Machine Learning (2024) 113:3653–3673
https://doi.org/10.1007/s10994-023-06421-x

Tracking treatment effect heterogeneity in evolving environments

Tian Qin[1] · **Long-Fei Li**[1] · **Tian-Zuo Wang**[1] · **Zhi-Hua Zhou**[1]

Received: 31 May 2023 / Revised: 2 August 2023 / Accepted: 3 October 2023 /
Published online: 11 January 2024

Abstract
Heterogeneous treatment effect (HTE) estimation plays a crucial role in developing personalized treatment plans across various applications. Conventional approaches assume that the observed data are independent and identically distributed (i.i.d.). In some real applications, however, the assumption does not hold: the environment may evolve, which leads to variations in HTE over time. To enable HTE estimation in evolving environments, we introduce and formulate the online HTE estimation problem. We propose an online ensemble-based HTE estimation method called ETHOS, which is capable of adapting to unknown evolving environments by ensembling the outputs of multiple base estimators that track environmental changes at different scales. Theoretical analysis reveals that ETHOS achieves an optimal expected dynamic regret $O(\sqrt{T(1+P_T)})$, where T denotes the number of observed examples and P_T characterizes the intensity of environment changes. The achieved dynamic regret ensures that our method consistently approaches the optimal online estimators as long as the evolution of the environment is moderate. We conducted extensive experiments on three common benchmark datasets with various environment evolving mechanisms. The results validate the theoretical analysis and the effectiveness of our proposed method.

Keywords Causal inference · Heterogeneous treatment effects · Treatment effect estimation · Evolving environments

Editors: Vu Nguyen, Dani Yogatama.

✉ Zhi-Hua Zhou
zhouzh@nju.edu.cn

Tian Qin
qint@lamda.nju.edu.cn

Long-Fei Li
lilf@lamda.nju.edu.cn

Tian-Zuo Wang
wangtz@lamda.nju.edu.cn

[1] National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China

1 Introduction

Treatment effect estimation is a fundamental problem in causal inference, with broad applications in fields such as healthcare (Shalit, 2019), advertising (Wang et al., 2015), and recommender systems (Schnabel et al., 2016). Among different levels of treatment effects, heterogeneous treatment effects (HTE) measure the relative effect of decisions on the outcome of individuals or subgroups of the population, providing valuable insights for creating personalized treatment plans. The estimation of HTE is challenging due to the fundamental difficulty of causal inference: we can only observe a factual outcome corresponding to the selected decision and can never observe the counterfactual outcomes that would have occurred if other decisions were made. Many studies have sought to overcome this challenge by leveraging machine learning models, leading to notable advancements in estimating HTE from observational data (Hill, 2011; Shalit et al., 2017; Wager & Athey, 2018; Yao et al., 2018; Yoon et al., 2018; Qin et al., 2021; Harada & Kashima, 2022).

These studies commonly assume that the observational data are independent and identically distributed (i.i.d.). However, in real-world problems, these assumptions can be strongly violated. First, the effect of treatments or the environment can evolve over time (Zhou, 2022a). For example, in the development of vaccines against COVID-19, the effectiveness of a vaccine decreased as the coronavirus evolved into different variants. Another example is that in recommender systems, the click-through rate of a recommended product can dramatically change as the user interest typically evolves over a range of time. In both examples, the treatment effects vary over time. Second, in real applications, the treatment policy and the distribution of observed units can vary across different time periods. For example, recommendation policies are often adjusted according to the popularity of products, and visitor populations of a shopping recommender system are usually different on weekdays and weekends. Moreover, the observational data may appear in online scenarios, where the evolution of effects happens continuously and gradually. As a result, explicitly dividing the data into disjoint parts and assuming stationary distributions on each of them for separate estimation is unreliable. Due to the aforementioned reasons, it is necessary to develop HTE methods that are capable of tracking treatment effect heterogeneity in evolving environments where the i.i.d. assumption does not hold.

Specifically, we formulate the online HTE estimation problem for T rounds. The *treatment effects*, the *probability of being treated*, and the *unit distribution* in each round may differ from previous ones. At each round, we submit an HTE estimator to the environment and subsequently observe a new example, which can be used to update the estimator. Note that the submitted estimator can be used to guide the treatment probability of the subsequent example, resembling many online decision processes such as online recommendation. Apart from the difficulties posed by the unobserved counterfactuals in conventional HTE estimation problems, the most challenging aspect of this problem is that we do not know how the environment would evolve in advance: The method must be able to adapt to any possible changes. Furthermore, it is crucial for the overall method to have strong theoretical guarantees, especially considering that HTE is frequently utilized in critical decision-making tasks such as medicating.

To address the online HTE estimation problem, we propose a method called ETHOS. We resolve the difficulty raised by unobserved counterfactuals by replacing the incalculable true estimation error with a surrogate loss function. By minimizing this surrogate loss, we effectively minimize the original true losses in expectation. To tackle the unknown environmental changes, we base ETHOS on the online ensemble technique, which was

originally introduced to handle the non-stationarity in online convex optimizations (Zhang et al., 2018; Zhao et al., 2021). Roughly, ETHOS maintains multiple base HTE estimators and optimally ensembles their outputs as the final estimator. Each base estimator optimizes the surrogate loss with a unique learning rate, which intuitively captures environmental changes at different scales. By aggregating the outputs of base estimators, ETHOS follows the base estimator that best adapts to the undergoing environmental change, thus providing reliable estimations throughout the whole process.

We measure the estimator performance with a variant of dynamic regret (Zinkevich, 2003), which compares the generalization loss of the learned estimators with any possible estimator sequence, making it suitable for characterizing the performance in evolving environments. We prove that our proposed method achieves an optimal expected dynamic regret of $O(\sqrt{T(1+P_T)})$, where P_T characterizes the accumulated changes of a comparator sequence, which can reflect the variations of the environment. The regret bound indicates that the proposed method consistently approaches any optimal estimator as long as $P_T \leq o(T)$, which is satisfied by environments undergoing mild changes. In this case, the average regret of the estimator in each round is $O(\sqrt{(1+P_T)/T})$, which decreases to zero as T approaches infinity.

In summary, our primary contributions are threefold. Firstly, to the best of our knowledge, this is the first work that introduces and formulates the online HTE estimation problem in evolving environments, which is crucial for deploying HTE estimators in real online applications. Secondly, we present an HTE estimation method for the problem with strong theoretical guarantees, supported by extensive experimental validation. Finally, we derive a problem lower bound, which matches the achieved dynamic regret of our proposed method, demonstrating the optimality of the method.

The remainder of the paper proceeds as follows. Section 2 discusses related work. Section 3 formulates the online HTE estimation problem. The proposed method and theoretical analysis are presented in Sects. 4 and 5, respectively. Empirical evaluations are provided in Sect. 6. Finally, we conclude in Sect. 7.

2 Related work

There have been considerable efforts in incorporating machine learning models into HTE estimation. Notable methods include proposals based on Bayesian additive regression trees (Hill, 2011; Hahn et al., 2020), random forests (Wager & Athey, 2018; Athey et al., 2019), deep neural networks (Shalit et al., 2017; Louizos et al., 2017; Yao et al., 2018; Yoon et al., 2018; Zhang et al., 2021; Harada & Kashima, 2022), etc. Meta-algorithms that can leverage any supervised learning or statistical regression methods have also been proposed (Künzel et al., 2019; Nie & Wager, 2020). These methods typically rely on the i.i.d. assumption, which is not assumed in this work.

Some studies investigate the mismatch of distributions in HTE estimation problems within the framework of transfer learning and domain adaptation (Künzel et al., 2018; Johansson et al., 2018; Shi et al., 2021; Bica & van der Schaar, 2022). These studies generally assumed the existence of at least two stationary domains and studied how to leverage the information of source domains to benefit HTE estimation on the target domain. As mentioned in Sect. 1, the observational data come in an online manner and does not have explicit domains with distinct distributions. It is also not appropriate to manually divide the accumulated data into several domains and apply transfer learning

since transfer learning requires the data on each domain to be i.i.d. while in the online HTE estimation problem, the distribution change can happen continuously, which means that it is unclear whether the data accumulated within a specific time period is identically distributed.

Brodersen et al. (2015) and Li and Buhlmann (2018) considered estimating treatment effects in observational studies that run over a certain period of time. They assume that after receiving a treatment, the effect can vary with time and constituents a non-stationary time series. The online HTE estimation problem considered in this work is different from theirs in that the outcome variable is real-valued instead of a time series, and that the we explore non-stationarity of observed distributions rather than the evolution of outcomes after receiving treatments.

In the field of online convex optimization, the concept of dynamic regret (Zinkevich, 2003) compares the cumulative loss of a learner to that of any sequence of comparators, taking into account the changing environments. However, the original definition of dynamic regret focuses on the in-sample losses incurred by the learner, whereas HTE estimation requires the ability to generalize to unseen data. Therefore, a modified dynamic regret is necessary to assess the performance of HTE estimators in evolving environments. While Zhang et al. (2018) proposed a method that achieves minimax optimal dynamic regret in online convex optimization tasks when the learner has access to true gradients, our problem poses additional challenges. In our case, obtaining the true gradients, or even the true losses, is impossible due to the presence of unobserved counterfactuals, making the application of online optimization techniques difficult.

3 Problem setup

HTE quantifies the effect of a treatment $W \in \mathcal{W}$ on the real-valued outcome $Y \in \mathcal{Y}$ of a specific subgroup described by covariates $\boldsymbol{X} \in \mathcal{X}$. We consider binary treatments $\mathcal{W} = \{-1, 1\}$. A unit belongs to the treated group if $W = 1$ and the control group if $W = -1$. Under the potential outcome framework (Neyman, 1923; Rubin, 1974), HTE is also known as the conditional average treatment effect (CATE) and is defined as

$$\tau(\boldsymbol{x}) \triangleq \mathbb{E}[Y(1) - Y(-1) \mid \boldsymbol{X} = \boldsymbol{x}],$$

where $Y(w)$ denotes the potential outcome for treatment $W = w$, i.e., the value that Y would obtain had $\boldsymbol{x}$ received treatment w. One of the main challenges of this task is that we can only observe the factual outcome $Y(W)$ for a unit, but never the counterfactual outcome $Y(-W)$.

