

Appendix

This appendix is organized as follows:

- Appendix 8: summary of the notations used in the analysis.
- Appendix 9: a review of counterfactual mean embeddings that are instrumental in Section 2.
- Appendix 10: proof for the asymptotic analysis of the plug-in estimator presented in Section 3.
- Appendix 11: contains further details on the efficient influence function of our counterfactual policy mean embedding and the associated estimator presented in Section 4.
- Appendix 12: provides the analysis of the doubly robust kernel test of the distributional policy effect presented in Section 5.1.
- Appendix 13: does the same for the sampling algorithm presented in Section 5.2.
- Appendix 14: details on the implementation of the algorithms and additional experiment details, discussions and results.

All the code to reproduce our numerical simulations is provided in the supplementary material and will be open-sourced upon acceptance of the manuscript.

8 Notations

In this appendix, we recall for clarity some useful notations that are used throughout the paper.

Notations for distributional off-policy evaluation setting and finite samples

- y_i, a_i, x_i are realizations of the outcome, action, and context random variables Y, A, X for $i \in \{1, \dots, n\}$. Potential outcomes are written $\{Y(a)\}_{a \in \mathcal{A}}$.
- The distribution on the context space is written P_X , the distribution on outcomes is conditional to actions and contexts and is written $P_{Y|X,A}$. Distributions on actions A are policies π belonging to a set Π . In the logged dataset, actions are drawn from a logging policy π_0 . Resulting triplet distribution is written $P_\pi = P_{Y|X,A} \times \pi \times P_X$.
- The distribution $\nu(\pi)$ represents the marginal distribution of outcomes over $\pi \times P_X$.

Notations related to the kernel-based representations used to embed counterfactual outcome distributions

- $\mathcal{H}_{\mathcal{F}}$ is a generic RKHS associated with a domain \mathcal{F} .
- $\mathcal{H}_{\mathcal{AX}}$: RKHS on $\mathcal{A} \times \mathcal{X}$ with kernel $k_{\mathcal{AX}}$ and feature map $\phi_{\mathcal{AX}}(a, x) = k_{\mathcal{AX}}(\cdot, (a, x))$. Inner product: $\langle \cdot, \cdot \rangle_{\mathcal{H}_{\mathcal{AX}}}$.
- $\mathcal{H}_{\mathcal{Y}}$: RKHS on \mathcal{Y} with kernel $k_{\mathcal{Y}}$ and feature map $\phi_{\mathcal{Y}}(y) = k_{\mathcal{Y}}(\cdot, y)$. Inner product: $\langle \cdot, \cdot \rangle_{\mathcal{H}_{\mathcal{Y}}}$.
- Given a distribution P over \mathcal{F} , the kernel mean embedding is $\mu_P = \mathbb{E}_P[\phi_{\mathcal{F}}(F)] \in \mathcal{H}_{\mathcal{F}}$.
- For conditional $P_{F|G}$, the conditional mean embedding is $\mu_{P|G}(g) = \mathbb{E}[\phi_{\mathcal{F}}(F) | G = g] \in \mathcal{H}_{\mathcal{F}}$.
- The counterfactual policy mean embedding (CPME): $\chi(\pi) = \mathbb{E}_{P_\pi}[\phi_{\mathcal{Y}}(Y(a))]$.
- κ_{ax}, κ_y : bounds on kernels: $\sup_{a,x} \|\phi_{\mathcal{AX}}(a, x)\|_{\mathcal{H}_{\mathcal{AX}}} \leq \kappa_{a,x}$, $\sup_y \|\phi_{\mathcal{Y}}(y)\|_{\mathcal{H}_{\mathcal{Y}}} \leq \kappa_y$.
- $S_2(\mathcal{H}_{\mathcal{AX}}, \mathcal{H}_{\mathcal{Y}})$ denotes the Hilbert space of the Hilbert-Schmidt operators from $\mathcal{H}_{\mathcal{AX}}$ to $\mathcal{H}_{\mathcal{Y}}$.
- $\mathcal{C}_{Y|A,X} \in S_2(\mathcal{H}_{\mathcal{AX}}, \mathcal{H}_{\mathcal{Y}})$ is the conditional mean operator.
- μ_π : the kernel policy embedding in $\mathcal{H}_{\mathcal{AX}}$.
- c, b : source condition and spectral decay parameters.
- λ : regularization parameter for learning $\mathcal{C}_{Y|A,X}$.
- L : Kernel integral operator $Lh := \int k(\cdot, w)h(w) d\rho(w)$, mapping $L^2(\rho) \rightarrow L^2(\rho)$.
- $\{\eta_j\}_{j \geq 1}$: Eigenvalues of L , ordered decreasingly, assumed to satisfy a spectral decay assumption $\eta_j \leq Cj^{-b}$.
- $\{\varphi_j\}_{j \geq 1}$: Orthonormal eigenfunctions of L in $L^2(\rho)$, satisfying $L\varphi_j = \eta_j\varphi_j$.
- \mathcal{H}^c : Interpolation space of order c , defined as $\mathcal{H}^c := \left\{ f = \sum_j h_j \varphi_j \mid \sum_j h_j^2 / \eta_j^c < \infty \right\}$.

Notations related to estimators and asymptotic analysis

- $\hat{\chi}_{pi}(\pi)$ and $\hat{\chi}_{dr}(\pi)$: plug-in and doubly robust estimators of CPME.
- $\hat{\mu}_{Y|A,X}(a, x)$: estimator of the conditional mean embedding.

- 943 – $\hat{\pi}_0(a|x)$: estimator of the logging policy.
- 944 – $r_C(n, b, c)$: convergence rate of $\hat{C}_{Y|A, X}$.
- 945 – $r_{\pi_0}(n)$: convergence rate of $\hat{\pi}_0$.
- 946 – $o_p(1), O_p(1)$: standard probabilistic asymptotic notations.

947 Notations for differentiability and statistical models

- 948 – \mathcal{P} : statistical model on $\mathcal{Z} = \mathcal{Y} \times \mathcal{A} \times \mathcal{X}$.
- 949 – $L^2(\rho)$: space of square-integrable real-valued functions w.r.t. measure ρ .
- 950 – $L^2(P; \mathcal{H})$: Bochner space of \mathcal{H} -valued functions with norm $\|f\|_{L^2(P; \mathcal{H})} =$
951 $(\int \|f(z)\|_{\mathcal{H}}^2 dP(z))^{1/2}$.
- 952 – $\Pi_{\mathcal{H}}[h | \mathcal{W}]$: orthogonal projection of h onto closed subspace $\mathcal{W} \subset \mathcal{H}$.
- 953 – \mathcal{P}_{π_0} : submodel of \mathcal{P} with fixed treatment policy π_0 .
- 954 – $\dot{\mathcal{P}}_P$: tangent space at P .
- 955 – $\mathcal{S}(P, \mathcal{P}, s)$: smooth submodels of \mathcal{P} at P with score s .
- 956 – $s, s_X, s_{Y|A, X}, s_{A|X}$: score functions.
- 957 – $\chi(\pi)(P)$: value of the CPME at P .
- 958 – $\dot{\chi}_P^\pi$: local parameter of $\chi(\pi)$ at P .
- 959 – $\dot{\chi}_P^{\pi, *}$: adjoint (efficient influence operator).
- 960 – \mathcal{H}_P : image of $\dot{\chi}_P^{\pi, *}$.

961 Notations for efficient influence functions

- 962 – ψ_P^π : efficient influence function (EIF) at P .
- 963 – $\tilde{\psi}_P^\pi$: candidate EIF: $\tilde{\psi}_P^\pi(y, a, x) = \dot{\chi}_P^{\pi, *}(\phi_Y)(y, a, x)$.

964 Error decomposition of the one-step estimator

- 965 – P_n : empirical distribution of the sample $\{z_i\}_{i=1}^n$.
- 966 – \hat{P}_n : estimated distribution using nuisance estimators.
- 967 – $\mathcal{S}_n = (P_n - P)\psi^\pi$: empirical average term.
- 968 – $\mathcal{T}_n = (P_n - P)(\hat{\psi}_n^\pi - \psi^\pi)$: empirical process term.
- 969 – $\mathcal{R}_n = \chi(\hat{P}_n) + P\hat{\psi}_n^\pi - \chi(\pi)$: remainder term.

970 Notations for empirical processes and equicontinuity

- 971 – $\mathcal{T}_n(\varphi) := \sqrt{n}(P_n - P)(\varphi)$: empirical process acting on φ .
- 972 – \mathcal{G} : class of \mathcal{H}_Y -valued functions (e.g. $\hat{\psi}_n^\pi - \psi^\pi$).

973 Notations for hypothesis testing

- 974 – H_0 : null hypothesis — $\nu(\pi) = \nu(\pi')$.
- 975 – H_1 : alternative hypothesis — $\nu(\pi) \neq \nu(\pi')$.
- 976 – $\varphi_{\pi, \pi'}$: difference of EIFs for policies π and π' .
- 977 – $\hat{\varphi}_{\pi, \pi'}, \tilde{\varphi}_{\pi, \pi'}$: estimates of $\varphi_{\pi, \pi'}$ over disjoint subsets.
- 978 – $\hat{\beta}_\pi(x) := \int \hat{\mu}_{Y|A, X}(a, x)\pi(da | x)$: estimated conditional policy mean.
- 979 – $f_{\pi, \pi'}^\dagger(y, a, x)$: cross-U-statistic kernel.
- 980 – $\bar{f}_{\pi, \pi'}^\dagger, S_{\pi, \pi'}^\dagger$: empirical mean and std of f^\dagger .
- 981 – $T_{\pi, \pi'}^\dagger$: normalized test statistic.
- 982 – \mathbb{H} : limiting Gaussian process in \mathcal{H}_Y .
- 983 – $\langle \mathbb{H}, h \rangle_{\mathcal{H}_Y}$: projection onto direction h .
- 984 – Φ : CDF of standard normal.
- 985 – $p = 1 - \Phi(T_{\pi, \pi'}^\dagger)$: p-value.

986 Notations for sampling from counterfactual distributions

- 987 – $(\tilde{y}_j)_{j=1}^m$: deterministic samples generated via kernel herding.
- 988 – \tilde{P}_Y^m : empirical distribution over the \tilde{y}_j .
- 989 – $\tilde{P}_{Y, dr}^m, \tilde{P}_{Y, pi}^m$: empirical distributions generated from $\hat{\chi}_{dr}(\pi)$ and $\hat{\chi}_{pi}(\pi)$.

9 Review of Counterfactual Mean Embeddings

In this appendix, we provide a background section on counterfactual mean embeddings [16] and distributional treatment effects.

9.1 Reproducing kernel hilbert spaces and kernel mean embeddings

A scalar-valued RKHS $\mathcal{H}_{\mathcal{W}}$ is a Hilbert space of functions $h : \mathcal{W} \rightarrow \mathbb{R}$. The RKHS is fully characterized by its feature map, which takes a point w in the original space \mathcal{W} and maps it to a feature $\phi_{\mathcal{W}}(w)$ in RKHS $\mathcal{H}_{\mathcal{W}}$. The closure of $\text{span}\{\phi_{\mathcal{W}}(w)\}_{w \in \mathcal{W}}$ is RKHS $\mathcal{H}_{\mathcal{W}}$. In other words, $\{\phi_{\mathcal{W}}(w)\}_{w \in \mathcal{W}}$ can be viewed as the dictionary of basis functions for RKHS $\mathcal{H}_{\mathcal{W}}$. The kernel $k_{\mathcal{W}} : \mathcal{W} \times \mathcal{W} \rightarrow \mathbb{R}$ is the inner product of features $\phi_{\mathcal{W}}(w)$ and $\phi_{\mathcal{W}}(w')$.

$$k_{\mathcal{W}}(w, w') = \langle \phi_{\mathcal{W}}(w), \phi_{\mathcal{W}}(w') \rangle_{\mathcal{H}_{\mathcal{W}}}. \quad (15)$$

A real-valued kernel k is continuous, symmetric and positive definite. The essential property of a function h in an RKHS $\mathcal{H}_{\mathcal{W}}$ is the eponymous reproducing property:

$$h(w) = \langle h, \phi_{\mathcal{W}}(w) \rangle_{\mathcal{H}_{\mathcal{W}}} \quad (16)$$

In other words, to evaluate h at w , we take the RKHS inner product between h and the features $\phi_{\mathcal{W}}(w)$ for $\mathcal{H}_{\mathcal{W}}$. The reproducing property, importantly, allows to separate function h from features $\phi_{\mathcal{W}}(w)$ and thereby decouple the steps of nonparametric causal estimation. Notably, the RKHS is a practical hypothesis space for nonparametric regression.

Example 9.1. (Nonparametric regression) Consider the output $y \in \mathbb{R}$, the input $w \in \mathcal{W}$ and the goal of estimating the conditional expectation function $h(w) = \mathbb{E}(Y | W = w)$. A kernel ridge regression estimator of h is

$$\hat{h} = \arg \min_{h \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^n \{y_i - \langle h, \phi_{\mathcal{W}}(w_i) \rangle_{\mathcal{H}}\}^2 + \lambda \|h\|_{\mathcal{H}}^2, \quad (17)$$

where $\lambda > 0$ is a hyperparameter on the ridge penalty $\|h\|_{\mathcal{H}}^2$, which imposes smoothness in estimation. The solution to the optimization problem has a well-known closed form:

$$\hat{h}(w) = Y^T (K_{WW} + n\lambda I)^{-1} K_{Ww}. \quad (18)$$

The closed-form solution involves the kernel matrix $K_{WW} \in \mathbb{R}^{n \times n}$ with (i, j) th entry $k_{\mathcal{W}}(w_i, w_j)$, and the kernel vector $K_{Ww} \in \mathbb{R}^n$ with i th entry $k_{\mathcal{W}}(w_i, w)$.

In this work, we use kernels and RKHSs to represent, compare, and estimate probability distributions. This is enabled by the approach known as kernel mean embedding (KME) of distributions [20], which we briefly review here. Let $\mathcal{H}_{\mathcal{W}}$ be a RKHS with kernel $k_{\mathcal{W}}$ defined on a space \mathcal{W} , and assume that $\sup_{w \in \mathcal{W}} k_{\mathcal{W}}(w, w) < \infty$. Then, for a probability distribution P over \mathcal{W} , the kernel mean embedding is defined as the Bochner integral¹:

$$\mu : \mathcal{P} \rightarrow \mathcal{H}_{\mathcal{W}}, \quad P \mapsto \mu_P := \int k_{\mathcal{W}}(\cdot, w) dP(w).$$

The embedded element μ_P , also written μ_W when $W \sim P$, serves as a representation of P in $\mathcal{H}_{\mathcal{W}}$. If $\mathcal{H}_{\mathcal{W}}$ is characteristic [65, 22, 66], this mapping is injective: $\mu_P = \mu_Q$ if and only if $P = Q$. Thus, μ_P uniquely identifies P , preserving all distributional information. Common examples of characteristic kernels on \mathbb{R}^d include Gaussian, Matérn, and Laplace kernels [22, 66], while linear and polynomial kernels are not characteristic due to their finite-dimensional RKHSs.

The kernel mean embedding induces a popular distance between probability measures known as the maximum mean discrepancy (MMD) [67, 68, 23]. For distributions P and Q , it is defined by:

$$\text{MMD}[\mathcal{H}_{\mathcal{W}}, P, Q] := \|\mu_P - \mu_Q\|_{\mathcal{H}_{\mathcal{W}}} = \sup_{h \in \mathcal{H}_{\mathcal{W}}, \|h\| \leq 1} \left| \int h dP - \int h dQ \right|.$$

¹See, e.g., [63, Chapter 2] and [64, Chapter 1] for the definition of the Bochner integral.

1024 The second equality follows from the reproducing property and the structure of RKHSs as vector
 1025 spaces [23, Lemma 4]. If $\mathcal{H}_{\mathcal{W}}$ is characteristic, then $\text{MMD}[\mathcal{H}_{\mathcal{W}}, P, Q] = 0$ implies $P = Q$, so
 1026 MMD defines a proper metric on distributions.

1027 Given an i.i.d. sample $\{w_i\}_{i=1}^n$ from P , the kernel mean embedding can be estimated via the empirical
 1028 average:

$$\hat{\mu}_P := \frac{1}{n} \sum_{i=1}^n k_{\mathcal{W}}(\cdot, w_i).$$

1029 This estimator is \sqrt{n} -consistent: $\|\mu_P - \hat{\mu}_P\|_{\mathcal{H}_{\mathcal{W}}} = O_p(n^{-1/2})$ under mild assumptions [23, 69, 70].

1030 Given a second i.i.d. sample $\{w'_j\}_{j=1}^m$ from Q , the squared empirical MMD is

$$\begin{aligned} \widehat{\text{MMD}}^2[\mathcal{H}_{\mathcal{W}}, P, Q] &= \|\hat{\mu}_P - \hat{\mu}_Q\|_{\mathcal{H}_{\mathcal{W}}}^2 \\ &= \frac{1}{n^2} \sum_{i,j=1}^n k_{\mathcal{W}}(w_i, w_j) - \frac{2}{nm} \sum_{i,j=1}^{n,m} k_{\mathcal{W}}(w_i, w'_j) + \frac{1}{m^2} \sum_{i,j=1}^m k_{\mathcal{W}}(w'_i, w'_j). \end{aligned}$$

1031 This estimator is consistent and converges at the parametric rate $O_p(n^{-1/2} + m^{-1/2})$. It is biased
 1032 but simple to compute; an unbiased version is also available [23, Eq. 3].

1033 The KME framework extends naturally to conditional distributions [46, 71–73]. Let (W, V) be a
 1034 random variable on $\mathcal{W} \times \mathcal{V}$ with joint distribution P_{WV} . Using kernels $k_{\mathcal{W}}$ and $k_{\mathcal{V}}$ with RKHSs
 1035 $\mathcal{H}_{\mathcal{W}}, \mathcal{H}_{\mathcal{V}}$, the conditional mean embedding of $P_{V|W=w}$ is defined as:

$$\mu_{V|W=w} := \int k_{\mathcal{V}}(\cdot, v) dP(v|w) \in \mathcal{H}_{\mathcal{V}}.$$

1036 This representation preserves all information if $\mathcal{H}_{\mathcal{V}}$ is characteristic. Given a sample $\{(w_i, v_i)\}_{i=1}^n$,
 1037 the conditional embedding can be estimated as

$$\hat{\mu}_{V|W=w} := \sum_{i=1}^n \beta_i(w) k_{\mathcal{V}}(\cdot, v_i),$$

1038 with weights

$$\beta(w) = (K + n\lambda I)^{-1} k_{\mathcal{W}}(w), \quad k_{\mathcal{W}}(w) = (k_{\mathcal{W}}(w, w_1), \dots, k_{\mathcal{W}}(w, w_n))^{\top}.$$

1039 Here, K is the $n \times n$ kernel matrix with entries $K_{ij} = k_{\mathcal{W}}(w_i, w_j)$, and $\lambda > 0$ is a regularization
 1040 parameter. This estimator corresponds to kernel ridge regression from \mathcal{W} into $\mathcal{H}_{\mathcal{V}}$, where the target
 1041 functions are feature maps $k_{\mathcal{V}}(\cdot, v_i)$. To guarantee convergence, λ must decay appropriately as
 1042 $n \rightarrow \infty$ [47, 54].

