DOUBLY ROBUST DEEP KERNEL TESTS FOR STRUCTURED OUTCOMES

Anonymous authors
Paper under double-blind review

ABSTRACT

We propose a test of whether two interventions induce the same outcome distribution when outcomes are structured and high-dimensional, such as images, text and graphs. We represent the interventional distributions as mean embeddings on learned deep features, such that the treatment effect is a doubly robust representation of the difference in mean embeddings. Our test, DR-KTE-Deep, optimises the learned features so as to maximise an estimated power criterion on this treatment effect representation. To motivate our learning objective, we show that the power criterion characterises the asymptotic signal-to-noise ratio of the test under fixed alternatives, and we derive a uniform finite-sample bound for its regularised estimate. We show that, with the learned kernel fixed, a held-out test is asymptotically valid under the null given suitable nuisance-estimation rates. Empirically, DR-KTE-Deep achieves substantially higher power than fixed and bandwidth-optimised kernel baselines across image, text and graph outcomes, while maintaining well-calibrated Type-I error.

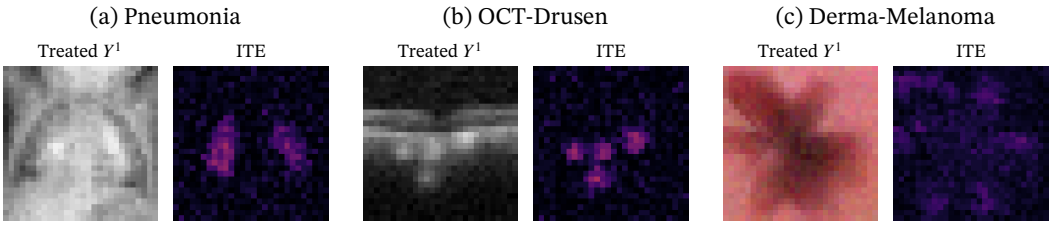

Figure 1: Semi-synthetic medical-image experiments adapted from MedMNIST (Yang et al., 2021; 2023). Each pair shows a treated potential outcome Y^1 and its ground-truth individual treatment effect (ITE).

1 INTRODUCTION

In causal problems involving structured and high-dimensional outcomes, such as images, text, and graphs, a mean outcome may be undefined or poorly aligned with the structure of the data. For example, comparing protein or text sequences requires choosing a numerical representation or a notion of similarity (Saigo et al., 2004; Lodhi et al., 2002; Reimers & Gurevych, 2019). For images, coordinate-wise averaging ignores the spatial structure of the data (LeCun et al., 1998). Distributional causal inference addresses this by defining the interventional distributions directly on the structured outcome space, without reducing outcomes to an arbitrary scalar summary or notion of averaging (Rothe, 2010; Chernozhukov et al., 2013; Kennedy et al., 2023; Muandet et al., 2021). Kernels define similarities on these spaces (Gärtner, 2003; Borgwardt et al., 2006; Shervashidze et al., 2011) and, through kernel mean embeddings, represent interventional distributions in a reproducing kernel Hilbert space (RKHS) (Smola et al., 2007).

Kernels define the geometry used to compare outcome distributions (Sriperumbudur et al., 2010; Simon-Gabriel & Schölkopf, 2018; Gretton, 2019). If this geometry is fixed a priori, the kernel may be poorly suited to detecting differences between the interventional distributions, leading to low test power (Gretton et al., 2012a; Sutherland et al., 2017). Deep kernels address this by using neural networks to learn a representation of the outcomes and compare distributions in the resulting

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

space (Wilson et al., 2016). Rather than learning this representation through a reconstruction task, the learned kernel can be optimised by directly maximising a test-power criterion (Sutherland et al., 2017).

We consider two soft interventions π and π' inducing outcome distributions P_Y^{π} and $P_Y^{\pi'}$. Determining whether an intervention changes the outcome distribution can be naturally formulated as a hypothesis testing problem, testing $H_0 : P_Y^{\pi} = P_Y^{\pi'}$ against $H_1 : P_Y^{\pi} \neq P_Y^{\pi'}$.

Treatment effects often need to be estimated from observational data, because experimentation can be costly, unethical or infeasible. In this setting, interventional distributions are not directly observed and must be recovered (Robins, 1986; Hernán & Robins, 2020; Muandet et al., 2021; Zenati et al., 2025). This can be done by modelling either treatment assignment or the conditional outcome embedding (Song et al., 2009; Grünewälder et al., 2012). However, either approach alone is sensitive to model misspecification. We therefore use a doubly robust estimator, which combines both approaches and remains consistent when either model is correctly specified (Bang & Robins, 2005; Luedtke & Chung, 2024).

In this paper, we introduce DR-KTE-Deep, which combines doubly robust interventional embeddings with deep kernel learning and testing. Our main contributions are: (i) we develop a method for learning deep kernels for tests of interventional distributions, estimating both the distributional discrepancy and the test-power criterion from observational data; (ii) we derive a uniform finite-sample bound for the estimated learning criterion, accounting for both kernel-class complexity and errors in estimating the treatment-assignment and outcome models; (iii) we establish conditional asymptotic validity of the held-out test after nuisance estimation and kernel training and selection; and (iv) we evaluate the method on image, text, and graph outcomes, showing substantial power gains over fixed and bandwidth-optimised doubly robust kernel tests, while maintaining Type-I error close to the nominal level.

The remainder of the paper is organised as follows. Section 3 develops the doubly robust test. Section 4 develops the deep-kernel learning objective. Section 5 evaluates its power and calibration across image, text, and graph outcomes. Section 6 concludes.

2 Related Work

Several works study distributional causal effects. Quantile treatment effects characterise changes across an ordered outcome distribution (Abadie et al., 2002; Chernozhukov & Hansen, 2005; Firpo, 2007), while counterfactual distribution and density methods target the full marginal law (Rothe, 2010; Chernozhukov et al., 2013; Kennedy et al., 2023). These methods are typically developed for scalar or finite-dimensional Euclidean outcomes. We instead consider structured outcomes whose interventional distributions must be compared without reducing them to low-dimensional summaries.

Kernel mean embeddings represent interventional outcome distributions in an RKHS. Muandet et al. (2021) introduce counterfactual mean embeddings, while Park et al. (2021), Näf et al. (2026), and Jain & Luedtke (2026) use kernel methods to study conditional distributional treatment effects. For marginal effects, Fawkes et al. (2024) and Martinez-Taboada et al. (2023) develop doubly robust kernel tests, and Zenati et al. (2025) extend them to general stochastic policies. Finally, Zenati et al. (2026) further consider adaptively collected data. These methods use prescribed outcome kernels, including aggregation over kernel bandwidths, without learning a deep outcome representation.

Other work adapts the kernel or outcome representation to improve test power. Gretton et al. (2012b) optimise the power of linear-time MMD tests over kernel mixtures and analyse convergence of the selection criterion. Sutherland et al. (2017) optimise kernel parameters for two-sample tests, while Liu et al. (2020) learn characteristic deep kernels and prove uniform convergence of their learning criterion. Other approaches combine multiple kernels through aggregation or fusion (Schrab et al., 2023; Biggs et al., 2023). In causal settings, Zenati & Gretton (2026) learn informative witness locations for distributional treatment effects, while Wibisono & Wang (2026) and Durand et al. (2025) learn outcome features or projections to detect causal effects. Our method instead learns the outcome kernel for an omnibus test of marginal interventional distributions. We estimate the learning criterion from observational data using causal nuisance models and analyse how errors in these models affect the criterion.

3 DOUBLY ROBUST KERNEL TEST

Problem Setting. We are given an observational dataset $\mathcal{D}$ of N independent and identically distributed (i.i.d.) samples $O_1, \dots, O_N \sim_{\text{i.i.d}} \mathbb{P}_0$, where $O_i = (X_i, A_i, Y_i) \in \mathcal{X} \times \mathcal{A} \times \mathcal{Y}$. $\mathcal{X}$, $\mathcal{A}$, and $\mathcal{Y}$ are standard Borel spaces. Write $\mathbb{P}_X$ for the X-marginal of $\mathbb{P}_0$, and $\mathbb{P}_{A,X}$ for its treatment-covariate marginal. Let ρ be a σ-finite reference measure on $\mathcal{A}$. The logging policy assigns $A \mid X = x$ with density $\pi_0(\cdot \mid x)$ relative to ρ. We consider a unit-level potential-outcome setting with no interference between units (Rubin, 2005; Imbens & Rubin, 2015). We compare two soft interventions π and π' with conditional densities $\pi(\cdot \mid x)$ and $\pi'(\cdot \mid x)$ respectively.

Identification. In our causal setting, we do not have direct access to samples from the interventional outcome distributions. The outcome distributions therefore need to be identified from observational data. We use a single jointly measurable response mechanism to define both fixed-treatment and soft-intervention outcomes. We make the following assumptions, standard in the causal literature.

Assumption 3.1 (Identification)**.** *There exist a standard Borel space $\mathcal{U}$, a $\mathcal{U}$-valued random variable U, and a jointly measurable map $F : \mathcal{A} \times \mathcal{X} \times \mathcal{U} \to \mathcal{Y}$ such that:*

1. *Consistency and structural response: $Y = F(A, X, U)$ almost surely; for every $a \in \mathcal{A}$, define $Y^a := F(a, X, U)$.*
2. *Conditional exogeneity: $U \perp\!\!\!\perp A \mid X$.*
3. *Positivity on intervention support: $\pi_0(a \mid x) > 0$ whenever $\pi(a \mid x) + \pi'(a \mid x) > 0$, up to a $\mathbb{P}_X \otimes \rho$-null set.*

For $p \in \{\pi, \pi'\}$, draw $A^p \mid X$ with density $p(\cdot \mid X)$ independently of U given X, and define $Y^p := F(A^p, X, U)$ and $P_Y^p := \mathcal{L}(Y^p)$. We test $H_0 : P_Y^\pi = P_Y^{\pi'}$ against $H_1 : P_Y^\pi \neq P_Y^{\pi'}$.

Choose a regular conditional law $P_{U|X}$ of U given X. Define the conditional outcome distribution, common to both interventions, by

$$Q(B \mid a, x) := \int_{\mathcal{U}} \mathbf{1}\{F(a, x, u) \in B\} \, P_{U|X}(\mathrm{d}u \mid x),$$

for any measurable $B \subseteq \mathcal{Y}$. Under Assumption 3.1, Q agrees with the observational conditional law $\mathbb{P}_0(Y \in \cdot \mid A = a, X = x)$ for $\mathbb{P}_{A,X}$-almost every (a, x). For each fixed treatment $a \in \mathcal{A}$, $Q(\cdot \mid a, x)$ is the conditional distribution of Y^a given X, for $\mathbb{P}_X$-almost every x. The soft-intervention law is given by the g-formula (Robins, 1986; Hernán & Robins, 2020):

$$P_Y^p(B) = \int_{\mathcal{X}} \int_{\mathcal{A}} Q(B \mid a, x) p(a \mid x) \rho(\mathrm{d}a) \mathbb{P}_X(\mathrm{d}x).$$

Positivity ensures that the target treatment-covariate measure is absolutely continuous with respect to $\mathbb{P}_{A,X}$. Therefore, the g-formula does not depend on the chosen version of the observational conditional law. This identifies the soft-intervention law without requiring identification at every fixed treatment value. We prove these claims in Proposition B.1.

Kernel Formulation. To compare distributions over structured outcomes, we represent them as elements of an RKHS. Let $\Theta \subseteq \mathbb{R}^{d_\Theta}$ be a non-empty parameter set. For each $\theta \in \Theta$, let $k_\theta : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ be a positive-semidefinite kernel on $\mathcal{Y}$ parameterised by θ with separable RKHS $\mathcal{H}_\theta$ and strongly measurable canonical feature map $\varphi_\theta(y) := k_\theta(y, \cdot)$.

We define the interventional mean embedding (Muandet et al., 2021; Zenati et al., 2025), embedding contrast, and squared maximum mean discrepancy (MMD) (Gretton et al., 2012a) respectively by

$$\chi_\theta(p) := \int_{\mathcal{Y}} \varphi_\theta(y) P_Y^p(\mathrm{d}y), \quad \Delta_\theta(\pi, \pi') := \chi_\theta(\pi) - \chi_\theta(\pi'), \quad D_\theta(\pi, \pi') := \|\Delta_\theta(\pi, \pi')\|_{\mathcal{H}_\theta}^2.$$

Assumption 3.2 (Embedding integrability)**.** *For every $\theta \in \Theta$ and $p \in \{\pi, \pi'\}$, assume $\mathbb{E}_{\mathbb{P}_0} \sqrt{k_\theta(Y, Y)} < \infty$ and $\int_{\mathcal{Y}} \sqrt{k_\theta(y, y)} P_Y^p(\mathrm{d}y) < \infty$. These first moments ensure Bochner integrability under the observed and interventional laws.*

If k_θ is characteristic, testing H_0 is equivalent to testing $D_\theta(\pi, \pi') = 0$ (Gretton et al., 2012a, Section 2.2).

Doubly Robust Representation. To estimate $D_\theta(\pi, \pi')$ from observational data, we construct a doubly robust representation of the embedding contrast $\Delta_\theta(\pi, \pi')$, following Zenati et al. (2025). We further compare the choice of estimator in Section D.4.

For a fixed $\theta \in \Theta$, define the conditional outcome mean embedding (Grünewälder et al., 2012) by

$$\mu_\theta(a, x) := \int_{\mathcal{Y}} \varphi_\theta(y) Q(\,\mathrm{d}y \mid a, x).$$

We use the null-set convention described in Section B.1. This embedding satisfies $\mu_\theta(A, X) = \mathbb{E}_{\mathbb{P}_0}[\varphi_\theta(Y) \mid A, X]$ almost surely. For ease of notation, write $r(a, x) := \pi(a \mid x) - \pi'(a \mid x)$. Let e be a measurable candidate logging-policy density that is positive wherever $|r| > 0$. Let m_θ be a measurable candidate conditional outcome embedding. We set ratios and integrands weighted by r or $|r|$ to zero where $r = 0$.

Define the Hilbert-valued pseudo-feature map

$$g_\theta^{\pi,\pi'}(O; e, m_\theta) := \frac{r(A, X)}{e(A \mid X)} \{\varphi_\theta(Y) - m_\theta(A, X)\} + \int_{\mathcal{A}} r(a, X) m_\theta(a, X) \rho(\,\mathrm{d}a). \tag{1}$$

The second term estimates the embedding contrast through the conditional outcome model, while the first corrects it using propensity-weighted outcome residuals.

We suppress the intervention superscript when no ambiguity arises, $g_\theta(O; e, m_\theta) := g_\theta^{\pi,\pi'}(O; e, m_\theta)$. The oracle pseudo-feature is then $g_\theta(O) := g_\theta(O; \pi_0, \mu_\theta)$. Write $\mathbb{E}_0$ for expectation under $\mathbb{P}_0$.

Proposition 3.3 (Observed-data representation and double robustness)**.** *Suppose Assumptions 3.1 and 3.2 hold and the candidate nuisances satisfy the absolute-integrability conditions of Section B.2. Write $\Delta_\theta := \Delta_\theta(\pi, \pi')$. For candidate nuisance functions e and m_θ,*

$$\mathbb{E}_0\left[g_\theta(O; e, m_\theta)\right] - \Delta_\theta = \mathbb{E}_0\left[\int_{\mathcal{A}} r(a, X) \left\{\frac{\pi_0(a \mid X)}{e(a \mid X)} - 1\right\} \{\mu_\theta(a, X) - m_\theta(a, X)\} \rho(\,\mathrm{d}a)\right].$$

Thus, $\mathbb{E}_0\left[g_\theta(O; e, m_\theta)\right] = \Delta_\theta$ if either $e = \pi_0$ or $m_\theta = \mu_\theta$ on the support of $|r|$.

A proof is given in Section B.2. We examine the effect of outcome-model misspecification empirically in Section D.5, using a misspecified m_θ but oracle propensity $e \equiv \pi_0$.

Test Statistic. Since $D_\theta(\pi, \pi') = \|\Delta_\theta(\pi, \pi')\|_{\mathcal{H}_\theta}^2$, the embedding contrast vanishes under H_0 and the induced cross-U kernel over $\mathcal{H}_\theta$ is one-degenerate. We therefore estimate $D_\theta(\pi, \pi')$ by taking the inner product of two independent estimates of $\Delta_\theta(\pi, \pi')$, following the cross-U and cross-MMD constructions of Kim & Ramdas (2024) and Shekhar et al. (2022).

Let $\hat{\pi}_0$ and $\hat{\mu}_\theta$ be nuisance estimators fitted on a sample independent of the evaluation sample, and set $\hat{g}_{\theta,i} := g_\theta(O_i; \hat{\pi}_0, \hat{\mu}_\theta)$. We split the evaluation sample into two disjoint halves $\mathcal{I}_1$ and $\mathcal{I}_2$ of size m. Define $\bar{g}_{\theta,\ell} := \frac{1}{m} \sum_{i \in \mathcal{I}_\ell} \hat{g}_{\theta,i}$ for $\ell \in \{1, 2\}$. The cross-MMD and variance estimators are then, respectively,

$$\widehat{D}_\theta^{\mathrm{cross}} := \langle \bar{g}_{\theta,1}, \bar{g}_{\theta,2} \rangle_{\mathcal{H}_\theta} = \frac{1}{m^2} \sum_{i \in \mathcal{I}_1} \sum_{j \in \mathcal{I}_2} \langle \hat{g}_{\theta,i}, \hat{g}_{\theta,j} \rangle_{\mathcal{H}_\theta}, \quad \hat{s}_\theta^2 := \frac{1}{m} \sum_{j \in \mathcal{I}_2} \Big(\langle \bar{g}_{\theta,1}, \hat{g}_{\theta,j} \rangle_{\mathcal{H}_\theta} - \widehat{D}_\theta^{\mathrm{cross}} \Big)^2.$$

Our test statistic is the studentised cross-U statistic

$$T_\theta := \sqrt{m} \widehat{D}_\theta^{\mathrm{cross}} / \hat{s}_\theta. \tag{2}$$

We set $T_\theta = 0$ whenever $\hat{s}_\theta = 0$.

Let Φ denote the standard normal distribution function. For a fixed kernel, Theorem 4.3 gives, under the corresponding moment, projected non-degeneracy and nuisance-bias conditions of Assumption B.3, $T_\theta \to^d \mathcal{N}(0, 1)$ under H_0. Since $D_\theta(\pi, \pi') > 0$ for a characteristic kernel under H_1, large positive values of T_θ therefore provide evidence against H_0. We reject at level α when $\hat{p} = 1 - \Phi(T_\theta) < \alpha$.

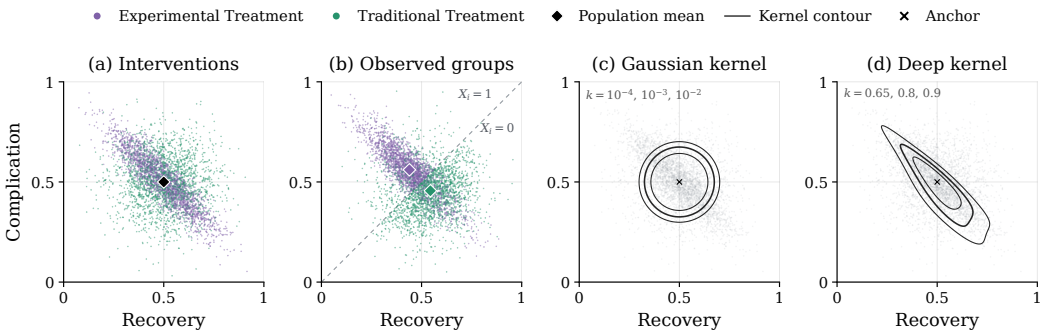

Figure 2: Synthetic example where treatment assignment is driven by an indicator of disease severity X. The potential-outcome distributions are Gaussians with the same mean, but the traditional-treatment distribution is isotropic while the experimental-treatment distribution is anisotropic. The deep kernel adapts to the geometry of the distributions, while the Gaussian kernel treats points uniformly throughout the outcome space.

4 Learning Deep Kernels

Learning Objective. The choice of outcome kernel affects both the population discrepancy D_θ and the sampling variability of its estimator. We therefore select θ using the signal-to-noise ratio (Gretton et al., 2012b; Sutherland et al., 2017; Liu et al., 2020)

$$J_\theta = D_\theta / s_\theta, \qquad \text{with } D_\theta := D_\theta(\pi, \pi'), \quad s_\theta^2 := \mathrm{Var}_{O\sim\mathbb{P}_0}\left[\langle \Delta_\theta, g_\theta(O)\rangle_{\mathcal{H}_\theta}\right]$$

whenever $s_\theta > 0$. J_θ is unchanged by a global rescaling of the kernel and governs the growth of the studentised test statistic under fixed alternatives.

Proposition 4.1 (Fixed-alternative growth rate)**.** *Suppose Assumptions 3.1 and 3.2 hold. For a fixed θ, suppose $D_\theta > 0$, $s_\theta > 0$, and $\mathbb{E}_{O\sim\mathbb{P}_0}[\|g_\theta^{\pi,\pi'}(O)\|_{\mathcal{H}_\theta}^4] < \infty$. Suppose the statistics use the true nuisance functions. Then the cross-MMD and variance estimators satisfy*

$$\widehat{D}_\theta^{\mathrm{cross}} \xrightarrow{\mathbb{P}} D_\theta, \qquad \hat{s}_\theta^2 \xrightarrow{\mathbb{P}} s_\theta^2, \qquad T_\theta/\sqrt{m} \xrightarrow{\mathbb{P}} J_\theta.$$

A proof is given in Section B.3. Section D.6 further studies the choice of learning criterion. This result motivates maximising J_θ; kernels with larger J_θ yield a test statistic that is asymptotically larger under fixed alternatives. Since D_θ and s_θ are unknown, we use an empirical criterion, along with regularisation (Jitkrittum et al., 2016; 2017).

$$\widehat{J}_{\lambda,\theta}^{\mathrm{DR}} = \widehat{D}_\theta^{\mathrm{cross}} / \hat{s}_{\theta,\lambda}, \tag{3}$$

where $\hat{s}_{\theta,\lambda}^2 := \hat{s}_\theta^2 + \lambda$ and $\lambda > 0$. The regularisation prevents the learning objective from becoming artificially large when the estimated variance $\hat{s}_\theta^2$ is close to zero. It is used only to stabilise kernel selection; the final test statistic remains unregularised.

While the learning criterion can be used to optimise the parameters of a simple kernel family, like the lengthscale of a Gaussian kernel, the kernel still compares outcomes isotropically throughout the outcome space, and may have low power against localised or anisotropic perturbations at finite sample size (Liu et al., 2020, Section 3) (Figure 2).

Deep Kernels. We instead learn a deep kernel combining a neural component with a base characteristic kernel. Since deep kernel optimisation is sensitive to random initialisation, we train and select an ensemble of M kernels, and average them. Let $\Xi \subseteq \mathbb{R}^{d_0}$ be the parameter space of a kernel member and let $\theta = (\vartheta_1, \ldots, \vartheta_M) \in \Theta$ collect each $\vartheta_b \in \Xi$ for $b \in [1, M]$. Define

$$k_{\vartheta_b}(y, y') = (1 - w_{\vartheta_b})\kappa_{\vartheta_b}\big(\zeta_{\vartheta_b}(y), \zeta_{\vartheta_b}(y')\big) + w_{\vartheta_b} k_{\vartheta_b}^{\mathrm{char}}(y, y'), \quad k_\theta(y, y') = \frac{1}{M}\sum_{b=1}^{M} k_{\vartheta_b}(y, y'). \tag{4}$$

Here $\zeta_\vartheta : \mathcal{Y} \to \mathbb{R}^d$ is a neural feature map, κ_ϑ is a Gaussian kernel on the learned feature space, $k_\vartheta^{\mathrm{char}}$ is a characteristic kernel on $\mathcal{Y}$, and $w_\vartheta \in [w_{\min}, 1]$ is a learned mixing parameter where $w_{\min} > 0$.

The learned component $\kappa_{\vartheta_b}(\zeta_{\vartheta_b}(y), \zeta_{\vartheta_b}(y'))$ adapts the geometry used to distinguish the interventional distributions, while the characteristic component preserves sensitivity to arbitrary distributional differences.

Lemma 4.2 (Characteristic ensemble safeguard)**.** *For every pair of probability laws P, Q for which the component mean embeddings exist,*

$$\mathrm{MMD}^2_{k_\theta}(P, Q) \geq \frac{w_{\min}}{M} \sum_{b=1}^{M} \mathrm{MMD}^2_{k^{\mathrm{char}}_{\vartheta_b}}(P, Q). \tag{5}$$

Consequently k_θ is characteristic for every $\theta \in \Theta$.

A proof is given in Section B.4. For every $\theta \in \Theta$, $D_\theta(\pi, \pi') = 0$ if and only if the two interventional outcome distributions are equal.

Training and Testing Procedure. We detail the training and testing procedure in Algorithm 1 and Section C.1. $\mathcal{D}$ is partitioned into independent learning, validation and testing samples (Gretton et al., 2012b; Liu et al., 2020). We optimise the kernel parameters on the learning sample, select a kernel on the validation sample, and compute T_θ on the testing sample. The learning sample is split into three disjoint sets $\mathcal{D}_{\mathrm{fit}}$, $\mathcal{D}_{\mathrm{sel},1}$, and $\mathcal{D}_{\mathrm{sel},2}$. We fit the nuisance estimators on $\mathcal{D}_{\mathrm{fit}}$ and compute $\widehat{J}^{\mathrm{DR}}_{\lambda,\vartheta_b}$ on $\mathcal{D}_{\mathrm{sel},1}$ and $\mathcal{D}_{\mathrm{sel},2}$. At each training step, we rotate the roles of the three folds and average the resulting criterion, then optimise each ϑ_b by gradient ascent. During training, we maintain an exponential moving average (EMA) of each member's kernel parameters (Tarvainen & Valpola, 2017) and select the EMA checkpoint with the largest unregularised validation statistic.

Let $\mathcal{F}_{\mathrm{learn}}$ denote the data and randomness used for nuisance estimation, kernel training and selection. The fitted nuisances and selected kernel are therefore fixed when conditioned on $\mathcal{F}_{\mathrm{learn}}$ and the test observations remain independent. We further impose the conditions of Assumption B.3, including conditional fourth-moment control, non-degenerate projected variance, and a conditional pseudo-feature bias $o_{\mathbb{P}}(m^{-1/2})$. The latter follows from a corresponding product-rate condition on the nuisance errors (Proposition 3.3).

Theorem 4.3 (Held-out null validity)**.** *Suppose Assumption B.3 holds, and set $T_{\hat{\theta}} = 0$ when $\hat{s}_{\hat{\theta}} = 0$. Under H_0, as the size m of each test half tends to infinity,*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(T_{\hat{\theta}} \leq t \mid \mathcal{F}_{\mathrm{learn}}) - \Phi(t) \right| \xrightarrow{\mathbb{P}} 0. \tag{6}$$

Consequently, $T_{\hat{\theta}} \xrightarrow{d} \mathcal{N}(0, 1)$ and the test has asymptotic level α for every fixed $\alpha \in (0, 1)$.

A proof is given in Section B.6. The result is conditional on the full learning procedure and does not require the optimised parameters to converge to a deterministic kernel.

Uniform Control of the Learning Criterion. We provide a uniform analysis of the learning criterion (see also Gretton et al. (2012b, Theorem 1), Liu et al. (2020, Theorem 11)) with the fitted nuisances $\hat{\pi}_0$ and $\hat{\mu}_\theta$, which will introduce a nuisance estimation error. The bound does not guarantee that training attains a global maximiser. Define

$$J_{\lambda,\theta} := \frac{D_\theta}{\sqrt{s^2_\theta + \lambda}}, \qquad \lambda > 0. \tag{7}$$

Let the nuisance-fitting sample have size n and each selection half have size m. Write $\mathcal{F}_{\mathrm{fit}}$ for the data and randomness used to fit the nuisance functions. The nuisance functions are therefore fixed when conditioned on $\mathcal{F}_{\mathrm{fit}}$, but retain their dependence on θ.