We consider an online process of HTE estimation that spans T rounds, where the environment, including the CATE function $\tau(\boldsymbol{x})$, the propensity score function $q(x) \triangleq P(W = 1 \mid \boldsymbol{x})$, and the distribution of $\boldsymbol{X}$, evolves over time. We estimate HTE with a function $h(\cdot;\boldsymbol{\theta}) : \mathcal{X} \to \mathbb{R}$, which is parametrized by $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Let $\boldsymbol{x}_t, w_t, y_t$, and $\tau_t(\boldsymbol{x})$ denote the observed covariates, treatment, factual outcome, and the underlying CATE function at round t, respectively. Let $p_t = P(w_t \mid \boldsymbol{x}_t)$. The online HTE estimation process is: At each round $t = 1, \ldots, T$,

1. The learner picks a parameter $\boldsymbol{\theta}_t \in \boldsymbol{\Theta}$, which constituents an HTE estimator $h(\cdot;\boldsymbol{\theta}_t)$;
2. The environment then reveals a new example $(\boldsymbol{x}_t, w_t, y_t, p_t)$ to the learner.

In this process, the learner determines the parameter based on data collected in previous rounds, and the environment randomly samples a new example $d_t = (\boldsymbol{x}_t, w_t, y_t)$ from the evolving distribution $\mathcal{D}_t$, which can differ for each round t. Note that we assume p_t is also revealed to the learner, which is reasonable since the deployed treatment policy is generally known in online applications such as online recommendation systems. Additionally, the environment can adjust the treatment probability according to the effect estimate $\hat{\tau}_t = h(\boldsymbol{x}_t;\boldsymbol{\theta}_t)$ before giving treatment to the unit $\boldsymbol{x}_t$, which resembles many real scenarios, e.g., an ad provider first estimates the effect on a user and then decides the probability of displaying a specific ad.

As the environment is evolving, the optimal HTE estimator can vary across different rounds. We measure the performance of an online estimator with a modified version of dynamic regret, which compares the cumulative loss of the estimator sequence output by an online algorithm with that of an arbitrary sequence $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T \in \boldsymbol{\Theta}$:

$$\mathbf{Reg}_T(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T) \triangleq \sum_{t=1}^{T} f_t(\boldsymbol{\theta}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{u}_t), \tag{1}$$

where the loss function $f_t(\boldsymbol{\theta}) \triangleq \mathbb{E}_{\boldsymbol{x}_t}\left[\frac{1}{2}(h(\boldsymbol{x}_t;\boldsymbol{\theta}) - \tau_t(\boldsymbol{x}_t))^2\right]$ is the expected squared error of $h(\cdot;\boldsymbol{\theta})$ on the underlying distribution at round t. The loss is commonly known as the squared precision in estimation of heterogeneous effects (Hill, 2011). We aim to design an algorithm that ensures that the generated sequence $\boldsymbol{\theta}_1 \ldots, \boldsymbol{\theta}_T$ has low dynamic regret, which means that it has comparable or better performance against any possible sequence $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T$, including the optimal one that best adapts to the changing environment. By achieving low dynamic regret, we can conclude that the algorithm successfully tracks the evolving treatment effect heterogeneity.

We consider h to be any function which satisfies that $(h(\boldsymbol{x};\boldsymbol{\theta}) - \tau)^2$ is convex in $\boldsymbol{\theta}$. For example, h could be any generalized linear models in the form of $h(\boldsymbol{x};\boldsymbol{\theta}) = \boldsymbol{\phi}(\boldsymbol{x})^\top \boldsymbol{\theta}$, where $\boldsymbol{\phi} : \mathcal{X} \to \mathcal{X}'$ is a feature transformation function that maps $\boldsymbol{x}$ from the original space to another feature space $\mathcal{X}'$. Although the true CATE function may not have the same form as h, with a properly specified $\boldsymbol{\phi}$, we can expect that the best estimator representable by h to well approximate the true function. We next state some standard assumptions in the treatment effect estimation (Imbens & Rubin, 2015) and the online convex optimization literature (Hazan, 2016; Zhang et al., 2018):

Assumption 1 *(SUTVA)* The potential outcomes for any unit do not vary with the treatments assigned to other units, and, for each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes.

Assumption 2 *(Consistency)* For all $t \in [T]$, the potential outcome $Y_t(w)$ equals the observed outcome Y_t if the actual treatment received is $W_t = w$.

Assumption 3 (ε-*strong ignorability*) For all $t \in [T]$, $\{Y_t(1), Y_t(-1)\} \perp W_t \mid \boldsymbol{X}_t$. There exists $0 < \varepsilon < 0.5$ such that $\varepsilon < P(W_t = 1 \mid \boldsymbol{x}_t) < 1 - \varepsilon$ holds for all $\boldsymbol{x}_t \in \mathcal{X}$.

Assumption 4 The covariates and the outcome have bounded norm, i.e., there exists A and B such that $\sup_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{x}\|_2 \leq A$ and $\sup_{y\in\mathcal{Y}} |y| \leq B$.

Assumption 5 The range and the magnitude of the gradient of $h(\boldsymbol{x};\boldsymbol{\theta})$ is bounded, i.e., there exists H and G such that for all $\boldsymbol{x} \in \mathcal{X}, \boldsymbol{\theta} \in \boldsymbol{\Theta}$, we have

$$-H \leq h(\boldsymbol{x};\boldsymbol{\theta}) \leq H \text{ and } \|\nabla_{\boldsymbol{\theta}} h(x;\boldsymbol{\theta})\|_2 \leq G.$$

Assumption 6 The domain $\boldsymbol{\Theta}$ is convex and $\mathbf{0} \in \boldsymbol{\Theta}$. There exist D such that

$$\sup_{\boldsymbol{\theta},\boldsymbol{\theta}' \in \boldsymbol{\Theta}} \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2 \leq D.$$

We make some explanations for the last three assumptions, which rarely occur in the causal inference literature. Assumption 4 assumes bounded norm of the covariates and the outcomes, which is a relatively weak requirement that can be met in many real scenarios. For instance, it holds when none of the considered variables can take on an infinite value, which is reasonable for common quantities with certain semantics, such as age, gender, blood pressure, etc. Assumption 6 is also commonly fulfilled, as we typically consider model parameters within a specific range. Regarding Assumption 5, it is important to note that it imposes restrictions on the estimating function h rather than the true CATE function. Hence, one can easily select a suitable function class that satisfies this assumption. For example, linear functions automatically meet the requirement given Assumptions 4 and 6. Although the above assumptions exclude some commonly used function classes like tree-based models, they facilitate the subsequent theoretical analysis and yield methods with strong theoretical guarantees.

4 Proposed method

We propose a method called ETHOS, standing for tracking Evolving Treatment effect Heterogeneity with Online enSemble, which tackles the evolving environments with online gradient descent and online ensemble techniques (Zhang et al., 2018).

In contrast to typical online optimization tasks where the loss values are easily available, the loss function $f_t(\cdot)$, even its empirical version, is not computable in the online HTE estimation problem due to the unobserved counterfactuals. To address this limitation and enable the learning of HTEs, we need a surrogate loss function as a substitute for the true loss. Noticing that $\hat{\tau}_t \triangleq w_t y_t / p_t$ is an unbiased estimator for the true effect, i.e., $\mathbb{E}_{w_t,y_t|\boldsymbol{x}_t}[\hat{\tau}_t] = \tau_t(\boldsymbol{x}_t)$, we define the following surrogate loss:

Definition 1 (*Surrogate loss*) The surrogate loss in the t-th round is defined as

$$\hat{f}_t(\boldsymbol{\theta}_t;\boldsymbol{x}_t,\hat{\tau}_t) \triangleq \frac{1}{2}\left(h(\boldsymbol{x}_t;\boldsymbol{\theta}_t) - \hat{\tau}_t\right)^2 = \frac{1}{2}\left(h(\boldsymbol{x}_t;\boldsymbol{\theta}_t) - \frac{w_t y_t}{p_t}\right)^2.$$

We observe that the surrogate loss function $\hat{f}_t$ is convex in $\boldsymbol{\theta}_t$ due to the properties of h. This convexity property allows for easy minimization using general online gradient descent methods (Hazan, 2016). Moreover, as Proposition 1 indicates, by optimizing $\hat{f}_t(\boldsymbol{\theta}_t;\boldsymbol{x}_t,\hat{\tau}_t) - \hat{f}_t(\boldsymbol{u}_t;\boldsymbol{x}_t,\hat{\tau}_t)$, we effectively optimize the difference between the true loss of $\boldsymbol{\theta}_t$ and $\boldsymbol{u}_t$ in expectation. Consequently, we can transform the objective of minimizing the

incalculable f_t into minimizing the surrogate loss $\hat{f}_t$. For simplicity, we sometimes write $\hat{f}_t(\cdot)$ instead of $\hat{f}_t(\cdot;\boldsymbol{x}_t,\hat{\tau}_t)$ in the remainder of the paper.

