LTCC: LOCALIZED TESTING OF COORDINATE CONTRIBUTIONS TO CONDITIONAL DISTRIBUTION DISCREPANCY

Anonymous authors
Paper under double-blind review

ABSTRACT

We study whether a prespecified response coordinate contributes to the discrepancy between two multivariate conditional distributions at a fixed covariate profile. We define a localized no-contribution hypothesis that requires preservation of the target coordinate's conditional marginal distribution and its marginal-standardized cross-association with the remaining responses. Population-specific conditional probability integral transforms separate association changes from marginal differences in other coordinates. Characteristic kernels yield an equivalent direct-sum reproducing kernel Hilbert space characterization. We derive an exact von Mises expansion accounting for estimated conditional distributions and construct a cross-fitted one-step estimator. We establish a Gaussian limit under fixed alternatives and a weighted-χ^2 limit under the null for the squared-norm statistic. A consistent multiplier bootstrap yields asymptotic level α and consistency against fixed alternatives. Simulations and applications to large language models and image classification illustrate the proposed testing procedure.

1 INTRODUCTION

Comparing multivariate conditional distributions of a response vector across populations at a prespecified covariate profile is a fundamental statistical problem (Yan et al., 2022; Wu et al., 2025). Beyond determining whether two conditional response distributions differ, many applications require identifying which response coordinates contribute to the discrepancy. For example, in auditing a large language model, one may ask whether a particular subject score contributes to the difference between a reference and an audited system at a given question-difficulty level. In image classification, one may ask whether a semantic output coordinate contributes to a background-induced distributional difference at a given foreground occupancy. Because the coordinates involved in a distributional discrepancy may vary with the covariate profile, unconditional or covariate-averaged comparisons do not generally answer such localized questions. We therefore study coordinate-specific inference for conditional distributional discrepancies at a fixed covariate profile.

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^\top$ denote the response vector and $\boldsymbol{X} \in \mathcal{X}$ the covariates. For $j \in [p]$, let Y_j be the coordinate of interest and $\boldsymbol{Y}_{-j}$ the remaining coordinates. At a fixed covariate profile $\boldsymbol{X} = \boldsymbol{x}_0$, comparing the conditional marginal distributions of Y_j alone is insufficient, since Y_j may contribute through changes in its association with $\boldsymbol{Y}_{-j}$ even when its marginal distribution is unchanged. Direct comparisons of association on the original response scale can be affected by coordinatewise marginal differences even when the dependence structure is unchanged (Póczos et al., 2012). Such marginal changes should not by themselves be attributed to Y_j. We therefore compare the association of Y_j with $\boldsymbol{Y}_{-j}$ after removing coordinatewise marginal effects.

This motivates the use of population-specific conditional probability integral transforms (PITs), which transform each response coordinate by its conditional marginal CDF at $\boldsymbol{x}_0$. Under continuous conditional marginals, the transformed coordinates have uniform marginals while preserving their conditional dependence structure (Veraverbeke et al., 2011), providing a common marginal scale for comparing associations across populations. We use the conditional PIT only for the association component, while retaining the original-scale conditional marginal distribution of Y_j as a separate component. On the PIT scale, we characterize the cross-association of Y_j with $\boldsymbol{Y}_{-j}$ relative

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.

to a reference distribution that preserves the joint distribution of the remaining transformed coordinates while removing their association with Y_j. We then define Y_j as having no contribution at $\boldsymbol{x}_0$ if both its conditional marginal distribution and marginal-standardized cross-association are preserved across populations. This definition captures marginal and dependence contributions while excluding discrepancies due solely to marginal changes in other coordinates.

We develop a kernel-based procedure for testing the localized no-contribution hypothesis. Using characteristic-kernel embeddings (Sriperumbudur et al., 2010), we combine the marginal and cross-association contrasts in a direct-sum reproducing kernel Hilbert space (RKHS), yielding a nonnegative criterion that vanishes exactly under the hypothesis. Because the conditional distributions are estimated, we derive an RKHS-valued von Mises expansion and construct a cross-fitted one-step estimator (Chernozhukov et al., 2018; Williamson et al., 2023). We establish the asymptotic distributions under the null and fixed alternatives and develop a multiplier bootstrap for valid inference.

Our contributions are threefold.
1. A localized coordinate-specific hypothesis. We formulate a no-contribution hypothesis at a prespecified covariate profile requiring the target coordinate's conditional marginal distribution and marginal-standardized cross-association with the remaining responses to be preserved. Population-specific conditional PITs distinguish association changes from non-target marginal changes, while the coordinate nulls jointly characterize equality of the two conditional joint distributions.
2. Kernel characterization and debiased estimation. We establish an equivalent direct-sum RKHS characterization and derive an exact RKHS-valued von Mises expansion that accounts for estimation of the conditional response distributions. The expansion has group-specific quadratic remainders and supports a cross-fitted one-step estimator that accommodates flexible conditional-distribution estimation.
3. Asymptotic theory and bootstrap calibration. We establish a Gaussian limit under fixed alternatives and a weighted-χ^2 limit under the null for the squared-norm statistic. We prove consistency of a multiplier bootstrap, yielding asymptotic level α and consistency against fixed alternatives. Simulation studies assess test size and power, while applications to large language models and image classification illustrate coordinate-specific detection across covariate profiles.

1.1 Related Work

Kernel and conditional distribution testing. Kernel mean embeddings and maximum mean discrepancy (MMD) provide general tools for nonparametric distributional comparison (Gretton et al., 2012; Sriperumbudur et al., 2010), with extensions to conditional distributions and conditional mean embeddings (Ren et al., 2016). Recent work studies global and local conditional two-sample testing using kernel, distance, conformal, and learning-based methods (Yan et al., 2022; Wu et al., 2025; Massiani et al., 2025; Moskvichev et al., 2026). Our focus is coordinate-specific: at a fixed covariate profile, we test whether a response coordinate contributes through its marginal distribution or its marginal-standardized cross-association with the remaining responses.

Feature detection and attribution of distribution changes. Feature-shift detection identifies changed features through conditional distribution comparisons (Kulinski et al., 2020), while related approaches explain distribution shifts through causal or transport-based representations (Budhathoki et al., 2021; Kulinski & Inouye, 2023). Our formulation instead localizes the discrepancy jointly with respect to the covariate profile and response coordinate, while allowing marginal distributions of the remaining responses to differ.

Variable importance and debiased inference. Prediction-based variable-importance methods quantify the contribution of variables to prediction or risk (Williamson et al., 2023; Chen et al., 2024; 2026). This differs from our target, which concerns whether a response coordinate participates in a conditional distributional discrepancy. Methodologically, our procedure builds on one-step estimation and cross-fitting (Chernozhukov et al., 2018; Williamson et al., 2023), together with inference for degenerate kernel statistics (Leucht & Neumann, 2013; Chwialkowski et al., 2014).

2 Localized no-contribution hypothesis

Fix a covariate profile $\boldsymbol{x}_0 \in \mathcal{X}$ and a target coordinate $j \in [p]$. For population $\ell \in \{1, 2\}$, write $F^{(\ell)} = F^{(\ell)}_{\boldsymbol{Y}|\boldsymbol{X}}(\cdot \mid \boldsymbol{x}_0)$ for the conditional joint distribution and $F^{(\ell)}_r = F^{(\ell)}_{Y_r|\boldsymbol{X}}(\cdot \mid \boldsymbol{x}_0)$ for its rth co-

ordinate marginal distribution, $r \in [p]$. We also use $F_r^{(\ell)}(t)$ to denote the corresponding conditional marginal CDF. Throughout, these conditional marginal CDFs are assumed to be continuous.

To compare associations after removing coordinatewise marginal effects, define the population-specific conditional PITs $U_r^{(\ell)} = F_r^{(\ell)}(Y_r^{(\ell)})$, $r \in [p]$, and $\boldsymbol{U}^{(\ell)} = (U_1^{(\ell)}, \ldots, U_p^{(\ell)})^\top$. This standardization is important: on the raw response scale, a change in the marginal distribution of a non-target coordinate can alter a cross-association measure even when the dependence structure involving Y_j is unchanged, see the example in Appendix A. Because the transformation is population-specific, the two populations are compared after their coordinatewise conditional marginals have been placed on the same uniform scale. Let $G^{(\ell)}$ and $G_{-j}^{(\ell)}$ denote the conditional distributions of $\boldsymbol{U}^{(\ell)}$ and $\boldsymbol{U}_{-j}^{(\ell)}$, respectively, given $\boldsymbol{X}^{(\ell)} = \boldsymbol{x}_0$. By continuity, each transformed coordinate has conditional distribution $\mathcal{U} = \mathrm{Unif}(0,1)$ at $\boldsymbol{x}_0$.

For $v \in [0,1]$ and $\boldsymbol{u}_{-j} \in [0,1]^{p-1}$, we use $(v, \boldsymbol{u}_{-j})$ for the vector obtained by inserting v into coordinate j. Product distributions such as $\mathcal{U} \otimes G_{-j}^{(\ell)}$ follow the same coordinate convention. We define the marginal-standardized cross-association measure by

$$C_j^{(\ell)} = G^{(\ell)} - \mathcal{U} \otimes G_{-j}^{(\ell)}, \qquad \ell \in \{1,2\}. \tag{1}$$

The reference distribution $\mathcal{U} \otimes G_{-j}^{(\ell)}$ preserves the joint distribution of the remaining transformed coordinates while making the target coordinate independent of them. The signed measure $C_j^{(\ell)}$ therefore characterizes cross-association on the marginal-standardized scale.

We define the localized no-contribution hypothesis as

$$H_{0,j}(\boldsymbol{x}_0): \quad F_j^{(1)} = F_j^{(2)} \quad \text{and} \quad C_j^{(1)} = C_j^{(2)}, \tag{2}$$

with alternative $H_{1,j}(\boldsymbol{x}_0)$ given by its complement. This definition combines an original-scale comparison of the target marginal with a marginal-standardized comparison of its cross-association with $\boldsymbol{Y}_{-j}$. It allows the remaining coordinate marginals to differ. Moreover, the common cross-association need not be zero: the hypothesis requires preservation of association, not independence. This separation is deliberate: the population-specific PIT removes coordinatewise marginal differences from the association component, while the original-scale comparison of $F_j^{(\ell)}$ retains marginal changes in the target coordinate itself.

To represent the two components of the null hypothesis by a single nonnegative population criterion, let $k_F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $k_C : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ be bounded characteristic kernels with RKHSs $\mathcal{H}_F$ and $\mathcal{H}_C$, and set $\mathcal{H}_j = \mathcal{H}_F \oplus \mathcal{H}_C$. For $k \in \{k_F, k_C\}$, write $\mu_k(\eta) = \int k(\cdot, z)\, \mathrm{d}\eta(z)$ for the kernel mean embedding of a finite signed measure η on the corresponding input space. Define the RKHS-valued coordinate contrast and the population criterion by

$$\gamma_j(\boldsymbol{x}_0) = \Big(\mu_{k_F}\big(F_j^{(2)} - F_j^{(1)}\big),\, \mu_{k_C}\big(C_j^{(2)} - C_j^{(1)}\big)\Big)^\top \in \mathcal{H}_j \quad \text{and} \quad Q_j(\boldsymbol{x}_0) = \|\gamma_j(\boldsymbol{x}_0)\|_{\mathcal{H}_j}^2. \tag{3}$$

The direct-sum norm adds the squared RKHS norms of the marginal and cross-association contrasts, so the two components cannot cancel each other.

Proposition 2.1. *If k_F is characteristic on $\mathbb{R}$ and k_C is characteristic on $\mathbb{R}^p$, then, for each $j \in [p]$,*

$$H_{0,j}(x_0) \quad \Longleftrightarrow \quad \gamma_j(x_0) = 0 \quad \Longleftrightarrow \quad Q_j(x_0) = 0.$$

Moreover,

$$F^{(1)} = F^{(2)} \quad \Longleftrightarrow \quad Q_j(x_0) = 0 \ \textit{for all } j \in [p].$$

Proposition 2.1 reduces each localized no-contribution hypothesis to the vanishing of a population criterion and shows that these criteria vanish jointly if and only if the two conditional joint distributions coincide, providing the basis for the estimation and testing procedures developed below. Figure 1 illustrates marginal and association contributions, and a marginal change in another coordinate that is not attributed to the target. Panels A and B distinguish the two contribution channels in (2): a change in the target marginal and a change in its marginal-standardized cross-association, respectively. Panel C highlights the role of the population-specific PIT: a change confined to a non-target marginal is removed from the association comparison when the underlying dependence structure is preserved. Appendix A gives the corresponding characterization and a raw-scale counterexample motivating the PIT standardization.

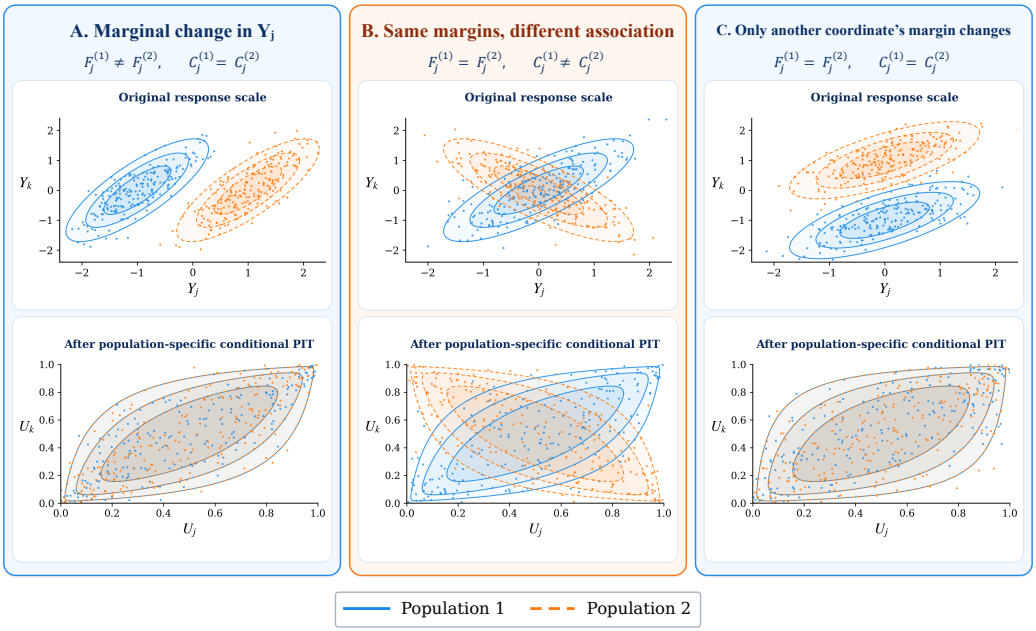

Figure 1: **Illustration of coordinate-specific contribution.** Let $\Sigma(s_j, s_k, \rho)$ be the 2×2 covariance matrix with diagonal entries s_j^2, s_k^2 and off-diagonal entries $\rho s_j s_k$. At $\boldsymbol{x}_0$, the conditional distributions of $(Y_j, Y_k)^\top$ in the two populations are **A**: $\mathcal{N}((-1,0)^\top, \Sigma(0.6, 0.8, 0.8))$ and $\mathcal{N}((1,0)^\top, \Sigma(0.6, 0.8, 0.8))$; **B**: $\mathcal{N}(0, \Sigma(0.8, 0.8, 0.8))$ and $\mathcal{N}(0, \Sigma(0.8, 0.8, -0.8))$; and **C**: $\mathcal{N}((0,-1)^\top, \Sigma(0.8, 0.6, 0.7))$ and $\mathcal{N}((0,1)^\top, \Sigma(0.8, 0.6, 0.7))$. Rows show the original and population-specific conditional-PIT scales. Hence $Q_j(\boldsymbol{x}_0) > 0$ in A–B and $Q_j(\boldsymbol{x}_0) = 0$ in C.

3 Debiased Estimation and Cross-Fitting

We observe independent samples $\{(\boldsymbol{X}_i^{(\ell)}, \boldsymbol{Y}_i^{(\ell)})\}_{i=1}^{n_\ell}$, $\ell \in \{1,2\}$, with i.i.d. observations within each population. To test $H_{0,j}(\boldsymbol{x}_0)$, we estimate the population contrast $\gamma_j(\boldsymbol{x}_0)$ and use its squared RKHS norm as the test statistic. For a pair of response distributions with continuous marginals, we write $\gamma_j(F^{(1)}, F^{(2)})$ for the contrast in (3), with each distribution supplying its own marginal CDFs for the PITs.

A plug-in estimator replaces the conditional response distributions with fitted distributions. For the association component, this affects both the kernel expectations and the PIT-transformed inputs, so estimation error can enter at first order. We address this error through an influence-function-based one-step correction and cross-fitting (Chernozhukov et al., 2018; Williamson et al., 2023).

3.1 Influence-Function Representation

We impose the following kernel conditions, satisfied by Gaussian kernels with fixed positive bandwidths, to derive the influence functions and control the von Mises remainder.