1043 Finally, we make use of the Hilbert space $S_2(\mathcal{H}_{\mathcal{W}_1}, \mathcal{H}_{\mathcal{W}_2})$ of Hilbert-Schmidt operators between
 1044 RKHSs. The conditional expectation operator $C : \mathcal{H}_{\mathcal{W}_1} \rightarrow \mathcal{H}_{\mathcal{W}_2}$ given by $h(\cdot) \mapsto \mathbb{E}[h(W_1)|W_2 = \cdot]$
 1045 is assumed to lie in $S_2(\mathcal{H}_{\mathcal{W}_1}, \mathcal{H}_{\mathcal{W}_2})$ and is estimated via ridge regression, by regressing $\phi_{\mathcal{W}_1}(W_1)$
 1046 on $\phi_{\mathcal{W}_2}(W_2)$ in $\mathcal{H}_{\mathcal{W}_2}$.

1047 9.2 Assumptions for consistency

1048 To prove consistency of our estimator, we rely on two standard approximation assumptions from
 1049 RKHS learning theory: *smoothness* of the target function and *spectral decay* of the kernel operator.
 1050 These are naturally formulated through the eigendecomposition of an associated integral operator,
 1051 which we introduce below. The results may be found in [74].

1052 **Kernel smoothing operator** Let $\mathcal{H}_{\mathcal{W}}$ be a reproducing kernel Hilbert space (RKHS) over a space
 1053 \mathcal{W} , with reproducing kernel with kernel $k_{\mathcal{W}} : \mathcal{W} \times \mathcal{W} \rightarrow \mathbb{R}$ consisting of functions of the form
 1054 $h : \mathcal{W} \rightarrow \mathbb{R}$. Let ρ be any Borel measure on \mathcal{W} . Let $L^2(\rho)$ be the space of square integrable functions
 1055 with respect to measure ρ . We define the integral operator L associated with the kernel $k_{\mathcal{W}}$ and the
 1056 measure ρ as:

$$L : L^2(\rho) \rightarrow L^2(\rho), h \mapsto \int k_{\mathcal{W}}(\cdot, w) h(w) d\rho(w) \quad (19)$$

Intuitively, this operator smooths a function h by averaging it with respect to the kernel $k_{\mathcal{W}}$ and the distribution ρ .

Remark 10. (*L as convolution*). If the kernel $k_{\mathcal{W}}$ is defined on $\mathcal{W} \subset \mathbb{R}^d$ and shift invariant, then L is a convolution of $k_{\mathcal{W}}$ and h . If $k_{\mathcal{W}}$ is smooth, then Lh is a smoothed version of h .

Spectral properties of the kernel smoothing operator The operator L is compact, self-adjoint, and positive semi-definite. Therefore, by the spectral theorem, L admits an orthonormal basis of eigenfunctions $(\varphi_j)_\rho$ in $\mathbb{L}_\rho^2(\mathcal{W})$, with corresponding non-negative eigenvalues (η_j) .

Assumption 11. (*Nonzero eigenvalues*). For simplicity, we assume $(\eta_j) > 0$ in this discussion; see [75, Remark 3] for the more general case.

Thus, for any $h \in L^2(\rho)$, we can write:

$$Lh = \sum_{j=1}^{\infty} \eta_j \langle \varphi_j, h \rangle_{\mathbb{L}_\rho^2} \varphi_j,$$

where each φ_j is defined up to ρ -almost-everywhere equivalence.

Feature map representation The following observations help to interpret this eigendecomposition.

Theorem 12. [76, Corollary 3.5] (*Mercer's Theorem*). The kernel $k_{\mathcal{W}}$ can be expressed as $k_{\mathcal{W}}(w, w') = \sum_{j=1}^{\infty} \eta_j \varphi_j(w) \varphi_j(w')$, where (w, w') are in the support of ρ , φ_j is a continuous element in the equivalence class $(\varphi_j)_\rho$, and the convergence is absolute and uniform.

Since the kernel $k_{\mathcal{W}}$ can be decomposed as:

$$k_{\mathcal{W}}(w, w') = \sum_{j=1}^{\infty} \eta_j \varphi_j(w) \varphi_j(w'),$$

with absolute and uniform convergence on compact subsets of the support of ρ , we can express the feature map $\phi_{\mathcal{W}}(w)$ associated with the RKHS as:

$$\phi_{\mathcal{W}}(w) = (\sqrt{\eta_1} \varphi_1(w), \sqrt{\eta_2} \varphi_2(w), \dots).$$

Thus, the inner product $\langle \phi_{\mathcal{W}}(w), \phi_{\mathcal{W}}(w') \rangle_{\mathcal{H}_{\mathcal{W}}}$ reproduces the kernel value $k_{\mathcal{W}}(w, w')$.

Both $L^2(\rho)$ and the RKHS \mathcal{H} can be described using the same orthonormal basis (φ_j) , but with different norms.

Remark 13. (*Comparison between \mathcal{H} and $\mathbb{L}_\rho^2(\mathcal{W})$*). A function $h \in L^2(\rho)$ has an expansion $h = \sum_j h_j \varphi_j$, and:

$$\|h\|_{L^2(\rho)}^2 = \sum_{j=1}^{\infty} h_j^2.$$

A function $h \in \mathcal{H}$ has the same expansion, but the RKHS norm is:

$$\|h\|_{\mathcal{H}}^2 = \sum_{j=1}^{\infty} \frac{h_j^2}{\eta_j}.$$

This means that functions with large coefficients on eigenfunctions associated with small eigenvalues are heavily penalized in \mathcal{H} , which enforces a notion of smoothness.

To summarize, the space \mathbb{L}_ρ^2 contains all square-integrable functions with respect to the measure ρ . In contrast, the RKHS \mathcal{H} is a subspace of \mathbb{L}_ρ^2 consisting of smoother functions—those whose spectral expansions put less weight on high-frequency eigenfunctions (i.e., those associated with small eigenvalues η_j).

This motivates two classical assumptions from statistical learning theory: the smoothness assumption, which constrains the target function via its spectral decay profile, and the spectral decay assumption, which characterizes the approximation capacity of the RKHS.

Remark 14. The smoothness assumption governs the approximation error (bias), while the spectral decay controls the estimation error (variance). These assumptions together determine the learning rate of kernel methods.

1093 **Source condition** To control the *bias* introduced by ridge regularization, we assume that the target
 1094 function lies in a smoother subspace of the RKHS. This is formalized by a *source condition*, a
 1095 common assumption in inverse problems and kernel learning theory [52, 77, 78].

1096 **Assumption 15.** (*Source Condition*) *There exists $c \in (1, 2]$ such that the target function h belongs to*
 1097 *the subspace*

$$\mathcal{H}^c = \left\{ f = \sum_{j=1}^{\infty} h_j \varphi_j : \sum_{j=1}^{\infty} \frac{h_j^2}{\eta_j^c} < \infty \right\} \subset \mathcal{H}.$$

1098

1099 When $c = 1$, this corresponds to assuming only that $h \in \mathcal{H}$. Larger values of c imply greater smooth-
 1100 ness: the function h can be well-approximated using only the leading eigenfunctions. Intuitively,
 1101 smoother targets lead to smaller bias and enable faster convergence of the estimator \hat{h} .

1102 **Variance and spectral decay** To control the *variance* of kernel ridge regression, we must also
 1103 constrain the complexity of the RKHS. This is done via a *spectral decay assumption*, which controls
 1104 the effective dimension of the RKHS by quantifying how quickly the eigenvalues η_j of the kernel
 1105 operator vanish.

1106 **Assumption 16.** (*Spectral Decay*) *We assume that there exists a constant $C > 0$ such that, for all j ,*

$$\eta_j \leq C j^{-b}, \quad \text{for some } b \geq 1.$$

1107

1108 This polynomial decay condition ensures that the contributions of high-frequency components
 1109 decrease rapidly. A bounded kernel implies that $b \geq 1$ [53, Lemma 10]. In the limit $b \rightarrow \infty$, the
 1110 RKHS becomes finite-dimensional. Intermediate values of b define how "large" or complex the RKHS
 1111 is, relative to the underlying measure ρ . A larger b corresponds to a smaller effective dimension and
 1112 thus a lower variance in estimation.

1113 **Space regularity** We can also require an additional assumption on the regularity of the domains.

1114 **Assumption 17.** (*Original Space Regularity Conditions*) *Assume that \mathcal{A}, \mathcal{X} (and \mathcal{Y}) are Polish*
 1115 *spaces.*

1116 A Polish space is a separable, completely metrizable topological space. This assumption covers a
 1117 broad range of settings, including discrete, continuous, and infinite-dimensional cases. When the
 1118 outcome Y is bounded, the moment condition is automatically satisfied.

1119 9.3 Further details on Counterfactual Policy Mean Embeddings

1120 To justify Proposition 2, we rely on the classical identification strategy established by Rosenbaum
 1121 and Rubin [43] and Robins [44]. Recall that the counterfactual policy mean embedding is defined as

$$\chi(\pi) := \mathbb{E}_{\pi \times P_X} [\phi_Y(Y(a))],$$

1122 which involves the unobserved potential outcome $Y(a)$. Under Assumption 1, we proceed to express
 1123 this quantity in terms of observed data.

1124 First, by the consistency assumption (also known as SUTVA), we have that for any realization where
 1125 $A = a$, the observed outcome satisfies $Y = Y(a)$. Second, by conditional exchangeability, we have
 1126 that $Y(a) \perp A \mid X$, which implies that the conditional distribution of $Y(a)$ given $X = x$ is equal to
 1127 the conditional distribution of Y given $A = a, X = x$. That is,

$$\mathbb{E}[\phi_Y(Y(a)) \mid X = x] = \mathbb{E}[\phi_Y(Y) \mid A = a, X = x] = \mu_{Y|A,X}(a, x).$$

1128 Finally, under the strong positivity assumption, the conditional density $\pi_0(a \mid x)$ is strictly bounded
 1129 away from zero for all $a \in \mathcal{A}, x \in \mathcal{X}$, ensuring that the conditional expectation $\mu_{Y|A,X}(a, x)$ is
 1130 identifiable throughout the support of $\pi \times P_X$. It follows that

$$\chi(\pi) = \mathbb{E}_{\pi \times P_X} [\mathbb{E}[\phi_Y(Y(a)) \mid X = x]] = \mathbb{E}_{\pi \times P_X} [\mu_{Y|A,X}(a, x)],$$

1131 which completes the identification argument.

10 Details and Analysis of the Plug-in Estimator

In this appendix, we provide further details on the analysis of the plug-in estimator proposed in Section 3.

10.1 Decoupling

We propose a plug-in estimator based on conditional mean operators for the nonparametric distribution of the outcome under policy a target policy π . Due to a decomposition property specific to the reproducing kernel Hilbert space, our plug-in estimator has a simple closed form solution.

Proposition 4 ((Decoupling via kernel mean embedding)). *Suppose Assumptions 1 and 3 hold. Then, the counterfactual policy mean embedding can be expressed as:*

$$\chi(\pi) = C_{Y|A,X} \mu_\pi$$

Proof. In Assumption 3, we impose that the scalar kernels are bounded. This assumption has several implications. First, the feature maps are Bochner integrable [74, see Definition A.5.20]. Bochner integrability permits us to interchange the expectation and inner product. Second, the mean embeddings exist. Third, the product kernel is also bounded and hence the tensor product RKHS inherits these favorable properties. By Proposition 2 and the linearity of expectation,

$$\begin{aligned} \chi(\pi) &= \int \mu_{Y|A,X}(a, x) d\pi(a|x) dP(x) \\ &= \int \mathcal{C}_{Y|A,X} \{ \phi_A(a) \otimes \phi_X(x) \} d\pi(a|x) dP(x) \\ &= \mathcal{C}_{Y|A,X} \int \phi_A(a) \otimes \phi_X(x) d\pi(a|x) dP(x) \\ &= \mathcal{C}_{Y|A,X} \mu_\pi. \end{aligned}$$

□

10.2 Analysis of the plug-in estimator

We will now present technical lemmas for kernel mean embeddings and conditional mean embeddings.

Kernel mean embedding For expositional purposes, we summarize classic results for the kernel mean embedding estimator $\hat{\mu}_z$ for $\mu_z = E\{\phi(Z)\}$.

Lemma 10.1. (Bennett inequality; Lemma 2 of Smale and Zhou [78]) *Let (ξ_i) be i.i.d. random variables drawn from the distribution P taking values in a real separable Hilbert space \mathcal{K} . Suppose there exists M such that $\|\xi_i\|_{\mathcal{K}} \leq M < \infty$ almost surely and $\sigma^2(\xi_i) = E(\|\xi_i\|_{\mathcal{K}}^2)$. Then for all $n \in \mathbb{N}$ and for all $\delta \in (0, 1)$,*

$$\Pr \left[\left\| \frac{1}{n} \sum_{i=1}^n \xi_i - E(\xi) \right\|_{\mathcal{K}} \leq \frac{2M \log(2/\delta)}{n} + \left\{ \frac{2\sigma^2(\xi) \log(2/\delta)}{n} \right\}^{1/2} \right] \geq 1 - \delta$$

We next provide a convergence result for the mean embedding, following from the above. This is included to make the paper self contained, however see [55, Proposition A.1] for an improved constant and a proof that the rate is minimax optimal.

Proposition 18. (Mean embedding Rate). *Suppose Assumptions 3 and 17 hold. Then with probability $1 - \delta$,*

$$\|\hat{\mu}_\pi - \mu_\pi\|_{\mathcal{H}} \leq r_{\mu_\pi}(n, \delta) = \frac{4\kappa_z \log(2/\delta)}{n^{1/2}}$$

Proof. The result follows from Lemma 10.1 with $\xi_i = \phi(Z_i)$, since

$$\left\| n^{-1} \sum_{i=1}^n \phi(Z_i) - E_{P_\pi} \{ \phi(Z) \} \right\|_{\mathcal{H}_Z} \leq \frac{2\kappa_z \log(2/\delta)}{n} + \left\{ \frac{2\kappa_z^2 \log(2/\delta)}{n} \right\}^{1/2} \leq \frac{4\kappa_z \log(2/\delta)}{n^{1/2}}$$

See [20, Theorem 2] for an alternative argument via Rademacher complexity.

□

1152 **Conditional mean embeddings** Below, we restate Assumptions 15 and 16 for the RKHS $\mathcal{H}_{\mathcal{A}\mathcal{X}}$,
 1153 which are used to establish the convergence rate of learning the conditional mean operator $C_{Y|A,X}$.
 1154 Our formulation of Assumption 15 differs slightly from the one in Appendix 9.2, but they are
 1155 equivalent due to [52, Remark 2].

Assumption 15 (Source condition.). *We define the (uncentered) covariance operator $\Sigma_{AX} = \mathbb{E}[\phi_{\mathcal{A}\mathcal{X}}(A, X) \otimes \phi_{\mathcal{A}\mathcal{X}}(A, X)]$. There exists a constant $B < \infty$ such that for a given $c \in (1, 3]$,*

$$\|C_{Y|A,X} \Sigma_{AX}^{-\frac{c-1}{2}}\|_{S_2(\mathcal{H}_{\mathcal{A}\mathcal{X}}, \mathcal{H}_{\mathcal{Y}})} \leq B$$

1156 In the above assumption, the smoothness parameter is allowed to range up to $c \leq 3$, in contrast to
 1157 prior work on kernel ridge regression, which typically restricts it to $c \leq 2$ [e.g. 53]. This extension is
 1158 justified by Meunier et al. [79, Remark 7 and Proposition 7], who showed that the saturation effect of
 1159 Tikhonov regularization can be extended to $c \leq 3$ when the error is measured in the RKHS norm, as
 1160 in Theorem 19, rather than the L^2 norm.