Proposition 1 *For any $t\in[T]$, $\boldsymbol{\theta}_t\in\boldsymbol{\Theta}$ generated by any online algorithm and any fixed comparator $\boldsymbol{u}_t\in\boldsymbol{\Theta}$, $\mathbb{E}_{d_t|d_{1:t-1}}\left[\hat{f}_t(\boldsymbol{\theta}_t;\boldsymbol{x}_t,\hat{\tau}_t)-\hat{f}_t(\boldsymbol{u}_t;\boldsymbol{x}_t,\hat{\tau}_t)\right]=f_t(\boldsymbol{\theta}_t)-f_t(\boldsymbol{u}_t)$.*

Proof All the distributions in this proof are conditioned on $d_{1:t-1}$. Denoting the variance of a random variable with $\mathbb{V}[\cdot]$, we have

$$\begin{aligned}
\mathbb{E}_{w_t,y_t|\boldsymbol{x}_t}\left[\hat{f}_t(\cdot;\boldsymbol{x}_t,\hat{\tau}_t)\right] &= \frac{1}{2}\mathbb{E}_{w_t,y_t|\boldsymbol{x}_t}\left[h(\boldsymbol{x}_t;\cdot)^2-2h(\boldsymbol{x}_t;\cdot)\hat{\tau}_t+\hat{\tau}_t^2\right]\\
&= \frac{1}{2}\left(h(\boldsymbol{x}_t;\cdot)^2-2h(\boldsymbol{x}_t;\cdot)\tau_t+\tau_t(\boldsymbol{x}_t)^2+\mathbb{V}_{w_t,y_t|\boldsymbol{x}_t}\left[\hat{\tau}_t\right]\right)\\
&= \frac{1}{2}\left(\left(h(\boldsymbol{x}_t;\cdot)-\tau_t(\boldsymbol{x}_t)\right)^2+\mathbb{V}_{w_t,y_t|\boldsymbol{x}_t}\left[\hat{\tau}_t\right]\right)\\
\Rightarrow\quad \mathbb{E}_{d_t}\left[\hat{f}_t(\cdot;\boldsymbol{x}_t,\hat{\tau}_t)\right] &= \mathbb{E}_{\boldsymbol{x}_t}\left[\mathbb{E}_{w_t,y_t|\boldsymbol{x}_t}\left[\hat{f}_t(\cdot;\boldsymbol{x}_t,\hat{\tau}_t)\right]\right]\\
&= \mathbb{E}_{\boldsymbol{x}_t}\left[\frac{1}{2}\left(h(\boldsymbol{x}_t;\cdot)-\tau_t(\boldsymbol{x}_t)\right)^2\right]+\frac{1}{2}\mathbb{E}_{\boldsymbol{x}_t}\left[\mathbb{V}_{w_t,y_t|\boldsymbol{x}_t}\left[\hat{\tau}_t\right]\right]\\
&= f_t(\cdot)+\frac{1}{2}\mathbb{E}_{\boldsymbol{x}_t}\left[\mathbb{V}_{w_t,y_t|\boldsymbol{x}_t}\left[\hat{\tau}_t\right]\right]\\
\Rightarrow\quad \mathbb{E}_{d_t}\left[\hat{f}_t(\boldsymbol{\theta}_t;\boldsymbol{x}_t,\hat{\tau}_t)-\hat{f}_t(\boldsymbol{u}_t;\boldsymbol{x}_t,\hat{\tau}_t)\right] &= f_t(\boldsymbol{\theta}_t)-f_t(\boldsymbol{u}_t).
\end{aligned}$$

□

The gradient of the surrogate loss with respect to $\boldsymbol{\theta}$ is

$$\hat{g}_t(\boldsymbol{\theta})\triangleq\nabla_{\boldsymbol{\theta}}\hat{f}_t(\boldsymbol{\theta};\boldsymbol{x}_t,\hat{\tau}_t)=\left(h(\boldsymbol{x}_t;\boldsymbol{\theta})-\frac{w_t y_t}{p_t}\right)\nabla_{\boldsymbol{\theta}}h(\boldsymbol{x}_t;\boldsymbol{\theta}),\tag{2}$$

which is upper bounded by a constant $\tilde{G}$, as shown in the following proposition:

Proposition 2 *Let $\tilde{G}\triangleq(H+B/\varepsilon)G$. For any $t\in[T]$, $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, we have $\|\hat{g}_t(\boldsymbol{\theta})\|_2\le\tilde{G}$.*

Proof The conclusion follows from Assumptions 3, 4, and 5. □

We can then optimize the surrogate loss with Algorithm 1, which simply performs online gradient descent with a fixed learning rate η. However, the outputs of Algorithm 1 heavily relies on the given learning rate, which can not be optimally specified if the environment changes are unknown (Zinkevich, 2003), which is exactly the most challenging part of the online HTE estimation problem. Hence, we need a more sophisticated method that adapts to any unknown scale of environmental changes.

Algorithm 1 Online base HTE estimator

Input: learning rate η
Initialize $\boldsymbol{\theta}_1^\eta$ with any point in $\boldsymbol{\Theta}$
for $t = 1, \ldots, T$ **do**
  Output $\boldsymbol{\theta}_t^\eta$
  Receive new example $(\boldsymbol{x}_t, w_t, y_t, p_t)$
  Compute the gradient $\hat{g}_t(\boldsymbol{\theta}_t^\eta)$ according to Eq. (2)
  Update $\boldsymbol{\theta}_{t+1}^\eta = \Pi_{\boldsymbol{\Theta}}[\boldsymbol{\theta}_t^\eta - \eta \hat{g}_t(\boldsymbol{\theta}_t^\eta)]$ ▷ $\Pi_{\boldsymbol{\Theta}}[\cdot]$ denotes projection onto $\boldsymbol{\Theta}$

To overcome the difficulty, we adopt the idea of ensemble learning (Zhou, 2012; Zhang et al., 2018; Zhao et al., 2021), where the learning is built on the wisdom of diverse base learners (Zhou & Tan, 2023). Intuitively, when performing gradient descent, a small learning rate is more suitable for environments that evolve gradually and a large learning rate can better capture abrupt changes. Therefore, if we optimize the surrogate loss with sufficiently many optimizers, each having a unique learning rate, we can ensure that we always have an optimizer that best adapts to the undergoing change of environments, even if the scale of change is unknown. Moreover, although we do not know which optimizer is the best, it is possible to combine all the optimizers with soft weights, for which we assign larger values if the corresponding optimizer suffers a smaller cumulative loss. In this way, the combined output can be comparable to the best one.

Based on the ensemble idea, we propose the ETHOS algorithm, which utilizes multiple base estimators with different learning rates and ensembles their outputs. The overall algorithm is outlined in Algorithm 2, which accepts some problem-dependent quantities as inputs and outputs $\boldsymbol{\theta}_t$ in each round. ETHOS invokes $N = O(\log T)$ base estimators from Algorithm 1 and aggregates their outputs using an exponential weighting scheme (Cesa-Bianchi & Lugosi, 2006), which assigns higher weights ω_t^η to base estimators with smaller losses, allowing the aggregated estimator to mimic the outputs of the best base estimators. At round t, ETHOS first receives the estimator parameters $\{\boldsymbol{\theta}_t^\eta\}_\eta$ from N base estimators, then computes a weighted average of the parameters as the output. After receiving a new example, base estimators update their parameters and ETHOS updates the weights accordingly. An advantage of ETHOS is that it only requires $O(\log T)$ storage space, significantly smaller than the size of data $O(T)$, which could have significant practical implications in real applications.

Algorithm 2 ETHOS

Input: Total number of rounds T, gradient upper bound $\tilde{G}$, decision diameter D

Set step size $\alpha = \sqrt{8/(T\tilde{G}^2D^2)}$

Set number of base estimators $N = \left\lceil \frac{1}{2}\log_2(1+\frac{4T}{7}) \right\rceil + 1$

Set a set of learning rates $\mathcal{H} = \left\{ \eta_i = \frac{2^{i-1}D}{\tilde{G}}\sqrt{\frac{7}{2T}} \;\middle|\; i = 1,\dots,N \right\}$

Set weight $\omega_1^\eta = \frac{1}{N}$ for each $\eta \in \mathcal{H}$

Activate an estimator from Algorithm 1 for each $\eta \in \mathcal{H}$, denoted by E^η

for $t = 1,\dots,T$ **do**

  Receive $\boldsymbol{\theta}_t^\eta$ from each E^η

  Output $\boldsymbol{\theta}_t = \sum_{\eta\in\mathcal{H}} \omega_t^\eta \boldsymbol{\theta}_t^\eta$

  Receive new example $(\boldsymbol{x}_t, w_t, y_t, p_t)$ and send to each E^η

  Update the weight of each base estimator by $\omega_{t+1}^\eta = \frac{\omega_t^\eta \exp\left(-\alpha \hat{f}_t(\boldsymbol{\theta}_t^\eta)\right)}{\sum_{\mu\in\mathcal{H}} \omega_t^\mu \exp\left(-\alpha \hat{f}_t(\boldsymbol{\theta}_t^\mu)\right)}$

The parameters in Algorithm 2 are carefully designed to make the algorithm enjoy an optimal upper bound on dynamic regret, as elaborated in the following theoretical section. Roughly, the step size α reflects how confident we are about the representativeness of the surrogate loss for the underlying environment: a large step size enables the meta-estimator to quickly catch up with the best base estimators, meaning that we believe that the surrogate loss faithfully reflects the evolvement of the true environment. The learning rates in $\mathcal{H}$ are set with a geometric series with a ratio 2, which ensures that a nearly optimal learning rate, and correspondingly, a nearly optimal HTE estimator is available for any possible scale of environment changes.

5 Theoretical guarantees

In this section, we present theoretical analysis for ETHOS and the online HTE estimation problem. In general, achieving a sublinear dynamic regret in T for online HTE estimation is impossible: in the worst case, the environment can evolve arbitrarily and make learning on previous data useless for generalizing on future environments. However, we can bound the regret with comparator-dependent quantities and obtain sublinear dynamic regret for benign environments. We consider the path length (Zinkevich, 2003) of a comparator sequence:

Definition 2 (*Path length*) The path length of a comparator sequence $\boldsymbol{u}_1, \dots, \boldsymbol{u}_T$ is

$$P_T(\boldsymbol{u}_1, \dots, \boldsymbol{u}_T) \triangleq \sum_{t=2}^{T} \|\boldsymbol{u}_t - \boldsymbol{u}_{t-1}\|_2.$$

The path length reflects the intensity of the environment change: a small path length of the optimal comparator sequence corresponds to a slowly evolving environment. In the remainder, we write the path length as P_T when the comparator sequence is clear from the context. We proceed to present Lemma 3, establishing that ETHOS achieves a comparator-dependent dynamic regret bound on the surrogate loss.