On a compact finite-dimensional parameter class, we impose the conditions of Assumption B.4. These require the fitted kernels, constructed from k_θ and the nuisance estimates, to be bounded and Lipschitz in θ, together with overlap and measurability conditions.

Theorem 4.4 (Uniform feasible learning criterion)**.** *Let Assumptions B.4 and 3.1 hold and fix $\lambda_0 > 0$. Let $\epsilon^2_{\pi,n}$ and $\epsilon^2_{\mu,n}$ be the nuisance errors conditional on $\mathcal{F}_{\mathrm{fit}}$ defined in Section B. For every $\lambda \geq \lambda_0$ and $\eta \in (0,1)$, there is a finite constant C, depending only on the constants in Assumption B.4 and λ_0, such that, with conditional probability at least $1-\eta$ over the two selection halves,*

$$\sup_{\theta\in\Theta} |\widehat{J}^{\mathrm{DR}}_{\lambda,\theta} - J_{\lambda,\theta}| \leq C \left\{ \frac{\sqrt{d_\Theta \log(1 + 2R_\Theta\sqrt{m}) + \log(4/\eta)} + 1}{\sqrt{m}} + \epsilon_{\pi,n}\epsilon_{\mu,n} + \epsilon_{\pi,n} + \epsilon^2_{\mu,n} \right\}. \tag{8}$$

For three cyclic rotations of the learning folds satisfying the same regularity constants, the averaged criterion obeys the analogous bound with unconditional probability at least $1-3\eta$, with the nuisance remainder averaged over rotations.

The uniform bound separates sampling error from nuisance-estimation error. The MMD numerator retains the doubly robust product remainder $\epsilon_{\pi,n}\epsilon_{\mu,n}$. Estimating s^2_θ introduces additional terms $\epsilon_{\pi,n}$ and $\epsilon^2_{\mu,n}$. For fixed class constants and confidence level, note that the sampling term is $O(\sqrt{\log m/m})$, which is asymptotically faster than the $O(m^{-1/3})$ of Gretton et al. (2012b, Theorem 1), Liu et al. (2020, Theorem 11).

Theorem 4.4 controls $J_{\lambda,\theta}$ uniformly over the full parameter class Θ for λ bounded away from zero. On subclasses where the variance is bounded below,

$$\Theta_s := \{\theta \in \Theta : s^2_\theta \geq s^2\},$$

the results in Section B allow $\lambda_m \to 0$ and yield uniform approximation to the unregularised criterion J_θ. The comparison with earlier rates follows the proof of Theorem B.12.

Since each kernel member is trained separately, we apply these results setting $M = 1$, $\Theta = \Xi$ and $d_\Theta = d_0$, under the corresponding assumptions. With $M > 1$, Proposition B.2 gives $J_{\lambda,\theta} \geq \min_b J_{\lambda,\vartheta_b}$ for $\lambda \geq 0$ with positive denominators. Conditional on the common learning and validation data and shared training randomness, independent and identically distributed member training reduces the training-randomness variance of the population discrepancy by a factor of $1/M$. Proposition B.2 establishes these population guarantees. Section D.8 studies the effect of ensemble size on test power. Section D.12 analyses the effect of jointly maximising the averaged criterion.

5 Experimental Evaluation

We evaluate our doubly robust deep kernel test, DR-KTE-Deep, on image, text and graph outcomes. Evaluations on low-dimensional outcomes are included in Section D. We compare our method with DR-KTE (Martinez-Taboada et al., 2023), which uses a fixed Gaussian kernel, and DR-KTE-Opt, which optimises the lengthscale of a Gaussian kernel. Experimental details are given in Section C. Section D.10 additionally compares our method to linear kernels. All methods are fitted and evaluated using the same total sample size.

We learn the deep kernel and nuisance parameters on $\mathcal{D}_{\text{learn}}$, sampled from the dataset's training pool, and select the kernel on $\mathcal{D}_{\text{val}}$, sampled from the validation pool. We then refit the nuisances on $\mathcal{D}_{\text{learn}} \cup \mathcal{D}_{\text{val}}$ and estimate the rejection rate at $\alpha = 0.05$ over 200 test datasets $\mathcal{D}_{\text{test}}$, sampled from the testing pool. We repeat this process over 20 independent outer repetitions and report the mean rejection rate $\pm$ one standard error across the outer repetitions.

Image Outcomes. We evaluate our method on mean-preserving distributional effects on medical-image outcomes. We extend the OCTMNIST construction of Zenati & Gretton (2026). Potential outcomes are constructed as $Y^a_i = B_i + R_i(a)$ where B_i is sampled from normal or healthy MedMNIST images (Yang et al., 2023) and $R_i(a)$ is a medically motivated perturbation representing a localised lesion, scattered deposits, or morphological changes. We scale the perturbation by an effect strength δ. The residual construction is mean-matched, $\mathbb{E}[Y^1_i - Y^0_i \mid X_i] = 0$, such that the effect is mainly distributional under the alternative. We consider ten constructions, three of which are illustrated in Figure 1. Section C.2 details the full generative process.

DR-KTE-Deep achieves substantially higher estimated power than the baselines on all constructions (Table 1), while DR-KTE-Opt only slightly improves power over the fixed kernel. Under H_0, rejec-

tion rates across the three methods remain close to the nominal level of $\alpha = 0.05$, with only a slight inflation for OCT-DME. In Figure 3, the power of the deep kernel grows with sample size and effect strength, while Gaussian baselines remain near the nominal rejection rate across the sample-size sweep.

Scenario	DR-KTE	DR-KTE-Opt	DR-KTE-Deep
Breast-Tumor	$0.749_{\pm 0.018}/0.053_{\pm 0.003}$	$0.761_{\pm 0.017}/0.051_{\pm 0.003}$	$\mathbf{0.989}_{\pm \mathbf{0.002}}/0.054_{\pm 0.003}$
Chest-Cardiomegaly	$0.156_{\pm 0.007}/0.050_{\pm 0.003}$	$0.158_{\pm 0.008}/0.050_{\pm 0.003}$	$\mathbf{0.928}_{\pm \mathbf{0.046}}/0.051_{\pm 0.004}$
Chest-Effusion	$0.231_{\pm 0.009}/0.053_{\pm 0.004}$	$0.245_{\pm 0.011}/0.053_{\pm 0.004}$	$\mathbf{0.999}_{\pm \mathbf{0.000}}/0.048_{\pm 0.004}$
Chest-Mass	$0.436_{\pm 0.018}/0.050_{\pm 0.003}$	$0.461_{\pm 0.019}/0.050_{\pm 0.003}$	$\mathbf{1.000}_{\pm \mathbf{0.000}}/0.056_{\pm 0.004}$
Derma-Melanoma	$0.171_{\pm 0.009}/0.051_{\pm 0.002}$	$0.172_{\pm 0.009}/0.050_{\pm 0.002}$	$\mathbf{0.898}_{\pm \mathbf{0.040}}/0.053_{\pm 0.004}$
OCT-CNV	$0.907_{\pm 0.014}/0.052_{\pm 0.003}$	$0.917_{\pm 0.014}/0.051_{\pm 0.003}$	$\mathbf{1.000}_{\pm \mathbf{0.000}}/0.053_{\pm 0.003}$
OCT-DME	$0.139_{\pm 0.008}/0.055_{\pm 0.004}$	$0.144_{\pm 0.008}/0.055_{\pm 0.004}$	$\mathbf{1.000}_{\pm \mathbf{0.000}}/0.063_{\pm 0.005}$
OCT-Drusen	$0.051_{\pm 0.003}/0.058_{\pm 0.004}$	$0.052_{\pm 0.003}/0.058_{\pm 0.004}$	$\mathbf{0.371}_{\pm \mathbf{0.080}}/0.058_{\pm 0.005}$
Pneumonia	$0.857_{\pm 0.021}/0.055_{\pm 0.003}$	$0.871_{\pm 0.020}/0.056_{\pm 0.003}$	$\mathbf{0.998}_{\pm \mathbf{0.001}}/0.053_{\pm 0.004}$
Retina-Severe	$0.285_{\pm 0.015}/0.050_{\pm 0.003}$	$0.288_{\pm 0.015}/0.049_{\pm 0.003}$	$\mathbf{0.959}_{\pm \mathbf{0.008}}/0.052_{\pm 0.003}$

Table 1: Power (H_1) and Type-I error (H_0) for medical-image outcomes at $N = 1500$ and $\alpha = 0.05$. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 independent repetitions.

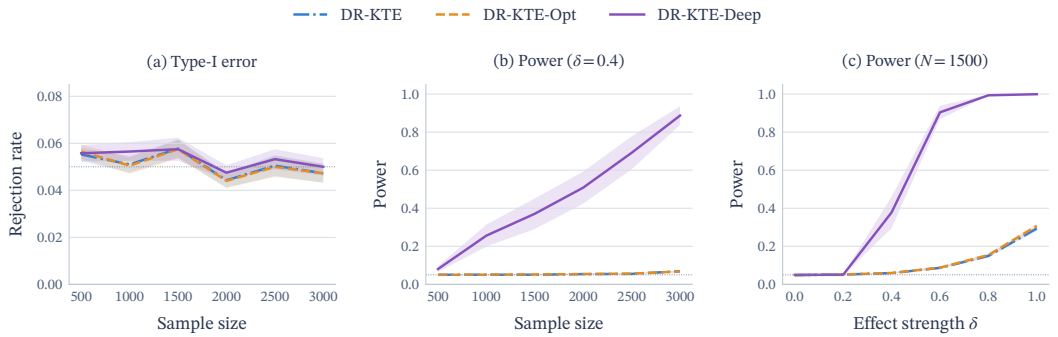

Figure 3: Results for OCT-Drusen. (a) Type-I error under H_0, (b) average power when varying N while keeping $\delta = 0.4$ fixed, (c) average power when varying δ while keeping $N = 1500$ fixed. Shaded bands show $\pm$ one standard error across 20 independent repetitions. The dotted grey line marks $\alpha = 0.05$.

Text Outcomes. We now evaluate our method on text outcomes, using a language model to generate synthetic data under the assumed causal structure (Bynum & Cho, 2025). Covariates and treatment generate a scalar pseudo-outcome, discretised into one of three tones used to condition the language model. We consider three scenarios: Oncology Notes, Movie Reviews and Clinical Notes. The generated text is represented as a sequence of tokens and used as input to each kernel method. Construction details are given in Section C.3. Section D.7 additionally compares each method through an alternative text embedding outcome representation.

Scenario	DR-KTE	DR-KTE-Opt	DR-KTE-Deep
Oncology Notes	$0.371_{\pm 0.015}/0.048_{\pm 0.003}$	$0.371_{\pm 0.015}/0.048_{\pm 0.003}$	$\mathbf{0.809}_{\pm \mathbf{0.023}}/0.055_{\pm 0.003}$
Movie Reviews	$0.255_{\pm 0.012}/0.050_{\pm 0.003}$	$0.254_{\pm 0.012}/0.050_{\pm 0.003}$	$\mathbf{0.449}_{\pm \mathbf{0.026}}/0.053_{\pm 0.004}$
Clinical Notes	$0.058_{\pm 0.004}/0.052_{\pm 0.003}$	$0.058_{\pm 0.004}/0.051_{\pm 0.003}$	$\mathbf{0.939}_{\pm \mathbf{0.013}}/0.058_{\pm 0.004}$

Table 2: Power (H_1) and Type-I error (H_0) for text outcomes, represented as a sequence of tokens, at $N = 1500$ and $\alpha = 0.05$. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 independent repetitions.

DR-KTE-Deep achieves the highest estimated power (Table 2), with substantial gains on Oncology Notes and Clinical Notes and smaller improvements on Movie Reviews. Under H_0, average rejection rates are close to the nominal level of $\alpha = 0.05$. In Figure 4, we vary the sample size N and a treatment-effect strength δ. DR-KTE-Deep maintains higher power across all sample sizes, with larger gains as δ increases.

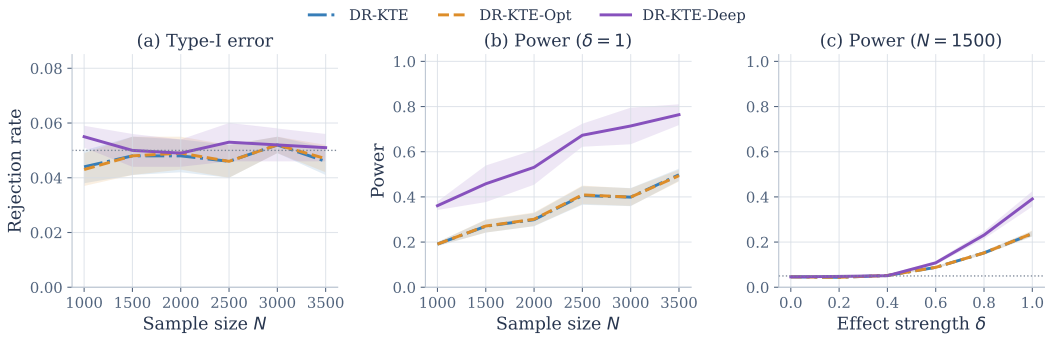

Figure 4: Results for Movie Reviews. (a) Type-I error under H_0, (b) average test power when varying N while keeping $\delta = 1.0$ fixed, (c) average test power when varying δ while keeping $N = 1500$ fixed. Shaded bands show $\pm$ one standard error across 20 independent repetitions. The dotted grey line marks $\alpha = 0.05$.

Graph Outcomes. We consider two experiments in which the outcome is a labelled graph. For LLM Code Repairs, we simulate the effect of using a large language model (LLM) to fix a buggy Python function represented as its abstract syntax tree (AST). For Contact Networks, we simulate the effect of social distancing on the spread of an infection using contact networks sampled from the Infectious SocioPatterns dataset (Isella et al., 2011). The graph constructions and encodings are detailed in Section C.4. In addition to Gaussian-based kernels, we evaluate Weisfeiler-Lehman (WL) subtree kernels (Shervashidze et al., 2011) in Section D.9.

DR-KTE-Deep achieves higher estimated average power on both LLM Code Repairs and Contact Networks, with smaller improvements on the latter. The Gaussian baselines already have high power on both constructions. Under H_0, rejection rates range from 0.048 to 0.052, close to the nominal level (Table 3).

Scenario	DR-KTE	DR-KTE-Opt	DR-KTE-Deep
LLM Code Repairs	$0.968_{\pm 0.008}/0.048_{\pm 0.004}$	$0.963_{\pm 0.009}/0.048_{\pm 0.004}$	$\mathbf{1.000}_{\pm \mathbf{0.000}}/0.051_{\pm 0.003}$
Contact Networks	$0.959_{\pm 0.006}/0.052_{\pm 0.003}$	$0.958_{\pm 0.006}/0.052_{\pm 0.003}$	$\mathbf{0.962}_{\pm \mathbf{0.006}}/0.051_{\pm 0.002}$

Table 3: Power (H_1) and Type-I error (H_0) for graph outcomes at $N = 1500$. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 outer repetitions.

6 Discussion

In this paper, we introduced DR-KTE-Deep, a doubly robust test of interventional distributions. We learned a deep outcome kernel by maximising an estimated power criterion, allowing the kernel to adapt to the geometry of the outcome space. We established conditional asymptotic null validity after kernel learning and selection, and a uniform finite-sample bound for the regularised learning criterion. DR-KTE-Deep achieved substantially higher empirical power than fixed and bandwidth-optimised baselines, while maintaining Type-I error close to the nominal level. Section D further motivates our choice of estimator and learning objective, and examines how outcome-model misspecification, outcome representation, ensemble size and outcome dimension affect the test. We use separate samples to learn the kernel and compute the test statistic. Future work could investigate whether reusing observations can further improve power while maintaining Type-I error control. We also restrict the test to two interventions, fixed in advance. Future work could extend our method to multiple candidate interventions.

AI USE STATEMENT

Language models were used to generate synthetic text outcomes for our experiments, as described in Section C.3. We also used generative AI for coding assistance, primarily for plotting and running scripts. We also used generative AI for proof strategies and checking theoretical proofs. Finally, we used generative AI to proofread and improve writing. The synthetic text datasets were manually checked by the authors for quality. LLM-generated code was verified and tested for correctness by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Alberto Abadie, Joshua Angrist, and Guido Imbens. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117, 2002. doi: 10.1111/1468-0262.00270.

Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in Brief*, 28:104863, 2020.

H. Bang and J. M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

Felix Biggs, Antonin Schrab, and Arthur Gretton. MMD-Fuse: Learning and combining kernels for two-sample testing without data splitting. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/edd00cead3425393baf13004de993017-Abstract.html.

Karsten M Borgwardt, Arthur Gretton, Malte J Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex J Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):e49–e57, 2006.

Lucius E.J. Bynum and Kyunghyun Cho. Language models as causal effect generators. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 2096–2115, Suzhou, China, 2025. Association for Computational Linguistics. doi: 10.18653/v1/2025.emnlp-main.107. URL https://aclanthology.org/2025.emnlp-main.107/.

Victor Chernozhukov and Christian Hansen. An IV model of quantile treatment effects. *Econometrica*, 73(1):245–261, 2005. doi: 10.1111/j.1468-0262.2005.00570.x.

Victor Chernozhukov, Iván Fernández-Val, and Blaise Melly. Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268, 2013.

Homer Durand, Gherardo Varando, and Gustau Camps-Valls. Learning causal response representations through direct effect analysis. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pp. 1135–1166, 2025. URL https://proceedings.mlr.press/v286/durand25a.html.

Ramtin Ehsani, Esteban Parra, Sonia Haiduc, and Preetha Chatterjee. Hierarchical knowledge injection for improving LLM-based program repair. In *Proceedings of the 40th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pp. 1440–1452. IEEE, 2025. doi: 10.1109/ASE63991.2025.00122.

Jake Fawkes, Robert Hu, Robin J. Evans, and Dino Sejdinovic. Doubly robust kernel statistics for testing distributional treatment effects. *Transactions on Machine Learning Research*, 2024.

Sergio Firpo. Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75(1):259–276, 2007. doi: 10.1111/j.1468-0262.2007.00738.x.

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

Arthur Gretton. Introduction to RKHS, and some simple kernel algorithms. Lecture notes, University College London, 2019. URL https://www.gatsby.ucl.ac.uk/~gretton/coursefiles/lecture4_introToRKHS.pdf.

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

Arthur Gretton, Dino Sejdinovic, Heiko Strathmann, Sivaraman Balakrishnan, Massimiliano Pontil, Kenji Fukumizu, and Bharath K. Sriperumbudur. Optimal kernel choice for large-scale two-sample tests. In *Advances in Neural Information Processing Systems*, volume 25, pp. 1205–1213, 2012b. URL https://proceedings.neurips.cc/paper/2012/hash/dbe272bab69f8e13f14b405e038deb64-Abstract.html.

Steffen Grünewälder, Guy Lever, Luca Baldassarre, Sam Patterson, Arthur Gretton, and Massimiliano Pontil. Conditional mean embeddings as regressors. In *International Conference on Machine Learning*, 2012.

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

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, 271(1):166–180, 2011. doi: 10.1016/j.jtbi.2010.11.033.

Saksham Jain and Alex Luedtke. Conditional distributional treatment effects: Doubly robust estimation and testing. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2603.16829. In press.

Wittawat Jitkrittum, Zoltán Szabó, Kacper P Chwialkowski, and Arthur Gretton. Interpretable distribution features with maximum testing power. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Wittawat Jitkrittum, Wenkai Xu, Zoltan Szabo, Kenji Fukumizu, and Arthur Gretton. A linear-time kernel goodness-of-fit test. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Edward H. Kennedy, Sivaraman Balakrishnan, and Larry Wasserman. Semiparametric counterfactual density estimation. *Biometrika*, 110(4):875–896, 2023. doi: 10.1093/biomet/asad017. URL https://doi.org/10.1093/biomet/asad017.

Ilmun Kim and Aaditya Ramdas. Dimension-agnostic inference using cross U-statistics. *Bernoulli*, 30(1):683–711, 2024.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. URL https://arxiv.org/abs/1412.6980.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

Feng Liu, Wenkai Xu, Jie Lu, Guangquan Zhang, Arthur Gretton, and Danica J. Sutherland. Learning deep kernels for non-parametric two-sample tests. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6316–6326, 2020. URL https://proceedings.mlr.press/v119/liu20m.html.

Ruhan Liu, Xiangning Wang, Qiang Wu, Ling Dai, Xi Fang, Tao Yan, Jaemin Son, Shiqi Tang, Jiang Li, Zijian Gao, Adrian Galdran, J. M. Poorneshwaran, Hao Liu, Jie Wang, Yerui Chen, Prasanna Porwal, Gavin Siew Wei Tan, Xiaokang Yang, Chao Dai, Haitao Song, Mingang Chen, Huating Li, Weiping Jia, Dinggang Shen, Bin Sheng, and Ping Zhang. DeepDRiD: Diabetic retinopathy-grading and image quality estimation challenge. *Patterns*, 3(6):100512, 2022. doi: 10.1016/j.patter.2022.100512.

Huma Lodhi, Craig Saunders, John Shawe-Taylor, Nello Cristianini, and Chris Watkins. Text classification using string kernels. *Journal of Machine Learning Research*, 2:419–444, 2002.

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

Alexander R. Luedtke, Marco Carone, and Mark J. van der Laan. An omnibus non-parametric test of equality in distribution for unknown functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(1):75–99, 2019. doi: 10.1111/rssb.12299.

Diego Martinez-Taboada, Aaditya Ramdas, and Edward H. Kennedy. An efficient doubly-robust test for the kernel treatment effect. In *Advances in Neural Information Processing Systems*, volume 36, pp. 59924–59952, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/bccdd196d798a51a4961989984a9ed4a-Abstract-Conference.html.

Krikamol Muandet, Motonobu Kanagawa, Sorawit Saengkyongam, and Sanparith Marukatat. Counterfactual mean embeddings. *Journal of Machine Learning Research*, 22(162):1–71, 2021.

Jeffrey Näf, Junhyung Park, and Herbert Susmann. Causal-drf: Conditional kernel treatment effect estimation using distributional random forest. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 1000–1008. PMLR, 2026.

Harish Nair. MTSamples. Hugging Face dataset, 2024. URL https://huggingface.co/datasets/harishnair04/mtsamples.

M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, Jul 2002. doi: 10.1103/PhysRevE.66.016128. URL https://link.aps.org/doi/10.1103/PhysRevE.66.016128.

Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 115–124, 2005. doi: 10.3115/1219840.1219855. URL https://aclanthology.org/P05-1015/.

Junhyung Park, Uri Shalit, Bernhard Schölkopf, and Krikamol Muandet. Conditional distributional treatment effect with kernel conditional mean embeddings and U-statistic regression. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8401–8412, 2021. URL https://proceedings.mlr.press/v139/park21c.html.

Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 3982–3992. Association for Computational Linguistics, 2019. doi: 10.18653/v1/D19-1410.

JM Robins. A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7: 1393–1512, 1986.

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

Donald B. Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005. doi: 10.1198/016214504000001880.

Hiroto Saigo, Jean-Philippe Vert, Nobuhisa Ueda, and Tatsuya Akutsu. Protein homology detection using string alignment kernels. *Bioinformatics*, 20(11):1682–1689, 2004. doi: 10.1093/bioinformatics/bth141.

Antonin Schrab, Ilmun Kim, Mélisande Albert, Béatrice Laurent, Benjamin Guedj, and Arthur Gretton. MMD aggregated two-sample test. *Journal of Machine Learning Research*, 24(194): 1–81, 2023. URL https://www.jmlr.org/papers/v24/21-1289.html.

Shubhanshu Shekhar, Ilmun Kim, and Aaditya Ramdas. A permutation-free kernel two-sample test. In *Advances in Neural Information Processing Systems*, volume 35, pp. 18168–18180, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/731b952bdc833485eb72f458cdd5c489-Abstract-Conference.html.

Nino Shervashidze, Pascal Schweitzer, Erik Jan van Leeuwen, Kurt Mehlhorn, and Karsten M. Borgwardt. Weisfeiler–Lehman graph kernels. *Journal of Machine Learning Research*, 12:2539–2561, 2011.

Carl-Johann Simon-Gabriel and Bernhard Schölkopf. Kernel distribution embeddings: Universal kernels, characteristic kernels and kernel metrics on distributions. *Journal of Machine Learning Research*, 19(44):1–29, 2018.

Alex Smola, Arthur Gretton, Le Song, and Bernhard Schölkopf. A Hilbert space embedding for distributions. In *International conference on algorithmic learning theory*, pp. 13–31. Springer, 2007.

Le Song, Jonathan Huang, Alex Smola, and Kenji Fukumizu. Hilbert space embeddings of conditional distributions with applications to dynamical systems. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 961–968, 2009.

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

Danica J. Sutherland, Hsiao-Yu Tung, Heiko Strathmann, Soumyajit De, Aaditya Ramdas, Alex Smola, and Arthur Gretton. Generative models and model criticism via optimized maximum mean discrepancy. In *International Conference on Learning Representations (ICLR)*, 2017. URL https://openreview.net/forum?id=HJWHIKqgl.

Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/68053af2923e00204c3ca7c6a3150cf7-Abstract.html.

Kevin Christian Wibisono and Yixin Wang. Causal inference with unstructured outcomes, 2026. URL https://arxiv.org/abs/2608.03085. Preprint.

Ratnadira Widyasari, Sheng Qin Sim, Camellia Lok, Haodi Qi, Jack Phan, Qijin Tay, Constance Tan, Fiona Wee, Jodie Ethelda Tan, Yuheng Yieh, Brian Goh, Ferdian Thung, Hong Jin Kang, Thong Hoang, David Lo, and Eng Lieh Ouh. BugsInPy: A database of existing bugs in Python programs to enable controlled testing and debugging studies. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, ESEC/FSE 2020, pp. 1556–1560. Association for Computing Machinery, 2020. doi: 10.1145/3368089.3417943.

Andrew Gordon Wilson, Zhiting Hu, Ruslan Salakhutdinov, and Eric P. Xing. Deep kernel learning. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 370–378, 2016.

Jiancheng Yang, Rui Shi, and Bingbing Ni. MedMNIST classification decathlon: A lightweight AutoML benchmark for medical image analysis. In *IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 191–195, 2021.

Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. MedMNIST v2—a large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scientific Data*, 10(1):41, 2023.

Houssam Zenati and Arthur Gretton. Semiparametric efficient test for interpretable distributional treatment effects, 2026. URL https://arxiv.org/abs/2605.08034.

Houssam Zenati, Bariscan Bozkurt, and Arthur Gretton. Doubly-robust estimation of counterfactual policy mean embeddings. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/49a77f5557668b87a568c2d9468d4d6b-Abstract-Conference.html.

Houssam Zenati, Bariscan Bozkurt, and Arthur Gretton. Kernel treatment effects with adaptively collected data. In *The 29th International Conference on Artificial Intelligence and Statistics*, 2026. URL https://openreview.net/forum?id=di2b5POqcE.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. Qwen3 embedding: Advancing text embedding and reranking through foundation models, 2025. URL https://arxiv.org/abs/2506.05176.

APPENDIX

Appendix A recalls the main notation. Appendix B contains the theoretical results and proofs. Appendix C describes the training and testing procedure, data constructions, architectures and hyperparameters. Appendix D presents additional experiments and ablations.

A NOTATIONS

We recall the main notation used throughout the paper.