1161 **Assumption 16** (Eigenvalue decay.). *Let $(\lambda_{1,i})_{i \geq 1}$ be the eigenvalues of Σ_{AX} . For some constant*
 1162 *$B > 0$ and parameter $b \in (0, 1]$ and for all $i \geq 1$,*

$$\lambda_{1,i} \leq C i^{-b}.$$

Theorem 19. (Theorem 3 [54]) *Suppose Assumptions, 3, 15, 16 and 17, hold and take $\lambda_1 = \Theta\left(n^{-\frac{1}{c+1/b}}\right)$. There is a constant $J_1 > 0$ independent of $n \geq 1$ and $\delta \in (0, 1)$ such that*

$$\left\| \hat{C}_{Y|A,X} - C_{Y|A,X} \right\|_{S_2(\mathcal{H}_{\mathcal{A}\mathcal{X}}, \mathcal{H}_{\mathcal{Y}})} \leq J_1 \log(4/\delta) \left(\frac{1}{\sqrt{n}} \right)^{\frac{c-1}{c+1/b}} =: r_C(\delta, n, b, c)$$

1163 *is satisfied for sufficiently large $n \geq 1$ with probability at least $1 - \delta$.*

1164 We will now appeal to these previous lemmas to prove the consistency of the causal function.

1165 **Theorem 5** ((Consistency of the plug-in estimator).). *Suppose Assumptions 1, 3, 15, 16 and 17. Set*
 1166 *$\lambda = n^{-1/(c+1/b)}$, which is rate optimal regularization. Then, with high probability,*

$$\|\hat{\chi}_{pi}(\pi) - \chi(\pi)\|_{\mathcal{H}_{\mathcal{Y}}} = O[r_C(n, \delta, b, c)] = O\left[n^{-(c-1)/\{2(c+1/b)\}}\right]$$

1167 *Proof of Theorem 5.* We note that

$$\begin{aligned} \hat{\chi}_{pi}(\pi) - \chi(\pi) &= \hat{C}_{Y|A,X} \hat{\mu}_\pi - C_{Y|A,X} \mu_\pi \\ &= \hat{C}_{Y|A,X} (\hat{\mu}_\pi - \mu_\pi) + (\hat{C}_{Y|A,X} - C_{Y|A,X}) \mu_\pi \\ &= (\hat{C}_{Y|A,X} - C_{Y|A,X}) (\hat{\mu}_\pi - \mu_\pi) + C_{Y|A,X} (\hat{\mu}_\pi - \mu_\pi) + (\hat{C}_{Y|A,X} - C_{Y|A,X}) \mu_\pi. \end{aligned}$$

1168 Therefore we can write with Cauchy-Schwartz inequality:

$$\begin{aligned} |\hat{\chi}_{pi}(\pi) - \chi(\pi)| &\leq \left\| \hat{C}_{Y|A,X} - C_{Y|A,X} \right\|_{S_2(\mathcal{H}_{\mathcal{A}\mathcal{X}}, \mathcal{H}_{\mathcal{Y}})} \|\hat{\mu}_\pi - \mu_\pi\|_{\mathcal{H}} \\ &\quad + \|C_{Y|A,X}\|_{S_2(\mathcal{H}_{\mathcal{A}\mathcal{X}}, \mathcal{H}_{\mathcal{Y}})} \|\hat{\mu}_\pi - \mu_\pi\|_{\mathcal{H}} \\ &\quad + \left\| \hat{C}_{Y|A,X} - C_{Y|A,X} \right\|_{S_2(\mathcal{H}_{\mathcal{A}\mathcal{X}}, \mathcal{H}_{\mathcal{Y}})} \|\mu_\pi\|_{\mathcal{H}} \end{aligned}$$

Therefore by Theorems 19 and 18, with probability $1 - 2\delta$,

$$|\hat{\chi}_{pi}(\pi) - \chi(\pi)| \leq r_C(n, \delta, b, c) \cdot r_\mu(n, \delta) + \|C_{Y|A,X}\|_{S_2(\mathcal{H}_{\mathcal{A}\mathcal{X}}, \mathcal{H}_{\mathcal{Y}})} \cdot r_\mu(n, \delta) + \kappa_{a,x} \cdot r_C(n, \delta, b, c).$$

1169 Using Assumption 15, we observe that $\|C_{Y|A,X}\|_{S_2(\mathcal{H}_{\mathcal{A}\mathcal{X}}, \mathcal{H}_{\mathcal{Y}})} \leq B\kappa^{c-1}$. As a result, the above
 1170 bound readily gives

$$|\hat{\chi}_{pi}(\pi) - \chi(\pi)| \lesssim n^{-\frac{1}{2} \frac{c-1}{c+1/b}}.$$

1171 □

1172 10.3 Further details and Estimation strategies for the kernel policy mean embedding

1173 **Discrete Action Spaces.** When the action space \mathcal{A} is discrete, we can directly compute the kernel
 1174 policy mean embedding by exploiting the known form of the target policy $\pi(a \mid x)$. For each logged
 1175 context x_i , we compute a convex combination of the feature maps $\phi_{\mathcal{A}\mathcal{X}}(a, x_i)$, weighted by the
 1176 policy $\pi(a \mid x_i)$. This leads to the following empirical estimator:

$$\hat{\mu}_\pi = \frac{1}{n} \sum_{i=1}^n \sum_{a \in \mathcal{A}} \pi(a \mid x_i) \phi_{\mathcal{A}\mathcal{X}}(a, x_i).$$

1177 The plug-in estimator for the counterfactual policy mean embedding then admits the following matrix
 1178 expression:

$$\begin{aligned} \hat{\chi}(\pi) &= \hat{C}_{Y|A,X} \hat{\mu}_\pi \\ &= (K_{AA} \odot K_{XX} + n\lambda I)^{-1} (\Phi_{\mathcal{A}} \otimes \Phi_{\mathcal{X}}) \left(\frac{1}{n} \sum_{i=1}^n \sum_{a \in \mathcal{A}} \pi(a \mid x_i) \phi_{\mathcal{A}\mathcal{X}}(a, x_i) \right) \\ &= (K_{AA} \odot K_{XX} + n\lambda I)^{-1} \underbrace{(\Phi_{\mathcal{A}} \otimes \Phi_{\mathcal{X}})(\Phi_\pi \otimes \Phi_{\mathcal{X}})}_{K_\pi \odot K_{XX}} \mathbf{1} \frac{1}{n} \\ &= (K_{AA} \odot K_{XX} + n\lambda I)^{-1} (K_\pi \odot K_{XX}) \mathbf{1} \frac{1}{n}, \end{aligned}$$

1179 where $K_\pi[i, j] = \sum_{a \in \mathcal{A}} k_{\mathcal{A}}(a_i, a) \pi(a \mid x_j)$, and Φ_π denotes the policy-weighted features.

Algorithm 3 Plug-in estimator of the CPME (Discrete actions)

Require: Kernels $k_{\mathcal{X}}, k_{\mathcal{A}}, k_{\mathcal{Y}}$, and regularization constant $\lambda > 0$.

Input: Logged data $(x_i, a_i, y_i)_{i=1}^n$, target policy $\pi(a \mid x)$.

- 1: Compute empirical kernel matrices $K_{AA}, K_{XX} \in \mathbb{R}^{n \times n}$ from the samples $\{(a_i, x_i)\}_{i=1}^n$
- 2: Compute the kernel outcome matrix $K_{yY} = [k_{\mathcal{Y}}(y_1, y), \dots, k_{\mathcal{Y}}(y_n, y)]$
- 3: Compute $K_\pi \in \mathbb{R}^{n \times n}$ with entries $K_\pi[i, j] = \sum_{a \in \mathcal{A}} \pi(a \mid x_j) \cdot k_{\mathcal{A}\mathcal{X}}((a_i, x_i), (a, x_j))$
- 4: Set $\tilde{K} = K_\pi \cdot \frac{1}{n} \cdot (1 \dots 1)^\top$

Output: An estimate $\hat{\chi}_{pi}(\pi)(y) = K_{yY} (K_{AA} \odot K_{XX} + n\lambda I)^{-1} \tilde{K}$.

1180 **Continuous Actions via Resampling.** When \mathcal{A} is continuous and no closed-form sum over actions
 1181 is available, we instead approximate the kernel policy mean embedding by resampling from $\pi(\cdot \mid x_i)$.
 1182 Specifically, for each logged covariate x_i , we sample $\tilde{a}_i \sim \pi(\cdot \mid x_i)$, and form the empirical estimate:

$$\hat{\mu}_\pi = \frac{1}{n} \sum_{i=1}^n \phi_{\mathcal{A}\mathcal{X}}(\tilde{a}_i, x_i).$$

1183 This leads to the following expression for the plug-in estimator:

$$\begin{aligned} \hat{\chi}(\pi) &= \hat{C}_{Y|A,X} \hat{\mu}_\pi \\ &= (K_{AA} \odot K_{XX} + n\lambda I)^{-1} (\Phi_{\mathcal{A}} \otimes \Phi_{\mathcal{X}}) \frac{1}{n} \sum_{i=1}^n \phi_{\mathcal{A}\mathcal{X}}(\tilde{a}_i, x_i) \\ &= (K_{AA} \odot K_{XX} + n\lambda I)^{-1} \underbrace{(\Phi_{\mathcal{A}} \otimes \Phi_{\mathcal{X}})(\Phi_{\tilde{\mathcal{A}}} \otimes \Phi_{\mathcal{X}})}_{K_{A\tilde{A}} \odot K_{XX}} \mathbf{1} \frac{1}{n} \\ &= (K_{AA} \odot K_{XX} + n\lambda I)^{-1} (K_{A\tilde{A}} \odot K_{XX}) \mathbf{1} \frac{1}{n}, \end{aligned}$$

1184 where $K_{A\tilde{A}}[i, j] = k_{\mathcal{A}}(a_i, \tilde{a}_j)$, and \tilde{a}_j is drawn from $\pi(\cdot \mid x_j)$.

Algorithm 4 Plug-in estimator of the CPME

Require: Kernels $k_{\mathcal{X}}, k_{\mathcal{A}}, k_{\mathcal{Y}}$, and regularization constant $\lambda > 0$.

Input: Logged data $(x_i, a_i, y_i)_{i=1}^n$, the target policy π ,

1: Compute empirical kernel matrices $K_{AA}, K_{XX} \in \mathbb{R}^{T \times T}$ from the empirical samples

2: Compute the kernel outcome matrix $K_{YY} = [k_{\mathcal{Y}}(y_1, y), \dots, k_{\mathcal{Y}}(y_n, y)]$

3: Compute \tilde{K} with resampling, $\tilde{K} = (K_{AA} \odot K_{XX}) \cdot (1 \dots 1)^\top \frac{1}{n}$ and $\tilde{A} \sim \pi(\cdot | X)$.

Output: An estimate $\hat{\chi}_{pi}(\pi)(y) = K_{YY} (K_{AA} \odot K_{XX} + n\lambda I)^{-1} \tilde{K}$.

1185 **Importance Sampling** This resampling procedure can be quite cumbersome however, and not
1186 appropriate for off-policy learning. When propensity scores are known, an optional alternative is to
1187 invoke an inverse propensity scoring method [80], which expresses the embedding under the target
1188 policy π as a reweighting of the observational distribution:

$$\mu_\pi = \mathbb{E}_{\pi_0 \times P_X} \left[\frac{\pi(a | x)}{\pi_0(a | x)} \phi_{\mathcal{AX}}(a, x) \right]. \quad (20)$$

1189 This formulation enables a direct estimator of μ_π from logged data $\{(x_i, a_i, y_i)\}_{i=1}^n$, using the known
1190 logging policy π_0 :

$$\hat{\mu}_\pi = \frac{1}{n} \sum_{i=1}^n \frac{\pi(a_i | x_i)}{\pi_0(a_i | x_i)} \phi_{\mathcal{AX}}(a_i, x_i). \quad (21)$$

1191 Let $W_\pi \in \mathbb{R}^n$ be the vector of importance weights $W_\pi[i] = \frac{\pi(a_i | x_i)}{\pi_0(a_i | x_i)}$, and let $\Phi_{\mathcal{AX}} =$
1192 $[\phi_{\mathcal{AX}}(a_1, x_1), \dots, \phi_{\mathcal{AX}}(a_n, x_n)]$. Then the estimator admits the vectorized form:

$$\hat{\mu}_\pi = \Phi_{\mathcal{AX}} \left(\frac{1}{n} W_\pi \right). \quad (22)$$

1193 Accordingly, the closed-form expression for the plug-in estimator becomes:

$$\begin{aligned} \hat{\chi}(\pi) &= \hat{C}_{Y|A, X} \hat{\mu}_\pi \\ &= (K_{AA} \odot K_{XX} + n\lambda I)^{-1} (\Phi_{\mathcal{A}} \otimes \Phi_{\mathcal{X}}) \cdot \Phi_{\mathcal{AX}} \left(\frac{1}{n} W_\pi \right) \\ &= (K_{AA} \odot K_{XX} + n\lambda I)^{-1} (K_{AA} \odot K_{XX}) W_\pi \cdot \frac{1}{n}. \end{aligned}$$

1194 This estimator leverages all observed samples without requiring resampling or external sampling
1195 procedures, and is especially suited to settings where both the logging and target policies are known
1196 or estimable. However, its stability critically depends on the variance of the importance weights W_π ,
1197 which may require regularization or clipping in practice. Moreover, this estimator is not compatible
1198 with the doubly robust estimator proposed in the next section.

1199 11 Details and Analysis of the Efficient Score Function based Estimator

1200 In this appendix, we provide background definitions and lemmas on the pathwise differentiability of
1201 RKHS-valued parameters [33, 40], followed by the derivation and analysis of a one-step estimator
1202 for the counterfactual policy mean embedding (CPME).

1203 As stated in Assumption 17, we work on a Polish space $(\mathcal{Z}, \mathcal{B})$ with $\mathcal{Z} = \mathcal{Y} \times \mathcal{A} \times \mathcal{X}$ and consider a
1204 collection of distributions \mathcal{P} defined on $(\mathcal{Z}, \mathcal{B})$. Let $z_1, \dots, z_n \sim P_0$ be an i.i.d. sample from some
1205 $P_0 \in \mathcal{P}$, and denote by P_n the empirical distribution. Let $\hat{P}_n \in \mathcal{P}$ be an estimate of P_0 . For a measure
1206 ρ on (\mathcal{X}, Σ) , the space $L^2(\rho)$ denotes the Hilbert space of ρ -almost surely equivalence classes of
1207 real-valued square-integrable functions, equipped with the inner product $\langle f, g \rangle_{L^2(\rho)} := \int f g d\rho$. For
1208 any Hilbert space \mathcal{H} , we write $L^2(P; \mathcal{H})$ for the space of Bochner-measurable functions $f : \mathcal{Z} \rightarrow \mathcal{H}$
1209 with finite norm

$$\|f\|_{L^2(P; \mathcal{H})} := \left(\int \|f(z)\|_{\mathcal{H}}^2 dP(z) \right)^{1/2}.$$

1210 If \mathcal{W} is a closed subspace of \mathcal{H} , we denote by $\Pi_{\mathcal{H}}[h | \mathcal{W}]$ the orthogonal projection of h onto \mathcal{W} .

1211 11.1 Background on pathwise differentiability of RKHS-valued parameters

1212 We begin with a brief review of the formalism used to characterize the smoothness of RKHS-valued
 1213 statistical parameters [33, 40]. Let \mathcal{P} be a model, i.e., a collection of probability distributions on the
 1214 Polish space $(\mathcal{Y} \times \mathcal{A} \times \mathcal{X}, \mathcal{B})$, dominated by a common σ -finite measure ρ .

1215 **Definition 11.1.** (*Quadratic mean differentiability*) A submodel $\{P_\epsilon : \epsilon \in [0, \delta)\} \subset \mathcal{P}$ is said to be
 1216 quadratic mean differentiable at P if there exists a score function $s \in L^2(P)$ such that

$$\left\| p_\epsilon^{1/2} - p^{1/2} - \frac{\epsilon}{2} s p^{1/2} \right\|_{L^2(\rho)} = o(\epsilon),$$

1217 where $p = \frac{dP}{d\rho}$ and $p_\epsilon = \frac{dP_\epsilon}{d\rho}$.

1218 We denote by $\mathcal{P}(P, \mathcal{P}, s)$ the set of submodels at P with score function s . The collection of such
 1219 $s \in L^2(P)$ for which $\mathcal{P}(P, \mathcal{P}, s) \neq \emptyset$ is called the *tangent set*, and its closed linear span is the
 1220 *tangent space* of \mathcal{P} at P , denoted $\dot{\mathcal{P}}_P$.

1221 We define $L_0^2(P) := \{s \in L^2(P) : \int s dP = 0\}$, the largest possible tangent space, and refer to
 1222 models with $\dot{\mathcal{P}}_P = L_0^2(P)$ for all $P \in \mathcal{P}$ as *locally nonparametric*.

1223 The parameter of interest is the counterfactual policy mean embedding and can written over the model
 1224 \mathcal{P} as $\chi(\pi) : \mathcal{P} \rightarrow \mathcal{H}_Y$, such that

$$\chi(\pi)(P) = \iint \mathbb{E}_P[\phi_Y(Y) \mid A = a, X = x] \pi(da \mid x) P_X(dx). \quad (23)$$

1225 **Definition 11.2.** (*Pathwise differentiability*) The parameter $\chi(\pi)$ is pathwise differentiable at P if
 1226 there exists a continuous linear map $\dot{\chi}_P^\pi : \dot{\mathcal{P}}_P \rightarrow \mathcal{H}_Y$ such that for all $\{P_\epsilon\} \in \mathcal{P}(P, \mathcal{P}, s)$,

$$\|\chi(\pi)(P_\epsilon) - \chi(\pi)(P) - \epsilon \dot{\chi}_P^\pi(s)\|_{\mathcal{H}_Y} = o(\epsilon).$$

1227 We refer to $\dot{\chi}_P^\pi$ as the local parameter of $\chi(\pi)$ at P , and its Hermitian adjoint $(\dot{\chi}_P^\pi)^* : \mathcal{H}_Y \rightarrow \dot{\mathcal{P}}_P$ as
 1228 the efficient influence operator. Its image, denoted $\dot{\mathcal{H}}_P$, is a closed subspace of \mathcal{H}_Y known as the
 1229 local parameter space.

1230 Next, we go on defining the efficient influence function of the parameter $\chi(\pi)$.

1231 **Definition 11.3.** (*Efficient influence function*) We say that $\chi(\pi)$ has an efficient influence function
 1232 (EIF) $\psi_P^\pi : \mathcal{Y} \times \mathcal{A} \times \mathcal{X} \rightarrow \mathcal{H}_Y$ if there exists a P -almost sure set such that

$$\dot{\chi}_P^{\pi,*}(h)(y, a, x) = \langle h, \psi_P^\pi(y, a, x) \rangle_{\mathcal{H}_Y} \quad \text{for all } h \in \mathcal{H}_Y.$$

1233 By the Riesz representation theorem, $\chi(\pi)$ admits an EIF if and only if $\dot{\chi}_P^{\pi,*}(\cdot)(y, a, x)$ defines a
 1234 bounded linear functional almost surely. In that case, $\psi_P^\pi(y, a, x)$ equals its Riesz representation in
 1235 \mathcal{H}_Y .

1236 In our case, since \mathcal{H}_Y is an RKHS over a space \mathcal{Y} , the local parameter space $\dot{\mathcal{H}}_P$ is itself an RKHS
 1237 over \mathcal{Y} , with associated feature map ϕ_Y . Define

$$\tilde{\psi}_P^\pi(y, a, x) := \dot{\chi}_P^{\pi,*}(\phi_Y)(y, a, x), \quad (24)$$

1238 which serves as a candidate representation of the EIF. The following result will serve us to show that
 1239 $\tilde{\psi}_P^\pi$ both provides the form of the EIF of χ , when it exists, and also a sufficient condition that can be
 1240 used to verify its existence.

1241 **Proposition 20.** [40, Theorem 1], *Form of the efficient influence function* Suppose χ is pathwise
 1242 differentiable at P and $\dot{\mathcal{H}}_P$ is an RKHS. Then:

1243 i) If an EIF ψ_P^π exists, then $\psi_P^\pi = \tilde{\psi}_P^\pi$ almost surely.

1244 ii) If $\|\tilde{\psi}_P^\pi\|_{L^2(P; \mathcal{H}_Y)} < \infty$, then $\chi(\pi)$ admits an EIF at P .

1245 Prior to that, we state below a result to show a sufficient condition for pathwise differentiability.

1246 **Lemma 11.1.** (Sufficient condition for pathwise differentiability) [81, Lemma 2] The parameter
 1247 $\chi : \mathcal{P} \rightarrow \mathcal{H}_Y$ is pathwise differentiable at P if:

1248 i) $\dot{\chi}_P$ is bounded and linear, and there exists a dense set of scores $\mathcal{S}(P)$ such that for all
 1249 $s \in \mathcal{S}(P)$, a submodel $\{P_\epsilon\} \in \mathcal{P}(P, \mathcal{P}, s)$ satisfies

$$\|\chi(P_\epsilon) - \chi(P) - \epsilon \dot{\chi}_P(s)\|_{\mathcal{H}_Y} = o(\epsilon),$$

1250 ii) and χ is locally Lipschitz at P , i.e., there exist $(c, \delta) > 0$ such that for all $P_1, P_2 \in B_\delta(P)$,

$$\|\chi(P_1) - \chi(P_2)\|_{\mathcal{H}_Y} \leq c H(P_1, P_2),$$

1251 where $H(\cdot, \cdot)$ denotes the Hellinger distance and $B_\delta(P)$ is the δ -neighborhood of P in
 1252 Hellinger distance.