Lemma 3 *For any comparator sequence $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T \in \mathcal{X}$, Algorithm 2 satisfies*

$$\sum_{t=1}^{T} \hat{f}_t(\boldsymbol{\theta}_t; \boldsymbol{x}_t, \hat{\tau}_t) - \sum_{t=1}^{T} \hat{f}_t(\boldsymbol{u}_t; \boldsymbol{x}_t, \hat{\tau}_t) \leq O\left(\sqrt{T\left(1 + P_T\right)}\right).$$

Proof For any $k \in [N]$, we have

$$\sum_{t=1}^{T} \hat{f}_t(\boldsymbol{\theta}_t; \boldsymbol{x}_t, \hat{\tau}_t) - \sum_{t=1}^{T} \hat{f}_t(\boldsymbol{u}_t; \boldsymbol{x}_t, \hat{\tau}_t) = \underbrace{\sum_{t=1}^{T} \hat{f}_t\left(\boldsymbol{\theta}_t\right) - \sum_{t=1}^{T} \hat{f}_t(\boldsymbol{\theta}_t^{\eta^k})}_{\texttt{term(a)}} + \underbrace{\sum_{t=1}^{T} \hat{f}_t(\boldsymbol{\theta}_t^{\eta^k}) - \sum_{t=1}^{T} \hat{f}_t\left(\boldsymbol{u}_t\right)}_{\texttt{term(b)}}.$$

The `term (a)` is the difference between the surrogate losses achieved by the meta-estimator and a base estimator, which can be bounded by using analysis of expert-based algorithms (Cesa-Bianchi & Lugosi, 2006). Based on Lemma 1 of Zhang et al. (2018), for any $\eta^k \in \mathcal{H}$, we have

$$\texttt{term(a)} \leq \frac{\tilde{G}D\sqrt{2T}}{4}\left(1 + \ln \frac{1}{\omega_1^{\eta^k}}\right) \leq \frac{\tilde{G}D\sqrt{2T}}{4}(1 + O(\log\log T)). \tag{3}$$

The `term (b)` is the regret achieved by online gradient descent with learning rate η^k on the convex surrogate functions (Zinkevich, 2003, Theorem 2), which is bounded by

$$\texttt{term(b)} \leq \frac{7D^2}{4\eta^k} + \frac{D}{\eta^k}\sum_{t=2}^{T} \|\boldsymbol{u}_t - \boldsymbol{u}_{t-1}\|_2 + \frac{\eta^k T\tilde{G}^2}{2} = \frac{7D^2}{4\eta^k} + \frac{DP_T}{\eta^k} + \frac{\eta^k T\tilde{G}^2}{2}.$$

The learning rate that minimizes `term(b)` is given by $\eta^\star = \sqrt{\frac{7D^2+4DP_t}{2T\tilde{G}^2}}$, which is not necessarily in $\mathcal{H}$. We choose $k = \lfloor \frac{1}{2}\log_2(1 + \frac{4P_T}{7D})\rfloor + 1$, which is no larger than N since $\frac{P_T}{D} \leq \frac{TD}{D} = T$, so that η_k is close to $\eta^\star$: $\eta_k = \frac{2^{k-1}D}{\tilde{G}}\sqrt{\frac{7}{2T}} \leq \eta^\star \leq 2\eta_k$, which gives

$$\texttt{term(b)} \leq \frac{7D^2}{2\eta^\star} + \frac{2DP_T}{\eta^\star} + \frac{\eta^\star T\tilde{G}^2}{2} = \frac{3\tilde{G}}{4}\sqrt{2T(7D^2 + 4DP_T)}. \tag{4}$$

Combining (3–4) and noticing $O(\log\log T)$ can be treated as a constant, we have

$$\begin{aligned}\texttt{term(a)} + \texttt{term(b)} &\leq \frac{\tilde{G}D\sqrt{2T}}{4}(1 + O(\log\log T)) + \frac{3\tilde{G}}{4}\sqrt{2T(7D^2 + 4DP_T)} \\ &\leq O\left(\sqrt{T(1 + P_T)}\right).\end{aligned}$$

□

Consequently, ETHOS achieves an expected dynamic regret on the true loss function, which is of the same order as the dynamic regret on the surrogate loss function:

Theorem 4 *For any comparator sequence $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T$, Algorithm 2 satisfies*

$$\mathbb{E}_{d_{1:T}}\left[\mathbf{Reg}_T(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_T)\right] \leq O\left(\sqrt{T(1+P_T)}\right).$$

Proof Noticing that the randomness of $\boldsymbol{\theta}_t$ generated from Algorithm 2 only comes from the data collected in previous $t-1$ rounds, we have

$$\begin{aligned}
\mathbb{E}_{d_{1:T}}\left[\mathbf{Reg}_T(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_T)\right] &= \sum_{t=1}^{T}\mathbb{E}_{d_{1:t-1}}\left[f_t(\boldsymbol{\theta}_t)-f_t(\boldsymbol{u}_t)\right] \\
&= \sum_{t=1}^{T}\mathbb{E}_{d_{1:t-1}}\left[\mathbb{E}_{d_t|d_{1:t-1}}\left[\hat{f}_t(\boldsymbol{\theta}_t)-\hat{f}_t(\boldsymbol{u}_t)\right]\right] \quad \text{(Proposition 1)} \\
&= \sum_{t=1}^{T}\mathbb{E}_{d_{1:t}}\left[\hat{f}_t(\boldsymbol{\theta}_t)-\hat{f}_t(\boldsymbol{u}_t)\right] \\
&= \mathbb{E}_{d_{1:T}}\left[\sum_{t=1}^{T}\hat{f}_t(\boldsymbol{\theta}_t)-\sum_{t=1}^{T}\hat{f}_t(\boldsymbol{u}_t)\right] \\
&\leq O\left(\sqrt{T(1+P_T)}\right). \quad \text{(Lemma 3)}
\end{aligned}$$

□

Remark 1 The $O(\sqrt{T(1+P_T)})$ bound shows that in expectation, we can have a sublinear dynamic regret and the loss of ETHOS can consistently approach any estimator sequence as long as the sequence does not undergo drastic changes, i.e., $P_T \leq o(T)$. This condition can be satisfied by optimal comparator sequences in a slowly evolving environment. In this case, the average regret of the estimator in each round is $O(\sqrt{(1+P_T)/T})$, diminishing as T tends to infinity.

The expected dynamic regret provides guarantees on the average performance over multiple runs. But it may not align with the goal of the online HTE estimation problem, where the focus is sequentially estimating the effects without running the process multiple times. Hence, it is more meaningful to derive a bound on the actual dynamic regret that holds with high probability:

Theorem 5 *For any comparator sequence $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_T$, Algorithm 2 satisfies that with probability at least $1-\delta$,*

$$\mathbf{Reg}_T(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_T) \leq O\left(\sqrt{T(1+P_T+\ln(1/\delta))}\right).$$

Proof Let $d_0 = 0$, $d_t = (\boldsymbol{x}_t, w_t, y_t)$ for all $t \geq 1$. Let $z_0 = 0$, $z_t = \sum_{i=1}^{t}(\hat{f}_i(\boldsymbol{\theta}_i) - f_i(\boldsymbol{\theta}_i) - \hat{f}_i(\boldsymbol{u}_i) + f_i(\boldsymbol{u}_i))$ for all $t \geq 1$. Using Proposition 1, we have

$$\begin{aligned}\mathbb{E}\left[z_{t+1} \mid d_{0:t}\right] &= \sum_{i=1}^{t}(\hat{f}_i(\boldsymbol{\theta}_i) - f_i(\boldsymbol{\theta}_i) - \hat{f}_i(\boldsymbol{u}_i) + f_i(\boldsymbol{u}_i)) \\ &\quad + \mathbb{E}\left[(\hat{f}_{t+1}(\boldsymbol{\theta}_{t+1}) - f_{t+1}(\boldsymbol{\theta}_{t+1})) - (\hat{f}_{t+1}(\boldsymbol{u}_{t+1}) - f_{t+1}(\boldsymbol{u}_{t+1})) \mid d_{0:t}\right] \\ &= \sum_{i=1}^{t}(\hat{f}_i(\boldsymbol{\theta}_i) - f_i(\boldsymbol{\theta}_i) - \hat{f}_i(\boldsymbol{u}_i) + f_i(\boldsymbol{u}_i)) = z_t,\end{aligned}$$

so $z_0, z_1, \ldots, z_T$ is a martingale w.r.t. $d_0, d_1, \ldots, d_T$. Under regularity conditions,

$$\begin{aligned}\nabla_{\boldsymbol{\theta}} f_t(\boldsymbol{\theta}) &= \nabla_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{x}_t}\left[\frac{1}{2}\left(h(\boldsymbol{x}_t;\boldsymbol{\theta}) - \tau_t(\boldsymbol{x}_t)\right)^2\right] = \mathbb{E}_{\boldsymbol{x}_t}\left[\frac{1}{2}\nabla_{\boldsymbol{\theta}}\left(h(\boldsymbol{x}_t;\boldsymbol{\theta}) - \tau_t(\boldsymbol{x}_t)\right)^2\right] \\ &= \mathbb{E}_{\boldsymbol{x}_t}\left[\left(h(\boldsymbol{x}_t;\boldsymbol{\theta}) - \tau_t(\boldsymbol{x}_t)\right)\nabla_{\boldsymbol{\theta}} h(\boldsymbol{x}_t;\boldsymbol{\theta})\right] = \mathbb{E}_{d_t}\left[\left(h(\boldsymbol{x}_t;\boldsymbol{\theta}) - w_t y_t/p_t\right)\nabla_{\boldsymbol{\theta}} h(\boldsymbol{x}_t;\boldsymbol{\theta})\right] \\ &= \mathbb{E}_{d_t}\left[\nabla_{\boldsymbol{\theta}} \hat{f}_t(\boldsymbol{\theta})\right] \le \tilde{G}.\end{aligned}$$