Data, interventions and samples

- $O_i = (X_i, A_i, Y_i) \in \mathcal{X} \times \mathcal{A} \times \mathcal{Y}$ denotes an observed covariate-treatment-outcome triplet; $\mathcal{D} = \{O_i\}_{i=1}^N$ is the observed dataset.
- $\mathbb{P}_0$ is the observed-data law, $\mathbb{P}_X$ its covariate marginal, $\mathbb{P}_{A,X}$ its treatment-covariate marginal, and $\mathbb{E}_0$ denotes expectation under $\mathbb{P}_0$.
- $\pi_0(a \mid x)$ is the logging-policy density; $\pi(a \mid x)$ and $\pi'(a \mid x)$ are the target intervention densities, all relative to the σ-finite treatment reference measure ρ.
- $U \in \mathcal{U}$ is the structural variable, with $U \perp\!\!\!\perp A \mid X$; the jointly measurable response $F : \mathcal{A} \times \mathcal{X} \times \mathcal{U} \to \mathcal{Y}$ defines $Y^a = F(a, X, U)$ and $Y = F(A, X, U)$.
- $A^p \mid X \sim p(\cdot \mid X)$, independently of U given X; $Y^p = F(A^p, X, U)$ has law P_Y^p. We test $H_0 : P_Y^\pi = P_Y^{\pi'}$.
- $P_{U|X}$ is a regular conditional law of U given X; $Q(B \mid a, x) = \int \mathbf{1}\{F(a, x, u) \in B\} P_{U|X}(\mathrm{d}u \mid x)$ is the common response kernel.
- $r(a, x) = \pi(a \mid x) - \pi'(a \mid x)$ is the policy contrast. Ratios and integrands weighted by r or $|r|$ are set to zero where $r = 0$.
- $\mathcal{D}_{\mathrm{learn}}$, $\mathcal{D}_{\mathrm{val}}$ and $\mathcal{D}_{\mathrm{test}}$ are the disjoint training, validation and test samples; $\mathcal{D}^{(1)}, \mathcal{D}^{(2)}, \mathcal{D}^{(3)}$ are the three learning folds.
- n is the nuisance-fitting sample size; $\mathcal{I}_1, \mathcal{I}_2$ index two evaluation halves of size m. In the test analysis, m denotes the size of each test half, also written m_{test}.

Kernels and interventional mean embeddings

- k_θ is an outcome kernel with parameter $\theta \in \Theta \subseteq \mathbb{R}^{d_\Theta}$, RKHS $\mathcal{H}_\theta$ and feature map $\varphi_\theta(y) = k_\theta(y, \cdot)$. Inner products and norms are in $\mathcal{H}_\theta$.
- $\chi_\theta(p) = \int \varphi_\theta(y) P_Y^p(\mathrm{d}y)$ is the mean embedding under intervention $p \in \{\pi, \pi'\}$.
- $\Delta_\theta = \chi_\theta(\pi) - \chi_\theta(\pi')$ is the embedding contrast and $D_\theta = \|\Delta_\theta\|_{\mathcal{H}_\theta}^2$ is the squared MMD. For characteristic k_θ, H_0 is equivalent to $D_\theta = 0$.
- $\mu_\theta(a, x) = \int \varphi_\theta(y) Q(\mathrm{d}y \mid a, x)$ is the conditional outcome mean embedding, with the null-set convention in Section B.1; it agrees $\mathbb{P}_{A,X}$-almost surely with the observational conditional mean.
- $\{k_\vartheta : \vartheta \in \Xi\}$, with $\Xi \subseteq \mathbb{R}^{d_0}$, is the family optimised by each training run. M is the ensemble size and $\theta = (\vartheta_1, \ldots, \vartheta_M) \in \Theta = \Xi^M$ collects its member parameters, with $d_\Theta = M d_0$.
- $\zeta_{\vartheta_b} : \mathcal{Y} \to \mathbb{R}^d$ is the neural feature map of member b and κ_{ϑ_b} is its Gaussian kernel on the learned features.
- $k_{\vartheta_b}^{\mathrm{char}}$ is the characteristic component, with weight $w_b \in [w_{\min}, 1]$ for $w_{\min} > 0$. The ensemble kernel is $k_\theta = M^{-1} \sum_{b=1}^M k_{\vartheta_b}$.

Nuisance estimation, kernel learning and testing

- e and m_θ are candidate logging-policy and conditional-embedding functions; $\hat{\pi}_0$ and $\hat{\mu}_\theta$ estimate π_0 and μ_θ, respectively.
- $g_\theta(O; e, m_\theta)$ is the doubly robust pseudo-feature in (1). Its oracle version is $g_\theta(O) = g_\theta(O; \pi_0, \mu_\theta)$, with $\mathbb{E}_0[g_\theta(O)] = \Delta_\theta$.
- $\widetilde{g}_\theta = g_\theta(\cdot; \hat{\pi}_0, \hat{\mu}_\theta)$ is the fitted pseudo-feature map, $\hat{g}_{\theta,i} = \widetilde{g}_\theta(O_i)$, and $\bar{g}_{\theta,\ell} = m^{-1} \sum_{i \in \mathcal{I}_\ell} \hat{g}_{\theta,i}$.
- $h_\theta(o, o') = \langle g_\theta(o), g_\theta(o') \rangle_{\mathcal{H}_\theta}$ is the oracle observed-data kernel; $\widetilde{h}_\theta$ is its fitted counterpart.
- $\widehat{D}_\theta^{\mathrm{cross}} = \langle \bar{g}_{\theta,1}, \bar{g}_{\theta,2} \rangle_{\mathcal{H}_\theta}$ is the cross-MMD estimator. $\hat{s}_\theta^2$ is the empirical variance of $U_{\theta,j} = \langle \bar{g}_{\theta,1}, \hat{g}_{\theta,j} \rangle_{\mathcal{H}_\theta}$ over $j \in \mathcal{I}_2$, with divisor m.

– $s_\theta^2 = \mathrm{Var}_{\mathbb{P}_0}(\langle \Delta_\theta, g_\theta(O)\rangle_{\mathcal{H}_\theta})$ is the oracle projected variance; $J_\theta = D_\theta / s_\theta$ is the fixed-alternative growth coefficient when $s_\theta > 0$.
– $\lambda > 0$ is the selection regulariser. The population and empirical criteria are $J_{\lambda,\theta} = D_\theta / \sqrt{s_\theta^2 + \lambda}$ and $\widehat{J}^{\mathrm{DR}}_{\lambda,\theta} = \widehat{D}^{\mathrm{cross}}_\theta / \hat{s}_{\theta,\lambda}$, where $\hat{s}^2_{\theta,\lambda} = \hat{s}^2_\theta + \lambda$.
– $\widetilde{\vartheta}_b$ denotes the EMA parameters and $\hat{\vartheta}_b$ the selected checkpoint of member b; $\hat{\theta} = (\hat{\vartheta}_1, \ldots, \hat{\vartheta}_M)$ specifies the selected ensemble.
– $T_\theta = \sqrt{m}\, \widehat{D}^{\mathrm{cross}}_\theta / \hat{s}_\theta$ is the test statistic. We reject H_0 at level α when $\hat{p} = 1 - \Phi(T_\theta) < \alpha$, where Φ is the standard normal distribution function.

Notation for the theoretical analysis

– $\mathcal{F}_{\mathrm{fit}}$ is the information used for nuisance fitting in the selection analysis; $\mathcal{F}_{\mathrm{learn}}$ contains all nuisance fitting, kernel training and selection information available before the final test. $\mathcal{F}_{\mathrm{data}}$ contains the common learning and validation data and shared training randomness, excluding member-specific randomness.
– $\widetilde{\Delta}_\theta$, $\widetilde{D}_\theta$ and $\widetilde{s}^2_\theta$ are the conditional population counterparts of Δ_θ, D_θ and s^2_θ formed from $\widetilde{g}_\theta$, holding the nuisance fits fixed. The label or marks empirical quantities computed with oracle pseudo-features.
– $\epsilon_{\pi,n}$ and $\epsilon_{\mu,n}$ are the conditional root-mean-square nuisance errors weighted by $|r|$: relative propensity error and conditional-embedding error, the latter uniform over Θ, as defined in (15).
– R_Θ bounds the Euclidean radius of Θ; $c_m(\eta)$ is the kernel-class sampling term in (16), with failure probability $\eta \in (0, 1)$.
– κ, W, A, B_μ bound the kernel diagonal, $|r|/\pi_0$, relative propensity error and fitted embedding norm, respectively. L and L_h are the parameter Lipschitz constants of the fitted and oracle observed-data kernels.
– $\Theta_s = \{\theta \in \Theta : s^2_\theta \geq s^2\}$, for $s > 0$, is the population comparison class used in the adaptation results; λ_m is a positive sequence of selection regularisers.
– $b_{\mathrm{test}} = \mathbb{E}_0[\hat{g}_{\hat{\theta}}(O) \mid \mathcal{F}_{\mathrm{learn}}]$ is the fitted-score bias under H_0, and $\xi(O) = \hat{g}_{\hat{\theta}}(O) - b_{\mathrm{test}}$ is the centred test score.
– $a_{\pi,\mathrm{test}}$ and $b_{\mu,\mathrm{test}}$ are the weighted nuisance errors for the selected kernel; their product bounds $\|b_{\mathrm{test}}\|_{\mathcal{H}_{\hat{\theta}}}$ under H_0.

B Theoretical Results and Proofs

We prove the results in Sections 3 and 4, then give further results on nuisance errors and the learned test under alternatives. We use the notation of the main text, suppressing the fixed-policy superscript on g_θ when unambiguous. We write $O_{\mathbb{P}}(1)$ for quantities bounded in probability and $o_{\mathbb{P}}(1)$ for quantities converging to zero in probability.

B.1 Identification of Soft-Intervention Outcome Laws

Proposition B.1 (Soft-intervention identification)**.** *Under Assumption 3.1, fix a regular conditional law $P_{U|X}$ and define $Q(B \mid a, x) := \int_{\mathcal{U}} \mathbf{1}\{F(a, x, u) \in B\} P_{U|X}(\mathrm{d}u \mid x)$. Then:*

(a) Q is a measurable probability kernel from $\mathcal{A} \times \mathcal{X}$ to $\mathcal{Y}$.

(b) Q is a version of the observational conditional law of Y given (A, X).

(c) For every fixed $a \in \mathcal{A}$, $Q(\cdot \mid a, X)$ is a version of the conditional law of Y^a given X.

(d) For any measurable conditional density p relative to ρ, draw $A^p \mid X$ from $p(\cdot \mid X)$ independently of U given X, and set $Y^p = F(A^p, X, U)$. Its law satisfies, for every measurable $B \subseteq \mathcal{Y}$,

$$P^p_Y(B) = \int_{\mathcal{X}} \int_{\mathcal{A}} Q(B \mid a, x) p(a \mid x) \rho(\mathrm{d}a) \mathbb{P}_X(\mathrm{d}x).$$

(e) If $p = 0$ wherever $\pi_0 = 0$, up to a $\mathbb{P}_X \otimes \rho$-null set, the integral in (d) is unchanged when Q is replaced by any observational conditional-law version. In particular, P^π_Y and $P^{\pi'}_Y$ are identified.

Proof. $\mathcal{X}$ and $\mathcal{U}$ are standard Borel spaces, hence a regular conditional law $P_{U|X}$ exists. For fixed (a, x), $Q(\cdot \mid a, x)$ is the distribution obtained by applying $u \mapsto F(a, x, u)$ to $P_{U|X}(\cdot \mid x)$, so it is a probability measure. For each measurable B, the map $(a, x, u) \mapsto \mathbf{1}\{F(a, x, u) \in B\}$ is measurable because F is jointly measurable. Integrating over $P_{U|X}(\,\mathrm{d}u \mid x)$ preserves measurability in (a, x). This holds for indicators of measurable rectangles and extends to nonnegative measurable functions by a monotone-class argument. Thus Q is a measurable probability kernel, proving (a).

For (b), conditional exogeneity implies that the joint law of (A, X, U) factors as $\mathbb{P}_{A,X}(\,\mathrm{d}a,\,\mathrm{d}x)P_{U|X}(\,\mathrm{d}u \mid x)$. Using consistency, $Y = F(A, X, U)$, we therefore obtain, for every measurable $C \subseteq \mathcal{A} \times \mathcal{X}$,

$$\begin{aligned}\mathbb{P}_0\{Y \in B, (A, X) \in C\} &= \int_C \int_{\mathcal{U}} \mathbf{1}\{F(a, x, u) \in B\} P_{U|X}(\,\mathrm{d}u \mid x)\mathbb{P}_{A,X}(\,\mathrm{d}a,\,\mathrm{d}x) \\ &= \int_C Q(B \mid a, x)\mathbb{P}_{A,X}(\,\mathrm{d}a,\,\mathrm{d}x).\end{aligned}$$

This is the defining property of an observational conditional-law version. For (c), let a be fixed. For every measurable $D \subseteq \mathcal{X}$, the definition $Y^a = F(a, X, U)$ gives

$$\Pr\{Y^a \in B, X \in D\} = \int_D Q(B \mid a, x)\mathbb{P}_X(\,\mathrm{d}x).$$

Thus $Q(\cdot \mid a, X)$ is a conditional law of Y^a given X for each fixed a; this does not identify it observationally at every action.

For (d), adjoin A^p on a probability-space extension using the kernel $p(a \mid x)\rho(\,\mathrm{d}a)$ while preserving (X, U). The resulting joint law is $\mathbb{P}_X(\,\mathrm{d}x)P_{U|X}(\,\mathrm{d}u \mid x)p(a \mid x)\rho(\,\mathrm{d}a)$. Integrating $\mathbf{1}\{F(a, x, u) \in B\}$ yields the stated formula by Tonelli's theorem.

For (e), the support condition and the logging density give

$$p(a \mid x)\rho(\,\mathrm{d}a)\mathbb{P}_X(\,\mathrm{d}x) = \frac{p(a \mid x)}{\pi_0(a \mid x)}\mathbb{P}_{A,X}(\,\mathrm{d}a,\,\mathrm{d}x),$$

with the ratio set to zero where $\pi_0 = 0$. Hence the target treatment-covariate measure is absolutely continuous with respect to $\mathbb{P}_{A,X}$. Any two observational conditional-law versions agree as probability measures outside a single $\mathbb{P}_{A,X}$-null set: apply conditional-expectation uniqueness to a countable determining class for the standard Borel space $\mathcal{Y}$. Absolute continuity makes this set null under the target measure as well, proving invariance of (d). In particular, any admissible choice of $P_{U|X}$ yields the same supported soft-intervention laws. □

Mean embeddings and integrability. For fixed θ, define $\mu_\theta(a, x) = \int \varphi_\theta(y)Q(\,\mathrm{d}y \mid a, x)$ wherever $\int \sqrt{k_\theta(y, y)}Q(\,\mathrm{d}y \mid a, x) < \infty$, and set it to zero elsewhere. This exceptional set is measurable and null under both $\mathbb{P}_{A,X}$ and the two target treatment-covariate measures by Assumption 3.2 and parts (b)-(d). Separability and strongly measurable approximation under kernel integration give a strongly measurable μ_θ. Part (b) identifies $\mu_\theta(A, X)$ with the observational Bochner conditional mean, and part (d) with Bochner Fubini gives

$$\chi_\theta(p) = \int_{\mathcal{X}} \int_{\mathcal{A}} \mu_\theta(a, x)p(a \mid x)\rho(\,\mathrm{d}a)\mathbb{P}_X(\,\mathrm{d}x).$$

The observed-law first moment in Assumption 3.2 is needed for the observational conditional mean, since target moments need not control actions outside the target support. Uniformly bounded kernels satisfy all these moment conditions.

For finite or countable $\mathcal{A}$ with counting measure ρ, the action integrals reduce to sums, giving the usual randomised mixture of atomic interventions.

B.2 Observed-Data Representation and Double Robustness

For Proposition 3.3, the candidate nuisances are assumed to satisfy

$$\mathbb{E}_0\left[\frac{|r(A,X)|}{e(A \mid X)}\|\varphi_\theta(Y) - m_\theta(A,X)\|_{\mathcal{H}_\theta}\right] < \infty,$$
$$\mathbb{E}_0\left[\int_{\mathcal{A}} |r(a,X)|\|m_\theta(a,X)\|_{\mathcal{H}_\theta}\rho(\,\mathrm{d}a)\right] < \infty.$$

These conditions make both terms of the pseudo-feature Bochner integrable and justify the conditional integration below; the target moments also ensure $\mathbb{E}_0 \int |r(a,X)|\|\mu_\theta(a,X)\|_{\mathcal{H}_\theta}\rho(\,\mathrm{d}a) < \infty$.

Proof of Proposition 3.3. Recall that $r(a,x) = \pi(a \mid x) - \pi'(a \mid x)$. As in Section 3, ratios and integrands weighted by r or $|r|$ are set to zero wherever $r = 0$.

Under Assumptions 3.1 and 3.2, Proposition B.1 and Bochner Fubini identify each target embedding through the common response kernel. Subtracting the two embeddings gives

$$\Delta_\theta(\pi,\pi') = \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}} r(a,X)\mu_\theta(a,X)\rho(\,\mathrm{d}a)\right].$$

The weight $r(A,X)/e(A \mid X)$ is fixed conditional on (A,X), and $\mu_\theta(A,X)$ agrees almost surely with the observational conditional mean. The integrability conditions above therefore give

$$\mathbb{E}_{\mathbb{P}_0}\left[\frac{r(A,X)}{e(A \mid X)}\{\varphi_\theta(Y) - m_\theta(A,X)\} \mid A,X\right] = \frac{r(A,X)}{e(A \mid X)}\{\mu_\theta(A,X) - m_\theta(A,X)\}.$$

Averaging over $A \mid X$, with density π_0, gives

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}_0}\left[g_\theta^{\pi,\pi'}(O;e,m_\theta)\right] \\
&\quad= \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}} \pi_0(a \mid X)\frac{r(a,X)}{e(a \mid X)}\{\mu_\theta(a,X) - m_\theta(a,X)\}\rho(\,\mathrm{d}a)\right] \\
&\qquad+ \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}} r(a,X)m_\theta(a,X)\rho(\,\mathrm{d}a)\right].
\end{aligned}$$

We now subtract the expression for $\Delta_\theta(\pi,\pi')$:

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}_0}\left[g_\theta^{\pi,\pi'}(O;e,m_\theta)\right] - \Delta_\theta(\pi,\pi') \\
&\quad= \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}} r(a,X)\left\{\frac{\pi_0(a \mid X)}{e(a \mid X)} - 1\right\}\{\mu_\theta(a,X) - m_\theta(a,X)\}\rho(\,\mathrm{d}a)\right].
\end{aligned}$$

This expression is zero if either $e = \pi_0$ or $m_\theta = \mu_\theta$ wherever $|r| > 0$. Thus, if at least one nuisance function is correct,

$$\mathbb{E}_{\mathbb{P}_0}\left[g_\theta^{\pi,\pi'}(O;e,m_\theta)\right] = \Delta_\theta(\pi,\pi').$$

Using the true nuisance functions gives $\mathbb{E}_{\mathbb{P}_0}[g_\theta^{\pi,\pi'}(O)] = \Delta_\theta(\pi,\pi')$. □

B.3 Fixed-Alternative Growth Rate

Proof of Proposition 4.1. Fix θ. We take all inner products and norms in $\mathcal{H}_\theta$. We assume the statistics use the true nuisance functions, so $\hat{g}_{\theta,i}^{\pi,\pi'} = g_\theta^{\pi,\pi'}(O_i)$ and

$$\bar{g}_{\theta,\ell} = \frac{1}{m}\sum_{i \in \mathcal{I}_\ell} g_\theta^{\pi,\pi'}(O_i), \qquad \ell \in \{1,2\}.$$

By Proposition 3.3, $\mathbb{E}_{\mathbb{P}_0}[g_\theta^{\pi,\pi'}(O_i)] = \Delta_\theta(\pi,\pi')$. The fourth-moment assumption gives finite first and second moments, so the RKHS law of large numbers gives $\bar{g}_{\theta,\ell} \xrightarrow{\mathbb{P}} \Delta_\theta(\pi,\pi')$ for $\ell \in \{1,2\}$. Continuity of the inner product gives

$$\widehat{D}_\theta^{\mathrm{cross}} = \langle \bar{g}_{\theta,1}, \bar{g}_{\theta,2} \rangle \xrightarrow{\mathbb{P}} \|\Delta_\theta(\pi,\pi')\|^2 = D_\theta.$$

For the variance estimator, define the real-valued variables

$$V_j := \langle \bar{g}_{\theta,1}, g_\theta^{\pi,\pi'}(O_j)\rangle, \qquad W_j := \langle \Delta_\theta(\pi,\pi'), g_\theta^{\pi,\pi'}(O_j)\rangle, \qquad j \in \mathcal{I}_2.$$

The V_j have sample mean $\widehat{D}_\theta^{\text{cross}}$. Expanding their sample variance gives

$$\hat{s}_\theta^2 = \frac{1}{m}\sum_{j\in\mathcal{I}_2} V_j^2 - \left(\widehat{D}_\theta^{\text{cross}}\right)^2.$$

We have already shown that $(\widehat{D}_\theta^{\text{cross}})^2 \xrightarrow{\mathbb{P}} D_\theta^2$. To find the limit of $\frac{1}{m}\sum_{j\in\mathcal{I}_2} V_j^2$, we compare V_j with W_j. By $|u^2 - w^2| = |u-w||u+w|$ and Cauchy-Schwarz,

$$\begin{aligned}\left|\frac{1}{m}\sum_{j\in\mathcal{I}_2}(V_j^2 - W_j^2)\right| &\le \|\bar{g}_{\theta,1} - \Delta_\theta(\pi,\pi')\|\big(\|\bar{g}_{\theta,1}\| + \|\Delta_\theta(\pi,\pi')\|\big)\frac{1}{m}\sum_{j\in\mathcal{I}_2}\|g_\theta^{\pi,\pi'}(O_j)\|^2 \\ &= o_{\mathbb{P}}(1).\end{aligned}$$

Here the first factor converges to zero, the second is bounded in probability, and the last converges to $\mathbb{E}_{\mathbb{P}_0}\|g_\theta^{\pi,\pi'}(O)\|^2 < \infty$ by the law of large numbers. The same law applied to W_j^2 gives

$$\frac{1}{m}\sum_{j\in\mathcal{I}_2} W_j^2 \xrightarrow{\mathbb{P}} \mathbb{E}_{\mathbb{P}_0}\left[\langle \Delta_\theta(\pi,\pi'), g_\theta^{\pi,\pi'}(O)\rangle^2\right].$$

We also have $\mathbb{E}_{\mathbb{P}_0}[W_j] = \langle \Delta_\theta(\pi,\pi'), \Delta_\theta(\pi,\pi')\rangle = D_\theta$. Combining these gives

$$\hat{s}_\theta^2 \xrightarrow{\mathbb{P}} \mathbb{E}_{\mathbb{P}_0}\left[\langle \Delta_\theta(\pi,\pi'), g_\theta^{\pi,\pi'}(O)\rangle^2\right] - D_\theta^2 = s_\theta^2.$$

Since $s_\theta > 0$, we have $\hat{s}_\theta \xrightarrow{\mathbb{P}} s_\theta$. By (2) and Slutsky's theorem,

$$\frac{T_\theta}{\sqrt{m}} = \frac{\widehat{D}_\theta^{\text{cross}}}{\hat{s}_\theta} \xrightarrow{\mathbb{P}} \frac{D_\theta}{s_\theta} = J_\theta.$$

We set $T_\theta = 0$ when $\hat{s}_\theta = 0$. Since this event has probability tending to zero, it does not affect the limit. □

B.4 Characteristic Ensemble Safeguard

This result is the analogue of Liu et al. (2020, Proposition 5). Their kernel in Eq. (1), however, additionally factors the characteristic component and the deep neural component.

Proof of Lemma 4.2. Let $Y, Y' \sim P$ and $Z, Z' \sim Q$ be mutually independent. Then

$$\mathrm{MMD}_k^2(P,Q) = \mathbb{E}[k(Y,Y')] + \mathbb{E}[k(Z,Z')] - 2\mathbb{E}[k(Y,Z)].$$

The squared MMD is therefore linear in the kernel. For each ensemble member b, write

$$k_{\vartheta_b}^{\text{deep}}(y,y') := \kappa_{\vartheta_b}\big(\zeta_{\vartheta_b}(y), \zeta_{\vartheta_b}(y')\big).$$

Substituting the ensemble kernel form in (4) yields

$$\mathrm{MMD}_{k_\theta}^2(P,Q) = \frac{1}{M}\sum_{b=1}^{M}\left[(1-w_b)\,\mathrm{MMD}_{k_{\vartheta_b}^{\text{deep}}}^2(P,Q) + w_b\,\mathrm{MMD}_{k_{\vartheta_b}^{\text{char}}}^2(P,Q)\right].$$

Every squared MMD is nonnegative, and $1 - w_b \ge 0$. Dropping the deep kernel terms and using $w_b \ge w_{\min}$ gives

$$\mathrm{MMD}_{k_\theta}^2(P,Q) \ge \frac{w_{\min}}{M}\sum_{b=1}^{M}\mathrm{MMD}_{k_{\vartheta_b}^{\text{char}}}^2(P,Q).$$

Each base kernel is characteristic and $w_{\min} > 0$, so the lower bound is strictly positive when $P \neq Q$. When $P = Q$, every squared MMD term is zero. Thus k_θ is characteristic. □

B.5 MEMBERWISE ENSEMBLE AGGREGATION

Proposition B.2 (Memberwise ensemble aggregation)**.** *Suppose Assumptions 3.1 and 3.2 hold. Let $k_{\vartheta_1}, \dots, k_{\vartheta_M}$ be fixed kernel members and $k_\theta = M^{-1}\sum_b k_{\vartheta_b}$. Use the oracle pseudo-features $g_{\vartheta_b} = g_{\vartheta_b}(\cdot; \pi_0, \mu_{\vartheta_b})$ based on the common response kernel Q, and assume each projection $\langle \Delta_{\vartheta_b}, g_{\vartheta_b}(O)\rangle_{\mathcal{H}_{\vartheta_b}}$ is square-integrable. For $\lambda \geq 0$, define $J_{\lambda,\vartheta_b} = D_{\vartheta_b}/\sqrt{s^2_{\vartheta_b} + \lambda}$ and analogously for the aggregate whenever the denominators are positive. Then*

$$D_\theta = \frac{1}{M}\sum_{b=1}^{M} D_{\vartheta_b}, \qquad \sqrt{s^2_\theta + \lambda} \leq \frac{1}{M}\sum_{b=1}^{M}\sqrt{s^2_{\vartheta_b} + \lambda}.$$

If all member and aggregate denominators are positive, then

$$J_{\lambda,\theta} \geq \frac{\sum_{b=1}^{M} D_{\vartheta_b}}{\sum_{b=1}^{M}\sqrt{s^2_{\vartheta_b} + \lambda}} \geq \min_{1\leq b\leq M} J_{\lambda,\vartheta_b}.$$

For learned parameters $\hat\theta = (\hat\vartheta_1, \dots, \hat\vartheta_M)$, let $\mathcal{F}_{\mathrm{data}}$ contain the common learning and validation data and all shared training randomness. If the selected member kernels are conditionally independent given $\mathcal{F}_{\mathrm{data}}$ and the $D_{\hat\vartheta_b}$ have finite conditional second moments, then

$$\mathrm{Var}(D_{\hat\theta} \mid \mathcal{F}_{\mathrm{data}}) = \frac{1}{M^2}\sum_{b=1}^{M}\mathrm{Var}(D_{\hat\vartheta_b} \mid \mathcal{F}_{\mathrm{data}}).$$

If the members are also conditionally identically distributed, this equals $M^{-1}\,\mathrm{Var}(D_{\hat\vartheta_1} \mid \mathcal{F}_{\mathrm{data}})$. This variance comes from member-specific training randomness, not sampling variability of the final test statistic.

