Kernel Treatment Effects with Adaptively Collected Data

Houssam Zenati[1] **Bariscan Bozkurt**[1] **Arthur Gretton**[1,2]

[1]Gatsby Computational Neuroscience Unit, University College London [2]Google DeepMind

Abstract

Adaptive experiments improve efficiency by adjusting treatment assignments based on past outcomes, but this adaptivity breaks the i.i.d. assumptions that underpin classical asymptotics. At the same time, many questions of interest are distributional, extending beyond average effects. Kernel treatment effects (KTE) provide a flexible framework by representing interventional outcome distributions in an RKHS and comparing them via kernel distances. We present the first kernel-based framework for distributional inference under adaptive data collection. Our method combines doubly robust RKHS scores with a witness function learned on one fold, and performs inference on a second fold using a projected, sequentially normalized scalar statistic with valid type-I error. Experiments show that the resulting procedure is well calibrated and effective for both mean shifts and higher-moment differences, outperforming adaptive baselines limited to scalar effects.

1 INTRODUCTION

Data in modern experiments are increasingly collected *adaptively*, with treatment assignments chosen sequentially in response to past outcomes, as in multi-armed and contextual bandits (Lattimore and Szepesvári, 2020), adaptive clinical trials (Chow and Chang, 2011), and dynamic pricing strategies in economics (Qiang and Bayati, 2016; Athey et al., 2022). Adaptivity improves participant welfare and accelerates learning during data collection, but it fundamentally alters the statistical properties of the data: allocation proportions and effective sample sizes become random and history-dependent (Bibaut and Kallus, 2025). This breaks the classical i.i.d. assumptions that underlie standard asymptotic theory (Vaart and Wellner, 1997), and as a consequence, estimators that are asymptotically normal under fixed designs may converge to non-normal limits or exhibit inflated variances (Bibaut and Kallus, 2025).

Simultaneously, reliance on the average effects is often insufficient, as many scientific and practical questions are inherently *distributional*. In medicine, clinicians care not only about mean efficacy but also about the distribution of side effects across patients (Rothe, 2010); in finance and operations, decision-makers evaluate policies using tail-sensitive criteria such as conditional value-at-risk (CVaR) (Rockafellar et al., 2000); and in reinforcement learning, distributional approaches explicitly target higher moments or quantiles of return distributions (Dabney et al., 2018). Existing statistical methods often rely on cumulative distribution functions (Chernozhukov et al., 2013; Huang et al., 2021), which become difficult to extend to high-dimensional or structured outcomes.

Kernel methods provide a powerful alternative. Kernel mean embeddings can represent interventional outcome distributions as elements of a reproducing kernel Hilbert space (RKHS) (Berlinet and Thomas-Agnan, 2011; Gretton, 2013; Muandet et al., 2021), enabling nonparametric comparison of distributions via kernel distances and supporting inference on complex outcomes such as images, sequences, or graphs (Gärtner, 2003). This framework has been used to define conditional distributional kernel treatment effects (Park et al., 2021), to design kernel-based hypothesis tests (Shekhar et al., 2023; Martinez Taboada et al., 2023; Fawkes et al., 2024), and to extend efficiency theory to Hilbert-space parameters (Luedtke and Chung, 2024). However, all existing KTE methods assume i.i.d. data, and it remains unknown how to conduct distributional causal inference when outcomes are observed under adaptive, history-dependent policies.

In this paper, we develop the first framework for *kernel*

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

treatment effect inference under adaptive data collection. Our contributions are as follows: *i)* we construct a doubly robust procedure that learns an RKHS witness on a first chronological fold and performs inference on a second fold using projected kernel scores and sequential normalization based only on past data; *ii)* we develop a reweighted plug-in estimator of the projected conditional standard deviation and prove its consistency under adaptive logging; *iii)* we obtain a sample-split adaptive test with valid Gaussian limits under the null; and *iv)* we validate the method in numerical simulations, including semi-synthetic benchmarks and high-dimensional structured outcomes such as images, showing that it detects distributional effects missed by scalar adaptive baselines.

Conceptually, our work connects kernel-based distributional causal inference with adaptive inference. Rather than deriving a Gaussian limit for the full RKHS-valued effect[1], we reduce inference to a scalar adaptive statistic while retaining sensitivity to distributional alternatives through a learned witness.

The remainder of the paper is structured as follows. Section 2 reviews related work. Section 3 formalizes the adaptive setting and KTE. Section 4 introduces our adaptive test statistic. We detail plug-in variance estimation in Section 5 and the practical procedure and power guarantees in Section 6. Section 7 reports simulations, and Section 8 concludes.

2 RELATED WORKS

Kernel mean embeddings (Smola et al., 2007) provide a nonparametric representation of distributions in RKHS and a way to compare them through inner products and norms (Kanagawa and Fukumizu, 2014; Sriperumbudur et al., 2010; Gretton et al., 2012a). Building on this framework, Muandet et al. (2021) introduced *Counterfactual Mean Embeddings* to represent full interventional outcome distributions and define distributional treatment effects under unconfoundedness; related work expressed ATE and CATE through conditional mean embeddings, yielding RKHS formulations of heterogeneous effects (Park et al., 2021; Singh et al., 2024). On the inferential side, Fawkes et al. (2024) proposed doubly robust kernel statistics for testing equality of interventional outcome distributions, Martinez Taboada et al. (2023) developed an efficient doubly robust kernel test with improved power and valid type-I error control, and (Luedtke and Chung, 2024; Zenati et al., 2025) established estimation guarantees for doubly robust estimators of interventional mean embeddings in non-adaptive settings.

Adaptive experimentation arises in multi-armed bandits, best-arm identification, adaptive clinical trials, contextual bandits, batch bandits, sequential policy learning, and dynamic pricing (Lattimore and Szepesvári, 2020; Garivier and Kaufmann, 2016; Chow and Chang, 2011; Zenati et al., 2022; Li et al., 2010; Perchet et al., 2016; Zenati et al., 2023; Qiang and Bayati, 2016). While such designs improve performance during data collection, they make inference harder because allocation proportions and effective sample sizes become random and history-dependent (Athey et al., 2022; Caria et al., 2023), breaking the classical i.i.d. asymptotic framework (van der Vaart, 1998; Hall and Heyde, 1980). For adaptive inference, Hadad et al. (2021) showed that suitable reweighting can recover approximate normality for policy evaluation, while related stabilization approaches include conditional-variance weighting and adaptive weighting without outcome models (Bibaut et al., 2021; Zhang et al., 2021); see also Zhang et al. (2020) for batched bandits, Howard et al. (2021); Waudby-Smith and Ramdas (2021) for always-valid inference, and Bibaut and Kallus (2025); Hirano and Porter (2023) for a broader view of when Gaussian limits fail or can be restored. Our work extends this literature to *distributional kernel treatment effects* under contextual adaptivity by combining a sample-split kernel witness, projected doubly robust scores, and sequential normalization.

3 PROBLEM STATEMENT

We formalize the estimation of *kernel treatment effects* (KTE) when data are collected via an *adaptive experiment* (e.g., contextual bandit algorithm).

3.1 Adaptive data collection setting

We consider a contextual decision-making system operating over T rounds. At each round $t \in \{1, \dots, T\}$, the agent observes a context $X_t \in \mathcal{X}$, sampled independently from an unknown distribution P_X, i.e., $X_t \sim P_X$. Given X_t, the agent selects an action $A_t \in \mathcal{A}$ according to a possibly adaptive policy $\pi_t \in \Pi$, such that $A_t \sim \pi_t(\cdot \mid \mathcal{F}_{t-1}, X_t)$, where $\mathcal{F}_{t-1} := \sigma(X_1, A_1, Y_1, \dots, X_{t-1}, A_{t-1}, Y_{t-1})$ denotes the filtration up to time $t-1$. The outcome $Y_t \in \mathcal{Y}$ is then generated according to a fixed, unknown outcome model $Y_t \sim P_{Y|X,A}(\cdot \mid X_t, A_t)$, depending only on the current context and action. We assume that the action space $\mathcal{A}$ is discrete and the outcome space $\mathcal{Y}$ may be either discrete or continuous, and that each policy π_t admits a density with respect to a base measure $\mu_{\mathcal{A}}$. The sequence of policies $\{\pi_t\}_{t=1}^T$ may depend on past

[1] An earlier version of this work pursued a Gaussian limit for the difference of interventional mean embeddings themselves, rather than for the test statistic directly. See Appendix C.

observations, rendering the overall data-generating process adaptive rather than i.i.d. The observed dataset consists of the trajectory $\mathcal{D}_T = \{(X_t, A_t, Y_t)\}_{t=1}^T$. We assume the existence of a potential outcome function $a \mapsto Y_t(a)$ such that $Y_t = Y_t(A_t)$, and that the collection $\{Y_t(a)\}_{a\in\mathcal{A}}$ is conditionally independent of A_t given X_t, i.e., conditional ignorability holds.

3.2 Target Estimand

We work throughout with an outcome kernel $k_\mathcal{Y}$ satisfying the following standing condition.

Assumption 3.1 (Outcome kernel)**.** *The kernel $k_\mathcal{Y}$ is bounded and characteristic. That is, there exists $\kappa < \infty$ such that $\sup_{y\in\mathcal{Y}} k_\mathcal{Y}(y, y) \le \kappa$, and the associated mean embedding of probability measures into $\mathcal{H}_\mathcal{Y}$ is injective.*

Let $\mathcal{H}_\mathcal{Y}$ denote the RKHS associated with $k_\mathcal{Y}$, and let $\phi_\mathcal{Y}(y) = k_\mathcal{Y}(\cdot, y)$ be its feature map. We first introduce the *interventional mean embedding* (IME) (Muandet et al., 2021)[2] of the interventional outcome distribution of $Y(a)$ for $a \in \mathcal{A}$:

$$\eta(a) := \mathbb{E}_{P_X \times P_{Y|X,A}} [\phi_\mathcal{Y}(Y(a))], \tag{IME}$$

where the expectation is taken over the fixed marginal distribution of contexts P_X and the conditional outcome distribution $P_{Y|X,A}$. Then, the *target estimand* is the generalized *kernel treatment effect* (KTE) which is defined as expressed as the MMD of the two interventional mean embeddings $\eta(a)$ and $\eta(a')$, that is the RKHS norm of the difference Ψ:

$$\tau(a, a') := \|\Psi(a, a')\|_{\mathcal{H}_\mathcal{Y}}, \tag{KTE}$$
$$\Psi(a, a') := \eta(a) - \eta(a'), \tag{1}$$

Because $k_\mathcal{Y}$ is characteristic, $\tau(a, a') = 0$ if and only if the two interventional outcome distributions under a and a' coincide. Now, define the following conditional mean embedding (CME) (Song et al., 2009) of the distribution $P_{Y|X,A}$:

$$\mu_{Y|A,X}(a, x) := \mathbb{E}_{P_{Y|X,A}}[\phi_\mathcal{Y}(Y) \mid A = a, X = x]. \tag{2}$$

Assumption 3.2 (Sequential ignorability and positivity)**.** *Fix the target actions $a, a' \in \mathcal{A}$. For each $b \in \{a, a'\}$ and each $t \ge 1$: i) Consistency: $Y_t = Y_t(b)$ on the event $\{A_t = b\}$. ii) Sequential exchangeability: $Y_t(b) \perp A_t \mid (X_t, \mathcal{F}_{t-1})$. iii) Positivity: $\pi_t(b \mid X_t) > 0$ almost surely.*

Under Assumption 3.2, the (IME) can be identified from observable data (Muandet et al., 2021; Zenati et al., 2025):

$$\eta(a) = \mathbb{E}_{P_X} \left[\mu_{Y|A,X}(a, x)\right]. \tag{3}$$

To construct doubly robust scores, let π be any conditional density on $\mathcal{A} \times \mathcal{X}$ and let $\bar{\mu} : \mathcal{A} \times \mathcal{X} \to \mathcal{H}_\mathcal{Y}$ be any measurable function. For discrete $\mathcal{A}$, define the uncentered doubly robust (DR) score

$$D'(\pi, \bar{\mu}, a)(X, A, Y) := \frac{\mathbb{1}\{A = a\}}{\pi(a \mid X)} \Big(\phi_\mathcal{Y}(Y) - \bar{\mu}(a, X)\Big) + \bar{\mu}(a, X). \tag{4}$$

Its centered version

$$D(\pi, \bar{\mu}, a) := D'(\pi, \bar{\mu}, a) - \eta(a)$$

coincides with the canonical gradient of the (IME) (Luedtke and Chung, 2024). In what follows, we work directly with the uncentered score D', since its conditional expectation equals the target embedding.

3.3 Why naive kernel tests fail under adaptivity

Let $(\widehat{\mu}^{(t)}_{Y|A,X})_{t\ge1}$ denote a sequence of estimators of the conditional mean embedding $\mu_{Y|A,X}$, where each $\widehat{\mu}^{(t)}_{Y|A,X} : \mathcal{A} \times \mathcal{X} \to \mathcal{H}_\mathcal{Y}$ is trained using data up to round t, and define the doubly robust RKHS score difference

$$\hat{\phi}_t := D'(\pi_t, \widehat{\mu}^{(t-1)}_{Y|A,X}, a)(X_t, A_t, Y_t) - D'(\pi_t, \widehat{\mu}^{(t-1)}_{Y|A,X}, a')(X_t, A_t, Y_t) \in \mathcal{H}_\mathcal{Y}. \tag{5}$$

In i.i.d. settings, averages of $(\hat{\phi}_t)$ admit Gaussian fluctuations in $\mathcal{H}_\mathcal{Y}$ under standard conditions. However, for testing

$$H_0 : \Psi(a, a') = 0,$$

the relevant object is a quadratic functional of this average, and the direct plug-in statistic

$$\left\|\widehat{\Psi}_T(a, a')\right\|^2_{\mathcal{H}_\mathcal{Y}}$$

is degenerate under the null, as in standard MMD testing. Accordingly, even in the i.i.d. case, valid kernel tests rely on sample splitting and studentization of a cross-U statistic (Kim and Ramdas, 2024). Under adaptive data collection, there is an additional difficulty: because π_t depends on the past, the sequence $(\hat{\phi}_t)$ is no longer i.i.d., and its conditional covariance

$$\Sigma_t := \mathrm{Cov}(\hat{\phi}_t \mid \mathcal{F}_{t-1})$$

varies with time. Thus the usual i.i.d. calibration need not survive under adaptive logging, even when the propensities are known. A scalar normalization based

[2] We prefer the term *interventional mean embedding* because our estimand is a population-level intervention distribution $P(Y \mid do(A = a))$, not a counterfactual quantity in Pearl's stricter rung-3 sense.

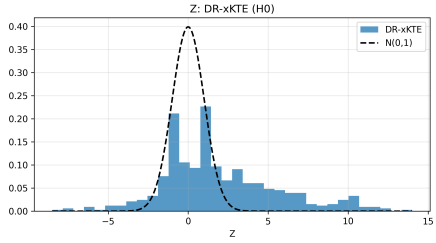

Figure 1: Histogram of the i.i.d.-style DR-xKTE statistic over 500 runs ($T = 700$, $d = 5$, $t_0 = 15$, $\varepsilon = 10^{-3}$) under adaptive logging with known propensities $\pi_t(1 \mid X_t)$. The resulting null distribution is visibly miscalibrated.

only on $\mathrm{Tr}(\Sigma_t)^{-1/2}$ controls the overall scale of the covariance operator but not its directional variation in $\mathcal{H}_\mathcal{Y}$. Since our goal is valid inference for a final scalar test statistic rather than a full Gaussian limit in $\mathcal{H}_\mathcal{Y}$ (see Appendix C), this motivates projecting onto a learned witness and applying a sequential, past-measurable variance normalization to the resulting scalar score process.

Example 3.3 (Contextual extension of (Bibaut and Kallus, 2025), Section 3.1.1)**.** *Let* $X_t \overset{i.i.d.}{\sim} \mathcal{N}(0, I_d)$*, and let*

$$Y_t(0) = f(X_t) + \varepsilon_t, \qquad Y_t(1) = f(X_t) + \Delta(X_t) + \varepsilon_t,$$

where $\Delta = Y_t(1) - Y_t(0)$ *is the treatment shift. The policy explores uniformly for* $t \le t_0$*, then commits to the empirically better arm with* ε*-randomization:*

$$\pi_t(1 \mid X_t) = \begin{cases} 0.5, & t \le t_0, \\ 1-\varepsilon, & t > t_0 \text{ and arm } 1 \text{ is selected}, \\ \varepsilon, & t > t_0 \text{ and arm } 0 \text{ is selected}. \end{cases}$$

Since the committed arm is history-dependent, the design is adaptive. Under $H_0 : \Delta \equiv 0$*, Bibaut and Kallus (2025) show that even the scalar ATE can have a non-Gaussian mixture limit. Figure 1 shows that the naive i.i.d.-style DR-xKTE statistic of Martinez Taboada et al. (2023) is similarly miscalibrated in this adaptive regime.*

Example 3.3 and Figure 1 illustrate the issue addressed in this paper: under adaptive treatment assignment, the usual i.i.d. studentization of a kernel test is no longer sufficient, and valid inference must account for the time-varying conditional variability of the score sequence using only past information.

4 ADAPTIVE TEST STATISTIC

We now introduce the adaptive test statistic studied in this paper. The construction has two stages. A first chronological fold is used to learn a witness in $\mathcal{H}_\mathcal{Y}$ for the direction of $\Psi(a, a')$. A second fold is then used for inference after projecting the doubly robust RKHS score onto that witness and normalizing the resulting scalar process by a predictable estimate of its conditional standard deviation.

4.1 Pilot witness and projected score

Throughout this section, write $T = 2n$ and use the chronological split

$$\mathcal{I}_1 := \{1, \dots, n\}, \qquad \mathcal{I}_2 := \{n+1, \dots, 2n\}.$$

The first fold is used to construct the witness, while inference is carried out on the second fold.

For $r \in \{1, 2\}$ and $t \in \mathcal{I}_r$, let $\widehat{\mu}^{(r)}_{t-1}$ be an $\mathcal{F}_{t-1}$-measurable estimator of $\mu_{Y|A,X}$, and define the foldwise doubly robust RKHS score

$$\begin{aligned} \hat{\phi}^{(r)}_t := {} & D'(\pi_t, \widehat{\mu}^{(r)}_{t-1}, a)(X_t, A_t, Y_t) \\ & - D'(\pi_t, \widehat{\mu}^{(r)}_{t-1}, a')(X_t, A_t, Y_t) \in \mathcal{H}_\mathcal{Y}. \end{aligned} \tag{6}$$

Throughout the analysis, we impose the standing envelope (e.g when learning a conditional mean embedding with fixed regularization parameter (Li et al., 2022))

$$\sup_{r\in\{1,2\}} \sup_{t\ge 1} \sup_{b\in\mathcal{A},\ x\in\mathcal{X}} \left\| \widehat{\mu}^{(r)}_{t-1}(b, x) \right\|_{\mathcal{H}_\mathcal{Y}} \le B_\mu \quad \text{a.s.} \tag{7}$$

The pilot witness is the first-fold average

$$v_n := \frac{1}{n} \sum_{t \in \mathcal{I}_1} \hat{\phi}^{(1)}_t \in \mathcal{H}_\mathcal{Y}. \tag{8}$$

To avoid the degenerate event $v_n = 0$, fix once and for all a unit vector $u_0 \in \mathcal{H}_\mathcal{Y}$, and define the tie-broken pilot direction

$$w_n := \begin{cases} v_n / \|v_n\|_{\mathcal{H}_\mathcal{Y}}, & v_n \neq 0, \\ u_0, & v_n = 0. \end{cases} \tag{9}$$

By construction, $\|w_n\|_{\mathcal{H}_\mathcal{Y}} = 1$ almost surely.

The second fold is then reduced to a scalar problem through the projected score

$$\xi_t(w_n) := \left\langle w_n, \hat{\phi}^{(2)}_t \right\rangle_{\mathcal{H}_\mathcal{Y}}, \qquad t \in \mathcal{I}_2. \tag{10}$$

The next lemma identifies the scalar target induced by the projection.