So $|z_t - z_{t-1}|$ is bounded for all $t \ge 1$:

$$\begin{aligned}|z_t - z_{t-1}| &= |(\hat{f}_t(\boldsymbol{\theta}_t) - \hat{f}_t(\boldsymbol{u}_t)) - (f_t(\boldsymbol{\theta}_t) - f_t(\boldsymbol{u}_t))| \le |\hat{f}_t(\boldsymbol{\theta}_t) - \hat{f}_t(\boldsymbol{u}_t)| + |f_t(\boldsymbol{\theta}_t) - f_t(\boldsymbol{u}_t)| \\ &\le \tilde{G}\|\boldsymbol{\theta}_t - \boldsymbol{u}_t\|_2 + \tilde{G}\|\boldsymbol{\theta}_t - \boldsymbol{u}_t\|_2 \le 2\tilde{G}D.\end{aligned}$$

Applying Azuma's inequality, we have that for any $\epsilon > 0$,

$$\mathbb{P}\left[z_T \ge \epsilon\right] = \mathbb{P}\left[z_T - z_0 \ge \epsilon\right] \le \exp\left(-\frac{\epsilon^2}{8T\tilde{G}^2D^2}\right).$$

Let $\delta = \exp\left(-\epsilon^2/(8T\tilde{G}^2D^2)\right)$. Then $\epsilon = 2\tilde{G}D\sqrt{2T\ln(1/\delta)}$. Hence with probability at least $1-\delta$,

$$\begin{aligned}\sum_{t=1}^{T} f_t(\boldsymbol{\theta}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{u}_t) &\le \sum_{t=1}^{T} \hat{f}_t(\boldsymbol{\theta}_t) - \sum_{t=1}^{T} \hat{f}_t(\boldsymbol{u}_t) + 2\tilde{G}D\sqrt{2T\ln(1/\delta)} \\ &\le O\left(\sqrt{T(1+P_T)}\right) + 2\tilde{G}D\sqrt{2T\ln(1/\delta)} \\ &\le O\left(\sqrt{T(1+P_T+\ln(1/\delta))}\right).\end{aligned}$$

□

Considering linear HTE estimators, we further prove a lower bound that matches the dynamic regret upper bound in Theorem 4, confirming the optimality of ETHOS.

Theorem 6 *Suppose that the HTE is estimated via $h(\boldsymbol{x};\boldsymbol{\theta}) = \boldsymbol{\theta}^\top \boldsymbol{x}$, which satisfies Assumption 5. Then for any $\tau \in [0, TD]$ and any online algorithm, there exists a sequence of distributions $\mathcal{D}_1, \ldots, \mathcal{D}_T$ satisfying Assumptions 3 and 4 and a sequence of comparators $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T \in \boldsymbol{\Theta}$ satisfying Assumption 6, such that*

$$P_T(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T) \le \tau \;\text{ and }\; \mathbb{E}_{d_{1:T}}\left[\mathbf{Reg}_T(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T)\right] \ge \Omega\left(\sqrt{T\left(1 + \frac{\tau}{D}\right)}\right).$$

To prove Theorem 6, we first present Lemma 7, which establishes a lower bound for estimating HTEs in the static environment setting.

Lemma 7 *Suppose that the HTE is estimated via* $h(\boldsymbol{x};\boldsymbol{\theta}) = \boldsymbol{\theta}^\top \boldsymbol{x}$, *which satisfies Assumption 5. Then for any online algorithm that outputs* $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T \in \boldsymbol{\Theta}$ *satisfying Assumption 6, there exists a sequence of distributions* $\mathcal{D}_1, \dots, \mathcal{D}_T$ *satisfying Assumptions 3 and 4 and a sequence of unchanged comparators* $\boldsymbol{u}_1 = \cdots = \boldsymbol{u}_T = \boldsymbol{u} \in \boldsymbol{\Theta}$, *such that the online HTE estimation process has*

$$\mathbb{E}_{d_{1:T}}\left[\mathbf{Reg}_T(\boldsymbol{u}_1, \dots, \boldsymbol{u}_T)\right] = \mathbb{E}_{d_{1:T}}\left[\sum_{t=1}^{T} f_t(\boldsymbol{\theta}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{u})\right] \geq B\sqrt{T}.$$

Proof Let A, B, D, ε be positive reals such that $B \geq 2$, $A \geq 4B/D$, $1/2 > \varepsilon > 0$. Let $\mathbb{B} = \left\{\boldsymbol{x} \in \mathbb{R}^T \mid \|\boldsymbol{x}\|_2 \leq 1\right\}$ denote the unit ball. Let $\boldsymbol{\Theta} = \frac{D}{2}\mathbb{B}$ which satisfies Assumption 6. Let $\mathcal{X} = A\mathbb{B}$ and $\mathcal{Y} = [-B, B]$ which satisfy Assumption 4. We have

$$\mathbb{E}_{d_{1:T}}\left[\sum_{t=1}^{T} f_t(\boldsymbol{\theta}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{u})\right] = \frac{1}{2}\mathbb{E}_{d_{1:T}}\left[\sum_{t=1}^{T} \left(\boldsymbol{\theta}_t^\top \boldsymbol{x}_t - \tau_t(\boldsymbol{x}_t)\right)^2 - \sum_{t=1}^{T} \left(\boldsymbol{u}^\top \boldsymbol{x}_t - \tau_t(\boldsymbol{x}_t)\right)^2\right].$$

Without loss of generality, we assume $T \geq 2$. Let $\boldsymbol{e}_1, \dots, \boldsymbol{e}_T$ be the standard basis of $\mathbb{R}^T$. Referring to the construction of Cesa-Bianchi et al. (1996), upon receiving $\boldsymbol{\theta}_t$, the environment chooses a distribution $\mathcal{D}_t$ where $\boldsymbol{x}_t \equiv A\boldsymbol{e}_t$, $\mathbb{P}(w_t = 1 \mid \boldsymbol{x}_t) = \varepsilon$, and

$$y_t(1) = -\mathrm{sgn}\left[\boldsymbol{\theta}_t^\top \boldsymbol{x}_t\right]\left(\frac{B}{\sqrt{T}} + \frac{1}{2}\right) \text{ and } y_t(-1) = \mathrm{sgn}\left[\boldsymbol{\theta}_t^\top \boldsymbol{x}_t\right]\left(\frac{B}{\sqrt{T}} + \frac{1}{2}\right),$$

where $\mathrm{sgn}[a]$ equals 1 if $a \geq 0$ and -1 otherwise. It is easy to verify that $y_t(-1) \in \mathcal{Y}$ and $y_t(1) \in \mathcal{Y}$. We have $\tau_t(\boldsymbol{x}_t) = y_t(1) - y_t(-1) = -\mathrm{sgn}\left[\boldsymbol{\theta}_t^\top \boldsymbol{x}_t\right]\left(\frac{2B}{\sqrt{T}} + 1\right)$, which gives

$$\sum_{t=1}^{T} \left(\boldsymbol{\theta}_t^\top \boldsymbol{x}_t - \tau_t(\boldsymbol{x}_t)\right)^2 \geq T\left(\frac{2B}{\sqrt{T}} + 1\right)^2 = \left(2B + \sqrt{T}\right)^2. \tag{5}$$

Let $\boldsymbol{u} = \left(\frac{-\mathrm{sgn}\left[\boldsymbol{\theta}_1^\top \boldsymbol{x}_1\right]2B}{A\sqrt{T}}, \dots, \frac{-\mathrm{sgn}\left[\boldsymbol{\theta}_T^\top \boldsymbol{x}_T\right]2B}{A\sqrt{T}}\right) \in \boldsymbol{\Theta}$.

We have

$$\begin{aligned}\sum_{t=1}^{T} \left(\boldsymbol{u}^\top \boldsymbol{x}_t - \tau_t(\boldsymbol{x}_t)\right)^2 &= \sum_{t=1}^{T}\left(\frac{-\mathrm{sgn}\left[\boldsymbol{\theta}_t^\top \boldsymbol{x}_t\right]2B}{\sqrt{T}} + \mathrm{sgn}\left[\boldsymbol{\theta}_t^\top \boldsymbol{x}_t\right]\left(\frac{2B}{\sqrt{T}} + 1\right)\right)^2 \\ &= \sum_{t=1}^{T}\left(\frac{2B}{\sqrt{T}} - \left(\frac{2B}{\sqrt{T}} + 1\right)\right)^2 = T.\end{aligned} \tag{6}$$

Combining (5) and (6), we have

$$\begin{aligned}\mathbb{E}_{d_{1:T}}\left[\mathbf{Reg}_T(\boldsymbol{u}_1, \dots, \boldsymbol{u}_T)\right] &= \frac{1}{2}\left(\sum_{t=1}^{T} \left(\boldsymbol{\theta}_t^\top \boldsymbol{x}_t - \tau_t(\boldsymbol{x}_t)\right)^2 - \sum_{t=1}^{T} \left(\boldsymbol{u}^\top \boldsymbol{x}_t - \tau_t(\boldsymbol{x}_t)\right)^2\right) \\ &\geq \frac{1}{2}\left(\left(2B + \sqrt{T}\right)^2 - T\right) \geq B\sqrt{T}.\end{aligned}$$

□

Proof of Theorem 6 Let A, B, D, ε be positive reals such that $B \geq 2$, $A \geq 4B/D, 1/2 > \varepsilon > 0$. Let $\boldsymbol{\Theta} = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \leq D/2\}$ be a ball with radius $D/2$ which satisfies Assumption 6. Given any $\tau \in [0, TD]$, we consider $\tau < D$ and $\tau \geq D$ respectively.

We first consider $\tau < D$. According to Lemma 7, for any online algorithm, there exists a sequence of distributions $\mathcal{D}_1, \ldots, \mathcal{D}_T$ and a sequence of unchanged comparator $\boldsymbol{u}_1 = \cdots = \boldsymbol{u}_T = \boldsymbol{u}$ such that $P_T(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T) = 0 < \tau$ and $\mathbb{E}_{d_{1:T}}\left[\mathbf{Reg}_T\right] \geq B\sqrt{T} = \max\left\{B\sqrt{T}, B\sqrt{\tau T/D}\right\}$.