1253 Finally, we will show that under suitable conditions, an estimator of the form

$$\hat{\chi}_n(\pi) := \chi(\pi)(\hat{P}_n) + P_n \psi_n^\pi$$

1254 achieves efficiency.

1255 11.2 Derivation of the Efficient Influence Function

1256 We now prove Lemma 4.1, which characterizes the existence and form of the efficient influence
 1257 function (EIF) of the CPME. We begin by restating the lemma for convenience.

1258 **Lemma 4.1** ((Existence and form of the efficient influence function)). Suppose Assumptions 1 and
 1259 17 hold. Then, the CPME $\chi(\pi)$ admits an EIF which is P -Bochner square integrable and takes the
 1260 form

$$\psi^\pi(y, a, x) = \frac{\pi(a | x)}{\pi_0(a | x)} \{ \phi_Y(y) - \mu_{Y|A,X}(a, x) \} + \int \mu_{Y|A,X}(a', x) \pi(da' | x) - \chi(\pi).$$

1261 The proof proceeds in two main steps. First, we establish that χ is pathwise differentiable in Lemma
 1262 11.2. Then, we derive the form of its EIF.

1263 **Lemma 11.2.** χ is pathwise differentiable relative to a locally nonparametric model \mathcal{P} at any $P \in \mathcal{P}$
 1264

1265 *Proof.* Fix $\pi \in \Pi$. To prove this lemma, we apply Lemma 11.1 to establish the pathwise differ-
 1266 entiability of χ relative to a restricted model \mathcal{P}_{π_0} . This model consists of all distributions P' such
 1267 that $\pi_{P'} = \pi_0$, and for which there exists $P \in \mathcal{P}$ with $P'_{Y|A,X} = P_{Y|A,X}$ and $P'_X = P_X$. Since
 1268 the functional $\chi(\pi)$ does not depend on the treatment assignment mechanism, we may then extend
 1269 pathwise differentiability from \mathcal{P}_{π_0} to the full, locally nonparametric model \mathcal{P} .

1270 Following the construction in Luedtke and Chung [40], we assume that for any $P \in \mathcal{P}$ and fixed
 1271 $\delta > 0$, the model \mathcal{P} contains submodels of the form $\{P_\epsilon : \epsilon \in [0, \delta)\}$, where the perturbations act
 1272 only on the marginal of X and the conditional of $Y | A, X$. Specifically,

$$\frac{dP_{\epsilon,X}}{dP_X}(x) = 1 + \epsilon s_X(x), \quad \frac{dP_{\epsilon,A|X}}{dP_{A|X}}(a | x) = 1, \quad \frac{dP_{\epsilon,Y|A,X}}{dP_{Y|A,X}}(y | a, x) = 1 + \epsilon s_{Y|A,X}(y | a, x),$$

1273 where s_X and $s_{Y|A,X}$ are measurable functions bounded in $(-\delta^{-1}, \delta^{-1})$, satisfying

$$\mathbb{E}_P[s_X(X)] = 0 \quad \text{and} \quad \mathbb{E}_P[s_{Y|A,X}(Y | A, X) | A, X] = 0 \quad \text{a.s..}$$

1274 **Step 1: Boundedness and quadratic mean differentiability of the local parameter** Let π_0 be
 1275 such that $\pi_0 = \pi_{P'}$ for some fixed $P' \in \mathcal{P}$. The local parameter $\dot{\chi}_P^\pi(s)$ can be expressed as

$$\dot{\chi}_P^\pi(s) = \int \frac{\pi(a | x)}{\pi_0(a | x)} \phi_Y(y) [s_{Y|A,X}(y | a, x) + s_X(x)] P(dz). \quad (25)$$

1276 **Boundedness.** We first verify that $\dot{\chi}_P^\pi$ is a bounded operator. This will establish the first part of
 1277 condition (i) of Lemma 11.1 for the model \mathcal{P} at P .

1278 Take any score function s in the tangent space $\dot{\mathcal{P}}_P$. Define

$$s_{Y|A,X}(y | a, x) := s(x, a, y) - \mathbb{E}_P[s(X, A, Y) | A = a, X = x],$$

1279

$$s_X(x) := \mathbb{E}_P[s(X, A, Y) | X = x].$$

1280 It is straightforward to verify that $\mathbb{E}_P[s(X, A, Y) | A, X] - \mathbb{E}_P[s(X, A, Y) | X] = 0$ P -almost
 1281 surely. Therefore, we have the decomposition $s = s_{Y|A,X} + s_X$. Since $s \in L^2(P)$, it follows that
 1282 both $s_{Y|A,X}$ and s_X are in $L^2(P)$ as well.

1283 Now, under the strong positivity assumption and the boundedness of the kernel κ , the integrand

$$(x, a, y) \mapsto \frac{\pi(a | x)}{\pi_0(a | x)} \phi_Y(y) [s_{Y|A,X}(y | a, x) + s_X(x)]$$

1284 belongs to $L^2(P; \mathcal{H}_Y)$. Hence, the local parameter $\dot{\chi}_P^\pi(s)$ is well-defined in \mathcal{H}_Y .

1285 To establish boundedness of the local parameter $\dot{\chi}_P^\pi$, we compute its squared RKHS norm:

$$\begin{aligned} \|\dot{\chi}_P^\pi(s)\|_{\mathcal{H}_Y}^2 &= \iint \frac{\pi(a | x)}{\pi_0(a | x)} \frac{\pi(a' | x')}{\pi_0(a' | x')} k_y(y, y') [s_{Y|A,X}(y | a, x) + s_X(x)] \\ &\quad \cdot [s_{Y|A,X}(y' | a', x') + s_X(x')] P^2(dz, dz') \end{aligned} \quad (26)$$

$$\begin{aligned} &\leq \iint \frac{\pi(a | x)}{\pi_0(a | x)} \frac{\pi(a' | x')}{\pi_0(a' | x')} \sqrt{k_y(y, y) k_y(y', y')} |s_{Y|A,X}(y | a, x) + s_X(x)| \\ &\quad \cdot |s_{Y|A,X}(y' | a', x') + s_X(x')| P^2(dz, dz') \end{aligned} \quad (27)$$

$$= \left[\int \frac{\pi(a | x)}{\pi_0(a | x)} \sqrt{k_y(y, y)} |s_{Y|A,X}(y | a, x) + s_X(x)| P(dz) \right]^2 \quad (28)$$

$$\leq \left[\int \frac{\pi^2(a | x)}{\pi_0^2(a | x)} |k_y(y, y)| P(dz) \right] \cdot \left[\int (s_{Y|A,X}(y | a, x) + s_X(x))^2 P(dz) \right] \quad (29)$$

$$\leq \frac{\sup_{a,x} \pi(a | x) \cdot \sup_{y \in \mathcal{Y}} |k_y(y, y)|}{\inf_{P' \in \mathcal{P}} \text{ess inf}_{a,x} \pi_{P'}(a | x)} \cdot \int (s_{Y|A,X}(y | a, x) + s_X(x))^2 P(dz) \quad (30)$$

$$\leq \frac{\sup_{a,x} \pi(a | x) \cdot \sup_{y \in \mathcal{Y}} |k_y(y, y)|}{\inf_{P' \in \mathcal{P}} \text{ess inf}_{a,x} \pi_{P'}(a | x)} \cdot \|s\|_{L^2(P)}^2. \quad (31)$$

1286 Here: the first inequality applies Jensen's inequality to pull absolute values inside, and
 1287 Cauchy–Schwarz on the kernel k_y . The second applies Cauchy–Schwarz to split the integrals. The
 1288 third uses Hölder's inequality with exponents $(1, \infty)$. The final inequality follows from decomposing
 1289 $s = s_{Y|A,X} + s_X + s_{A|X}$, where

$$s_{A|X}(a | x) := \mathbb{E}_P[s(Z) | A = a, X = x] - \mathbb{E}_P[s(Z) | X = x].$$

1290 We then use

$$\|s\|_{L^2(P)}^2 = \|s_{Y|A,X} + s_X\|_{L^2(P)}^2 + \|s_{A|X}\|_{L^2(P)}^2 \geq \|s_{Y|A,X} + s_X\|_{L^2(P)}^2.$$

1291 Since the kernel k_y is bounded and π_0 is uniformly bounded away from zero by the strong positivity
 1292 assumption, the bound in (31) is finite. Therefore, $\dot{\chi}_P^\pi$ is a bounded linear operator.

1293 **Quadratic mean differentiability.** We now establish that $\chi(\pi)$ is quadratic mean differentiable at
 1294 P with respect to the restricted model \mathcal{P}_{π_0} , assuming $\pi_0 = \pi_P$.

1295 As in Luedtke and Chung [40], we consider a smooth submodel $\{P_\epsilon : \epsilon \in [0, \delta]\} \subset \mathcal{P}_{\pi_0}$ of the form:

$$\frac{dP_{\epsilon,X}}{dP_X}(x) = 1 + \epsilon s_X(x), \quad \frac{dP_{\epsilon,A|X}}{dP_{A|X}}(a | x) = 1, \quad \frac{dP_{\epsilon,Y|A,X}}{dP_{Y|A,X}}(y | a, x) = 1 + \epsilon s_{Y|A,X}(y | a, x),$$

1296 where s_X and $s_{Y|A,X}$ are bounded in $[-\delta^{-1}/2, \delta^{-1}/2]$, satisfy $\mathbb{E}_P[s_X(X)] = 0$ and $\mathbb{E}_P[s_{Y|A,X}(Y |$
 1297 $A, X) | A, X] = 0$ almost surely. The score of this submodel at $\epsilon = 0$ is given by $s(x, a, y) =$
 1298 $s_X(x) + s_{Y|A,X}(y | a, x)$, and its $L^2(P)$ -closure spans the tangent space of \mathcal{P}_{π_0} at P .

1299 Letting $\dot{\chi}_P^\pi(s)$ be defined as in Equation (25), we compute

$$\begin{aligned}
& \|\chi(\pi)(P_\epsilon) - \chi(\pi)(P) - \epsilon \dot{\chi}_P^\pi(s)\|_{\mathcal{H}_Y}^2 \\
&= \left\| \iiint \phi_Y(y)(1 + \epsilon s_{Y|A,X}(y | a, x))(1 + \epsilon s_X(x)) P_{Y|A,X}(dy | a, x) \pi(da | x) P_X(dx) \right. \\
&\quad \left. - \iiint \phi_Y(y) P_{Y|A,X}(dy | a, x) \pi(da | x) P_X(dx) \right. \\
&\quad \left. - \epsilon \iiint \frac{\pi(a | x)}{\pi_0(a | x)} \phi_Y(y) [s_{Y|A,X}(y | a, x) + s_X(x)] P_{Y|A,X}(dy | a, x) \pi_0(da | x) P_X(dx) \right\|_{\mathcal{H}_Y}^2 \\
&= \epsilon^4 \left\| \iiint \phi_Y(y) s_{Y|A,X}(y | a, x) s_X(x) P_{Y|A,X}(dy | a, x) \pi(da | x) P_X(dx) \right\|_{\mathcal{H}_Y}^2 \\
&= \epsilon^4 \left\| \int \frac{\pi(a | x)}{\pi_0(a | x)} \phi_Y(y) s_{Y|A,X}(y | a, x) s_X(x) P(dz) \right\|_{\mathcal{H}_Y}^2.
\end{aligned}$$

1300 This is $o(\epsilon^2)$ provided that the last \mathcal{H}_Y -norm is finite. To verify this, observe that the integrand

$$(x, a, y) \mapsto \frac{\pi(a | x)}{\pi_0(a | x)} \phi_Y(y) s_{Y|A,X}(y | a, x) s_X(x)$$

1301 belongs to $L^2(P; \mathcal{H}_Y)$, since k_Y , $s_{Y|A,X}$, and s_X are bounded and π_0 satisfies the strong positivity
1302 assumption. Indeed if we compute its squared norm:

$$\begin{aligned}
& \left\| \int \frac{\pi(a | x)}{\pi_0(a | x)} \phi_Y(y) s_{Y|A,X}(y | a, x) s_X(x) P(dz) \right\|_{\mathcal{H}_Y}^2 \\
&= \iint \frac{\pi(a | x)}{\pi_0(a | x)} \frac{\pi(a' | x')}{\pi_0(a' | x')} k_Y(y, y') s_{Y|A,X}(y | a, x) s_X(x) s_{Y|A,X}(y' | a', x') s_X(x') P(dz) P(dz') \\
&< \infty.
\end{aligned}$$

1303 Thus, $\chi(\pi)$ is quadratic mean differentiable at P relative to \mathcal{P}_{π_0} .

1304 **Step 2: Local Lipschitzness.** Let $\pi_0 = \pi_{P'}$ for some fixed $P' \in \mathcal{P}$. We now verify that $\chi(\pi)$ is
1305 locally Lipschitz over the restricted model \mathcal{P}_{π_0} .

1306 Fix any $P, \tilde{P} \in \mathcal{P}_{\pi_0}$. Define the π -reweighted distributions:

$$P^\pi(z) := \frac{\pi(a | x)}{\pi_0(a | x)} P(z), \quad \tilde{P}^\pi(z) := \frac{\pi(a | x)}{\pi_0(a | x)} \tilde{P}(z),$$

1307 where $z = (x, a, y)$. Then:

$$\begin{aligned}
\|\chi(\pi)(P) - \chi(\pi)(\tilde{P})\|_{\mathcal{H}_Y}^2 &= \iint k_Y(y, y') (P^\pi - \tilde{P}^\pi)(dz) (P^\pi - \tilde{P}^\pi)(dz') \\
&= \iint k_Y(y, y') \frac{\pi(a | x)}{\pi_0(a | x)} (P - \tilde{P})(dz) \frac{\pi(a' | x')}{\pi_0(a' | x')} (P - \tilde{P})(dz') \\
&= \iint k_Y(y, y') \left[\sqrt{dP(z)} - \sqrt{d\tilde{P}(z)} \right] \left[\sqrt{dP(z')} - \sqrt{d\tilde{P}(z')} \right] \\
&\quad \times \left[\sqrt{dP(z)} + \sqrt{d\tilde{P}(z)} \right] \left[\sqrt{dP(z')} + \sqrt{d\tilde{P}(z')} \right] \\
&\quad \times \frac{\pi(a | x)}{\pi_0(a | x)} \frac{\pi(a' | x')}{\pi_0(a' | x')}.
\end{aligned}$$

1308 Applying the Cauchy–Schwarz inequality yields:

$$\begin{aligned} \|\chi(\pi)(P) - \chi(\pi)(\tilde{P})\|_{\mathcal{H}_Y}^2 &\leq \left(\iint k_Y^2(y, y') \left[\frac{\pi(a|x)}{\pi_0(a|x)} \frac{\pi(a'|x')}{\pi_0(a'|x')} \right]^2 \left[\sqrt{dP(z)} + \sqrt{d\tilde{P}(z)} \right]^2 \right. \\ &\quad \cdot \left. \left[\sqrt{dP(z')} + \sqrt{d\tilde{P}(z')} \right]^2 \right)^{1/2} \\ &\quad \cdot \left(\iint \left[\sqrt{dP(z)} - \sqrt{d\tilde{P}(z)} \right]^2 \left[\sqrt{dP(z')} - \sqrt{d\tilde{P}(z')} \right]^2 \right)^{1/2} \\ &= (\Lambda)^{1/2} \cdot H^2(P, \tilde{P}). \end{aligned}$$

1309 Where $\Lambda = \iint k_Y^2(y, y') \left[\frac{\pi(a|x)}{\pi_0(a|x)} \frac{\pi(a'|x')}{\pi_0(a'|x')} \right]^2 \left[\sqrt{dP(z)} + \sqrt{d\tilde{P}(z)} \right]^2 \left[\sqrt{dP(z')} + \sqrt{d\tilde{P}(z')} \right]^2$.

1310 Using the inequality $(b+c)^2 \leq 2(b^2+c^2)$ and applying Hölder's inequality:

$$\begin{aligned} \Lambda &\leq 2 \iint k_Y^2(y, y') \left[\frac{\pi(a|x)}{\pi_0(a|x)} \right]^2 \left[\frac{\pi(a'|x')}{\pi_0(a'|x')} \right]^2 (P + \tilde{P})(dz)(P + \tilde{P})(dz') \\ &\leq \frac{2 \sup_{x,a} \pi^2(a|x) \cdot \sup_{y,y'} k_Y^2(y, y')}{\inf_{P' \in \mathcal{P}} \inf_{x,a} \pi_{P'}^2(a|x)}. \end{aligned}$$

1311 This upper bound is finite under the strong positivity assumption and the boundedness of the kernel
1312 k_Y . Therefore, $\chi(\pi)$ is locally Lipschitz over \mathcal{P}_{π_0} . This establishes part (ii) of Lemma 11.1 and
1313 therefore finishes the proof.

1314 □

1315 Now that we have proved Lemma 11.2, we establish Lemma 4.1 and derive the form of the efficient
1316 influence function.

1317 *Proof.* To prove Lemma 4.1, we first recall that the local parameter takes the form, for $s \in \dot{\mathcal{P}}_P$

$$\dot{\chi}_P^\pi(s) = \iiint \phi_Y(y) [s_{Y|A,X}(y|a, x) + s_{A|X}(a|x) + s_X(x)] P_{Y|A,X}(dy|a, x) \pi(da|x) P_X(dx)$$

1318 Therefore, the efficient influence operator takes the form for $h \in \mathcal{H}_Y$

$$\dot{\chi}_P^{\pi,*}(h)(y, a, x) = \frac{\pi(a|x)}{\pi_0(a|x)} \{h(y) - \mathbb{E}_P[h(Y) | A = a, X = x]\} \quad (32)$$

$$+ \int \mathbb{E}_P[h(Y) | A = a', X = x] \pi(da' | x) \quad (33)$$

$$- \iint \mathbb{E}_P[h(Y) | A = a', X = x'] \pi(da' | x) P_X(dx') \quad (34)$$

By Proposition 20, the EIF is given by evaluating the efficient influence operator at the representer $\phi_Y(y')$, that is

$$\begin{aligned} \psi_P^\pi(z)(y') &= \dot{\chi}_P^{\pi,*}(\phi_Y(y'))(y, a, x) = \frac{\pi(a|x)}{\pi_0(a|x)} \{ \phi_Y(y') - \mathbb{E}_P[\phi_Y(y') | A = a, X = x] \} \\ &\quad + \int \mathbb{E}_P[\phi_Y(y') | A = a', X = x] \pi(da' | x) \\ &\quad - \iint \mathbb{E}_P[\phi_Y(y') | A = a', X = x'] \pi(da' | x) P_X(dx'). \end{aligned}$$

Indeed this function belongs to $L^2(P; \mathcal{H}_Y)$. Recalling the definition of the conditional mean embedding $\mu_{Y|A,X}(a, x)$ in (3) and noting that $\mathbb{E}_P [\mu_{Y|A,X}(a, x)] = \chi(\pi)(P)$, we can rewrite the above as follows:

$$\psi_P^\pi(z) = \frac{\pi(a | x)}{\pi_0(a | x)} [\phi_Y(y) - \mu_{Y|A,X}(a, x)] + \int \mu_{Y|A,X}(a', x) \pi(da' | x) - \chi(\pi)(P)$$

Finally, since the kernel k_Y is bounded and π_0 is bounded away from zero by Assumption 1, it follows that $\psi_P^\pi \in L^2(P; \mathcal{H}_Y)$.

