\leq \mathbb{E}^{f,\pi}[Y] - \mathbb{E}^{f,\pi_{t_0}}[Y] && \text{since } f \in [e^{\text{tst}}] \\
&= \mathbb{E}^f[\sum_{t\in\mathcal{T}} (\mathbb{E}^{f,\pi_t}[Y \mid X] - \mathbb{E}^{f,\pi_{t_0}}[Y \mid X])\pi(t \mid X)] \\
&= \mathbb{E}^f[\sum_{t\in\mathcal{T}} \tau_f(X, t)\pi(t \mid X)] \\
&\leq \mathbb{E}^f[\max_{t\in\mathcal{T}} \tau_f(X, t)] \\
&= \mathbb{E}^f[\max_{t\in\mathcal{T}} \tau_f^{S^\diamond}(X^{S^\diamond}, t)] && \text{from (45)} \\
&= V^{[e^{\text{tst}}]}(\pi^{\text{e-inv}}) && \text{from (53),}
\end{aligned}$$

which concludes the proof. ∎

A.5 Proof of Proposition 10

Proof Let $S \subseteq \{1, \dots, d\}$, $\alpha \in (0, 1)$, $\tilde{X} := [1 \quad X]^\top$ and $\tilde{X}^S := [1 \quad X^S]^\top$ and $\hat{B}$ be the estimator solving the equation $\sum_i^n G_i(\alpha, A, \beta, B) = 0$. Assume Setting 1 and assume the following regularity conditions (these are similar to the ones required by Boruvka et al. (2018) with the difference that we require them to hold for all $e \in \mathcal{E}^{\text{tr}}$).

Assumption 6 (Regularity conditions) *For all $e \in \mathcal{E}^{\text{tr}}$ it holds that*

(i) $\mathbb{E}^{e,\pi^{\text{tr}}}[Y^4] < \infty$ and $\max_{j\in\{1,\dots,d+1\}} \mathbb{E}^e[(\tilde{X}^j)^4] < \infty$,

(ii) the matrices $\mathbb{E}^e[\tilde{X}^S(\tilde{X}^S)^\top]$ and

$$\mathbb{E}^e\left[\sum_{t\in\mathcal{T}}\tilde{\pi}(t|X^S)\begin{bmatrix}\tilde{X}\\u_e\tilde{X}\\v_t-\tilde{\pi}(1|X^S)\\(v_t-\tilde{\pi}(1|X^S))u_e\end{bmatrix}\begin{bmatrix}\tilde{X}\\u_e\tilde{X}\\v_t-\tilde{\pi}(1|X^S)\\(v_t-\tilde{\pi}(1|X^S))u_e\end{bmatrix}^\top\right]$$

are invertible.

From Proposition 3.1 in Boruvka et al. (2018), we have that $\sqrt{n}(\hat{B}-B)\xrightarrow{d}\mathcal{N}(0,\mathbb{V}[B])$, where $\hat{\mathbb{V}}$ defined in (23) is a consistent estimator of $\mathbb{V}[B]$. Next, from Theorem 8.3 in Boos et al. (2013), we have that $T_n := n\hat{B}\hat{\mathbb{V}}\hat{B}\xrightarrow{d}\chi^2_{|\text{vec}(B)|}$. Let q_α be the $(1-\alpha)$-quantile of $\chi^2_{|\text{vec}(B)|}$ and $\psi_n^{\text{Wd}}(D^{\text{tr}},S,\alpha):=\mathbb{1}(T_n>q_\alpha)$. We can then conclude that

$$\sup_{\mathbb{P}\in H_{0,S}^{\text{tr}}}\limsup_{n\to\infty}\mathbb{P}(\psi_n^{\text{Wd}}(D^{\text{tr}},S,\alpha)=1)\leq\alpha.$$

∎

A.6 Proof of Proposition 11

Proof For all $x\in\mathcal{X},t,w\in\mathcal{T}=\{0,1\},y\in\mathcal{Y}$ and $e\in\mathcal{E}$, let $Z_w^e(x,t,y):=\bar{\mu}^e(x,w)+\frac{\mathbb{1}(t=w)(y-\bar{\mu}^e(x,w))}{\bar{\pi}(w|x)}$. (This implies that $O^e(X,T,Y)=Z_1^e(X,T,Y)-Z_0^e(X,T,Y)$.) We now show that for all $w\in\mathcal{T}=\{0,1\}$ and $S\subseteq\{1,\ldots,d\}$, $\mathbb{E}^{e,\pi^{\text{tr}}}[Z_w^e(X,T,Y)\mid X^S]=\mathbb{E}^{e,\pi_w}[Y\mid X^S]$ if one of the models is correct.