Assumption 3.1. *(i) The kernels k_F and k_C are bounded and characteristic on $\mathbb{R}$ and $\mathbb{R}^p$, respectively, with separable RKHSs $\mathcal{H}_F$ and $\mathcal{H}_C$. (ii) The map $\boldsymbol{u} \mapsto k_C(\cdot, \boldsymbol{u})$ from $\mathbb{R}^p$ into $\mathcal{H}_C$ is twice continuously differentiable, with first derivatives uniformly bounded and Lipschitz in the $\mathcal{H}_C$ norm.*

The next theorem gives the influence-function representation and von Mises expansion, with explicit componentwise formulas deferred to Appendix B. Let $\|\cdot\|_{\mathrm{TV}}$ denote the total variation norm.

Theorem 3.1. *Under Assumption 3.1, the contrast $\gamma_j(F^{(1)}, F^{(2)})$ admits influence functions $\phi_j^{(1)}$ and $\phi_j^{(2)}$ with respect to $F^{(1)}$ and $F^{(2)}$, respectively, of the form*

$$\phi_j^{(\ell)}(\boldsymbol{y}; F^{(\ell)}) = \left(\phi_{F,j}^{(\ell)}(\boldsymbol{y}; F^{(\ell)}),\, \phi_{C,j}^{(\ell)}(\boldsymbol{y}; F^{(\ell)})\right)^\top, \qquad \ell \in \{1,2\}, \tag{4}$$

where the components are given in (19)–(20). They satisfy $\int \phi_j^{(\ell)}(\boldsymbol{y}; F^{(\ell)})\,\mathrm{d}F^{(\ell)}(\boldsymbol{y}) = 0$ *for each* $\ell \in \{1,2\}$. *For any reference distributions* $\bar{F}^{(1)}$ *and* $\bar{F}^{(2)}$ *with continuous coordinate marginals,*

$$\gamma_j(F^{(1)}, F^{(2)}) = \gamma_j(\bar{F}^{(1)}, \bar{F}^{(2)}) + \sum_{\ell=1}^{2} \int \phi_j^{(\ell)}(\boldsymbol{y}; \bar{F}^{(\ell)})\,\mathrm{d}F^{(\ell)}(\boldsymbol{y}) + R_{\gamma,j}, \tag{5}$$

where

$$\|R_{\gamma,j}\|_{\mathcal{H}_j} = O\left(\|F^{(1)} - \bar{F}^{(1)}\|_{\mathrm{TV}}^2 + \|F^{(2)} - \bar{F}^{(2)}\|_{\mathrm{TV}}^2\right). \tag{6}$$

The marginal component is linear in the response distributions and contributes no remainder; the quadratic remainder arises from the association component. This motivates correcting the plug-in contrast by estimating the two influence functions using held-out observations localized around $\boldsymbol{x}_0$.

3.2 Cross-Fitted One-Step Estimator

Partition the population-ℓ sample into a fixed number $K_\ell \geq 2$ of folds $I_1^{(\ell)}, \ldots, I_{K_\ell}^{(\ell)}$. Let $\widehat{F}_{-r}^{(\ell)}(\cdot \mid \boldsymbol{x}_0)$ denote the conditional response distribution estimated using observations outside $I_r^{(\ell)}$, and abbreviate it as $\widehat{F}_{-r}^{(\ell)}$. Its coordinate marginal CDFs define the corresponding fitted PITs. Write $\widehat{\phi}_{j,-r}^{(\ell)}(\boldsymbol{y}) = \phi_j^{(\ell)}(\boldsymbol{y}; \widehat{F}_{-r}^{(\ell)})$ for the fitted influence function.

Suppose $\mathcal{X} \subseteq \mathbb{R}^d$. For a univariate localization kernel g, set $G(\boldsymbol{u}) = \prod_{q=1}^d g(u_q)$ and $G_h(\boldsymbol{u}) = h^{-d}G(\boldsymbol{u}/h)$. Given a bandwidth h_ℓ, define the localized fold average of an $\mathcal{H}_j$-valued function f by

$$\mathbb{P}_{\ell,r}^{h_\ell} f = \sum_{i \in I_r^{(\ell)}} G_{h_\ell}(\boldsymbol{X}_i^{(\ell)} - \boldsymbol{x}_0) f(\boldsymbol{Y}_i^{(\ell)}) \Big/ \sum_{i \in I_r^{(\ell)}} G_{h_\ell}(\boldsymbol{X}_i^{(\ell)} - \boldsymbol{x}_0).$$

This estimates $\mathbb{E}\{f(\boldsymbol{Y}^{(\ell)}) \mid \boldsymbol{X}^{(\ell)} = \boldsymbol{x}_0\}$ using observations in $I_r^{(\ell)}$. Motivated by (5), for each fold pair (s,t) define

$$\widehat{\gamma}_j^{(s,t)}(\boldsymbol{x}_0) = \gamma_j\left(\widehat{F}_{-s}^{(1)}, \widehat{F}_{-t}^{(2)}\right) + \mathbb{P}_{1,s}^{h_1}\widehat{\phi}_{j,-s}^{(1)} + \mathbb{P}_{2,t}^{h_2}\widehat{\phi}_{j,-t}^{(2)}. \tag{7}$$

Each correction is evaluated on observations excluded from fitting its corresponding distribution. We aggregate the corrected contrasts using fold-size weights and then take the squared norm:

$$\widehat{\gamma}_j(\boldsymbol{x}_0) = n_1^{-1}n_2^{-1}\sum_{s=1}^{K_1}\sum_{t=1}^{K_2} |I_s^{(1)}|\,|I_t^{(2)}|\widehat{\gamma}_j^{(s,t)}(\boldsymbol{x}_0), \qquad \widehat{Q}_j(\boldsymbol{x}_0) = \|\widehat{\gamma}_j(\boldsymbol{x}_0)\|_{\mathcal{H}_j}^2. \tag{8}$$

4 Testing the Localized No-Contribution Hypothesis

By Proposition 2.1, it is equivalent to test $H_{0,j}(\boldsymbol{x}_0) : Q_j(\boldsymbol{x}_0) = 0$ versus $H_{1,j}(\boldsymbol{x}_0) : Q_j(\boldsymbol{x}_0) > 0$. We therefore reject $H_{0,j}(\boldsymbol{x}_0)$ for sufficiently large values of $\widehat{Q}_j(\boldsymbol{x}_0)$. We derive the asymptotic distributions of $\widehat{Q}_j(\boldsymbol{x}_0)$ and construct a multiplier bootstrap to approximate the critical value.

4.1 Asymptotic Distributions

Write $n = n_1 + n_2$ and define $a_n^2 = (n_1 h_1^d)^{-1} + (n_2 h_2^d)^{-1}$. Let $f_X^{(\ell)}$ denote the covariate density in population ℓ. We impose the following regularity conditions.

Assumption 4.1. *The dimensions* d *and* p *are fixed. In addition, (i)* $n_1 \asymp n_2$*; (ii) the number of folds* K_1 *and* K_2 *are fixed, and each fold proportion* $|I_r^{(\ell)}|/n_\ell$ *converges to a limit in* $(0,1)$*; (iii) the fold partitions are deterministic or generated independently of the observations.*

Assumption 4.2. *The localization kernel* g *is bounded, compactly supported, and of order* $\nu \geq 2$*:* $\int g(u)\,\mathrm{d}u = 1$*,* $\int u^q g(u)\,\mathrm{d}u = 0$ *for* $q = 1, \ldots, \nu-1$*, and* $\int u^\nu g(u)\,\mathrm{d}u \neq 0$*.*

Assumption 4.3. *For each* $\ell \in \{1,2\}$*,* $f_X^{(\ell)}(\boldsymbol{x}_0) > 0$*, and* $f_X^{(\ell)}$ *is* ν *times continuously differentiable with bounded derivatives in a neighborhood of* $\boldsymbol{x}_0$*. The map* $\boldsymbol{u} \mapsto F_{\boldsymbol{Y}|\boldsymbol{X}}^{(\ell)}(\cdot \mid \boldsymbol{u})$ *is* ν *times continuously differentiable in the total variation norm, with bounded derivatives on the same neighborhood.*

Assumption 4.4. *There exists a deterministic sequence $r_n \downarrow 0$ such that*

$$\max_{\ell\in\{1,2\}} \max_{r\in[K_\ell]} \left\| \widehat{F}^{(\ell)}_{-r}(\cdot \mid \boldsymbol{x}_0) - F^{(\ell)}_{\boldsymbol{Y}|\boldsymbol{X}}(\cdot \mid \boldsymbol{x}_0) \right\|_{\mathrm{TV}} = O_p(r_n).$$

With probability tending to one, all fitted distributions have continuous coordinate marginal CDFs.

Assumption 4.5. *As $n \to \infty$, (i) $h_\ell \to 0$ and $n_\ell h_\ell^d \to \infty$ for $\ell = 1, 2$; (ii) $n_\ell^{1/2} h_\ell^{\nu+d/2} \to 0$ for $\ell = 1, 2$, and $(n_1 h_1^d)/(n_2 h_2^d) \to c \in (0, \infty)$; (iii) $a_n^{-1} r_n^2 \to 0$.*

Assumptions 4.2 and 4.3 control the local smoothing bias, which is negligible relative to a_n under the undersmoothing condition in Assumption 4.5(ii). Assumption 4.4 imposes a high-level condition on the first-stage conditional distribution estimation error, while Assumption 4.5(iii) ensures that the quadratic von Mises remainder in (6) is $o_p(a_n)$, and thus negligible.

For $\ell \in \{1, 2\}$, write $\phi_j^{(\ell)}(\boldsymbol{y}, \boldsymbol{x}_0) = \phi_j^{(\ell)}(\boldsymbol{y}; F^{(\ell)})$, and let $R_G = \int_{\mathbb{R}^d} G(\boldsymbol{u})^2 \, \mathrm{d}\boldsymbol{u}$. Let $\mathbb{G}_j(\boldsymbol{x}_0)$ be a mean-zero Gaussian random element in $\mathcal{H}_j$ with covariance operator $V_j(\boldsymbol{x}_0)$ defined, for $u, v \in \mathcal{H}_j$, by

$$\begin{aligned}\langle V_j(\boldsymbol{x}_0)u, v\rangle_{\mathcal{H}_j} =& \frac{R_G}{(1+c) f_X^{(1)}(\boldsymbol{x}_0)} \mathbb{E}\Big[\langle \phi_j^{(1)}(\boldsymbol{Y}^{(1)}, \boldsymbol{x}_0), u\rangle_{\mathcal{H}_j} \langle \phi_j^{(1)}(\boldsymbol{Y}^{(1)}, \boldsymbol{x}_0), v\rangle_{\mathcal{H}_j} \,\Big|\, \boldsymbol{X}^{(1)} = \boldsymbol{x}_0\Big] \\ &+ \frac{cR_G}{(1+c) f_X^{(2)}(\boldsymbol{x}_0)} \mathbb{E}\Big[\langle \phi_j^{(2)}(\boldsymbol{Y}^{(2)}, \boldsymbol{x}_0), u\rangle_{\mathcal{H}_j} \langle \phi_j^{(2)}(\boldsymbol{Y}^{(2)}, \boldsymbol{x}_0), v\rangle_{\mathcal{H}_j} \,\Big|\, \boldsymbol{X}^{(2)} = \boldsymbol{x}_0\Big].\end{aligned} \tag{9}$$

Theorem 4.1. *Suppose Assumption 3.1 and Assumptions 4.1–4.5 hold.*

(i) *Under any fixed alternative $Q_j(\boldsymbol{x}_0) > 0$,*

$$a_n^{-1}\{\widehat{Q}_j(\boldsymbol{x}_0) - Q_j(\boldsymbol{x}_0)\} \rightsquigarrow N\big(0, \sigma_j^2(\boldsymbol{x}_0)\big), \tag{10}$$

where $\sigma_j^2(\boldsymbol{x}_0) = 4\langle V_j(\boldsymbol{x}_0)\gamma_j(\boldsymbol{x}_0), \gamma_j(\boldsymbol{x}_0)\rangle_{\mathcal{H}_j}$.

(ii) *Under $H_{0,j}(\boldsymbol{x}_0)$,*

$$a_n^{-2}\widehat{Q}_j(\boldsymbol{x}_0) \rightsquigarrow \|\mathbb{G}_j(\boldsymbol{x}_0)\|^2_{\mathcal{H}_j}. \tag{11}$$

If $\lambda_{j,1}(\boldsymbol{x}_0), \lambda_{j,2}(\boldsymbol{x}_0), \ldots$ are the eigenvalues of $V_j(\boldsymbol{x}_0)$, counted with multiplicity, then

$$\|\mathbb{G}_j(\boldsymbol{x}_0)\|^2_{\mathcal{H}_j} \overset{d}{=} \sum_{r=1}^{\infty} \lambda_{j,r}(\boldsymbol{x}_0) Z_r^2, \tag{12}$$

where $Z_1, Z_2, \ldots$ are independent standard normal random variables.

The different normalizations arise from $\widehat{Q}_j(\boldsymbol{x}_0) - Q_j(\boldsymbol{x}_0) = 2\langle \gamma_j(\boldsymbol{x}_0), \widehat{\gamma}_j(\boldsymbol{x}_0) - \gamma_j(\boldsymbol{x}_0)\rangle_{\mathcal{H}_j} + \|\widehat{\gamma}_j(\boldsymbol{x}_0) - \gamma_j(\boldsymbol{x}_0)\|^2_{\mathcal{H}_j}$. Under a fixed alternative with $\sigma_j^2(\boldsymbol{x}_0) > 0$, the linear term yields a nondegenerate Gaussian limit. Under $H_{0,j}(\boldsymbol{x}_0)$, $\gamma_j(\boldsymbol{x}_0) = 0$, so the linear term vanishes and the quadratic term yields the weighted-χ^2 limit in (12).

4.2 Multiplier bootstrap calibration

The limiting null distribution in (12) depends on the unknown eigenvalues of $V_j(\boldsymbol{x}_0)$. We propose a multiplier bootstrap to approximate the null distribution and implement the test.

For each population $\ell \in \{1, 2\}$, let $r(i)$ denote the fold containing observation i, and write $\widehat{\phi}^{(\ell)}_{j,i}(\boldsymbol{x}_0) = \widehat{\phi}^{(\ell)}_{j,-r(i)}(\boldsymbol{Y}_i^{(\ell)})$. Define the kernel density estimator $\widehat{f}_{\ell,n}(\boldsymbol{x}_0) = n_\ell^{-1} \sum_{i=1}^{n_\ell} G_{h_\ell}(\boldsymbol{X}_i^{(\ell)} - \boldsymbol{x}_0)$. Conditionally on the data, generate standard normal multipliers $\xi_i^{(\ell)}$, independently across observations and populations, and set $\bar{\xi}_\ell = n_\ell^{-1} \sum_{i=1}^{n_\ell} \xi_i^{(\ell)}$. For $\ell = 1, 2$, define

$$\widehat{\mathbb{G}}^*_{\ell n,j}(\boldsymbol{x}_0) = \frac{h_\ell^{d/2}}{\sqrt{n_\ell}\, \widehat{f}_{\ell,n}(\boldsymbol{x}_0)} \sum_{i=1}^{n_\ell} (\xi_i^{(\ell)} - \bar{\xi}_\ell) G_{h_\ell}(\boldsymbol{X}_i^{(\ell)} - \boldsymbol{x}_0) \widehat{\phi}^{(\ell)}_{j,i}(\boldsymbol{x}_0).$$

The same multiplier is applied to both components of each fitted influence function to retain their estimated cross-covariance. Combine the two population-specific bootstrap terms as

$$\widehat{\mathbb{G}}^*_{j,n}(\boldsymbol{x}_0) = \left(\frac{n_2 h_2^d}{n_1 h_1^d + n_2 h_2^d}\right)^{1/2} \widehat{\mathbb{G}}^*_{1n,j}(\boldsymbol{x}_0) + \left(\frac{n_1 h_1^d}{n_1 h_1^d + n_2 h_2^d}\right)^{1/2} \widehat{\mathbb{G}}^*_{2n,j}(\boldsymbol{x}_0). \tag{13}$$

Let $T^*_{j,n}(\boldsymbol{x}_0) = \|\widehat{\mathbb{G}}^*_{j,n}(\boldsymbol{x}_0)\|^2_{\mathcal{H}_j}$, and let $\mathbb{P}^*$ denote probability conditional on the data. For a fixed significance level $\alpha \in (0,1)$, denote the conditional $(1-\alpha)$ quantile of $T^*_{j,n}(\boldsymbol{x}_0)$ by $c^*_{j,1-\alpha}(\boldsymbol{x}_0)$. The proposed test rejects $H_{0,j}(\boldsymbol{x}_0)$ when $a_n^{-2}\widehat{Q}_j(\boldsymbol{x}_0) > c^*_{j,1-\alpha}(\boldsymbol{x}_0)$.