Proof. In the Hilbert direct sum $\bigoplus_{b=1}^{M}\mathcal{H}_{\vartheta_b}$, the averaged kernel has the isometric feature representation

$$\varphi_\theta(y) = M^{-1/2}\big(\varphi_{\vartheta_1}(y), \dots, \varphi_{\vartheta_M}(y)\big).$$

Linearity of Bochner integration gives $\Delta_\theta = M^{-1/2}(\Delta_{\vartheta_1}, \dots, \Delta_{\vartheta_M})$ and the same identity for the conditional embeddings. Using the common π_0 and Q in (1) therefore gives $g_\theta(O) = M^{-1/2}(g_{\vartheta_1}(O), \dots, g_{\vartheta_M}(O))$. Taking squared norms proves the discrepancy identity, and

$$\langle \Delta_\theta, g_\theta(O)\rangle_{\mathcal{H}_\theta} = \frac{1}{M}\sum_{b=1}^{M}\langle \Delta_{\vartheta_b}, g_{\vartheta_b}(O)\rangle_{\mathcal{H}_{\vartheta_b}}.$$

Let $Z_b := \langle \Delta_{\vartheta_b}, g_{\vartheta_b}(O)\rangle_{\mathcal{H}_{\vartheta_b}} - D_{\vartheta_b}$. By Proposition 3.3, $\mathbb{E}_0 Z_b = 0$, so $s_{\vartheta_b} = \|Z_b\|_{L^2(\mathbb{P}_0)}$ and $s_\theta = \|M^{-1}\sum_b Z_b\|_{L^2(\mathbb{P}_0)}$. The triangle inequality in $L^2(\mathbb{P}_0)\oplus\mathbb{R}$ yields

$$\sqrt{s^2_\theta + \lambda} = \left\|\frac{1}{M}\sum_{b=1}^{M}(Z_b, \sqrt{\lambda})\right\| \leq \frac{1}{M}\sum_{b=1}^{M}\|(Z_b, \sqrt{\lambda})\| = \frac{1}{M}\sum_{b=1}^{M}\sqrt{s^2_{\vartheta_b} + \lambda}.$$

Since $D_{\vartheta_b} \geq 0$, the denominator bound proves the first efficacy inequality. The intervening ratio is a weighted average of the member efficacies with positive weights proportional to $\sqrt{s^2_{\vartheta_b} + \lambda}$, proving the second. No independence of the projected scores on a common observation is required. Finally, the discrepancy identity holds for every realised collection of learned kernels; conditional independence eliminates its cross-covariances, and identical conditional distributions give the $1/M$ formula. □

B.6 HELD-OUT NULL VALIDITY

The sample-splitting principle also appears in Gretton et al. (2012b); Liu et al. (2020). Here we additionally control bias from nuisance estimation.

Let $\mathcal{F}_{\text{learn}}$ contain the data and randomness used for nuisance fitting, kernel training and selection. Conditional on $\mathcal{F}_{\text{learn}}$, the fitted map $\hat{g}_{\hat{\theta}}(O) := g_{\hat{\theta}}^{\pi,\pi'}(O; \hat{\pi}_0, \hat{\mu}_{\hat{\theta}})$ is fixed. Write m for the size of each test half. For a new observation $O \sim \mathbb{P}_0$ independent of $\mathcal{F}_{\text{learn}}$, define

$$b_{\text{test}} := \mathbb{E}_{\mathbb{P}_0}[\hat{g}_{\hat{\theta}}(O) \mid \mathcal{F}_{\text{learn}}],$$
$$\xi(O) := \hat{g}_{\hat{\theta}}(O) - b_{\text{test}}, \qquad \bar{\xi}_\ell := \frac{1}{m}\sum_{i \in \mathcal{I}_\ell} \xi(O_i), \qquad \ell \in \{1, 2\}.$$

b_{test} is the bias from nuisance estimation under H_0.

Let $\mathcal{G}_m$ denote the information from the learning procedure $\mathcal{F}_{\text{learn}}$ and the first test half. $\bar{\xi}_1$ is therefore fixed when conditioned on $\mathcal{G}_m$. For a new observation $O \sim \mathbb{P}_0$ independent of $\mathcal{G}_m$, define

$$V(O) := \langle \bar{\xi}_1, \xi(O) \rangle_{\mathcal{H}_{\hat{\theta}}}, \qquad v_{\text{test}} := \mathbb{E}_{\mathbb{P}_0}[V(O)^2 \mid \mathcal{G}_m].$$

Since $V(O)$ has conditional mean zero, v_{test} is its variance conditioned on $\mathcal{G}_m$.

Assumption B.3 (Held-out regularity)**.** *Let $m \to \infty$. Suppose the two test halves are mutually independent i.i.d. samples from $\mathbb{P}_0$ and are independent of $\mathcal{F}_{\text{learn}}$. Assume that $\hat{g}_{\hat{\theta}}$ is strongly measurable and conditionally Bochner integrable. Under H_0, suppose that:*

1. *The conditional fourth moment of the centred pseudo-feature is bounded in probability:*
$$\mathbb{E}_{\mathbb{P}_0}[\|\xi(O)\|_{\mathcal{H}_{\hat{\theta}}}^4 \mid \mathcal{F}_{\text{learn}}] = O_{\mathbb{P}}(1).$$

2. *The projected variance is nonzero with probability tending to one, $\mathbb{P}(v_{\text{test}} > 0) \to 1$, and*
$$(m v_{\text{test}})^{-1} = O_{\mathbb{P}}(1), \qquad \frac{\mathbb{E}_{\mathbb{P}_0}[V(O)^4 \mid \mathcal{G}_m]}{m v_{\text{test}}^2} \xrightarrow{\mathbb{P}} 0.$$
The inverse and ratio are defined to be zero when $v_{\text{test}} = 0$.

3. *The conditional mean tends to zero faster than $m^{-1/2}$:*
$$\sqrt{m}\|b_{\text{test}}\|_{\mathcal{H}_{\hat{\theta}}} \xrightarrow{\mathbb{P}} 0.$$

Double robustness gives a sufficient nuisance-rate condition for condition 3 of Assumption B.3. Under Assumptions 3.1 and 3.2, define the weighted nuisance errors

$$a_{\pi,\text{test}}^2 := \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}} |r(a,X)| \left\{\frac{\pi_0(a \mid X)}{\hat{\pi}_0(a \mid X)} - 1\right\}^2 \rho(\mathrm{d}a) \middle| \mathcal{F}_{\text{learn}}\right],$$
$$b_{\mu,\text{test}}^2 := \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}} |r(a,X)| \|\hat{\mu}_{\hat{\theta}}(a,X) - \mu_{\hat{\theta}}(a,X)\|_{\mathcal{H}_{\hat{\theta}}}^2 \rho(\mathrm{d}a) \middle| \mathcal{F}_{\text{learn}}\right].$$

Under H_0, $\Delta_{\hat{\theta}}(\pi, \pi') = 0$. If $\hat{\pi}_0$ is positive on the support of $|r|$ and the terms in the double robustness identity are integrable, taking norms in the identity of Proposition 3.3 and applying Cauchy-Schwarz with weight $|r(a,X)|$ gives

$$\|b_{\text{test}}\|_{\mathcal{H}_{\hat{\theta}}} \le a_{\pi,\text{test}} b_{\mu,\text{test}}. \tag{9}$$

$a_{\pi,\text{test}} b_{\mu,\text{test}} = o_{\mathbb{P}}(m^{-1/2})$ is therefore sufficient for condition 3.

Proof of Theorem 4.3. This argument also covers a fixed kernel by taking $\hat{\theta} = \theta$ deterministically and retaining nuisance-fitting data and randomness in $\mathcal{F}_{\text{learn}}$. Write $b = b_{\text{test}}$ and $v = v_{\text{test}}$. All inner products and norms are in $\mathcal{H}_{\hat{\theta}}$.

Let $V_j := V(O_j)$ for $j \in \mathcal{I}_2$, and define

$$\widehat{D}^{\text{c}} := \frac{1}{m}\sum_{j \in \mathcal{I}_2} V_j = \langle \bar{\xi}_1, \bar{\xi}_2 \rangle,$$
$$(\hat{s}^{\text{c}})^2 := \frac{1}{m}\sum_{j \in \mathcal{I}_2} V_j^2 - (\widehat{D}^{\text{c}})^2, \qquad T^{\text{c}} := \frac{\sqrt{m}\widehat{D}^{\text{c}}}{\hat{s}^{\text{c}}}.$$

Set $T^{\mathrm{c}} = 0$ when $\hat{s}^{\mathrm{c}} = 0$.

V_j are i.i.d., conditional on $\mathcal{G}_m$, with mean zero and variance v. On $\{v > 0\}$, the fourth-moment ratio in condition 2 is the Lyapunov condition, so $\sqrt{m}\widehat{D}^{\mathrm{c}}/\sqrt{v}$ has a conditional standard normal limit. The sample-variance identity above and conditional Chebyshev and Markov inequalities give, for every $\delta > 0$,

$$\mathbb{P}\left(\left|\frac{(\hat{s}^{\mathrm{c}})^2}{v} - 1\right| > \delta \middle| \mathcal{G}_m\right) \le \frac{4\mathbb{E}_{\mathbb{P}_0}[V(O)^4 \mid \mathcal{G}_m]}{\delta^2 m v^2} + \frac{2}{m\delta} \xrightarrow{\mathbb{P}} 0.$$

Here the two bounds control the sample second moment and the squared sample mean, using $\mathbb{E}_{\mathbb{P}_0}[(\widehat{D}^{\mathrm{c}})^2/v \mid \mathcal{G}_m] = 1/m$. Applying Slutsky's theorem conditionally gives

$$d_m := \sup_{t\in\mathbb{R}} |\mathbb{P}(T^{\mathrm{c}} \le t \mid \mathcal{G}_m) - \Phi(t)| \xrightarrow{\mathbb{P}} 0.$$

Since $\mathbb{P}(v > 0) \to 1$ and $(\hat{s}^{\mathrm{c}})^2/v \xrightarrow{\mathbb{P}} 1$, we have $\mathbb{P}(\hat{s}^{\mathrm{c}} > 0) \to 1$. On $\{\hat{s}^{\mathrm{c}} > 0\}$, $(mv)^{-1} = O_{\mathbb{P}}(1)$ also gives $(\hat{s}^{\mathrm{c}})^{-1} = O_{\mathbb{P}}(\sqrt{m})$.

Conditional independence and centring give

$$\mathbb{E}_{\mathbb{P}_0}[\|\bar{\xi}_\ell\|^2 \mid \mathcal{F}_{\mathrm{learn}}] = \frac{1}{m}\mathbb{E}_{\mathbb{P}_0}[\|\xi(O)\|^2 \mid \mathcal{F}_{\mathrm{learn}}] = O_{\mathbb{P}}(m^{-1}).$$

By condition 1 and conditional Markov's inequality,

$$\|\bar{\xi}_\ell\| = O_{\mathbb{P}}(m^{-1/2}), \qquad \frac{1}{m}\sum_{j\in\mathcal{I}_2} \|\xi(O_j)\|^2 = O_{\mathbb{P}}(1).$$

Since the fitted half-sample means are $\bar{\xi}_\ell + b$, condition 3 and Cauchy-Schwarz give

$$\begin{aligned}\left|\widehat{D}^{\mathrm{cross}}_{\hat{\theta}} - \widehat{D}^{\mathrm{c}}\right| &= \left|\langle b, \bar{\xi}_1 + \bar{\xi}_2\rangle + \|b\|^2\right| \\ &\le \|b\|\left(\|\bar{\xi}_1\| + \|\bar{\xi}_2\|\right) + \|b\|^2 = o_{\mathbb{P}}(m^{-1}).\end{aligned}$$

Applying the reverse triangle inequality to the residual vectors and Cauchy-Schwarz gives

$$\begin{aligned}|\hat{s}_{\hat{\theta}} - \hat{s}^{\mathrm{c}}| &\le \|b\| \left\{\frac{1}{m}\sum_{j\in\mathcal{I}_2} \|\xi(O_j) - \bar{\xi}_2\|^2\right\}^{1/2} \\ &\le \|b\| \left\{\frac{1}{m}\sum_{j\in\mathcal{I}_2} \|\xi(O_j)\|^2\right\}^{1/2} = o_{\mathbb{P}}(m^{-1/2}).\end{aligned}$$

On $\{\hat{s}^{\mathrm{c}} > 0\}$, combining these bounds with $(\hat{s}^{\mathrm{c}})^{-1} = O_{\mathbb{P}}(\sqrt{m})$ yields

$$\frac{\sqrt{m}(\widehat{D}^{\mathrm{cross}}_{\hat{\theta}} - \widehat{D}^{\mathrm{c}})}{\hat{s}^{\mathrm{c}}} = o_{\mathbb{P}}(1), \qquad \frac{\hat{s}_{\hat{\theta}}}{\hat{s}^{\mathrm{c}}} \xrightarrow{\mathbb{P}} 1.$$

Hence $\mathbb{P}(\hat{s}_{\hat{\theta}} > 0) \to 1$. Since $T^{\mathrm{c}} = O_{\mathbb{P}}(1)$, the fitted statistic satisfies $T_{\hat{\theta}} - T^{\mathrm{c}} = o_{\mathbb{P}}(1)$.

Since $0 \le d_m \le 1$ and $d_m \xrightarrow{\mathbb{P}} 0$, we have $\mathbb{E}[d_m] \to 0$. Therefore $\mathbb{E}[d_m \mid \mathcal{F}_{\mathrm{learn}}] \xrightarrow{\mathbb{P}} 0$. For any fixed $\delta > 0$,

$$q_m(\delta) := \mathbb{P}\left(|T_{\hat{\theta}} - T^{\mathrm{c}}| > \delta \mid \mathcal{F}_{\mathrm{learn}}\right) \xrightarrow{\mathbb{P}} 0,$$

by Markov's inequality, since $\mathbb{E}[q_m(\delta)] = \mathbb{P}(|T_{\hat{\theta}} - T^{\mathrm{c}}| > \delta) \to 0$.

Averaging over the first test half gives the bound $\mathbb{E}[d_m \mid \mathcal{F}_{\mathrm{learn}}]$ for the normal-approximation error of T^{c}. The fitted and centred statistics differ by at most δ, except with conditional probability $q_m(\delta)$. Comparing their distribution functions at $t - \delta$ and $t + \delta$ gives

$$\begin{aligned}&\sup_{t\in\mathbb{R}} \left|\mathbb{P}(T_{\hat{\theta}} \le t \mid \mathcal{F}_{\mathrm{learn}}) - \Phi(t)\right| \\ &\qquad \le \mathbb{E}[d_m \mid \mathcal{F}_{\mathrm{learn}}] + q_m(\delta) + \frac{\delta}{\sqrt{2\pi}},\end{aligned}$$

where the last term follows because the standard normal density is bounded by $1/\sqrt{2\pi}$. Letting $m \to \infty$ and then $\delta \to 0$ proves (6). The left-hand side is bounded by one, so taking expectations gives the unconditional normal limit and asymptotic level α. □

B.7 Uniform Control of the Learning Criterion

We prove Theorem 4.4 by separating sampling error from nuisance error. Throughout this subsection, expectations, variances and norms involving fitted functions are conditional on $\mathcal{F}_{\mathrm{fit}}$, the data and randomness used to fit the nuisance estimators. This fixes the fitted functions for every θ. Write n for the fitting-sample size and m for the size of each selection half. The two halves are mutually independent i.i.d. samples from $\mathbb{P}_0$ and are independent of $\mathcal{F}_{\mathrm{fit}}$. For a separable Hilbert space $\mathcal{H}$, let $\|f\|^2_{L^2(\mathbb{P}_0;\mathcal{H})} := \mathbb{E}_0\|f(O)\|^2_{\mathcal{H}}$.

Write $\mathcal{O} := \mathcal{X} \times \mathcal{A} \times \mathcal{Y}$. For the ensemble in (4), let $\Theta = \Xi^M$, where $\Xi \subseteq \mathbb{R}^{d_0}$ is the member parameter set and $d_\Theta = Md_0$. For the memberwise optimisation in Algorithm 1, we apply the learning bound with $M = 1$, $\Theta = \Xi$ and $d_\Theta = d_0$. For $\vartheta \in \Xi$, write $k^{\mathrm{deep}}_\vartheta$ for the learned component and w_ϑ for its mixture weight, with $w_{\vartheta_b} = w_b$.

Let $\widetilde{g}_\theta := g_\theta(\cdot; \hat{\pi}_0, \hat{\mu}_\theta)$ denote the fitted pseudo-feature. Its conditional population quantities are

$$\widetilde{\Delta}_\theta := \mathbb{E}_0[\widetilde{g}_\theta(O)], \qquad \widetilde{D}_\theta := \|\widetilde{\Delta}_\theta\|^2_{\mathcal{H}_\theta},$$
$$\widetilde{s}^2_\theta := \mathrm{Var}_{\mathbb{P}_0}\big(\langle \widetilde{\Delta}_\theta, \widetilde{g}_\theta(O)\rangle_{\mathcal{H}_\theta}\big).$$

A subscript or denotes an empirical quantity computed with oracle pseudo-features.

Measurable versions. We use the common response kernel from Proposition B.1:

$$\mu_\theta(a, x) = \int_{\mathcal{Y}} \varphi_\theta(y) Q(\,\mathrm{d}y \mid a, x). \tag{10}$$

The measurable null-set convention in Section B.1 applies whenever only first moments are assumed. Under the uniform diagonal bound below, the integral exists for every (θ, a, x) and satisfies $\|\mu_\theta(a, x)\|_{\mathcal{H}_\theta} \le \sqrt{\kappa}$; no parameter-dependent exceptional sets are needed.

Write the estimators from Section 3 in terms of the fitted observed-data kernel $\widetilde{h}_\theta(o, o') := \langle \widetilde{g}_\theta(o), \widetilde{g}_\theta(o')\rangle_{\mathcal{H}_\theta}$. Set $U_{\theta,j} := m^{-1} \sum_{i \in \mathcal{I}_1} \widetilde{h}_\theta(O_i, O_j)$. Then

$$\widehat{D}^{\mathrm{cross}}_\theta = \frac{1}{m} \sum_{j \in \mathcal{I}_2} U_{\theta,j}, \tag{11}$$

and

$$\hat{s}^2_\theta = \frac{1}{m} \sum_{j \in \mathcal{I}_2} U^2_{\theta,j} - \big(\widehat{D}^{\mathrm{cross}}_\theta\big)^2. \tag{12}$$

Assumption B.4 (Selection regularity)**.** *The set $\Theta \subset \mathbb{R}^{d_\Theta}$ is compact and lies in a Euclidean ball of radius R_Θ. There are fixed $\kappa, W, A, B_\mu, L < \infty$ such that, almost surely over the nuisance-fitting sample, $k_\theta(y, y) \le \kappa$ for every (θ, y) and, outside one common $(\mathbb{P}_X \otimes \rho)$-null set, whenever $|r(a, x)| > 0$, $\pi_0(a \mid x), \hat{\pi}_0(a \mid x) > 0$ and, for every $\theta \in \Theta$,*

$$\frac{|r(a, x)|}{\pi_0(a \mid x)} \le W, \quad \left|\frac{\pi_0(a \mid x)}{\hat{\pi}_0(a \mid x)} - 1\right| \le A, \quad \|\hat{\mu}_\theta(a, x)\|_{\mathcal{H}_\theta} \le B_\mu. \tag{13}$$

For every θ, $\hat{\mu}_\theta$ and $\widetilde{g}_\theta$ are strongly measurable. Moreover, $\widetilde{h}_\theta(o, o') := \langle \widetilde{g}_\theta(o), \widetilde{g}_\theta(o')\rangle_{\mathcal{H}_\theta}$ is jointly measurable and, almost surely over the nuisance-fitting sample, on a common $\mathbb{P}_0^2$-full set,

$$|\widetilde{h}_\theta(o, o') - \widetilde{h}_{\theta'}(o, o')| \le L\|\theta - \theta'\|_2. \tag{14}$$

Conditional on $\mathcal{F}_{\mathrm{fit}}$, define the nuisance errors

$$\begin{aligned} \epsilon^2_{\pi,n} &:= \mathbb{E}_0\left[\int_{\mathcal{A}} |r(a, X)| \left\{\frac{\pi_0(a \mid X)}{\hat{\pi}_0(a \mid X)} - 1\right\}^2 \rho(\,\mathrm{d}a)\right], \\ \epsilon^2_{\mu,n} &:= \sup_{\theta \in \Theta} \mathbb{E}_0\left[\int_{\mathcal{A}} |r(a, X)| \|\hat{\mu}_\theta(a, X) - \mu_\theta(a, X)\|^2_{\mathcal{H}_\theta} \rho(\,\mathrm{d}a)\right]. \end{aligned} \tag{15}$$

We assume that $\epsilon^2_{\mu,n}$ is measurable with respect to $\mathcal{F}_{\mathrm{fit}}$.

For $\lambda > 0$, the empirical and fitted-population criteria are continuous on compact Θ under the measurability and Lipschitz assumptions on $\widetilde{h}_\theta$. We can take the supremum of their absolute difference over a countable dense subset of Θ, so it is measurable. Measurability of the error relative to $J_{\lambda,\theta}$ follows from continuity of $f \mapsto \sup_{\theta\in\Theta} |f(\theta) - J_{\lambda,\theta}|$ on $C(\Theta)$. This map is Lipschitz in the supremum norm because $J_{\lambda,\theta}$ is deterministic and bounded.

For a failure probability $\eta \in (0,1)$, define the kernel-class sampling term

$$c_m(\eta) := \frac{\sqrt{d_\Theta \log(1 + 2R_\Theta\sqrt{m}) + \log(4/\eta)} + 1}{\sqrt{m}}. \tag{16}$$

B.7.1 Kernel-Class Regularity

We give sufficient conditions for regularity of k_θ, the oracle kernel h_θ and the fitted kernel $\widetilde{h}_\theta$.

For $o, o' \in \mathcal{O}$, define the oracle induced observed-data kernel

$$h_\theta(o, o') := \langle g_\theta^{\pi,\pi'}(o), g_\theta^{\pi,\pi'}(o')\rangle_{\mathcal{H}_\theta}. \tag{17}$$

Assumption B.5 (Oracle induced-kernel class)**.** *The parameter set Θ is compact and contained in the Euclidean ball of radius R_Θ. There exist a measurable set $\mathcal{O}_0 \subseteq \mathcal{O}$ with $\mathbb{P}_0(\mathcal{O}_0) = 1$ and constants $B_g, L_h < \infty$ such that, simultaneously for every $o, o' \in \mathcal{O}_0$ and $\theta, \theta' \in \Theta$,*

$$\|g_\theta^{\pi,\pi'}(o)\|_{\mathcal{H}_\theta} \le B_g, \qquad |h_\theta(o,o') - h_{\theta'}(o,o')| \le L_h\|\theta - \theta'\|_2.$$

The scalar map $(\theta, o, o') \mapsto h_\theta(o, o')$ is jointly measurable on $\Theta \times \mathcal{O}_0^2$. Consequently, $|h_\theta(o,o')| \le B_g^2$ on this set; write $B_h := B_g^2$.

Remark B.6 (Sufficient conditions for neural kernels). For Gaussian kernels on neural features, a bounded outcome domain, bounded parameter norms, positive bandwidth lower bounds and uniform Lipschitz control of the feature map in its parameters suffice for raw-kernel regularity. A bounded activation alone need not control parameter sensitivity on unbounded inputs. Related bounds appear in Liu et al. (2020, Propositions 21 and 23).

Lemma B.7 (Regularity of the deep-kernel ensemble)**.** *Suppose there exist finite constants*

$$K_{\text{deep}},\ K_{\text{char}},\ L_{\text{deep}},\ L_{\text{char}},\ L_w < \infty$$

such that, for all $\vartheta, \vartheta' \in \Xi$ and $y, y' \in \mathcal{Y}$,

$$|k_\vartheta^{\text{deep}}(y,y')| \le K_{\text{deep}}, \qquad |k_\vartheta^{\text{char}}(y,y')| \le K_{\text{char}},$$

$$|k_\vartheta^{\text{deep}}(y,y') - k_{\vartheta'}^{\text{deep}}(y,y')| \le L_{\text{deep}}\|\vartheta - \vartheta'\|_2, \qquad |k_\vartheta^{\text{char}}(y,y') - k_{\vartheta'}^{\text{char}}(y,y')| \le L_{\text{char}}\|\vartheta - \vartheta'\|_2,$$

and

$$|w_\vartheta - w_{\vartheta'}| \le L_w\|\vartheta - \vartheta'\|_2.$$

Writing $L_0 := L_{\text{deep}} + L_{\text{char}} + L_w(K_{\text{deep}} + K_{\text{char}})$, for $\theta = (\vartheta_1, \dots, \vartheta_M)$ and $\theta' = (\vartheta'_1, \dots, \vartheta'_M)$,

$$|k_\theta(y,y') - k_{\theta'}(y,y')| \le \frac{L_0}{\sqrt{M}}\|\theta - \theta'\|_2.$$

Moreover,

$$\sup_{\theta,y} k_\theta(y,y) \le \max\{K_{\text{deep}}, K_{\text{char}}\}.$$

Proof. For one member, add and subtract $(1 - w_\vartheta)k_{\vartheta'}^{\text{deep}} + w_\vartheta k_{\vartheta'}^{\text{char}}$. Since $0 \le w_\vartheta \le 1$,

$$\begin{aligned}
&\left|(1-w_\vartheta)k_\vartheta^{\text{deep}} + w_\vartheta k_\vartheta^{\text{char}} - (1-w_{\vartheta'})k_{\vartheta'}^{\text{deep}} - w_{\vartheta'}k_{\vartheta'}^{\text{char}}\right| \\
&\le (1-w_\vartheta)|k_\vartheta^{\text{deep}} - k_{\vartheta'}^{\text{deep}}| + w_\vartheta|k_\vartheta^{\text{char}} - k_{\vartheta'}^{\text{char}}| + |w_\vartheta - w_{\vartheta'}||k_{\vartheta'}^{\text{char}} - k_{\vartheta'}^{\text{deep}}| \\
&\le L_0\|\vartheta - \vartheta'\|_2.
\end{aligned}$$

Average this bound over members and apply Cauchy-Schwarz:

$$|k_\theta - k_{\theta'}| \le \frac{L_0}{M}\sum_{b=1}^{M}\|\vartheta_b - \vartheta'_b\|_2 \le \frac{L_0}{\sqrt{M}}\|\theta - \theta'\|_2.$$

The diagonal bound follows first from convexity within each member and then from averaging the members. □

Proposition B.8 (Primitive conditions for the induced kernel)**.** *Suppose Θ is compact and contained in a Euclidean ball of finite radius, Assumption 3.1 holds, the map $(\theta, y, y') \mapsto k_\theta(y, y')$ is jointly measurable, and, for finite κ, L_k,*

$$\sup_{\theta\in\Theta}\sup_{y\in\mathcal{Y}} k_\theta(y,y) \le \kappa, \qquad \sup_{y,y'\in\mathcal{Y}} |k_\theta(y,y') - k_{\theta'}(y,y')| \le L_k\|\theta-\theta'\|_2.$$

Let

$$W := \operatorname*{ess\,sup}_{\mathbb{P}_X\otimes\rho} \frac{|r(a,x)|}{\pi_0(a \mid x)}.$$

If $W < \infty$, then there exists a measurable set $\mathcal{O}_0 \subseteq \mathcal{O}$ with $\mathbb{P}_0(\mathcal{O}_0) = 1$ such that

$$\sup_{\theta\in\Theta} \|g_\theta^{\pi,\pi'}(o)\|_{\mathcal{H}_\theta} \le B_g := 2(W+1)\sqrt{\kappa} \qquad \text{for every } o \in \mathcal{O}_0,$$

and the induced kernels in (17) *satisfy Assumption B.5 with*

$$B_h = B_g^2 = 4(W+1)^2\kappa, \qquad L_h = 4(W+1)^2 L_k.$$

Proof. Choose a common full-measure set on which the contrast-overlap bound defining W holds and whose x-sections have full ρ-measure, and let $\mathcal{O}_0$ be the corresponding full-measure subset of observed triples. Positive definiteness and the diagonal bound imply

$$|k_\theta(y,y')| \le \sqrt{k_\theta(y,y)k_\theta(y',y')} \le \kappa.$$

Hence $\|\varphi_\theta(y)\|_{\mathcal{H}_\theta} \le \sqrt{\kappa}$ and, by Jensen's inequality, $\|\mu_\theta(a,x)\|_{\mathcal{H}_\theta} \le \sqrt{\kappa}$. Since $\int |r(a,x)|\rho(\,\mathrm{d}a) \le 2$,

$$\begin{aligned}\|g_\theta^{\pi,\pi'}(o)\|_{\mathcal{H}_\theta} &\le \frac{|r(a,x)|}{\pi_0(a \mid x)}\|\varphi_\theta(y) - \mu_\theta(a,x)\|_{\mathcal{H}_\theta} + \int |r(\widetilde{a},x)|\|\mu_\theta(\widetilde{a},x)\|_{\mathcal{H}_\theta}\rho(\,\mathrm{d}\widetilde{a}) \\ &\le 2W\sqrt{\kappa} + 2\sqrt{\kappa} = 2(W+1)\sqrt{\kappa}.\end{aligned}$$

Cauchy-Schwarz gives $|h_\theta(o,o')| \le B_g^2 = B_h$.