□

11.3 Analysis of the one-step estimator

In this section we provide the analysis of the one-step estimator. We start by restating Theorem 6.

Theorem 6 ((Consistency of the doubly robust estimator).). *Suppose Assumptions 1, 3, 15, 16 and 17. Set $\lambda = n^{-1/(c+1/b)}$, which is rate optimal regularization. Then, with high probability,*

$$\|\hat{\chi}_{dr}(\pi) - \chi(\pi)\|_{\mathcal{H}_Y} = \mathcal{O} \left[n^{-1/2} + r_{\pi_0}(n) \cdot r_C(n, \delta, b, c) \right]$$

For this Theorem, we will begin by decomposing the error terms .

$$\hat{\chi}_{dr} - \chi(P) = \chi(\hat{P}_n) + P_n \hat{\psi}_n - \chi(P) = (P_n - P) \hat{\psi}_n + \chi(\hat{P}_n) + P \hat{\psi}_n - \chi(P) \quad (35)$$

$$= (P_n - P) \psi^\pi + (P_n - P) (\hat{\psi}_n^\pi - \psi^\pi) + \chi(\hat{P}_n) + P \hat{\psi}_n - \chi(\pi) \quad (36)$$

$$= \mathcal{S}_n + \mathcal{T}_n + \mathcal{R}_n \quad (37)$$

where $\mathcal{S}_n = (P_n - P) \psi^\pi$, $\mathcal{T}_n = (P_n - P) (\hat{\psi}_n^\pi - \psi^\pi)$ and $\mathcal{R}_n = \chi(\hat{P}_n) + P \hat{\psi}_n - \chi(\pi)$. \mathcal{S}_n is a sample average of a fixed function. We call \mathcal{R}_n the remainder terms and \mathcal{T}_n the empirical process term. The remainder terms \mathcal{R}_n , quantify the error in the approximation of the one-step estimator across the samples. The following result provides a reasonable condition under which the drift terms will be negligible.

11.3.1 Bounding the empirical process term

As explained in Appendix 11.4, Luedtke and Chung [40] proposed a cross-fitted version of the one-step estimator. However, splitting the data may lead to a loss in power. We are therefore interested in identifying a sufficient condition under which the empirical term \mathcal{T}_n becomes asymptotically negligible *without* sample splitting.

In the scalar-valued case, a Donsker class assumption ensures the empirical process term is asymptotically negligible [31]. However, directly extending this notion to \mathcal{H}_Y -valued functions is not straightforward, since standard entropy-based arguments rely on the total ordering of \mathbb{R} [60]. Fortunately, Park and Muandet [60] introduces a notion of *asymptotic equicontinuity* adapted to Banach- or Hilbert-space valued empirical processes, which we adopt in this setting.

Definition 11.4. (*Asymptotic equicontinuity*). *We say that the empirical process $\{\mathcal{T}_n(\varphi) = \sqrt{n} (P_n - P) \varphi : \varphi \in \mathcal{G}\}$ with values in \mathcal{H} and indexed by \mathcal{G} is asymptotic equicontinuous at $\varphi_0 \in \mathcal{G}$ if, for every sequence $\{\hat{\varphi}_n\} \subset \mathcal{G}$ with $\|\hat{\varphi}_n - \varphi_0\| \xrightarrow{P} 0$, we have*

$$\|\mathcal{T}_n(\hat{\varphi}_n) - \mathcal{T}_n(\varphi_0)\|_{\mathcal{H}} \xrightarrow{P} 0. \quad (38)$$

Note that (38) is equivalent to $\mathcal{T}_n = (P_n - P) (\hat{\psi}_n^\pi - \psi^\pi) = o_P \left(\frac{1}{\sqrt{n}} \right)$. Park and Muandet [60] gives sufficient conditions for asymptotic equicontinuity to hold that we will leverage to show asymptotic equicontinuity. First we state the following result on the convergence of the efficient influence function estimator.

Assumption 21. (*Estimated Positivity*) *There exists a constant $\eta > 0$ such that, with high probability as $n \rightarrow \infty$,*

$$\hat{\pi}_0(a | x) \geq \eta, \quad \text{for all } (a, x) \in \mathcal{A} \times \mathcal{X}.$$

1351 **Lemma 11.3.** (*Influence Function Error*). Suppose that the conditions of Lemma 4.1 hold, as well as
 1352 Assumptions 3, 21. Then the following bound holds:

$$\|\psi_P^\pi - \psi_0^\pi\|_{L^2(P_0; \mathcal{H}_Y)} = O_P \left(\left\| \frac{1}{\hat{\pi}_0} - \frac{1}{\pi_0} \right\|_{L^2(\pi_{P_X})} + \|\mu_{Y|A,X} - \hat{\mu}_{Y|A,X}\|_{L^2(\pi_{P_X}; \mathcal{H}_Y)} \right)$$

1353 *Proof.* We expand the difference between the estimated and oracle influence functions:

$$\begin{aligned} \psi_P^\pi(z) - \psi_0^\pi(z) &= \left(\frac{\pi(a|x)}{\hat{\pi}_0(a|x)} - \frac{\pi(a|x)}{\pi_0(a|x)} \right) (\phi_Y(y) - \mu_{Y|A,X}(a, x)) \\ &\quad + \frac{\pi(a|x)}{\hat{\pi}_0(a|x)} (\mu_{Y|A,X}(a, x) - \hat{\mu}_{Y|A,X}(a, x)) \\ &\quad + \int (\hat{\mu}_{Y|A,X}(a', x) - \mu_{Y|A,X}(a', x)) \pi(da' | x). \end{aligned}$$

1354 Taking the $L^2(P_0; \mathcal{H}_Y)$ norm and applying the triangle inequality yields:

$$\|\psi_P^\pi - \psi_0^\pi\|_{L^2(P_0; \mathcal{H}_Y)} \leq \text{(I)} + \text{(II)} + \text{(III)},$$

1355 where:

$$\begin{aligned} \text{(I)} &= \left\| \left(\frac{\pi(a|x)}{\hat{\pi}_0(a|x)} - \frac{\pi(a|x)}{\pi_0(a|x)} \right) (\phi_Y(y) - \mu_{Y|A,X}(a, x)) \right\|_{L^2(P_0; \mathcal{H}_Y)}, \\ \text{(II)} &= \left\| \frac{\pi(a|x)}{\hat{\pi}_0(a|x)} (\mu_{Y|A,X}(a, x) - \hat{\mu}_{Y|A,X}(a, x)) \right\|_{L^2(P_0; \mathcal{H}_Y)}, \\ \text{(III)} &= \left\| \int (\hat{\mu}_{Y|A,X}(a', x) - \mu_{Y|A,X}(a', x)) \pi(da' | x) \right\|_{L^2(P_0; \mathcal{H}_Y)}. \end{aligned}$$

1356 First, we consider the term

$$\text{(I)} = \left\| \left(\frac{\pi(a|x)}{\hat{\pi}_0(a|x)} - \frac{\pi(a|x)}{\pi_0(a|x)} \right) (\phi_Y(y) - \mu_{Y|A,X}(a, x)) \right\|_{L^2(P_0; \mathcal{H}_Y)}.$$

1357 Let $\Delta(a, x) := \frac{\pi(a|x)}{\hat{\pi}_0(a|x)} - \frac{\pi(a|x)}{\pi_0(a|x)}$ and $h(a, x, y) := \phi_Y(y) - \mu_{Y|A,X}(a, x) \in \mathcal{H}_Y$. Then,

$$\text{(I)}^2 = \int \|\Delta(a, x) \cdot h(a, x, y)\|_{\mathcal{H}_Y}^2 dP_0(a, x, y) = \int \Delta^2(a, x) \cdot \|h(a, x, y)\|_{\mathcal{H}_Y}^2 dP_0.$$

1358 Applying the Cauchy–Schwarz inequality gives:

$$\text{(I)} \leq \left(\int \Delta^2(a, x) dP_0 \right)^{1/2} \left(\int \|\phi_Y(y) - \mu_{Y|A,X}(a, x)\|_{\mathcal{H}_Y}^2 dP_0 \right)^{1/2}.$$

1359 Noting that $P_0(da, dx) = \pi_0(a|x)P_X(dx)$ and using the change of measure:

$$\int r^2(a, x) dP_0 = \int \left(\pi_0(a|x) \pi(a|x) \left(\frac{1}{\hat{\pi}_0(a|x)} - \frac{1}{\pi_0(a|x)} \right) \right)^2 \pi(a|x) P_X(dx),$$

1360 we obtain:

$$\text{(I)} \leq \left\| \pi_0 \pi \left(\frac{1}{\hat{\pi}_0} - \frac{1}{\pi_0} \right) \right\|_{L^2(\pi_{P_X})} \cdot \left(\int \|\phi_Y(y) - \mu_{Y|A,X}(a, x)\|_{\mathcal{H}_Y}^2 dP_0 \right)^{1/2}.$$

1361 Using that the kernel is bounded in Assumption 3, then the second factor is finite, and:

$$\text{(I)} = O_P \left(\left\| \left(\frac{1}{\hat{\pi}_0} - \frac{1}{\pi_0} \right) \right\|_{L^2(\pi_{P_X})} \right),$$

1362 for some constant depending on the kernel and outcome variance.

1363 Second, we analyze the term

$$(II) = \left\| \frac{\pi(a | x)}{\hat{\pi}_0(a | x)} (\mu_{Y|A,X}(a, x) - \hat{\mu}_{Y|A,X}(a, x)) \right\|_{L^2(P_0; \mathcal{H}_Y)}.$$

1364 By definition of the $L^2(P_0; \mathcal{H}_Y)$ norm, we have:

$$\begin{aligned} (II)^2 &= \int \left\| \frac{\pi(a | x)}{\hat{\pi}_0(a | x)} (\mu_{Y|A,X}(a, x) - \hat{\mu}_{Y|A,X}(a, x)) \right\|_{\mathcal{H}_Y}^2 dP_0(a, x) \\ &= \int \left(\frac{\pi(a | x)}{\hat{\pi}_0(a | x)} \right)^2 \left\| \mu_{Y|A,X}(a, x) - \hat{\mu}_{Y|A,X}(a, x) \right\|_{\mathcal{H}_Y}^2 \pi_0(a | x) P_X(dx). \end{aligned}$$

1365 Changing the measure to $\pi(a | x) P_X(dx)$ and bounding the weight by positivity assumptions yields:

$$(II)^2 = \int \left\| \mu_{Y|A,X}(a, x) - \hat{\mu}_{Y|A,X}(a, x) \right\|_{\mathcal{H}_Y}^2 \cdot w(a, x) \cdot \pi(a | x) P_X(dx),$$

1366 where $w(a, x) := \frac{\pi_0(a|x)\pi(a|x)}{\hat{\pi}_0^2(a|x)}$. If $\hat{\pi}_0 \geq \eta > 0$, because of Assumption 21, then

$$(II) = O_P \left(\left\| \mu_{Y|A,X} - \hat{\mu}_{Y|A,X} \right\|_{L^2(\pi P_X; \mathcal{H}_Y)} \right),$$

1367 for some constant depending on the inverse propensity bound.

1368 Eventually, we bound the term

$$(III) = \left\| \int (\hat{\mu}_{Y|A,X}(a', x) - \mu_{Y|A,X}(a', x)) \pi(da' | x) \right\|_{L^2(P_0; \mathcal{H}_Y)}.$$

1369 Simply, the interm does Using Jensen's inequality in the Hilbert space \mathcal{H}_Y [82, Chapter 6], for each
1370 fixed x , we have:

$$\left\| \int (\hat{\mu}(a', x) - \mu(a', x)) \pi(da' | x) \right\|_{\mathcal{H}_Y} \leq \int \|\hat{\mu}(a', x) - \mu(a', x)\|_{\mathcal{H}_Y} \pi(da' | x).$$

1371 Now square both sides and integrate over $P_0(a, x) = \pi_0(a | x) P_X(dx)$. Since the integrand is
1372 independent of a , this is equivalent to integrating over P_X with the density $\pi_0(a | x)$ marginalized
1373 out:

$$\begin{aligned} (III)^2 &= \int \left\| \int (\hat{\mu}(a', x) - \mu(a', x)) \pi(da' | x) \right\|_{\mathcal{H}_Y}^2 \pi_0(a | x) P_X(dx) \\ &= \int \left\| \int (\hat{\mu}(a', x) - \mu(a', x)) \pi(da' | x) \right\|_{\mathcal{H}_Y}^2 P_X(dx) \\ &\leq \int \left(\int \|\hat{\mu}(a', x) - \mu(a', x)\|_{\mathcal{H}_Y} \pi(da' | x) \right)^2 P_X(dx) \\ &\leq \int \int \|\pi(a' | x) (\hat{\mu}(a', x) - \mu(a', x))\|_{\mathcal{H}_Y}^2 \pi(da' | x) P_X(dx), \end{aligned}$$

1374 Therefore, using that π is bounded

$$(III) \leq \left\| \sup_{x,a} \pi(a | x) \right\| \left\| \hat{\mu}_{Y|A,X} - \mu_{Y|A,X} \right\|_{L^2(\pi P_X; \mathcal{H}_Y)}.$$

1375 Combining the bounds yields the desired result. \square

1376 Then, we are now in position to state:

1377 **Lemma 11.4.** (Asymptotic equicontinuity of the empirical process term) Suppose that Assump-
1378 tions 1, 3, 15, 17, 21 hold. Moreover, assume k_Y is a C^∞ Mercer kernel. Then the empirical process
1379 term satisfies $\|\mathcal{T}_n\|_{\mathcal{H}_Y} = o_P(n^{-1/2})$.

1380 *Proof.* Under Assumptions 1, 3, 15, 17, 21, the functions $\hat{\psi}_n^\pi(y, a, x) - \psi^\pi(y, a, x)$ lie in a finite and
 1381 shrinking ball of the RKHS \mathcal{H}_Y , therefore if k_Y is a C^∞ Mercer kernel, we can apply [75, Theorem
 1382 D] on the class $\mathcal{G} := \{\hat{\psi}_n^\pi - \psi^\pi\} \subset L^2(P; \mathcal{H}_Y)$ to verify the conditions of Theorem 6 of Park and
 1383 Muandet [60].

1384 Then, by Lemma 11.3 and with consistency of the nuisance parameters, $\|\hat{\psi}_n^\pi - \psi^\pi\|_{L^2(P; \mathcal{H}_Y)} \rightarrow 0$,
 1385 and by their stochastic equicontinuity result in Corollary 8, [60], we readily have:

$$\|(P_n - P)(\hat{\psi}_n^\pi - \psi^\pi)\|_{\mathcal{H}_Y} \rightarrow 0 \quad \text{in probability.}$$

1386 Hence, $\|\mathcal{T}_n\|_{\mathcal{H}_Y} = o_P(n^{-1/2})$, completing the proof. \square

1387 11.3.2 Bounding the remainder term

1388 **Lemma 11.5.** (*Remainder term bound*). *Assumptions 1, 3, 15, 16, 17, 21, then*
 1389 $\|\mathcal{R}_n\|_{\mathcal{H}_Y} = O_p(r_C(n, \delta, b, c)r_{\pi_0}(n))$.

1390 *Proof.* From the definitions, the remainder term can be written as

$$\begin{aligned} \mathcal{R}_n &= \chi(\hat{P}_n) + P_0 \hat{\psi}_n^\pi - \chi(\pi) \\ &= \mathbb{E}_{P_0} \left[\frac{\pi(a | x)}{\hat{\pi}_0(a | x)} (\phi_Y(y) - \hat{\mu}_{Y|A,X}(a, x)) + \int \hat{\mu}_{Y|A,X}(a', x) \pi(da' | x) \right] \\ &\quad - \mathbb{E}_{P_0} \left[\frac{\pi(a | x)}{\pi_0(a | x)} (\phi_Y(y) - \mu_{Y|A,X}(a, x)) + \int \mu_{Y|A,X}(a', x) \pi(da' | x) \right] \\ &= \mathbb{E}_{P_0} \left[\frac{\pi(a | x)}{\hat{\pi}_0(a | x)} (\mathbb{E}[\phi_Y(y) | A = a, X = x] - \hat{\mu}_{Y|A,X}(a, x)) \right] \\ &\quad + \mathbb{E}_{P_0} \left[+ \int (\hat{\mu}_{Y|A,X}(a', x) - \mu_{Y|A,X}(a', x)) \pi(da' | x) \right] \\ &\quad - \mathbb{E}_{P_0} \left[\frac{\pi(a | x)}{\pi_0(a | x)} (\mathbb{E}[\phi_Y(y) | A = a, X = x] - \mu_{Y|A,X}(a, x)) \right] \\ &= \mathbb{E}_{P_0} \left[\frac{\pi(a | x)}{\hat{\pi}_0(a | x)} (\mu_{Y|A,X}(a, x) - \hat{\mu}_{Y|A,X}(a, x)) + \int (\hat{\mu}_{Y|A,X}(a', x) - \mu_{Y|A,X}(a', x)) \pi(da' | x) \right] \end{aligned}$$

1391 We can expand the expectation into the following:

$$\begin{aligned} \mathcal{R}_n &= \iint \left[\frac{\pi(a | x)}{\hat{\pi}_0(a | x)} (\mu_{Y|A,X}(a, x) - \hat{\mu}_{Y|A,X}(a, x)) \right] \pi_0(da | x) P_X(dx) \\ &\quad + \iint \int (\hat{\mu}_{Y|A,X}(a', x) - \mu_{Y|A,X}(a', x)) \pi(da' | x) \pi_0(da | x) P_X(dx) \\ &= \iint \frac{\pi_0(a | x)}{\hat{\pi}_0(a | x)} (\mu_{Y|A,X}(a, x) - \hat{\mu}_{Y|A,X}(a, x)) \pi(a | x) P_X(dx) \\ &\quad + \iint (\hat{\mu}_{Y|A,X}(a', x) - \mu_{Y|A,X}(a', x)) \pi(da' | x) P_X(dx) \\ &= \iint \left(\frac{\pi_0(a | x)}{\hat{\pi}_0(a | x)} - 1 \right) (\mu_{Y|A,X}(a, x) - \hat{\mu}_{Y|A,X}(a, x)) \pi(da | x) P_X(dx) \\ &= \iint \pi_0(a | x) \left(\frac{1}{\hat{\pi}_0(a | x)} - \frac{1}{\pi_0(a | x)} \right) (\mu_{Y|A,X}(a, x) - \hat{\mu}_{Y|A,X}(a, x)) \pi(da | x) P_X(dx) \end{aligned}$$

1392 By the Cauchy–Schwarz inequality, we have

$$\left\| \iint \pi_0(a | x) \left(\frac{1}{\hat{\pi}_0(a | x)} - \frac{1}{\pi_0(a | x)} \right) (\mu_{Y|A,X}(a, x) - \hat{\mu}_{Y|A,X}(a, x)) \pi(da | x) P_X(dx) \right\|_{\mathcal{H}_Y}$$

1393

$$\leq \left(\iint \pi_0^2(a | x) \left(\frac{1}{\hat{\pi}_0(a | x)} - \frac{1}{\pi_0(a | x)} \right)^2 \pi(da | x) P_X(dx) \right)^{1/2} \|\mu_{Y|A,X} - \hat{\mu}_{Y|A,X}\|_{\mathcal{H}_Y}.$$

1394

$$\leq \left\| \pi_0 \left(\frac{1}{\hat{\pi}_0} - \frac{1}{\pi_0} \right) \right\|_{L^2(\pi P_X)} \cdot \|\mu_{Y|A,X} - \hat{\mu}_{Y|A,X}\|_{\mathcal{H}_Y}$$

1395 If we write $r_{\pi_0}(n) = \left\| \frac{1}{\hat{\pi}_0} - \frac{1}{\pi_0} \right\|_{L^2(\pi P_X)}$ an error bound on the estimation of the inverse propen-
 1396 sity scores, and noting that by Theorem 19, the regression error on $\|\mu_{Y|A,X} - \hat{\mu}_{Y|A,X}\|_{\mathcal{H}_Y}$ is
 1397 $O_p(r_C(n, \delta, b, c))$, and we conclude the proof.