Lemma 4.1 (Projected doubly robust identity)**.** *For every* $t \in \mathcal{I}_2$*,*

$$\mathbb{E}[\xi_t(w_n) \mid \mathcal{F}_{t-1}] = \langle w_n, \Psi(a, a') \rangle_{\mathcal{H}_\mathcal{Y}} =: \theta_n(a, a'). \tag{11}$$

In particular, under $H_0 : \eta(a) = \eta(a')$*,*

$$\mathbb{E}[\xi_t(w_n) \mid \mathcal{F}_{t-1}] = 0.$$

4.2 Sequential standard deviation estimation

For $t \in \mathcal{I}_2$, define the oracle conditional standard deviation of the projected score by

$$\sigma_{0,t}(w_n) := \mathrm{Var}(\xi_t(w_n) \,|\, \mathcal{F}_{t-1})^{1/2}. \tag{12}$$

We estimate $\sigma_{0,t}(w_n)$ using only observations from the past of the inferential fold. The key point is that a past observation X_s, A_s, Y_s is re-evaluated under the current policy π_t but with its own historical nuisance estimate $\widehat{\mu}^{(2)}_{s-1}$, as noted in (Bibaut et al., 2021).

Fix $t \in \mathcal{I}_2$, and let

$$S_t := \{\, s \in \mathcal{I}_2 : s < t \,\}, \qquad m_t := |S_t|.$$

For every $s \in S_t$, define the history-recycled RKHS score

$$\begin{aligned}\check{\phi}^{(2)}_{s,t} := D'(\pi_t, \widehat{\mu}^{(2)}_{s-1}, a)(X_s, A_s, Y_s) \\ - D'(\pi_t, \widehat{\mu}^{(2)}_{s-1}, a')(X_s, A_s, Y_s) \in \mathcal{H}_\mathcal{Y}.\end{aligned} \tag{13}$$

Note that the sth observation is paired with the historical nuisance $\widehat{\mu}^{(2)}_{s-1}$ and not with the current nuisance $\widehat{\mu}^{(2)}_{t-1}$. We correct the discrepancy between the logging law at time s and the evaluation law at time t using the importance ratio

$$w_{s,t} := \frac{\pi_t(A_s \mid X_s)}{\pi_s(A_s \mid X_s)}. \tag{14}$$

The corresponding projected score is

$$\check{\xi}_{s,t}(w_n) := \left\langle w_n, \check{\phi}^{(2)}_{s,t} \right\rangle_{\mathcal{H}_\mathcal{Y}}. \tag{15}$$

Conditionally on $\mathcal{F}_{t-1}$, the pair (π_t, w_n) is fixed, and the oracle projected moments are

$$M_{1,t}(w_n) := \mathbb{E}[\xi_t(w_n) \,|\, \mathcal{F}_{t-1}] = \theta_n(a, a'), \tag{16}$$

$$M_{2,t}(w_n) := \mathbb{E}\left[\xi_t(w_n)^2 \,\middle|\, \mathcal{F}_{t-1}\right]. \tag{17}$$

By definition,

$$\sigma^2_{0,t}(w_n) = M_{2,t}(w_n) - M_{1,t}(w_n)^2. \tag{18}$$

For $t = n+1$, there is no past data in the inferential fold. We therefore set

$$\widehat{M}_{1,n+1}(w_n) := 0, \ \widehat{M}_{2,n+1}(w_n) := 1, \ \widehat{\sigma}^2_{n+1}(w_n) := 1.$$

This single convention is asymptotically negligible.

For $t \geq n+2$, we estimate the moments by the importance-weighted empirical averages

$$\widehat{M}_{1,t}(w_n) := \frac{1}{m_t} \sum_{s \in S_t} w_{s,t}\, \check{\xi}_{s,t}(w_n), \tag{19}$$

$$\widehat{M}_{2,t}(w_n) := \frac{1}{m_t} \sum_{s \in S_t} w_{s,t}\, \check{\xi}_{s,t}(w_n)^2. \tag{20}$$

The plug-in projected conditional variance estimator is then

$$\widehat{\sigma}^2_t(w_n) := \left(\widehat{M}_{2,t}(w_n) - \widehat{M}_{1,t}(w_n)^2\right)_+, \qquad t \geq n+2, \tag{21}$$

where $(x)_+ := \max(x, 0)$. All quantities above are $\mathcal{F}_{t-1}$-measurable, so the resulting statistic remains fully predictable on the inferential fold.

To avoid division by zero in finite samples, let $(\epsilon_n)_{n \geq 1}$ be a deterministic clip sequence with $\epsilon_n \downarrow 0$, and define the clipped feasible standard deviation

$$\widetilde{\sigma}_t(w_n) := \widehat{\sigma}_t(w_n) \vee \epsilon_n. \tag{22}$$

Finally, the adaptive doubly robust kernel treatment effect test statistic is

$$T_n(w_n) := \frac{1}{\sqrt{n}} \sum_{t \in \mathcal{I}_2} \widetilde{\sigma}_t(w_n)^{-1}\, \xi_t(w_n). \tag{ADR-KTE}$$

4.3 Null asymptotic normality

We now state the null asymptotic normality theorem. We now state mild assumptions on the adaptive data collection.

Assumption 4.2 (Exploration floor)**.** *There exist $c_\pi > 0$ and $\alpha \in [0, 1)$ such that*

$$\pi_t(b \mid x) \geq c_\pi\, t^{-\alpha} \tag{23}$$

for every $t \geq 1$, every $b \in \mathcal{A}$, and P_X-almost every $x \in \mathcal{X}$.

The assumption above is milder than assuming lower bounded propensities, and verified by many bandit algorithms (Lattimore and Szepesvári, 2020).

Assumption 4.3 (Asymptotically negligible clipping)**.** *For the deterministic clip sequence (ϵ_n),*

$$\frac{1}{n} \sum_{t \in \mathcal{I}_2} \mathbb{1}\{\sigma_{0,t}(w_n) \leq 2\epsilon_n\} \xrightarrow{p} 0. \tag{24}$$

Theorem 4.4 (Null asymptotic normality of the adaptive test)**.** *Assume Assumptions 3.2, 3.1, (7), 4.2, 4.3, and the assumptions of Theorem 5.4. Let $\nu(\alpha, \beta, p) > 0$ denote the rate constant from Theorem 5.4, and choose*

$$\epsilon_n = n^{-\rho} \ \text{ with } \ 0 < \rho < \min\left\{\frac{1-3\alpha}{4}, \frac{\nu(\alpha, \beta, p)}{2}\right\}.$$

Then the statistic $T_n(w_n)$ satisfies

$$T_n(w_n) \overset{d}{\Rightarrow} \mathcal{N}(0, 1) \qquad \text{under } H_0 : \eta(a) = \eta(a').$$

The proof is deferred to Appendix D. The next section is devoted to constructing $\widehat{\sigma}_t(w_n)$ and proving the required average consistency of the feasible predictable variances by reducing the RKHS problem to a projected scalar variance-estimation problem.

5 SEQUENTIAL CONDITIONAL STANDARD DEVIATION ESTIMATION

The statistic $T_n(w_n)$ depends on the predictable conditional standard deviation of the projected score $\xi_t(w_n)$. We now study the estimator introduced above. The key point is that after conditioning on the pilot fold, the projection direction w_n is fixed and the inferential-fold problem becomes scalar.

5.1 Sufficient conditions for projected variance consistency

We first require a rate of convergence of the inferential-fold nuisance sequence to a fixed limit.

Assumption 5.1 (Nuisance stabilization rate)**.** *There exist $\mu_\infty : \mathcal{A} \times \mathcal{X} \to \mathcal{H}_\mathcal{Y}$ and $\beta > 0$ such that*

$$\left\|\widehat{\mu}^{(2)}_{t-1} - \mu_\infty\right\|_{L_2(P_X \times \mu_\mathcal{A}; \mathcal{H}_\mathcal{Y})} = O(t^{-\beta}) \qquad a.s. \quad (25)$$

Note that above, we only require the nuisance estimate to converge to a fix limit which may be mispecified. This is allowed by our use of DR RKHS scores. Next, we assume the condition below which will quantify the complexity of the logging policies through their density ratios with respect a reference policy π_{ref}.

Assumption 5.2 (Logging-policy class regularity)**.** *There exist a reference policy π_{ref} and constants $G < \infty$ and $p > 0$ such that*

$$\sup_{g \in \mathcal{G}_{\text{ref}}} \|g\|_\infty \le G, \quad (26)$$

and

$$\log N_{[]}(\varepsilon, \mathcal{G}_{\text{ref}}, \|\cdot\|_{2,\text{ref}}) \lesssim \varepsilon^{-p}. \quad (27)$$

where we define $\mathcal{G}_{\text{ref}} := \left\{\frac{\pi}{\pi_{\text{ref}}} : \pi \in \Pi\right\}$, and equip this class with the norm

$$\|h\|^2_{2,\text{ref}} := \mathbb{E}_{X \sim P_X}\left[\int h(b,X)^2\, \pi_{\text{ref}}(b \mid X)\, \mathrm{d}\mu_\mathcal{A}(b)\right].$$

Finally, for consistency of the variance estimator we require the same strengthened exploration floor as in Bibaut et al. (2021).

Assumption 5.3 (Strengthened exploration floor)**.** *Assumption 4.2 holds with exponent*

$$\alpha < \alpha_\star(\beta, p) := \min\left\{\frac{1}{3+p}, \frac{1}{1+2p}, \beta\right\}. \quad (28)$$

Theorem 5.4 (Average consistency of the projected conditional standard deviation estimator)**.** *Assume 3.2, 3.1, (7), 5.1, 5.2, and 5.3. Then there exists a constant $\nu(\alpha, \beta, p) > 0$ such that*

$$\frac{1}{n}\sum_{t \in \mathcal{I}_2} \left|\widehat{M}_{1,t}(w_n) - M_{1,t}(w_n)\right| = O_{\text{a.s.}}\left(n^{-\nu(\alpha,\beta,p)}\right), \quad (29)$$

$$\frac{1}{n}\sum_{t \in \mathcal{I}_2} \left|\widehat{M}_{2,t}(w_n) - M_{2,t}(w_n)\right| = O_{\text{a.s.}}\left(n^{-\nu(\alpha,\beta,p)}\right), \quad (30)$$

and consequently

$$\frac{1}{n}\sum_{t \in \mathcal{I}_2} \left|\widehat{\sigma}^2_t(w_n) - \sigma^2_{0,t}(w_n)\right| = O_{\text{a.s.}}\left(n^{-\nu(\alpha,\beta,p)}\right). \quad (31)$$

Corollary 5.5 (Average inverse-standard-deviation consistency)**.** *Assume the conditions of Theorem 5.4 and Assumption 4.3. If*

$$\epsilon_n = n^{-\rho} \qquad \text{with} \qquad 0 < \rho < \frac{\nu(\alpha,\beta,p)}{2},$$

then

$$\frac{1}{n}\sum_{t \in \mathcal{I}_2} \left|\frac{\sigma^2_{0,t}(w_n)}{\widetilde{\sigma}^2_t(w_n)} - 1\right| \xrightarrow{p} 0. \quad (32)$$

In particular,

$$\frac{1}{n}\sum_{t \in \mathcal{I}_2} \frac{\sigma^2_{0,t}(w_n)}{\widetilde{\sigma}^2_t(w_n)} \xrightarrow{p} 1. \quad (33)$$

The proof is deferred to Appendix E. The proof is purely scalar after conditioning on the pilot fold and follows the a similar decomposition as that of the variance-estimation proof of Bibaut et al. (2021), now applied to the projected outcome.

6 PRACTICAL PROCEDURE AND POWER

Sections 4.3 and 5 separate the analysis into two parts: a null asymptotic normality result for the normalized statistic $T_n(w_n)$, and a projected variance-estimation theorem proving the predictable-variation consistency required by that CLT. We now combine these pieces and study the behavior of the test under fixed alternatives.

6.1 Practical adaptive KTE test

The practical procedure of the (ADR-KTE) test is summarized in Algorithm 1 (more details in Appendix G). It uses a chronological pilot fold to construct an RKHS witness and an inferential fold to compute the normalized projected score.

Corollary 6.1 (Asymptotic validity of the adaptive KTE test)**.** *Under the assumptions of Theorem 4.4,*

$$T_n(w_n) \overset{d}{\Rightarrow} \mathcal{N}(0,1) \qquad \text{under } H_0 : \eta(a) = \eta(a').$$

Algorithm 1 Adaptive KTE test (ADR-KTE)

Input: adaptive data $\mathcal{D}_T$, policies $\{\pi_t\}_{t=1}^T$, target actions (a, a'), significance level γ
Split $\{1, \dots, T\}$ chronologically into $\mathcal{I}_1 = \{1, \dots, n\}$ and $\mathcal{I}_2 = \{n+1, \dots, 2n\}$
for each $t \in \mathcal{I}_1$ **do**
  construct a predictable nuisance estimate $\widehat{\mu}_{t-1}^{(1)}$
  compute the pilot RKHS score $\hat{\phi}_t^{(1)}$ from (6)
Form the pilot witness $v_n = \frac{1}{n} \sum_{t \in \mathcal{I}_1} \hat{\phi}_t^{(1)}$
Set the tie-broken pilot direction

$$w_n = \begin{cases} v_n / \|v_n\|_{\mathcal{H}_\mathcal{Y}}, & v_n \neq 0, \\ u_0, & v_n = 0, \end{cases}$$

where $u_0 \in \mathcal{H}_\mathcal{Y}$ is fixed with $\|u_0\|_{\mathcal{H}_\mathcal{Y}} = 1$
for each $t \in \mathcal{I}_2$ **do**
  construct a predictable nuisance estimate $\widehat{\mu}_{t-1}^{(2)}$
  compute $\xi_t(w_n) = \langle w_n, \hat{\phi}_t^{(2)} \rangle_{\mathcal{H}_\mathcal{Y}}$
  if $t = n+1$ **then**
    set $\widehat{\sigma}_t(w_n) = 1$
  else
    let $S_t = \{s \in \mathcal{I}_2 : s < t\}$
    for each $s \in S_t$ **do**
      compute $w_{s,t} = \pi_t(A_s \mid X_s) / \pi_s(A_s \mid X_s)$
      compute $\check{\xi}_{s,t}(w_n) = \langle w_n, \check{\phi}_{s,t}^{(2)} \rangle_{\mathcal{H}_\mathcal{Y}}$
    compute $\widehat{M}_{1,t}(w_n)$ and $\widehat{M}_{2,t}(w_n)$ from (19)–(20)
    set $\widehat{\sigma}_t^2(w_n) = \big(\widehat{M}_{2,t}(w_n) - \widehat{M}_{1,t}(w_n)^2\big)_+$
    set $\widehat{\sigma}_t(w_n) = \sqrt{\widehat{\sigma}_t^2(w_n)}$
  set $\widetilde{\sigma}_t(w_n) = \widehat{\sigma}_t(w_n) \vee \epsilon_n$
Set $T_n(w_n) = \frac{1}{\sqrt{n}} \sum_{t \in \mathcal{I}_2} \widetilde{\sigma}_t(w_n)^{-1} \xi_t(w_n)$
Reject H_0 if $T_n(w_n) > z_{1-\gamma}$

Consequently, the rejection rule

$$\phi_n = \mathbb{1}\{T_n(w_n) > z_{1-\gamma}\}$$

has asymptotic level γ.

Corollary 6.1 gives the final null calibration guarantee. The statistic is one-dimensional, but its signal is inherited from the RKHS-valued effect $\Psi(a, a')$ through the pilot witness v_n.

6.2 Witness consistency

We next show that the pilot witness consistently estimates $\Psi(a, a')$. This result is separate from the variance-estimation analysis and requires only the boundedness of the kernel, the nuisance envelope (7), and the exploration floor.

Theorem 6.2 (Pilot witness consistency)**.** *Under Assumptions 3.2, 3.1, 4.2, and the nuisance envelope (7), the pilot witness (8) satisfies*

$$\mathbb{E}\big[\|v_n - \Psi(a, a')\|_{\mathcal{H}_\mathcal{Y}}^2\big] \lesssim n^{\alpha - 1}. \tag{34}$$

In particular, if $\alpha < 1$, *then*

$$v_n \xrightarrow{L_2(\mathcal{H}_\mathcal{Y})} \Psi(a, a'), \qquad v_n \xrightarrow{p} \Psi(a, a').$$

The proof is deferred to Appendix F.

Corollary 6.3 (Consistency of the pilot direction)**.** *Assume the conditions of Theorem 6.2 and let* $\Psi(a, a') \neq 0$. *Then*

$$w_n \xrightarrow{p} \frac{\Psi(a, a')}{\|\Psi(a, a')\|_{\mathcal{H}_\mathcal{Y}}}.$$

The proof is deferred to Appendix F.

Theorem 6.2 is the basic reason the projected statistic remains meaningful under fixed alternatives. The witness is learned from a raw RKHS average. Stabilization is only needed on the inferential fold, where the goal is valid asymptotic normality of the final normalized statistic.

6.3 Consistency under fixed alternatives

We now turn to the behavior of the test under H_1.

Corollary 6.4 (Consistency under fixed alternatives)**.** *Assume the conditions of Corollary 6.1 and let* $\Psi(a, a') \neq 0$. *Then*

$$T_n(w_n) \xrightarrow{p} +\infty,$$

and therefore the rejection rule in Corollary 6.1 is consistent.

Remark 6.5 (Deterministic limit direction under H_1). When $\Psi(a, a') \neq 0$, write

$$w_\star := \frac{\Psi(a, a')}{\|\Psi(a, a')\|_{\mathcal{H}_\mathcal{Y}}}.$$

The power proof uses only that $w_n \xrightarrow{p} w_\star$. No uniform lower variance bound over random pilot directions is imposed.

The proof is deferred to Appendix F. Corollary 6.4 shows that the test remains sensitive to fixed distributional alternatives. The chronology of the method is therefore simple: the first fold learns the witness, the second fold estimates the conditional standard deviation of the projected score using the projected CADR construction, and the resulting normalized statistic is asymptotically valid under the null and consistent under fixed alternatives.

7 NUMERICAL SIMULATIONS

In this section, we study the empirical calibration and power of our proposed test ADR-KTE under *adaptive* data collection. We observe a stream $\{(X_t, A_t, Y_t)\}_{t=1}^T$ generated by a bandit-style logging policy $\pi_t(\cdot \mid X_t)$. We evaluate both calibration (Scenario I) and power (Scenarios II–IV) at a significance level of $\alpha = 0.05$. Additional details and results appear in Appendix H. Our implementation is available on GitHub.[3]

[3] https://github.com/houssamzenati/adaptive-KTE

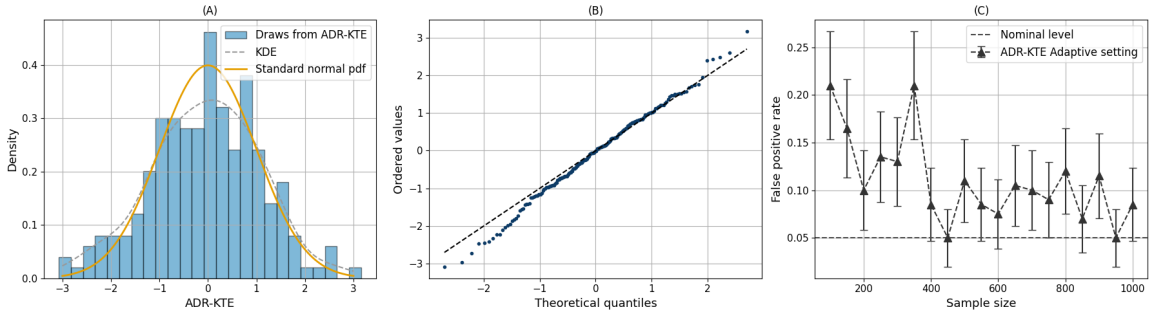

Figure 2: Illustration of 200 simulations of ADR-KTE under the null in the adaptive setting with $T = 950$: (A) Histogram with KDE and standard normal pdf, (B) Normal Q-Q plot, (C) False positives against sample sizes. The results show approximate Gaussian behaviour and controlled type-I error.