To compare h_θ across θ, define for $o = (x,a,y)$ the finite signed measure

$$\begin{aligned}M_o(B) := \frac{r(a,x)}{\pi_0(a \mid x)}\{\mathbf{1}(y \in B) - Q(B \mid a,x)\} \\ + \int_{\mathcal{A}} r(\widetilde{a},x)Q(B \mid \widetilde{a},x)\rho(\,\mathrm{d}\widetilde{a}).\end{aligned}$$

The common kernel Q makes M_o independent of θ. Its total variation satisfies $\|M_o\|_{\mathrm{TV}} \le 2(W+1)$. Moreover,

$$g_\theta^{\pi,\pi'}(o) = \int_{\mathcal{Y}} \varphi_\theta(u)M_o(\,\mathrm{d}u),$$

so

$$h_\theta(o,o') = \int_{\mathcal{Y}}\int_{\mathcal{Y}} k_\theta(u,v)M_o(\,\mathrm{d}u)M_{o'}(\,\mathrm{d}v).$$

Measurability of Q and joint measurability of the kernel family imply that $(\theta,o,o') \mapsto h_\theta(o,o')$ is jointly measurable on $\Theta \times \mathcal{O}_0^2$. Therefore

$$\begin{aligned}|h_\theta(o,o') - h_{\theta'}(o,o')| &\le \|M_o\|_{\mathrm{TV}}\|M_{o'}\|_{\mathrm{TV}} \sup_{u,v} |k_\theta(u,v) - k_{\theta'}(u,v)| \\ &\le 4(W+1)^2 L_k\|\theta-\theta'\|_2.\end{aligned}$$

□

Proposition B.9 (Primitive conditions for the fitted CME family)**.** *Condition on $\mathcal{F}_{\mathrm{fit}}$. Retain the compact parameter set and overlap and fitted propensity-ratio bounds of Assumption B.4, with a measurable $\hat{\pi}_0$ independent of θ. Suppose k_θ is jointly measurable, has diagonal bound κ and uniform parameter Lipschitz constant L_k as in Proposition B.8, and each φ_θ is strongly measurable. Let*

$$\hat{\mu}_\theta(a,x) = \sum_{i=1}^n \beta_i(a,x)\varphi_\theta(Y_i), \qquad \sup_{a,x}\sum_{i=1}^n |\beta_i(a,x)| \le B_\beta,$$

where the fitting outcomes Y_i and measurable coefficient functions β_i are fixed conditional on $\mathcal{F}_{\mathrm{fit}}$ and the coefficients do not depend on θ. Suppose these bounds hold almost surely with fixed constants. Then Assumption B.4 holds with

$$B_\mu = B_\beta\sqrt{\kappa}, \qquad L = C_\beta^2 L_k, \qquad C_\beta := W(1+A)(1+B_\beta) + 2B_\beta,$$

and $|\widetilde{h}_\theta(o,o')| \le \kappa C_\beta^2$ on a common $\mathbb{P}_0^2$-full set.

Proof. The CME bound follows by the triangle inequality. Define the finite signed measure

$$\widehat{M}_o = \frac{r(a,x)}{\hat{\pi}_0(a \mid x)}\left(\delta_y - \sum_i \beta_i(a,x)\delta_{Y_i}\right) + \sum_i \left\{\int r(t,x)\beta_i(t,x)\rho(\,\mathrm{d}t)\right\}\delta_{Y_i},$$

with the ratio set to zero where $r = 0$. Since $|r|/\hat{\pi}_0 \le W(1+A)$ and $\int |r(t,x)|\rho(\,\mathrm{d}t) \le 2$, its total variation is at most C_β on a common full set. It is independent of θ, and

$$\widetilde{g}_\theta(o) = \int \varphi_\theta(u)\widehat{M}_o(\,\mathrm{d}u), \qquad \widetilde{h}_\theta(o,o') = \iint k_\theta(u,v)\widehat{M}_o(\,\mathrm{d}u)\widehat{M}_{o'}(\,\mathrm{d}v).$$

These representations give the envelope and Lipschitz bounds by $|k_\theta| \le \kappa$ and the assumed bound on $k_\theta - k_{\theta'}$. The finite expansions and measurable coefficient integrals give strong measurability of the fitted functions and joint measurability of the scalar kernel. □

For kernel ridge regression with a fixed input kernel and ridge, the coefficients $\beta_i(a,x)$ do not depend on θ, although the fitted embedding still changes through φ_θ. With the $K + n\lambda_\mu I$ normalisation, a bounded input kernel and λ_μ bounded away from zero give the uniform coefficient bound required above.

B.7.2 Oracle Concentration and Adaptation

Let $(\lambda_m)_{m\ge1}$ be a deterministic sequence of positive selection regularisers. Let two independent oracle selection samples each contain m observations. Construct the oracle versions of (11), (12), and (3) by replacing the fitted pseudo-features with $g_\theta^{\pi,\pi'}$, and denote the resulting quantities by

$$\widehat{D}^{\mathrm{cross}}_{\theta,\mathrm{or}}, \qquad \widehat{s}^2_{\theta,\mathrm{or}}, \qquad \widehat{J}^{\mathrm{or}}_{\lambda_m,\theta}.$$

For $\delta \in (0,1)$, define

$$r_{D,m}(\delta) := C\left\{B_h c_m(\delta) + \frac{L_h}{\sqrt{m}}\right\}, \tag{18}$$

$$r_{s,m}(\delta) := C\left\{B_h^2 c_m(\delta) + \frac{B_h L_h}{\sqrt{m}} + \frac{B_h^2}{m}\right\}, \tag{19}$$

where $C > 0$ is a universal constant. For $s > 0$, let

$$\Theta_s := \{\theta \in \Theta : s_\theta^2 \ge s^2\}.$$

Membership in Θ_s depends on unknown population variances.

Lemma B.10 (Uniform oracle concentration)**.** *Under Assumptions 3.1 and 3.2 and Assumption B.5, with probability at least $1-\delta$,*

$$\sup_{\theta\in\Theta} |\widehat{D}^{\mathrm{cross}}_{\theta,\mathrm{or}} - D_\theta| \le r_{D,m}(\delta), \qquad \sup_{\theta\in\Theta} |\widehat{s}^2_{\theta,\mathrm{or}} - s^2_\theta| \le r_{s,m}(\delta).$$

Proof. We adapt the covering and bounded-difference arguments of Liu et al. (2020, Propositions 15–16 and Lemmas 17–19) to the cross-statistic. Earlier uniform bounds for kernel mixtures appear in Gretton et al. (2012b, Supplement, Appendix A). Let $\tau_m = m^{-1/2}$. Since Θ lies in a Euclidean ball of radius R_Θ, it admits a τ_m-net $\mathcal{N}_{\tau_m}$ with

$$|\mathcal{N}_{\tau_m}| \le \{1 + 2R_\Theta\sqrt{m}\}^{d_\Theta}.$$

For fixed θ,

$$\widehat{D}_\theta = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} h_\theta(O_i^{(1)}, O_j^{(2)}), \qquad \mathbb{E}_0\widehat{D}_\theta = D_\theta.$$

Replacing one observation in either half changes $\widehat{D}_\theta$ by at most $2B_h/m$. McDiarmid's inequality therefore gives

$$\mathbb{P}\{|\widehat{D}_\theta - D_\theta| > t\} \le 2\exp\left(-\frac{mt^2}{4B_h^2}\right).$$

A union bound over $\mathcal{N}_{\tau_m}$ controls the maximum over the net by a universal constant times

$$B_h c_m(\delta).$$

Since both $\widehat{D}_\theta$ and D_θ are L_h-Lipschitz, extending from the net adds at most $2L_h/\sqrt{m}$. This proves the numerator bound.

For the variance estimator, write

$$\bar{G}_1 := \frac{1}{m}\sum_{i=1}^{m} g_\theta(O_i^{(1)}), \qquad \bar{G}_2 := \frac{1}{m}\sum_{j=1}^{m} g_\theta(O_j^{(2)}),$$

and

$$Z_j := \langle \bar{G}_1, g_\theta(O_j^{(2)}) - \bar{G}_2\rangle_{\mathcal{H}_\theta}.$$

Then $\widehat{s}_\theta^2 = m^{-1}\sum_j Z_j^2$ and $|Z_j| \le 2B_h$. Replacing one observation in either half changes $\widehat{s}_\theta^2$ by at most CB_h^2/m for a universal constant C. Hence

$$\mathbb{P}\{|\widehat{s}_\theta^2 - \mathbb{E}_0\widehat{s}_\theta^2| > t\} \le 2\exp\left(-c\frac{mt^2}{B_h^4}\right)$$

for a universal $c > 0$.

It remains to control the bias. Let $H = g_\theta(O)$, $\Delta = \mathbb{E}_0 H$, $\xi = H - \Delta$, and let $\Sigma := \mathbb{E}_0[\xi \otimes \xi]$. Conditional on $\bar{G}_1$, $\widehat{s}_\theta^2$ is the sample variance, with denominator m, of $\langle \bar{G}_1, H_j\rangle$. Therefore

$$\mathbb{E}_0[\widehat{s}_\theta^2 \mid \bar{G}_1] = \frac{m-1}{m}\langle \bar{G}_1, \Sigma\bar{G}_1\rangle.$$

Since $\bar{G}_1 = \Delta + m^{-1}\sum_i \xi_i$,

$$\mathbb{E}_0\langle \bar{G}_1, \Sigma\bar{G}_1\rangle = s_\theta^2 + \frac{1}{m}\mathbb{E}_0\langle \xi, \Sigma\xi\rangle.$$

Thus

$$\mathbb{E}_0\widehat{s}_\theta^2 - s_\theta^2 = -\frac{s_\theta^2}{m} + \frac{m-1}{m^2}\mathbb{E}_0\langle \xi, \Sigma\xi\rangle,$$

and boundedness of h_θ yields

$$|\mathbb{E}_0\widehat{s}_\theta^2 - s_\theta^2| \le C\frac{B_h^2}{m}.$$

Finally, $\theta \mapsto \widehat{s}_\theta^2$ is $8B_hL_h$-Lipschitz and $\theta \mapsto s_\theta^2$ is $4B_hL_h$-Lipschitz. Applying the preceding concentration inequality on the net, adding the bias, and extending from the net proves the variance bound. □

Lemma B.11 (Stable ratio bound)**.** *Let $D, \widehat{D} \in \mathbb{R}$, $v, \widehat{v} \in \mathbb{R}_+$, and $s, B > 0$. Suppose*

$$0 \le D \le B, \qquad v \ge s^2, \qquad |\widehat{v} - v| \le s^2/2.$$

Then, for every $\lambda \ge 0$,

$$\left|\frac{\widehat{D}}{\sqrt{\widehat{v}+\lambda}} - \frac{D}{\sqrt{v}}\right| \le \frac{\sqrt{2}}{s}|\widehat{D} - D| + \frac{\sqrt{2}B}{s^3}\{|\widehat{v} - v| + \lambda\}. \tag{20}$$

Proof. The variance condition gives $\widehat{v} \geq s^2/2$. Hence

$$\left|\frac{\widehat{D}}{\sqrt{\widehat{v}+\lambda}} - \frac{D}{\sqrt{v}}\right| \leq \frac{|\widehat{D}-D|}{\sqrt{\widehat{v}+\lambda}} + D\left|(\widehat{v}+\lambda)^{-1/2} - v^{-1/2}\right|$$
$$\leq \frac{\sqrt{2}}{s}|\widehat{D}-D| + \frac{\sqrt{2}D}{s^3}|\widehat{v}+\lambda-v|,$$

where the second line uses the mean-value theorem on $[s^2/2, \infty)$. Since $D \leq B$, (20) follows. □

Theorem B.12 (Variance-stable oracle adaptation)**.** *Let $s > 0$ satisfy $\Theta_s \neq \varnothing$, and suppose Assumptions 3.1 and 3.2 and Assumption B.5 hold. If $r_{s,m}(\delta) \leq s^2/2$, then, with probability at least $1-\delta$,*

$$\sup_{\theta\in\Theta_s} |\widehat{J}^{\mathrm{or}}_{\lambda_m,\theta} - J_\theta| \leq \epsilon_m(\delta), \tag{21}$$

where

$$\epsilon_m(\delta) := \frac{\sqrt{2}}{s} r_{D,m}(\delta) + \frac{\sqrt{2}B_h}{s^3}\{r_{s,m}(\delta) + \lambda_m\}. \tag{22}$$

If B_h, L_h, s are fixed,

$$\frac{d_\Theta \log(1+2R_\Theta\sqrt{m})+1}{m} \longrightarrow 0, \qquad \lambda_m = O\left(\sqrt{\frac{d_\Theta \log(1+2R_\Theta\sqrt{m})+1}{m}}\right),$$

then

$$\sup_{\theta\in\Theta_s} |\widehat{J}^{\mathrm{or}}_{\lambda_m,\theta} - J_\theta| = O_{\mathbb{P}}\left(\sqrt{\frac{d_\Theta \log(1+2R_\Theta\sqrt{m})+1}{m}}\right).$$

Proof of Theorem B.12. On the event in Lemma B.10, every $\theta \in \Theta_s$ satisfies

$$\widehat{s}^2_{\theta,\mathrm{or}} \geq s^2_\theta - r_{s,m}(\delta) \geq s^2/2.$$

Moreover, $0 \leq D_\theta = \mathbb{E}_0 h_\theta(O_1, O_2) \leq B_h$. Apply Lemma B.11 pointwise with

$$D = D_\theta, \quad \widehat{D} = \widehat{D}^{\mathrm{cross}}_{\theta,\mathrm{or}}, \quad v = s^2_\theta, \quad \widehat{v} = \widehat{s}^2_{\theta,\mathrm{or}}, \quad B = B_h,$$

and take the supremum over Θ_s. This gives (21) with $\epsilon_m(\delta)$ as defined in (22). The stochastic rate follows from the definitions of $r_{D,m}$ and $r_{s,m}$. □

Theorem 1 of Gretton et al. (2012b) and Theorem 11 of Liu et al. (2020) choose a ridge of order $t^{-1/3}$ for training size t. Their ratio bounds amplify root-t estimation errors, up to logarithmic and class-complexity factors, by $\lambda_t^{-1/2}$ and balance this against $O(\lambda_t)$ bias. Both assume a positive population variance lower bound. Here, for sufficiently large samples, the uniform concentration event bounds the empirical variance away from zero on Θ_s, so Lemma B.11 avoids this amplification. The same refinement applies on the variance-bounded subclass of Liu et al. using their Propositions 15–16 and $\lambda_t = O(\sqrt{\log t/t})$ for fixed class constants and a fixed positive variance lower bound. Thus the cube-root bounds do not establish an intrinsic rate limitation.

B.7.3 Nuisance Perturbation

We now transfer the oracle concentration bounds to fitted pseudo-features.

Lemma B.13 (Score envelopes from selection regularity)**.** *Under Assumption B.4, on a common full-measure set,*

$$\sup_{\theta\in\Theta}\left\{\|g^{\pi,\pi'}_\theta(o)\|_{\mathcal{H}_\theta} \vee \|\widetilde{g}_\theta(o)\|_{\mathcal{H}_\theta}\right\} \leq M_g,$$

where

$$M_g := \max\left\{2(W+1)\sqrt{\kappa}, W(1+A)(\sqrt{\kappa}+B_\mu)+2B_\mu\right\}.$$

Consequently $|\widetilde{h}_\theta(o,o')| \leq M_g^2$.

Proof. The diagonal kernel bound and Jensen's inequality give $\|\varphi_\theta(y)\| \le \sqrt{\kappa}$ and $\|\mu_\theta(a,x)\| \le \sqrt{\kappa}$. Also $\int |r(a,x)|\rho(\,\mathrm{d}a) \le 2$. The oracle bound therefore follows from

$$\|g_\theta(o)\| \le \frac{|r(a,x)|}{\pi_0(a \mid x)}\|\varphi_\theta(y) - \mu_\theta(a,x)\| + \int |r(\widetilde{a},x)|\|\mu_\theta(\widetilde{a},x)\|\rho(\,\mathrm{d}\widetilde{a}).$$

On the support of $|r|$,

$$\frac{|r|}{\hat{\pi}_0} = \frac{|r|}{\pi_0}\frac{\pi_0}{\hat{\pi}_0} \le W(1+A),$$

which gives the fitted bound in the same way. Cauchy-Schwarz gives the final claim. □

Write $\widetilde{B}_h := M_g^2$. Let $\widetilde{r}_{D,m}(\delta)$ and $\widetilde{r}_{s,m}(\delta)$ be defined by (18)-(19), with (B_h, L_h) replaced by $(\widetilde{B}_h, L)$.

Lemma B.14 (Exact nuisance decomposition)**.** *Under Assumptions B.4 and 3.1, define*

$$\alpha(a,x) := \frac{\pi_0(a \mid x)}{\hat{\pi}_0(a \mid x)} - 1, \qquad d_\theta(a,x) := \hat{\mu}_\theta(a,x) - \mu_\theta(a,x), \qquad q(a,x) := \frac{r(a,x)}{\pi_0(a \mid x)},$$

where $q := 0$ off $\{|r| > 0\}$. Writing $\xi_\theta(O) := \varphi_\theta(Y) - \mu_\theta(A,X)$ and $t_{\theta,d}(X) := \int r(a,X)d_\theta(a,X)\rho(\,\mathrm{d}a)$, the fitted-score error has the exact decomposition

$$\widetilde{g}_\theta - g_\theta = q\alpha\xi_\theta + \{t_{\theta,d} - qd_\theta\} - q\alpha d_\theta. \tag{23}$$

The first term has conditional mean zero given (A,X), and the second has conditional mean zero given X. Hence, for $B_\theta := \widetilde{\Delta}_\theta - \Delta_\theta$,

$$B_\theta = -\mathbb{E}_0\left[\int_{\mathcal{A}} r(a,X)\alpha(a,X)d_\theta(a,X)\rho(\,\mathrm{d}a)\right], \qquad \sup_{\theta\in\Theta}\|B_\theta\| \le \epsilon_{\pi,n}\epsilon_{\mu,n}. \tag{24}$$

Moreover, the squared $L^2(\mathbb{P}_0;\mathcal{H}_\theta)$ norms of the three terms in (23) *are bounded, respectively, by*

$$4\kappa W\epsilon_{\pi,n}^2, \qquad W\epsilon_{\mu,n}^2, \qquad W(B_\mu + \sqrt{\kappa})^2\epsilon_{\pi,n}^2. \tag{25}$$

Proof. Because $r/\hat{\pi}_0 = q(1+\alpha)$ and $\hat{\mu}_\theta = \mu_\theta + d_\theta$, subtracting the oracle score from the fitted score gives

$$\begin{aligned}\widetilde{g}_\theta - g_\theta &= q(1+\alpha)(\xi_\theta - d_\theta) + \int r(\mu_\theta + d_\theta)\,\mathrm{d}\rho - q\xi_\theta - \int r\mu_\theta\,\mathrm{d}\rho \\ &= q\alpha\xi_\theta + t_{\theta,d} - qd_\theta - q\alpha d_\theta,\end{aligned}$$

where $\int f\,\mathrm{d}\rho$ abbreviates integration over $\mathcal{A}$. Since $\mathbb{E}_0[\xi_\theta \mid A,X] = 0$, the first conditional-mean statement follows. Also

$$\mathbb{E}_0[qd_\theta(A,X) \mid X] = \int r(a,X)d_\theta(a,X)\rho(\,\mathrm{d}a) = t_{\theta,d}(X),$$

which proves the second. Only the last term contributes to the unconditional mean, and conditioning on X yields the identity for B_θ. Cauchy-Schwarz with respect to the finite measure $|r(a,x)|\mathbb{P}_X(\,\mathrm{d}x)\rho(\,\mathrm{d}a)$ proves the product bound.

For the component bounds, $\|\xi_\theta\| \le 2\sqrt{\kappa}$ and $r^2/\pi_0 \le W|r|$, whence

$$\mathbb{E}_0\|q\alpha\xi_\theta\|^2 \le 4\kappa W\epsilon_{\pi,n}^2.$$

The middle term is the negative of qd_θ centred conditionally on X, so conditional variance contraction gives

$$\mathbb{E}_0\|t_{\theta,d} - qd_\theta\|^2 \le \mathbb{E}_0\|qd_\theta\|^2 \le W\epsilon_{\mu,n}^2.$$

Finally, $\|d_\theta\| \le B_\mu + \sqrt{\kappa}$ and the same change of measure gives the third bound in (25). □

Lemma B.15 (Sharp population nuisance perturbation)**.** *Under Assumptions B.4 and 3.1, for every* $\theta \in \Theta$*,*

$$\widetilde{D}_\theta - D_\theta = 2\langle \Delta_\theta, B_\theta \rangle + \|B_\theta\|^2, \qquad |\widetilde{D}_\theta - D_\theta| \le 4\sqrt{\kappa}\epsilon_{\pi,n}\epsilon_{\mu,n} + \epsilon_{\pi,n}^2\epsilon_{\mu,n}^2. \tag{26}$$

There is a finite C*, depending only on* (κ, W, A, B_μ)*, such that*

$$\sup_{\theta\in\Theta} |\widetilde{s}_\theta^2 - s_\theta^2| \le C\{\epsilon_{\pi,n} + \epsilon_{\pi,n}^2 + \epsilon_{\pi,n}\epsilon_{\mu,n} + \epsilon_{\mu,n}^2\} \le C\{\epsilon_{\pi,n} + \epsilon_{\mu,n}^2\}. \tag{27}$$

If $\hat{\pi}_0 = \pi_0$ *on the support of* $|r|$*, then* $B_\theta = 0$ *and*

$$\widetilde{D}_\theta = D_\theta, \qquad \widetilde{s}_\theta^2 = s_\theta^2 + \mathbb{E}_0\left[\langle \Delta_\theta, \widetilde{g}_\theta(O) - g_\theta(O)\rangle^2\right], \tag{28}$$

so $0 \le \widetilde{s}_\theta^2 - s_\theta^2 \le 4\kappa W \epsilon_{\mu,n}^2$.

Proof. The numerator identity follows from $\widetilde{\Delta}_\theta = \Delta_\theta + B_\theta$. Since $\|\Delta_\theta\| \le 2\sqrt{\kappa}$, (24) gives the stated bound.

To bound the variance error, we treat the propensity and outcome components separately. Suppress θ and write

$$G := g - \Delta, \qquad e := \widetilde{g} - g, \qquad \bar{e} := e - B, \qquad Z := \langle \Delta, G \rangle.$$

The centred fitted projection is

$$\widetilde{Z} := \langle \Delta + B, G + \bar{e} \rangle = Z + R, \qquad R := \langle \Delta, \bar{e} \rangle + \langle B, G + \bar{e} \rangle.$$

Both Z and R are centred. Therefore

$$\widetilde{s}^2 - s^2 = 2\mathbb{E}_0[ZR] + \mathbb{E}_0[R^2]. \tag{29}$$

Let $p := q\alpha\xi$, $u := t_d - qd$, and $v := -q\alpha d$ denote the three terms in (23). Also set $t(X) := \int r(a, X)\mu(a, X)\rho(\,\mathrm{d}a)$. Then

$$G = q\xi + t(X) - \Delta.$$

The conditional-embedding term has the exact cancellation

$$\mathbb{E}_0\left[Z\langle \Delta, u \rangle\right] = 0. \tag{30}$$

Indeed, u is (A, X)-measurable and $\mathbb{E}_0[\xi \mid A, X] = 0$, so the covariance with the residual part $q\xi$ vanishes. The remaining part $t(X) - \Delta$ is X-measurable, while $\mathbb{E}_0[u \mid X] = 0$. Thus there is no term linear in $\epsilon_{\mu,n}$.

The same calculation does not cancel the propensity contribution. Instead,

$$\mathbb{E}_0\left[Z\langle \Delta, p \rangle\right] = \mathbb{E}_0\left[q(A, X)^2\alpha(A, X)\langle \Delta, \xi(O)\rangle^2\right], \tag{31}$$

which need not be zero. Since $r^2/\pi_0 \le W|r|$, bounded features and Cauchy-Schwarz give

$$|\mathbb{E}_0[Z\langle \Delta, p \rangle]| \le C\mathbb{E}_0 \int |r(a, X)||\alpha(a, X)|\rho(\,\mathrm{d}a) \le C\epsilon_{\pi,n},$$

where $\mathbb{E}_0 \int |r| \,\mathrm{d}\rho \le 2$.

For the mixed term,

$$\mathbb{E}_0[v \mid X] = -\int r(a, X)\alpha(a, X)d(a, X)\rho(\,\mathrm{d}a).$$

The residual part $q\xi$ is again orthogonal to v, while Cauchy-Schwarz gives

$$\mathbb{E}_0\left\|\int r\alpha d \,\mathrm{d}\rho\right\| \le \epsilon_{\pi,n}\epsilon_{\mu,n}.$$

Since $t(X)$, Δ, and the scores are uniformly bounded by Lemma B.13,

$$|\mathbb{E}_0[Z\langle \Delta, v - B \rangle]| + |\mathbb{E}_0[Z\langle B, G + \bar{e} \rangle]| \le C\epsilon_{\pi,n}\epsilon_{\mu,n}.$$

It remains to control the quadratic term in (29). From (25),

$$\mathbb{E}_0\|\bar{e}\|^2 \le \mathbb{E}_0\|e\|^2 \le C(\epsilon_{\pi,n}^2 + \epsilon_{\mu,n}^2).$$

Using $\|B\| \le \epsilon_{\pi,n}\epsilon_{\mu,n}$ and the score envelope,

$$\mathbb{E}_0[R^2] \le C\{\epsilon_{\pi,n}^2 + \epsilon_{\mu,n}^2 + \epsilon_{\pi,n}^2\epsilon_{\mu,n}^2\}.$$

Substitution into (29) proves the first inequality in (27). Under Assumption B.4, both nuisance errors are bounded by constants: $|\alpha| \le A$, $\|d\| \le B_\mu + \sqrt{\kappa}$, and the $|r|$-weighted measure has mass at most two. Hence the higher-order and mixed terms are absorbed into $C(\epsilon_{\pi,n} + \epsilon_{\mu,n}^2)$, proving the second inequality.

If $\hat{\pi}_0 = \pi_0$, then $p = v = B = 0$ and $e = u$. Equation (30) yields

$$\widetilde{s}^2 = \mathbb{E}_0\{Z + \langle \Delta, u\rangle\}^2 = s^2 + \mathbb{E}_0\langle \Delta, u\rangle^2.$$

The middle component bound in (25) and $\|\Delta\|^2 \le 4\kappa$ give the final inequality. □

B.7.4 Uniform Criterion Bound and Consequences

Proof of Theorem 4.4. Conditional on $\mathcal{F}_{\mathrm{fit}}$, Lemma B.13 and the covering argument in Lemma B.10 give, with probability at least $1 - \eta$,

$$\sup_{\theta\in\Theta} |\widehat{D}_\theta^{\mathrm{cross}} - \widetilde{D}_\theta| \le \widetilde{r}_{D,m}(\eta), \qquad \sup_{\theta\in\Theta} |\widehat{s}_\theta^2 - \widetilde{s}_\theta^2| \le \widetilde{r}_{s,m}(\eta). \tag{32}$$

The definitions of these two radii and of c_m imply $\widetilde{r}_{D,m}(\eta) + \widetilde{r}_{s,m}(\eta) \le Cc_m(\eta)$ for fixed class constants. Combining (32) with (26) and (27) gives the componentwise bounds

$$\sup_{\theta\in\Theta} |\widehat{D}_\theta^{\mathrm{cross}} - D_\theta| \le C\{c_m(\eta) + \epsilon_{\pi,n}\epsilon_{\mu,n}\}, \tag{33}$$

$$\sup_{\theta\in\Theta} |\hat{s}_\theta^2 - s_\theta^2| \le C\{c_m(\eta) + \epsilon_{\pi,n} + \epsilon_{\mu,n}^2\}. \tag{34}$$

Here the bounded nuisance envelopes absorb $(\epsilon_{\pi,n}\epsilon_{\mu,n})^2$ into the product term.