1398

□

1399 11.3.3 Consistency proof

1400 We are now in position to prove Theorem 6.

1401 *Proof.* The decomposition in Eq (37) provides:

$$\|\hat{\chi}_{dr} - \chi(P_0)\|_{\mathcal{H}} \leq \|\mathcal{T}_n\|_{\mathcal{H}} + \|\mathcal{S}_n\|_{\mathcal{H}} + \|\mathcal{R}_n\|_{\mathcal{H}}, \quad (39)$$

1402 The sample average \mathcal{S}_n converges to 0 by the central limit theorem for Hilbert-valued random
 1403 variables (see [83], see also Examples 1.4.7 and 1.8.5 in [84]), that is $\|\mathcal{S}_n\|_{\mathcal{H}_Y} = o_P(n^{-1/2})$.

1404 Then by combining the results of Lemma 11.4 (or Lemma 11.6) and Lemma 11.5, we obtain readily
 1405 that:

$$\|\hat{\chi}_{dr} - \chi(P_0)\|_{\mathcal{H}} = O_p\left(n^{-1/2} + r_C(n, \delta, b, c)r_{\pi_0}(n)\right).$$

1406

□

1407 11.4 Additional details on the cross-fitted estimator

1408 We now describe how cross-fitting [39, 36, 37, 85], can be used for our one-step estimator, following
 1409 Luedtke and Chung [40]. Let P_n^j denote the empirical distribution on the j -th fold of the samples and
 1410 let $\hat{P}_n^j \in \mathcal{P}$ denote an estimate of P_0 based on the remaining $j - 1$ folds. The cross-fitted one-step
 1411 estimator takes the form

$$\bar{\chi}_{dr}(\pi) = \frac{1}{k} \sum_{j=1}^k \left[\chi(\hat{P}_n^j) + P_n^j \hat{\psi}_n^j \right]. \quad (40)$$

1412 Using a similar decomposition as in Eq. (37), we obtain:

$$\bar{\chi}_{dr}(\pi) - \chi(\pi)(P) = \frac{1}{k} \sum_{j=1}^k (P_n^j - P) \psi^\pi + \frac{1}{k} \sum_{j=1}^k (P_n^j - P) (\hat{\psi}_n^{j,\pi} - \psi^\pi) \quad (41)$$

$$+ \frac{1}{k} \sum_{j=1}^k (\chi(\hat{P}_n^j) + P \hat{\psi}_n^{j,\pi} - \chi(\pi)(P)) \quad (42)$$

1413 Then, to prove the consistency of the estimator, we use the following triangular inequality.

$$\|\bar{\chi}_{dr}(\pi) - \chi(\pi)(P)\|_{\mathcal{H}} \leq \max_j \|\mathcal{T}_n^j\|_{\mathcal{H}} + \max_j \|\mathcal{S}_n^j\|_{\mathcal{H}} + \max_j \|\mathcal{R}_n^j\|_{\mathcal{H}}, \quad (43)$$

1414 where $\mathcal{S}_n^j := (P_n^j - P)\psi^\pi$, $\mathcal{T}_n^j := (P_n^j - P)(\hat{\psi}_n^{j,\pi} - \psi^\pi)$, $\mathcal{R}_n^j = \chi(\hat{P}_n^j) + P\hat{\psi}_n^{j,\pi} - \chi(\pi)$ We call
 1415 \mathcal{R}_n^j the remainder terms and \mathcal{T}_n^j the empirical process terms, $j \in \{1, k\}$.

1416 **Lemma 11.6.** [40, Lemma 3](Sufficient condition for negligible empirical process terms). Suppose
 1417 that χ is pathwise differentiable at P_0 with EIF ψ_0 . For each $j \in \{1, k\}$, $\|\psi_n^j - \psi_0\|_{L^2(P; \mathcal{H})} = o_p(1)$
 1418 implies that $\|\mathcal{T}_n^j\|_{\mathcal{H}} = o_p(n^{-1/2})$.

1419 Luedtke and Chung [81] proves this lemma via a conditioning argument that makes use of Cheby-
 1420 shev's inequality for Hilbert-valued random variables [86] and the dominated convergence theorem.

1421 Then, to prove the sufficient condition, we recall the result of Lemma 11.3, which now allows to
 1422 show that the cross-fitted CPME is consistent.

1423 12 Details and Analysis of the Doubly-Robust Test for the Distributional 1424 Policy Effect

1425 **Theorem 7** ((Asymptotic normality of the test statistic)). Suppose that the con-
 1426 ditions of Theorem 6 hold. Suppose that $\mathbb{E}_{P_0} [\|\varphi_{\pi, \pi'}(y, a, x)\|^4]$ is finite, that
 1427 $\mathbb{E}_{P_0} [\varphi_{\pi, \pi'}(y, a, x)] = 0$ and $\mathbb{E}_{P_0} [\langle \varphi_{\pi, \pi'}(y, a, x), \varphi_{\pi, \pi'}(y', a', x') \rangle] > 0$. Suppose also that
 1428 $r_{\pi_0}(n, \delta) \cdot r_C(n, \delta, b, c) = \mathcal{O}(n^{-1/2})$. Set $\lambda = n^{-1/(c+1/b)}$ and $m = \lfloor n/2 \rfloor$. then it follows that

$$T_{\pi, \pi'}^\dagger \xrightarrow{d} \mathcal{N}(0, 1).$$

1429 The proof uses the steps of Kim and Ramdas [61] and Martinez Taboada et al. [27], but is restated
 1430 as it leverage the theorems and assumptions relevant to CPME. Specifically we provide a result
 1431 similar on asymptotic normality to that of Luedtke and Chung [40, Theorem 2], which holds for the
 1432 non-cross fitted estimator.

Lemma 12.1. (Asymptotic linearity and weak convergence of the one-step estimator). Suppose that
 the conditions of Theorem 6 hold. Suppose also that $r_{\pi_0}(n, \delta) \cdot r_C(n, \delta, b, c) = \mathcal{O}(n^{-1/2})$. Set
 $\lambda = n^{-1/(c+1/b)}$ Under these conditions,

$$n^{1/2} [\hat{\chi}_{dr}(\pi) - \chi(\pi)] \rightsquigarrow \mathbb{H},$$

1433 where \mathbb{H} is a tight \mathcal{H} -valued Gaussian random variable that is such that, for each $h \in \mathcal{H}$, the
 1434 marginal distribution $\langle \mathbb{H}, h \rangle_{\mathcal{H}}$ is $N(0, E_0[\langle \psi^\pi(y, a, x), h \rangle_{\mathcal{H}}^2])$.

1435 This lemma can be obtained following the arguments of Luedtke and Chung [40], where the cross-
 1436 fitted estimator essentially requires for $j \in \{1, 2\}$, $\mathcal{R}_n^j = o_p(n^{-1/2})$ and $\mathcal{T}_n^j = o_p(n^{-1/2})$ to apply
 1437 Slutsky's lemma and a central limit theorem for Hilbert-valued random variables [84].

1438 *Proof.* We split the dataset $\{(x_i, a_i, y_i)\}_{i=1}^n$ into two disjoint parts:

$$\mathcal{D} = \{(x_i, a_i, y_i)\}_{i=1}^m, \quad \tilde{\mathcal{D}} = \{(x_j, a_j, y_j)\}_{j=m+1}^n.$$

Further,

$$f_{\pi, \pi'}(y, a, x) = \frac{1}{n-m} \sum_{j=m+1}^n \langle \varphi_{\pi, \pi'}(y, a, x), \varphi_{\pi, \pi'}(y_j, a_j, x_j) \rangle, \quad T_{\pi, \pi'} = \frac{\sqrt{n} \bar{f}_{\pi, \pi'}}{S_{\pi, \pi'}}$$

where $\bar{f}_{\pi, \pi'}$ and $S_{\pi, \pi'}^2$ are the empirical mean and variance respectively:

$$\bar{f}_{\pi, \pi'} = \frac{1}{n} \sum_{i=1}^n f_{\pi, \pi'}(y_i, a_i, x_i), \quad S_{\pi, \pi'}^2 = \frac{1}{n} \sum_{i=1}^n (f_{\pi, \pi'}(y_i, a_i, x_i) - \bar{f}_{\pi, \pi'})^2$$

1439 We define the test statistic using the doubly robust estimators $\hat{\varphi}_{\pi, \pi'}$ and $\tilde{\varphi}_{\pi, \pi'}$, which are computed
 1440 respectively from \mathcal{D} and $\tilde{\mathcal{D}}$:

$$f_{\pi, \pi'}^\dagger(y_i, a_i, x_i) := \frac{1}{n-m} \sum_{j=m+1}^n \langle \hat{\varphi}_{\pi, \pi'}(y_i, a_i, x_i), \tilde{\varphi}_{\pi, \pi'}(y_j, a_j, x_j) \rangle,$$

1441

$$\bar{f}_{\pi,\pi'}^\dagger := \frac{1}{m} \sum_{i=1}^m f_{\pi,\pi'}^\dagger(Z_i), \quad (S_{\pi,\pi'}^\dagger)^2 := \frac{1}{m} \sum_{i=1}^m \left(f_{\pi,\pi'}^\dagger(Z_i) - \bar{f}_{\pi,\pi'}^\dagger \right)^2,$$

1442

$$T_{\pi,\pi'}^\dagger := \frac{\sqrt{m} \bar{f}_{\pi,\pi'}^\dagger}{S_{\pi,\pi'}^\dagger}.$$

1443 As [27, 61], the asymptotic normality results in four steps:

1444

$$1. \text{ Consistency of the mean: } m\bar{f}_{\pi,\pi'}^\dagger = m\bar{f}_{\pi,\pi'} + o_{\mathbb{P}}(1)$$

1445

$$2. \text{ Consistency of the variance: } m(S_{\pi,\pi'}^\dagger)^2 = m(S_{\pi,\pi'})^2 + o_{\mathbb{P}}(1)$$

1446

$$3. \text{ Bounded variance under conditional law: } \frac{1}{\mathbb{E}[mf_{\pi,\pi'}^\dagger(Z)^2 | \mathcal{D}_2]} = \mathcal{O}_{\mathbb{P}}(1)$$

1447

$$4. \text{ Conclude with asymptotic normality: } T_{\pi,\pi'}^\dagger \xrightarrow{d} \mathcal{N}(0, 1)$$

1448 **Consistency of the mean** We follow the same outline as Martinez Taboada et al. [27] did, using
 1449 Lemma 12.1 for the asymptotic normality of $\varphi_{\pi,\pi'}$.

1450 **Consistency of the variance** We follow the same outline as Martinez Taboada et al. [27] did.

1451 **Bounded variance** We now show that the denominator in the normalization of $T_{\pi,\pi'}^\dagger$ is bounded
 1452 away from zero in probability:

$$\frac{1}{\mathbb{E} \left[mf_{\pi,\pi'}^2(Z) \mid \tilde{\mathcal{D}} \right]} = \mathcal{O}_P(1).$$

For compactness, we define:

$$\tau = \frac{1}{\sqrt{m}} \sum_{i=1}^m \varphi_{\pi,\pi'}(y_i, a_i, x_i), \quad \gamma = \frac{1}{\sqrt{n-m}} \sum_{j=m+1}^n \varphi_{\pi,\pi'}(y_j, a_j, x_j),$$

1453 and

$$\hat{\tau} = \frac{1}{\sqrt{m}} \sum_{i=1}^m \hat{\varphi}_{\pi,\pi'}(y_i, a_i, x_i), \quad \tilde{\gamma} = \frac{1}{\sqrt{n-m}} \sum_{j=m+1}^n \tilde{\varphi}_{\pi,\pi'}(y_j, a_j, x_j)$$

1454 so that:

1455

$$m\bar{f}_{\pi,\pi'} = \langle \tau, \gamma \rangle, \quad m\bar{f}_{\pi,\pi'}^\dagger = \langle \hat{\tau}, \tilde{\gamma} \rangle,$$

1456

$$(\sqrt{m}S_{\pi,\pi'})^2 = \sum_{i=1}^m \langle \varphi_{\pi,\pi'}(y_i, a_i, x_i), \gamma \rangle^2 - m(\bar{f}_{\pi,\pi'})^2,$$

$$(\sqrt{m}S_{\pi,\pi'}^\dagger)^2 = \sum_{i=1}^m \langle \hat{\varphi}_{\pi,\pi'}(y_i, a_i, x_i), \tilde{\gamma} \rangle^2 - m(\bar{f}_{\pi,\pi'}^\dagger)^2.$$

1457 Recall that $f_{\pi,\pi'}(Z) = \langle \varphi_{\pi,\pi'}(Z), \gamma \rangle$, and that $\gamma = \frac{1}{\sqrt{n-m}} \sum_{j=m+1}^n \varphi_{\pi,\pi'}(Z_j) \in \mathcal{H}$ is a random
 1458 element measurable with respect to $\tilde{\mathcal{D}}$. Conditional on $\tilde{\mathcal{D}}$, the variance of the test statistic is:

$$\mathbb{E} \left[mf_{\pi,\pi'}^2(Z) \mid \tilde{\mathcal{D}} \right] = \langle C\gamma, \gamma \rangle,$$

1459 where $C = \mathbb{E}[\varphi(Z) \otimes \varphi(Z)]$ is the covariance operator over \mathcal{H} , which is compact, self-adjoint and
 1460 positive semi-definite.

1461 Using the eigendecomposition of C (see Section 9.1), we write:

$$C = \sum_{j=1}^{\infty} \lambda_j v_j \otimes v_j, \quad \gamma = \sum_{j=1}^{\infty} \beta_j v_j,$$

1462 so that

$$\mathbb{E} \left[m f_{\pi, \pi'}^2(Z) \mid \tilde{\mathcal{D}} \right] = \sum_{j=1}^{\infty} \lambda_j \beta_j^2.$$

1463 From Assumption 16, we know that the eigenvalues satisfy $\lambda_j \leq C j^{-b}$ for some $b \geq 1$. This decay
1464 implies that the kernel is not degenerate and the operator C has at least one strictly positive eigenvalue:
1465 $\lambda_1 > 0$.