Adaptive data collection. Actions follow an ε-greedy contextual bandit with per–arm online ridge. At time t, with features $Z_t = [1, X_t]$, each arm a has $S_a = \lambda I + \sum_{s \le t: A_s = a} Z_s Z_s^\top$, $b_a = \sum_{s \le t: A_s = a} Z_s Y_s$, $\hat{\theta}_a = S_a^\dagger b_a$, and score $q_a(t) = Z_t^\top \hat{\theta}_a$. The propensity is

$$\pi_t(1 \mid X_t) = \begin{cases} 1 - \frac{1}{2}\varepsilon_t, & q_1(t) > q_0(t), \\ \frac{1}{2}\varepsilon_t, & q_1(t) < q_0(t), \\ \frac{1}{2}, & \text{otherwise}, \end{cases}$$

with $\varepsilon_0 \in (0,1)$, $\varepsilon_{\min} > 0$, $\beta \in (0,1]$. We sample $A_t \sim \pi_t$, observe Y_t, and store $\pi_t(A_t \mid X_t)$. For sample-splitting we use non-overlapping time folds: by default an *alternating* split ($\mathcal{I}_0 = \{t \text{ odd}\}$, $\mathcal{I}_1 = \{t \text{ even}\}$). Each fold is evaluated in temporal order so all nuisance weights remain predictable.

Baselines. We compare to two adaptive inference methods: **CADR** (Bibaut et al., 2021), stabilizes the DR score with history-measurable weights that estimate its conditional variance from past data, yielding a martingale CLT, and **AW-AIPW** (Hadad et al., 2021), enforces deterministic quadratic variation in adaptive experiments by reweighting AIPW scores with variance-stabilizing allocations, guaranteeing asymptotic normality. *Both are scalar, targeting mean effects* (i.e., contrasts of $\mathbb{E}[Y(a)]$), whereas our ADR-KTE directly targets *the full outcome distribution* via RKHS mean embeddings. We use the authors' open-source implementations and fit the regression nuisances with kernel ridge regressions; details are in Appendix H.

7.1 Synthetic data.

We adapt the synthetic designs of Martinez Taboada et al. (2023) to the adaptive setting by replacing i.i.d. assignment with an ϵ-greedy policy $\{\pi_t\}$ as described above. Each replicate simulates covariates $X \in \mathbb{R}^5$, draws T rounds under π_t. The potential outcomes are defined as $Y_t(A_t) = \cos(\beta^\top X_t) + \Delta(s)\mathbf{1}(A_t = 1) + \epsilon_t$, with $\beta = (0.1, 0.2, 0.3, 0.4, 0.5)^\top$, independent noises $\epsilon_t \sim \mathcal{N}(0, 0.5)$, and the shift random variable $\Delta(s)$ varied to match each scenario s. Four scenarios are considered for $\Delta(s)$: (I) no effect; (II) mean shift only; (III–IV) higher-moment changes at equal means. Additional details and other forms of potential outcome function $Y_t(A_t)$ experimented are given in Appendix H.

In Scenario I, ADR-KTE is well calibrated (see the empirical histogram, QQ-plot and false positive rate in Figure 2). Across Scenarios II–IV (Figure 3), it attains high power for both mean and higher-moment shifts. By contrast, ATE-focused baselines (CADR, AW-AIPW) match only under mean shifts (II) and fail under purely distributional changes (III–IV).

7.2 IHDP dataset

We evaluate our method on the Infant Health and Development Program (IHDP) data (Hill, 2011), following the same design as in (Martinez Taboada et al., 2023): after removing missing rows we retain 908 units with 18 covariates (9 continuous, 9 categorical). In our experiment, treatments are assigned adaptively via the ϵ-greedy policy described earlier. The outcome construction mirrors the simulation design of previous Scenarios (I)–(IV), where potential outcomes are similarly defined as $Y_t(A_t) = \cos(\beta^\top X_t) + \Delta(s)\mathbf{1}(A_t = 1) + \epsilon_t$, with $\beta = (1, \ldots, 1)^\top$, independent Gaussian noises $\epsilon_t \sim \mathcal{N}(0, 0.5)$, and the shift random variable $\Delta(s)$ varied to match each scenario s (zero under the null, mean shift in II, equal-mean distributional changes in III–IV). Full implementation details are provided in Appendix H.

Table 1 reports true positive rates (mean $\pm$ standard error). ADR-KTE achieves near-perfect power across Scenarios II–IV, illustrating the benefits of our distributional kernel test under adaptivity. Conversely, CADR and AW-AIPW succeed only on the mean shift (II), largely failing (rejection rates $\approx \alpha$) under equal-mean distributional shifts (III–IV).

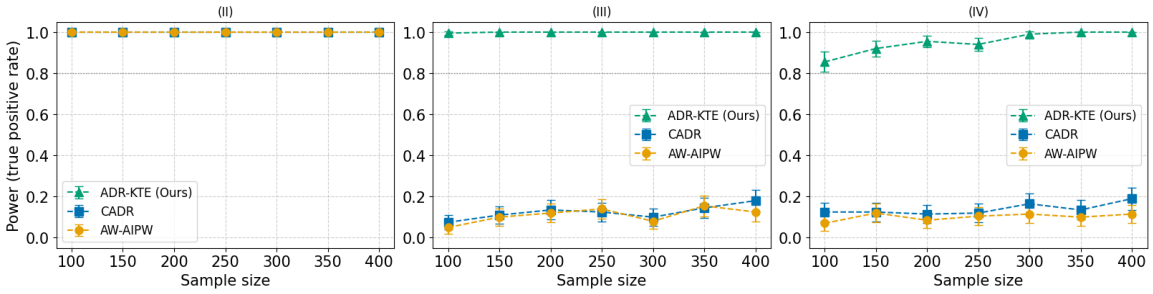

Figure 3: True positive rates (200 simulations, Scenarios II–IV). Mean-focused baselines (CADR/AW-AIPW) achieve matching performance on II; ADR-KTE shows markedly higher power on III–IV (higher-moment shifts).

Table 1: True positive rates (mean ± se) for IHDP on 200 simulations and a sample size $T = 908$.

	II	III	IV
ADR-KTE	$\mathbf{1.0 \pm 0.0}$	$\mathbf{1.0 \pm 0.0}$	$\mathbf{1.0 \pm 0.0}$
CADR	1.0 ± 0.0	0.075 ± 0.04	0.085 ± 0.04
AW-AIPW	1.0 ± 0.0	0.07 ± 0.04	0.045 ± 0.03

7.3 dSprite dataset

We evaluate our kernel test on the *dSprites* dataset (Matthey et al., 2017) with structured outcomes $Y \in \mathbb{R}^{64\times64}$. Contexts $X \sim \mathrm{Unif}([0,1]^2)$ are mapped to images Y by a deterministic renderer $g(X, A)$ that places a white heart shape in a black image based on X, A. We study two regimes: Scenario I (null), where both treatments induce the same image distribution, and Scenario IV (shift), where $A = 1$ translates the heart shape relative to $A = 0$ (a spatial change with unchanged mean intensity). Logged data are collected by an *adaptive* ε-greedy policy with per-arm online ridge. Our test, ADR-KTE, operates directly on flattened images. By contrast, baselines (CADR and AW-AIPW) require scalar outcomes, forcing us to use the mean pixel per image, which inherently cannot detect the spatial shift in Scenario IV.

Table 2: True positive rates (mean ± se) for dSprite on 200 simulations and a sample size of $T = 1000$.

	I	IV
ADR-KTE	$\mathbf{0.04 \pm 0.03}$	$\mathbf{1.00 \pm 0.00}$
CADR	0.19 ± 0.05	0.19 ± 0.05
AW-AIPW	0.10 ± 0.04	0.10 ± 0.04

ADR-KTE shows near-nominal type-I error in Scenario I and perfect power in Scenario IV, detecting the spatial shift in the full image distribution. In contrast, CADR and AW-AIPW (fed only the mean pixel) exhibit non-trivial false positives under the null and no power in the shift scenario, underscoring the value of testing for structured outcomes.

7.4 Comparison with non-adaptive baselines

To complement the comparison with adaptive baselines, we also evaluate against kernel procedures designed for i.i.d. data, namely DR-xKTE (Martinez Taboada et al., 2023) and KTE (Muandet et al., 2021). Appendix H.5 reports results under both i.i.d. and adaptive designs. Under i.i.d. sampling, all three methods are correctly calibrated and attain full power. Under an ε-greedy adaptive design, which maintains persistent exploration and falls in a stable adaptive regime (Lai and Wei, 1982), the picture changes: DR-xKTE becomes anti-conservative under the null, KTE remains roughly calibrated but suffers a substantial loss of power, and ADR-KTE achieves the strongest calibration-power tradeoff, with null rejection rates moving toward the nominal level as T increases and power close to one. These results show that even under a stable adaptive design, non-adaptive kernel tests are not reliable without explicit variance control.

8 DISCUSSION

We introduced ADR-KTE, a kernel-based test for distributional treatment effects under adaptive data collection. Our method uses a learn-then-test construction: a first chronological fold estimates a witness direction in the outcome RKHS, and inference is then performed on the associated projected score. The resulting scalar statistic remains sensitive to general distributional differences, including shifts beyond the mean. Recent work (Shen et al., 2026) shows that, under a stability condition (Lai and Wei, 1982), estimators that are asymptotically normal in the i.i.d. setting may remain so under adaptive designs without additional stabilization. A key open question is whether a similar principle holds for kernel-based distributional estimators beyond the projected setting considered here. Other promising directions include conditional effects and richer embedding regressors.

Acknowledgements

Houssam Zenati, Bariscan Bozkurt, and Arthur Gretton are supported by the Gatsby Charitable Foundation. The authors thank Zikai Shen for relevant remarks on an earlier version of this work. The authors also thank the AISTATS 2026 reviewers and the Area Chair for their constructive feedback and helpful discussions during the review process.

References

Athey, S., Eckles, D., and Imbens, G. W. (2022). Design and analysis of experiments in the digital age. *Annual Review of Economics*, 14:779–806.

Berlinet, A. and Thomas-Agnan, C. (2011). *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media.

Bibaut, A., Dimakopoulou, M., Kallus, N., Chambaz, A., and van Der Laan, M. (2021). Post-contextual-bandit inference. *Advances in neural information processing systems*, 34:28548–28559.

Bibaut, A. and Kallus, N. (2025). Demystifying inference after adaptive experiments. *Annual Review of Statistics and its Application*, 12(1):407–423.

Caria, S., Gordon, B., Kasy, M., et al. (2023). Adaptive experiments in economics. *Annual Review of Economics*, 15:615–647.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268.

Chow, S.-C. and Chang, M. (2011). *Adaptive Design Methods in Clinical Trials*. Chapman & Hall/CRC, 2nd edition.

Dabney, W., Ostrovski, G., Silver, D., and Munos, R. (2018). Implicit quantile networks for distributional reinforcement learning. In *International conference on machine learning*, pages 1096–1105. PMLR.

Fawkes, J., Hu, R., Evans, R. J., and Sejdinovic, D. (2024). Doubly robust kernel statistics for testing distributional treatment effects. *Transactions on Machine Learning Research*.

Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Conference on Learning Theory (COLT)*, pages 998–1027.

Gärtner, T. (2003). A survey of kernels for structured data. *ACM SIGKDD explorations newsletter*, 5(1):49–58.

Gretton, A. (2013). Introduction to rkhs, and some simple kernel algorithms. *Adv. Top. Mach. Learn. Lecture Conducted from University College London*, 16(5-3):2.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012a). A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012b). A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773.

Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S., and Athey, S. (2021). Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the national academy of sciences*, 118(15):e2014602118.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic press.

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

Hirano, K. and Porter, J. R. (2023). Asymptotic representations for sequential decisions, adaptive experiments, and batched bandits. *arXiv preprint arXiv:2302.03117*.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. (2021). Time-uniform chernoff bounds via nonnegative supermartingales. *Annals of Statistics*, 49(2):1055–1080.

Huang, A., Leqi, L., Lipton, Z., and Azizzadenesheli, K. (2021). Off-policy risk assessment in contextual bandits. In *Advances in Neural Information Processing Systems*, volume 34, pages 23714–23726.

Kanagawa, M. and Fukumizu, K. (2014). Recovering Distributions from Gaussian RKHS Embeddings. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, volume 33.

Kim, I. and Ramdas, A. (2024). Dimension-agnostic inference using cross u-statistics. *Bernoulli*, 30(1):683–711.

Lai, T. L. and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics*, 10(1):154–166.

Lattimore, T. and Szepesvári, C. (2020). *Bandit Algorithms*. Cambridge University Press.

Lattimore, T. and Szepesvári, C. (2020). *Bandit Algorithms*. Cambridge University Press.

Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web (WWW)*, pages 661–670.

Li, Z., Meunier, D., Mollenhauer, M., and Gretton, A. (2022). Optimal rates for regularized conditional mean embedding learning. *Advances in Neural Information Processing Systems*, 35:4433–4445.

Luedtke, A. and Chung, I. (2024). One-step estimation of differentiable Hilbert-valued parameters. *The Annals of Statistics*, 52(4):1534 – 1563.

Martinez Taboada, D., Ramdas, A., and Kennedy, E. (2023). An efficient doubly-robust test for the kernel treatment effect. In *Advances in Neural Information Processing Systems*, volume 36, pages 59924–59952.

Matthey, L., Higgins, I., Hassabis, D., and Lerchner, A. (2017). dsprites: Disentanglement testing sprites dataset. https://github.com/deepmind/dsprites-dataset/.

Muandet, K., Kanagawa, M., Saengkyongam, S., and Marukatat, S. (2021). Counterfactual mean embeddings. *Journal of Machine Learning Research*, 22(162):1–71.

Park, J., Shalit, U., Schölkopf, B., and Muandet, K. (2021). Conditional distributional treatment effect with kernel conditional mean embeddings and u-statistic regression. In *International conference on machine learning*, pages 8401–8412.

Perchet, V., Rigollet, P., Chassang, S., and Snowberg, E. (2016). Batched bandit problems. *The Annals of Statistics*, 44:660–681.

Qiang, S. and Bayati, M. (2016). Dynamic pricing with demand learning and strategic consumers: An application to online retail. *Operations Research*, 64(4):931–944.

Rockafellar, R. T., Uryasev, S., et al. (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42.

Rothe, C. (2010). Nonparametric estimation of distributional policy effects. *Journal of Econometrics*, 155(1):56–70.

Shekhar, S., Kim, I., and Ramdas, A. (2023). A permutation-free kernel independence test. *Journal of Machine Learning Research*, 24(369):1–68.

Shen, Z., Zenati, H., Kallus, N., Gretton, A., Khamaru, K., and Bibaut, A. (2026). Efficient inference after directionally stable adaptive experiments.

Singh, R., Xu, L., and Gretton, A. (2024). Kernel methods for causal functions: dose, heterogeneous and incremental response curves. *Biometrika*, 111(2):497–516.

Smola, A., Gretton, A., Song, L., and Schölkopf, B. (2007). A hilbert space embedding for distributions. In *International conference on algorithmic learning theory*, pages 13–31. Springer.

Song, L., Huang, J., Smola, A., and Fukumizu, K. (2009). Hilbert space embeddings of conditional distributions with applications to dynamical systems. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 961–968.

Sriperumbudur, B., Gretton, A., Fukumizu, K., Schölkopf, B., and Lanckriet, G. (2010). Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11:1517–1561.

Vaart, A. v. d. and Wellner, J. A. (1997). Weak convergence and empirical processes with applications to statistics. *Journal of the Royal Statistical Society-Series A Statistics in Society*, 160(3):596–608.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Waudby-Smith, I. and Ramdas, A. (2021). Time-uniform central limit theorems and confidence sequences. In *International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 10663–10672.

Xu, L. and Gretton, A. (2023). Causal benchmark based on disentangled image dataset.

Zenati, H., Bietti, A., Diemert, E., Mairal, J., Martin, M., and Gaillard, P. (2022). Efficient kernelized ucb for contextual bandits. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*.

Zenati, H., Bozkurt, B., and Gretton, A. (2025). Doubly-robust estimation of counterfactual policy mean embeddings. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Zenati, H., Diemert, E., Martin, M., Mairal, J., and Gaillard, P. (2023). Sequential counterfactual risk minimization. In *International Conference on Machine Learning*, pages 40681–40706. PMLR.

Zhang, K., Janson, L., and Murphy, S. (2020). Inference for batched bandits. *Advances in neural information processing systems*, 33:9818–9829.

Zhang, K., Janson, L., and Murphy, S. (2021). Statistical inference with m-estimators on adaptively collected data. *Advances in neural information processing systems*, 34:7460–7471.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model.

[Yes] Sections 3–6 define the adaptive data-collection setting, the kernel treatment effect estimand, the projected adaptive test statistic ADR-KTE, and the assumptions used in the analysis. Algorithm 1 summarizes the practical procedure.

(b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Partially] The paper provides a statistical analysis of the procedure, including null asymptotic normality, variance-estimation consistency, witness consistency, and consistency under fixed alternatives. It does not provide a dedicated computational complexity analysis in the main text, although implementation details are given in Appendix G.

(c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes] Implementation code is provided via the GitHub link in Section 7, and Appendix H gives additional implementation details.

2. For any theoretical claim, check if you include:

(a) Statements of the full set of assumptions of all theoretical results. [Yes] See Assumptions 3.2, 3.1, 4.2, 4.3, 5.1, 5.2, and 5.3.

(b) Complete proofs of all theoretical results. [Yes] Proofs are deferred to the appendix, including Appendix D, Appendix E, and Appendix F.

(c) Clear explanations of any assumptions. [Yes] The paper provides discussion around the main assumptions, including the role of exploration, the misspecified nuisance limit, the policy-class regularity condition, and the RKHS-specific discussion in Appendix C.

3. For all figures and tables that present empirical results, check if you include:

(a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes] A code repository is linked in Section 7, and Appendix H provides additional implementation details.

(b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes] Experimental setups, adaptive policies, fold splits, kernels, regularization choices, and data-generation details are described in Section 7 and Appendix H.

(c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes] The paper reports rejection rates / true positive rates with standard errors in tables and describes Monte Carlo repetition counts in the experimental sections and appendix. Calibration plots and power plots are also provided.

(d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes] See the computation-infrastructure paragraph in Appendix H.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

(a) Citations of the creator if your work uses existing assets. [Yes] The paper cites the IHDP dataset (Hill, 2011), dSprites (Matthey et al., 2017), and the baseline methods used for comparison.

(b) The license information of the assets, if applicable. [No] The current draft cites the datasets and methods but does not explicitly list asset-license information.

(c) New assets either in the supplemental material or as a URL, if applicable. [Yes] The implementation is made available through the GitHub repository linked in Section 7.

(d) Information about consent from data providers/curators. [Not Applicable]

(e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable] The experiments use standard research benchmark datasets and simulated data, and do not involve sensitive content.

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

(a) The full text of instructions given to participants and screenshots. [Not Applicable]

(b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Kernel Treatment Effects with Adaptively Collected Data: Appendix

This appendix is organized as follows:

- Appendix A: summary of the notations used in the paper and in the analysis.
- Appendix B: background on reproducing kernel Hilbert spaces and martingale tools used in the proofs.
- Appendix C: discussion of direct RKHS stabilization and its limitations under adaptive data collection.
- Appendix D: proofs for the adaptive projected test statistic introduced in Section 4.
- Appendix E: proof of consistency of the projected conditional standard deviation estimator from Section 5.
- Appendix F: proofs for asymptotic validity, pilot witness consistency, and consistency under fixed alternatives from Section 6.
- Appendix G: implementation details.
- Appendix H: additional experimental results.

A NOTATIONS

We collect here the notation used in the paper.

Adaptive data and filtration

- $t \in \{1, \dots, T\}$: round index.
- $X_t \in \mathcal{X}$, $A_t \in \mathcal{A}$, $Y_t \in \mathcal{Y}$: context, action, and outcome at round t.
- $\mathcal{F}_t := \sigma(X_1, A_1, Y_1, \dots, X_t, A_t, Y_t)$: observed-data filtration.
- $X_t \sim P_X$ i.i.d., and $Y_t \sim P_{Y|X,A}(\cdot \mid X_t, A_t)$.
- $\pi_t(\cdot \mid X_t)$: adaptive logging policy at time t, possibly $\mathcal{F}_{t-1}$-measurable.
- $\mu_{\mathcal{A}}$: base measure on the discrete action space $\mathcal{A}$.
- $\mathcal{D}_T := \{(X_t, A_t, Y_t)\}_{t=1}^T$: observed trajectory.