Next, we consider $\tau \geq D$. Without loss of generality, we assume $\lceil \tau/D \rceil$ divides T and divide the overall length-T online estimation problem into $\lceil \tau/D \rceil$ consecutive subproblems $S_1, \ldots, S_{\lceil \tau/D \rceil}$, each of which has length $L = T/\lceil \tau/D \rceil$. We apply Lemma 7 to each $S_i, i \in [\lceil \tau/D \rceil]$, and get a sequence of distributions $\mathcal{D}_1, \ldots, \mathcal{D}_T$ and a sequence of comparators $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T$ such that for all $i \in [\lceil \tau/D \rceil]$, $\boldsymbol{u}_{(i-1)L+1} = \boldsymbol{u}_{(i-1)L+2} = \cdots = \boldsymbol{u}_{iL}$, and that $\mathbb{E}_{d_{(i-1)L+1:iL}}\left[\mathbf{Reg}_{(i-1)L+1:iL}\right] \geq B\sqrt{L}$. Therefore, we have

$$\begin{aligned}\mathbb{E}_{d_{1:T}}\left[\mathbf{Reg}_T\right] &= \sum_{i=1}^{\lceil \tau/D \rceil} \mathbb{E}_{d_{(i-1)L+1:iL}}\left[\mathbf{Reg}_{(i-1)L+1:iL}\right] \\ &\geq \lceil \tau/D \rceil B\sqrt{L} \geq B\sqrt{\tau T/D} = \max\left\{B\sqrt{T}, B\sqrt{\tau T/D}\right\}.\end{aligned}$$

$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T$ changes at most $\lceil \tau/D \rceil - 1 \leq \tau/D$ times, so $P_T(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T) \leq \tau/D \cdot D = \tau$.

Combining above results, we have that for any $\tau \in [0, TD]$, there exists a sequence of distributions $\mathcal{D}_1, \ldots, \mathcal{D}_T$ and a sequence of comparators $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T$ such that $P_T(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T) \leq \tau$ and

$$\mathbb{E}_{d_{1:T}}\left[\mathbf{Reg}_T\right] \geq \max\left\{B\sqrt{T}, B\sqrt{\frac{\tau T}{D}}\right\} \geq \Omega\left(B\sqrt{T\left(1 + \frac{\tau}{D}\right)}\right).$$

□

6 Experiments

Due to the nature of the HTE estimation problem, we do not have access to the counterfactuals or the true CATE values from real-world observational data. And although the COVID example and recommender system motivate this work, the evolving treatment effects are ubiquitous in other scenarios as well. Therefore, we evaluate the performance of ETHOS on three semi-synthetic datasets that are commonly used in the causal inference literature with carefully designed environment evolving mechanisms. We first describe the data generation and the environment evolving mechanisms, then implementations and baselines, and finally the results and analysis.

IHDP. This dataset was first introduced for HTE estimation tasks by Hill (2011). The data come from a randomized experiment studying the effects of specialist home visits on future cognitive test scores. An imbalanced observational dataset is created by removing all children with non-white mothers in the treated group. The dataset consists of 25 covariates and records of 139 treated and 608 controlled units.

Twins. This dataset is constructed on the data of twins birth in the USA between 1989 and 1991. A total of 11,984 twin pairs are selected into records (Louizos et al., 2017). We

have 46 pre-treatment covariates. The treatment is being the heavier one in the twins, and the outcome is 1-year mortality.

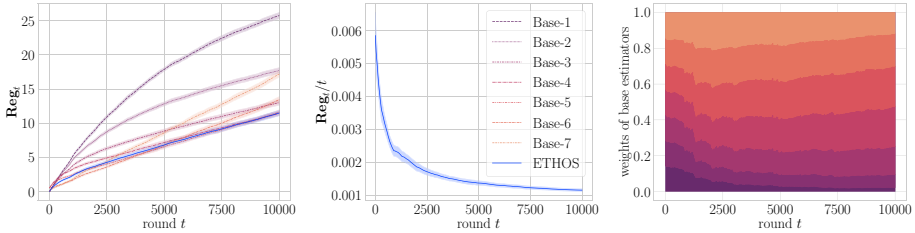
Jobs. This dataset is the combination of Lalonde experiment data and the PSID comparison group (Shalit et al., 2017). The covariates include 8 variables such as age, education, ethnicity, etc. The people in the treated group join a job training program. The outcome is employment status. It contains 297 treated and 2915 controlled units.

Outcome generation and treatment assignments. We use the covariates in all three datasets and simulate the outcomes and the treatment assignments. The covariates are scaled to have Euclidean norms smaller than 1. We consider *linear* and *nonlinear* responses for each dataset. The response surfaces are modified from Hill (2011) with proper scaling. The linear setting uses $Y(-1) \sim \mathcal{N}(\boldsymbol{X}^\top \boldsymbol{\beta}_1/\|\boldsymbol{\beta}_1\|, 0.1)$ and $Y(1) \sim \mathcal{N}(\boldsymbol{X}^\top \boldsymbol{\beta}_2/\|\boldsymbol{\beta}_2\|, 0.1)$, where the coefficients in $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ are randomly sampled values (0, 1, 2, 3, 4) with probabilities (0.5, 0.2, 0.15, 0.1, 0.05). The nonlinear setting uses $Y(-1) \sim \mathcal{N}(\exp((\boldsymbol{X}+0.05)^\top \boldsymbol{\beta}_3)/\exp(\|\boldsymbol{\beta}_3\|), 0.1)$, $Y(1) \sim \mathcal{N}(\boldsymbol{X}^\top \boldsymbol{\beta}_3/\|\boldsymbol{\beta}_3\|, 0.1)$, where the coefficients in $\boldsymbol{\beta}_3$ are randomly sampled values (0, 0.1, 0.2, 0.3, 0.4) with probabilities (0.6, 0.1, 0.1, 0.1, 0.1). As for the treatment assignments, we fit a logistic regression model $\ell : \mathcal{X} \to [0, 1]$ on a randomly sub-sampled data and each unit is treated with probability $\min\{\max\{\ell(\boldsymbol{X}), \varepsilon\}, 1-\varepsilon\}$ to satisfy Assumption 3.

Environment evolving mechanisms. We consider *switching* and *linear* mechanisms for the evolving environment. The T rounds are divided into K contiguous segments of length $L = T/K$. For each segment, we randomly sample 50% from a dataset as the test set, on which the true dynamic regret is computed. The samples revealed to the online HTE learner in the L rounds are sampled from the left 50%. In the switching case, for every segment, we re-generate the data following the procedure described in the previous paragraph, which mimics sudden environmental changes. In the linear evolving case, the related functions in each segment are linear combinations of two sets of functions, mimicking gradual changes in the environment. For example, let v_i, v_{i+1} denote CATE functions generated for the i-th and $(i+1)$-th segments respectively, then the CATE function at round $t \in [iL, (i+1)L)$ is $\tau_t(\cdot) = (1-\lambda_t)v_i(\cdot) + \lambda_t v_{i+1}(\cdot)$, where $\lambda_t = \max\{0, (t-(i+0.5)L)/(0.5L)\}$. Other related functions are also similarly generated.

Implementation. For all settings, we set $T = 10000$, $D = G = H = B = 1$, $\varepsilon = 0.05$. $\tilde{G}$ is computed as in Proposition 2. We use $h(\boldsymbol{x};\boldsymbol{\theta}) = \boldsymbol{\phi}(\boldsymbol{x})^\top \boldsymbol{\beta}$ as the HTE estimator. For linear responses, we set $\boldsymbol{\phi}(\boldsymbol{x}) = (\boldsymbol{x}, 1)$ to adjust for possible bias term. For nonlinear responses, $\boldsymbol{\phi}(\boldsymbol{x})$ is instantiated via a randomly initialized ReLU-activated multi-layer perceptron (MLP) without any training. The numbers of neurons in each layer are set to 256, 256, 256, and 2048. The parameters of the MLP are not updated during the online process, which fulfills the requirements of the theoretical analysis.

Baselines. Because the online HTE estimation problem is first proposed by this work, there do not exist any published baseline methods. Inspired by the sliding-window technique which is commonly used to handle non-stationarity (Cheung et al., 2019), we adopt common HTE estimation methods with different window size, which only learns from data within the nearest window. We make a simplification that only retrains an HTE estimator for disjoint windows due to the huge costs of repeated learning. The baselines include ordinary least squares with treatment as a feature (OLS), k-nearest neighbor (k-NN), propensity score matching (PSM), random forest (RF), BART (Chipman et al., 2010; Hill, 2011), causal forest (CF) (Wager & Athey, 2018; Athey et al., 2019), balancing neural network (BNN) (Johansson et al., 2016), treatment-agnostic representation network (TARNet)



(a) (b) (c)

Fig. 1 The figures record the running process of ETHOS for the setting with linear responses, switching environments, and $K = 100$. The results are averaged over 10 replications. The legend in the middle is shared by three figures. The bands depict standard errors. **a** plots the dynamic regret of ETHOS and its seven base estimators against the optimal comparator sequences, showing that ETHOS is able to catch up with the best base estimator. **b** plots the ratio between the dynamic regret of ETHOS in t rounds and t, which almost decreases to zero as t increases, verifying the sublinear dynamic regret bound in Theorem 4. **c** shows the change of base estimator weights $\{\omega_i\}_{i=1}^7$ during the running, where the weights of best estimators keep increasing

Table 1 The average cumulative loss (squared PEHE) and standard errors of ETHOS and baselines for the setting with linear responses and switching environments

Method	IHDP		Twins		Jobs	
	$K = 10$	$K = 100$	$K = 10$	$K = 100$	$K = 10$	$K = 100$
OLS	14.4 ± 0.5	17.2 ± 0.6	4.3 ± 0.2	9.9 ± 0.3	8.1 ± 0.5	9.0 ± 0.3
k-NN	29.6 ± 0.6	36.6 ± 0.8	20.2 ± 0.2	22.7 ± 0.3	23.9 ± 0.6	24.4 ± 0.7
PSM	32.3 ± 0.8	36.0 ± 0.9	20.5 ± 0.3	22.9 ± 0.3	23.8 ± 0.6	24.3 ± 0.7
RF	24.4 ± 0.7	31.0 ± 0.6	10.2 ± 0.4	13.4 ± 0.3	5.2 ± 0.3	5.4 ± 0.2
BART	15.4 ± 0.6	19.1 ± 0.6	4.9 ± 0.2	9.5 ± 0.3	4.1 ± 0.3	4.5 ± 0.1
CF	13.6 ± 0.5	16.2 ± 0.5	4.6 ± 0.1	8.1 ± 0.2	7.6 ± 0.5	8.2 ± 0.2
BNN	57.0 ± 1.9	54.6 ± 0.5	18.8 ± 1.3	19.7 ± 0.6	8.4 ± 0.2	11.6 ± 0.1
TARNet	32.8 ± 0.8	34.0 ± 0.9	6.9 ± 0.3	13.1 ± 0.4	8.9 ± 0.3	9.2 ± 0.4
CFR	30.5 ± 0.8	33.2 ± 1.2	6.9 ± 0.2	12.7 ± 0.4	8.8 ± 0.3	9.0 ± 0.4
ETHOS	**10.5 ± 0.5**	**11.5 ± 0.4**	**4.0 ± 0.2**	**5.3 ± 0.2**	**1.0 ± 0.1**	**1.3 ± 0.1**

Results with the smallest average losses are displayed in bold font

(Shalit et al., 2017), and counterfactual regression with Wasserstein distance (CFR) (Shalit et al., 2017). The window sizes include 25, 50, and 100.