1466 Moreover, by Lemma 12.1 and the Central Limit Theorem in separable Hilbert spaces [83], the
1467 limiting distribution of γ is Gaussian:

$$\gamma \xrightarrow{d} \sum_{j=1}^{\infty} \sqrt{\lambda_j} N_j v_j, \quad \text{where } N_j \sim \mathcal{N}(0, 1).$$

1468 Hence,

$$\beta_1 = \langle \gamma, v_1 \rangle \xrightarrow{d} \sqrt{\lambda_1} N_1, \quad \Rightarrow \quad \lambda_1 \beta_1^2 \xrightarrow{d} \lambda_1^2 N_1^2.$$

1469 Therefore, the conditional variance is lower bounded:

$$\mathbb{E} \left[m f_{\pi, \pi'}^2(Z) \mid \tilde{\mathcal{D}} \right] = \sum_{j=1}^{\infty} \lambda_j \beta_j^2 \geq \lambda_1 \beta_1^2 \xrightarrow{d} \lambda_1^2 N_1^2.$$

1470 This shows that the variance remains bounded away from zero in probability. More formally, for any
1471 $\epsilon > 0$, we can find $M > 0$ such that:

$$\mathbb{P} \left(\frac{1}{\mathbb{E} \left[m f_{\pi, \pi'}^2(Z) \mid \tilde{\mathcal{D}} \right]} > M \right) < \epsilon.$$

1472 Hence,

$$\frac{1}{\mathbb{E} \left[m f_{\pi, \pi'}^2(Z) \mid \tilde{\mathcal{D}} \right]} = \mathcal{O}_P(1).$$

1473 **Asymptotic normality** We now conclude the asymptotic normality of $T_{\pi, \pi'}^\dagger$, following Mar-
1474 tinez Taboada et al. [27]. Suppose that $\mathbb{E}_{P_0} [\|\varphi_{\pi, \pi'}(y, a, x)\|^4]$ is finite, that $\mathbb{E}_{P_0} [\varphi_{\pi, \pi'}(y, a, x)] = 0$
1475 and $\mathbb{E}_{P_0} [\langle \varphi_{\pi, \pi'}(y, a, x), \varphi_{\pi, \pi'}(y', a', x') \rangle] > 0$, from Kim and Ramdas [61], we have:

$$\frac{\sqrt{m} \bar{f}_{\pi, \pi'}}{\sqrt{\mathbb{E}[f_{\pi, \pi'}^2(Z) \mid \tilde{\mathcal{D}}]}} \xrightarrow{d} \mathcal{N}(0, 1), \quad \frac{S_{\pi, \pi'}^2}{\mathbb{E}[f_{\pi, \pi'}^2(Z) \mid \tilde{\mathcal{D}}]} \xrightarrow{p} 1.$$

1476 Using the previous steps, we have:

$$m \bar{f}_{\pi, \pi'}^\dagger = m \bar{f}_{\pi, \pi'} + o_P(1), \quad m S_{\pi, \pi'}^{\dagger 2} = m S_{\pi, \pi'}^2 + o_P(1),$$

1477 which implies:

$$\frac{m \bar{f}_{\pi, \pi'}^\dagger}{\sqrt{\mathbb{E}[m f_{\pi, \pi'}^2(Z) \mid \tilde{\mathcal{D}}]}} = \frac{m \bar{f}_{\pi, \pi'}}{\sqrt{\mathbb{E}[m f_{\pi, \pi'}^2(Z) \mid \tilde{\mathcal{D}}]}} + o_P(1) \xrightarrow{d} \mathcal{N}(0, 1),$$

1478 Moreover,

$$\frac{m S_{\pi, \pi'}^{\dagger 2}}{\mathbb{E}[m f_{\pi, \pi'}^2(Z) \mid \tilde{\mathcal{D}}]} \xrightarrow{p} 1.$$

1479 Taking square roots on both sides (which preserves convergence in probability by the continuous
1480 mapping theorem), we obtain:

$$\frac{\sqrt{\mathbb{E}[mf_{\pi,\pi'}^2(Z) \mid \tilde{\mathcal{D}}]}}{\sqrt{m}S_{\pi,\pi'}^\dagger} \xrightarrow{p} 1.$$

1481 By Slutsky's theorem by combining the last two:

$$T_{\pi,\pi'}^\dagger = \frac{\sqrt{m}f_{\pi,\pi'}^\dagger}{S_{\pi,\pi'}^\dagger} \xrightarrow{d} \mathcal{N}(0, 1). \quad \square$$

1482

□

1483 13 Details and Analysis of the sampling from the counterfactual distribution

1484 **Proposition 9** ((Convergence of MMD of herded samples, weak convergence to the counter-
1485 factual outcome distribution).). *Suppose the conditions of Lemma 4.1 and Assumption 8 hold.*
1486 *Let $(\tilde{y}_{dr,j})$ and $\tilde{P}_{Y,dr}^m$ (resp. $(\tilde{y}_{pi,j})$, $\tilde{P}_{Y,pi}^m$) be generated from $\hat{\chi}_{dr}(\pi)$ (resp. $\hat{\chi}_{pi}(\pi)$) via Al-*
1487 *gorithm 2. Then, with high probability, $\text{MMD}(\tilde{P}_{Y,pi}^m, \nu(\pi)) = O_p(r_C(n, b, c) + m^{-1/2})$ and*
1488 *$\text{MMD}(\tilde{P}_{Y,dr}^m, \nu(\pi)) = O_p(n^{-1/2} + r_{\pi_0}(n)r_C(n, b, c) + m^{-1/2})$. Moreover, $(\tilde{y}_{dr,j}) \rightsquigarrow \nu(\pi)$ and*
1489 *$(\tilde{y}_{pi,j}) \rightsquigarrow \nu(\pi)$.*

1490 *Proof.* Fix $\pi \in \Pi$. By Theorem 6, the estimated embedding $\hat{\chi}_{dr}(\pi)$ satisfies:

$$\|\hat{\chi}_{dr}(\pi) - \chi(\pi)\|_{\mathcal{H}_Y} = O_p\left(n^{-1/2} + r_{\pi_0}(n) \cdot r_C(n, b, c)\right).$$

1491 Let $\{\tilde{y}_t\}_{t=1}^m$ be the herded samples generated from $\hat{\chi}_{dr}(\pi)$ using Algorithm 2. According to Bach
1492 et al. [87, Section 4.2], the empirical mean embedding of these samples approximates $\hat{\chi}_{dr}(\pi)$ at rate:

$$\left\| \hat{\chi}_{dr}(\pi) - \frac{1}{m} \sum_{t=1}^m \phi_Y(\tilde{y}_t) \right\|_{\mathcal{H}_Y} = \mathcal{O}(m^{-1/2}).$$

1493 By the triangle inequality:

$$\left\| \frac{1}{m} \sum_{t=1}^m \phi_Y(\tilde{y}_t) - \chi(\pi) \right\|_{\mathcal{H}_Y} = O_p\left(n^{-1/2} + r_{\pi_0}(n)r_C(n, b, c) + m^{-1/2}\right).$$

1494 By definition of MMD and the reproducing property, we have:

$$\text{MMD}(\tilde{P}_Y^m, \nu(\pi)) = \left\| \frac{1}{m} \sum_{t=1}^m \phi_Y(\tilde{y}_t) - \chi(\pi) \right\|_{\mathcal{H}_Y},$$

1495 so the same rate applies.

1496 For the plug-in estimator $\hat{\chi}_{pi}(\pi)$, which does not involve nuisance estimation, we obtain:

$$\|\hat{\chi}_{pi}(\pi) - \chi(\pi)\|_{\mathcal{H}_Y} = O_p(r_C(n, b, c)),$$

1497 yielding, with the same arguments

$$\text{MMD}(\tilde{P}_{Y,pi}^m, \nu(\pi)) = O_p(r_C(n, b, c) + m^{-1/2}).$$

1498 Finally, weak convergence of the empirical measures \tilde{P}_Y^m to $\nu(\pi)$ follows from convergence in MMD
1499 norm with a characteristic kernel; see Simon-Gabriel et al. [62, Theorem 1.1] and Sriperumbudur
1500 [88]. □

14 Experiment details

In this Appendix we provide additional details on the simulated settings as well as additional experiment results.

14.1 Testing experiments

We are given a logged dataset $\mathcal{D}_{\text{init}} = \{(x_i, a_i, y_i)\}_{i=1}^n \sim P_0$, collected under a logging policy π_0 . For two target policies π and π' , the objective is to test the null hypothesis:

$$H_0 : \nu(\pi) = \nu(\pi'), \quad \text{vs.} \quad H_1 : \nu(\pi) \neq \nu(\pi'),$$

where $\nu(\pi)$ and $\nu(\pi')$ denote the counterfactual distributions of outcomes under π and π' , respectively.

14.1.1 Baseline

We use baselines to evaluate the ability of our framework to detect differences in counterfactual outcome distributions induced by different target policies, compared to alternative approaches.

Kernel Policy Test (KPT). An adaptation of the kernel treatment effect test of Muandet et al. [16], extended to the OPE setting. It tests whether the counterfactual distributions $\nu(\pi)$ and $\nu(\pi')$ differ by comparing reweighted outcome samples using the maximum mean discrepancy (MMD). The key idea is to view both outcome distributions as being implicitly represented by importance-weighted samples from the logging distribution.

Given two importance weight vectors w_π and $w_{\pi'}$ corresponding to the target policies π and π' , respectively, the test computes the unbiased squared MMD statistic:

$$\widehat{\text{MMD}}_u^2 = \frac{1}{n(n-1)} \sum_{i \neq j} \left[w_i^\pi w_j^\pi k(y_i, y_j) + w_i^{\pi'} w_j^{\pi'} k(y_i, y_j) - 2w_i^\pi w_j^{\pi'} k(y_i, y_j) \right],$$

where $k(y_i, y_j)$ is a positive definite kernel on the outcome space (typically RBF). To obtain a p -value, KPT uses a permutation-based null distribution. It repeatedly permutes the correspondence between samples and their importance weights (thus preserving the outcome data while randomizing their "assignment") and recomputes the MMD statistic under each permutation. The p -value is estimated as the proportion of permuted statistics that exceed the observed MMD. As Muandet et al. [16], we use 10000 permutations.

Average Treatment Effect Test (PT-linear). A simple variant of KPT using linear kernels, testing only for differences in means. It serves as a reference for detecting average treatment differences.

Doubly Robust Kernel Policy Test (DR-KPT). We construct a doubly robust test statistic based on the difference of efficient influence functions:

$$T_{\pi, \pi'}^\dagger = \frac{\sqrt{m} \bar{f}_{\pi, \pi'}^\dagger}{S_{\pi, \pi'}^\dagger},$$

where \bar{f}^\dagger is the empirical mean of pairwise inner products of influence function differences across data splits, and S^\dagger the empirical standard deviation. The null is rejected when $T_{\pi, \pi'}^\dagger$ exceeds a standard normal threshold.

14.1.2 Model selection and tuning

We repeat each experiment 100 times and report test powers with 95% confidence intervals. For DR-KPT and KPT, the kernel k_Y is RBF. For DR-KPT the regularization parameter λ is selected via 3-fold cross-validation in the range $\{10^{-4}, \dots, 10^0\}$, as done in [16]. We use the median heuristic for the lengthscales of the kernel $k_{\mathcal{A}}$, $k_{\mathcal{X}}$ and k_Y .

1536 14.1.3 Simulated Setting

1537 The experiments are conducted in a synthetic continuous treatment setting. Covariates $x_i \in \mathbb{R}^d$
 1538 are sampled independently from a multivariate standard normal distribution $\mathcal{N}(0, I_d)$. Treatments
 1539 $a_i \in \mathbb{R}$ are drawn from a Gaussian logging policy $\pi_0(a | x) = \mathcal{N}(x^\top w, 1)$, where the weight vector
 1540 is fixed as $w = \frac{1}{\sqrt{d}} \mathbf{1}_d$. Outcomes are generated according to a linear outcome model with additive
 1541 noise:

$$y_i = x_i^\top \beta + \gamma a_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$

1542 where $\beta \in \mathbb{R}^d$ is a linearly increasing vector and $\gamma \in \mathbb{R}$ controls the treatment effect strength.

1543 We evaluate four distinct scenarios, each specifying a different relationship between the target policies
 1544 π , π' , and the logging policy π_0 . These scenarios are designed to induce progressively more complex
 1545 shifts in the treatment distribution, affecting the downstream outcome distribution. We set the
 1546 covariate dimension to $d = 5$, $\gamma = 1$ and evaluate β in the grid $\beta = [0.1, 0.2, 0.3, 0.4, 0.5]$. β
 1547 is taken at different values across samples to reflect heterogeneity in user features and outcome
 1548 interactions.

1549 **Scenario I (Null).** This is the calibration setting in which $\pi = \pi'$. The two policies generate treat-
 1550 ments from the same Gaussian distribution with shared mean and variance, ensuring no counterfactual
 1551 distributional shift. Under the null hypothesis, we expect all tests to maintain the nominal Type I
 1552 error rate.

1553 **Scenario II (Mean Shift).** Here, the target policy π remains identical to the logging policy, while
 1554 the alternative policy π' is a Gaussian with the same variance but a shifted mean. Specifically, π' uses
 1555 a weight vector $w' = w + \delta$, with $\delta = 2 \cdot \mathbf{1}_d$. This results in a systematic mean shift in treatment
 1556 assignment, causing a change in the marginal distribution of outcomes through the linear outcome
 1557 model. This tests whether the methods can detect simple, mean-level differences in counterfactual
 1558 outcomes.

1559 **Scenario III (Mixture).** In this case, the policy π remains a standard Gaussian as in previous
 1560 scenarios, while the alternative π' is a 50/50 mixture of two Gaussian policies with opposing shifts in
 1561 their means: $w_1 = w + \mathbf{1}_d$, $w_2 = w - \mathbf{1}_d$. Although the resulting treatment distribution is bimodal,
 1562 its overall mean matches that of π . This scenario introduces a change in higher-order structure (e.g.,
 1563 variance, modality) without altering the first moment, allowing us to test whether the methods detect
 1564 distributional differences beyond the mean.

1565 **Scenario IV (Shifted Mixture).** This is the most complex scenario. As in Scenario III, the
 1566 alternative policy π' is a mixture of two Gaussian components, but this time only one component is
 1567 shifted: $w_1 = w + 2 \cdot \mathbf{1}_d$, $w_2 = w$. The resulting treatment distribution under π' differs from π in
 1568 both mean and higher-order moments. This scenario combines characteristics of Scenarios II and III
 1569 and evaluates whether the tests remain sensitive to subtle and structured counterfactual shifts.

1570 Across all scenarios, we generate $n = 1000$ samples per run and estimate importance weights for π
 1571 and π' using fitted models based on the observed data. Specifically, we fit a linear regression model
 1572 to the logged treatments T as a function of the covariates X to estimate the mean of the logging
 1573 policy π_0 , and evaluate its Gaussian density to obtain estimated propensities. This experimental
 1574 design enables evaluation of the calibration and power of distributional tests under a range of realistic
 1575 divergences.

1576 In all scenarios (Tables 1–4), **DR-KPT consistently demonstrates the best computational effi-**
 1577 **ciency**, with runtimes typically two orders of magnitude lower than both KPT and PT-linear. This
 1578 efficiency stems from the closed-form structure of its test statistic, which avoids repeated resampling
 1579 or kernel matrix permutations. In contrast, KPT relies on costly permutation-based MMD calculations,
 1580 and PT-linear, while simpler, still requires repeated reweighting. For readability and to emphasize
 1581 this computational advantage, we reorder the tables so that DR-KPT appears in the last row of each
 1582 scenario.

1583 Next, to empirically illustrate the benefits of sample-splitting in the test statistic provided in Section
 1584 5.1, we provide below in Figure 4 the same histograms as given in Figure 1. Concretely, instead of
 1585 splitting the samples in m and $n - m$, we use all the samples in the definition of $T_{\pi, \pi'}^\dagger, f_{\pi, \pi'}^\dagger(y_i, a_i, x_i)$

Table 1: Average runtime (in seconds) for Scenario I. Values are reported as mean \pm std over 100 runs.

Method	100	150	200	250	300	350	400
KPT	0.495 \pm 0.070	0.740 \pm 0.039	1.134 \pm 0.081	1.623 \pm 0.075	2.257 \pm 0.074	3.204 \pm 0.118	4.180 \pm 0.136
PT-linear	0.592 \pm 0.061	0.774 \pm 0.038	1.060 \pm 0.051	1.553 \pm 0.076	2.373 \pm 0.202	3.384 \pm 0.160	4.358 \pm 0.251
DR-KPT	0.004 \pm 0.005	0.007 \pm 0.004	0.010 \pm 0.009	0.008 \pm 0.002	0.013 \pm 0.007	0.025 \pm 0.023	0.019 \pm 0.007

Table 2: Average runtime (in seconds) for Scenario II. Values are reported as mean \pm std over 100 runs.

Method	100	150	200	250	300	350	400
KPT	0.559 \pm 0.044	0.794 \pm 0.040	1.173 \pm 0.063	1.764 \pm 0.093	2.301 \pm 0.085	3.342 \pm 0.126	4.204 \pm 0.182
PT-linear	0.486 \pm 0.035	0.767 \pm 0.037	1.071 \pm 0.030	1.630 \pm 0.062	2.405 \pm 0.182	3.738 \pm 0.251	4.767 \pm 0.228
DR-KPT	0.004 \pm 0.003	0.007 \pm 0.005	0.012 \pm 0.006	0.014 \pm 0.008	0.023 \pm 0.012	0.022 \pm 0.009	0.027 \pm 0.031

and in the test statistics in Eq. (14). As we can see, the resulting distribution is not normal, the QQ plot does not conclude and the test is not at all calibrated.

14.2 Sampling experiments

We study whether our estimated counterfactual policy mean embeddings (CPMEs) can be used to generate samples that approximate the true counterfactual outcome distribution. Formally, given a logged dataset $\mathcal{D}_{\text{init}} = \{(x_i, a_i, y_i)\}_{i=1}^n \sim P_0$ and a target policy π , we aim to generate samples $\{\tilde{y}_j\}_{j=1}^m$ such that their empirical distribution \tilde{P}_Y^m approximates the counterfactual outcome distribution $\nu(\pi)$ under π .

14.2.1 Procedure

We employ kernel herding to deterministically sample from the estimated embedding $\hat{\chi}(\pi)$ in RKHS. The algorithm sequentially selects samples $\tilde{y}_1, \dots, \tilde{y}_m$ that approximate the target embedding via greedy maximization:

$$\tilde{y}_t = \arg \max_{y \in \mathcal{Y}} \left\{ \hat{\chi}(\pi)(y) - \frac{1}{t-1} \sum_{\ell=1}^{t-1} k_{\mathcal{Y}}(\tilde{y}_\ell, y) \right\},$$

where $k_{\mathcal{Y}}$ is a universal kernel on the outcome space.

Since no comparable baselines for counterfactual sampling are available in the literature, we focus on comparing the quality of samples generated from two estimators of $\chi(\pi)$: the plug-in estimator and the doubly robust estimator. Both versions yield distinct herded samples, which we evaluate against ground truth samples generated under the target policy π .

14.2.2 Model selection and tuning

To report the distance metrics, we repeat each experiment 100 times and report the associated metric with 95% confidence intervals. For both plug-in and DR estimators, the kernel $k_{\mathcal{Y}}$ is RBF and the regularization parameter λ is selected via 3-fold cross-validation in the range $\{10^{-4}, \dots, 10^0\}$, as done in the sampling experiments of Muandet et al. [16]. We use the median heuristic for the lengthscales of the kernel $k_{\mathcal{A}}$, $k_{\mathcal{X}}$ and $k_{\mathcal{Y}}$.

14.2.3 Simulated Setting

We simulate logged data under different outcome models and logging policies. Covariates $x_i \in \mathbb{R}^d$ are sampled from a standard Gaussian distribution. Treatments $a_i \in \mathbb{R}$ are drawn either from a uniform distribution or from a logistic policy whose parameters depend on x_i . Outcomes y_i are then generated via one of the following nonlinear functions:

$$\text{Nonlinear: } y = \sin(x^\top \beta) + a^2 + \varepsilon, \quad \text{Quadratic: } y = (x^\top \beta)^2 + a^2 + \varepsilon,$$

where β is a fixed coefficient vector and $\varepsilon \sim \mathcal{N}(0, 1)$. For each synthetic setup, we generate logged data under the logging policy π_0 and obtain oracle samples under the target policy π for evaluation.

Table 3: Average runtime (in seconds) for Scenario III. Values are reported as mean \pm std over 100 runs.

Method	100	150	200	250	300	350	400
KPT	0.523 \pm 0.063	0.836 \pm 0.025	1.161 \pm 0.018	1.596 \pm 0.008	2.157 \pm 0.042	3.174 \pm 0.014	4.044 \pm 0.021
PT-linear	0.505 \pm 0.052	0.802 \pm 0.015	1.134 \pm 0.013	1.577 \pm 0.014	2.142 \pm 0.043	3.181 \pm 0.041	4.051 \pm 0.024
DR-KPT	0.004 \pm 0.003	0.008 \pm 0.009	0.011 \pm 0.005	0.015 \pm 0.009	0.020 \pm 0.010	0.025 \pm 0.013	0.025 \pm 0.014

Table 4: Average runtime (in seconds) for Scenario IV. Values are reported as mean \pm std over 100 runs.