Theorem 4.2. *Suppose Assumption 3.1 and Assumptions 4.1–4.5 hold. (i) If $H_{0,j}(\boldsymbol{x}_0)$ holds and $\mathrm{tr}\{V_j(\boldsymbol{x}_0)\} > 0$, then*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}^*\{T^*_{j,n}(\boldsymbol{x}_0) \le t\} - \mathbb{P}\{\|\mathbb{G}_j(\boldsymbol{x}_0)\|^2_{\mathcal{H}_j} \le t\} \right| \xrightarrow{p} 0. \tag{14}$$

Consequently, $c^*_{j,1-\alpha}(\boldsymbol{x}_0) \xrightarrow{p} c_{j,1-\alpha}(\boldsymbol{x}_0)$, where $c_{j,1-\alpha}(\boldsymbol{x}_0)$ is the $(1-\alpha)$ quantile of $\|\mathbb{G}_j(\boldsymbol{x}_0)\|^2_{\mathcal{H}_j}$, and $\mathbb{P}\{a_n^{-2}\widehat{Q}_j(\boldsymbol{x}_0) > c^*_{j,1-\alpha}(\boldsymbol{x}_0)\} \longrightarrow \alpha$. Thus, the test has asymptotic level α. (ii) Under any fixed alternative $Q_j(\boldsymbol{x}_0) > 0$, $\mathbb{P}\{a_n^{-2}\widehat{Q}_j(\boldsymbol{x}_0) > c^*_{j,1-\alpha}(\boldsymbol{x}_0)\} \longrightarrow 1$. Thus, the test is consistent against fixed alternatives.

5 Experiments

We present simulation studies and two data applications in this section. The simulations examine empirical size, power, localization over the covariate profile, and sensitivity to changes in conditional dependence. We then apply the proposed test to Qwen2.5 responses on MMLU-Pro and CLIP semantic scores on ImageNet-9. Additional experiments in Appendix D compare coordinate contribution with predictive variable importance, assess plug-in versus one-step estimation, and provide an Airfoil semi-synthetic study.

Unless otherwise stated, all tests are conducted at nominal level $\alpha = 0.05$. In the simulation studies, we set $d = 1$, $p = 5$, $X^{(1)}, X^{(2)} \overset{\text{iid}}{\sim} \mathrm{Unif}[-1,1]$, and take $j = 1$ as the target coordinate. We use Gaussian kernels for both response components and estimate the two conditional response distributions with conditional RealNVP models (Papamakarios et al., 2021) under two-fold cross-fitting. Inference is based on the multiplier-bootstrap procedure in Section 4.2, and the population criterion $Q_1(x_0)$ is reported when analytically available. Experiment-specific data-generating mechanisms are described below, with additional simulation details provided in Appendix D.

5.1 Size, Power, and Localization

We first study the five-dimensional conditional Gaussian model

$$Y^{(1)} \mid X = x \sim \mathcal{N}(m(x), \Sigma), \qquad Y^{(2)} \mid X = x \sim \mathcal{N}(m(x) + \Delta(x)e_1, \Sigma),$$

where $e_1 = (1, 0, \ldots, 0)^\top$, $m(x) = \big(0.4\sin(\pi x),\, 0.3x,\, 0.25\cos(\pi x),\, 0.2(x^2 - 1/3),\, -0.2x\big)^\top$ and $\Sigma = (\Sigma_{rs})_{1\le r,s\le 5}$ with $\Sigma_{rs} = 0.4^{|r-s|}$.

We consider an exact null $\Delta(x) = 0$, a global alternative $\Delta(x) = 0.45$, and a localized alternative $\Delta(x) = 0.75 b_{0.3}(x)$, where $b_{0.3}$ is a smooth bump supported on $(-0.3, 0.3)$; its explicit form is given in Appendix D. Hence $Q_1(x) = 0$ everywhere under the exact null, $Q_1(x) > 0$ everywhere under the global alternative, and $Q_1(x) > 0$ exactly when $|x| < 0.3$ under the localized alternative. We use $n_1 = n_2 \in \{200, 500, 1000\}$.

Figure 2 shows that empirical size remains close to the nominal level at profiles where $Q_1(x_0) = 0$, both under the exact null and outside the support of the localized alternative. At profiles where $Q_1(x_0) > 0$, empirical power increases with sample size. Under the localized alternative, power is concentrated in the region where the coordinate null fails, while empirical size returns to the nominal level outside that region. These results show near-nominal empirical size under the examined null settings, increasing power under fixed alternatives, and localization of the region where $Q_1(x_0) > 0$.

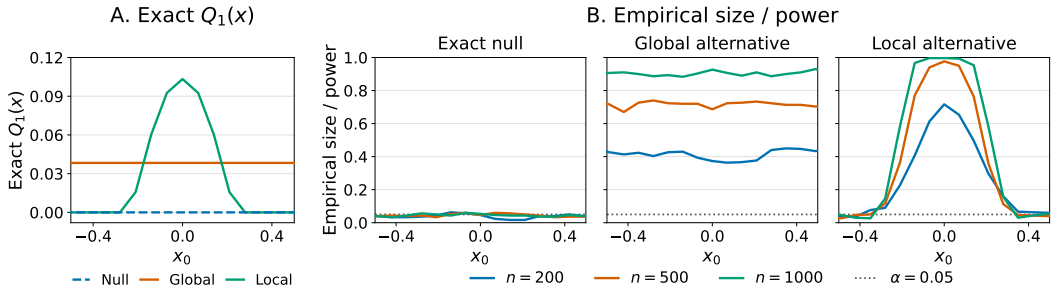

Figure 2: **Pointwise size, power, and localization.** A: Exact $Q_1(x)$ under the null, global alternative, and local alternative. B: Corresponding empirical size and power for $n_1 = n_2 \in \{200, 500, 1000\}$: size where $Q_1(x_0) = 0$ and power where $Q_1(x_0) > 0$. The dotted line marks $\alpha = 0.05$.

5.2 Coordinate Contributions Through Conditional Dependence

To isolate dependence changes, we generate $Y^{(\ell)} \mid X = x \sim \mathcal{N}(0, \Sigma(\rho_\ell(x)))$, where $\Sigma(r)$ has unit diagonal, $\Sigma_{1k}(r) = \Sigma_{k1}(r) = r$ for $k = 2, \ldots, 5$, and all other off-diagonal entries equal to zero. Let $\rho_1(x) = -\rho_2(x) = 0.5\Delta_\rho \exp(-x^2/0.18)$. Thus, every coordinate has conditional marginal distribution $N(0, 1)$ in both groups, while only the dependence between Y_1 and Y_{-1} changes. We consider $n_1 = n_2 \in \{500, 1000\}$. For the location study, we set $\Delta_\rho = 0.8$; for the signal-strength study at $x_0 = 0$, we use $\Delta_\rho \in \{0, 0.2, 0.4, 0.6, 0.8\}$.

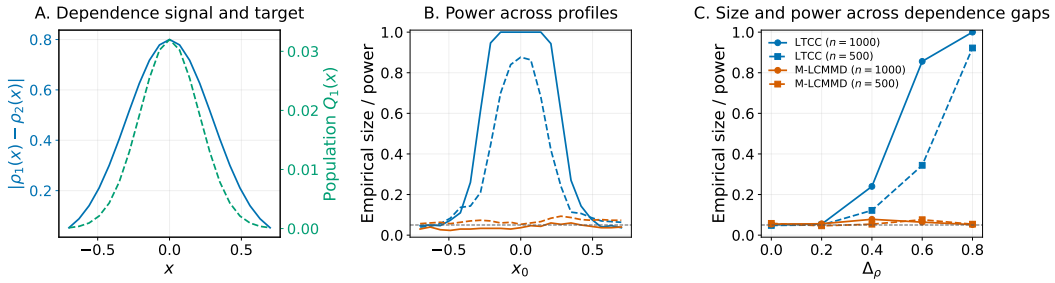

Figure 3: **Coordinate contributions through conditional dependence.** A: Dependence discrepancy and population $Q_1(x)$. B: Empirical power across x_0 for $n_1 = n_2 \in \{500, 1000\}$ at $\Delta_\rho = 0.8$. C: Empirical size at $\Delta_\rho = 0$ and power for $\Delta_\rho > 0$ at $x_0 = 0$.

Figure 3 shows that LTCC detects dependence-only changes despite identical conditional marginals. Its power is highest near $x_0 = 0$, where $Q_1(x_0)$ is largest. For $\Delta_\rho > 0$, power increases with both Δ_ρ and the sample size. In contrast, Marginal Local-CMMD (Yan et al., 2022) remains near the nominal level across signal strengths and sample sizes because the conditional marginal distribution of Y_1 is unchanged by construction. Thus, LTCC can identify coordinate contributions arising purely through changes in dependence.

5.3 Conditional Response Auditing with an LLM

We study conditional response auditing using precomputed Qwen2.5-7B-Instruct (Qwen et al., 2025) responses on MMLU-Pro (Wang et al., 2024). The conditioning variable X indexes a reference-model-based question profile, and the five-dimensional response contains scores for Math, Physics, Chemistry, Engineering, and Biology. We introduce an advice-routing mechanism whose activation probability depends on both the routing intensity ρ and the conditioning profile X, and examine how the resulting coordinate contributions vary across profiles.

Group 1 always uses the neutral response, whereas in group 2 the Math and Physics coordinates may use responses generated with auxiliary advice. The probability of receiving advice increases with

the routing intensity ρ and varies with X, with a higher probability at $x_0 = 0.75$ than at $x_0 = 0.25$. The exact routing mechanism is given in Appendix D.5. We generate independent groups with $n_1 = n_2 = 500$ and test the Physics coordinate at $x_0 \in \{0.25, 0.75\}$ for $\rho \in \{0, 0.05, 0.10, 0.20, 0.30\}$.

Table 1: **Empirical size and power under varying routing intensity.** Entries at $\rho = 0$ report empirical size and those at $\rho > 0$ report empirical power, based on 100 independently resampled snapshots.

Routing intensity ρ	0.00	0.05	0.10	0.20	0.30
Physics, $x_0 = 0.25$	0.070	0.040	0.050	0.110	0.280
Physics, $x_0 = 0.75$	0.060	0.080	0.140	0.650	0.910

At $\rho = 0$, the empirical sizes are 0.07 and 0.06 at $x_0 = 0.25$ and $x_0 = 0.75$, respectively. For $\rho > 0$, empirical power increases with routing intensity at both profiles and rises substantially faster at $x_0 = 0.75$, reaching 0.28 and 0.91 at $\rho = 0.30$. This pattern is consistent with the larger local advice probability at the higher conditioning profile. Further experimental details are given in Appendix D.5.

5.4 ImageNet-9 Background Sensitivity

We ask which CLIP semantic-score coordinates contribute to the conditional distributional discrepancy between Mixed-Same and Mixed-Rand, and how these contributions vary with foreground coverage. Using ImageNet-9 (Xiao et al., 2021), Mixed-Same pairs foregrounds with backgrounds from the same class, whereas Mixed-Rand uses backgrounds sampled without matching the foreground class. For a frozen CLIP ViT-B/32 model (Radford et al., 2021), X represents foreground coverage and $Y \in \mathbb{R}^9$ contains the cosine similarities between the image embedding and the text embeddings for the nine ImageNet-9 classes. Using non-overlapping foreground samples, we perform pointwise tests at foreground coverages $x_0 \in \{6.0\%, 25.6\%, 69.3\%\}$. Further details are given in Appendix D.5.

Table 2: Pointwise p-values for the ImageNet-9 background comparison. Bold entries indicate $p < 0.05$.

x_0	Dog	Bird	Vehicle	Reptile	Carnivore	Insect	Instrument	Primate	Fish
Low (6.0%)	**0.017**	0.060	0.565	0.194	**0.037**	**0.030**	0.566	0.150	**0.028**
Middle (25.6%)	0.212	0.137	0.933	0.481	0.815	0.349	0.229	**0.015**	0.273
High (69.3%)	0.716	0.953	0.423	0.669	0.755	0.704	0.709	0.582	0.927

At the nominal pointwise 0.05 level, Dog, Carnivore, Insect, and Fish have $p < 0.05$ at low foreground coverage, while Primate has $p < 0.05$ at middle coverage; none falls below 0.05 at high coverage. The pointwise pattern is concentrated at lower foreground coverage, consistent with stronger background sensitivity when the foreground occupies less of the image.

6 Conclusion

We studied coordinate contributions to conditional distributional discrepancies. LTCC identifies which coordinates of a multivariate response contribute to a conditional distributional difference at a given covariate profile through changes in the target marginal distribution or its marginal-standardized cross-association with the remaining responses. Population-specific conditional PITs are central to this formulation, as they separate association changes from marginal changes in other coordinates. We introduced a coordinate-level distributional contrast, developed a cross-fitted one-step estimator compatible with flexible conditional distribution estimators, and established valid inference using a multiplier bootstrap. Across synthetic and real-data experiments, LTCC identifies contributing response coordinates across different covariate profiles, complementing global conditional two-sample testing by revealing which coordinates carry the discrepancy and how their contributions vary with the conditioning profile.

AI USE STATEMENT

We used generative AI tools to aid in polishing the writing of this paper, including improving grammar, clarity, and readability. All AI-assisted edits were carefully reviewed by the authors before inclusion in the final manuscript.

REFERENCES

Thomas Brooks, D. Pope, and Michael Marcolini. Airfoil Self-Noise. UCI Machine Learning Repository, 1989. DOI: https://doi.org/10.24432/C5VW2C.

Kailash Budhathoki, Dominik Janzing, Patrick Bloebaum, and Hoiyi Ng. Why did the distribution change? In *International Conference on Artificial Intelligence and Statistics*, pp. 1666–1674. PMLR, 2021.

Geng Chen, Yinxu Jia, Guanghui Wang, and Changliang Zou. Zipper: Addressing degeneracy in algorithm-agnostic inference. *Advances in Neural Information Processing Systems*, 37:65884–65911, 2024.

Xingshu Chen, Yifeng Guo, and Jin-Hong Du. Flow-disentangled feature importance. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=Fx8AtzGlTS.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Kacper Chwialkowski, Dino Sejdinovic, and Arthur Gretton. A wild bootstrap for degenerate kernel tests. *Advances in neural information processing systems*, 27, 2014.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander J. Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(1):723–773, 2012. URL https://jmlr.org/papers/v13/gretton12a.html.

Sean Kulinski and David I Inouye. Towards explaining distribution shifts. In *International conference on machine learning*, pp. 17931–17952. PMLR, 2023.

Sean Kulinski, Saurabh Bagchi, and David I Inouye. Feature shift detection: Localizing which features have shifted via conditional distribution tests. *Advances in neural information processing systems*, 33:19523–19533, 2020.

Anne Leucht and Michael H Neumann. Dependent wild bootstrap for degenerate u-and v-statistics. *Journal of Multivariate Analysis*, 117:257–280, 2013.

Pierre-François Massiani, Christian Fiedler, Lukas Haverbeck, Friedrich Solowjow, and Sebastian Trimpe. Kernel conditional tests from learning-theoretic bounds. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=hJKDwf32Xu.

Peter Moskvichev, Siu Lun Chau, and Dino Sejdinovic. Measuring differences between conditional distributions using kernel embeddings. *arXiv preprint arXiv:2605.02260*, 2026.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

Barnabás Póczos, Zoubin Ghahramani, and Jeff Schneider. Copula-based kernel dependency measures. In *Proceedings of the 29th International Conference on Machine Learning*, 2012.

Qwen, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li,

Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 Technical Report, January 2025. URL http://arxiv.org/abs/2412.15115. arXiv:2412.15115 [cs].

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PMLR, 2021.

Yong Ren, Jun Zhu, Jialian Li, and Yucen Luo. Conditional generative moment-matching networks. *Advances in Neural Information Processing Systems*, 29, 2016.

Xinwei Shen and Nicolai Meinshausen. Engression: extrapolation through the lens of distributional regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(3):653–677, 2025.

Bharath K Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert RG Lanckriet. Hilbert space embeddings and metrics on probability measures. *The Journal of Machine Learning Research*, 11:1517–1561, 2010.

Noël Veraverbeke, Marek Omelka, and Irène Gijbels. Estimation of a conditional copula and association measures. *Scandinavian Journal of Statistics*, 38(4):766–780, 2011. doi: 10.1111/j.1467-9469.2011.00744.x.

Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran Arulraj, Xuan He, Ziyan Jiang, et al. Mmlu-pro: A more robust and challenging multi-task language understanding benchmark. *Advances in Neural Information Processing Systems*, 37:95266–95290, 2024.

Brian D Williamson, Peter B Gilbert, Noah R Simon, and Marco Carone. A general framework for inference on algorithm-agnostic variable importance. *Journal of the American Statistical Association*, 118(543):1645–1658, 2023.

Xiaoyang Wu, Lin Lu, Zhaojun Wang, and Changliang Zou. Conditional testing based on localized conformal p-values. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Ip6UwB35uT.

Kai Yuanqing Xiao, Logan Engstrom, Andrew Ilyas, and Aleksander Madry. Noise or signal: The role of image backgrounds in object recognition. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=gl3D-xY7wLq.

Jian Yan, Zhuoxi Li, and Xianyang Zhang. Distance and kernel-based measures for global and local two-sample conditional distribution testing. *arXiv preprint arXiv:2210.08149*, 2022.