Results and analysis. For each setting, we conducted the experiment 10 times with different random seeds and report the average results and standard error. Figure 1 illustrates several quantities observed during the execution of ETHOS in a particular setting. The effectiveness of the online ensemble technique is evident as the weights of good base estimators keep increasing and the performance of ETHOS is as good as the best base estimator. Furthermore, the dynamic regret clearly demonstrates a sublinear trend, providing further validation of the theoretical analysis in Sect. 5.

Next, we compared ETHOS with baselines. We ran all baseline methods using three different window sizes, but due to space limitations, we only present the results for the

Table 2 Results for the setting with linear responses and linear changes

Method	IHDP		Twins		Jobs	
	$K = 10$	$K = 100$	$K = 10$	$K = 100$	$K = 10$	$K = 100$
OLS	12.2 ± 0.5	15.5 ± 0.4	4.3 ± 0.2	6.6 ± 0.3	6.7 ± 0.2	8.4 ± 0.4
k-NN	26.2 ± 0.6	30.9 ± 0.4	17.1 ± 0.2	18.1 ± 0.2	20.6 ± 0.7	21.5 ± 0.5
PSM	28.3 ± 0.7	31.5 ± 0.4	17.5 ± 0.2	18.0 ± 0.2	20.5 ± 0.7	21.5 ± 0.5
RF	20.3 ± 0.3	26.3 ± 0.4	9.1 ± 0.3	11.0 ± 0.3	3.9 ± 0.3	4.9 ± 0.2
BART	13.2 ± 0.6	15.8 ± 0.5	3.9 ± 0.2	7.2 ± 0.3	3.3 ± 0.2	4.3 ± 0.1
CF	11.9 ± 0.5	14.4 ± 0.4	4.9 ± 0.1	5.4 ± 0.1	6.5 ± 0.3	8.0 ± 0.3
BNN	55.9 ± 1.9	53.9 ± 0.4	18.2 ± 1.1	19.4 ± 0.6	7.3 ± 0.2	7.5 ± 0.1
TARNet	31.1 ± 1.2	35.7 ± 1.4	6.2 ± 0.3	9.9 ± 0.4	7.8 ± 0.5	8.5 ± 0.5
CFR	29.2 ± 1.2	33.6 ± 1.3	6.1 ± 0.3	9.8 ± 0.4	7.8 ± 0.5	8.4 ± 0.5
ETHOS	**9.3 ± 0.4**	**10.3 ± 0.3**	**3.8 ± 0.3**	**4.9 ± 0.2**	**0.8 ± 0.1**	**1.0 ± 0.1**

Results with the smallest average losses are displayed in bold font

Table 3 Results for the setting with nonlinear responses and switching changes

Method	IHDP		Twins		Jobs	
	$K = 10$	$K = 100$	$K = 10$	$K = 100$	$K = 10$	$K = 100$
k-NN	33.5 ± 0.7	48.4 ± 1.1	23.0 ± 0.2	27.2 ± 0.6	53.0 ± 2.8	111.5 ± 3.1
PSM	33.9 ± 0.9	48.1 ± 1.2	23.0 ± 0.3	27.0 ± 0.6	53.0 ± 2.8	111.3 ± 3.1
RF	34.5 ± 0.8	48.4 ± 1.4	27.6 ± 0.4	32.7 ± 0.4	58.7 ± 2.7	102.2 ± 3.7
BART	22.6 ± 1.5	44.4 ± 1.2	20.0 ± 0.7	27.9 ± 0.7	43.2 ± 2.7	105.3 ± 5.0
CF	16.8 ± 0.7	30.5 ± 0.6	**10.2 ± 0.5**	**16.2 ± 0.4**	30.8 ± 1.5	61.1 ± 2.4
BNN	122.3 ± 2.9	119.2 ± 2.6	90.3 ± 2.2	89.3 ± 4.4	230.1 ± 7.2	236.7 ± 5.0
TARNet	101.1 ± 2.6	114.3 ± 1.3	59.4 ± 2.7	70.7 ± 1.1	143.2 ± 3.2	168.5 ± 2.9
CFR	76.8 ± 3.9	92.9 ± 2.0	53.6 ± 2.3	64.4 ± 1.0	137.8 ± 3.6	162.7 ± 3.8
ETHOS	**15.4 ± 0.8**	**22.9 ± 1.0**	45.0 ± 4.2	44.0 ± 1.9	**19.8 ± 0.8**	**34.7 ± 1.6**

Results with the smallest average losses are displayed in bold font

Table 4 Results for the setting with nonlinear responses and linear changes

Method	IHDP		Twins		Jobs	
	$K = 10$	$K = 100$	$K = 10$	$K = 100$	$K = 10$	$K = 100$
k-NN	28.7 ± 0.8	32.0 ± 1.0	19.5 ± 0.3	20.0 ± 0.3	42.3 ± 2.6	69.2 ± 2.1
PSM	28.8 ± 1.0	32.1 ± 1.1	19.5 ± 0.3	20.2 ± 0.3	41.9 ± 2.7	69.2 ± 2.1
RF	29.1 ± 0.8	33.1 ± 0.6	24.3 ± 0.4	24.9 ± 0.2	47.6 ± 2.5	67.9 ± 3.4
BART	19.3 ± 1.3	28.7 ± 1.2	16.8 ± 0.6	20.1 ± 0.8	33.8 ± 3.0	61.9 ± 2.1
CF	16.2 ± 0.7	27.5 ± 0.8	**8.9 ± 0.4**	**10.9 ± 0.3**	25.4 ± 1.4	36.5 ± 1.3
BNN	122.0 ± 2.9	119.9 ± 2.5	90.0 ± 2.3	88.2 ± 2.5	230.3 ± 4.9	232.4 ± 4.9
TARNet	98.6 ± 2.9	102.1 ± 2.1	58.4 ± 2.2	62.4 ± 1.2	141.8 ± 4.1	147.6 ± 2.7
CFR	74.3 ± 2.3	79.8 ± 1.6	53.2 ± 2.3	55.1 ± 1.5	139.2 ± 4.0	142.6 ± 2.9
ETHOS	**14.8 ± 1.0**	**20.5 ± 0.6**	34.3 ± 2.7	40.3 ± 2.1	**17.3 ± 0.9**	**27.1 ± 1.2**

Results with the smallest average losses are displayed in bold font

best configurations in Tables 1, 2, 3 and 4. Tables 1 and 2 show the results for linear responses with switching and linear environment changes, respectively, where the environment changes occurred 10 or 100 times. ETHOS consistently achieves significantly lower cumulative loss compared to all baselines across all settings. Notably, the performance gap between ETHOS and OLS, which is also a linear model, indicates that ETHOS's superior performance is not solely due to the assumption of linearity in the responses but also its ability to handle unknown environmental changes effectively. Tables 3 and 4 report the results on nonlinear responses with respective switching and linear environment changes. ETHOS still outperforms all baselines on two datasets, despite the true HTE in these settings now falling outside the hypothesis space of ETHOS. It is worth noting that while neural network-based methods, such as CFR, have a similar hypothesis space, ETHOS outperforms them, indicating its strong adaptability to evolving environments. Improved performance in nonlinear settings could potentially be achieved by carefully designing the nonlinear mapping function ϕ or initializing the model with pre-trained models using auxiliary data. Additionally, we observed that tree-based methods, such as CF, perform better than neural network-based methods in these settings, which could be attributed to the robustness of trees in low-data regimes, where deep neural networks are prone to overfitting.

Moreover, we show the cumulative PEHEs for the linear setting in Table 5, showcasing that while ETHOS is derived from theoretical analysis on the squared PEHE loss (the mean squared error), it also performs well when evaluated with PEHE (the root mean squared error) as the performance measure.