For the ratio, decompose

$$\begin{aligned} |\widehat{J}_{\lambda,\theta}^{\mathrm{DR}} - J_{\lambda,\theta}| &\le \frac{|\widehat{D}_\theta^{\mathrm{cross}} - D_\theta|}{\sqrt{\lambda}} + D_\theta \left|(\widehat{s}_\theta^2 + \lambda)^{-1/2} - (s_\theta^2 + \lambda)^{-1/2}\right| \\ &\le \frac{|\widehat{D}_\theta^{\mathrm{cross}} - D_\theta|}{\sqrt{\lambda_0}} + \frac{2\kappa}{\lambda_0^{3/2}}|\widehat{s}_\theta^2 - s_\theta^2|, \end{aligned}$$

where $D_\theta \le 4\kappa$ and the derivative of $v \mapsto (v + \lambda)^{-1/2}$ has magnitude at most $(2\lambda_0^{3/2})^{-1}$. Taking the supremum and using (33) to (34) proves (8).

For cyclic rotation c, apply the conditional result given its own nuisance-fitting fold and write its errors as $\epsilon_{\pi,n}^{(c)}$ and $\epsilon_{\mu,n}^{(c)}$. Integrating each conditional failure probability and applying an unconditional union bound yields simultaneous control with probability at least $1 - 3\eta$. Averaging the three inequalities gives, for example,

$$\sup_{\theta\in\Theta} \left| \frac{1}{3}\sum_{c=1}^{3} \widehat{J}_{\lambda,\theta}^{\mathrm{DR},(c)} - J_{\lambda,\theta} \right| \le C\left[c_m(\eta) + \frac{1}{3}\sum_{c=1}^{3} \left\{ \epsilon_{\pi,n}^{(c)}\epsilon_{\mu,n}^{(c)} + \epsilon_{\pi,n}^{(c)} + (\epsilon_{\mu,n}^{(c)})^2 \right\} \right].$$

No independence among the three rotation events is required. □

Joint product-class complexity. If $\Xi \subseteq \{v : \|v\|_2 \le R_0\}$, then $d_\Theta = Md_0$ and $R_\Theta \le \sqrt{M}R_0$. Hence

$$c_m(\eta) = O\left(\sqrt{\frac{Md_0 \log(1 + 2R_0\sqrt{Mm}) + \log(1/\eta) + 1}{m}} \right).$$

For fixed class constants and nuisance rates $\epsilon_{\pi,n} = O_{\mathbb{P}}(a_{\pi,n})$ and $\epsilon_{\mu,n} = O_{\mathbb{P}}(a_{\mu,n})$, Theorem 4.4 gives criterion error

$$O_{\mathbb{P}}\left(\sqrt{\frac{Md_0 \log(1 + 2R_0\sqrt{Mm}) + 1}{m}} + a_{\pi,n} + a_{\mu,n}^2 \right).$$

If M increases, the sampling term vanishes when $Md_0 \log(1 + 2R_0\sqrt{Mm}) = o(m)$, provided the regularity constants remain bounded. The memberwise procedure in Algorithm 1 uses the $M = 1$ bound for each trained member.

Theorem B.16 (Approximation of unregularised efficacy on a variance-stable class)**.** *Let $\Theta_s := \{\theta \in \Theta : s_\theta^2 \geq s^2\}$ be nonempty for $s > 0$, and suppose Assumptions B.4 and 3.1 hold. There exist finite constants C_v, C_s, where C_v depends only on the constants in Assumption B.4 and C_s may additionally depend on s, such that the following holds. On every $\mathcal{F}_{\mathrm{fit}}$ realisation satisfying*

$$C_v\{c_m(\eta) + \epsilon_{\pi,n} + \epsilon_{\mu,n}^2\} \leq s^2/2,$$

for every $\lambda_m > 0$, with conditional probability at least $1 - \eta$,

$$\sup_{\theta \in \Theta_s} |\widehat{J}^{\mathrm{DR}}_{\lambda_m,\theta} - J_\theta| \leq C_s\{c_m(\eta) + \epsilon_{\pi,n}\epsilon_{\mu,n} + \epsilon_{\pi,n} + \epsilon_{\mu,n}^2 + \lambda_m\}. \tag{35}$$

Suppose also that deterministic sequences $a_{\pi,n}, a_{\mu,n} \to 0$ satisfy $\epsilon_{\pi,n} = O_{\mathbb{P}}(a_{\pi,n})$ and $\epsilon_{\mu,n} = O_{\mathbb{P}}(a_{\mu,n})$, and that

$$\frac{d_\Theta \log(1 + 2R_\Theta\sqrt{m}) + 1}{m} \longrightarrow 0, \qquad \lambda_m = O\left(\sqrt{\frac{d_\Theta \log(1 + 2R_\Theta\sqrt{m}) + 1}{m}}\right).$$

Then the margin holds with probability tending to one for fixed s and fixed η, and the left-hand side of (35) *is*

$$O_{\mathbb{P}}\left(\sqrt{\frac{d_\Theta \log(1 + 2R_\Theta\sqrt{m}) + 1}{m}} + a_{\pi,n} + a_{\mu,n}^2\right).$$

Proof. The proof of (34) supplies the constant C_v independently of the ridge. On its concentration event, the assumed inequality gives $\widehat{s}_\theta^2 \geq s^2/2$ throughout Θ_s. Apply Lemma B.11 with $D = D_\theta$, $\widehat{D} = \widehat{D}_\theta^{\mathrm{cross}}$, $v = s_\theta^2$, $\widehat{v} = \widehat{s}_\theta^2$, and $B = 4\kappa$. Equations (33) to (34) then give (35). The rate statement follows from (16). □

Under the preceding ridge choice, for $n \asymp m$ and fixed class constants, sufficient conditions for a uniform approximation rate of $O_{\mathbb{P}}(\sqrt{\log m/m})$ for the unregularised criterion J_θ on Θ_s are $\epsilon_{\pi,n} = O_{\mathbb{P}}(\sqrt{\log m/m})$ and $\epsilon_{\mu,n} = O_{\mathbb{P}}((\log m/m)^{1/4})$. These learner rates, including uniformity over θ for the regression, are assumed rather than proved here.

Corollary B.17 (Known logging policy)**.** *Under the conditions of Theorem 4.4, if $\hat{\pi}_0 = \pi_0$ on the support of $|r|$, then, for every $\lambda \geq \lambda_0$, with conditional probability at least $1 - \eta$,*

$$\sup_{\theta \in \Theta} |\widehat{J}^{\mathrm{DR}}_{\lambda,\theta} - J_{\lambda,\theta}| \leq C\{c_m(\eta) + \epsilon_{\mu,n}^2\}.$$

If, in addition, Θ_s is nonempty and

$$C_v\{c_m(\eta) + \epsilon_{\mu,n}^2\} \leq s^2/2,$$

then, for every $\lambda_m > 0$, with conditional probability at least $1 - \eta$,

$$\sup_{\theta \in \Theta_s} |\widehat{J}^{\mathrm{DR}}_{\lambda_m,\theta} - J_\theta| \leq C_s\{c_m(\eta) + \epsilon_{\mu,n}^2 + \lambda_m\}.$$

Proof. With known logging, $\epsilon_{\pi,n} = 0$, the population nuisance perturbation in the numerator vanishes, and (28) bounds the population variance perturbation by $4\kappa W\epsilon_{\mu,n}^2$. Combining these facts with (32) proves the fixed-ridge result. The stable-ratio argument in the proof of Theorem B.16 proves the second result. □

Remark B.18 (Approximate optimisation and held-out inference). The comparison below follows Liu et al. (2020, Corollary 13), with an optimisation error as allowed in their Corollary 12, footnote 7. Let $\lambda > 0$. On an event where $\sup_{\theta \in \Theta} |\widehat{J}^{\mathrm{DR}}_{\lambda,\theta} - J_{\lambda,\theta}| \leq \epsilon$, suppose a selected $\hat{\theta} \in \Theta$ satisfies

$$\widehat{J}^{\mathrm{DR}}_{\lambda,\hat{\theta}} \geq \sup_{\theta \in \Theta} \widehat{J}^{\mathrm{DR}}_{\lambda,\theta} - \delta_{\mathrm{opt}}, \qquad \delta_{\mathrm{opt}} \geq 0.$$

Then

$$\begin{aligned} J_{\lambda,\hat{\theta}} &\geq \widehat{J}^{\mathrm{DR}}_{\lambda,\hat{\theta}} - \epsilon \\ &\geq \sup_{\theta\in\Theta} \widehat{J}^{\mathrm{DR}}_{\lambda,\theta} - \delta_{\mathrm{opt}} - \epsilon \\ &\geq \sup_{\theta\in\Theta} J_{\lambda,\theta} - 2\epsilon - \delta_{\mathrm{opt}}. \end{aligned}$$

Theorem 4.4 bounds ϵ for $\lambda \geq \lambda_0 > 0$, but we do not bound the optimisation error δ_{opt} of the training procedure. Separate member training gives no guarantee of joint optimisation over Ξ^M.

For $\lambda_m \to 0$, Theorem B.16 gives the same comparison for J_θ on Θ_s, provided $\hat{\theta} \in \Theta_s$ and the empirical optimisation inequality holds over Θ_s. Held-out validity under Theorem 4.3 does not require global optimisation.

B.8 Null Second-Order Structure

We use the conditional notation and nuisance errors from Section B.7.

Proposition B.19 (Null second-order structure)**.** *Fix θ and assume all terms below are integrable. Consider a regular perturbation $t \mapsto \mathbb{P}_t$ through $\mathbb{P}_0$ for which the embedding contrast satisfies $\Delta_{\theta,t} = \Delta_{\theta,0} + t\dot{\Delta}_\theta + o(t)$ in $\mathcal{H}_\theta$. Under H_0,*

$$D_{\theta,t} = t^2\|\dot{\Delta}_\theta\|^2_{\mathcal{H}_\theta} + o(t^2). \tag{36}$$

If the perturbation has score $\ell \in L^2_0(\mathbb{P}_0)$ and $\dot{\Delta}_\theta = \mathbb{E}_0[g_\theta(O)\ell(O)]$, then the quadratic form in (36) *is induced by*

$$h_{\theta,0}(o_1, o_2) := \langle g_\theta(o_1), g_\theta(o_2)\rangle_{\mathcal{H}_\theta}.$$

Equivalently, $D''_{\theta,0}[\ell,\ell] = 2\mathbb{E}_0^{\otimes 2}[h_{\theta,0}(O_1,O_2)\ell(O_1)\ell(O_2)]$; thus, under the conventional $t^2/2$ Taylor normalisation, the order-two coefficient is $2h_{\theta,0}$, whereas $h_{\theta,0}$ is the kernel averaged by the corresponding U-statistic. Moreover,

$$\mathbb{E}_0[h_{\theta,0}(O_1,O_2) \mid O_1] = \mathbb{E}_0[h_{\theta,0}(O_1,O_2) \mid O_2] = 0 \quad \textit{almost surely},$$

so $h_{\theta,0}$ is one-degenerate.

For the fitted cross-product, suppose Assumptions B.4 and 3.1 hold, let $\xi_{\theta,i} := \widetilde{g}_\theta(O_i) - \widetilde{\Delta}_\theta$ and let $\bar{\xi}_{\theta,\ell}$ be its mean over selection half ℓ. Under H_0, $\widetilde{\Delta}_\theta = B_\theta$ and

$$\widehat{D}^{\mathrm{cross}}_\theta = \|B_\theta\|^2 + \langle B_\theta, \bar{\xi}_{\theta,1} + \bar{\xi}_{\theta,2}\rangle + \langle \bar{\xi}_{\theta,1}, \bar{\xi}_{\theta,2}\rangle, \tag{37}$$

$$\widehat{s}^2_\theta = \frac{1}{m}\sum_{j\in\mathcal{I}_2} \left\langle B_\theta + \bar{\xi}_{\theta,1}, \xi_{\theta,j} - \bar{\xi}_{\theta,2}\right\rangle^2. \tag{38}$$

Consequently, $\mathbb{E}_0[\widehat{D}^{\mathrm{cross}}_\theta \mid \mathcal{F}_{\mathrm{fit}}] = \|B_\theta\|^2 \leq \epsilon^2_{\pi,n}\epsilon^2_{\mu,n}$ and, for this fixed θ,

$$\widehat{D}^{\mathrm{cross}}_\theta = O_{\mathbb{P}}\left(\{\epsilon_{\pi,n}\epsilon_{\mu,n} + m^{-1/2}\}^2 \middle| \mathcal{F}_{\mathrm{fit}}\right), \qquad \widehat{s}^2_\theta = O_{\mathbb{P}}\left(\{\epsilon_{\pi,n}\epsilon_{\mu,n} + m^{-1/2}\}^2 \middle| \mathcal{F}_{\mathrm{fit}}\right). \tag{39}$$

Proof. Under H_0, $\Delta_{\theta,0} = 0$. Squaring the Hilbert-space expansion gives

$$\|t\dot{\Delta}_\theta + o(t)\|^2 = t^2\|\dot{\Delta}_\theta\|^2 + o(t^2),$$

which proves (36). If $\dot{\Delta}_\theta = \mathbb{E}_0[g_\theta\ell]$, then

$$\|\dot{\Delta}_\theta\|^2 = \mathbb{E}_0^{\otimes 2}\left[h_{\theta,0}(O_1,O_2)\ell(O_1)\ell(O_2)\right].$$

Multiplying this identity by two gives the asserted second-derivative convention. Since $\mathbb{E}_0 g_\theta = \Delta_{\theta,0} = 0$,

$$\mathbb{E}_0[h_{\theta,0}(O_1,O_2) \mid O_1] = \langle g_\theta(O_1), \mathbb{E}_0 g_\theta(O_2)\rangle = 0 \quad \text{almost surely},$$

and similarly conditioning on O_2. This is the asserted one-degeneracy.

For the empirical identities, write the two fitted half-means as $B_\theta + \bar{\xi}_{\theta,1}$ and $B_\theta + \bar{\xi}_{\theta,2}$ and expand their inner product. Subtracting the second-half projected mean from each projected observation gives (38). Independence and centring of the two halves then give the conditional expectation. Finally, the Hilbert-space second-moment bound yields $\bar{\xi}_{\theta,\ell} = O_{\mathbb{P}}(m^{-1/2} \mid \mathcal{F}_{\mathrm{fit}})$, while $m^{-1}\sum_{j\in\mathcal{I}_2}\|\xi_{\theta,j} - \bar{\xi}_{\theta,2}\|^2 = O_{\mathbb{P}}(1 \mid \mathcal{F}_{\mathrm{fit}})$. Combining these facts with (24) proves (39). □

This expansion parallels Luedtke et al. (2019, Theorem 2), whose U-statistic uses all off-diagonal pairs. Here the cross-product uses independent halves, with conditional studentisation as in Kim & Ramdas (2024); Shekhar et al. (2022). With fitted nuisances, only the centred term in (37) is degenerate; the B_θ terms account for nuisance bias.

B.9 Held-Out Fixed-Alternative Scaling

Proposition 4.1 concerns a fixed kernel and true nuisance functions. We now extend its scaling limit to the selected kernel with fitted nuisances, conditional on the learning procedure. Write m_{test} for the size of each test half.

Theorem B.20 (Held-out fixed-alternative scaling)**.** *Suppose Assumptions 3.1 and 3.2 hold. Consider a joint sequence with $m_{\mathrm{test}} \to \infty$ under a fixed alternative. Assume that the final fitted score is $\mathcal{F}_{\mathrm{learn}}$-measurable and conditionally Bochner integrable, and that the two test halves consist of i.i.d. observations from $\mathbb{P}_0$, are mutually independent, and are independent of $\mathcal{F}_{\mathrm{learn}}$. For a random fitted score $f : \mathcal{O} \to \mathcal{H}_{\hat\theta}$, write*

$$\|f\|^2_{2,\mathrm{learn}} := \mathbb{E}_0[\|f(O)\|^2_{\mathcal{H}_{\hat\theta}} \mid \mathcal{F}_{\mathrm{learn}}].$$

For some fixed $s > 0$, suppose $\mathbb{P}(\hat\theta \in \Theta_s) \to 1$, define $J_{\hat\theta} := 0$ on $\{\hat\theta \notin \Theta_s\}$, and assume

$$\|\widehat{g}_{\hat\theta} - g^{\pi,\pi'}_{\hat\theta}\|_{2,\mathrm{learn}} = o_{\mathbb{P}}(1), \qquad \mathbb{E}_0[\|\widehat{g}_{\hat\theta}(O)\|^4_{\mathcal{H}_{\hat\theta}} \mid \mathcal{F}_{\mathrm{learn}}] = O_{\mathbb{P}}(1).$$

Then

$$\frac{T_{\hat\theta}}{\sqrt{m_{\mathrm{test}}}} - J_{\hat\theta} \xrightarrow{\mathbb{P}} 0. \tag{40}$$

Proof of Theorem B.20. Condition on $\mathcal{F}_{\mathrm{learn}}$ and write

$$G := \widehat{g}_{\hat\theta}, \qquad g := g^{\pi,\pi'}_{\hat\theta}, \qquad \mu_G := \mathbb{E}_0[G(O) \mid \mathcal{F}_{\mathrm{learn}}], \qquad \Delta := \Delta_{\hat\theta}.$$

The conditional fourth-moment bound and conditional laws of large numbers on the two independent test halves give

$$\widehat{D}^{\mathrm{cross}}_{\hat\theta} - \|\mu_G\|^2_{\mathcal{H}_{\hat\theta}} = o_{\mathbb{P}}(1),$$
$$\hat{s}^2_{\hat\theta} - \mathrm{Var}_{\mathbb{P}_0}\left(\langle \mu_G, G(O)\rangle_{\mathcal{H}_{\hat\theta}} \middle| \mathcal{F}_{\mathrm{learn}}\right) = o_{\mathbb{P}}(1).$$

For the second line, first condition additionally on the first test half; the second-half studentiser is a scalar sample variance. The conditional fourth-moment bound makes its law of large numbers uniform along the joint sequence, and replacing the first-half mean by μ_G costs $o_{\mathbb{P}}(1)$.

Since $\mathbb{E}_0[g(O) \mid \mathcal{F}_{\mathrm{learn}}] = \Delta$, conditional Jensen's inequality and the assumed score convergence give

$$\|\mu_G - \Delta\|_{\mathcal{H}_{\hat\theta}} \le \|G - g\|_{2,\mathrm{learn}} = o_{\mathbb{P}}(1).$$

Moreover,

$$\begin{aligned}&\left\|\langle \mu_G, G(O)\rangle_{\mathcal{H}_{\hat\theta}} - \langle \Delta, g(O)\rangle_{\mathcal{H}_{\hat\theta}}\right\|_{L^2(\mathbb{P}_0|\mathcal{F}_{\mathrm{learn}})} \\ &\quad \le \|\mu_G - \Delta\|_{\mathcal{H}_{\hat\theta}}\|G\|_{2,\mathrm{learn}} + \|\Delta\|_{\mathcal{H}_{\hat\theta}}\|G - g\|_{2,\mathrm{learn}} = o_{\mathbb{P}}(1).\end{aligned}$$

Here $\|G\|_{2,\mathrm{learn}} = O_{\mathbb{P}}(1)$ follows from the conditional fourth-moment bound, and $\|\Delta\| = O_{\mathbb{P}}(1)$ follows from the preceding display and $\|\mu_G\| \le \|G\|_{2,\mathrm{learn}}$. Thus the fitted-score numerator and variance differ from $D_{\hat\theta}$ and $s^2_{\hat\theta}$, respectively, by $o_{\mathbb{P}}(1)$. On $\{\hat\theta \in \Theta_s\}$, division is stable because $s_{\hat\theta} \ge s$. In particular, $\mathbb{P}(\hat{s}_{\hat\theta} = 0) \to 0$, and the zero-denominator convention in (2) does not affect the limit. It follows that

$$\frac{T_{\hat\theta}}{\sqrt{m_{\mathrm{test}}}} = J_{\hat\theta} + o_{\mathbb{P}}(1),$$

which proves (40). □

C EXPERIMENTAL DETAILS

We now detail our experimental setting. We first describe the training and testing procedure, then specify the construction of each experiment. Finally, we detail the neural network architectures and hyperparameters used for the optimisation.

C.1 TRAINING AND TESTING PROCEDURE

We describe the training and testing procedure in Algorithm 1. Each ensemble member is trained separately by maximising its own learning criterion, using the same learning sample $\mathcal{D}_{\mathrm{learn}}$ but a different random initialisation. We estimate the outcome embedding using kernel ridge regression and the propensity estimate using logistic regression for binary treatments and a Gaussian logging-policy model for continuous treatments. For each member, we average the learning criteria across three rotations of the learning folds and optimise ϑ_b by gradient ascent. At each epoch, we update an exponential moving average (EMA) of the kernel parameters and compute its unregularised test statistic on $\mathcal{D}_{\mathrm{val}}$, using nuisance estimators fitted on $\mathcal{D}_{\mathrm{learn}}$. We retain the EMA checkpoint with the highest validation statistic for each member, then average the selected kernels as in (4). Finally, we refit the nuisance estimators on $\mathcal{D}_{\mathrm{learn}} \cup \mathcal{D}_{\mathrm{val}}$ for testing.

We keep the selected kernels and fitted nuisance estimators fixed and evaluate all methods on the same 200 fresh test datasets $\mathcal{D}_{\mathrm{test}}$ at $\alpha = 0.05$. We perform 20 independent outer repetitions, drawing new learning and validation samples and repeating kernel training, checkpoint selection and nuisance estimation. We report the mean rejection rate $\pm$ one standard error across outer repetitions. The reported sample size is $N := |\mathcal{D}_{\mathrm{learn}}| + |\mathcal{D}_{\mathrm{val}}| + |\mathcal{D}_{\mathrm{test}}|$ where each sample contains $N/3$ observations, up to integer rounding. Here, $|\mathcal{D}_{\mathrm{test}}|$ denotes the size of each test dataset.

For experiments with binary treatments, we compare two deterministic interventions $\pi(1 \mid x) = 1$ and $\pi'(0 \mid x) = 1$.

For DR-KTE, we use the Gaussian outcome kernel $k(y, y') = \exp\{-\|\phi(y) - \phi(y')\|^2/s^2\}$ where $\phi(y)$ is a numerical representation of the structured outcome. We use flattened pixel values for images, token IDs for text, node-label and parent-index encodings for graphs. We set s to the median pairwise Euclidean distance over the combined learning and validation outcomes, except on Causal-Blob, where we use the fifth percentile.

Algorithm 1 Training and testing procedure of DR-KTE-Deep

Require: Logged data $\mathcal{D}$, target interventions π, π', ensemble size M, member kernel family $\{k_\vartheta : \vartheta \in \Xi\}$, selection regularisation $\lambda > 0$, epochs $T_{\max}$, EMA decay γ

Split $\mathcal{D}$ into disjoint samples $\mathcal{D}_{\text{learn}}$, $\mathcal{D}_{\text{val}}$, and $\mathcal{D}_{\text{test}}$
Split $\mathcal{D}_{\text{learn}}$ into three equal folds $\mathcal{D}^{(1)}, \mathcal{D}^{(2)}, \mathcal{D}^{(3)}$
for $b = 1, \dots, M$ **do**
    Initialise $\vartheta_b \in \Xi$ independently across members and set $\widetilde{\vartheta}_b \leftarrow \vartheta_b$
    for $t = 1, \dots, T_{\max}$ **do**
        for $r = 1, 2, 3$ **do**
            Assign the folds cyclically to $\mathcal{D}^{(r)}_{\text{fit}}, \mathcal{D}^{(r)}_{\text{sel},1}, \mathcal{D}^{(r)}_{\text{sel},2}$
            Fit $\hat{\pi}^{(r)}_0$ and $\hat{\mu}^{(r)}_{\vartheta_b}$ on $\mathcal{D}^{(r)}_{\text{fit}}$
            Compute pseudo-features on $\mathcal{D}^{(r)}_{\text{sel},1}$ and $\mathcal{D}^{(r)}_{\text{sel},2}$
            Compute $\widehat{J}^{\text{DR},(r)}_{\lambda,\vartheta_b}$ using (3)
        Update ϑ_b by gradient ascent on $\frac{1}{3}\sum_{r=1}^{3} \widehat{J}^{\text{DR},(r)}_{\lambda,\vartheta_b}$
        Update the EMA parameters $\widetilde{\vartheta}_b$
        Fit nuisances on $\mathcal{D}_{\text{learn}}$ and evaluate $T_{\widetilde{\vartheta}_b}$ on $\mathcal{D}_{\text{val}}$
    Set $\hat{\vartheta}_b$ to the EMA checkpoint with the largest validation statistic on $\mathcal{D}_{\text{val}}$
Set $\hat{\theta} = (\hat{\vartheta}_1, \dots, \hat{\vartheta}_M)$ and form $k_{\hat{\theta}}$ using (4)
Refit the nuisance estimators on the combined learning and validation samples $\mathcal{D}_{\text{learn}} \cup \mathcal{D}_{\text{val}}$
Split $\mathcal{D}_{\text{test}}$ into two halves $\mathcal{I}_1, \mathcal{I}_2$ of size m_{test}
Compute the test pseudo-features and form $T_{\hat{\theta}}$ using (2)
return $T_{\hat{\theta}}$ and the asymptotic p-value $1 - \Phi(T_{\hat{\theta}})$

C.2 Image Outcomes

We detail the construction of experiments on image outcomes from Section 5. We extend the OCTMNIST experiments of Zenati & Gretton (2026). We apply semi-synthetic perturbations to baseline images sampled from MedMNIST v2 (Yang et al., 2023), released under CC BY-NC 4.0 for DermaMNIST and CC BY 4.0 for the remaining datasets.

For each unit, we sample a synthetic covariate

$$X_i \in \mathbb{R}^{d_X}, \quad X_i \sim \mathcal{N}(0, I_{d_X}), \tag{41}$$

and assign treatment as

$$A_i \mid X_i \sim \text{Bernoulli}\{\pi_0(1 \mid X_i)\}, \quad \pi_0(1 \mid X_i) = \text{clip}\{\sigma(\boldsymbol{\alpha}^\top X_i), 0.07, 0.93\}. \tag{42}$$

The potential outcomes are generated as

$$Y_i^a = B_i + R_i(a), \quad a \in \{0, 1\},$$

where B_i is a baseline image sampled from a healthy or normal-labelled class and $R_i(a)$ is a treatment dependent residual. We generate a potential outcome Y_i^a for each treatment a and retain the outcome under the assigned treatment, $Y_i = Y_i^{A_i}$.

To construct the residual, let

$$q_i = q_{\min} + (q_{\max} - q_{\min})\sigma(\boldsymbol{\beta}^\top X_i)$$

where $q_i \in (q_{\min}, q_{\max})$ determines the probability of a severe case. We sample

$$S_i \mid X_i \sim \text{Bernoulli}(q_i),$$

define mean-one multiplicative jitter factors

$$J_{ia} = \exp(\tau U_{ia} - \tau^2/2), \quad U_{ia} \sim \mathcal{N}(0, 1),$$

and independent pixel noise $\varepsilon_{ia} \sim \mathcal{N}(0, \sigma_\varepsilon^2 I)$. The residuals under both treatments are then

$$R_i(0) = q_i \delta J_{i0} M_i + \varepsilon_{i0}, \quad R_i(1) = S_i \delta J_{i1} M_i + \varepsilon_{i1},$$

where M_i is a mask describing one of three perturbations: a localised lesion, scattered deposits, and a morphological change. Localised perturbations add a single blob to the baseline image, as in Chest-Effusion, Chest-Mass, OCT-DME, OCT-CNV and Pneumonia. Scattered perturbations add several small blobs to plausible locations, as in OCT-Drusen and Retina-Severe. Morphological perturbations change an existing structure; Chest-Cardiomegaly enlarges the cardiac silhouette, while Derma-Melanoma distorts and desaturates a mole. Finally, Breast-Tumor distorts an added localised lesion, combining a localised and morphological perturbation.