A A GAUSSIAN CHARACTERIZATION OF COORDINATE CONTRIBUTION AND THE ROLE OF PIT STANDARDIZATION

We give a complementary Gaussian characterization of coordinate contribution. Throughout, $\boldsymbol{x}_0$ is fixed and suppressed, and Φ denotes the standard normal cumulative distribution function. Let

$$\boldsymbol{Y}^{(1)} \sim \mathcal{N}_3\big(\mathbf{0}, \Sigma(\rho_1)\big), \qquad \boldsymbol{Y}^{(2)} \sim \mathcal{N}_3\big((0,0,\delta)^\top, \Sigma(\rho_2)\big), \qquad \Sigma(\rho) = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

with $\rho_1, \rho_2 \in (-1, 1)$.

Proposition A.1. *Suppose that k_F and k_C are characteristic. For the model above,*

$$Q_1(\boldsymbol{x}_0) = 0 \iff \rho_1 = \rho_2, \qquad Q_2(\boldsymbol{x}_0) = 0 \iff \rho_1 = \rho_2, \qquad Q_3(\boldsymbol{x}_0) = 0 \iff \delta = 0.$$

Proof. Write $m_3^{(1)} = 0$ and $m_3^{(2)} = \delta$. In population ℓ, the coordinatewise PIT variables are

$$U_1^{(\ell)} = \Phi\big(Y_1^{(\ell)}\big), \qquad U_2^{(\ell)} = \Phi\big(Y_2^{(\ell)}\big), \qquad U_3^{(\ell)} = \Phi\big(Y_3^{(\ell)} - m_3^{(\ell)}\big),$$

so $U_r^{(\ell)} \sim \mathcal{U}$ for every $r \in \{1,2,3\}$. Moreover, $Y_3^{(\ell)}$ is independent of $(Y_1^{(\ell)}, Y_2^{(\ell)})$, hence $U_3^{(\ell)}$ is independent of $(U_1^{(\ell)}, U_2^{(\ell)})$. Let

$$G_{12}(\rho) = \mathcal{L}\big\{\Phi(Z_1), \Phi(Z_2)\big\}, \qquad (Z_1, Z_2) \sim \mathcal{N}_2\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

denote the bivariate Gaussian copula with correlation parameter ρ. Since the third coordinate is independent of the first two, the transformed joint distribution factorizes as

$$G^{(\ell)} = G_{12}(\rho_\ell) \otimes \mathcal{U}. \tag{15}$$

Coordinate $j = 1$. Because $U_2^{(\ell)}$ and $U_3^{(\ell)}$ are independent uniform, $G_{-1}^{(\ell)} = \mathcal{U} \otimes \mathcal{U}$, and (1) together with (15) gives

$$C_1^{(\ell)} = \big\{G_{12}(\rho_\ell) - \mathcal{U} \otimes \mathcal{U}\big\} \otimes \mathcal{U}.$$

If $\rho_1 = \rho_2$, then $C_1^{(1)} = C_1^{(2)}$. Conversely, suppose that $C_1^{(1)} = C_1^{(2)}$ and evaluate both signed measures at the rectangle $A = [0, 1/2] \times [0, 1/2] \times [0, 1]$. Since $\mathcal{U}([0,1]) = 1$,

$$C_1^{(\ell)}(A) = \big(G_{12}(\rho_\ell) - \mathcal{U} \otimes \mathcal{U}\big)\big([0,1/2]^2\big) = \mathbb{P}\big(Y_1^{(\ell)} \le 0,\, Y_2^{(\ell)} \le 0\big) - \frac{1}{4},$$

where the second equality uses that $U_r^{(\ell)} \le 1/2$ if and only if $Y_r^{(\ell)} \le 0$. For the standard bivariate normal distribution with correlation ρ_ℓ,

$$\mathbb{P}\big(Y_1^{(\ell)} \le 0,\, Y_2^{(\ell)} \le 0\big) = \frac{1}{4} + \frac{\arcsin(\rho_\ell)}{2\pi},$$

hence $C_1^{(\ell)}(A) = \arcsin(\rho_\ell)/(2\pi)$. Since $\arcsin$ is strictly increasing on $(-1, 1)$, $\rho_1 \neq \rho_2$ implies $C_1^{(1)}(A) \neq C_1^{(2)}(A)$ and therefore $C_1^{(1)} \neq C_1^{(2)}$. We conclude that $C_1^{(1)} = C_1^{(2)}$ if and only if $\rho_1 = \rho_2$. On the other hand, $F_1^{(1)} = F_1^{(2)} = \mathcal{N}(0,1)$ for all ρ_1, ρ_2, so $H_{0,1}(\boldsymbol{x}_0)$ holds if and only if $\rho_1 = \rho_2$, and Proposition 2.1 gives $Q_1(\boldsymbol{x}_0) = 0$ if and only if $\rho_1 = \rho_2$.

Coordinate $j = 2$. By symmetry between the first two coordinates, $F_2^{(1)} = F_2^{(2)} = \mathcal{N}(0,1)$ and $C_2^{(1)} = C_2^{(2)}$ if and only if $\rho_1 = \rho_2$, so $Q_2(\boldsymbol{x}_0) = 0$ if and only if $\rho_1 = \rho_2$.

Coordinate $j = 3$. Because $Y_3^{(\ell)}$ is independent of $\boldsymbol{Y}_{-3}^{(\ell)}$, the transformed joint distribution factorizes as $G^{(\ell)} = \mathcal{U} \otimes G_{-3}^{(\ell)}$, with the coordinates placed in their natural positions. Therefore $C_3^{(\ell)} = 0$ for $\ell \in \{1,2\}$, and $C_3^{(1)} = C_3^{(2)}$ regardless of ρ_1 and ρ_2. The coordinate marginals are $F_3^{(1)} = \mathcal{N}(0,1)$ and $F_3^{(2)} = \mathcal{N}(\delta, 1)$, so $F_3^{(1)} = F_3^{(2)}$ if and only if $\delta = 0$. Hence $H_{0,3}(\boldsymbol{x}_0)$ holds if and only if $\delta = 0$, and Proposition 2.1 gives $Q_3(\boldsymbol{x}_0) = 0$ if and only if $\delta = 0$. □

Proposition A.1 makes the attribution pattern of the proposed criterion explicit. If $\rho_1 = \rho_2$ and $\delta \neq 0$, then only coordinate 3 contributes through its marginal distribution. If $\rho_1 \neq \rho_2$ and $\delta = 0$, then coordinates 1 and 2 contribute through changes in their marginal-standardized cross-association. If both parameters change, all three coordinates are detected. The next result shows why the PIT standardization is needed for this attribution.

Proposition A.2. *Consider the same Gaussian model and define the raw-scale analogue of the cross-association measure by*

$$\widetilde{C}_j^{(\ell)} = F^{(\ell)} - F_j^{(\ell)} \otimes F_{-j}^{(\ell)}.$$

Suppose that $\rho_1 = \rho_2 = \rho \neq 0$ and $\delta \neq 0$. Then the only population difference is the marginal distribution of Y_3. However,

$$\widetilde{C}_1^{(1)} \neq \widetilde{C}_1^{(2)}, \qquad \widetilde{C}_2^{(1)} \neq \widetilde{C}_2^{(2)}.$$

Thus, a raw-scale association comparison would incorrectly attribute the marginal change in Y_3 to coordinates 1 and 2.

Proof. When $\rho_1 = \rho_2 = \rho$, the joint distribution of (Y_1, Y_2) is identical across populations and Y_3 is independent of (Y_1, Y_2). Hence

$$F^{(\ell)} = F_{12,\rho} \otimes F_3^{(\ell)},$$

where $F_{12,\rho}$ denotes the common distribution of (Y_1, Y_2). Because

$$F_1^{(\ell)} = F_2^{(\ell)} = \Phi,$$

we have

$$\widetilde{C}_1^{(\ell)} = \{F_{12,\rho} - \Phi \otimes \Phi\} \otimes F_3^{(\ell)}.$$

Consider the rectangle

$$A = (-\infty, 0] \times (-\infty, 0] \times (-\infty, 0].$$

For the first two coordinates,

$$\{F_{12,\rho} - \Phi \otimes \Phi\}\big((-\infty, 0]^2\big) = \Pr(Y_1 \leq 0, Y_2 \leq 0) - \frac{1}{4} = \frac{\arcsin(\rho)}{2\pi}.$$

Therefore,

$$\widetilde{C}_1^{(1)}(A) = \frac{\arcsin(\rho)}{2\pi}\frac{1}{2},$$

whereas

$$\widetilde{C}_1^{(2)}(A) = \frac{\arcsin(\rho)}{2\pi}\Phi(-\delta).$$

Since $\delta \neq 0$, $\Phi(-\delta) \neq 1/2$, and hence

$$\widetilde{C}_1^{(1)}(A) \neq \widetilde{C}_1^{(2)}(A).$$

Thus $\widetilde{C}_1^{(1)} \neq \widetilde{C}_1^{(2)}$. The result for coordinate 2 follows by symmetry. □

This example illustrates the role of the PIT standardization. Although the dependence structure involving Y_1 and Y_2 is unchanged, a raw-scale association measure changes because the marginal distribution of another coordinate changes. In contrast, the population-specific PIT maps all coordinate marginals to the same uniform scale. Therefore, changes in $C_j^{(\ell)}$ reflect changes in marginal-standardized association rather than differences induced solely by nuisance marginal distributions of other coordinates.

B Influence-Function and Estimation Details

This appendix records the explicit influence functions, the exact von Mises remainder, and the Gaussian-kernel derivatives used in Section 3.

B.1 EXPLICIT INFLUENCE FUNCTION

This subsection gives the explicit componentwise influence functions used in Theorem 3.1. The marginal component follows directly from linearity, whereas the association component also accounts for the perturbation of the marginal cdfs entering the PIT map. For a finite signed measure η, let $\|\eta\|_{\mathrm{TV}}$ denote its total variation norm, and let $\delta_{\boldsymbol{y}}$ denote the Dirac measure at $\boldsymbol{y}$. For $\boldsymbol{z} = (z_1, \dots, z_p)^\top$, define

$$\boldsymbol{u}^{(\ell)}(\boldsymbol{z}) = \big(F_1^{(\ell)}(z_1), \dots, F_p^{(\ell)}(z_p)\big)^\top.$$

Fix $\boldsymbol{y} = (y_1, \dots, y_p)^\top$ and consider the contamination path

$$F_{\varepsilon,\boldsymbol{y}}^{(\ell)} = (1-\varepsilon)F^{(\ell)} + \varepsilon\delta_{\boldsymbol{y}}. \tag{16}$$

Its rth marginal cdf and the associated PIT map are

$$F_{\varepsilon,\boldsymbol{y},r}^{(\ell)}(z_r) = F_r^{(\ell)}(z_r) + \varepsilon\{\mathbf{1}(y_r \le z_r) - F_r^{(\ell)}(z_r)\}, \tag{17}$$

$$\boldsymbol{u}_{\varepsilon,\boldsymbol{y}}^{(\ell)}(\boldsymbol{z}) = \Big(F_{\varepsilon,\boldsymbol{y},1}^{(\ell)}(z_1), \dots, F_{\varepsilon,\boldsymbol{y},p}^{(\ell)}(z_p)\Big)^\top. \tag{18}$$

For the perturbation calculation, $C_j(F_{\varepsilon,\boldsymbol{y}}^{(\ell)})$ is evaluated by the same transformed-joint-distribution minus independent-reference expression as in (1).

The marginal components of the two influence functions are

$$\phi_{F,j}^{(1)}(\boldsymbol{y}; F^{(1)}) = -\mu_{k_F}\big(\delta_{y_j} - F_j^{(1)}\big), \qquad \phi_{F,j}^{(2)}(\boldsymbol{y}; F^{(2)}) = \mu_{k_F}\big(\delta_{y_j} - F_j^{(2)}\big). \tag{19}$$

The association components are

$$\begin{aligned}
\phi_{C,j}^{(1)}(\boldsymbol{y}; F^{(1)}) = \; -\Bigg[& k_C\big(\cdot, \boldsymbol{u}^{(1)}(\boldsymbol{y})\big) - \int_0^1 k_C\big(\cdot, (v, \boldsymbol{u}_{-j}^{(1)}(\boldsymbol{y}))\big)\,\mathrm{d}v - \mu_{k_C}\big(C_j^{(1)}\big) \\
& + \int \frac{\partial}{\partial\varepsilon} k_C\big(\cdot, \boldsymbol{u}_{\varepsilon,\boldsymbol{y}}^{(1)}(\boldsymbol{z})\big)\Big|_{\varepsilon=0} \mathrm{d}F^{(1)}(\boldsymbol{z}) \\
& - \int\int_0^1 \frac{\partial}{\partial\varepsilon} k_C\big(\cdot, (v, \boldsymbol{u}_{\varepsilon,\boldsymbol{y},-j}^{(1)}(\boldsymbol{z}))\big)\Big|_{\varepsilon=0} \mathrm{d}v\,\mathrm{d}F^{(1)}(\boldsymbol{z})\Bigg], \\
\phi_{C,j}^{(2)}(\boldsymbol{y}; F^{(2)}) = \; \Bigg[& k_C\big(\cdot, \boldsymbol{u}^{(2)}(\boldsymbol{y})\big) - \int_0^1 k_C\big(\cdot, (v, \boldsymbol{u}_{-j}^{(2)}(\boldsymbol{y}))\big)\,\mathrm{d}v - \mu_{k_C}\big(C_j^{(2)}\big) \\
& + \int \frac{\partial}{\partial\varepsilon} k_C\big(\cdot, \boldsymbol{u}_{\varepsilon,\boldsymbol{y}}^{(2)}(\boldsymbol{z})\big)\Big|_{\varepsilon=0} \mathrm{d}F^{(2)}(\boldsymbol{z}) \\
& - \int\int_0^1 \frac{\partial}{\partial\varepsilon} k_C\big(\cdot, (v, \boldsymbol{u}_{\varepsilon,\boldsymbol{y},-j}^{(2)}(\boldsymbol{z}))\big)\Big|_{\varepsilon=0} \mathrm{d}v\,\mathrm{d}F^{(2)}(\boldsymbol{z})\Bigg].
\end{aligned} \tag{20}$$

The last two lines in each expression account for the perturbation of the marginal cdfs entering the PIT transformation.

For the remainder calculations, the smoothness requirement in Assumption 3.1 includes the following technical conditions: along every line segment $\{\boldsymbol{u} + t\boldsymbol{a} : t \in [0,1]\} \subset [0,1]^p$, the first two derivatives below exist in $\mathcal{H}_C$ and satisfy, for finite constants L_1, L_2,

$$\begin{aligned}
\sup_{t\in[0,1]} \left\| \frac{\partial}{\partial t} k_C(\cdot, \boldsymbol{u} + t\boldsymbol{a}) \right\|_{\mathcal{H}_C} &\le L_1 \|\boldsymbol{a}\|_\infty, \\
\sup_{t\in[0,1]} \left\| \frac{\partial^2}{\partial t^2} k_C(\cdot, \boldsymbol{u} + t\boldsymbol{a}) \right\|_{\mathcal{H}_C} &\le L_2 \|\boldsymbol{a}\|_\infty^2.
\end{aligned} \tag{21}$$

These bounds control perturbations of the PIT arguments and the second-order remainder. They hold, in particular, for the Gaussian kernel with any fixed positive bandwidth.