7 Conclusion

In this paper, we introduce and formulate the online HTE estimation problem in evolving environments. Leveraging the online ensemble technique, we propose a novel method called ETHOS, which enables the tracking of treatment effect heterogeneity in evolving environments with unknown environmental changes. The optimality of ETHOS is

Table 5 The average cumulative PEHE and standard errors for the setting with linear responses and switching environment changes

Method	IHDP		Twins		Jobs	
	$K = 10$	$K = 100$	$K = 10$	$K = 100$	$K = 10$	$K = 100$
OLS	3.8 ± 0.1	4.1 ± 0.1	2.1 ± 0.0	3.1 ± 0.1	2.8 ± 0.1	3.0 ± 0.1
k-NN	5.4 ± 0.1	6.1 ± 0.1	4.5 ± 0.0	4.8 ± 0.0	4.9 ± 0.1	4.9 ± 0.1
PSM	5.7 ± 0.1	6.0 ± 0.1	4.5 ± 0.0	4.8 ± 0.0	4.9 ± 0.1	4.9 ± 0.1
RF	4.9 ± 0.1	5.6 ± 0.1	3.2 ± 0.1	3.7 ± 0.0	2.3 ± 0.1	2.3 ± 0.0
BART	3.9 ± 0.1	4.4 ± 0.1	2.2 ± 0.0	3.1 ± 0.0	2.0 ± 0.1	2.1 ± 0.0
CF	3.7 ± 0.1	4.0 ± 0.1	2.1 ± 0.0	2.8 ± 0.0	2.7 ± 0.1	2.9 ± 0.0
BNN	7.5 ± 0.1	7.4 ± 0.0	4.3 ± 0.1	4.4 ± 0.1	3.1 ± 0.1	3.2 ± 0.0
TARNet	5.7 ± 0.1	5.8 ± 0.1	2.6 ± 0.1	3.6 ± 0.1	3.0 ± 0.1	3.0 ± 0.1
CFR	5.5 ± 0.1	5.8 ± 0.1	2.6 ± 0.0	3.6 ± 0.1	3.0 ± 0.1	3.0 ± 0.1
ETHOS	**3.2 ± 0.1**	**3.4 ± 0.1**	**1.9 ± 0.1**	**2.3 ± 0.0**	**1.0 ± 0.1**	**1.1 ± 0.0**

Results with the smallest average losses are displayed in bold font

established through a problem lower bound that matches its achieved expected dynamic regret. Experimental results provide empirical evidence supporting the effectiveness of the proposed method and the validity of the theoretical analysis. We believe that our proposed method will benefit HTE estimation in diverse real online applications and hope that this work would inspire further investigations.

The challenging online HTE estimation problem in evolving environments offers ample opportunities for diverse future research. For the specific problem proposed in this work, potential future work may include exploring alternative measures to characterize the evolution of the environment, rather than relying solely on P_T, to better suit various scenarios; designing more efficient methods, e.g., with fewer base estimators; and extending the analysis from convex functions to non-convex ones, which could provide greater flexibility in modeling the treatment effects. Moreover, we can consider the broader problem of causal inference in evolving environments, where potential directions may include considering evolving sets of confounding covariates or the existence of latent confounders (Wang et al., 2020, 2023a, b); improving HTE estimations via learning the confounder structure (Lv et al., 2021; Qin et al., 2023); and grounding decision-making (Zhou, 2022; Wang et al., 2022) with HTE estimators in dynamic environments.

Acknowledgements This research was supported by the National Key R&D Program of China (2022ZD0114800) and National Science Foundation of China (61921006).

Author Contributions Tian Qin conceived and developed the procedure, finished the proofs, conducted the experiments and wrote the manuscript. Long-Fei Li helped with theoretical analysis. Tian-Zuo Wang revised the manuscript. Zhi-Hua Zhou conceived the study and was in charge of the overall direction and planning. All authors discussed the results and contributed to the final manuscript.

Data availibility The data used in this work are all public.

Code availability The code will be released after publishing.

Declarations

Conflict of interest All authors declare that have no conflict of interest.

Consent to participate Not applicable.

Consent for publication Not applicable.

Ethics approval Not applicable.

References

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics, 47*(2), 1148–1178.

Bica, I., & van der Schaar, M. (2022). Transfer learning on heterogeneous feature spaces for treatment effects estimation. *Advances in Neural Information Processing Systems, 35*, 37184.

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics, 9*(1), 247–274.

Cesa-Bianchi, N., Long, P., & Warmuth, M. (1996). Worst-case quadratic loss bounds for prediction using linear functions and gradient descent. *IEEE Transactions on Neural Networks, 7*(3), 604–619. https://doi.org/10.1109/72.501719

Cesa-Bianchi, N., & Lugosi, G. (2006). *Prediction, Learning, and Games*. Cambridge University Press.

Cheung, W. C., Simchi-Levi, D., & Zhu, R. (2019). Learning to optimize under non-stationarity. In *Proceedings of the 22nd international conference on artificial intelligence and statistics*, pp. 1079–1087.
Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). 03. BART: Bayesian additive regression trees. *Annals of Applied Statistics, 4*(1), 266–298.
Hahn, P. R., Murray, J. S., & Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects. *Bayesian Analysis, 15*(3), 965–1056.
Harada, S., & Kashima, H. (2022). InfoCEVAE: Treatment effect estimation with hidden confounding variables matching. *Machine Learning*. https://doi.org/10.1007/s10994-022-06246-0
Hazan, E. (2016). Introduction to online convex optimization. *Foundations and Trends in Optimization, 2*(3–4), 157–325.
Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics, 20*(1), 217–240.
Imbens, G. W., & Rubin, D. B. (2015). In: *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.
Johansson, F. D., Kallus, N., Shalit, U., Sontag, D. A. (2018). Learning weighted representations for generalization across designs. arXiv:abs/1802.08598.
Johansson, F. D., Shalit, U., & Sontag, D. A. (2016). Learning representations for counterfactual inference. In *Proceedings of the 33rd international conference on machine learning*, pp. 3020–3029.
Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences, 116*(10), 4156–4165.
Künzel, S. R., Stadie, B. C., Vemuri, N., Ramakrishnan, V., Sekhon, J. S., & Abbeel, P. (2018). Transfer learning for estimating causal effects using neural networks. arXiv:abs/1808.07804.
Li, S., & Buhlmann, P. (2018). Estimating heterogeneous treatment effects in nonstationary time series with state-space models. arXiv:abs/1812.04063.
Louizos, C., Shalit, U., Mooij, J. M., Sontag, D. A., Zemel, R. S., & Welling, M. (2017). Causal effect inference with deep latent-variable models. *Advances in Neural Information Processing Systems, 30*, 6446–6456.
Lv, Y., Miao, J., Liang, J., Chen, L., & Qian, Y. (2021). BIC-based node order learning for improving Bayesian network structure learning. *Frontiers of Computer Science, 15*(6), 156337. https://doi.org/10.1007/s11704-020-0268-6
Neyman, J. (1923). Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczych, 10*, 1–51.
Nie, X., & Wager, S. (2020). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika, 108*, 299.
Qin, T., Wang, T. Z., & Zhou, Z. H. (2021). Budgeted heterogeneous treatment effect estimation. In *Proceedings of the 38th international conference on machine learning*, pp. 8693–8702.
Qin, T., Wang, T. Z., & Zhou, Z. H. (2023). Learning causal structure on mixed data with tree-structured functional models. In *Proceedings of the 23rd SIAM international conference on data mining*, pp. 613–621.
Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688.
Schnabel, T., Swaminathan, A., Singh, A., Chandak, N., & Joachims, T. (2016). Recommendations as treatments: Debiasing learning and evaluation. In *Proceedings of the 33rd international conference on machine learning*, pp. 1670–1679.
Shalit, U. (2019). Can we learn individual-level treatment policies from clinical data? *Biostatistics, 21*(2), 359–362.
Shalit, U., Johansson, F. D., & Sontag, D. A. (2017). Estimating individual treatment effect: Generalization bounds and algorithms. In *Proceedings of the 34th international conference on machine learning*, pp. 3076–3085.
Shi, C., Veitch, V., & Blei, D. M. (2021). Invariant representation learning for treatment effect estimation. In *Proceedings of the 37th conference on uncertainty in artificial intelligence*, pp. 1546–1555.
Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association, 113*(523), 1228–1242.
Wang, H., Yu, Y., & Jiang, Y. (2022). Review of the progress of communication-based multi-agent reinforcement learning. *SCIENTIA SINICA Informationis, 52*(5), 742–764. https://doi.org/10.1360/SSI-2020-0180
Wang, P., Sun, W., Yin, D., Yang, J., & Chang, Y. (2015). Robust tree-based causal inference for complex ad effectiveness analysis. In *Proceedings of the 8th ACM international conference on web search and data mining*, pp. 67–76.

Wang, T. Z., Qin, T., & Zhou, Z. H. (2023a). Estimating possible causal effects with latent variables via adjustment. In *Proceedings of the 40th international conference on machine learning*, pp. 36308–36335.

Wang, T. Z., Qin, T., & Zhou, Z. H. (2023). Sound and complete causal identification with latent variables given local background knowledge. *Artificial Intelligence, 322*, 103964. https://doi.org/10.1016/j.artint.2023.103964

Wang, T. Z., Wu, X. Z., Huang, S. J., & Zhou, Z. H. (2020). Cost-effectively identifying causal effects when only response variable is observable. In *Proceedings of the 37th international conference on machine learning*, pp. 10060–10069.

Yao, L., Li, S., Li, Y., Huai, M., Gao, J., & Zhang, A. (2018). Representation learning for treatment effect estimation from observational data. In *Advances in neural information processing systems*, pp. 2638–2648.

Yoon, J., Jordon, J., & van der Schaar, M. (2018). GANITE: Estimation of individualized treatment effects using generative adversarial nets. In *Proceedings of the 6th international conference on learning representations*.

Zhang, L., Lu, S., & Zhou, Z. H. (2018). Adaptive online learning in dynamic environments. In *Advances in Neural Information Processing Systems 31*, pp. 1330–1340.

Zhang, W., Liu, L., & Li, J. (2021). Treatment effect estimation with disentangled latent factors. In *35th AAAI conference on artificial intelligence*, pp. 10923–10930.

Zhao, P., Wang, G., Zhang, L., & Zhou, Z. H. (2021). Bandit convex optimization in non-stationary environments. *Journal of Machine Learning Research, 22*, 1–45.

Zhao, P., Zhang, Y. J., Zhang, L., & Zhou, Z. H. (2021). Adaptivity and non-stationarity: Problem-dependent dynamic regret for online convex optimization. arXiv:abs/2112.14368.

Zhou, Z. H. (2012). *Ensemble methods: Foundations and algorithms*. CRC Press.

Zhou, Z. H. (2022). Open-environment machine learning. *National Science Review, 9*(8), 123. https://doi.org/10.1093/nsr/nwac123

Zhou, Z. H. (2022). Rehearsal: Learning from prediction to decision. *Frontiers of Computer Science, 16*(4), 164352.

Zhou, Z. H., & Tan, Z. H. (2023). Learnware: Small models do big. *Science China Information Sciences*. https://doi.org/10.1007/s11432-023-3823-6

Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning*, pp. 928–936.