We use $d_X = 6$, $\boldsymbol{\alpha} = (0.90, -0.75, 0.55, -0.40, 0.30, -0.20)^\top$ and $\boldsymbol{\beta} = (0.60, -0.50, 0.40, -0.30, 0.20, -0.10)^\top$ with $q_{\min} = 0.10$, $q_{\max} = 0.40$, $\tau = 0.20$ and $\sigma_\varepsilon = 0.025$. Baseline images are sampled with replacement from the corresponding training, validation or test split.

For Table 1 under H_1, we use $\delta = 0.7$ for Breast-Tumor, 1.4 for Chest-Cardiomegaly, 1.2 for Chest-Effusion, 1.0 for Chest-Mass, 1.4 for Derma-Melanoma, 0.8 for OCT-CNV, 0.4 for OCT-DME and OCT-Drusen, and 1.0 for Pneumonia and Retina-Severe. We set $\delta = 0$ under H_0.

C.3 Text Outcomes

We now detail the generation of text outcomes in Section 5. Given some covariate X_i, we sample a treatment A_i according to Equation (42) and generate pseudo-outcomes

$$\tilde{Y}_i^0 = \boldsymbol{\beta}^\top X_i + \varepsilon_{i0}, \quad \tilde{Y}_i^1 = \boldsymbol{\beta}^\top X_i + b_i(\delta) + \varepsilon_{i1}, \quad \varepsilon_{i0}, \varepsilon_{i1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2). \tag{43}$$

where $b_i(\delta) = \delta(2Z_i - 1)$, with $Z_i \sim_{\text{i.i.d.}} \text{Bernoulli}(0.5)$. Each pseudo-outcome is then discretised into one of three tones. The tone is used as conditioning to generate the outcome text Y^a for each treatment a from a large language model. We retain the potential outcome under the assigned treatment, $Y_i = Y_i^{A_i}$.

Oncology Notes. We sample a five-dimensional covariate $X_i \sim \mathcal{N}(0, I_5)$. The covariate is rendered as a breast-cancer patient profile containing age, cancer stage, ECOG performance status, a biomarker score, and a comorbidity index. For each profile, we prompt the language model to write three post-treatment clinical notes describing respectively disease progression, stabilised disease and positive treatment response.

Movie Reviews. For each unit, we first sample a movie review from the Rotten Tomatoes dataset (Pang & Lee, 2005) and prompt the language model to rewrite the review as positive, negative and neutral. We construct the covariate $X_i \in \mathbb{R}^5$ by applying a fixed Gaussian random projection to the sampled review's embedding.

Clinical Notes. For each unit, we first sample a clinical note from the MTSamples dataset (Nair, 2024). We then append to the note two language model generated sentences describing a positive, negative or neutral clinical outcome. As with movie reviews, we construct $X_i \in \mathbb{R}^5$ from a fixed Gaussian random projection of the note's embedding.

We use $\boldsymbol{\alpha} = (0.90, -0.75, 0.55, -0.40, 0.30)^\top$, $\boldsymbol{\beta} = (0.60, -0.50, 0.40, -0.30, 0.20)^\top$, and $\sigma_\varepsilon = 0.1$. Pseudo-outcomes below -1 select the negative variant, in $[-1, 1)$ select the neutral variant, and at or above 1 select the positive text variant. For H_1, we use $\delta = 1.0$ for Oncology Notes and Movie Reviews and $\delta = 2.6$ for Clinical Notes. For H_0, we use $\delta = 0$.

We generate the text variants with Qwen3-8B, with thinking disabled, temperature 0.7 and $p = 0.9$. Outcomes in Section D.7 and text covariates are embedded using Qwen3-Embedding-0.6B.

C.4 Graph Outcomes

Finally, we detail the graph outcome constructions of Section 5. For a pre-treatment graph G_i, we extract and standardise a scalar feature c_i from G_i to obtain the covariate X_i. As before, treatment is assigned according to Equation (42) with $\alpha = 1$. Here, treatment modifies G_i, from which we construct the outcome graph Y_i.

Contact Networks. Each unit's pre-treatment graph G_i is a contact network obtained by sampling a connected subgraph from the Infectious SocioPatterns dataset (Isella et al., 2011). Each node is an individual, and each edge is weighted by the number of 20-second intervals during which two individuals were in close proximity. We use mean excess degree as a proxy for epidemic risk, motivated by its role in random-network epidemic thresholds (Newman, 2002):

$$c_i = \frac{\sum_v d_v^2}{\sum_v d_v} - 1,$$

where d_v is the number of distinct contacts of individual v. Contact networks with a larger epidemic risk are therefore more likely to receive the intervention.

We sample an infection source with probability proportional to each node's weighted degree. Under $A_i = 0$, the contact network is unchanged. Under $A_i = 1$, we remove 20% of the contact edges, prioritising edges with more neighbours, greater contact weight, and greater weighted degree. We then sample potential transmission edges, with probability increasing with contact weight, and construct a transmission tree from each network. The two potential outcomes use the same infection source and sampled potential transmission edges. Under H_0, the transmission trees are generated from the unchanged contact network for both potential outcomes.

LLM Code Repairs. Each unit's pre-treatment graph G_i is a buggy Python function from the BugsInPy dataset (Widyasari et al., 2020), represented as its abstract syntax tree (AST).

Under $A_i = 0$, we retain the original AST, $Y_i^0 = G_i$. Under $A_i = 1$, the outcome Y_i^1 is the AST of the corresponding LLM-generated repair attempt from Ehsani et al. (2025). We choose c_i as the log-size

$$c_i = \log(1 + |\mathcal{V}_i|),$$

where $|\mathcal{V}_i|$ is the number of nodes in the original AST. Longer functions are therefore more likely to receive an LLM-generated repair.

Under H_0, we set both potential outcomes to the original AST, $Y_i^0 = Y_i^1 = G_i$, while retaining the same covariates and treatment-assignment mechanism. Under H_1, the treated potential outcome is the corresponding repair candidate.

Each outcome tree is encoded as a padded sequence of node labels and parent indices. This representation is injective for the labelled tree, although constructing the tree may discard information from the original program or contact network. The WL subtree kernel is not generally characteristic, and therefore does not provide the same omnibus guarantees as the Gaussian kernel tests.

C.5 Architectures and Hyperparameters

We describe the neural network architectures and hyperparameters used for DR-KTE-Deep. For scalar outcomes, Causal-Blob, and text outcomes when represented as an embedding, ζ_θ is a four-layer fully connected network with softplus activations (Figure 5). Text outcomes, when represented as a sequence of tokens, use a two-layer transformer (Figure 7). For image outcomes, we use a four-layer convolutional network with LeakyReLU activations (Figure 6). Finally for graph outcomes, we use a three-layer message-passing graph neural network (GNN) (Figure 8).

We use an Adam optimiser (Kingma & Ba, 2015), with separate learning rates for the neural features and the remaining parameters. Hyperparameters are summarised in Tables 4 and 5.

DR-KTE-Opt shares the same optimiser, regularisation parameters and number of epochs, but only optimises the lengthscale of its kernel, using the same learning rate as the non-neural parameters of DR-KTE-Deep.

Hyperparameter	Value
Training variance regularisation λ	10^{-4}
Validation and test variance regularisation	0
CME regularisation λ_μ	10^{-3}
EMA decay γ	0.95
Minimum mixture weight $w_{\min}$	10^{-9}
Learning-rate schedule	Cosine decay to 1% of initial rate
Gradient norm clipping	1.0

Table 4: Shared hyperparameters for our experiments.

Experiment	Epochs	Ensemble Size M	Feature LR	Other LR
Image outcomes	500	3	2×10^{-2}	10^{-3}
Text outcomes (token sequences)	1000	3	10^{-3}	10^{-4}
Graph outcomes	1000	3	2×10^{-2}	10^{-3}
Text outcomes (embedded)	1000	3	10^{-5}	10^{-4}
Scalar outcomes	1000	5	2×10^{-2}	10^{-3}
Causal-Blob	1000	5	2×10^{-2}	10^{-2}

Table 5: Experiment-specific hyperparameters.

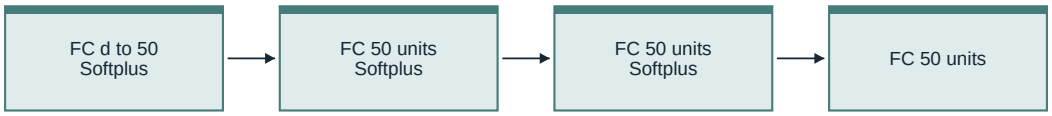

Figure 5: Four-layer fully connected feature extractor for scalar outcomes, Causal-Blob, and text outcomes under an embedded representation.

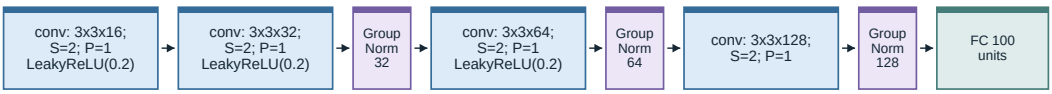

Figure 6: Four-layer convolutional feature extractor for image outcomes.

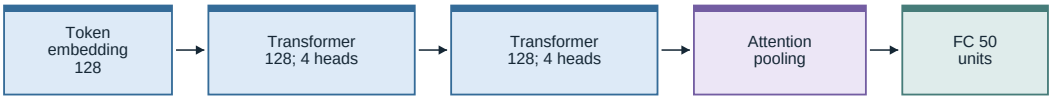

Figure 7: Two-layer transformer feature extractor for text outcomes represented as a sequence of tokens.

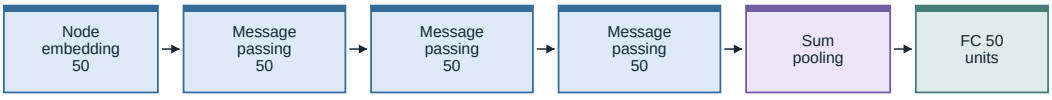

Figure 8: Three-layer message-passing feature extractor for graph outcomes.

D Additional Experiments

D.1 Additional Results on Image Outcomes

Medical-Image Outcomes. We run additional simulations on medical-image outcomes, varying both sample size and effect strength. On OCT-DME, deep-kernel power is already near one throughout the sample-size sweep, while the Gaussian baselines have low power (Figure 9). On Derma-Melanoma, increasing N substantially raises deep-kernel power, with smaller gains for the Gaussian baselines (Figure 10).

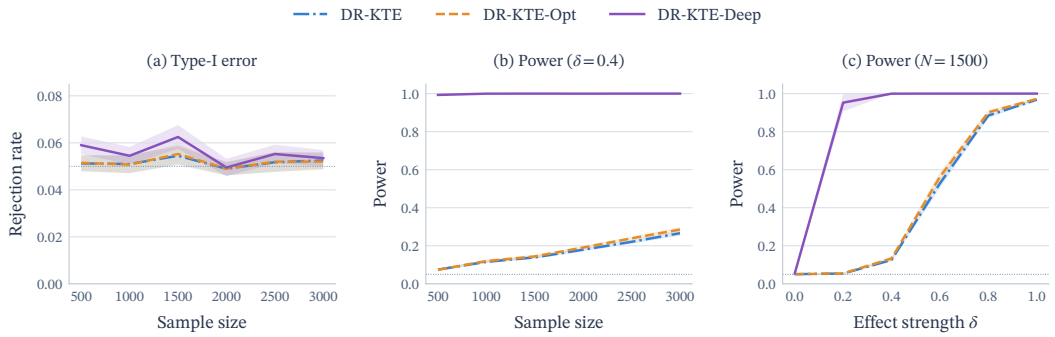

Figure 9: Results for OCT-DME. From left to right: (a) Type-I error under H_0 ($\delta = 0$), (b) average test power when varying N while keeping $\delta = 0.4$ fixed, (c) average test power when varying δ while keeping $N = 1500$ fixed. Shaded bands show $\pm$ one standard error across 20 independent outer repetitions. The dotted grey line marks $\alpha = 0.05$.

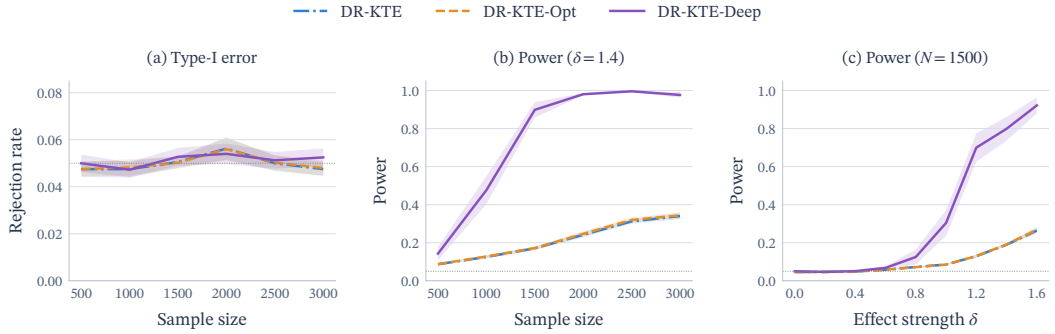

Figure 10: Results for Derma-Melanoma. From left to right: (a) Type-I error under H_0 ($\delta = 0$), (b) average test power when varying N with $\delta = 1.4$ fixed, (c) average test power when varying δ with $N = 1500$ fixed. Shaded bands show $\pm$ one standard error across 20 independent outer repetitions. The dotted grey line marks $\alpha = 0.05$.

MNIST, RetinaMNIST, BUSI. We next study a construction where treatment changes the sampling frequency of image labels of the MNIST (LeCun et al., 1998), RetinaMNIST (Yang et al., 2021; 2023; Liu et al., 2022), and BUSI (Al-Dhabyani et al., 2020) datasets.

Covariates and treatments are sampled as in Equation (41) and Equation (42). We define the scalar pseudo-outcomes

$$\tilde{Y}_i^0 = \boldsymbol{\beta}^\top X_i + \varepsilon_{i0}, \quad \tilde{Y}_i^1 = \boldsymbol{\beta}^\top X_i + b_i(\delta) + \varepsilon_{i1}, \quad \varepsilon_{i0}, \varepsilon_{i1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2), \tag{44}$$

where $d_X = 5$, $\boldsymbol{\alpha} = (0.90, -0.75, 0.55, -0.40, 0.30)^\top$, $\boldsymbol{\beta} = (0.60, -0.50, 0.40, -0.30, 0.20)^\top$ and $\sigma_\varepsilon = 0.1$.

We consider the three scenarios of Muandet et al. (2021). For Scenario I, $b_i(\delta) = 0$. For Scenario II, $b_i(\delta) = \delta$. For Scenario III, $b_i(\delta) = \delta(2Z_i - 1)$, where $Z_i \sim_{\text{i.i.d.}} \text{Bernoulli}(0.5)$. Scenario I thus

tests H_0, while Scenarios II and III test H_1. The pseudo-outcomes are converted to discrete image labels ℓ_i^a within each dataset's range. Y_i^a is then generated by sampling an image with label ℓ_i^a. We retain the outcome under the assigned treatment, $Y_i = Y_i^{A_i}$. The generative process is summarised in Figure 11, and Figure 12 illustrates the construction.

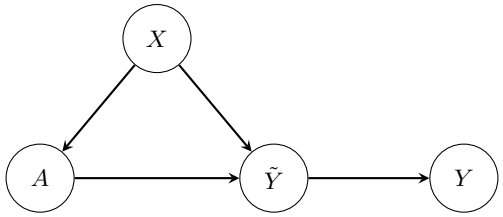

Figure 11: Data-generating process for MNIST, RetinaMNIST and BUSI. The covariate X and treatment A generate a scalar pseudo-outcome $\tilde{Y}$, which is discretised to select the label of the sampled image outcome Y.

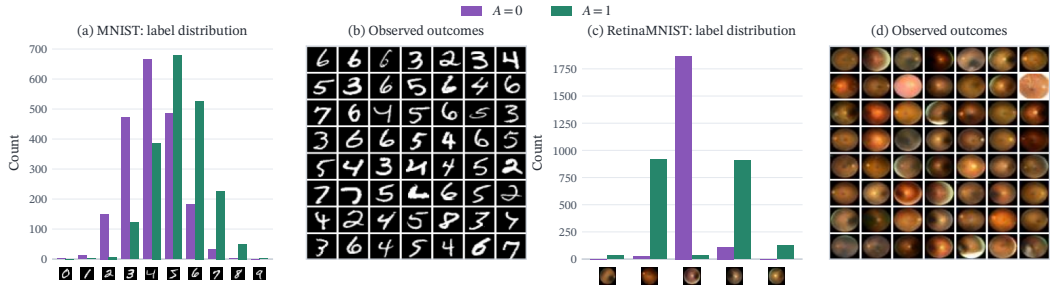

Figure 12: MNIST (Scenario II, left) and RetinaMNIST (Scenario III, right). Panels (a) and (c) show label distributions under each treatment. Panels (b) and (d) show sampled image outcomes.

We report average Type-I error for Scenario I and average power for Scenarios II and III in Table 6 for the three datasets at $N = 1500$ and $N = 3000$. Type-I error ranges from 0.045 to 0.054, close to the nominal rate of $\alpha = 0.05$. Under the alternatives, DR-KTE-Deep has higher estimated power than the baselines in every setting, with the largest improvement on RetinaMNIST in Scenario III.

Dataset	N	Scenario	DR-KTE	DR-KTE-Opt	DR-KTE-Deep
MNIST	1500	I	$0.046_{\pm 0.003}$	$0.046_{\pm 0.003}$	$0.051_{\pm 0.003}$
		II	$0.731_{\pm 0.026}$	$0.731_{\pm 0.026}$	$\mathbf{0.870_{\pm 0.024}}$
		III	$0.721_{\pm 0.024}$	$0.721_{\pm 0.024}$	$\mathbf{0.803_{\pm 0.023}}$
	3000	I	$0.049_{\pm 0.003}$	$0.049_{\pm 0.003}$	$0.049_{\pm 0.004}$
		II	$0.948_{\pm 0.008}$	$0.948_{\pm 0.008}$	$\mathbf{0.986_{\pm 0.002}}$
		III	$0.938_{\pm 0.010}$	$0.937_{\pm 0.010}$	$\mathbf{0.973_{\pm 0.006}}$
RetinaMNIST	1500	I	$0.053_{\pm 0.003}$	$0.052_{\pm 0.003}$	$0.051_{\pm 0.004}$
		II	$0.585_{\pm 0.017}$	$0.587_{\pm 0.017}$	$\mathbf{0.647_{\pm 0.022}}$
		III	$0.474_{\pm 0.020}$	$0.486_{\pm 0.022}$	$\mathbf{0.732_{\pm 0.035}}$
	3000	I	$0.053_{\pm 0.004}$	$0.054_{\pm 0.003}$	$0.050_{\pm 0.003}$
		II	$0.798_{\pm 0.013}$	$0.799_{\pm 0.013}$	$\mathbf{0.881_{\pm 0.011}}$
		III	$0.743_{\pm 0.014}$	$0.759_{\pm 0.014}$	$\mathbf{0.939_{\pm 0.009}}$
BUSI	1500	I	$0.048_{\pm 0.003}$	$0.049_{\pm 0.003}$	$0.046_{\pm 0.004}$
		III	$0.878_{\pm 0.014}$	$0.880_{\pm 0.014}$	$\mathbf{0.911_{\pm 0.013}}$
	3000	I	$0.045_{\pm 0.002}$	$0.046_{\pm 0.002}$	$0.049_{\pm 0.004}$
		III	$0.954_{\pm 0.006}$	$0.956_{\pm 0.006}$	$\mathbf{0.973_{\pm 0.005}}$

Table 6: Average rejection rates for MNIST, RetinaMNIST, and BUSI image outcomes at $\alpha = 0.05$. Scenario I reports Type-I error; Scenarios II and III report power. Effect strengths are $\delta = 1.0$ and 1.5 for MNIST II and III, $\delta = 2.0$ and 4.0 for RetinaMNIST II and III, and $\delta = 4.0$ for BUSI III. Values are means $\pm$ one standard error across 20 independent outer repetitions.

Fake-MNIST. We now consider a treatment that changes the source dataset of the sampled image, adapting the comparison of Liu et al. (2020) between real MNIST images and images generated by a generative adversarial network (GAN). We use a single digit covariate $X_i \in \{0, \dots, 9\}$, sampled uniformly, and assign treatment according to

$$A_i \mid X_i \sim \text{Bernoulli}\{\pi_0(1 \mid X_i)\}, \qquad \pi_0(1 \mid X_i) = \text{clip}\left\{\sigma\left(s_\pi \frac{X_i - 4.5}{\sqrt{8.25}}\right), 0.07, 0.93\right\},$$

where 4.5 and $\sqrt{8.25}$ are the respective mean and standard deviation of a uniformly sampled digit, and s_π is a scaling hyperparameter.

Under H_0, we sample real MNIST images of digit X_i under both interventions. Under H_1, we sample a GAN-generated image of digit X_i with probability

$$r(A_i, X_i) = A_i \frac{X_i + 1}{s_r},$$

and a real MNIST image otherwise, where s_r is a scaling hyperparameter. Outcomes under $A_i = 0$ are thus always real images, while outcomes under $A_i = 1$ can be real or GAN-generated. Larger digits are more likely to be assigned $A_i = 1$ and, under $A_i = 1$, to be replaced by a GAN-generated image.

Average Type-I error and power at $N = 1500$ and $N = 3000$ are reported in Table 7. Under H_1, DR-KTE-Deep exceeds all baselines at both sample sizes. DR-KTE-Opt slightly improves power over DR-KTE. Under H_0, rejection rates range from 0.050 to 0.054 across all methods and sample sizes, close to the nominal level.

N	Scenario	DR-KTE	DR-KTE-Opt	DR-KTE-Deep
1500	H_0	$0.051_{\pm 0.003}$	$0.052_{\pm 0.003}$	$0.050_{\pm 0.004}$
	H_1	$0.225_{\pm 0.008}$	$0.229_{\pm 0.008}$	$\mathbf{0.380_{\pm 0.050}}$
3000	H_0	$0.054_{\pm 0.004}$	$0.054_{\pm 0.004}$	$0.050_{\pm 0.003}$
	H_1	$0.502_{\pm 0.012}$	$0.514_{\pm 0.012}$	$\mathbf{0.961_{\pm 0.019}}$

Table 7: Type-I error (H_0) and power (H_1) for Fake-MNIST at $\alpha = 0.05$. Values are means $\pm$ one standard error across 20 independent outer repetitions.

D.2 Causal-Blob

We evaluate our method on a Gaussian mixture dataset adapted from the Blob dataset of Liu et al. (2020) to a causal setting with covariates. For each unit, we sample a mode index $M_i \in \{1, \dots, 9\}$ uniformly with centres $\{\mu_m\}_{m=1}^{9} = \{(j,k) : j,k \in \{0,1,2\}\}$. The covariate is then a noisy version of the mode location:

$$X_i = \mu_{M_i} + \varepsilon_X, \quad \varepsilon_X \sim \mathcal{N}(0, \sigma_X^2 I_2).$$

Treatment is sampled as in Equation (42). The potential outcomes follow the mode-centred Gaussians

$$Y_i^0 \mid M_i = m \sim \mathcal{N}(\mu_m, \Sigma_P), \qquad Y_i^1 \mid M_i = m \sim \mathcal{N}(\mu_m, \Sigma_{Q,m}),$$

where $\Sigma_P = 0.03 I_2$ is isotropic and shared across modes, while $\Sigma_{Q,m}$ has the same diagonal entries as Σ_P but a mode-dependent off-diagonal entry ρ_m. We use $\sigma_X = 0.3$ and $\boldsymbol{\alpha} = (1.2, -0.8)^\top$. The two outcome distributions thus have the same mean but differ in higher-order moments under H_1.

Table 8 shows that DR-KTE-Deep achieves the highest power on Causal-Blob at both sample sizes, with the largest gain over DR-KTE-Opt at $N = 3000$. DR-KTE-Opt also improves power over DR-KTE. Type-I error for all three tests remains close to the nominal level at both sample sizes.

	Power/Type-I error (H_1/H_0)		
N	DR-KTE	DR-KTE-Opt	DR-KTE-Deep
1500	$0.322_{\pm 0.014}/0.049_{\pm 0.004}$	$0.351_{\pm 0.016}/0.051_{\pm 0.003}$	$\mathbf{0.382}_{\pm 0.024}/0.050_{\pm 0.004}$
3000	$0.710_{\pm 0.020}/0.046_{\pm 0.003}$	$0.780_{\pm 0.022}/0.048_{\pm 0.003}$	$\mathbf{0.891}_{\pm 0.019}/0.045_{\pm 0.003}$

Table 8: Power and Type-I error for Causal-Blob at $\alpha = 0.05$. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 independent outer repetitions.

D.3 Scalar Outcomes

We also evaluate our method on scalar outcomes. We adapt the scalar-outcome scenarios of Muandet et al. (2021) with binary treatments. Covariates and treatments are sampled as in Equations (41) and (42), and we use the scalar outcomes of Equation (44) directly. Scenarios I, II and III cover the null hypothesis, a mean shift and a mean-preserving distributional shift respectively. We additionally consider Scenario IV of Zenati et al. (2025), with $b_i(\delta) = 4\delta Z_i$, where $Z_i \sim \mathrm{Bernoulli}(0.5)$. This represents a change in both the mean and variance of the outcome distributions.

Table 9 reports average Type-I error for Scenario I and average power for Scenarios II-IV. Type-I error is close to the nominal level for the kernel tests at both sample sizes. Power is close to one in Scenarios II and IV, with little improvement from learning a deep kernel. Scenario III is more challenging, although power increases substantially with sample size. DR-KTE-Opt achieves the highest estimated power in Scenario III at both sample sizes.

D.4 Estimator Comparison

We study how the choice of estimator affects test power and Type-I error, using the same learned outcome kernel for all methods. We consider the continuous-treatment construction of Zenati et al.

N	Scenario	Power/Type-I error (H_1/H_0)		
		DR-KTE	DR-KTE-Opt	DR-KTE-Deep
1500	II	$0.999_{\pm 0.000}/0.055_{\pm 0.004}$	$\mathbf{1.000}_{\pm 0.000}/0.059_{\pm 0.004}$	$0.995_{\pm 0.002}/0.055_{\pm 0.004}$
	III	$0.611_{\pm 0.028}/0.055_{\pm 0.004}$	$\mathbf{0.653}_{\pm 0.027}/0.059_{\pm 0.004}$	$0.614_{\pm 0.028}/0.055_{\pm 0.004}$
	IV	$0.980_{\pm 0.007}/0.055_{\pm 0.004}$	$\mathbf{0.985}_{\pm 0.005}/0.059_{\pm 0.004}$	$0.981_{\pm 0.004}/0.055_{\pm 0.004}$
3000	II	$\mathbf{1.000}_{\pm 0.000}/0.048_{\pm 0.003}$	$\mathbf{1.000}_{\pm 0.000}/0.046_{\pm 0.003}$	$\mathbf{1.000}_{\pm 0.000}/0.048_{\pm 0.003}$
	III	$0.903_{\pm 0.013}/0.048_{\pm 0.003}$	$\mathbf{0.927}_{\pm 0.012}/0.046_{\pm 0.003}$	$0.914_{\pm 0.013}/0.048_{\pm 0.003}$
	IV	$\mathbf{0.994}_{\pm 0.002}/0.048_{\pm 0.003}$	$\mathbf{0.994}_{\pm 0.002}/0.046_{\pm 0.003}$	$0.990_{\pm 0.002}/0.048_{\pm 0.003}$

Table 9: Power and Type-I error for scalar outcomes at $\alpha = 0.05$, with $\delta = 0.8$ and $M = 5$. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 independent outer repetitions.