B.2 Von Mises expansion and second-order remainder

We now make explicit the remainder $R_{\gamma,j}$ in the two-sample von Mises expansion (5). The marginal component is linear in the underlying distribution and hence has no remainder; only the association component contributes a second-order term. For arbitrary reference distributions $\bar{F}^{(1)}$ and $\bar{F}^{(2)}$, set $\Delta^{(\ell)} = F^{(\ell)} - \bar{F}^{(\ell)}$ and

$$\rho_\ell = \max_{r\in[p]} \sup_{s\in\mathbb{R}} \left| F_r^{(\ell)}(s) - \bar{F}_r^{(\ell)}(s) \right|.$$

For each population define

$$F_t^{(\ell)} = \bar{F}^{(\ell)} + t\Delta^{(\ell)}, \qquad t \in [0,1],$$

$$\boldsymbol{u}_t^{(\ell)}(\boldsymbol{z}) = \left(\bar{F}_1^{(\ell)}(z_1) + t\{F_1^{(\ell)}(z_1) - \bar{F}_1^{(\ell)}(z_1)\}, \ldots, \bar{F}_p^{(\ell)}(z_p) + t\{F_p^{(\ell)}(z_p) - \bar{F}_p^{(\ell)}(z_p)\} \right)^\top,$$

and

$$H_{j,t}^{(\ell)}(\boldsymbol{z}) = k_C\big(\cdot, \boldsymbol{u}_t^{(\ell)}(\boldsymbol{z})\big) - \int_0^1 k_C\big(\cdot, (v, \boldsymbol{u}_{t,-j}^{(\ell)}(\boldsymbol{z}))\big)\, \mathrm{d}v. \tag{22}$$

Then the exact association remainder is recorded below.

$$R_{\gamma,j} = \begin{pmatrix} 0 \\ R_{C,j}^{(2)} - R_{C,j}^{(1)} \end{pmatrix}, \tag{23}$$

$$R_{C,j}^{(\ell)} = \int_0^1 (1-t) \left[2 \int \frac{\partial}{\partial t} H_{j,t}^{(\ell)}(\boldsymbol{z})\, \mathrm{d}\Delta^{(\ell)}(\boldsymbol{z}) + \int \frac{\partial^2}{\partial t^2} H_{j,t}^{(\ell)}(\boldsymbol{z})\, \mathrm{d}F_t^{(\ell)}(\boldsymbol{z}) \right] \mathrm{d}t. \tag{24}$$

Moreover,

$$\left\| R_{C,j}^{(\ell)} \right\|_{\mathcal{H}_C} \le 2L_1\rho_\ell \left\| \Delta^{(\ell)} \right\|_{\mathrm{TV}} + L_2\rho_\ell^2 \le (2L_1 + L_2) \left\| \Delta^{(\ell)} \right\|_{\mathrm{TV}}^2, \tag{25}$$

and hence

$$\|R_{\gamma,j}\|_{\mathcal{H}_j} \le (2L_1 + L_2) \left(\left\| \Delta^{(1)} \right\|_{\mathrm{TV}}^2 + \left\| \Delta^{(2)} \right\|_{\mathrm{TV}}^2 \right). \tag{26}$$

B.3 Closed-form evaluation for the Gaussian kernel

For the Gaussian association kernel used in our experiments, the derivatives appearing in the influence function and the remainder bounds admit closed-form expressions. This subsection records these formulas and verifies the derivative bounds required in Assumption 3.1. Suppose that the association component uses

$$k_C(\boldsymbol{w}, \boldsymbol{u}) = \exp\left\{ -\frac{\|\boldsymbol{w} - \boldsymbol{u}\|^2}{2\sigma^2} \right\}, \qquad \boldsymbol{w}, \boldsymbol{u} \in \mathbb{R}^p, \quad \sigma > 0.$$

By the reproducing property and differentiation of the Gaussian kernel, for every $\boldsymbol{u}, \boldsymbol{a} \in \mathbb{R}^p$ and $t \in \mathbb{R}$,

$$\left\| \frac{\partial}{\partial t} k_C(\cdot, \boldsymbol{u} + t\boldsymbol{a}) \right\|_{\mathcal{H}_C} = \frac{\|\boldsymbol{a}\|_2}{\sigma} \le \frac{\sqrt{p}}{\sigma} \|\boldsymbol{a}\|_\infty,$$

$$\left\| \frac{\partial^2}{\partial t^2} k_C(\cdot, \boldsymbol{u} + t\boldsymbol{a}) \right\|_{\mathcal{H}_C} = \frac{\sqrt{3}\,\|\boldsymbol{a}\|_2^2}{\sigma^2} \le \frac{\sqrt{3}\,p}{\sigma^2} \|\boldsymbol{a}\|_\infty^2.$$

Thus, in (21), one may take

$$L_1 = \frac{\sqrt{p}}{\sigma}, \qquad L_2 = \frac{\sqrt{3}\,p}{\sigma^2}.$$

These bounds hold uniformly on $\mathbb{R}^p$. For any $\boldsymbol{w} \in \mathbb{R}^p$, equations (17)–(18) give

$$\frac{\partial}{\partial \varepsilon} k_C\big(\boldsymbol{w}, \boldsymbol{u}_{\varepsilon,\boldsymbol{y}}^{(\ell)}(\boldsymbol{z})\big)\Big|_{\varepsilon=0} = \frac{k_C\big(\boldsymbol{w}, \boldsymbol{u}^{(\ell)}(\boldsymbol{z})\big)}{\sigma^2} \sum_{r=1}^p \{w_r - F_r^{(\ell)}(z_r)\}\{\mathbf{1}(y_r \le z_r) - F_r^{(\ell)}(z_r)\}. \tag{27}$$

Similarly,

$$\frac{\partial}{\partial \varepsilon} k_C\big(\boldsymbol{w}, (v, \boldsymbol{u}^{(\ell)}_{\varepsilon,\boldsymbol{y},-j}(\boldsymbol{z}))\big)\Big|_{\varepsilon=0} = \frac{k_C\big(\boldsymbol{w}, (v, \boldsymbol{u}^{(\ell)}_{-j}(\boldsymbol{z}))\big)}{\sigma^2} \sum_{r\neq j}\{w_r - F_r^{(\ell)}(z_r)\} \times \{\mathbf{1}(y_r \le z_r) - F_r^{(\ell)}(z_r)\}. \tag{28}$$

Thus, the influence function requires only marginal cdfs, indicator comparisons, and derivatives of the association kernel k_C; no marginal density estimation is needed.

C Proofs and Technical Details

C.1 Proof of Proposition 2.1

Proof. By the direct-sum norm, $Q_j(\boldsymbol{x}_0) = 0$ if and only if both components of $\gamma_j(\boldsymbol{x}_0)$ vanish. Since k_F is characteristic on $\mathbb{R}$,

$$\mu_{k_F}\big(F_j^{(2)} - F_j^{(1)}\big) = 0 \quad \Longleftrightarrow \quad F_j^{(1)} = F_j^{(2)}.$$

For the association component,

$$\mu_{k_C}\big(C_j^{(2)} - C_j^{(1)}\big) = 0$$

is equivalent to

$$\mu_{k_C}\big(G^{(2)} + \mathcal{U} \otimes G^{(1)}_{-j}\big) = \mu_{k_C}\big(G^{(1)} + \mathcal{U} \otimes G^{(2)}_{-j}\big).$$

Both sides are embeddings of finite positive measures with total mass two. Dividing both measures by two and using that k_C is characteristic on $\mathbb{R}^p$ gives

$$G^{(2)} + \mathcal{U} \otimes G^{(1)}_{-j} = G^{(1)} + \mathcal{U} \otimes G^{(2)}_{-j},$$

which is equivalent to $C_j^{(1)} = C_j^{(2)}$. The converse implications follow immediately from the definition of γ_j. Therefore,

$$H_{0,j}(\boldsymbol{x}_0) \quad \Longleftrightarrow \quad \gamma_j(\boldsymbol{x}_0) = 0 \quad \Longleftrightarrow \quad Q_j(\boldsymbol{x}_0) = 0.$$

It remains to prove the joint characterization. If $F^{(1)} = F^{(2)}$, then all coordinate marginals and the population-specific PIT distributions coincide, and hence $C_j^{(1)} = C_j^{(2)}$ for every j. The first part therefore gives $Q_j(\boldsymbol{x}_0) = 0$ for all j.

Conversely, suppose that $Q_j(\boldsymbol{x}_0) = 0$ for every $j \in [p]$. By the first part,

$$F_j^{(1)} = F_j^{(2)} \quad \text{and} \quad C_j^{(1)} = C_j^{(2)}, \qquad j = 1, \dots, p.$$

Let

$$D = G^{(2)} - G^{(1)}$$

denote the difference between the two PIT joint distributions, and let D_A denote its marginal on a coordinate set $A \subseteq [p]$. Since

$$C_j^{(\ell)} = G^{(\ell)} - \mathcal{U} \otimes G^{(\ell)}_{-j},$$

the equality $C_j^{(1)} = C_j^{(2)}$ implies

$$D = \mathcal{U} \otimes D_{-j}, \qquad j = 1, \dots, p,$$

with the uniform coordinate inserted in position j.

Taking $j = 1$ gives

$$D = \mathcal{U}_1 \otimes D_{\{2,\dots,p\}}.$$

For $m = 2, \dots, p-1$, marginalizing the identity $D = \mathcal{U}_m \otimes D_{-m}$ over coordinates $1, \dots, m-1$ gives

$$D_{\{m,\dots,p\}} = \mathcal{U}_m \otimes D_{\{m+1,\dots,p\}}.$$

Applying these identities successively yields

$$D = \mathcal{U}_1 \otimes \cdots \otimes \mathcal{U}_{p-1} \otimes D_{\{p\}}.$$

Each PIT coordinate is uniform in both populations, so

$$D_{\{p\}} = G_p^{(2)} - G_p^{(1)} = \mathcal{U} - \mathcal{U} = 0.$$

Hence $D = 0$, and therefore

$$G^{(1)} = G^{(2)}.$$

We have already shown that $F_j^{(1)} = F_j^{(2)}$ for every j. Since the conditional marginal distributions are continuous, $G^{(\ell)}$ is the copula associated with $F^{(\ell)}$. Thus, by Sklar's theorem, equality of all coordinate marginals together with $G^{(1)} = G^{(2)}$ implies

$$F^{(1)} = F^{(2)}.$$

This proves the final equivalence. □

C.2 Proof of Theorem 3.1

Proof. Step 1: influence functions. Fix $\boldsymbol{y} \in \mathbb{R}^p$ and perturb one population distribution at a time along (16), holding the other population distribution fixed. For the marginal component, linearity gives

$$\frac{\partial}{\partial \varepsilon}\left\{\mu_{k_F}(F_j^{(2)}) - \mu_{k_F}(F_{\varepsilon,\boldsymbol{y},j}^{(1)})\right\}\Big|_{\varepsilon=0} = -\mu_{k_F}\big(\delta_{y_j} - F_j^{(1)}\big) = \phi_{F,j}^{(1)}(\boldsymbol{y}; F^{(1)}),$$
$$\frac{\partial}{\partial \varepsilon}\left\{\mu_{k_F}(F_{\varepsilon,\boldsymbol{y},j}^{(2)}) - \mu_{k_F}(F_j^{(1)})\right\}\Big|_{\varepsilon=0} = \mu_{k_F}\big(\delta_{y_j} - F_j^{(2)}\big) = \phi_{F,j}^{(2)}(\boldsymbol{y}; F^{(2)}),$$

For the association component, fix $\ell \in \{1, 2\}$. The definition of $C_j(F_{\varepsilon,\boldsymbol{y}}^{(\ell)})$ gives

$$\mu_{k_C}\big(C_j(F_{\varepsilon,\boldsymbol{y}}^{(\ell)})\big) = (1-\varepsilon)\int\left[k_C\big(\cdot, \boldsymbol{u}_{\varepsilon,\boldsymbol{y}}^{(\ell)}(\boldsymbol{z})\big) - \int_0^1 k_C\big(\cdot, (v, \boldsymbol{u}_{\varepsilon,\boldsymbol{y},-j}^{(\ell)}(\boldsymbol{z}))\big)\,\mathrm{d}v\right]\mathrm{d}F^{(\ell)}(\boldsymbol{z})$$
$$+\,\varepsilon\left[k_C\big(\cdot, \boldsymbol{u}_{\varepsilon,\boldsymbol{y}}^{(\ell)}(\boldsymbol{y})\big) - \int_0^1 k_C\big(\cdot, (v, \boldsymbol{u}_{\varepsilon,\boldsymbol{y},-j}^{(\ell)}(\boldsymbol{y}))\big)\,\mathrm{d}v\right].$$

Differentiating at $\varepsilon = 0$ and then differentiating the two-sample contrast itself yields

$$\frac{\partial}{\partial \varepsilon}\left\{\mu_{k_C}(C_j^{(2)}) - \mu_{k_C}\big(C_j(F_{\varepsilon,\boldsymbol{y}}^{(1)})\big)\right\}\Big|_{\varepsilon=0} = \phi_{C,j}^{(1)}(\boldsymbol{y}; F^{(1)}),$$
$$\frac{\partial}{\partial \varepsilon}\left\{\mu_{k_C}\big(C_j(F_{\varepsilon,\boldsymbol{y}}^{(2)})\big) - \mu_{k_C}(C_j^{(1)})\right\}\Big|_{\varepsilon=0} = \phi_{C,j}^{(2)}(\boldsymbol{y}; F^{(2)}),$$

The right-hand sides are exactly the expressions in (20). This proves (4).

Step 2: centering. For either population, the unsigned marginal derivative has mean zero because

$$\mu_{k_F}\left(\int \delta_{y_j}\,\mathrm{d}F^{(\ell)}(\boldsymbol{y}) - F_j^{(\ell)}\right) = 0.$$

Consequently, both $\phi_{F,j}^{(1)}$ and $\phi_{F,j}^{(2)}$ are centered under their corresponding distributions. For the first line of (20),

$$\int\left[k_C\big(\cdot, \boldsymbol{u}^{(\ell)}(\boldsymbol{y})\big) - \int_0^1 k_C\big(\cdot, (v, \boldsymbol{u}_{-j}^{(\ell)}(\boldsymbol{y}))\big)\,\mathrm{d}v\right]\mathrm{d}F^{(\ell)}(\boldsymbol{y}) = \mu_{k_C}\big(C_j^{(\ell)}\big).$$

For each fixed $\boldsymbol{z}$ and coordinate r,

$$\int\left\{\mathbf{1}(y_r \le z_r) - F_r^{(\ell)}(z_r)\right\}\mathrm{d}F^{(\ell)}(\boldsymbol{y}) = 0.$$

Because the derivative of the kernel along the path in (18) is linear in these coordinatewise increments, both derivative terms in (20) also integrate to zero. The unsigned association derivative is

therefore centered; so are both signed expressions in (20). This proves the centering statement in Theorem 3.1.

Step 3: one-group von Mises expansion. Fix one population and suppress the superscript ℓ. Let $\Delta = F - \bar{F}$, $F_t = \bar{F} + t\Delta$, and let $\boldsymbol{u}_t$ and $H_{j,t}$ be defined as in (22). Write

$$B_j(t) = \mu_{k_C}\big(C_j(F_t)\big) = \int H_{j,t}(\boldsymbol{z})\,\mathrm{d}F_t(\boldsymbol{z}).$$

Ordinary differentiation gives

$$\frac{\partial}{\partial t}B_j(t) = \int \frac{\partial}{\partial t}H_{j,t}(\boldsymbol{z})\,\mathrm{d}F_t(\boldsymbol{z}) + \int H_{j,t}(\boldsymbol{z})\,\mathrm{d}\Delta(\boldsymbol{z}), \tag{29}$$

and

$$\frac{\partial^2}{\partial t^2}B_j(t) = \int \frac{\partial^2}{\partial t^2}H_{j,t}(\boldsymbol{z})\,\mathrm{d}F_t(\boldsymbol{z}) + 2\int \frac{\partial}{\partial t}H_{j,t}(\boldsymbol{z})\,\mathrm{d}\Delta(\boldsymbol{z}). \tag{30}$$

We next identify the first derivative at $t = 0$. The direct-observation part of the influence function satisfies

$$\int \big\{H_{j,0}(\boldsymbol{y}) - \mu_{k_C}(C_j(\bar{F}))\big\}\,\mathrm{d}F(\boldsymbol{y}) = \int H_{j,0}(\boldsymbol{y})\,\mathrm{d}\Delta(\boldsymbol{y}).$$

Moreover, for every fixed $\boldsymbol{z}$ and coordinate r,

$$\int \big\{\mathbf{1}(y_r \le z_r) - \bar{F}_r(z_r)\big\}\,\mathrm{d}F(\boldsymbol{y}) = F_r(z_r) - \bar{F}_r(z_r).$$

Averaging the two derivative terms in (20) with respect to $F(\boldsymbol{y})$ therefore gives

$$\int \frac{\partial}{\partial t}H_{j,t}(\boldsymbol{z})\bigg|_{t=0}\,\mathrm{d}\bar{F}(\boldsymbol{z}).$$

Combining the preceding displays with (29) identifies the first-order term with the integral of the positive-sign association expression in (20). Taylor's theorem with integral remainder then gives the corresponding one-group expansion, where (30) gives exactly the remainder $R_{C,j}(F, \bar{F})$ in (24).

The marginal component has no remainder because

$$\mu_{k_F}(F_j) - \mu_{k_F}(\bar{F}_j) = \int \mu_{k_F}\big(\delta_{y_j} - \bar{F}_j\big)\,\mathrm{d}F(\boldsymbol{y}).$$

Applying the one-group expansion with a minus sign to population 1 and a plus sign to population 2, with these signs already built into (4), proves equations (5) and (23).