Kernel and embedding notation

- $k_{\mathcal{Y}}$: bounded characteristic kernel on $\mathcal{Y}$ with RKHS $\mathcal{H}_{\mathcal{Y}}$.
- $\phi_{\mathcal{Y}}(y) := k_{\mathcal{Y}}(\cdot, y)$: outcome feature map.
- $\mu_{Y|A,X}(a, x) := \mathbb{E}[\phi_{\mathcal{Y}}(Y) \mid A = a, X = x] \in \mathcal{H}_{\mathcal{Y}}$: conditional mean embedding.
- $\eta(a) := \mathbb{E}_{P_X}[\mu_{Y|A,X}(a, X)] \in \mathcal{H}_{\mathcal{Y}}$: interventional mean embedding.
- $\Psi(a, a') := \eta(a) - \eta(a') \in \mathcal{H}_{\mathcal{Y}}$: embedding difference.
- $\tau(a, a') := \|\Psi(a, a')\|_{\mathcal{H}_{\mathcal{Y}}}$: kernel treatment effect.

Doubly robust score

- For a policy π and nuisance $\bar{\mu} : \mathcal{A} \times \mathcal{X} \to \mathcal{H}_{\mathcal{Y}}$,

$$D'(\pi, \bar{\mu}, a)(X, A, Y) := \frac{\mathbb{1}\{A = a\}}{\pi(a \mid X)} \big(\phi_{\mathcal{Y}}(Y) - \bar{\mu}(A, X)\big) + \bar{\mu}(a, X).$$

- Foldwise score difference:

$$\hat{\phi}_t^{(r)} := D'(\pi_t, \hat{\mu}_{t-1}^{(r)}, a)(X_t, A_t, Y_t) - D'(\pi_t, \hat{\mu}_{t-1}^{(r)}, a')(X_t, A_t, Y_t) \in \mathcal{H}_{\mathcal{Y}}.$$

Sample split and projected statistic

- We write $T = 2n$ and use the chronological split

$$\mathcal{I}_1 = \{1, \dots, n\}, \qquad \mathcal{I}_2 = \{n + 1, \dots, 2n\}.$$

- Raw pilot witness:

$$v_n := \frac{1}{n} \sum_{t \in \mathcal{I}_1} \hat{\phi}_t^{(1)} \in \mathcal{H}_{\mathcal{Y}}.$$

– Tie-broken pilot direction:

$$w_n := \begin{cases} v_n/\|v_n\|_{\mathcal{H}_\mathcal{Y}}, & v_n \neq 0, \\ u_0, & v_n = 0, \end{cases} \qquad \|u_0\|_{\mathcal{H}_\mathcal{Y}} = 1.$$

– Projected inferential-fold score:

$$\xi_t(w_n) := \langle w_n, \hat{\phi}_t^{(2)} \rangle_{\mathcal{H}_\mathcal{Y}}, \qquad t \in \mathcal{I}_2.$$

– Projected target:

$$\theta_n(a, a') := \langle w_n, \Psi(a, a') \rangle_{\mathcal{H}_\mathcal{Y}}.$$

Sequential conditional variance estimation

– For $t \in \mathcal{I}_2$, the oracle projected conditional standard deviation is

$$\sigma_{0,t}(w_n) := \mathrm{Var}(\xi_t(w_n) \mid \mathcal{F}_{t-1})^{1/2}.$$

– Past index set in the inferential fold:

$$S_t := \{s \in \mathcal{I}_2 : s < t\}, \qquad m_t := |S_t|.$$

– History-recycled RKHS score:

$$\check{\phi}_{s,t}^{(2)} := D'(\pi_t, \hat{\mu}_{s-1}^{(2)}, a)(X_s, A_s, Y_s) - D'(\pi_t, \hat{\mu}_{s-1}^{(2)}, a')(X_s, A_s, Y_s).$$

– Importance ratio:

$$w_{s,t} := \frac{\pi_t(A_s \mid X_s)}{\pi_s(A_s \mid X_s)}.$$

– History-recycled projected score:

$$\check{\xi}_{s,t}(w_n) := \langle w_n, \check{\phi}_{s,t}^{(2)} \rangle_{\mathcal{H}_\mathcal{Y}}.$$

– Plug-in projected moments for $t \geq n + 2$:

$$\hat{M}_{1,t}(w_n) := \frac{1}{m_t} \sum_{s \in S_t} w_{s,t} \check{\xi}_{s,t}(w_n),$$

$$\hat{M}_{2,t}(w_n) := \frac{1}{m_t} \sum_{s \in S_t} w_{s,t} \check{\xi}_{s,t}(w_n)^2.$$

– Initial convention at the first inferential time:

$$\hat{M}_{1,n+1}(w_n) := 0, \qquad \hat{M}_{2,n+1}(w_n) := 1, \qquad \hat{\sigma}_{n+1}^2(w_n) := 1.$$

– Plug-in conditional variance and standard deviation for $t \geq n + 2$:

$$\hat{\sigma}_t^2(w_n) := \big(\hat{M}_{2,t}(w_n) - \hat{M}_{1,t}(w_n)^2\big)_+, \qquad \hat{\sigma}_t(w_n) := \sqrt{\hat{\sigma}_t^2(w_n)}.$$

– Clipped feasible standard deviation:

$$\tilde{\sigma}_t(w_n) := \hat{\sigma}_t(w_n) \vee \epsilon_n.$$

Final adaptive test statistic

– Final adaptive test statistic:

$$T_n(w_n) := \frac{1}{\sqrt{n}} \sum_{t \in \mathcal{I}_2} \tilde{\sigma}_t(w_n)^{-1} \xi_t(w_n).$$

Reference policy class used in Appendix E

– π_{ref}: fixed strictly positive reference density.
– $\mathcal{G}_{\mathrm{ref}} := \{\pi/\pi_{\mathrm{ref}} : \pi \in \Pi\}$.
–

$$\|h\|_{2,\mathrm{ref}}^2 := \mathbb{E}_{X \sim P_X}\left[\int h(b, X)^2 \, \pi_{\mathrm{ref}}(b \mid X) \, \mathrm{d}\mu_{\mathcal{A}}(b)\right].$$

B BACKGROUND

This appendix collects the background material used in the current proof stack.

B.1 Review of Reproducing Kernel Hilbert Spaces

A *positive definite kernel* on a set $\mathcal{F}$ is a function $k : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ such that for any $m \in \mathbb{N}$, any $w_1, \ldots, w_m \in \mathcal{F}$, and any $c_1, \ldots, c_m \in \mathbb{R}$,

$$\sum_{i,j=1}^{m} c_i c_j \, k(w_i, w_j) \geq 0.$$

By the Moore–Aronszajn theorem, k induces a unique Hilbert space $\mathcal{H}_\mathcal{F}$ of functions on $\mathcal{F}$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_\mathcal{F}}$ such that $k(\cdot, w) \in \mathcal{H}_\mathcal{F}$ for every $w \in \mathcal{F}$, and the reproducing property holds:

$$f(w) = \langle f, k(\cdot, w) \rangle_{\mathcal{H}_\mathcal{F}} \qquad \forall f \in \mathcal{H}_\mathcal{F}, \ \forall w \in \mathcal{F}.$$

We write the associated feature map as

$$\phi_\mathcal{F}(w) := k(\cdot, w) \in \mathcal{H}_\mathcal{F}.$$

Kernel mean embeddings. Let $W \sim P$ be a random element of $\mathcal{F}$ such that $\mathbb{E}[\sqrt{k(W, W)}] < \infty$. The *kernel mean embedding* of P into $\mathcal{H}_\mathcal{F}$ is

$$\mu_P := \mathbb{E}[\phi_\mathcal{F}(W)] \in \mathcal{H}_\mathcal{F}.$$

Given observations $w_1, \ldots, w_n$, the empirical embedding is

$$\widehat{\mu}_P = \frac{1}{n} \sum_{i=1}^{n} \phi_\mathcal{F}(w_i).$$

Conditional mean embeddings. Let $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ be random variables with RKHSs $\mathcal{H}_\mathcal{X}$ and $\mathcal{H}_\mathcal{Y}$, and associated feature maps $\phi_\mathcal{X}$ and $\phi_\mathcal{Y}$. Define the covariance operators

$$C_{YX} := \mathbb{E}[\phi_\mathcal{Y}(Y) \otimes \phi_\mathcal{X}(X)], \qquad C_{XX} := \mathbb{E}[\phi_\mathcal{X}(X) \otimes \phi_\mathcal{X}(X)].$$

When C_{XX} is injective, the conditional mean operator is

$$\mathcal{C}_{Y|X} := C_{YX} C_{XX}^{-1},$$

and the conditional mean embedding satisfies

$$\mu_{Y|X}(x) = \mathcal{C}_{Y|X} \phi_\mathcal{X}(x) = \mathbb{E}[\phi_\mathcal{Y}(Y) \mid X = x].$$

With data $\{(x_i, y_i)\}_{i=1}^n$, a standard regularized estimator is

$$\widehat{\mathcal{C}}_{Y|X} = \Phi_Y (K_X + \lambda I_n)^{-1} \Phi_X^\top, \qquad \widehat{\mu}_{Y|X}(x) = \widehat{\mathcal{C}}_{Y|X} \phi_\mathcal{X}(x),$$

where K_X is the Gram matrix on $\{x_i\}_{i=1}^n$, and Φ_X, Φ_Y collect the feature maps in their columns; see Song et al. (2009); Li et al. (2022).

Maximum mean discrepancy. For two distributions P, Q on $\mathcal{F}$, the maximum mean discrepancy is

$$\mathrm{MMD}(P, Q) := \|\mu_P - \mu_Q\|_{\mathcal{H}_\mathcal{F}}.$$

If k is characteristic, then $\mathrm{MMD}(P, Q) = 0$ if and only if $P = Q$ (Gretton et al., 2012b). Under i.i.d. sampling, direct quadratic kernel statistics are degenerate under the null, and recent cross-U constructions restore asymptotic normality by sample splitting (Kim and Ramdas, 2024). Our procedure adopts the same broad philosophy, but under adaptive data collection and after projection onto a pilot witness.

Interventional mean embeddings and KTE. In our setting, $\mathcal{F} = \mathcal{Y}$ and the corresponding RKHS is $\mathcal{H}_\mathcal{Y}$. The interventional mean embedding under action $a \in \mathcal{A}$ is

$$\eta(a) := \mathbb{E}_{P_X}[\mu_{Y|A,X}(a, X)] \in \mathcal{H}_\mathcal{Y},$$

and the kernel treatment effect between actions a and a' is

$$\tau(a, a') = \|\eta(a) - \eta(a')\|_{\mathcal{H}_\mathcal{Y}}.$$

B.2 Martingale Tools

Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration. A sequence $(Z_t)_{t\geq 1}$ of $\mathcal{H}$-valued random elements is a martingale difference sequence if Z_t is $\mathcal{F}_t$-measurable and

$$\mathbb{E}[Z_t \mid \mathcal{F}_{t-1}] = 0.$$

Martingale orthogonality. If (Z_t) is square-integrable and $\mathcal{H}$-valued, then for $s \neq t$,

$$\mathbb{E}\langle Z_s, Z_t\rangle_\mathcal{H} = 0. \tag{35}$$

Indeed, if $s < t$, then Z_s is $\mathcal{F}_{t-1}$-measurable, so

$$\mathbb{E}\langle Z_s, Z_t\rangle = \mathbb{E}[\langle Z_s, \mathbb{E}[Z_t \mid \mathcal{F}_{t-1}]\rangle] = 0.$$

As a consequence, if (Z_t) is square-integrable, then

$$\mathbb{E}\left\|\sum_{t=1}^n Z_t\right\|_\mathcal{H}^2 = \sum_{t=1}^n \mathbb{E}\,\|Z_t\|_\mathcal{H}^2.$$

Theorem B.1 (Strong law for martingale sums (Hall and Heyde, 1980, Thm. 2.18))**.** *Let $\{S_n = \sum_{i=1}^n X_i, \mathcal{F}_n, n \geq 1\}$ be a scalar martingale and let $\{U_n, n \geq 1\}$ be a nondecreasing sequence of positive random variables such that U_n is $\mathcal{F}_{n-1}$-measurable for each n. If $1 \leq p \leq 2$, then*

$$\lim_{n\to\infty} U_n^{-1} S_n = 0$$

almost surely on the set

$$\left\{\lim_{n\to\infty} U_n = \infty,\ \sum_{i=1}^\infty U_i^{-p}\,\mathbb{E}(|X_i|^p \mid \mathcal{F}_{i-1}) < \infty\right\}.$$

Remark B.2 (How we use Theorem B.1). Taking $p = 2$ and $U_n = n$ yields

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{\text{a.s.}} 0$$

whenever

$$\sum_{i=1}^\infty i^{-2}\,\mathbb{E}[X_i^2 \mid \mathcal{F}_{i-1}] < \infty.$$

This is the form used to control centered scalar remainder terms in the variance-estimation argument.

Theorem B.3 (Scalar martingale CLT with predictable-variation normalization)**.** *Let $(X_t, \mathcal{F}_t)_{t\geq 1}$ be a square-integrable scalar martingale difference sequence and define*

$$V_n^2 := \sum_{t=1}^n \mathbb{E}[X_t^2 \mid \mathcal{F}_{t-1}].$$

Assume $V_n^2 \to \infty$ in probability and

$$\frac{\max_{1\leq t\leq n} |X_t|}{V_n} \xrightarrow{p} 0.$$

Then

$$\frac{\sum_{t=1}^n X_t}{V_n} \Rightarrow \mathcal{N}(0, 1).$$

Theorem B.4 (Scalar martingale CLT for triangular arrays)**.** *Let* $\{(X_{n,t}, \mathcal{F}_{n,t}) : 1 \le t \le n,\ n \ge 1\}$ *be a square-integrable scalar martingale-difference array. Assume*

$$\sum_{t=1}^{n} \mathbb{E}[X_{n,t}^2 \mid \mathcal{F}_{n,t-1}] \xrightarrow{p} 1$$

and that there exists $\delta > 0$ *such that*

$$\sum_{t=1}^{n} \mathbb{E}[|X_{n,t}|^{2+\delta} \mid \mathcal{F}_{n,t-1}] \xrightarrow{p} 0.$$

Then

$$\sum_{t=1}^{n} X_{n,t} \ \Rightarrow\ \mathcal{N}(0,1).$$

C DIRECT RKHS STABILIZATION AND ITS LIMITATIONS

A natural route for adaptive inference with kernel treatment effects is to work directly with the RKHS-valued score process

$$\hat{\phi}_t = D'(\pi_t, \widehat{\mu}_{t-1}, a)(X_t, A_t, Y_t) - D'(\pi_t, \widehat{\mu}_{t-1}, a')(X_t, A_t, Y_t) \in \mathcal{H}_{\mathcal{Y}},$$

and to study asymptotic normality of a normalized RKHS-valued average. This appendix discusses the scope and limitations of that approach.

C.1 Uniform Positivity and Bounded RKHS Scores

Under uniform positivity, the raw RKHS score sequence is already uniformly bounded. In that regime, one can study the unscaled RKHS average directly through standard martingale arguments.

Proposition C.1 (Uniform positivity yields bounded RKHS scores)**.** *Assume:*

(i) Assumption 3.2,
(ii) the bounded-kernel part of Assumption 3.1,
(iii) the nuisance envelope

$$\sup_{t \ge 1} \sup_{a \in \mathcal{A},\ x \in \mathcal{X}} \|\widehat{\mu}_{t-1}(a,x)\|_{\mathcal{H}_{\mathcal{Y}}} \le B_\mu \quad a.s., \tag{36}$$

(iv) a uniform positivity condition: there exists $c > 0$ *such that*

$$\pi_t(a \mid x) \ge c \qquad \forall t \ge 1,\ \forall a \in \mathcal{A},\ \textit{for } P_X\textit{-a.e. } x.$$

Then there exists $C < \infty$ *such that*

$$\|\hat{\phi}_t\|_{\mathcal{H}_{\mathcal{Y}}} \le C \qquad a.s.\ \textit{for all } t \ge 1.$$

Proof. By boundedness of the kernel,

$$\|\phi_{\mathcal{Y}}(Y_t)\|_{\mathcal{H}_{\mathcal{Y}}} \le \sqrt{\kappa}.$$

Hence for any $b \in \mathcal{A}$,

$$\big\|D'(\pi_t, \widehat{\mu}_{t-1}, b)(X_t, A_t, Y_t)\big\|_{\mathcal{H}_{\mathcal{Y}}} \le \frac{1}{c}\big(\sqrt{\kappa} + B_\mu\big) + B_\mu.$$

The claim follows by the triangle inequality applied to the difference between the two actions a and a'. □

Proposition C.1 shows that, under uniform positivity, the score process is already uniformly controlled in norm. In such settings, variance normalization is not needed to prevent score increments from becoming large.

C.2 Trace Normalization and Operator Geometry

Let

$$\Sigma_t := \mathrm{Cov}(\hat{\phi}_t \mid \mathcal{F}_{t-1})$$

denote the conditional covariance operator of the RKHS score. A natural scalar normalization is obtained from its trace:

$$\omega_t^{-2} := \mathrm{Tr}(\Sigma_t), \qquad \widetilde{\Sigma}_t := \omega_t^2 \Sigma_t = \frac{\Sigma_t}{\mathrm{Tr}(\Sigma_t)}.$$

This normalization fixes the total conditional variance:

$$\mathrm{Tr}(\widetilde{\Sigma}_t) = 1.$$

However, it does not determine the directional geometry of the covariance operator in $\mathcal{H}_\mathcal{Y}$. Even after normalization by trace, the operators $\widetilde{\Sigma}_t$ may continue to vary across different directions of the RKHS.

For $u, v \in \mathcal{H}_\mathcal{Y}$, recall that the rank-one operator $u \otimes v$ is defined by

$$(u \otimes v)h := \langle v, h \rangle_{\mathcal{H}_\mathcal{Y}}\, u, \qquad h \in \mathcal{H}_\mathcal{Y}.$$

This notation is sufficient to make the issue explicit.

C.3 A Finite-Rank Illustration

Example C.2 (Directional variation after trace normalization)**.** *Let $u_1, u_2 \in \mathcal{H}_\mathcal{Y}$ be orthonormal, let $(\lambda_t)_{t\geq 1}$ be any positive scalar sequence, and define*

$$\Sigma_t = \lambda_t\, u_1 \otimes u_1 \quad \textit{for } t \in B_{2m-1}, \qquad \Sigma_t = \lambda_t\, u_2 \otimes u_2 \quad \textit{for } t \in B_{2m},$$

where the block lengths satisfy $|B_m| = 2^{2^m}$. Then

$$\omega_t^{-2} = \mathrm{Tr}(\Sigma_t) = \lambda_t, \qquad \widetilde{\Sigma}_t = \frac{\Sigma_t}{\mathrm{Tr}(\Sigma_t)} = \begin{cases} u_1 \otimes u_1, & t \in B_{2m-1}, \\ u_2 \otimes u_2, & t \in B_{2m}. \end{cases}$$

Thus trace normalization removes the scalar factor λ_t, but the normalized covariance still alternates between two distinct directions in $\mathcal{H}_\mathcal{Y}$. Because the block lengths grow super-exponentially, the Cesàro averages of $(\widetilde{\Sigma}_t)$ along the terminal times of odd and even blocks have different limits. In particular, $(\widetilde{\Sigma}_t)$ has no Cesàro limit.

Example C.2 shows that trace normalization controls the scale of the covariance operator but not its directional variation. Consequently, scalar normalization by $\mathrm{Tr}(\Sigma_t)$ alone is not sufficient to guarantee stabilization of an RKHS-valued covariance process.

Remark C.3 (Why we do not use a CADR-type nondegenerate efficiency bound in the RKHS). The scalar nondegeneracy condition used in CADR (Bibaut et al., 2021) is natural in a one-dimensional problem. Its literal RKHS analogue would require a lower bound of the form

$$\Sigma_t \succeq cI$$

for the conditional covariance operator of the RKHS score. In an infinite-dimensional RKHS, conditional covariance operators are typically compact, so such a lower bound fails on any infinite-dimensional subspace. This is why the main null theorem is formulated through direct control of the feasible predictable variation of the projected statistic rather than through a uniform operator-level nondegeneracy assumption.