(i) Assume that $\bar{\mu}^e$ is correct, i.e., $\bar{\mu}^e(x,t)=\mathbb{E}^e[Y\mid X=x,T=t]$ for all $x\in\mathcal{X}$, $t\in\mathcal{T}$ and $e\in\mathcal{E}^{\text{tr}}$. We then have for all $w\in\mathcal{T}$

$$\mathbb{E}^{e,\pi^{\text{tr}}}\left[\frac{\mathbb{1}(T=w)(Y-\bar{\mu}^e(X,w))}{\bar{\pi}(w\mid X)}\mid X^S\right]=\mathbb{E}^e\left[\mathbb{E}^{e,\pi^{\text{tr}}}\left[\frac{\mathbb{1}(T=w)(Y-\bar{\mu}^e(X,w))}{\bar{\pi}(w|X)}\mid X\right]\mid X^S\right]$$

$$=\mathbb{E}^e\left[\frac{\pi^{\text{tr}}(w|X)}{\bar{\pi}(w|X)}\mathbb{E}^e\left[(Y-\bar{\mu}^e(X,w))\mid X,T=w\right]\mid X^S\right]$$

$$=0.\tag{54}$$

Next, we have

$$\mathbb{E}^{e,\pi^{\text{tr}}}\left[\bar{\mu}^e(X,w)\mid X^S\right]=\mathbb{E}^e\left[\mathbb{E}^e\left[Y\mid X,T=w\right]\mid X^S\right]$$

$$=\mathbb{E}^e\left[\mathbb{E}^{e,\pi_w}\left[Y\mid X\right]\mid X^S\right]$$

$$=\mathbb{E}^{e,\pi_w}\left[Y\mid X^S\right].\tag{55}$$

Then, from (54) and (55), we have that $\mathbb{E}^{e,\pi^{\text{tr}}}[Z_w^e(X,T,Y)\mid X^S]=\mathbb{E}^{e,\pi_w}[Y\mid X^S]$ and it thus holds for all $x\in\mathcal{X}^S$ that

$$\mathbb{E}^{e,\pi^{\text{tr}}}[O^e(X,T,Y)\mid X^S=x]=\mathbb{E}^{e,\pi^{\text{tr}}}[Z_1^e(X,T,Y)-Z_0^e(X,T,Y)\mid X^S=x]=\tau_e^S(x,1).$$

(ii) Assume that $\bar{\pi}$ is correct, i.e., $\bar{\pi}(t \mid x) = \pi^{\mathrm{tr}}(t \mid x)$ for all $x \in \mathcal{X}$, $t \in \mathcal{T}$. We then have for all $w \in \mathcal{T}$

$$\mathbb{E}^{e,\pi^{\mathrm{tr}}}\Big[\frac{\mathbb{1}(T=w)(Y-\bar{\mu}^e(X,w))}{\bar{\pi}(w\mid X)} \mid X^S\Big] = \mathbb{E}^e\Big[\mathbb{E}^{e,\pi^{\mathrm{tr}}}\Big[\frac{\mathbb{1}(T=w)(Y-\bar{\mu}^e(X,w))}{\pi^{\mathrm{tr}}(w\mid X)} \mid X\Big] \mid X^S\Big]$$

$$= \mathbb{E}^e\big[\mathbb{E}^e\big[(Y-\bar{\mu}^e(X,w)) \mid X, T=w\big] \mid X^S\big]$$

$$= \mathbb{E}^e\big[\mathbb{E}^{e,\pi_w}\big[Y-\bar{\mu}^e(X,w) \mid X\big] \mid X^S\big]$$