Step 4: remainder bound. For the path from $\bar{F}$ to F, set

$$\rho = \max_{r\in[p]}\sup_{s\in\mathbb{R}}|F_r(s) - \bar{F}_r(s)|.$$

The direction of $\boldsymbol{u}_t(\boldsymbol{z})$ has sup norm at most ρ. By (21) and the two terms in the definition of $H_{j,t}$,

$$\left\|\frac{\partial}{\partial t}H_{j,t}(\boldsymbol{z})\right\|_{\mathcal{H}_C} \le 2L_1\rho, \qquad \left\|\frac{\partial^2}{\partial t^2}H_{j,t}(\boldsymbol{z})\right\|_{\mathcal{H}_C} \le 2L_2\rho^2.$$

Using (24), $\int_0^1(1-t)\,\mathrm{d}t = 1/2$, and the fact that F_t is a probability measure gives

$$\begin{aligned}\|R_{C,j}(F,\bar{F})\|_{\mathcal{H}_C} &\le \frac{1}{2}\left\{4L_1\rho\|\Delta\|_{\mathrm{TV}}\right\} + L_2\rho^2 \\ &= 2L_1\rho\|\Delta\|_{\mathrm{TV}} + L_2\rho^2.\end{aligned}$$

Since marginalization contracts total variation,

$$\rho \le \max_{r\in[p]}\|F_r - \bar{F}_r\|_{\mathrm{TV}} \le \|F - \bar{F}\|_{\mathrm{TV}}.$$

This proves (25). Applying the triangle inequality to $R^{(2)}_{C,j} - R^{(1)}_{C,j}$ proves (26). □

C.3 ASYMPTOTIC LINEAR REPRESENTATION

For $\ell \in \{1, 2\}$, define the full-sample local average

$$\mathbb{P}_{\ell,n}^{h_\ell} f = \frac{\sum_{i=1}^{n_\ell} G_{h_\ell}(\boldsymbol{X}_i^{(\ell)} - \boldsymbol{x}_0) f(\boldsymbol{Y}_i^{(\ell)})}{\sum_{i=1}^{n_\ell} G_{h_\ell}(\boldsymbol{X}_i^{(\ell)} - \boldsymbol{x}_0)}, \tag{31}$$

and let

$$\mathbb{P}_{n,j}(\boldsymbol{x}_0) = \mathbb{P}_{1,n}^{h_1} \phi_j^{(1)}(\cdot, \boldsymbol{x}_0) + \mathbb{P}_{2,n}^{h_2} \phi_j^{(2)}(\cdot, \boldsymbol{x}_0). \tag{32}$$

Under Assumption 3.1, define the finite uniform bound

$$\bar{\kappa} = \max\left\{ \sup_{t \in \mathbb{R}} k_F(t,t), \sup_{\boldsymbol{u} \in [0,1]^p} k_C(\boldsymbol{u}, \boldsymbol{u}) \right\} < \infty.$$

The boundedness and Lipschitz smoothness in Assumption 3.1 imply the following stability properties of the influence function.

Lemma C.1. *There exist finite constants B_j and L_j, depending only on the kernel bounds and p, such that, for $\ell \in \{1, 2\}$, every probability distribution P in a sufficiently small total-variation neighborhood of $F^{(\ell)}$, and every y,*

$$\|\phi_j^{(\ell)}(y; P)\|_{\mathcal{H}_j} \le B_j, \tag{33}$$

and, for any two such distributions P, Q with continuous coordinate marginals,

$$\sup_y \|\phi_j^{(\ell)}(y; P) - \phi_j^{(\ell)}(y; Q)\|_{\mathcal{H}_j} \le L_j \|P - Q\|_{\mathrm{TV}}. \tag{34}$$

Proof. The marginal component satisfies

$$\|\mu_{k_F}(\delta_{y_j} - P_j)\|_{\mathcal{H}_F} \le 2\sqrt{\bar{\kappa}},$$

and its difference under P and Q is bounded by $\sqrt{\bar{\kappa}}\|P - Q\|_{\mathrm{TV}}$.

For the association component, write

$$u_P(z) = \{P_1(z_1), \ldots, P_p(z_p)\}^\top$$

and

$$H_{j,P}(z) = k_C(\cdot, u_P(z)) - \int_0^1 k_C(\cdot, (v, u_{P,-j}(z)))\, \mathrm{d}v.$$

The first-derivative bound and $\sup_z \|u_P(z) - u_Q(z)\|_\infty \le \|P - Q\|_{\mathrm{TV}}$ give

$$\sup_z \|H_{j,P}(z) - H_{j,Q}(z)\| \le 2L_1 \|P - Q\|_{\mathrm{TV}}. \tag{35}$$

Moreover, $\sup_z \|H_{j,P}(z)\| \le 2\sqrt{\bar{\kappa}}$. Hence

$$\left\| \int H_{j,P}\, \mathrm{d}P - \int H_{j,Q}\, \mathrm{d}Q \right\| \le C \|P - Q\|_{\mathrm{TV}}.$$

For the two derivative terms in (20), define their combined integrand by

$$J_{j,P}(y, z) = \frac{\partial}{\partial \varepsilon} \left[k_C\{\cdot, u_{P,\varepsilon,y}(z)\} - \int_0^1 k_C\{\cdot, (v, u_{P,\varepsilon,y,-j}(z))\}\, \mathrm{d}v \right] \Bigg|_{\varepsilon = 0},$$

where

$$u_{P,\varepsilon,y,r}(z) = P_r(z_r) + \varepsilon\{\mathbf{1}(y_r \le z_r) - P_r(z_r)\}.$$

The direction in this derivative has sup norm at most one. The first-derivative bound therefore gives $\sup_{y,z} \|J_{j,P}(y, z)\| \le 2L_1$. The mean-value theorem, together with the bounded first and second derivatives of $u \mapsto k_C(\cdot, u)$, gives

$$\sup_{y,z} \|J_{j,P}(y, z) - J_{j,Q}(y, z)\| \le C \|P - Q\|_{\mathrm{TV}}.$$

The difference between P and Q changes both the base point of the kernel derivative and its direction; the two changes are controlled respectively by the second- and first-derivative bounds. It follows that

$$\sup_y \left\| \int J_{j,P}(y,z)\,\mathrm{d}P(z) - \int J_{j,Q}(y,z)\,\mathrm{d}Q(z) \right\| \le C\|P-Q\|_{\mathrm{TV}}.$$

Combining these bounds with the population-2 representation

$$\phi^{(2)}_{C,j}(y;P) = H_{j,P}(y) - \int H_{j,P}(z)\,\mathrm{d}P(z) + \int J_{j,P}(y,z)\,\mathrm{d}P(z)$$

proves both assertions for $\ell = 2$. The population-1 expression is the negative of the displayed right-hand side, so the same norm bounds apply. □

Lemma C.2. *Suppose Assumption 3.1 and Assumptions 4.1– 4.5 hold. Then*

$$\widehat{\gamma}_j(\boldsymbol{x}_0) - \gamma_j(\boldsymbol{x}_0) = \mathbb{P}_{n,j}(\boldsymbol{x}_0) + \mathrm{Rem}_{j,n}(\boldsymbol{x}_0), \tag{36}$$

where

$$\|\,\mathrm{Rem}_{j,n}(\boldsymbol{x}_0)\|_{\mathcal{H}_j} = O_p\left(r_n^2 + r_n h_1^\nu + r_n h_2^\nu + \frac{r_n}{\sqrt{n_1 h_1^d}} + \frac{r_n}{\sqrt{n_2 h_2^d}} + \frac{1}{n_1 h_1^d} + \frac{1}{n_2 h_2^d} \right). \tag{37}$$

In particular,

$$a_n^{-1}\|\,\mathrm{Rem}_{j,n}(\boldsymbol{x}_0)\|_{\mathcal{H}_j} = o_p(1). \tag{38}$$

Moreover,

$$a_n^{-1}\mathbb{P}_{n,j}(\boldsymbol{x}_0) \rightsquigarrow \mathbb{G}_j(\boldsymbol{x}_0) \qquad \textit{in } \mathcal{H}_j. \tag{39}$$

Consequently,

$$a_n^{-1}\{\widehat{\gamma}_j(\boldsymbol{x}_0) - \gamma_j(\boldsymbol{x}_0)\} \rightsquigarrow \mathbb{G}_j(\boldsymbol{x}_0) \qquad \textit{in } \mathcal{H}_j. \tag{40}$$

Proof. Fix a fold pair (s,t). Theorem 3.1, with $\widehat{F}^{(1)}_{-s}$ and $\widehat{F}^{(2)}_{-t}$ as the two reference distributions, gives

$$\gamma_j(\boldsymbol{x}_0) = \gamma_j\left(\widehat{F}^{(1)}_{-s}, \widehat{F}^{(2)}_{-t}\right) + \mathbb{E}\left[\widehat{\phi}^{(1)}_{j,-s}(\boldsymbol{Y}^{(1)}) \mid \boldsymbol{X}^{(1)} = \boldsymbol{x}_0\right] + \mathbb{E}\left[\widehat{\phi}^{(2)}_{j,-t}(\boldsymbol{Y}^{(2)}) \mid \boldsymbol{X}^{(2)} = \boldsymbol{x}_0\right] + R_{j,st}. \tag{41}$$

where

$$\max_{s\in[K_1],\, t\in[K_2]} \|R_{j,st}\|_{\mathcal{H}_j} = O_p(r_n^2) \tag{42}$$

by Assumption 4.4 and (6). Combining this expansion with the fold-pair estimator in (7), and adding and subtracting the two true influence functions, and using the centering property in Theorem 3.1, yields

$$\begin{aligned} \widehat{\gamma}^{(s,t)}_j(\boldsymbol{x}_0) - \gamma_j(\boldsymbol{x}_0) = {} & \mathbb{P}^{h_1}_{1,s}\phi^{(1)}_j(\cdot;F^{(1)}) + \mathbb{P}^{h_2}_{2,t}\phi^{(2)}_j(\cdot;F^{(2)}) \\ & + \left[\mathbb{P}^{h_1}_{1,s} - \mathbb{E}\{\,\cdot \mid \boldsymbol{X}^{(1)} = \boldsymbol{x}_0\}\right]\left\{\widehat{\phi}^{(1)}_{j,-s} - \phi^{(1)}_j(\cdot;F^{(1)})\right\} \\ & + \left[\mathbb{P}^{h_2}_{2,t} - \mathbb{E}\{\,\cdot \mid \boldsymbol{X}^{(2)} = \boldsymbol{x}_0\}\right]\left\{\widehat{\phi}^{(2)}_{j,-t} - \phi^{(2)}_j(\cdot;F^{(2)})\right\} - R_{j,st}. \end{aligned} \tag{43}$$

It remains to control the terms arising from replacement of the true influence functions by their cross-fitted estimates. Lemma C.1 and Assumption 4.4 imply

$$\max_{\ell\in\{1,2\}} \max_{r\in[K_\ell]} \sup_y \left\| \widehat{\phi}^{(\ell)}_{j,-r}(y) - \phi^{(\ell)}_j(y;F^{(\ell)}) \right\|_{\mathcal{H}_j} = O_p(r_n). \tag{44}$$

Conditionally on the corresponding training sample, standard local-kernel empirical bounds give

$$\left\| (\mathbb{P}^{h_\ell}_{\ell,r} - \mathbb{P}^{h_\ell}_{\ell})\left\{\widehat{\phi}^{(\ell)}_{j,-r} - \phi^{(\ell)}_j(\cdot;F^{(\ell)})\right\} \right\|_{\mathcal{H}_j} = O_p\left(\frac{r_n}{\sqrt{n_\ell h_\ell^d}}\right), \tag{45}$$

where

$$\mathbb{P}_\ell^{h_\ell} f = \frac{\mathbb{E}[G_{h_\ell}(\boldsymbol{X}^{(\ell)} - \boldsymbol{x}_0) f(\boldsymbol{Y}^{(\ell)})]}{\mathbb{E}[G_{h_\ell}(\boldsymbol{X}^{(\ell)} - \boldsymbol{x}_0)]}.$$

Assumption 4.3, the order-ν kernel expansion, and (44) give

$$\left\| \mathbb{P}_\ell^{h_\ell} \left\{ \widehat{\phi}_{j,-r}^{(\ell)} - \phi_j^{(\ell)}(\cdot; F^{(\ell)}) \right\} - \mathbb{E}\Big[\left\{ \widehat{\phi}_{j,-r}^{(\ell)} - \phi_j^{(\ell)}(\cdot; F^{(\ell)}) \right\} (\boldsymbol{Y}^{(\ell)}) \,\Big|\, \boldsymbol{X}^{(\ell)} = \boldsymbol{x}_0 \Big] \right\|_{\mathcal{H}_j} = O_p(r_n h_\ell^\nu). \tag{46}$$

Combining (45) and (46) controls the fitted-influence-function terms in (43). We next replace the weighted average of the fold-specific leading terms by the corresponding full-sample local average. Since the number of folds is fixed and the true influence functions are uniformly bounded by Lemma C.1, the usual ratio expansion for normalized kernel averages gives

$$\left\| \sum_{r=1}^{K_\ell} \omega_{\ell r} \mathbb{P}_{\ell,r}^{h_\ell} \phi_j^{(\ell)}(\cdot; F^{(\ell)}) - \mathbb{P}_{\ell,n}^{h_\ell} \phi_j^{(\ell)}(\cdot; F^{(\ell)}) \right\|_{\mathcal{H}_j} = O_p\{(n_\ell h_\ell^d)^{-1}\}. \tag{47}$$

Averaging (43) over all fold pairs with weights $\omega_{1s}\omega_{2t}$, and using (42) and (44)–(47), yields (36)–(37). Under Assumption 4.5,

$$a_n^{-1} r_n^2 = o(1), \qquad a_n^{-1} r_n h_\ell^\nu = o(1), \qquad a_n^{-1} \frac{r_n}{\sqrt{n_\ell h_\ell^d}} = O(r_n) = o(1),$$

and $a_n^{-1}(n_\ell h_\ell^d)^{-1} = o(1)$. Hence (38) follows.

It remains to establish the limit of the leading term. Fix $\ell \in \{1, 2\}$. By Lemma C.1,

$$\sup_y \|\phi_j^{(\ell)}(y, \boldsymbol{x}_0)\|_{\mathcal{H}_j} < \infty.$$

The centering property in Theorem 3.1 and Assumption 4.3 imply that the bias of the normalized local average is $O(h_\ell^\nu)$, which is negligible after multiplication by $\sqrt{n_\ell h_\ell^d}$ under Assumption 4.5. The boundedness of the localization kernel and of the influence function verifies the Hilbert-space Lindeberg condition. A change of variables and dominated convergence give, for arbitrary $u, v \in \mathcal{H}_j$, the limiting covariance

$$\frac{R_G}{f_X^{(\ell)}(\boldsymbol{x}_0)} \mathbb{E}\Big[\langle \phi_j^{(\ell)}(\boldsymbol{Y}^{(\ell)}, \boldsymbol{x}_0), u \rangle \langle \phi_j^{(\ell)}(\boldsymbol{Y}^{(\ell)}, \boldsymbol{x}_0), v \rangle \mid \boldsymbol{X}^{(\ell)} = \boldsymbol{x}_0 \Big].$$

The same calculation for squared norms gives trace convergence. The triangular-array Lindeberg–Feller theorem in the separable Hilbert space $\mathcal{H}_j$ therefore yields

$$\sqrt{n_\ell h_\ell^d}\, \mathbb{P}_{\ell,n}^{h_\ell} \phi_j^{(\ell)}(\cdot, \boldsymbol{x}_0) \rightsquigarrow \mathbb{G}_{\ell,j}(\boldsymbol{x}_0). \tag{48}$$

The two group-specific limits are independent. Therefore,

$$a_n^{-1} \mathbb{P}_{n,j}(\boldsymbol{x}_0) = \left(\frac{n_2 h_2^d}{n_1 h_1^d + n_2 h_2^d} \right)^{1/2} \sqrt{n_1 h_1^d}\, \mathbb{P}_{1,n}^{h_1} \phi_j^{(1)} + \left(\frac{n_1 h_1^d}{n_1 h_1^d + n_2 h_2^d} \right)^{1/2} \sqrt{n_2 h_2^d}\, \mathbb{P}_{2,n}^{h_2} \phi_j^{(2)}.$$

Combining (48) with the stable sample-size ratio in Assumption 4.5 gives (39) with covariance operator (9). The claimed limit (40) then follows from (36), (38), and Slutsky's theorem. □

C.4 Proof of Theorem 4.1

Proof. The Hilbert-space convergence (40) is established in Lemma C.2.

Part (i): fixed alternative. By the identity $\widehat{Q}_j(\boldsymbol{x}_0) - Q_j(\boldsymbol{x}_0) = 2\langle \gamma_j(\boldsymbol{x}_0), \widehat{\gamma}_j(\boldsymbol{x}_0) - \gamma_j(\boldsymbol{x}_0) \rangle_{\mathcal{H}_j} + \|\widehat{\gamma}_j(\boldsymbol{x}_0) - \gamma_j(\boldsymbol{x}_0)\|_{\mathcal{H}_j}^2$, we have

$$\begin{aligned} a_n^{-1}\{\widehat{Q}_j(\boldsymbol{x}_0) - Q_j(\boldsymbol{x}_0)\} &= 2\left\langle \gamma_j(\boldsymbol{x}_0), a_n^{-1}\{\widehat{\gamma}_j(\boldsymbol{x}_0) - \gamma_j(\boldsymbol{x}_0)\} \right\rangle_{\mathcal{H}_j} \\ &\quad + a_n^{-1} \|\widehat{\gamma}_j(\boldsymbol{x}_0) - \gamma_j(\boldsymbol{x}_0)\|_{\mathcal{H}_j}^2. \end{aligned}$$

(40) implies $\widehat{\gamma}_j(\boldsymbol{x}_0) - \gamma_j(\boldsymbol{x}_0) = O_p(a_n)$, so the second term is $O_p(a_n) = o_p(1)$. The continuous-mapping theorem applied to the first term gives

$$a_n^{-1}\{\widehat{Q}_j(\boldsymbol{x}_0) - Q_j(\boldsymbol{x}_0)\} \rightsquigarrow 2\langle \gamma_j(\boldsymbol{x}_0), \mathbb{G}_j(\boldsymbol{x}_0)\rangle_{\mathcal{H}_j}.$$

The limit is mean-zero Gaussian with variance $4\langle V_j(\boldsymbol{x}_0)\gamma_j(\boldsymbol{x}_0), \gamma_j(\boldsymbol{x}_0)\rangle_{\mathcal{H}_j}$, proving (10).