(2025). We compare doubly robust estimation with inverse-propensity weighting (IPW) and a kernel policy test. For each unit, we sample the covariate as

$$X_i \sim \mathcal{N}(0, I_5),$$

and treatment is assigned as

$$A_i \mid X_i \sim \mathcal{N}(X_i^\top w, 1).$$

The outcome is then

$$Y_i = X_i^\top \beta + A_i + 0.1\varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1),$$

where $w, \beta \in \mathbb{R}^5$ are fixed coefficient vectors.

We compare two interventions π and π' and set $\pi := \pi_0$. We consider four policy-test scenarios adapted from Zenati et al. (2025) with continuous treatments. Scenario I tests the null hypothesis with $\pi'(\cdot \mid x) = \mathcal{N}(x^\top w_{\mathrm{I}}, 1)$, where $w_{\mathrm{I}} \neq w$ and $\|\beta + w_{\mathrm{I}}\| = \|\beta + w\|$, so the two interventions induce the same marginal outcome distribution. In Scenario II, $\pi'(\cdot \mid x) = \mathcal{N}(x^\top(w + 2\mathbf{1}_5), 1)$. In Scenario III, $\pi'(\cdot \mid x)$ is an equal mixture of the distributions $\mathcal{N}(x^\top(w + \mathbf{1}_5), 1)$ and $\mathcal{N}(x^\top(w - \mathbf{1}_5), 1)$. Finally, in Scenario IV, $\pi'(\cdot \mid x)$ is an equal mixture of the distributions $\mathcal{N}(x^\top(w + 2\mathbf{1}_5), 1)$ and $\mathcal{N}(x^\top w, 1)$.

We compare our doubly robust estimator, DR-KTE-Deep, to a cross-U IPW estimator, IPW-Cross, obtained by setting $\hat{\mu}_\theta(a, x) := 0$ in the plug-in estimator of Equation (1). We additionally compare our method to a kernel policy test, KPT, following Zenati et al. (2025). All methods use the same deep outcome kernel, learned with the doubly robust criterion. We divide the data into four equal samples for learning, validation, nuisance fitting and testing. For this comparison, checkpoint selection uses variance regularisation 10^{-4}, and the final cross-U statistics are unregularised. KPT uses 10,000 permutations to calibrate its test. These choices hold the learned kernel and data budget fixed while comparing the test estimators.

Table 10 shows that DR-KTE-Deep achieves higher estimated power than IPW-Cross and KPT in all three alternative scenarios, with the largest gains in Scenario III. Under the null hypothesis, its rejection rate is close to the nominal level of $\alpha = 0.05$, while IPW-Cross and KPT have higher Type-I error.

Scenario	Power/Type-I error (H_1/H_0)		
	DR-KTE-Deep	IPW-Cross	KPT
Scenario II	$\mathbf{0.961}_{\pm 0.007}/0.048_{\pm 0.003}$	$0.920_{\pm 0.012}/0.085_{\pm 0.009}$	$0.638_{\pm 0.022}/0.070_{\pm 0.013}$
Scenario III	$\mathbf{0.739}_{\pm 0.011}/0.048_{\pm 0.003}$	$0.574_{\pm 0.008}/0.085_{\pm 0.009}$	$0.291_{\pm 0.017}/0.070_{\pm 0.013}$
Scenario IV	$\mathbf{0.945}_{\pm 0.009}/0.048_{\pm 0.003}$	$0.903_{\pm 0.013}/0.085_{\pm 0.009}$	$0.607_{\pm 0.027}/0.070_{\pm 0.013}$

Table 10: Power and Type-I error at $\alpha = 0.05$, $N = 1600$ and $M = 1$ for different choices of estimator. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 independent outer repetitions.

D.5 Robustness to Outcome Model Misspecification

We analyse the contribution of fitting the conditional outcome embedding with the logging policy known. We compare the fitted embedding with $\hat{\mu}_\theta(a, x) \equiv 0$, while maintaining $\hat{\pi}_0(a \mid x) \equiv \pi_0(a \mid$

x) and a learned deep outcome kernel. We evaluate the comparison on Causal-Blob (Section D.2) and the four scalar-outcome scenarios of Section D.3.

Table 11 shows similar Type-I error for the fitted and zero outcome embeddings, with rejection rates close to the nominal level. Using the fitted outcome embedding improves power on Causal-Blob and in Scenarios II and III, with the largest gain in Scenario III. Both estimators achieve power close to one in Scenario IV.

	Power/Type-I error (H_1/H_0)	
Environment	Fitted $\hat{\mu}_\theta$	$\hat{\mu}_\theta \equiv 0$
Causal-Blob	$0.318_{\pm 0.024}/0.047_{\pm 0.006}$	$0.247_{\pm 0.017}/0.047_{\pm 0.007}$
Scenario II	$0.999_{\pm 0.001}/0.056_{\pm 0.004}$	$0.927_{\pm 0.012}/0.057_{\pm 0.003}$
Scenario III	$0.616_{\pm 0.015}/0.056_{\pm 0.004}$	$0.349_{\pm 0.013}/0.057_{\pm 0.003}$
Scenario IV	$1.000_{\pm 0.000}/0.056_{\pm 0.004}$	$0.999_{\pm 0.001}/0.057_{\pm 0.003}$

Table 11: Power and Type-I error at $\alpha = 0.05$, $N = 1500$ and $M = 1$. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 independent outer repetitions.

D.6 Training Objective Comparison

We test whether our power learning objective improves detection compared to a reconstruction objective. On OCT-Drusen and Fake-MNIST, we compare DR-KTE-Deep with a frozen randomly initialised CNN and a pixel autoencoder. All variants share the same encoder architecture, initial encoder weights, learning and validation observations, and test datasets.

DR-KTE-Deep has the highest power on both constructions under H_1 (Table 12), while Type-I error remains close to the nominal level for all methods.

Scenario		Random Weights	Pixel Autoencoder	DR-KTE-Deep
OCT-Drusen	H_1	0.057 ± 0.004	0.052 ± 0.004	$\mathbf{0.371 \pm 0.080}$
	H_0	0.054 ± 0.004	0.050 ± 0.003	0.058 ± 0.005
Fake-MNIST	H_1	0.204 ± 0.008	0.200 ± 0.012	$\mathbf{0.380 \pm 0.050}$
	H_0	0.053 ± 0.004	0.055 ± 0.004	0.050 ± 0.004

Table 12: Power under H_1 and Type-I error under H_0 at $N = 1500$. Values are means and one standard error across 20 independent outer repetitions.

D.7 Text Representation Ablation

We consider an alternative representation for the text outcomes described in Section 5, mapping each outcome to a fixed-length vector using the Qwen3-Embedding-0.6B (Zhang et al., 2025) embedding model. Since a language-model embedding can capture semantic information, a simple Gaussian kernel may be well suited to detecting differences in outcome distributions.

Table 13 shows that all three methods achieve high, similar power using this representation. Type-I error remains close to the nominal level for Oncology Notes and Clinical Notes, while all three methods are conservative on Movie Reviews.

However, this embedding is not guaranteed to be injective: distinct texts can collapse to the same embedding. A Gaussian kernel, even if characteristic on the embedding space, might therefore not detect every difference between the original text outcome distributions. Additionally, this representation can be more costly than a sequence of tokens, as it requires inference over an embedding model.

Scenario	Power/Type-I error (H_1/H_0)		
	DR-KTE	DR-KTE-Opt	DR-KTE-Deep
Oncology Notes	$\mathbf{0.828}_{\pm 0.035}/0.057_{\pm 0.010}$	$0.827_{\pm 0.035}/0.058_{\pm 0.010}$	$\mathbf{0.828}_{\pm 0.036}/0.058_{\pm 0.010}$
Movie Reviews	$0.892_{\pm 0.009}/0.029_{\pm 0.004}$	$\mathbf{0.895}_{\pm 0.008}/0.029_{\pm 0.004}$	$\mathbf{0.895}_{\pm 0.008}/0.029_{\pm 0.004}$
Clinical Notes	$\mathbf{0.894}_{\pm 0.027}/0.049_{\pm 0.007}$	$\mathbf{0.894}_{\pm 0.027}/0.049_{\pm 0.007}$	$\mathbf{0.894}_{\pm 0.027}/0.050_{\pm 0.007}$

Table 13: Power (H_1) and Type-I error (H_0) at $N = 1500$ and $\alpha = 0.05$ for text outcomes represented as text embeddings.

D.8 Effect of Ensemble Size

We evaluate the effect of memberwise averaging by varying the number of separately trained kernels included in the ensemble. We perform five independent batches, each consisting of 20 independent outer repetitions. In each repetition, we train five members and form ensembles with $M \in \{1, 2, 3, 5\}$, using the first M members. All ensemble sizes share the same observations, nuisance estimates and member kernels. We use Causal-Blob and OCT-Drusen at $N = 1500$ and $\delta = 1.0$, following the procedure in Section C.1.

For each outer repetition and ensemble size, we estimate power as the rejection rate over 200 fresh test datasets under H_1, at $\alpha = 0.05$. We then compute the mean and sample standard deviation of the 20 power estimates within each batch. For each statistic, we report its mean and one standard error across five independent batches.

On OCT-Drusen, power increases slightly with ensemble size, and its standard deviation is lower at $M = 5$ than at $M = 1$. On Causal-Blob, average power increases from 0.350 to 0.394, while variability does not decrease consistently (Table 14 and Figure 13).

While Proposition B.2 establishes variance reduction for the population discrepancy with the training data fixed, it does not imply that empirical power becomes less variable across repetitions.

Dataset	Statistic	$M = 1$	$M = 2$	$M = 3$	$M = 5$
Causal-Blob	Power	$0.350_{\pm 0.003}$	$0.373_{\pm 0.007}$	$0.383_{\pm 0.007}$	$0.394_{\pm 0.008}$
	SD	$0.150_{\pm 0.011}$	$0.157_{\pm 0.015}$	$0.150_{\pm 0.014}$	$0.151_{\pm 0.014}$
OCT-Drusen	Power	$0.993_{\pm 0.000}$	$0.995_{\pm 0.000}$	$0.995_{\pm 0.000}$	$0.996_{\pm 0.001}$
	SD	$0.011_{\pm 0.001}$	$0.009_{\pm 0.002}$	$0.009_{\pm 0.002}$	$0.007_{\pm 0.001}$

Table 14: Mean and standard deviation (SD) of empirical power at $N = 1500$ and $\delta = 1.0$. Each statistic is computed across 20 independent outer repetitions within each batch; values are means $\pm$ one standard error across five independent batches.

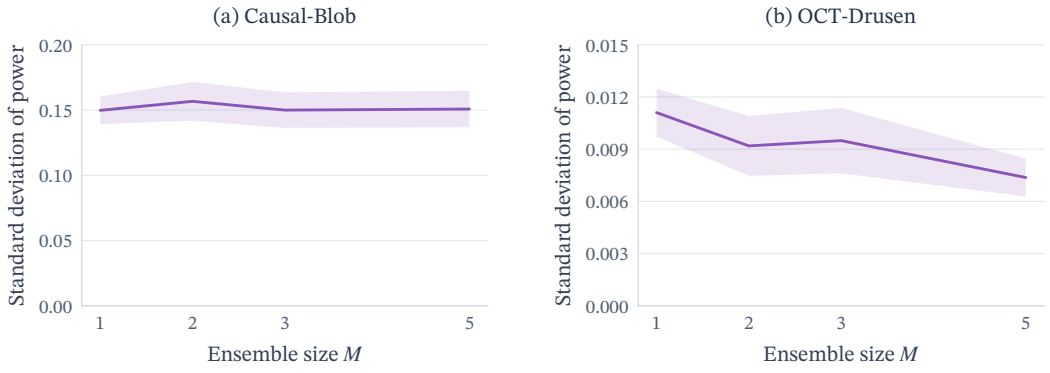

Figure 13: Average standard deviation of empirical power against ensemble size at $N = 1500$ and $\delta = 1.0$. Shaded bands show $\pm$ one standard error across five independent batches.

D.9 Weisfeiler-Lehman Outcome Kernel

We consider an alternative base kernel for the graph-outcome experiments. We replace the Gaussian kernel used as k^{char} with a Weisfeiler-Lehman (WL) subtree kernel (Shervashidze et al., 2011). We denote the fixed-kernel test by DR-KTE-WL and the learned variant by DR-KTE-WL-Deep. Since the WL kernel is not generally characteristic, these variants do not retain the same omnibus guarantee as the Gaussian-based methods.

Table 15 reports average power under H_1 and average Type-I error under H_0 for both WL-based methods. Learning the outcome kernel increases power from 0.743 to 0.964 on LLM Code Repairs. On Contact Networks, the two variants achieve similar power, with a slightly higher average for the learned kernel. Type-I error remains close to the nominal level for both constructions and methods.

	Power/Type-I error (H_1/H_0)	
Scenario	DR-KTE-WL	DR-KTE-WL-Deep
LLM Code Repairs	$0.743_{\pm 0.023}/0.059_{\pm 0.006}$	$\mathbf{0.964}_{\pm 0.008}/0.050_{\pm 0.004}$
Contact Networks	$0.971_{\pm 0.005}/0.057_{\pm 0.003}$	$\mathbf{0.972}_{\pm 0.005}/0.056_{\pm 0.004}$

Table 15: Power (H_1) and Type-I error (H_0) for graph outcomes using Weisfeiler-Lehman base kernels at $N = 1500$. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 outer repetitions. Gaussian-kernel results are given in Table 3.

D.10 Linear Outcome Kernel

We compare DR-KTE-Deep with DR-ATE, using a linear outcome kernel $k(y, y') = \phi(y)^\top \phi(y')$, where $\phi(y)$ is the numerical representation of the structured outcome. DR-ATE therefore only tests for average treatment effects.

For the medical-image experiments, Causal-Blob and the scalar outcomes Scenario III, which construct a mean-matched but distributional effect, DR-KTE-Deep has high power, while DR-ATE rejects around the nominal rate (Tables 16, 20 and 21). In the remaining experiments, where the construction introduces a mean-shift in interventional outcomes, DR-ATE often has high power, but DR-KTE-Deep always achieves higher rejection rates (Tables 17 to 19).

	Power/Type-I error (H_1/H_0)	
Scenario	DR-ATE	DR-KTE-Deep
Breast-Tumor	$0.052_{\pm 0.003}/0.047_{\pm 0.004}$	$\mathbf{0.989}_{\pm 0.002}/0.054_{\pm 0.003}$
Chest-Cardiomegaly	$0.044_{\pm 0.003}/0.043_{\pm 0.003}$	$\mathbf{0.928}_{\pm 0.046}/0.051_{\pm 0.004}$
Chest-Effusion	$0.046_{\pm 0.005}/0.049_{\pm 0.003}$	$\mathbf{0.999}_{\pm 0.000}/0.048_{\pm 0.004}$
Chest-Mass	$0.052_{\pm 0.003}/0.051_{\pm 0.004}$	$\mathbf{1.000}_{\pm 0.000}/0.056_{\pm 0.004}$
Derma-Melanoma	$0.045_{\pm 0.004}/0.048_{\pm 0.002}$	$\mathbf{0.898}_{\pm 0.040}/0.053_{\pm 0.004}$
OCT-CNV	$0.051_{\pm 0.004}/0.056_{\pm 0.003}$	$\mathbf{1.000}_{\pm 0.000}/0.053_{\pm 0.003}$
OCT-DME	$0.052_{\pm 0.003}/0.059_{\pm 0.004}$	$\mathbf{1.000}_{\pm 0.000}/0.063_{\pm 0.005}$
OCT-Drusen	$0.046_{\pm 0.003}/0.053_{\pm 0.005}$	$\mathbf{0.371}_{\pm 0.080}/0.058_{\pm 0.005}$
Pneumonia	$0.053_{\pm 0.003}/0.047_{\pm 0.002}$	$\mathbf{0.998}_{\pm 0.001}/0.053_{\pm 0.004}$
Retina-Severe	$0.056_{\pm 0.003}/0.044_{\pm 0.003}$	$\mathbf{0.959}_{\pm 0.008}/0.052_{\pm 0.003}$

Table 16: Comparison with DR-ATE for medical-image outcomes at $N = 1500$. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 outer repetitions.

Dataset	Scenario	DR-ATE	DR-KTE-Deep
MNIST	I	$0.045_{\pm 0.003}$	$0.051_{\pm 0.003}$
	II	$0.677_{\pm 0.027}$	$\mathbf{0.870}_{\pm 0.024}$
	III	$0.695_{\pm 0.023}$	$\mathbf{0.803}_{\pm 0.023}$
RetinaMNIST	I	$0.052_{\pm 0.003}$	$0.051_{\pm 0.004}$
	II	$0.434_{\pm 0.016}$	$\mathbf{0.647}_{\pm 0.022}$
	III	$0.235_{\pm 0.010}$	$\mathbf{0.732}_{\pm 0.035}$
BUSI	I	$0.054_{\pm 0.003}$	$0.046_{\pm 0.004}$
	III	$0.771_{\pm 0.016}$	$\mathbf{0.911}_{\pm 0.013}$
Fake-MNIST	H_0	$0.052_{\pm 0.003}$	$0.050_{\pm 0.004}$
	H_1	$0.114_{\pm 0.004}$	$\mathbf{0.380}_{\pm 0.050}$

Table 17: Comparison with DR-ATE for MNIST, RetinaMNIST, BUSI and Fake-MNIST at $N = 1500$. Scenario I and Fake-MNIST H_0 report Type-I error; Scenarios II and III and Fake-MNIST H_1 report power. Values are means $\pm$ one standard error across 20 outer repetitions.

	Power/Type-I error (H_1/H_0)	
Scenario	DR-ATE	DR-KTE-Deep
Oncology Notes	$0.375_{\pm 0.015}/0.051_{\pm 0.003}$	$\mathbf{0.809}_{\pm 0.023}/0.055_{\pm 0.003}$
Movie Reviews	$0.141_{\pm 0.009}/0.048_{\pm 0.003}$	$\mathbf{0.449}_{\pm 0.026}/0.053_{\pm 0.004}$
Clinical Notes	$0.053_{\pm 0.004}/0.053_{\pm 0.003}$	$\mathbf{0.939}_{\pm 0.013}/0.058_{\pm 0.004}$

Table 18: Comparison with DR-ATE for text outcomes at $N = 1500$. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 outer repetitions.

	Power/Type-I error (H_1/H_0)	
Scenario	DR-ATE	DR-KTE-Deep
LLM Code Repairs	$0.874_{\pm 0.019}/0.056_{\pm 0.004}$	$\mathbf{1.000}_{\pm 0.000}/0.051_{\pm 0.003}$
Contact Networks	$0.923_{\pm 0.009}/0.059_{\pm 0.004}$	$\mathbf{0.962}_{\pm 0.006}/0.051_{\pm 0.002}$

Table 19: Comparison with DR-ATE for graph outcomes at $N = 1500$. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 outer repetitions.

	Power/Type-I error (H_1/H_0)	
Scenario	DR-ATE	DR-KTE-Deep
Causal-Blob	$0.049_{\pm 0.003}/0.043_{\pm 0.002}$	$\mathbf{0.382}_{\pm 0.024}/0.050_{\pm 0.004}$

Table 20: Comparison with DR-ATE on Causal-Blob at $N = 1500$. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 outer repetitions.

	Power/Type-I error (H_1/H_0)	
Scenario	DR-ATE	DR-KTE-Deep
Scenario II	$0.939_{\pm 0.009}/0.097_{\pm 0.007}$	$\mathbf{0.995}_{\pm 0.002}/0.055_{\pm 0.004}$
Scenario III	$0.050_{\pm 0.004}/0.097_{\pm 0.007}$	$\mathbf{0.614}_{\pm 0.028}/0.055_{\pm 0.004}$
Scenario IV	$0.972_{\pm 0.006}/0.097_{\pm 0.007}$	$\mathbf{0.981}_{\pm 0.004}/0.055_{\pm 0.004}$

Table 21: Comparison with DR-ATE for scalar outcomes at $N = 1500$. Each alternative is paired with the shared Scenario I null. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 outer repetitions.

D.11 Scaling with Outcome Dimension

We compare DR-ATE, DR-KTE, DR-KTE-Opt and DR-KTE-Deep as the outcome dimension increases. We use Causal-Blob (Section D.2), where we vary the outcome dimension by appending independent noise coordinates to the two-dimensional outcomes. For $d > 2$, the augmented outcome is

$$Y_i^{(d)} = \begin{pmatrix} Y_i^{(2)} \\ Z_i \end{pmatrix}, \quad Z_i \sim \mathcal{N}(0, 0.17^2 I_{d-2}), Y_i^{(2)} = (1 - A_i)Y_i^0 + A_i Y_i^1,$$

where Z_i is independent of the covariates, treatment and original outcome $Y_i^{(2)}$. Since the appended coordinates Z_i have the same distribution under both treatments, the treatment effect is confined to the first two coordinates.

Figure 14 shows that DR-KTE-Opt improves power over DR-KTE, while DR-KTE-Deep achieves substantially higher power through $d = 8$. DR-ATE has little power beyond its null rejection rate throughout, as expected for this mean-preserving alternative. At $d = 16$ and $d = 32$, all four methods have little power beyond their null rejection rates. Type-I error is above the nominal level for all four methods.

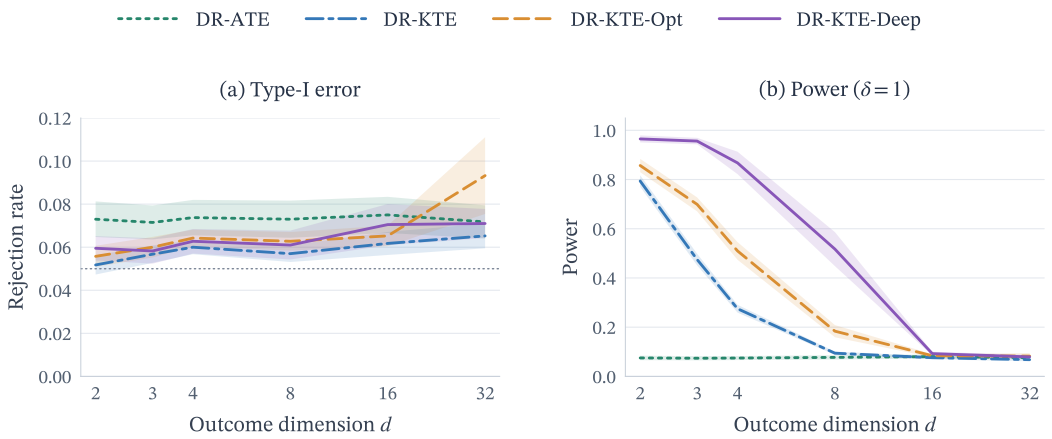

Figure 14: Causal-Blob at $N = 4000$ when varying the outcome dimension, with $\delta = 0$ under H_0 and $\delta = 1.0$ under H_1. Shaded bands show $\pm$ one standard error across 20 independent outer repetitions. The dotted grey line marks $\alpha = 0.05$.

D.12 Effect of Joint Maximisation

We compare DR-KTE-Deep to a variant that jointly maximises the power learning criterion of the averaged ensemble. In DR-KTE-Deep, each kernel member is trained and selected separately. The final ensemble can therefore be composed of members selected from different training epochs. The joint maximisation variant instead selects a single ensemble checkpoint on the validation sample. Joint maximisation therefore removes the flexibility to select each member at its own best validation checkpoint.

Joint maximisation generally yields similar or lower estimated power than DR-KTE-Deep, while improving on DR-KTE (Table 22). OCT-Drusen is the exception, with higher power under joint maximisation. Type-I error remains close to the nominal level.

	Power/Type-I error (H_1/H_0)		
Scenario	DR-KTE	DR-KTE-Deep	Joint maximisation
Breast-Tumor	$0.749_{\pm 0.018}/0.053_{\pm 0.003}$	$\mathbf{0.989}_{\pm 0.002}/0.054_{\pm 0.003}$	$\mathbf{0.989}_{\pm 0.003}/0.053_{\pm 0.003}$
Chest-Cardiomegaly	$0.156_{\pm 0.007}/0.050_{\pm 0.003}$	$\mathbf{0.928}_{\pm 0.046}/0.051_{\pm 0.004}$	$0.763_{\pm 0.077}/0.051_{\pm 0.004}$
Chest-Effusion	$0.231_{\pm 0.009}/0.053_{\pm 0.004}$	$\mathbf{0.999}_{\pm 0.000}/0.048_{\pm 0.004}$	$0.998_{\pm 0.001}/0.056_{\pm 0.004}$
Chest-Mass	$0.436_{\pm 0.018}/0.050_{\pm 0.003}$	$\mathbf{1.000}_{\pm 0.000}/0.056_{\pm 0.004}$	$0.999_{\pm 0.000}/0.050_{\pm 0.004}$
Derma-Melanoma	$0.171_{\pm 0.009}/0.051_{\pm 0.002}$	$\mathbf{0.898}_{\pm 0.040}/0.053_{\pm 0.004}$	$0.795_{\pm 0.060}/0.046_{\pm 0.003}$
OCT-CNV	$0.907_{\pm 0.014}/0.052_{\pm 0.003}$	$\mathbf{1.000}_{\pm 0.000}/0.053_{\pm 0.003}$	$\mathbf{1.000}_{\pm 0.000}/0.053_{\pm 0.003}$
OCT-DME	$0.139_{\pm 0.008}/0.055_{\pm 0.004}$	$\mathbf{1.000}_{\pm 0.000}/0.063_{\pm 0.005}$	$\mathbf{1.000}_{\pm 0.000}/0.058_{\pm 0.004}$
OCT-Drusen	$0.051_{\pm 0.003}/0.058_{\pm 0.004}$	$0.371_{\pm 0.080}/0.058_{\pm 0.005}$	$\mathbf{0.483}_{\pm 0.089}/0.053_{\pm 0.004}$
Pneumonia	$0.857_{\pm 0.021}/0.055_{\pm 0.003}$	$\mathbf{0.998}_{\pm 0.001}/0.053_{\pm 0.004}$	$0.996_{\pm 0.001}/0.051_{\pm 0.003}$
Retina-Severe	$0.285_{\pm 0.015}/0.050_{\pm 0.003}$	$\mathbf{0.959}_{\pm 0.008}/0.052_{\pm 0.003}$	$0.952_{\pm 0.009}/0.049_{\pm 0.003}$
Oncology Notes	$0.371_{\pm 0.015}/0.048_{\pm 0.003}$	$\mathbf{0.809}_{\pm 0.023}/0.055_{\pm 0.003}$	$0.770_{\pm 0.033}/0.050_{\pm 0.005}$
Movie Reviews	$0.255_{\pm 0.012}/0.050_{\pm 0.003}$	$\mathbf{0.449}_{\pm 0.026}/0.053_{\pm 0.004}$	$0.423_{\pm 0.031}/0.054_{\pm 0.005}$
Clinical Notes	$0.058_{\pm 0.004}/0.052_{\pm 0.003}$	$\mathbf{0.939}_{\pm 0.013}/0.058_{\pm 0.004}$	$0.933_{\pm 0.014}/0.062_{\pm 0.004}$
LLM Code Repairs	$0.968_{\pm 0.008}/0.048_{\pm 0.004}$	$\mathbf{1.000}_{\pm 0.000}/0.051_{\pm 0.003}$	$\mathbf{1.000}_{\pm 0.000}/0.048_{\pm 0.004}$
Contact Networks	$0.959_{\pm 0.006}/0.052_{\pm 0.003}$	$\mathbf{0.962}_{\pm 0.006}/0.051_{\pm 0.002}$	$\mathbf{0.962}_{\pm 0.006}/0.051_{\pm 0.003}$

Table 22: Effect of joint maximisation at $N = 1500$ and $\alpha = 0.05$. Each cell reports H_1/H_0, with means $\pm$ one standard error across 20 outer repetitions.