$$= \mathbb{E}^{e,\pi_w}\big[Y-\bar{\mu}^e(X,w) \mid X^S\big]. \tag{56}$$

Next, we have

$$\mathbb{E}^{e,\pi^{\mathrm{tr}}}\big[\bar{\mu}^e(X,w) \mid X^S\big] = \mathbb{E}^{e,\pi_w}\big[\bar{\mu}^e(X,w) \mid X^S\big], \tag{57}$$

since the expectation is only over X and does not depend on the treatment T. Then, from (56) and (57), we have that $\mathbb{E}^{e,\pi^{\mathrm{tr}}}[Z_w^e(X,T,Y) \mid X^S] = \mathbb{E}^{e,\pi_w}[Y \mid X^S]$ and it thus holds for all $x \in \mathcal{X}^S$ that

$$\mathbb{E}^{e,\pi^{\mathrm{tr}}}[O^e(X,T,Y) \mid X^S = x] = \mathbb{E}^{e,\pi^{\mathrm{tr}}}[Z_1^e(X,T,Y) - Z_0^e(X,T,Y) \mid X^S = x] = \tau_e^S(x,1).$$

∎

Appendix B. Few-shot policy generalization in linear models

Example 2 (Few-shot policy generalization for linear CATE functions) *Let $S \in \mathbf{S}_{\mathcal{E}^{\mathrm{tr}}}^{\mathrm{e\text{-}inv}}$ and $N := \{1, \ldots, d\} \setminus S$ and recall that $\mathcal{T} = \{1, \ldots, k\}$. We assume that $\mathcal{H}$ is a class of linear functions parameterized by $\Theta \subseteq \mathbb{R}^{k \times d}$, i.e., $\mathcal{H} := \{\tau \mid \exists \theta \in \Theta \text{ s.t. } \forall t \in \mathcal{T} : \tau(\cdot, t) \equiv \theta_t(\cdot)\}$.*

By the linearity of $\mathcal{H}$, Assumption 5 is satisfied, that is, there exists $\theta^S \in \Theta^S \subseteq \mathbb{R}^{k \times |S|}$ and $\theta^N \in \Theta^N \subseteq \mathbb{R}^{k \times |N|}$ such that

$$\forall x \in \mathcal{X}, t \in \mathcal{T} : \tau_{e^{\mathrm{tst}}}(x,t) = \theta_t^S x^S + \theta_t^N x^N. \tag{58}$$

Under Assumption 2, we then have

$$\forall x \in \mathcal{X}^S, t \in \mathcal{T} : \theta_t^S x = \tau_{\mathrm{tr}}^S(x,t) - \theta_t^N \mathbb{E}^{e^{\mathrm{tst}}}[X^N \mid X^S = x], \tag{59}$$

and hence

$$\forall x \in \mathcal{X}, t \in \mathcal{T} : \tau_{e^{\mathrm{tst}}}(x,t) = \tau_{\mathrm{tr}}^S(x^S,t) + \theta_t^N(x^N - \mathbb{E}^{e^{\mathrm{tst}}}[X^N \mid X^S = x^S]). \tag{60}$$

Next, let $\hat{q}$ be an estimator of $\mathbb{E}^{e^{\mathrm{tst}}}[X^N \mid X^S = \cdot]$ and, for all $\theta \in \Theta^N$, define $\tau_\theta^S : (x,t) \mapsto \tau_{\mathrm{tr}}^S(x^S,t) + \theta_t(x^N - \hat{q}(x^S))$. Importantly, the estimand we are estimating by $\hat{q}$ does not change with θ^N. We then consider the unconstrained optimization

$$\hat{\theta} \in \operatorname*{argmin}_{\theta \in \Theta^N} \sum_{i=1}^m \ell(Y_i^{\mathrm{tst}}, X_i^{\mathrm{tst}}, T_i^{\mathrm{tst}}, \pi_i^{\mathrm{tst}}, \tau_\theta^S). \tag{61}$$

Here, by utilizing the e-invariance information τ_{tr}^S along with Assumption 2, we now optimize over the restricted function class $\Theta^N \subset \Theta$.