C.4 Projection and Scalar Stabilization

The preceding discussion motivates reducing the inferential problem before normalization. Once the score is projected onto a fixed direction $v_n \in \mathcal{H}_\mathcal{Y}$,

$$\xi_t(v_n) = \langle v_n, \hat{\phi}_t \rangle_{\mathcal{H}_\mathcal{Y}},$$

the resulting sequence is scalar. At that point, normalization by a conditional standard deviation becomes a direct stabilization of the object entering the final test.

This projection step preserves the distributional nature of the problem through the choice of v_n, while avoiding the need to stabilize an operator-valued process in the full RKHS. The resulting inference problem is therefore governed by the scalar sequence $(\xi_t(v_n))$ rather than by the full RKHS score process $(\hat{\phi}_t)$.

The discussion above does not rule out RKHS-valued stabilization in principle. It shows, rather, that trace normalization alone does not control the operator-level variation needed for RKHS-valued Gaussian limits under general adaptive sampling, whereas projection reduces the normalization problem to a scalar one.

D PROOFS FOR THE ADAPTIVE TEST

This appendix contains the proofs for the results stated in Section 4. We first prove the projected doubly robust identity. We then establish projected moment bounds, verify the predictable variation and Lyapunov conditions of the feasible statistic, and finally prove null asymptotic normality.

D.1 Projected doubly robust identity

Proof of Lemma 4.1. Since w_n is $\mathcal{F}_n$-measurable and $n < t$, it is $\mathcal{F}_{t-1}$-measurable. Therefore

$$\mathbb{E}[\xi_t(w_n) \,|\, \mathcal{F}_{t-1}] = \left\langle w_n, \mathbb{E}\left[\hat{\phi}_t^{(2)} \mid \mathcal{F}_{t-1}\right] \right\rangle_{\mathcal{H}_\mathcal{Y}}.$$

It suffices to show that

$$\mathbb{E}\left[\hat{\phi}_t^{(2)} \mid \mathcal{F}_{t-1}\right] = \Psi(a, a').$$

Fix $b \in \{a, a'\}$. For any predictable nuisance $\bar{\mu}$,

$$\mathbb{E}[D'(\pi_t, \bar{\mu}, b)(X_t, A_t, Y_t) \,|\, \mathcal{F}_{t-1}] = \eta(b),$$

by the standard doubly robust identity. Taking the difference between $b = a$ and $b = a'$ yields

$$\mathbb{E}\left[\hat{\phi}_t^{(2)} \mid \mathcal{F}_{t-1}\right] = \eta(a) - \eta(a') = \Psi(a, a').$$

Hence

$$\mathbb{E}[\xi_t(w_n) \,|\, \mathcal{F}_{t-1}] = \langle w_n, \Psi(a, a') \rangle_{\mathcal{H}_\mathcal{Y}} =: \theta_n(a, a').$$

Under $H_0 : \eta(a) = \eta(a')$, this equals 0. □

D.2 Projected moment bounds

Lemma D.1 (Projected q-th moment bound)**.** *Assume Assumptions 3.2, 3.1, (7), and 4.2. Then for every $q \geq 2$ there exists $C_q < \infty$ such that, for every $t \in \mathcal{I}_2$,*

$$\mathbb{E}[|\xi_t(w_n)|^q \,|\, \mathcal{F}_{t-1}] \leq C_q \, t^{\alpha(q-1)} \qquad a.s.$$

In particular,

$$\mathbb{E}\left[\xi_t(w_n)^2 \,\middle|\, \mathcal{F}_{t-1}\right] \lesssim t^\alpha, \qquad \mathbb{E}\left[\xi_t(w_n)^4 \,\middle|\, \mathcal{F}_{t-1}\right] \lesssim t^{3\alpha},$$

and therefore

$$\sigma_{0,t}^2(w_n) \lesssim t^\alpha \qquad a.s.$$

Proof. Since $\|w_n\|_{\mathcal{H}_\mathcal{Y}} = 1$,

$$|\xi_t(w_n)|^q = |\langle w_n, \hat{\phi}_t^{(2)} \rangle_{\mathcal{H}_\mathcal{Y}}|^q \leq \|\hat{\phi}_t^{(2)}\|_{\mathcal{H}_\mathcal{Y}}^q.$$

Write

$$\hat{\phi}_t^{(2)} = G_t(a) - G_t(a'), \qquad G_t(b) := D'(\pi_t, \widehat{\mu}_{t-1}^{(2)}, b)(X_t, A_t, Y_t).$$

By $(u+v)^q \le 2^{q-1}(u^q+v^q)$,

$$\|\hat{\phi}_t^{(2)}\|_{\mathcal{H}_\mathcal{Y}}^q \le 2^{q-1}\|G_t(a)\|_{\mathcal{H}_\mathcal{Y}}^q + 2^{q-1}\|G_t(a')\|_{\mathcal{H}_\mathcal{Y}}^q.$$

Fix $b \in \{a, a'\}$. By boundedness of the kernel and the nuisance envelope,

$$\left\|\phi_\mathcal{Y}(Y_t) - \widehat{\mu}_{t-1}^{(2)}(A_t, X_t)\right\|_{\mathcal{H}_\mathcal{Y}} \le \sqrt{\kappa} + B_\mu =: C_\mu,$$

and

$$\|\widehat{\mu}_{t-1}^{(2)}(b, X_t)\|_{\mathcal{H}_\mathcal{Y}} \le B_\mu.$$

Therefore

$$\|G_t(b)\|_{\mathcal{H}_\mathcal{Y}} \le C_\mu \frac{\mathbb{1}\{A_t = b\}}{\pi_t(b \mid X_t)} + B_\mu,$$

so

$$\|G_t(b)\|_{\mathcal{H}_\mathcal{Y}}^q \lesssim \frac{\mathbb{1}\{A_t = b\}}{\pi_t(b \mid X_t)^q} + 1.$$

Taking conditional expectations and using

$$\mathbb{E}\left[\frac{\mathbb{1}\{A_t = b\}}{\pi_t(b \mid X_t)^q} \,\middle|\, \mathcal{F}_{t-1}\right] = \mathbb{E}_{X_t}\left[\pi_t(b \mid X_t)^{-(q-1)}\right] \lesssim t^{\alpha(q-1)}$$

by Assumption 4.2, we obtain

$$\mathbb{E}\left[\|G_t(b)\|_{\mathcal{H}_\mathcal{Y}}^q \,\middle|\, \mathcal{F}_{t-1}\right] \lesssim t^{\alpha(q-1)}.$$

The same bound therefore holds for $\|\hat{\phi}_t^{(2)}\|_{\mathcal{H}_\mathcal{Y}}^q$ and hence for $|\xi_t(w_n)|^q$. The variance bound follows from

$$\sigma_{0,t}^2(w_n) \le \mathbb{E}\left[\xi_t(w_n)^2 \mid \mathcal{F}_{t-1}\right]. \qquad \square$$

D.3 Predictable variation of the feasible statistic

Define

$$X_{n,t} := \frac{1}{\sqrt{n}}\,\widetilde{\sigma}_t(w_n)^{-1}\,\xi_t(w_n), \qquad t \in \mathcal{I}_2.$$

Under H_0, Lemma 4.1 shows that $(X_{n,t}, \mathcal{F}_t)_{t\in\mathcal{I}_2}$ is a scalar martingale-difference array.

Lemma D.2 (Predictable variation convergence)**.** *Assume the conditions of Theorem 4.4. Then*

$$V_n^2 := \sum_{t\in\mathcal{I}_2} \mathbb{E}\left[X_{n,t}^2 \,\middle|\, \mathcal{F}_{t-1}\right] = \frac{1}{n}\sum_{t\in\mathcal{I}_2} \frac{\sigma_{0,t}^2(w_n)}{\widetilde{\sigma}_t^2(w_n)} \xrightarrow{p} 1.$$

Proof. This is exactly (33) from Corollary 5.5. $\square$

D.4 Lyapunov condition for the feasible statistic

Lemma D.3 (Feasible Lyapunov condition)**.** *Assume the conditions of Theorem 4.4. Then*

$$\sum_{t\in\mathcal{I}_2} \mathbb{E}\left[|X_{n,t}|^4 \,\middle|\, \mathcal{F}_{t-1}\right] \xrightarrow{p} 0.$$

Proof. By definition of $X_{n,t}$,

$$\sum_{t\in\mathcal{I}_2} \mathbb{E}\left[|X_{n,t}|^4 \,\middle|\, \mathcal{F}_{t-1}\right] = \frac{1}{n^2}\sum_{t\in\mathcal{I}_2} \widetilde{\sigma}_t(w_n)^{-4}\,\mathbb{E}\left[\xi_t(w_n)^4 \,\middle|\, \mathcal{F}_{t-1}\right].$$

Since $\widetilde{\sigma}_t(w_n) \ge \epsilon_n = n^{-\rho}$, Lemma D.1 yields

$$\sum_{t\in\mathcal{I}_2} \mathbb{E}\left[|X_{n,t}|^4 \,\middle|\, \mathcal{F}_{t-1}\right] \lesssim n^{-2}\epsilon_n^{-4} \sum_{t=n+1}^{2n} t^{3\alpha} \lesssim n^{-2+4\rho}\cdot n^{1+3\alpha} = n^{3\alpha-1+4\rho}.$$

By (4.4), $\rho < (1-3\alpha)/4$, so the right-hand side converges to 0. $\square$

D.5 Proof of the null theorem

Proof of Theorem 4.4. Under H_0, Lemma 4.1 implies that $(X_{n,t}, \mathcal{F}_t)_{t\in\mathcal{I}_2}$ is a scalar martingale-difference array. By Lemma D.2,

$$\sum_{t\in\mathcal{I}_2} \mathbb{E}\big[X_{n,t}^2 \,\big|\, \mathcal{F}_{t-1}\big] \xrightarrow{p} 1.$$

By Lemma D.3,

$$\sum_{t\in\mathcal{I}_2} \mathbb{E}\big[|X_{n,t}|^4 \,\big|\, \mathcal{F}_{t-1}\big] \xrightarrow{p} 0.$$

Applying the scalar martingale CLT for triangular arrays, Theorem B.4, gives

$$\sum_{t\in\mathcal{I}_2} X_{n,t} \;\xRightarrow{d}\; \mathcal{N}(0,1).$$

By definition,

$$\sum_{t\in\mathcal{I}_2} X_{n,t} = \frac{1}{\sqrt{n}} \sum_{t\in\mathcal{I}_2} \widetilde{\sigma}_t(w_n)^{-1}\xi_t(w_n) = T_n(w_n).$$

This proves the claim. □

E PROOF OF THE PROJECTED CONDITIONAL STANDARD DEVIATION CONSISTENCY THEOREM

This appendix proves Theorem 5.4. The proof does not use RKHS empirical-process argument. Conditionally on the pilot fold, the direction w_n is fixed and the inferential-fold problem becomes scalar (Bibaut et al., 2021).

Throughout this appendix, condition on $\mathcal{F}_n$. By construction,

$$\|w_n\|_{\mathcal{H}_\mathcal{Y}} = 1. \tag{37}$$

All almost-sure statements below are understood conditionally on $\mathcal{F}_n$.

E.1 Scalarization of the projected inferential-fold problem

For $s \in \mathcal{I}_2$, define the scalar projected outcome

$$\widetilde{Y}_s := \langle w_n, \phi_\mathcal{Y}(Y_s)\rangle_{\mathcal{H}_\mathcal{Y}},$$

and the projected nuisance sequence

$$\widetilde{\mu}_{s-1}(b,x) := \Big\langle w_n, \widehat{\mu}^{(2)}_{s-1}(b,x)\Big\rangle_{\mathcal{H}_\mathcal{Y}}, \qquad b\in\mathcal{A},\ x\in\mathcal{X}.$$

Also define the projected nuisance contrast

$$\Delta\widetilde{\mu}_{s-1}(x) := \widetilde{\mu}_{s-1}(a,x) - \widetilde{\mu}_{s-1}(a',x).$$

Since $\|w_n\|_{\mathcal{H}_\mathcal{Y}} = 1$, boundedness of the kernel and the nuisance envelope imply

$$|\widetilde{Y}_s| \le \sqrt{\kappa}, \qquad \sup_{s\ge 1}\sup_{b\in\mathcal{A},\ x\in\mathcal{X}} |\widetilde{\mu}_{s-1}(b,x)| \le B_\mu \quad \text{a.s.} \tag{38}$$

Likewise, Assumption 5.1 implies

$$\big\|\widetilde{\mu}_{t-1} - \widetilde{\mu}_\infty\big\|_{L_2(P_X\times\mu_\mathcal{A})} = O(t^{-\beta}) \qquad \text{a.s.}, \tag{39}$$

where

$$\widetilde{\mu}_\infty(b,x) := \langle w_n, \mu_\infty(b,x)\rangle_{\mathcal{H}_\mathcal{Y}}.$$

For $s < t$ in $\mathcal{I}_2$, the history-recycled projected score can be written as

$$\begin{aligned}\check{\xi}_{s,t}(w_n) = &\frac{\mathbb{1}\{A_s = a\}}{\pi_t(a \mid X_s)}\Big(\widetilde{Y}_s - \widetilde{\mu}_{s-1}(a, X_s)\Big) \\ &- \frac{\mathbb{1}\{A_s = a'\}}{\pi_t(a' \mid X_s)}\Big(\widetilde{Y}_s - \widetilde{\mu}_{s-1}(a', X_s)\Big) \\ &+ \Delta\widetilde{\mu}_{s-1}(X_s).\end{aligned} \tag{40}$$

Thus, conditionally on $\mathcal{F}_n$, the projected inferential-fold score is exactly a scalar doubly robust contrast between the two target actions a and a'.

E.2 Projected $f_1, \dots, f_5$ decomposition

Let $g_t := \pi_t/\pi_{\mathrm{ref}}$. For any $g \in \mathcal{G}_{\mathrm{ref}}$ and any scalar nuisance $m : \mathcal{A} \times \mathcal{X} \to \mathbb{R}$, define

$$\Delta m(x) := m(a, x) - m(a', x),$$

$$\begin{aligned}f_1(g, m)(X, A, Y) := &\frac{\mathbb{1}\{A = a\}}{g(a \mid X)\pi_{\mathrm{ref}}(a \mid X)^2}\Big(\widetilde{Y} - m(a, X)\Big)^2 \\ &+ \frac{\mathbb{1}\{A = a'\}}{g(a' \mid X)\pi_{\mathrm{ref}}(a' \mid X)^2}\Big(\widetilde{Y} - m(a', X)\Big)^2,\end{aligned} \tag{41}$$

$$\begin{aligned}f_2(m)(X, A, Y) := 2\Bigg[&\frac{\mathbb{1}\{A = a\}}{\pi_{\mathrm{ref}}(a \mid X)}\Big(\widetilde{Y} - m(a, X)\Big) \\ &- \frac{\mathbb{1}\{A = a'\}}{\pi_{\mathrm{ref}}(a' \mid X)}\Big(\widetilde{Y} - m(a', X)\Big)\Bigg]\Delta m(X),\end{aligned} \tag{42}$$

$$f_3(g, m)(X, A, Y) := g(A \mid X)\,\Delta m(X)^2, \tag{43}$$

$$\begin{aligned}f_4(m)(X, A, Y) := &\frac{\mathbb{1}\{A = a\}}{\pi_{\mathrm{ref}}(a \mid X)}\Big(\widetilde{Y} - m(a, X)\Big) \\ &- \frac{\mathbb{1}\{A = a'\}}{\pi_{\mathrm{ref}}(a' \mid X)}\Big(\widetilde{Y} - m(a', X)\Big),\end{aligned} \tag{44}$$

$$f_5(g, m)(X, A, Y) := g(A \mid X)\,\Delta m(X). \tag{45}$$

Lemma E.1 (Projected weighted-moment decomposition)**.** *For every $s < t$ in $\mathcal{I}_2$,*

$$g_t(A_s \mid X_s)\,\check{\xi}_{s,t}(w_n)^2 = f_1(g_t, \widetilde{\mu}_{s-1})(O_s) + f_2(\widetilde{\mu}_{s-1})(O_s) + f_3(g_t, \widetilde{\mu}_{s-1})(O_s), \tag{46}$$

$$g_t(A_s \mid X_s)\,\check{\xi}_{s,t}(w_n) = f_4(\widetilde{\mu}_{s-1})(O_s) + f_5(g_t, \widetilde{\mu}_{s-1})(O_s). \tag{47}$$

Proof. This is a direct expansion of (40). The mixed product between the a- and a'-terms vanishes because the indicators $\mathbb{1}\{A_s = a\}$ and $\mathbb{1}\{A_s = a'\}$ are disjoint. □

E.3 Empirical and target moment pieces

For $t \in \mathcal{I}_2$ and $t \geq n + 2$, define

$$\widehat{\Phi}^{(k)}_{1,t} := \frac{1}{m_t}\sum_{s \in S_t}\frac{\pi_{\mathrm{ref}}(A_s \mid X_s)}{\pi_s(A_s \mid X_s)}\,f_k(g_t, \widetilde{\mu}_{s-1})(O_s), \qquad k \in \{1, 3, 5\},$$

$$\widehat{\Phi}^{(2)}_{1,t} := \frac{1}{m_t}\sum_{s \in S_t}\frac{\pi_{\mathrm{ref}}(A_s \mid X_s)}{\pi_s(A_s \mid X_s)}\,f_2(\widetilde{\mu}_{s-1})(O_s),$$

$$\widehat{\Phi}^{(4)}_{2,t} := \frac{1}{m_t}\sum_{s \in S_t}\frac{\pi_{\mathrm{ref}}(A_s \mid X_s)}{\pi_s(A_s \mid X_s)}\,f_4(\widetilde{\mu}_{s-1})(O_s).$$

Define the corresponding Cesàro targets

$$\overline{\Phi}^{(k)}_{1,t} := \frac{1}{m_t}\sum_{s\in S_t} P_{\mathrm{ref}}\big[f_k(g_t,\widetilde{\mu}_{s-1})(O)\big], \qquad k\in\{1,3,5\},$$

$$\overline{\Phi}^{(2)}_{1,t} := \frac{1}{m_t}\sum_{s\in S_t} P_{\mathrm{ref}}\big[f_2(\widetilde{\mu}_{s-1})(O)\big], \qquad \overline{\Phi}^{(4)}_{2,t} := \frac{1}{m_t}\sum_{s\in S_t} P_{\mathrm{ref}}\big[f_4(\widetilde{\mu}_{s-1})(O)\big].$$

Finally, define the current-time targets

$$\Phi^{(1)}_{0,1,t} := P_{\mathrm{ref}}\big[f_1(g_t,\widetilde{\mu}_{t-1})(O)\big], \qquad \Phi^{(2)}_{0,1,t} := P_{\mathrm{ref}}\big[f_2(\widetilde{\mu}_{t-1})(O)\big], \qquad \Phi^{(3)}_{0,1,t} := P_{\mathrm{ref}}\big[f_3(g_t,\widetilde{\mu}_{t-1})(O)\big],$$

$$\Phi^{(4)}_{0,2,t} := P_{\mathrm{ref}}\big[f_4(\widetilde{\mu}_{t-1})(O)\big], \qquad \Phi^{(5)}_{0,2,t} := P_{\mathrm{ref}}\big[f_5(g_t,\widetilde{\mu}_{t-1})(O)\big].$$

By Lemma E.1,

$$\widehat{M}_{1,t}(w_n) = \widehat{\Phi}^{(4)}_{2,t} + \widehat{\Phi}^{(5)}_{1,t}, \qquad M_{1,t}(w_n) = \Phi^{(4)}_{0,2,t} + \Phi^{(5)}_{0,2,t}, \tag{48}$$

and

$$\widehat{M}_{2,t}(w_n) = \widehat{\Phi}^{(1)}_{1,t} + \widehat{\Phi}^{(2)}_{1,t} + \widehat{\Phi}^{(3)}_{1,t}, \qquad M_{2,t}(w_n) = \Phi^{(1)}_{0,1,t} + \Phi^{(2)}_{0,1,t} + \Phi^{(3)}_{0,1,t}. \tag{49}$$

E.4 Proof of Theorem 5.4

Reference law, truncated policy class, and intermediate Cesàro moments. For any measurable $h : \mathcal{X}\times\mathcal{A}\times\mathcal{Y}\to\mathbb{R}$, define

$$P_{\mathrm{ref}}h := \mathbb{E}_{X\sim P_X,\ A\sim\pi_{\mathrm{ref}}(\cdot|X),\ Y\sim P_{Y|X,A}(\cdot|X,A)}\big[h(X,A,Y)\big].$$

Equivalently,

$$P_{\mathrm{ref}} = P_X\otimes\pi_{\mathrm{ref}}(\cdot\mid X)\otimes P_{Y|X,A}(\cdot\mid X,A).$$

For $\delta>0$, define the truncated policy subclass

$$\Pi_e(\delta) := \Big\{\pi\in\Pi : \inf_{b\in\mathcal{A},\ x\in\mathcal{X}} \pi(b\mid x)\ge\delta\Big\},$$

and its ratio representation

$$\mathcal{G}_e(\delta) := \Big\{\pi/\pi_{\mathrm{ref}} : \pi\in\Pi_e(\delta)\Big\}\subseteq\mathcal{G}_{\mathrm{ref}}.$$

For $t\ge n+2$, define the intermediate Cesàro moments

$$\overline{M}_{1,t}(w_n) := \overline{\Phi}^{(4)}_{2,t} + \overline{\Phi}^{(5)}_{1,t}, \qquad \overline{M}_{2,t}(w_n) := \overline{\Phi}^{(1)}_{1,t} + \overline{\Phi}^{(2)}_{1,t} + \overline{\Phi}^{(3)}_{1,t}.$$

Then, for $j\in\{1,2\}$,

$$\widehat{M}_{j,t}(w_n) - M_{j,t}(w_n) = \big(\widehat{M}_{j,t}(w_n) - \overline{M}_{j,t}(w_n)\big) + \big(\overline{M}_{j,t}(w_n) - M_{j,t}(w_n)\big).$$

For $k\in\{1,3,5\}$ and $\delta>0$, define the projected localized classes

$$\mathcal{F}_{k,t}(\delta) := \Big\{\big(f_k(g,\widetilde{\mu}_{s-1})(O_s)\big)_{s\in S_t} :\ g\in\mathcal{G}_e(\delta)\Big\}.$$

Proof of Theorem 5.4. Condition on $\mathcal{F}_n$. Then w_n is deterministic and $\|w_n\|_{\mathcal{H}_\mathcal{Y}}=1$. Define

$$\widetilde{Y}_s := \langle w_n, \phi_\mathcal{Y}(Y_s)\rangle_{\mathcal{H}_\mathcal{Y}}, \qquad \widetilde{\mu}_{s-1}(b,x) := \langle w_n, \widehat{\mu}^{(2)}_{s-1}(b,x)\rangle_{\mathcal{H}_\mathcal{Y}}.$$

By contraction,

$$|\widetilde{Y}_s|\le\sqrt{\kappa}, \qquad \sup_{s,b,x}|\widetilde{\mu}_{s-1}(b,x)|\le B_\mu, \qquad \|\widetilde{\mu}_{t-1}-\widetilde{\mu}_\infty\|_{L_2(P_X\times\mu_\mathcal{A})} = O(t^{-\beta}) \quad \text{a.s.}$$

Thus the inferential-fold problem is scalar.