Method	100	150	200	250	300	350	400
KPT	0.548 \pm 0.065	0.839 \pm 0.014	1.171 \pm 0.012	1.611 \pm 0.013	2.176 \pm 0.042	3.239 \pm 0.032	4.142 \pm 0.032
PT-linear	0.523 \pm 0.062	0.831 \pm 0.008	1.160 \pm 0.014	1.626 \pm 0.058	2.385 \pm 0.127	3.282 \pm 0.115	4.153 \pm 0.043
DR-KPT	0.004 \pm 0.005	0.009 \pm 0.007	0.015 \pm 0.008	0.015 \pm 0.010	0.018 \pm 0.010	0.023 \pm 0.011	0.025 \pm 0.015

We set the covariate dimension to $d = 5$ and evaluate β in the grid $\beta = [0.1, 0.2, 0.3, 0.4, 0.5]$. β is taken at different values across samples to reflect heterogeneity in user features and outcome interactions.

Figure 5 illustrates the counterfactual outcome distributions recovered via kernel herding using both PI-CPME and DR-CPME estimators under different logging policies and outcome functions.

To assess the fidelity of the sampled distributions, we compare the empirical distribution \tilde{P}_Y^m of herded samples to the true counterfactual distribution using two metrics:

- **Wasserstein distance** between the sampled and ground truth outcomes,
- **Maximum Mean Discrepancy (MMD)** with a Gaussian kernel.

Table 5: Wasserstein distance between herded samples and samples from the oracle counterfactual distribution

Method	logistic-nonlinear	logistic-quadratic	uniform-nonlinear	uniform-quadratic
Plug-in	1.29e-01 \pm 2.6e-01	1.41e-01 \pm 4.9e-02	9.08e-02 \pm 3.7e-01	6.78e-02 \pm 1.9e-02
DR	8.60e-02 \pm 2.2e-02	1.36e-01 \pm 3.9e-02	5.00e-02 \pm 1.5e-02	6.63e-02 \pm 1.6e-02

Results in Table 5, 6 show that samples obtained from the doubly robust estimator exhibit lower discrepancy to the oracle distribution.

14.3 Off-policy evaluation

We are given a dataset of n i.i.d. *logged* observations $\{(x_i, a_i, y_i)\}_{i=1}^n \sim P_0$. Given only this logged data from P_0 , the goal of *off-policy evaluation* is to estimate $R(\pi)$, the expected outcomes induced by a target policy π belonging to the policy set Π :

$$R(\pi) = \mathbb{E}_{P_\pi} [Y(a)]. \quad (44)$$

After identification, the risk of the policy simply boils down to $R(\pi) = \mathbb{E}_{P_\pi} [Y(a)]$, and the CPME $\chi(\pi) = \mathbb{E}_{P_\pi} [\phi_Y(Y(a))]$ describes the risk when the feature map $\phi_Y = y$ is linear.

14.3.1 Baselines

We compare our method against the following baseline estimators on synthetic datasets.

Direct Method (DM). The direct method [32] fits a regression model $\hat{\eta} : \mathcal{U} \times \mathcal{A} \rightarrow \mathbb{R}$ on the logged dataset $\mathcal{D}_{\text{init}} = \{(y_i, a_i, x_i)\}_{i=1}^n$, and estimates the expected reward under a target policy π as

$$\hat{R}_{\text{DM}}(\pi) = \frac{1}{n} \sum_{i=1}^n \int \hat{\eta}(x_i, a) \pi(a|x_i) da.$$

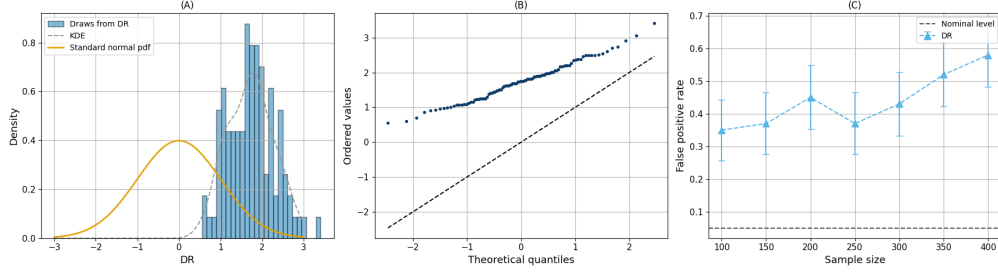


Figure 4: Illustration of 100 simulations of the non-sample-split DR-KPT under the null: (A) Histogram of DR-KPT alongside the pdf of a standard normal for $n = 400$, (B) Normal Q-Q plot of DR-KPT for $n = 400$, (C) False positive rate of DR-KPT against different sample sizes.

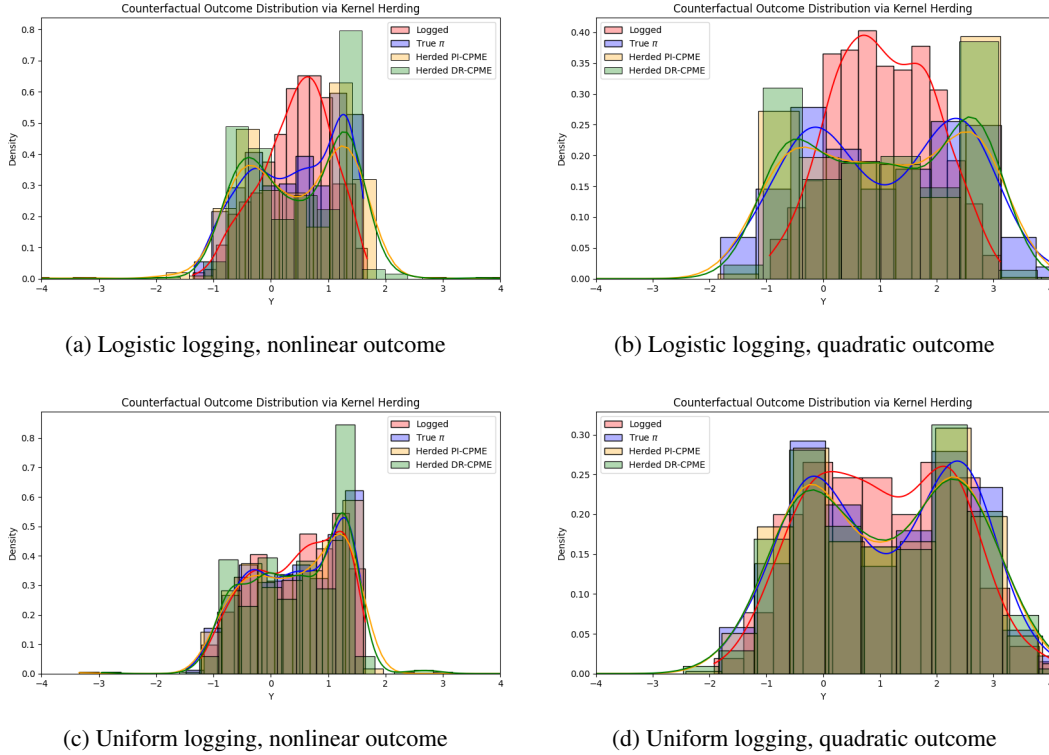


Figure 5: Counterfactual outcome distributions estimated via kernel herding from PI-CPME and DR-CPME samples, compared to the logged and true outcome distributions.

1637 Since the evaluated policy differs from the logging policy $\pi_0 \neq \pi$, a covariate shift is induced over the
 1638 joint space $\mathcal{A} \times \mathcal{X}$. It is well known that under the covariate shift, a parametric regression model may
 1639 produce a significant bias [89]. To demonstrate this, we use a 3-layer feedforward neural network as
 1640 the regressor and call it DM-NN.

1641 **Weighted Inverse Propensity Score (wIPS).** This estimator reweights logged rewards using
 1642 inverse propensity scores [80]:

$$\hat{R}_{\text{wIPS}}(\pi) = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}, \quad w_i = \frac{\pi(a_i | u_i)}{\pi_0(a_i | u_i)}.$$

1643 This estimator is unbiased when the true propensities are known.

Table 6: MMD distance between herded samples and samples from the oracle counterfactual distribution

Method	logistic-nonlinear	logistic-quadratic	uniform-nonlinear	uniform-quadratic
Plug-in	$1.11\text{e-}03 \pm 5.9\text{e-}03$	$9.85\text{e-}04 \pm 6.0\text{e-}04$	$1.92\text{e-}04 \pm 1.2\text{e-}03$	$3.31\text{e-}04 \pm 2.5\text{e-}04$
DR	$4.38\text{e-}04 \pm 3.6\text{e-}04$	$9.80\text{e-}04 \pm 6.0\text{e-}04$	$6.49\text{e-}05 \pm 4.4\text{e-}05$	$3.51\text{e-}04 \pm 2.5\text{e-}04$

Doubly Robust (DR). The DR estimator [32] combines the two previous methods, that is $\hat{\eta}$ and w_i using:

$$\hat{R}_{\text{DR}} = \frac{1}{n} \sum_{i=1}^n \left(\int \hat{\eta}(x_i, a) \pi(a|x_i) da + w_i(y_i - \hat{\eta}(x_i, a_i)) \right),$$

and remains consistent if either $\hat{\eta}$ or π_0 is correctly specified. We use the same parametrization for $\hat{\eta}$ as we do for the DM method and therefore call this doubly robust approach DR-NN.

Counterfactual Policy Mean Embeddings (CPME). We define a product kernel $k_{\mathcal{AX}}((a, x), (a', x')) = k_{\mathcal{A}}(a, a')k_{\mathcal{X}}(x, x')$, with Gaussian kernels on a and x . The outcome kernel $k_{\mathcal{Y}}$ is linear.

Relation to DM. When $\hat{\eta}$ is fit via kernel ridge regression (see Exemple 9.1), the DM estimate becomes:

$$\hat{R}_{\text{DM}}(\pi) \approx Y^\top (K + n\lambda I)^{-1} \cdot \frac{1}{n} \sum_{i=1}^n k_{\mathcal{AX}}(\tilde{a}_i, x_i)$$

where $K_{ij} = k_{\mathcal{AX}}((a_i, x_i), (a_j, x_j))$, and $\tilde{a}_i \sim \pi(\cdot | x_i)$. This matches the CME form proposed in [16], showing that CME/CPME is as a nonparametric version of the DM. Because kernel methods mitigate covariate shift, CPME is consistent and asymptotically unbiased. We will therefore refer to the plug-in $\hat{\chi}_{\text{pi}}(\pi)$ and the doubly robust $\hat{\chi}_{\text{dr}}(\pi)$ estimators as DM-CPME and DR-CPME.

14.3.2 Model selection and tuning

Each estimator is tuned by 5-fold cross-validation procedure for OPE setting introduced in [16, Appendix B]: For the DM and DR-NN models, we vary the number of hidden units $n_h \in \{50, 100, 150, 200\}$. For CPME and DR-CPME, the regularization parameter λ is selected from the range $\{10^{-8}, \dots, 10^{-3}\}$. We repeat each experiment 30 times and report mean squared error (MSE) with 95% confidence intervals. For CPME, the kernel $k_{\mathcal{Y}}$ is linear, and the regularization parameter λ is selected via cross-validation. We use the median heuristic for the lengthscales of the kernel $k_{\mathcal{A}}$ and $k_{\mathcal{X}}$.

14.3.3 Simulated setting

We simulate the recommendation scenario of Muandet et al. [16] where users receive ordered lists of K items drawn from a catalog of M items. Each item $m \in \{1, \dots, M\}$ is represented by a feature vector $v_m \in \mathbb{R}^d$, and each user $j \in \{1, \dots, N\}$ is assigned a feature vector $x_j \in \mathbb{R}^d$, both sampled i.i.d. from $\mathcal{N}(0, I_d)$. A recommendation $a = (v_{m_1}, \dots, v_{m_K}) \in \mathbb{R}^{d \times K}$ is formed by sampling items without replacement.

The user receives a binary outcome based on whether they click on any item in the recommended list. Formally, given a recommendation a_i and a user feature vector x_j , the probability of a click is defined as

$$\theta_{ij} = \frac{1}{1 + \exp(-\bar{a}_i^\top x_j + \epsilon_{ij})},$$

where \bar{a}_i is the average of the K item vectors in the list a_i , and $\epsilon_{ij} \sim \mathcal{N}(0, 1)$ is independent noise. The binary reward is then sampled as $y_{ij} \sim \text{Bernoulli}(\theta_{ij})$.

In our experiment, a target policy $\pi(a | x)$ generates a recommendation list $a = (v_{m_1}, \dots, v_{m_K})$ by sampling K items without replacement from the M -item catalog, where sampling is governed by a multinomial distribution. For a given user j , each item's selection probability is proportional to

1679 $\exp(b_j^\top v_l)$, where b_j is the user-specific parameter vector. If we set $b_j = x_j$, the policy is optimal in
 1680 the sense that it aligns with user preferences.

1681 To construct the policies for the experiment, we first generate user features x_1, \dots, x_N . The target
 1682 policy π uses $b_j^* = p_j \odot x_j$, where $p_j \in \{0, 1\}^d$ is a binary mask with i.i.d. Bernoulli(0.5) entries,
 1683 zeroing out about half the dimensions of x_j . The logging policy π_0 is then defined by scaling:
 1684 $b_j = \alpha b_j^*$ with $\alpha \in [-1, 1]$. The parameter α controls policy similarity: $\alpha = 1$ recovers $\pi_0 = \pi$,
 1685 while $\alpha = -1$ results in maximal divergence.

1686 We generate two datasets $\mathcal{D}_{\text{init}} = \{(y_i, a_i, x_i)\}_{i=1}^n$ and $\mathcal{D}_{\text{target}} = \{(\tilde{y}_i, \tilde{a}_i, x_i)\}_{i=1}^n$, using π_0 and π
 1687 respectively, with shared user features x_i . The target outcomes \tilde{y}_i are reserved for evaluation.

1688 We evaluate performance across five setting where we vary the the values of: (i) number of observa-
 1689 tions (n), (ii) number of recommendations (K), (iii) number of users (N), (iv) dimension of context
 1690 (d), (v) policy similarity (α). Results (log scale) are shown in Figure 6.

1691 We observe:

- 1692 • All estimators generally show improved performance as the number of observations increases,
 1693 except for IPS, which exhibits a slight decline between $n = 2000$ and $n = 5000$.
- 1694 • The performance of all estimators deteriorates as either the number of recommendations (K) or the
 1695 context dimension (d) increases.
- 1696 • All estimators degrade as $\alpha \rightarrow -1$, with IPS and CPME/DR-CPME demonstrating the better
 1697 robustness.
- 1698 • CPME and DR-CPME consistently outperform the other estimators across most settings.
- 1699 • Our proposed doubly robust method, DR-CPME, offers a performance improvement over the
 1700 CPME algorithm.

1701 14.4 Computation infrastructure

1702 We ran our experiments on local CPUs of desktops and on a GPU-enabled node (in a remote server)
 1703 with the following specifications:

- 1704 • **Operating System:** Linux (kernel version 6.8.0-55-generic)
- 1705 • **GPU:** NVIDIA RTX A4500
 - 1706 – Driver Version: 560.35.05
 - 1707 – CUDA Version: 12.6
 - 1708 – Memory: 20 GB GDDR6

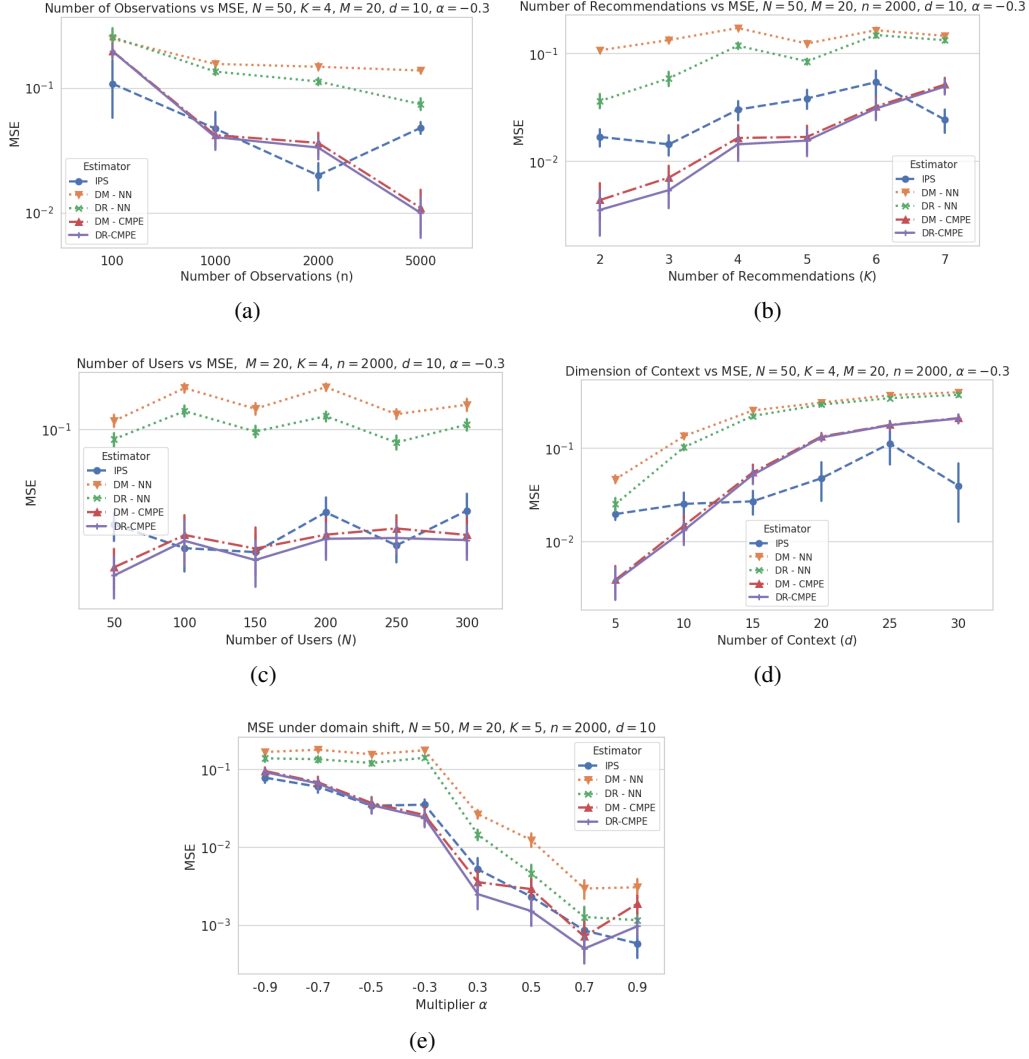


Figure 6: Mean squared error results for the off-policy evaluation experiment described in Appendix 14.3.3, reported across variations in: (a) the number of observations n , (b) the number of recommendations K , (c) the number of users N , (d) the context dimension d , and (e) the policy shift multiplier α .