Part (ii): coordinate null. Under $H_{0,j}(\boldsymbol{x}_0)$, Proposition 2.1 gives $\gamma_j(\boldsymbol{x}_0) = 0$. Hence

$$a_n^{-2}\widehat{Q}_j(\boldsymbol{x}_0) = \|a_n^{-1}\widehat{\gamma}_j(\boldsymbol{x}_0)\|^2_{\mathcal{H}_j}.$$

The continuous-mapping theorem and (40) give (11). Lemma C.1 implies $\mathrm{tr}\{V_j(\boldsymbol{x}_0)\} < \infty$, so $V_j(\boldsymbol{x}_0)$ is positive, self-adjoint, and trace class on $\mathcal{H}_j$. Its Karhunen–Loève representation yields (12). □

C.5 Proof of Theorem 4.2

Proof. For brevity, write $\phi_{j,i}^{(\ell)} = \phi_j^{(\ell)}(\boldsymbol{Y}_i^{(\ell)}, \boldsymbol{x}_0)$, $\widehat{\phi}_{j,i}^{(\ell)} = \widehat{\phi}_{j,i}^{(\ell)}(\boldsymbol{x}_0)$, and $\delta\phi_{j,i}^{(\ell)} = \widehat{\phi}_{j,i}^{(\ell)} - \phi_{j,i}^{(\ell)}$.

Step 1: replacement of the estimated influence functions. Lemma C.1, Assumption 4.4, and the fixed number of folds imply

$$\max_{\ell\in\{1,2\}} \max_{1\le i\le n_\ell} \|\delta\phi_{j,i}^{(\ell)}\|_{\mathcal{H}_j} = O_p(r_n) = o_p(1). \tag{49}$$

Define the oracle group process

$$\mathbb{G}^{*,0}_{\ell n,j}(\boldsymbol{x}_0) = \frac{h_\ell^{d/2}}{\sqrt{n_\ell}\,\widehat{f}_{\ell,n}(\boldsymbol{x}_0)} \sum_{i=1}^{n_\ell} (\xi_i^{(\ell)} - \bar{\xi}_\ell) G_{h_\ell}(\boldsymbol{X}_i^{(\ell)} - \boldsymbol{x}_0)\phi_{j,i}^{(\ell)}. \tag{50}$$

Conditionally on the data, the centered multipliers are Gaussian, so

$$\mathbb{E}^*\left\|\widehat{\mathbb{G}}^*_{\ell n,j}(\boldsymbol{x}_0) - \mathbb{G}^{*,0}_{\ell n,j}(\boldsymbol{x}_0)\right\|^2_{\mathcal{H}_j} \le \frac{h_\ell^d}{n_\ell \widehat{f}_{\ell,n}(\boldsymbol{x}_0)^2} \sum_{i=1}^{n_\ell} G_{h_\ell}(\boldsymbol{X}_i^{(\ell)} - \boldsymbol{x}_0)^2 \|\delta\phi_{j,i}^{(\ell)}\|^2_{\mathcal{H}_j}.$$

The kernel density denominator converges to $f_X^{(\ell)}(\boldsymbol{x}_0) > 0$, while $h_\ell^d n_\ell^{-1}\sum_i G_{h_\ell}^2 = O_p(1)$. Together with (49), this gives

$$\mathbb{E}^*\left\|\widehat{\mathbb{G}}^*_{\ell n,j}(\boldsymbol{x}_0) - \mathbb{G}^{*,0}_{\ell n,j}(\boldsymbol{x}_0)\right\|^2_{\mathcal{H}_j} = o_p(1). \tag{51}$$

Step 2: conditional Gaussian limit of the oracle process. Conditionally on the data, (50) is a mean-zero Gaussian element in $\mathcal{H}_j$. Its covariance operator equals

$$\widehat{V}^0_{\ell n,j} = \frac{h_\ell^d}{n_\ell \widehat{f}_{\ell,n}(\boldsymbol{x}_0)^2} \sum_{i=1}^{n_\ell} a_{\ell i,n} \otimes_{\mathrm{op}} a_{\ell i,n} - \frac{h_\ell^d}{\widehat{f}_{\ell,n}(\boldsymbol{x}_0)^2} \bar{a}_{\ell n} \otimes_{\mathrm{op}} \bar{a}_{\ell n},$$

where

$$a_{\ell i,n} = G_{h_\ell}(\boldsymbol{X}_i^{(\ell)} - \boldsymbol{x}_0)\phi_{j,i}^{(\ell)}, \qquad \bar{a}_{\ell n} = n_\ell^{-1}\sum_{i=1}^{n_\ell} a_{\ell i,n}.$$

For every $u, v \in \mathcal{H}_j$, a change of variables, dominated convergence, and the boundedness from Lemma C.1 imply convergence in probability of the corresponding bilinear forms to

$$\frac{R_G}{f_X^{(\ell)}(\boldsymbol{x}_0)} \mathbb{E}\Big[\langle \phi_j^{(\ell)}(\boldsymbol{Y}^{(\ell)}, \boldsymbol{x}_0), u\rangle \langle \phi_j^{(\ell)}(\boldsymbol{Y}^{(\ell)}, \boldsymbol{x}_0), v\rangle \mid \boldsymbol{X}^{(\ell)} = \boldsymbol{x}_0\Big].$$

The same argument applied to squared norms gives trace convergence. The centering and bias bounds from Lemma C.2 imply that the second operator above is negligible in trace norm. Projection onto the eigenbasis of the limiting trace-class operator then gives conditional tightness. Consequently, each oracle group process converges conditionally to the same Gaussian limit $\mathbb{G}_{\ell,j}(\boldsymbol{x}_0)$ as in (48).

The group multipliers are independent. Combining the two oracle group processes with the weights in (13) therefore gives, for $\mathrm{BL}_1(\mathcal{H}_j)$ denoting the unit bounded-Lipschitz class,

$$\sup_{f\in \mathrm{BL}_1(\mathcal{H}_j)} \left| \mathbb{E}^* f\{\mathbb{G}^{*,0}_{j,n}(\boldsymbol{x}_0)\} - \mathbb{E} f\{\mathbb{G}_j(\boldsymbol{x}_0)\} \right| \xrightarrow{p} 0. \tag{52}$$

Step 3: feasible bootstrap process. (51), conditional Cauchy–Schwarz, and the bounded combination weights imply that replacing the oracle influence functions by the cross-fitted estimated influence functions changes the combined process by $o_p(1)$ in conditional mean norm. Together with (52), this proves conditional weak convergence of $\widehat{\mathbb{G}}^*_{j,n}(\boldsymbol{x}_0)$ to $\mathbb{G}_j(\boldsymbol{x}_0)$. The conditional continuous-mapping theorem then yields

$$T^*_{j,n}(\boldsymbol{x}_0) = \|\widehat{\mathbb{G}}^*_{j,n}(\boldsymbol{x}_0)\|^2_{\mathcal{H}_j} \rightsquigarrow \|\mathbb{G}_j(\boldsymbol{x}_0)\|^2_{\mathcal{H}_j}$$

conditionally on the data, in probability. If $\mathrm{tr}\{V_j(\boldsymbol{x}_0)\} > 0$, the weighted chi-square limit has a continuous distribution function. Conditional weak convergence therefore implies the uniform cdf convergence in (14) and consistency of the conditional quantile.

Step 4: size and power. Under $H_{0,j}(\boldsymbol{x}_0)$, Theorem 4.1(ii) and quantile consistency give

$$\mathbb{P}\left\{ a_n^{-2}\widehat{Q}_j(\boldsymbol{x}_0) > c^*_{j,1-\alpha}(\boldsymbol{x}_0) \right\} \longrightarrow \alpha.$$

Under any fixed alternative, Theorem 4.1(i) gives $\widehat{Q}_j(\boldsymbol{x}_0) = Q_j(\boldsymbol{x}_0) + O_p(a_n) \xrightarrow{p} Q_j(\boldsymbol{x}_0) > 0$, so $a_n^{-2}\widehat{Q}_j(\boldsymbol{x}_0) \xrightarrow{p} \infty$. Boundedness of the estimated influence functions and the conditional second-moment bound for the multiplier process imply $c^*_{j,1-\alpha}(\boldsymbol{x}_0) = O_p(1)$, proving consistency. □

D Additional Experiments and Experimental Details

This appendix collects all supplementary experiments and experimental details. Section D.1 provides the common setup and full specifications for the simulation studies in the main text. Sections D.2–D.3 present two additional experiments, respectively distinguishing coordinate contribution from predictive variable importance and studying coordinate–profile localization in an Airfoil-based semi-synthetic setting. Section D.4 reports first-stage robustness, a plug-in versus one-step estimation comparison, bandwidth sensitivity, and numerical implementation details. Section D.5 provides supplementary results and implementation details for the LLM and ImageNet-9 applications in Section 5.

D.1 Simulation Details for the Main-Text Experiments

Common setup. Unless stated otherwise, the synthetic experiments use $d = 1$, $p = 5$, and

$$X^{(1)}, X^{(2)} \overset{\text{iid}}{\sim} \mathrm{Unif}[-1, 1].$$

The target coordinate is $j = 1$. We use Gaussian kernels for the two response components:

$$k_F(s,t) = \exp\left\{ -\frac{(s-t)^2}{2\sigma_F^2} \right\}, \qquad s,t \in \mathbb{R},$$

and

$$k_C(\boldsymbol{u},\boldsymbol{v}) = \exp\left\{ -\frac{\|\boldsymbol{u}-\boldsymbol{v}\|_2^2}{2\sigma_C^2} \right\}, \qquad \boldsymbol{u},\boldsymbol{v} \in \mathbb{R}^p.$$

For the reported synthetic experiments, we set $\sigma_F = 1$ and $\sigma_C = 0.6$.

The two conditional response distributions are estimated independently using two-fold cross-fitting. Our default first-stage estimator is a conditional RealNVP normalizing flow (Papamakarios et al., 2021). In implementation, expectations under the fitted conditional distributions appearing in the one-step estimator are evaluated by Monte Carlo sampling from the fitted conditional generators. Unless otherwise stated, we use $M = 512$ conditional draws; the dependence-shift and bandwidth-sensitivity experiments use $M = 1024$. For localization in the covariate, we use the Epanechnikov kernel

$$g(u) = \frac{3}{4}(1-u^2)\mathbf{1}\{|u| \le 1\}$$

and bandwidth

$$h_\ell = c_h n_\ell^{-1/4}, \qquad \ell = 1, 2.$$

We use $c_h = 1$ in the reported synthetic experiments. All coordinate tests use nominal level $\alpha = 0.05$. The RKHS inner products are evaluated using Gaussian-kernel Gram matrices.

Size, power, and localization. For Figure 2, we generate

$$Y^{(1)} \mid X = x \sim N\{m(x), \Sigma\}, \qquad Y^{(2)} \mid X = x \sim N\{m(x) + \Delta(x)e_1, \Sigma\},$$

where $e_1 = (1, 0, \ldots, 0)^\top$, $m(x) = \big(0.4\sin(\pi x),\, 0.3x,\, 0.25\cos(\pi x),\, 0.2(x^2 - 1/3),\, -0.2x\big)^\top$ and $\Sigma = (\Sigma_{rs})_{1 \le r,s \le 5}$ with $\Sigma_{rs} = 0.4^{|r-s|}$.

We consider an exact null $\Delta(x) = 0$, a global alternative $\Delta(x) = 0.45$, and a local alternative $\Delta(x) = 0.75 b_{0.3}(x)$. The smooth bump is defined by

$$b_w(x) = \begin{cases} \exp\left\{1 - \dfrac{1}{1 - (x/w)^2}\right\}, & |x| < w, \\ 0, & |x| \ge w. \end{cases}$$

Thus $Q_1(x) = 0$ everywhere under the exact null, $Q_1(x) > 0$ everywhere under the global alternative, and $Q_1(x) > 0 \iff |x| < 0.3$ under the local alternative. We use $n_1 = n_2 \in \{200, 500, 1000\}$.

The dense curves in Figure 2 use 300 Monte Carlo replications and 499 multiplier-bootstrap draws. The higher-precision pointwise evaluation in Table 3 uses 1000 replications and 999 bootstrap draws.

Table 3: Empirical size and power for the Gaussian location-shift experiment.

Setting	x_0	$n = 200$	$n = 500$	$n = 1000$
Exact null	0	0.058	0.044	0.053
Local alternative, outside support	0.65	0.040	0.036	0.040
Global alternative	0	0.426	0.723	0.921
Local alternative, center	0	0.660	0.961	0.999

Dependence-mediated coordinate changes. For Figure 3, we generate

$$Y^{(\ell)} \mid X = x \sim N\{0, \Sigma(\rho_\ell(x))\},$$

where

$$\Sigma(r) = \begin{pmatrix} 1 & r & r & r & r \\ r & 1 & 0 & 0 & 0 \\ r & 0 & 1 & 0 & 0 \\ r & 0 & 0 & 1 & 0 \\ r & 0 & 0 & 0 & 1 \end{pmatrix},$$

and

$$\rho_1(x) = \frac{\Delta_\rho}{2} a(x), \qquad \rho_2(x) = -\frac{\Delta_\rho}{2} a(x), \qquad a(x) = \exp\left\{-\frac{x^2}{2(0.3)^2}\right\}.$$

Every coordinate therefore has conditional marginal distribution $N(0, 1)$ in both groups, while only the associations between Y_1 and Y_{-1} change across groups; the coordinates in Y_{-1} remain mutually independent. For the location curve, we use $n_1 = n_2 \in \{500, 1000\}$, $\Delta_\rho = 0.8$, 300 Monte Carlo replications, 499 bootstrap draws, and 21 evaluation points on $[-0.7, 0.7]$. For the signal curve at $x_0 = 0$, we vary

$$\Delta_\rho \in \{0, 0.2, 0.4, 0.6, 0.8\}$$

and, for each sample size $n_1 = n_2 \in \{500, 1000\}$, use 500 Monte Carlo replications and 999 bootstrap draws. Marginal Local-CMMD applies the local conditional two-sample procedure of Yan et al. (2022) only to the scalar response Y_1. It therefore tests a different null from the proposed coordinate test.

D.2 Coordinate Contribution versus Predictive Importance

To distinguish coordinate contribution from predictive variable importance, we consider

$$Z^{(1)} \sim \mathcal{N}(0,1), \qquad Z^{(2)} \sim \mathcal{N}(\delta,1),$$

and, for each group $\ell \in \{1,2\}$, define

$$Y_1^{(\ell)} = Z^{(\ell)}, \qquad Y_k^{(\ell)} = \rho Z^{(\ell)} + \sqrt{1-\rho^2}\,\varepsilon_k^{(\ell)}, \quad k = 2,\dots,5,$$

where the $\varepsilon_k^{(\ell)}$ are independent standard normal variables. We set $\delta = 0.7$, $n_1 = n_2 = 500$, and vary $\rho \in \{0, 0.25, 0.5, 0.75, 0.9, 0.97, 0.99\}$. For every $\rho < 1$, $F_1^{(1)} \neq F_1^{(2)}$, so $Q_1(0) > 0$. Because the marginal discrepancy in Y_1 is fixed and the dependence structure is common across groups, $Q_1(0)$ is constant in ρ. As ρ increases, however, Y_1 becomes increasingly redundant with Y_{-1}, and its leave-one-coordinate-out predictive importance decreases.

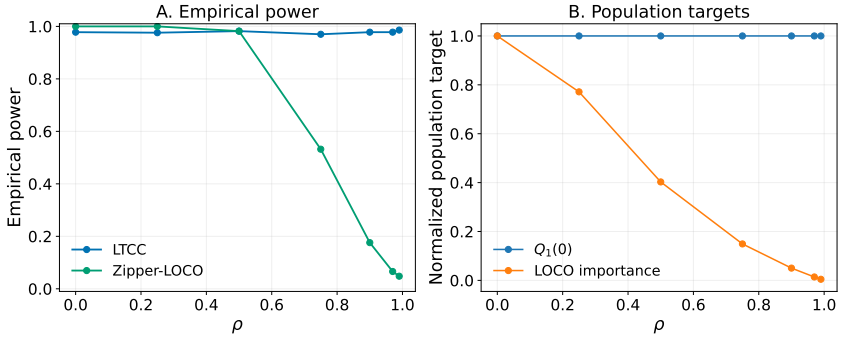

Figure 4: **Coordinate contribution versus predictive variable importance.** A: Empirical power of LTCC and Zipper leave-one-coordinate-out (LOCO) against their respective alternatives. B: Corresponding population targets, each normalized by its own maximum over ρ. As ρ increases, $Q_1(0)$ remains constant, while the LOCO predictive-importance target decreases as Y_1 becomes more predictable from Y_{-1}.

Figure 4 compares LTCC with Zipper-LOCO as the redundancy between Y_1 and Y_{-1} increases. LTCC maintains high power throughout the examined range of ρ, whereas the power of Zipper-LOCO decreases as ρ approaches one. The corresponding population targets exhibit the same contrast. The marginal discrepancy in Y_1 is fixed across ρ, and the marginal-standardized association structure is preserved between the two groups; consequently, $Q_1(0)$ remains constant. In contrast, the LOCO target decreases with ρ, since Y_{-1} becomes increasingly informative about Y_1 and the additional predictive value of including Y_1 diminishes.