Let $\delta_t := c_\pi t^{-\alpha}$. For the empirical-process part, consider the classes

$$\mathcal{F}_{k,t}(\delta_t) = \Big\{ \big(f_k(g, \widetilde{\mu}_{s-1})(O_s)\big)_{s\in S_t} : g \in \mathcal{G}_e(\delta_t) \Big\}, \qquad k \in \{1,3,5\}.$$

The same monotone bracketing argument as in (Bibaut et al., 2021) applies after projection: the map $u \mapsto \langle w_n, u\rangle$ is a contraction, so projection does not enlarge the entropy of the policy-ratio class. Hence the martingale empirical-process terms satisfy

$$\big|\widehat{M}_{j,t}(w_n) - \overline{M}_{j,t}(w_n)\big| = O_{\text{a.s.}}(m_t^{-\nu_{\text{emp}}})$$

for some $\nu_{\text{emp}} > 0$, $j = 1, 2$.

For the approximation part, use the projected analogue of Lemma 5 of Bibaut et al. (2021): for any $g, g_1 \in \mathcal{G}_e(\delta)$ and any scalar nuisances m, m_1,

$$|\Phi^{(1)}_{0,1}(g, m) - \Phi^{(1)}_{0,1}(g_1, m_1)| \lesssim \delta^{-2}\|g - g_1\|_{L_1(P_X\times\mu_{\mathcal{A}})} + \delta^{-1}\|m - m_1\|_{L_1(P_X\times\mu_{\mathcal{A}})},$$

while the remaining four pieces are Lipschitz in m with no worse than $O(1)$ constant. Since $\overline{\Phi}$ and Φ_0 are evaluated at the same current policy g_t, only the nuisance difference remains. Therefore

$$\big|\overline{M}_{j,t}(w_n) - M_{j,t}(w_n)\big| \lesssim \frac{1}{m_t}\sum_{s\in S_t} \delta_t^{-1}\|\widetilde{\mu}_{s-1} - \widetilde{\mu}_{t-1}\|_{L_1(P_X\times\mu_{\mathcal{A}})} = O_{\text{a.s.}}(t^{-(\beta-\alpha)}),$$

because $\alpha < \beta$.

Let

$$\nu := \min\{\nu_{\text{emp}}, \beta - \alpha\} > 0.$$

Then

$$\big|\widehat{M}_{j,t}(w_n) - M_{j,t}(w_n)\big| = O_{\text{a.s.}}(m_t^{-\nu}), \qquad j = 1, 2.$$

Averaging over $t \in \mathcal{I}_2$, and using $m_t = t - n - 1$,

$$\frac{1}{n}\sum_{t\in\mathcal{I}_2} m_t^{-\nu} = \frac{1}{n}\sum_{u=1}^{n-1} u^{-\nu} = O(n^{-\nu'})$$

for some $\nu' > 0$. Renaming ν' as $\nu(\alpha, \beta, p)$ yields (29) and (30).

Finally,

$$\widehat{\sigma}_t^2(w_n) = \big(\widehat{M}_{2,t}(w_n) - \widehat{M}_{1,t}(w_n)^2\big)_+, \qquad \sigma_{0,t}^2(w_n) = \big(M_{2,t}(w_n) - M_{1,t}(w_n)^2\big)_+.$$

Since $x \mapsto x_+$ is 1-Lipschitz and $M_{1,t}(w_n) = \theta_n(a, a')$ is uniformly bounded, while $\widehat{M}_{1,t}(w_n) \to M_{1,t}(w_n)$ a.s., the map $(m_1, m_2) \mapsto (m_2 - m_1^2)_+$ is eventually Lipschitz on the relevant set. Hence (31) follows. □

F PROOFS FOR VALIDITY, WITNESS CONSISTENCY, AND POWER

This appendix contains the proofs for the results in Section 6. We first prove asymptotic validity by combining the null theorem with the variance-consistency result. We then prove consistency of the pilot witness and of the pilot direction. Finally, we prove consistency of the adaptive test under fixed alternatives.

F.1 Proof of Corollary 6.1

Proof of Corollary 6.1. This is immediate from Theorem 4.4. □

F.2 Proof of Theorem 6.2

Proof of Theorem 6.2. Recall

$$v_n = \frac{1}{n} \sum_{t \in \mathcal{I}_1} \hat{\phi}_t^{(1)}.$$

By the doubly robust identity, for every $t \in \mathcal{I}_1$,

$$\mathbb{E}\left[\hat{\phi}_t^{(1)} \,\middle|\, \mathcal{F}_{t-1}\right] = \Psi(a, a').$$

Therefore

$$v_n - \Psi(a, a') = \frac{1}{n} \sum_{t \in \mathcal{I}_1} \Delta_t, \qquad \Delta_t := \hat{\phi}_t^{(1)} - \mathbb{E}[\hat{\phi}_t^{(1)} \mid \mathcal{F}_{t-1}], \tag{50}$$

and $(\Delta_t)_{t \in \mathcal{I}_1}$ is a sequence of $\mathcal{H}_\mathcal{Y}$-valued martingale differences.

It remains to bound $\mathbb{E}[\|\Delta_t\|_{\mathcal{H}_\mathcal{Y}}^2]$. Since

$$\|\Delta_t\|_{\mathcal{H}_\mathcal{Y}}^2 \le 2\|\hat{\phi}_t^{(1)}\|_{\mathcal{H}_\mathcal{Y}}^2 + 2\|\Psi(a, a')\|_{\mathcal{H}_\mathcal{Y}}^2,$$

it is enough to bound $\mathbb{E}[\|\hat{\phi}_t^{(1)}\|_{\mathcal{H}_\mathcal{Y}}^2]$. By the same argument as in Lemma D.1 with $q = 2$ and $r = 1$ in place of $r = 2$,

$$\mathbb{E}\left[\|\hat{\phi}_t^{(1)}\|_{\mathcal{H}_\mathcal{Y}}^2 \,\middle|\, \mathcal{F}_{t-1}\right] \lesssim t^\alpha.$$

Hence

$$\mathbb{E}\left[\|\Delta_t\|_{\mathcal{H}_\mathcal{Y}}^2\right] \lesssim t^\alpha.$$

Martingale orthogonality then yields

$$\begin{aligned} \mathbb{E}\left[\|v_n - \Psi(a, a')\|_{\mathcal{H}_\mathcal{Y}}^2\right] &= \frac{1}{n^2} \sum_{t \in \mathcal{I}_1} \mathbb{E}\left[\|\Delta_t\|_{\mathcal{H}_\mathcal{Y}}^2\right] \\ &\lesssim \frac{1}{n^2} \sum_{t=1}^{n} t^\alpha \asymp n^{\alpha - 1}. \end{aligned} \tag{51}$$

This proves (34). Since $\alpha < 1$, the right-hand side tends to zero, so

$$v_n \xrightarrow{L_2(\mathcal{H}_\mathcal{Y})} \Psi(a, a') \qquad \text{and hence} \qquad v_n \xrightarrow{p} \Psi(a, a'). \qquad \square$$

F.3 Proof of Corollary 6.3

Proof of Corollary 6.3. Since $\Psi(a, a') \neq 0$, Theorem 6.2 implies

$$v_n \xrightarrow{p} \Psi(a, a').$$

In particular,

$$\Pr(v_n = 0) \le \Pr(\|v_n - \Psi(a, a')\|_{\mathcal{H}_\mathcal{Y}} \ge \|\Psi(a, a')\|_{\mathcal{H}_\mathcal{Y}}) \longrightarrow 0.$$

On the event $\{v_n \neq 0\}$, we have $w_n = v_n / \|v_n\|_{\mathcal{H}_\mathcal{Y}}$. Since the map

$$v \mapsto \frac{v}{\|v\|_{\mathcal{H}_\mathcal{Y}}}$$

is continuous away from 0, the continuous mapping theorem gives

$$\frac{v_n}{\|v_n\|_{\mathcal{H}_\mathcal{Y}}} \xrightarrow{p} \frac{\Psi(a, a')}{\|\Psi(a, a')\|_{\mathcal{H}_\mathcal{Y}}}.$$

Because $\Pr(v_n = 0) \to 0$, the tie-breaker case is asymptotically negligible, and therefore

$$w_n \xrightarrow{p} \frac{\Psi(a, a')}{\|\Psi(a, a')\|_{\mathcal{H}_\mathcal{Y}}}.$$

$\square$

F.4 Proof of Corollary 6.4

Proof of Corollary 6.4. Let

$$\theta_n(a, a') := \langle w_n, \Psi(a, a') \rangle_{\mathcal{H}_\mathcal{Y}}.$$

By Corollary 6.3,

$$\theta_n(a, a') \xrightarrow{p} \|\Psi(a, a')\|_{\mathcal{H}_\mathcal{Y}} > 0. \tag{52}$$

Write

$$\begin{aligned} T_n(w_n) = \frac{1}{\sqrt{n}} \sum_{t \in \mathcal{I}_2} \widetilde{\sigma}_t(w_n)^{-1} \Big(\xi_t(w_n) - \theta_n(a, a') \Big) \\ + \frac{\theta_n(a, a')}{\sqrt{n}} \sum_{t \in \mathcal{I}_2} \widetilde{\sigma}_t(w_n)^{-1}. \end{aligned} \tag{53}$$

We first control the stochastic term. Since

$$\mathbb{E}[\xi_t(w_n) - \theta_n(a, a') \,|\, \mathcal{F}_{t-1}] = 0,$$

the first term in (53) is a martingale array with predictable variation

$$\frac{1}{n} \sum_{t \in \mathcal{I}_2} \frac{\sigma_{0,t}^2(w_n)}{\widetilde{\sigma}_t^2(w_n)} \xrightarrow{p} 1$$

by Corollary 5.5. Thus the first term in (53) is $O_p(1)$.

We now lower bound the deterministic drift. For $s < t$ in $\mathcal{I}_2$, boundedness of the kernel, the nuisance envelope, and Assumption 4.2 give

$$|\check{\xi}_{s,t}(w_n)| \lesssim \frac{\mathbb{1}\{A_s = a\}}{\pi_t(a \mid X_s)} + \frac{\mathbb{1}\{A_s = a'\}}{\pi_t(a' \mid X_s)} + 1 \lesssim t^\alpha.$$

Moreover,

$$w_{s,t} = \frac{\pi_t(A_s \mid X_s)}{\pi_s(A_s \mid X_s)} \lesssim s^\alpha \lesssim t^\alpha,$$

since $\pi_t(A_s \mid X_s) \leq 1$ and $\pi_s(A_s \mid X_s) \gtrsim s^{-\alpha}$. Therefore

$$\widehat{M}_{2,t}(w_n) \lesssim t^\alpha \cdot t^{2\alpha} = t^{3\alpha},$$

so

$$\widehat{\sigma}_t(w_n) \lesssim t^{3\alpha/2} \qquad \text{and hence} \qquad \widetilde{\sigma}_t(w_n)^{-1} \gtrsim t^{-3\alpha/2}.$$

Consequently,

$$\frac{1}{\sqrt{n}} \sum_{t \in \mathcal{I}_2} \widetilde{\sigma}_t(w_n)^{-1} \gtrsim \frac{1}{\sqrt{n}} \sum_{t=n+1}^{2n} t^{-3\alpha/2} \asymp n^{1/2 - 3\alpha/2}.$$

By Assumption 5.3, $\alpha < \alpha_\star(\beta, p) \leq 1/3$, so $1/2 - 3\alpha/2 > 0$. Therefore the drift term in (53) diverges to $+\infty$ in probability by (52), while the stochastic term remains $O_p(1)$. Hence

$$T_n(w_n) \xrightarrow{p} +\infty.$$

This proves consistency of the rejection rule. □

G CLOSED FORMS FOR SAMPLE-SPLIT AND PROJECTED ADAPTIVE STATISTICS

G.1 Projected adaptive DR-KTE (PADR-KTE)

We now give the kernel-matrix closed form for the projected adaptive statistic introduced in Sections 4–6. Unlike the previous cross-fold variance-stabilized construction, the new estimator is not bilinear in two stabilized folds. The first fold is used only to construct the RKHS witness, while the second fold is reduced to a scalar sequential adaptive-DR problem after projection.

For simplicity we write the binary contrast $a = 1$ versus $a' = 0$. For a general pair (a, a'), replace the indicators and propensities accordingly.

Chronological split and fold-local indexing. Write $T = 2n$ and split chronologically:

$$\mathcal{I}_1 = \{1, \dots, n\}, \qquad \mathcal{I}_2 = \{n+1, \dots, 2n\}.$$

For convenience, index each fold locally by $u = 1, \dots, n$:

$$(X_{1,u}, A_{1,u}, Y_{1,u}) := (X_u, A_u, Y_u), \qquad (X_{2,u}, A_{2,u}, Y_{2,u}) := (X_{n+u}, A_{n+u}, Y_{n+u}).$$

Let $\pi_u^{(1)}(\cdot \mid x) := \pi_u(\cdot \mid x)$ and $\pi_u^{(2)}(\cdot \mid x) := \pi_{n+u}(\cdot \mid x)$.

Let $k_{\mathcal{Y}}$ be a PD kernel on outcomes with RKHS $\mathcal{H}_{\mathcal{Y}}$ and feature map $\varphi_{\mathcal{Y}}(y) = k_{\mathcal{Y}}(\cdot, y)$. For a fold $r \in \{1, 2\}$, define the feature operator

$$\Phi_{\mathcal{Y},r} c = \sum_{u=1}^{n} c_u \, \varphi_{\mathcal{Y}}(Y_{r,u}) \in \mathcal{H}_{\mathcal{Y}}, \qquad \langle \Phi_{\mathcal{Y},r} c, \Phi_{\mathcal{Y},r'} d \rangle_{\mathcal{H}_{\mathcal{Y}}} = c^\top K_{\mathcal{Y}}^{(r,r')} d,$$

where $K_{\mathcal{Y}}^{(r,r')} = [k_{\mathcal{Y}}(Y_{r,u}, Y_{r',v})]_{u,v=1}^{n}$. Let $K_{\mathcal{X}}^{(r,r)} = [k_{\mathcal{X}}(X_{r,u}, X_{r,v})]_{u,v=1}^{n}$ be the within-fold covariate Gram block.

Past-only KRR coefficient vectors. Fix a fold $r \in \{1, 2\}$, a local time $u \in \{1, \dots, n\}$, and an arm $b \in \{0, 1\}$. Define the past arm-specific index set

$$J_{r,u}^{(b)} := \{\, v \in \{1, \dots, u-1\} : A_{r,v} = b \,\}.$$

Let $e_u \in \mathbb{R}^n$ be the u-th canonical basis vector, and let $S_{r,u}^{(b)} \in \mathbb{R}^{n \times |J_{r,u}^{(b)}|}$ be the selector that inserts a vector indexed by $J_{r,u}^{(b)}$ into the full fold coordinates. With ridge $\lambda > 0$, define the past-only KRR coefficient vector

$$h_{r,u}^{(b)} := S_{r,u}^{(b)} \Big(K_{\mathcal{X}}^{(r,r)}[J_{r,u}^{(b)}, J_{r,u}^{(b)}] + \lambda I \Big)^{-1} K_{\mathcal{X}}^{(r,r)}[J_{r,u}^{(b)}, u] \in \mathbb{R}^n,$$

with the convention $h_{r,u}^{(b)} = 0$ if $J_{r,u}^{(b)} = \emptyset$. Then

$$\widehat{\mu}_{u-1}^{(r)}(b, X_{r,u}) = \Phi_{\mathcal{Y},r} \, h_{r,u}^{(b)}.$$

Fold-1 pilot witness. For the contrast $1 - 0$, the fold-1 RKHS score at local time u is

$$\hat{\phi}_u^{(1)} = \Phi_{\mathcal{Y},1} \, d_u^{\mathrm{pil}},$$

where

$$d_u^{\mathrm{pil}} = \frac{\mathbf{1}\{A_{1,u} = 1\}}{\pi_u^{(1)}(1 \mid X_{1,u})} \big(e_u - h_{1,u}^{(1)}\big) - \frac{\mathbf{1}\{A_{1,u} = 0\}}{\pi_u^{(1)}(0 \mid X_{1,u})} \big(e_u - h_{1,u}^{(0)}\big) + h_{1,u}^{(1)} - h_{1,u}^{(0)}.$$

Hence

$$\bar{d}_1 := \frac{1}{n} \sum_{u=1}^{n} d_u^{\mathrm{pil}}, \qquad v_n = \Phi_{\mathcal{Y},1} \bar{d}_1, \qquad \|v_n\|_{\mathcal{H}_{\mathcal{Y}}}^2 = \bar{d}_1^\top K_{\mathcal{Y}}^{(1,1)} \bar{d}_1.$$

If $v_n \neq 0$, define the witness coefficient vector

$$c^{\mathrm{wit}} := \frac{\bar{d}_1}{\sqrt{\bar{d}_1^\top K_{\mathcal{Y}}^{(1,1)} \bar{d}_1}}, \qquad w_n = \Phi_{\mathcal{Y},1} c^{\mathrm{wit}}.$$

Then the projected scalar outcomes on fold 2 are

$$z := K_{\mathcal{Y}}^{(2,1)} c^{\mathrm{wit}} \in \mathbb{R}^n, \qquad z_u = \langle w_n, \varphi_{\mathcal{Y}}(Y_{2,u}) \rangle_{\mathcal{H}_{\mathcal{Y}}}.$$

If $v_n = 0$, replace w_n by the fixed tie-breaker u_0 and compute $z_u = \langle u_0, \varphi_{\mathcal{Y}}(Y_{2,u}) \rangle_{\mathcal{H}_{\mathcal{Y}}}$.

Fold-2 projected scalar nuisances. After projection, the inferential fold is scalar. Using the same past-only KRR vectors $h_{2,u}^{(0)}, h_{2,u}^{(1)}$, define the scalar nuisance evaluations

$$m_u^{(b)} := z^\top h_{2,u}^{(b)} = \langle w_n, \widehat{\mu}_{u-1}^{(2)}(b, X_{2,u}) \rangle_{\mathcal{H}_{\mathcal{Y}}}, \qquad b \in \{0,1\}.$$

Current projected score. The projected DR score at inferential local time $u \in \{1, \dots, n\}$ is

$$\xi_u = z^\top c_u,$$

where

$$c_u = \frac{\mathbf{1}\{A_{2,u} = 1\}}{\pi_u^{(2)}(1 \mid X_{2,u})} \big(e_u - h_{2,u}^{(1)}\big) - \frac{\mathbf{1}\{A_{2,u} = 0\}}{\pi_u^{(2)}(0 \mid X_{2,u})} \big(e_u - h_{2,u}^{(0)}\big) + h_{2,u}^{(1)} - h_{2,u}^{(0)}.$$

Equivalently,

$$\xi_u = \frac{\mathbf{1}\{A_{2,u} = 1\}}{\pi_u^{(2)}(1 \mid X_{2,u})} \big(z_u - m_u^{(1)}\big) - \frac{\mathbf{1}\{A_{2,u} = 0\}}{\pi_u^{(2)}(0 \mid X_{2,u})} \big(z_u - m_u^{(0)}\big) + m_u^{(1)} - m_u^{(0)}.$$

History-recycled projected scores. For $1 \leq s < t \leq n$, the historical nuisance at time s is kept fixed, while only the evaluation-time propensity is changed to $\pi_t^{(2)}$. Define

$$\check{\xi}_{s,t} = z^\top c_{s,t},$$

where

$$c_{s,t} = \frac{\mathbf{1}\{A_{2,s} = 1\}}{\pi_t^{(2)}(1 \mid X_{2,s})} \big(e_s - h_{2,s}^{(1)}\big) - \frac{\mathbf{1}\{A_{2,s} = 0\}}{\pi_t^{(2)}(0 \mid X_{2,s})} \big(e_s - h_{2,s}^{(0)}\big) + h_{2,s}^{(1)} - h_{2,s}^{(0)}.$$

Equivalently,

$$\check{\xi}_{s,t} = \frac{\mathbf{1}\{A_{2,s} = 1\}}{\pi_t^{(2)}(1 \mid X_{2,s})} \big(z_s - m_s^{(1)}\big) - \frac{\mathbf{1}\{A_{2,s} = 0\}}{\pi_t^{(2)}(0 \mid X_{2,s})} \big(z_s - m_s^{(0)}\big) + m_s^{(1)} - m_s^{(0)}.$$

Sequential conditional variance estimator. For $1 \leq s < t \leq n$, define the importance ratio

$$\rho_{s,t} := \frac{\pi_t^{(2)}(A_{2,s} \mid X_{2,s})}{\pi_s^{(2)}(A_{2,s} \mid X_{2,s})}.$$

For a fixed inferential time $t \geq 2$, collect the recycled score coefficients into

$$C_t := \big[c_{1,t}, \dots, c_{t-1,t}\big] \in \mathbb{R}^{n \times (t-1)}, \qquad \rho_t := (\rho_{1,t}, \dots, \rho_{t-1,t})^\top.$$

Then

$$\check{\xi}_t := (\check{\xi}_{1,t}, \dots, \check{\xi}_{t-1,t})^\top = C_t^\top z,$$

and the sequential moments are

$$\widehat{M}_{1,t} = \frac{1}{t-1} \rho_t^\top C_t^\top z, \qquad \widehat{M}_{2,t} = \frac{1}{t-1} \rho_t^\top \big(C_t^\top z\big)^{\odot 2}, \qquad t \geq 2.$$

Set the initial convention

$$\widehat{\sigma}_1^2 := 1.$$

For $t \geq 2$, define

$$\widehat{\sigma}_t^2 = \Big(\widehat{M}_{2,t} - \widehat{M}_{1,t}^2\Big)_+, \qquad \widetilde{\sigma}_t = \sqrt{\widehat{\sigma}_t^2} \vee \epsilon_n.$$

For $t = 1$, set $\widetilde{\sigma}_1 := 1 \vee \epsilon_n$.

PADR-KTE statistic. The projected adaptive DR-KTE statistic is

$$T_{\text{PADR-KTE}} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widetilde{\sigma}_t^{-1} \, \xi_t.$$

Efficient evaluation. The vectors $h_{1,u}^{(b)}$ and $h_{2,u}^{(b)}$ are computed once from covariate Gram blocks. The pilot fold enters the inferential fold only through the witness coefficients c^{wit} and the projected outcome vector $z = K_{\mathcal{Y}}^{(2,1)} c^{\text{wit}}$. After that point, all inferential computations are scalar: only the evaluation-time denominators $\pi_t^{(2)}(b \mid X_{2,s})$ and the ratios $\rho_{s,t}$ vary with t; all historical smoother vectors $h_{2,s}^{(b)}$ are reused.

H SUPPLEMENTARY ON NUMERICAL SIMULATIONS

This section provides a detailed supplement to the numerical simulations presented in Section 7. We first specify the kernel function leveraged in our method. Following this, we discuss the baseline algorithms against which our approach was compared, and conclude by detailing additional experimental setups and presenting supplementary numerical results.

H.1 Kernel

In our experiments, we employed the Gaussian kernel (also known as the Radial Basis Function or RBF kernel), defined for all $h_i, h_j \in \mathbb{R}^{d_{\mathcal{H}}}$ as:

$$k_{\mathcal{H}}(h_i, h_j) = \exp\left(-\frac{\|h_i - h_j\|_2^2}{2\gamma^2}\right).$$

The parameter γ is the length-scale of the kernel, which controls the smoothness of the resulting function space. The Gaussian kernel is widely used in practice and satisfies the crucial properties of boundedness, continuity, and characteristicity (Sriperumbudur et al., 2010). For both the covariate space $\mathcal{X}$ and the outcome space $\mathcal{Y}$, we utilized the Gaussian kernel, setting the length-scales based on the median of the pairwise Euclidean distances from the given data. Specifically, for a dataset $\{h_i\}_{i=1}^T$, the median pairwise distance is given by

$$\gamma_{\text{median}} = \text{median}\{\|h_i - h_j\|_2 \mid 1 \le i < j \le T\}.$$

In particular, we chose the length-scale for the covariate kernel ($k_{\mathcal{X}}$) to be equal to the median pairwise distance, and for the outcome kernel ($k_{\mathcal{Y}}$), we set the length-scale to be one half of the calculated median distance.

H.2 Baselines

(i) CADR (Contextual Adaptive Doubly Robust): CADR is a *stabilized DR* estimator specifically designed for data that is both contextual (dependent on covariates X) and adaptively collected (where the data collection process changes over time). The estimator operates by forming a canonical gradient $D'(g_t, \bar{Q}_{t-1})(X_t, A_t, Y_t)$—a term that incorporates both the policy and an outcome model. This gradient is then aggregated across time using history-measurable inverse standard-deviation weights, $\hat{\sigma}_t^{-1}$. The components are defined as follows:

- g_t is the logging policy at time t.
- $\bar{Q}_t : \mathcal{A} \times \mathcal{X} \to \mathcal{Y}$: An estimate of the Conditional Outcome Model $\mathbb{E}[Y \mid A = \cdot, X = \cdot]$. Crucially, for every t, $\bar{Q}_t$ is trained using only data observed up to time t.
- $\hat{\sigma}_t^{-1}$: The inverse of $\hat{\sigma}_t$, which estimates the conditional standard deviation $\sigma_{0,t} = \text{Var}\big(D'(g_t, \bar{Q}_{t-1})(O_t) \mid O_{1:t-1}\big)^{1/2}$. These weights stabilize the variance of the overall estimate.
- O_t: The set of observed variables at time t, $O_t = (X_t, A_t, Y_t)$, and $O_{1:t-1} = (O(1), \ldots, O(t-1))$.

The stabilized estimate is constructed as

$$\widehat{\Psi}_T = \Big(\tfrac{1}{T}\sum_{t=1}^{T}\hat{\sigma}_t^{-1}\Big)^{-1} \cdot \tfrac{1}{T}\sum_{t=1}^{T}\hat{\sigma}_t^{-1}\, D'\big(g_t, \bar{Q}_{t-1}\big)(O_t),$$

with asymptotic normality under consistency of the *conditional* standard-deviation estimators $\hat{\sigma}_t$ (each trained on past data only) and a mild exploration condition ($g_t(a \mid x) \gtrsim t^{-1/2}$); see (Bibaut et al., 2021, Algorithm 1; Theorem 1; Section 3).[4] We implement CADR exactly as specified with fold-wise, predictable nuisance fits and $\hat{\sigma}_t$ built from past data only.

(ii) Variance-stabilized AIPW of Hadad et al. (2021). Hadad et al. (2021) propose an *adaptively-weighted AIPW* family for non-contextual adaptive experiments that ensures martingale variance convergence via *variance-stabilizing weights*. Let Γ_t denote the (A)IPW score for a fixed arm and e_t its propensity. Weights $\{h_t\}$ are chosen so that $\sum_t h_t^2/e_t$ is *deterministic* (stick-breaking), which yields a studentized statistic with a standard normal limit. Two named allocation schemes are: *constant allocation* $\lambda_t^{\text{const}} = \frac{1}{T-t+1}$ (giving $h_t \propto \sqrt{e_t/T}$), and the *two-point allocation* $\lambda_t^{\text{two-point}}$ that interpolates between high-propensity and vanishing-propensity regimes using a heuristic for future propensities; both satisfy the sufficient bounds of their Theorem 3. We implement this baseline as **AW-AIPW (Hadad)** with both `constant` and `two_point` allocation options, and with AIPW scores; see (Hadad et al., 2021, Section 2.2–2.3; Theorem 2–3; Equation (12)–(18)).

H.3 Additional description of the experiments

In this Appendix, we provide additional details and descriptions for the experiments in our main text.

H.3.1 Synthetic data

All data (covariates, treatments, responses) is simulated. Each round draws a context $X_t \in \mathbb{R}^5$ i.i.d. from $\mathcal{N}(0, I_5)$. We consider three cases for the underlying function f that generates the potential outcome:

(i) *cosine model* with $f(x) = \cos(\beta^\top x)$ and $\beta = (0.1, 0.2, 0.3, 0.4, 0.5)$;

(ii) *linear model* with $f(x) = \beta^\top x$ and the same β; and

(iii) *sigmoidal model* with $f(x) = \sigma(\beta^\top x)$ where $\sigma(z) = ln(|16z - 8| + 1) \cdot \text{sign}(z - 0.5)$ and the same β.

Then, potential outcomes are generated as $Y_t(0) = f(X_t) + \varepsilon_t$ and $Y_t(1) = f(X_t) + \delta_t + \varepsilon_t$, with i.i.d. noise $\varepsilon_t \sim \mathcal{N}(0, 0.5)$.

Scenarios. We use the four scenarios of Martinez Taboada et al. (2023) through the treatment effect δ_t: Scenario I (null) uses $\delta_t = 0$; Scenario II (mean shift) uses $\delta_t = 2$; Scenario III (symmetric mixture) uses $\delta_t = 2\,S_t$ with $S_t \in \{-1, +1\}$ Rademacher(0.5); Scenario IV (random scale) uses $\delta_t \sim \text{Uniform}[-4, 4]$. These match the no-effect, constant-mean, symmetric mixture, and random-scale shifts respectively with exact constant values in (Martinez Taboada et al., 2023).

Adaptive data collection (two arms, ε-greedy with online ridge). Each arm $a \in \{0, 1\}$ maintains an online ridge model for the potential outcome $Y_t(a)$ based on an augmented design vector $x_t^{\text{aug}} = (1, X_t)$ that includes an unpenalized intercept. The ridge state for each arm is a pair (S_a, b_a), where $S_a \in \mathbb{R}^{6\times 6}$ is initialized as

$$S_a = \text{diag}(0, \lambda, \ldots, \lambda), \qquad b_a = 0,$$

with $\lambda = 10^{-2}$ applied to the $d = 5$ non-bias coordinates. At each round t, the current model parameters are updated by solving the linear system

$$S_a \theta_a = b_a, \qquad a \in \{0, 1\},$$

[4] CADR constructs $\hat{\sigma}_t^2$ via importance-reweighting across past policies g_s using ratios g_t/g_s and proves almost-sure consistency of $\hat{\sigma}_t^2$ under a bracketing-entropy bound on the logging policy class and a rate for the outcome-regression sequence $\bar{Q}_t$.

yielding the estimated regression weights θ_a. The predicted rewards are

$$q_a(t) = \langle \theta_a, x_t^{\text{aug}} \rangle, \qquad a \in \{0, 1\}.$$

The exploration probability decays with time according to

$$\varepsilon_t = \max\big(\varepsilon_{\min}, \varepsilon_0/(t+1)^p\big), \quad \text{with } \varepsilon_0 = 0.2,\ \varepsilon_{\min} = 0.05,\ p = 0.99.$$

Given $(q_0(t), q_1(t))$, the ε-greedy decision rule defines the logging propensities as

$$\pi_t(1 \mid X_t) = \begin{cases} 1 - \frac{1}{2}\varepsilon_t, & q_1(t) > q_0(t), \\ \frac{1}{2}\varepsilon_t, & q_1(t) < q_0(t), \\ 0.5, & q_1(t) = q_0(t), \end{cases} \qquad \pi_t(0 \mid X_t) = 1 - \pi_t(1 \mid X_t).$$

An action $A_t \in \{0, 1\}$ is then sampled according to these propensities, and the observed reward is $Y_t = Y_t(A_t)$. The scalar weight used in subsequent estimators is the realized propensity,

$$w_t = \begin{cases} \pi_t(1 \mid X_t), & A_t = 1, \\ \pi_t(0 \mid X_t), & A_t = 0. \end{cases}$$

After observing (X_t, A_t, Y_t), only the chosen arm's ridge state is updated as

$$S_{A_t} \leftarrow S_{A_t} + x_t^{\text{aug}} (x_t^{\text{aug}})^\top, \qquad b_{A_t} \leftarrow b_{A_t} + x_t^{\text{aug}} Y_t.$$

This sequential rule generates a non-i.i.d. adaptive trajectory with time-varying propensities $\pi_t(1 \mid X_t)$ that progressively concentrate as the regression parameters stabilize.

Propensity matrices for foldwise evaluation. For test statistics that require foldwise policy-on-fold propensities, we snapshot θ_a over time to build matrices that map each decision time to propensities evaluated on all contexts within the same fold. Concretely, we split the trajectory into two non-adaptive folds using the default *alternating* split (odd vs. even indices, chronological within each). For each fold r and each in-fold time t, we compute $\pi_t(1 \mid X_s)$ for all in-fold contexts X_s using the θ_a snapshot at time t, yielding dense $|\mathcal{I}_r| \times |\mathcal{I}_r|$ propensity matrices per fold (with the same greedy/non-greedy/tie rule as above). These matrices, together with the realized w_t, are passed to the test procedures.

Kernels and run lengths. Outcome similarities use an RBF kernel with bandwidth set as $\gamma = 1/\sigma^2$ (i.e., $\gamma = 2.0$ when $\sigma^2 = 0.5$), unless otherwise stated. Each experiment uses a trajectory length $T = 1000$ and we run 200 Monte-Carlo replications per configuration. All other defaults follow the description above.

H.3.2 IHDP data

To evaluate our proposed method on a real-world benchmark, we generate a semi-synthetic dataset based on the Infant Health and Development Program (IHDP) data (Hill, 2011). The original IHDP data originates from a randomized experiment on the effects of specialist home visits on cognitive test scores.

Following the preprocessing steps used in (Martinez Taboada et al., 2023), we retain 908 samples with 18 covariates (9 continuous, 9 categorical), resulting in $X_t \in \mathbb{R}^{18}$ for all t. We synthesize the adaptive policies π_t with two arms, using an ϵ-greedy with online ridge regression. This policy structure is identical to the one discussed in the preceding section, and it results in binary treatments, $A_t \in \{0, 1\}$.

The potential outcomes are generated according to the following equations:

$$Y_t(0) = \cos(\beta^\top X_t) + \epsilon_t, \quad Y_t(1) = \cos(\beta^\top X_t) + \delta_t + \epsilon_t.$$

Here, the term δ_t is used to control the treatment effect, defining four different experimental scenarios. The noise term $\epsilon_t \sim \mathcal{N}(0, 0.5)$ is an i.i.d. Gaussian random variable with zero mean and variance 0.5, i.e., $\epsilon_t \sim \mathcal{N}(0, 0.5)$.

Figure 4: Observational samples from the dSprite data in Scenario IV

Scenarios. We utilize the same four scenarios, that we adapted for the synthetic data experiments from Martinez Taboada et al. (2023), by defining the treatment effect term δ_t: (i) Scenario I (Null): The treatment has no effect, defined by $\delta_t = 0$; (ii) Scenario II (Mean Shift): The treatment introduces a constant positive shift, defined by $\delta_t = 2$; (iii) Scenario III (Symmetric Mixture): The treatment effect is a symmetric mixture, defined by $\delta_t = 2\,S_t$ with $S_t \in \{-1, +1\}$ Rademacher(0.5); (iv) Scenario IV (Random Scale): The treatment effect is randomly scaled, defined by $\delta_t \sim \text{Uniform}[-4, 4]$.

Evaluation protocol. We evaluated our method's performance across varying sample sizes. This was done by running experiments on the IHDP dataset using subsampling without replacement, where the subset size was varied uniformly within the set $\{100, 150, 200, \ldots, 850, 900, 908\}$, with 908 representing the full available dataset. We utilized the non-adaptive alternating fold splitting protocol, consistent with our synthetic dataset experiments, and ran each distinct experiment over 200 Monte-Carlo replications.

For the Gaussian kernels used, we followed a median heuristic: the length-scale for the covariate kernel was set equal to the median pairwise distance, while the length-scale for the outcome kernel was set to one half of that median distance. The regularization parameter λ was set to 10^{-2}.

The true positive rates for Scenarios II-IV, utilizing the full available dataset, are presented in Table 1. A separate discussion detailing additional results that incorporate varying data sizes is provided in Section H.4.

H.3.3 dSprite dataset

We adapt the structured image benchmark of Xu and Gretton (2023) and adapt it to the two-scenario setting of our adaptive kernel test. Each outcome $Y \in [0,1]^{64\times 64}$ is a grayscale image of a heart shape on a black background, rendered from latent coordinates $(\text{posX}, \text{posY}) \in [0,1]^2$. Contexts $X_t = (x_t^{(1)}, x_t^{(2)})$ are sampled uniformly from $\text{Unif}([0,1]^2)$, and images are generated through a deterministic renderer

$$Y_t(a) = g(X_t, a) \in [0,1]^{64\times 64},$$

where $a \in \{0,1\}$ indexes the treatment and g draws a white heart centered at position $(x_t^{(1)} + \Delta_a^{(1)},\, x_t^{(2)} + \Delta_a^{(2)})$ with fixed scale and rotation. The offsets $(\Delta_a^{(1)}, \Delta_a^{(2)})$ define the two experimental regimes:

$$\begin{aligned} \text{Scenario I (null):} \quad & (\Delta_0^{(1)}, \Delta_0^{(2)}) = (0,0), \quad (\Delta_1^{(1)}, \Delta_1^{(2)}) = (0,0); \\ \text{Scenario IV (shift):} \quad & (\Delta_0^{(1)}, \Delta_0^{(2)}) = (0,0), \quad (\Delta_1^{(1)}, \Delta_1^{(2)}) = (\delta,0), \end{aligned}$$

where $\delta = 0.15$ induces a rightward translation of the heart under $A = 1$ while preserving mean pixel intensity. Gaussian pixel noise $\mathcal{N}(0, 0.01)$ is added to each image. Hence, the marginal intensity distributions of $Y(0)$ and $Y(1)$ coincide, but their spatial structure differs. Figure 4 shows observational samples generated under Scenario IV, where the adaptive policy produces trajectories with spatially translated outcomes. Figure 5 depicts corresponding counterfactual image pairs $(Y(0), Y(1))$, confirming that the treatment $A = 1$ only shifts the heart horizontally without altering overall brightness or shape.

Adaptive data collection. Logged trajectories $\{(X_t, A_t, Y_t)\}_{t=1}^T$ are generated by an ε-greedy contextual policy with two arms and per-arm online ridge regression, identical to the adaptive linear setting in §H.3.1. Each arm $a \in \{0,1\}$ maintains the sufficient statistics

$$S_a = \text{diag}(0, \lambda, \ldots, \lambda), \qquad b_a = 0,$$

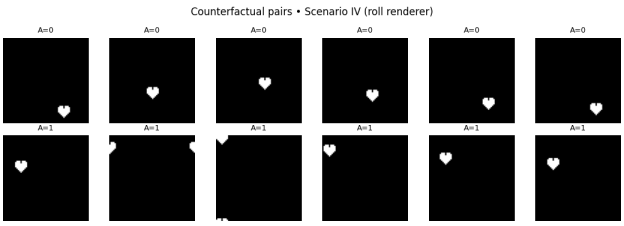

Figure 5: Counterfactual pairs from the dSprite data in Scenario IV

with $\lambda = 10^{-2}$ and features $x_t^{\text{aug}} = (1, X_t) \in \mathbb{R}^3$. At each round t, the arm parameters $\theta_a = S_a^{-1} b_a$ yield predictions $q_a(t) = \langle \theta_a, x_t^{\text{aug}} \rangle$. The exploration rate follows

$$\varepsilon_t = \max(\varepsilon_{\min},\, \varepsilon_0/(t+1)^p), \qquad \varepsilon_0 = 0.2,\ \varepsilon_{\min} = 0.05,\ p = 0.99.$$

Actions are sampled according to

$$\pi_t(1|X_t) = \begin{cases} 1 - \frac{1}{2}\varepsilon_t, & q_1(t) > q_0(t), \\ \frac{1}{2}\varepsilon_t, & q_1(t) < q_0(t), \\ 0.5, & q_1(t) = q_0(t), \end{cases} \qquad \pi_t(0|X_t) = 1 - \pi_t(1|X_t).$$

After observing (X_t, A_t, Y_t), only the chosen arm is updated:

$$S_{A_t} \leftarrow S_{A_t} + x_t^{\text{aug}} (x_t^{\text{aug}})^\top, \qquad b_{A_t} \leftarrow b_{A_t} + x_t^{\text{aug}} Y_t.$$

The sequence $\{\pi_t(1|X_t)\}$ is stored to compute the stabilized kernel test statistics.

Foldwise evaluation. To enable cross-fold variance stabilization, we use an *alternating* split $(\mathcal{I}_0, \mathcal{I}_1)$ and record fold-specific propensity matrices $\Pi_{r \leftarrow r}$ computed from the parameter snapshots $\{\theta_a^{(t)}\}_{t \in \mathcal{I}_r}$. Each matrix encodes, for every evaluation time t in a fold, the propensities $\pi_t(A_s|X_s)$ for all contexts s within the same fold.

Evaluation protocol. Each experiment runs for $T = 1000$ adaptive rounds and is repeated over 200 Monte-Carlo replications. For each test, empirical type-I error is the proportion of rejections at level 0.05 under Scenario I, and empirical power is the proportion of rejections under Scenario IV. All tests use a Gaussian RBF kernel on outcomes with bandwidth chosen by the median heuristic and $\lambda = 10^{-2}$ regularization. ADR-KTE operates directly on flattened images $Y_t \in \mathbb{R}^{4096}$, while baseline methods (CADR, AW-AIPW) are restricted to the mean pixel intensity as scalar outcome.

H.4 Additional results

Synthetic dataset. To complete the presentation of our synthetic dataset experiments, this section provides the comparative results for our proposed method and the baseline algorithms under two alternative potential outcome generating functions: the linear model and the sigmoidal model, both discussed in Section H.3.1.

- **Linear model results:** The calibration of our proposed method, ADR-KTE, in the linear case (Scenario I) is demonstrated in Figure 6. The collected metrics—including the empirical histogram, Q-Q plot, and false positive rate across varying data sizes—collectively confirm that our method is well-calibrated.

  Figure 7 provides the comparison of ADR-KTE with the baselines CADR and AW-AIPW across Scenarios II-IV. Consistent with our preceding findings, the baselines achieve matching performance in Scenario II (mean shift) and even show slightly better results in the small data size regime. Crucially, however, our method significantly outperforms the baselines in scenarios characterized by purely distributional changes with an identical mean (Scenarios III and IV).

- **Sigmoidal model results:** The findings for the sigmoidal case similarly mirror these results. The calibration of ADR-KTE in Scenario I is shown in Figure 8, while the comparative power results across Scenarios II-IV are displayed in Figure 9. In both model structures, our method maintains its superior power in detecting distributional differences where mean-based methods fail.

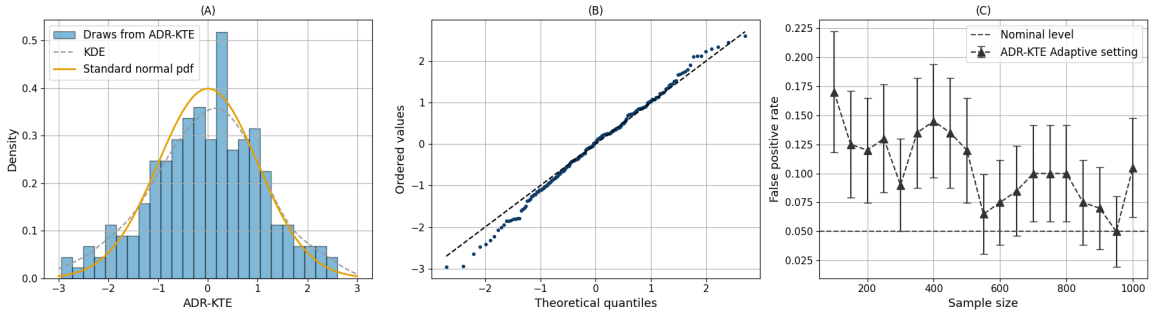

Figure 6: Calibration of ADR-KTE under the null hypothesis (Scenario I) in the adaptive setting for the linear model (based on 200 simulations). (A): Empirical histogram vs. standard normal PDF ($T = 900$); (B): Normal Q-Q plot; (C): False Positive Rate across sample sizes. The results confirm approximate Gaussian asymptotics and controlled type-I error.

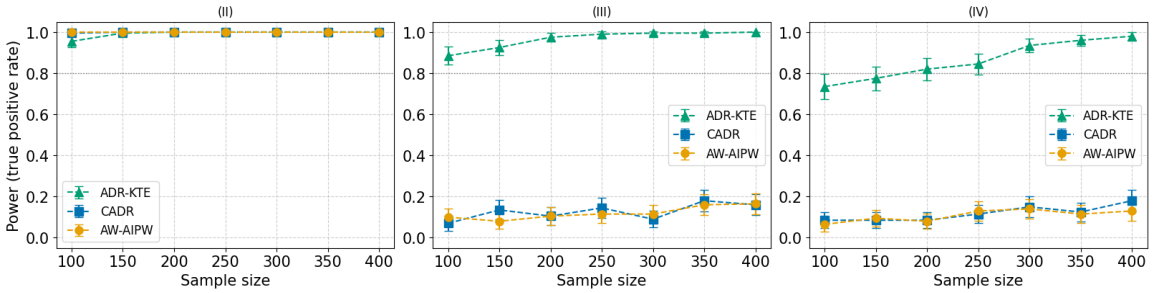

Figure 7: Power comparison (true positive rates) for the linear model across Scenarios II–IV, based on 200 simulations. Mean-focused baselines (CADR/AW-AIPW) achieve matching power on Scenario II (mean shift). ADR-KTE demonstrates markedly higher power in detecting higher-moment shifts (Scenarios III–IV).

IHDP dataset: We now present the results from the numerical simulations conducted on the IHDP dataset, focusing on the method's performance across varying sample sizes.

Figure 10 illustrates the calibration of our proposed method under the null hypothesis (Scenario I), based on 200 Monte-Carlo runs. This figure presents the histogram of test statistics, the Q-Q plot, and the type-I error across varying sample sizes.

The power of our method in comparison with the baselines for Scenarios II-IV is demonstrated across varying data sizes in Figure 11. These results show that, in particular, our method exhibits a significant advantage in power for detecting distributional effects, in contrast to the mean-focused baselines.

H.5 Comparison and Discussion with Non-adaptive Kernel Baselines

To rigorously assess the impact of data adaptivity on standard kernel methods, we benchmarked our ADR-KTE test against two prominent estimators designed for i.i.d. data: DR-xKTE (Martinez Taboada et al., 2023) and the standard KTE (Muandet et al., 2021). We evaluate these across $T \in \{100, 150, \ldots, 500\}$ to demonstrate the stability of our approach under varying sample sizes.

Data Generating Process. For both i.i.d. and adaptive settings, we utilize a sigmoidal baseline function $h(x) = \mathrm{sgn}(x - 0.5)\log(|16x - 8| + 1)$ to generate the outcome base. The potential outcomes are then defined as $Y_t(0) = h(X_t^\top \beta) + \varepsilon_t$ and $Y_t(1) = h(X_t^\top \beta) + \Delta_t + \varepsilon_t$, where $\beta = [0.1, 0.2, 0.3, 0.4, 0.5]^\top$ and $\varepsilon_t \sim \mathcal{N}(0, 0.5)$. In the i.i.d. case, actions are sampled from a fixed logistic policy $\pi(1|x) = \mathrm{logit}^{-1}(x^\top \beta)$. In the adaptive case, we employ the ε-greedy contextual bandit described in Section 7.1, which updates its arm parameters and exploration rate ε_t at every time step t.

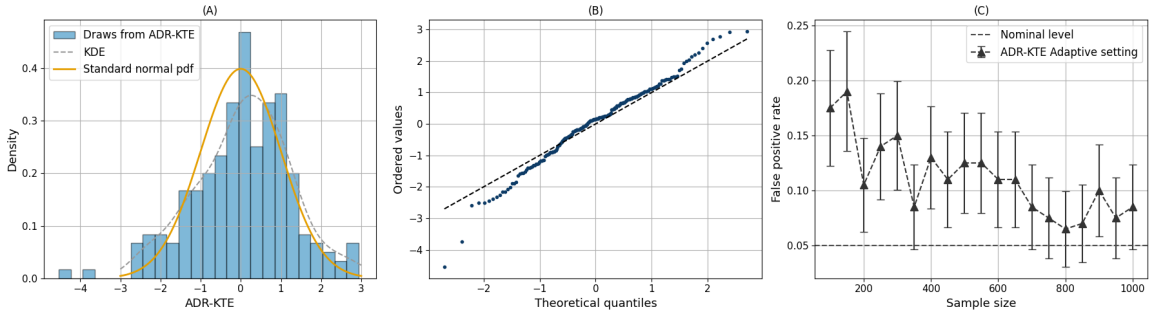

Figure 8: Demonstration of the Calibration of ADR-KTE in the adaptive setting for the sigmoidal model under the null hypothesis (Scenario I), based on 200 replications. (A): Histogram of test statistics compared to the standard normal PDF (shown for $T = 850$); (B): Normal Q-Q plot; (C): type-I error (False Positive Rate) evolution across sample sizes.

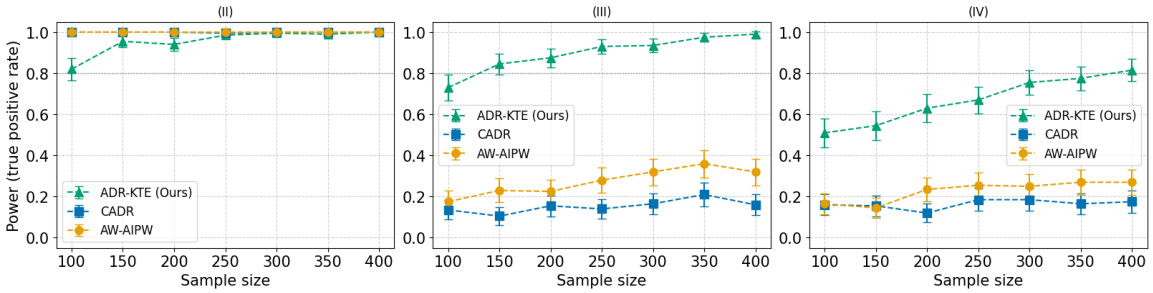

Figure 9: Comparative Power results (true positive rates) for the sigmoidal model across Scenarios II–IV, using 200 Monte-Carlo runs. Baselines focused on mean effects (CADR/AW-AIPW) achieve matching performance for the mean shift in Scenario II. In contrast, ADR-KTE displays a significantly greater ability to detect distributional differences characterized by higher-moment shifts (Scenarios III–IV).

Case 1: Verification under i.i.d. Sampling We first verify that all estimators are correctly calibrated under a standard i.i.d. sampling scheme. As shown in Table 3, all procedures exhibit nominal type-I error under the null (Scenario I) and full power under the alternative (Scenario II). This confirms that in the absence of adaptivity, our proposed ADR-KTE maintains the standard effectiveness of i.i.d. methods.

Case 2: Evaluation under Stable Adaptive Data Collection. We next evaluate the methods under the ε-greedy adaptive policy. In contrast to highly anisotropic exploration rules, this policy maintains a uniform exploration floor, so the empirical action-feature covariance is naturally viewed as operating in a stable adaptive regime, and in our setting it is reasonable to regard it as satisfying the stronger full-matrix stability condition (Lai and Wei, 1982). Recent theory shows that, under such stability conditions, estimators that are efficient in the i.i.d. setting remain asymptotically normal and efficient under adaptive sampling for scalar pathwise differentiable targets (Shen et al., 2026). Although that result is stated for scalar targets, the underlying mechanism is the stabilization of the predictable quadratic variation, which suggests that an analogous phenomenon should also hold for Hilbert-valued and RKHS-based scores.

Table 4 is broadly consistent with this picture, while also showing that finite-sample behavior remains method-dependent. DR-xKTE is clearly anti-conservative under the null at small and moderate horizons, so its large rejection rates under the alternative are not directly interpretable as reliable power. KTE stays roughly calibrated, but suffers a substantial loss of power throughout. In contrast, ADR-KTE delivers by far the strongest overall performance: it attains high power across all horizons, and its null rejection rate decreases steadily toward the nominal level as T grows. Overall, these results support the view that our ε-greedy design lies in a stable adaptive regime, and that explicit variance control remains important in practice for RKHS-based inference, even when the underlying adaptive scheme is asymptotically well-behaved.

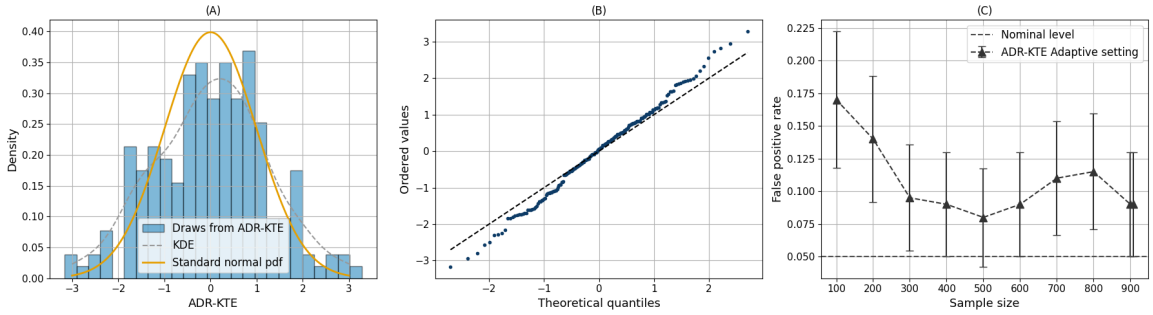

Figure 10: Assessment of the Calibration of ADR-KTE under the null hypothesis (Scenario I) in the adaptive setting, using the IHDP dataset (200 replications). (A): Distribution of test statistics (histogram versus standard normal PDF, shown for the full sample size $T = 908$); (B): Normal Q-Q plot; (C): type-I error (False Positive Rate) control across varying sample sizes.

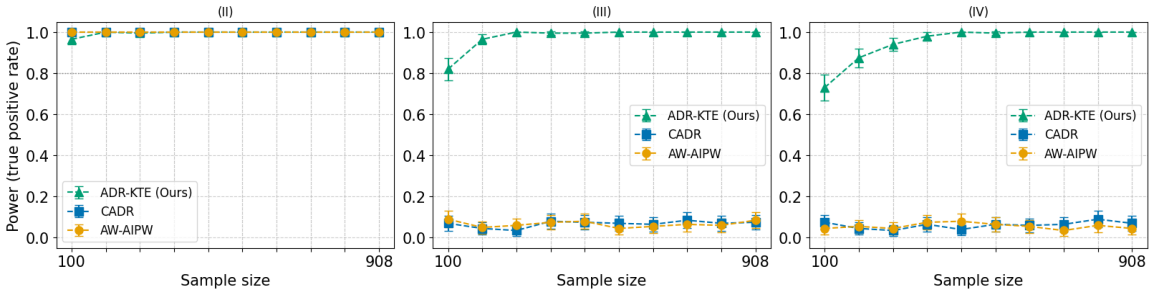

Figure 11: Comparative Power Analysis (true positive rates) for the IHDP dataset across Scenarios II–IV, based on 200 Monte-Carlo runs. The mean-focused baselines (CADR/AW-AIPW) show matching detection capability for the pure mean shift in Scenario II. Conversely, ADR-KTE exhibits a substantially improved power profile for identifying distributional disparities stemming from higher-moment changes (Scenarios III–IV).

H.6 Computation infrastructure

We ran our experiments on local CPUs of desktops and on a GPU-enabled node (in a remote server) with the following specifications:

- **Operating System:** Linux (kernel version 6.8.0-55-generic)
- **GPU:** NVIDIA RTX A4500
  - Driver Version: 560.35.05
  - CUDA Version: 12.6
  - Memory: 20 GB GDDR6

Table 3: Rejection rates (mean ± se) under an i.i.d. sampling scheme for Scenario I (Null) and Scenario II (Alternative). All methods are correctly calibrated and achieve full power when data collection is not adaptive.

Scenario I (Null)	100	150	200	250	300	350	400	450	500
DR-xKTE	0.085	0.105	0.050	0.045	0.075	0.065	0.055	0.075	0.055
KTE	0.050	0.010	0.025	0.055	0.055	0.070	0.040	0.025	0.040
ADR-KTE	0.085	0.040	0.090	0.085	0.050	0.080	0.085	0.070	0.065
Scenario II (Alt)	100	150	200	250	300	350	400	450	500
DR-xKTE	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
KTE	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ADR-KTE	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4: Rejection rates under the ε-greedy adaptive policy. Because this policy maintains persistent exploration, the resulting adaptive design is naturally viewed as stable, plausibly in the stronger full-matrix sense. The table shows that finite-sample performance nevertheless depends strongly on the test statistic: DR-xKTE is anti-conservative under the null, KTE is roughly calibrated but markedly underpowered, and ADR-KTE provides the best calibration-power tradeoff, with null rejection rates moving toward the nominal level as T increases while power remains high.

Scenario I (Null)	100	150	200	250	300	350	400	450	500
DR-xKTE	0.165	0.150	0.125	0.100	0.080	0.090	0.105	0.095	0.075
KTE	0.065	0.055	0.055	0.035	0.040	0.045	0.060	0.035	0.035
ADR-KTE	0.140	0.125	0.125	0.120	0.145	0.115	0.065	0.095	0.060
Scenario II (Alt)	100	150	200	250	300	350	400	450	500
DR-xKTE	0.960	1.000	0.985	0.995	1.000	1.000	1.000	1.000	1.000
KTE	0.260	0.200	0.220	0.180	0.155	0.225	0.145	0.190	0.245
ADR-KTE	0.885	0.945	0.950	0.955	0.990	0.995	1.000	1.000	1.000