Thus, LTCC continues to identify Y_1 as a contributing coordinate even when its incremental predictive importance becomes very small. At $\rho = 0.99$, the normalized LOCO target is below 1% of its maximum, whereas the normalized LTCC target remains equal to one. The two population targets are normalized separately over the examined ρ-grid.

D.3 Airfoil Semi-Synthetic Experiment

We use the UCI Airfoil Self-Noise data (Brooks et al., 1989) to construct a semi-synthetic experiment for evaluating coordinate and profile localization under a data-driven conditional response distribution. The dataset contains $1{,}503$ observations of six measured quantities. We take the scaled sound pressure level as the one-dimensional conditioning variable X and define the five-dimensional response as

$$Y = \big(\log(\text{frequency}),\ \text{angle of attack},\ \text{chord length},\ \text{free-stream velocity},\ \log(\text{suction-side displacement thickness})\big).$$

After the two log transformations, X and the five response coordinates are standardized using the full observed dataset. We regard the standardized observations as a sample from an underlying joint

distribution of (X, Y) and use them to construct data-driven approximations to the corresponding conditional response distributions.

Let $\{(x_i, y_i) : i = 1, \ldots, N_{\mathrm{AF}}\}$, with $N_{\mathrm{AF}} = 1503$, denote the standardized observations. We use local kernel weighting of these observations to obtain an empirical approximation to the underlying conditional response distribution at $X = x$. Specifically, with φ_b denoting the density of $N(0, b^2)$ and $b = 0.20$, observation i receives conditional sampling probability

$$w_i(x) = \frac{\varphi_b(x - x_i)}{\sum_{m=1}^{N_{\mathrm{AF}}} \varphi_b(x - x_m)}.$$

Thus, observations with covariate values closer to x receive greater weight in the empirical approximation to the conditional response distribution.

Population 1 draws X from the Gaussian-kernel-smoothed empirical distribution of the observed x_i's; population 2 draws X from the exponentially tilted version with tilt proportional to $e^{0.50x}$. These continuous X-marginals differ across populations. In each Monte Carlo replication, we generate independent samples with $n_1 = n_2 = 500$. We use 100 Monte Carlo replications for each size, power, and localization setting. For LTCC inference, the conditional response distributions are estimated using a conditional kernel density estimator under two-fold cross-fitting.

For the size experiment, the two populations are constructed to share the same underlying conditional response distribution while having different marginal distributions of X. We use the locally weighted Airfoil sample as an empirical approximation to this common distribution. In implementation, given $X = x$, a donor observation is sampled according to $w_i(x)$, and its complete five-dimensional response vector is retained so that the empirical approximation preserves the observed cross-coordinate dependence. By construction, the population-level conditional distributions satisfy

$$F^{(1)}_{Y|X}(\cdot \mid x) = F^{(2)}_{Y|X}(\cdot \mid x)$$

for every x. Hence, the coordinate null holds for every response coordinate and evaluation profile. Across the three conditioning profiles and five response coordinates, empirical size ranges from 0.01 to 0.09.

For the power and localization experiments, we construct a controlled coordinate-specific alternative using data-driven approximations to the underlying Airfoil conditional marginals. The five coordinates are generated independently from their locally estimated conditional marginals, and both populations share the same baseline conditional construction before the signal is introduced. The resulting alternative introduces coordinate-specific marginal changes and complements the dependence-shift experiment in Section 5.2.

Let $x_\star$ denote the empirical median of the standardized conditioning variable and define

$$\beta_{\mathrm{AF}}(x) = b_{0.35}(x - x_\star),$$

where b_w is the smooth bump defined in Section D.1. Population 1 is left unchanged, whereas in population 2 we set

$$Y_r^{(2)} = Y_r^{(2),0} + c\,\beta_{\mathrm{AF}}(X^{(2)})\mathbf{1}\{r \in \{2, 4\}\}, \qquad r = 1, \ldots, 5.$$

Thus, the alternative changes only the conditional marginals of Y_2 (angle of attack) and Y_4 (free-stream velocity), and only within the support of β_{AF}. The dependence structure is identical across the two populations by construction. Hence, for $c > 0$, the coordinate null holds for Y_1, Y_3, Y_5 throughout, while it fails for Y_2, Y_4 wherever $\beta_{\mathrm{AF}}(x) > 0$.

Empirical power increases with signal strength for both active coordinates. At $c = 0.50$, the rejection probabilities for Y_2 and Y_4 are 0.66 and 0.39, respectively; at $c = 0.75$, they increase to 0.92 and 0.86. At $c = 1$, they reach 1.00 and 0.98.

We further evaluate localization over the conditioning profile by fixing $c = 0.75$ and testing at 17 target points corresponding to empirical quantiles

$$q = 0.10, 0.15, \ldots, 0.90$$

of the sound-pressure distribution. Because β_{AF} is compactly supported around $x_\star$, the coordinate discrepancy for Y_2 and Y_4 is localized around the median conditioning profile. Figure 5 shows

the injected local signal and the resulting pointwise size and power. At the median sound-pressure profile ($x_0 = 125.721$ on the original scale), empirical power is 0.93 and 0.88 for the two active coordinates Y_2 and Y_4, respectively, while the empirical sizes for the inactive coordinates Y_1, Y_3, Y_5 are 0.02, 0.07, and 0.03. Outside the support of the perturbation, all five coordinate nulls hold by construction.

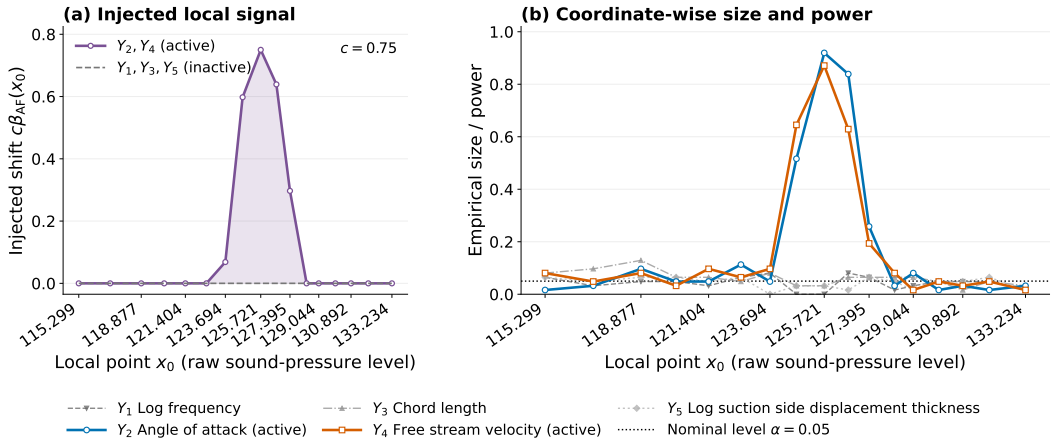

Figure 5: **Coordinate and profile localization in the Airfoil semi-synthetic experiment. A:** Injected local marginal shift with signal strength $c = 0.75$, acting on Y_2 (angle of attack) and Y_4 (free-stream velocity). **B:** Pointwise empirical power for the two perturbed coordinates within the signal region and empirical size at coordinates and profiles satisfying the coordinate null. Results are based on 100 Monte Carlo replications.

D.4 Robustness, Sensitivity, and Implementation Details

Sensitivity to conditional-distribution estimation. We assess sensitivity to the first-stage conditional-distribution estimator under a model with nonlinear location, heteroscedastic scale, and non-Gaussian innovations. We generate

$$X^{(1)}, X^{(2)} \overset{\text{iid}}{\sim} \text{Unif}[-1, 1].$$

For $r = 1, \ldots, 5$, let

$$m_r(x) = a_r \sin(\pi x) + b_r x, \qquad s_r(x) = 0.75 + 0.20 \sin^2\left(\pi x + \frac{(r-1)\pi}{5}\right),$$

where

$$a = (0.8, -0.7, 0.6, -0.5, 0.4)^\top, \qquad b = (0.30, -0.25, 0.20, 0.15, -0.10)^\top.$$

We further define

$$\pi_0(x) = 0.20 + 0.05 \sin(\pi x),$$

and use the standardized non-Gaussian innovation

$$\varepsilon(q) = \frac{\{B - q\}/\sqrt{q(1-q)} + \tau Z}{\sqrt{1 + \tau^2}}, \qquad B \sim \text{Bernoulli}(q), \quad Z \sim N(0,1), \quad \tau = 0.25,$$

with B and Z independent. This innovation has mean zero and variance one for every q, so varying q changes its distributional shape without changing its first two moments.

For population 1, all coordinates use $q = 0.50$. For population 2, let

$$\pi_r^{(2)}(x; \delta) = \begin{cases} 0.50 + \delta\{\pi_0(x) - 0.50\}, & r = 1, \\ \pi_0(x), & r = 2, \ldots, 5, \end{cases}$$

and generate, conditional on $X = x$,

$$Y_r^{(1)} = m_r(x) + s_r(x)\varepsilon_r^{(1)}(0.50),$$
$$Y_r^{(2)} = m_r(x) + \kappa x^2 + s_r(x)\varepsilon_r^{(2)}\left(\pi_r^{(2)}(x;\delta)\right), \qquad \kappa = 2.5.$$

Innovations are independent across coordinates and observations.

We evaluate coordinate 1 at $x_0 = 0$. Since $\pi_0(0) = 0.20$, when $\delta = 0$ the target-coordinate innovation probability is 0.50 in both populations, and the additional location term κx^2 vanishes at $x_0 = 0$. Hence the conditional marginals of Y_1 coincide at the target profile. Conditional independence also gives

$$C_1^{(1)} = C_1^{(2)} = 0,$$

so $Q_1(0) = 0$. At the same time, coordinates $2, \ldots, 5$ have different conditional distributions across the two populations. Thus, $\delta = 0$ defines an exact coordinate null without imposing a joint conditional null. For $\delta > 0$,

$$\pi_1^{(2)}(0;\delta) = 0.50 - 0.30\delta,$$

which introduces a shape change in the conditional distribution of Y_1 while preserving the first two innovation moments.

We use $n_1 = n_2 = 1000$, 300 Monte Carlo replications, and

$$\delta \in \{0, 0.2, 0.35, 0.5, 0.65, 0.8\}.$$

We compare the oracle conditional sampler with a linear-Gaussian model, a kernel-based conditional distribution estimator, a conditional normalizing flow (CNF) (Papamakarios et al., 2021), a generative moment matching network (GMMN) (Ren et al., 2016), and Engression (Shen & Meinshausen, 2025). All estimators use the same data and cross-fitting splits within each replication, with matched random-number streams for conditional sampling and multiplier-bootstrap draws.

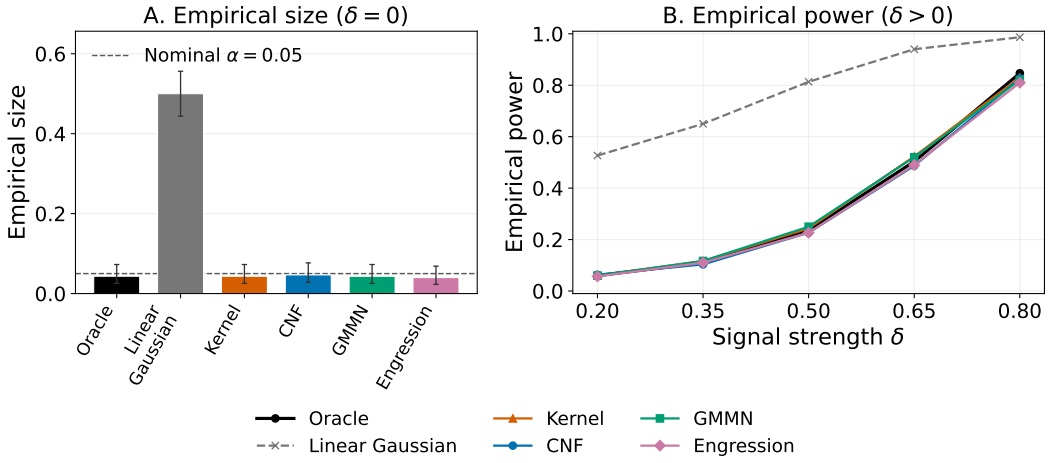

Figure 6: **Sensitivity to first-stage conditional-distribution estimation.** Empirical size at the coordinate null $\delta = 0$ and empirical power under the shape alternatives $\delta > 0$ for a nonlinear, heteroscedastic, and non-Gaussian conditional response model. The horizontal dashed line denotes the nominal level $\alpha = 0.05$.

Figure 6 shows that the oracle and flexible conditional-distribution estimators attain empirical size close to the nominal level at $\delta = 0$, with power increasing as the shape difference grows. In contrast, the misspecified linear-Gaussian first stage exhibits substantial size distortion under the nonlinear, heteroscedastic, and non-Gaussian conditional model. This result highlights the importance of adequately estimating the conditional response distribution for calibration. The flexible estimators remain substantially closer to the oracle benchmark while retaining increasing power as the signal strengthens.

Plug-in versus one-step estimation. We further examine the finite-sample effect of the influence-function correction by comparing the cross-fitted one-step estimator $\widehat{Q}_1(x_0)$ with the corresponding plug-in estimator. The two estimators use the same fitted conditional distributions, cross-fitting splits, and conditional draws.

We use a sparse five-covariate nonlinear design. The coordinates of $X \in [-1,1]^5$ are independent uniform variables, but only X_1 affects the conditional response distribution. All five covariates are supplied to the conditional RealNVP estimator, while localization is performed on X_1. Conditional on $X_1 = x$, let

$$Z^{(\ell)} \sim N_5\{0, \Sigma_\ell(x)\},$$

where the diagonal entries are one,

$$\operatorname{corr}(Z_1, Z_3) = 0.25,$$

and

$$\operatorname{corr}(Z_1, Z_2) = \begin{cases} 0.20 + 0.10\sin(\pi x), & \ell = 1, \\ 0.45 + 0.10\sin(\pi x), & \ell = 2, \end{cases}$$

with all remaining off-diagonal correlations equal to zero. For $r = 1, \ldots, 5$, define

$$m_r(x) = 0.35\sin\left(\pi x + \frac{(r-1)\pi}{5}\right) + 0.15(r-3)x,$$

and

$$s_r(x) = 0.9 + 0.08r + 0.10\cos\left(\pi x + \frac{(r-1)\pi}{5}\right).$$

Let

$$d = (0, 0.35, -0.30, 0.25, -0.20)^\top$$

and

$$\Delta(x) = 1.2\exp\left\{-\frac{1}{2}\left(\frac{x-0.5}{0.45}\right)^2\right\}.$$

The responses are generated as

$$Y_r^{(1)} = m_r(x) + s_r(x)\sinh(0.7Z_r^{(1)}),$$

and

$$Y_r^{(2)} = m_r(x) + d_r + s_r(x)\sinh(0.7Z_r^{(2)}) + \Delta(x)\mathbf{1}\{r = 1\}.$$

Thus, the design combines nonlinear conditional location and scale, a non-Gaussian response transformation, nuisance-coordinate shifts, and a localized marginal shift in the target coordinate.

We evaluate $Q_1(x_0)$ at $x_0 = (0.5, 0, 0, 0, 0)$; because the true conditional response distribution depends only on X_1, the population target is determined by $X_1 = 0.5$. The population target is evaluated independently using numerical integration, giving $Q_1(x_0) = 0.3752$. We use $n_1 = n_2 \in \{200, 500, 1000\}$, 500 paired Monte Carlo replications, two-fold cross-fitting, and $M = 512$ conditional draws. Estimation accuracy is measured by mean absolute error (MAE) and root mean squared error (RMSE) relative to the population target.

Table 4 shows that the one-step estimator has substantially lower MAE and RMSE than the plug-in estimator at all three sample sizes in this setting, with its estimation error decreasing as the sample size increases.

Table 4: **Plug-in versus one-step estimation accuracy.** MAE and RMSE for estimating $Q_1(x_0)$, based on 500 paired Monte Carlo replications and an independently evaluated population target.

$n_1 = n_2$	Plug-in MAE	One-step MAE	Plug-in RMSE	One-step RMSE
200	0.288	0.079	0.290	0.101
500	0.277	0.053	0.278	0.067
1000	0.245	0.042	0.247	0.053

Bandwidth sensitivity. We examine sensitivity to the localization bandwidth multiplier c_h and the association-kernel bandwidth σ_C, while fixing $\sigma_F = 1$. We consider

$$c_h \in \{0.8, 1.0, 1.2\}, \qquad \sigma_C \in \{0.4, 0.6, 0.8\}.$$

The experiment uses the dependence-shift design of Section D.1 at $x_0 = 0$, with $\Delta_\rho = 0$ under the exact null and $\Delta_\rho = 0.6$ under the fixed alternative. The main synthetic specification corresponds to

$$(c_h, \sigma_C) = (1.0, 0.6).$$

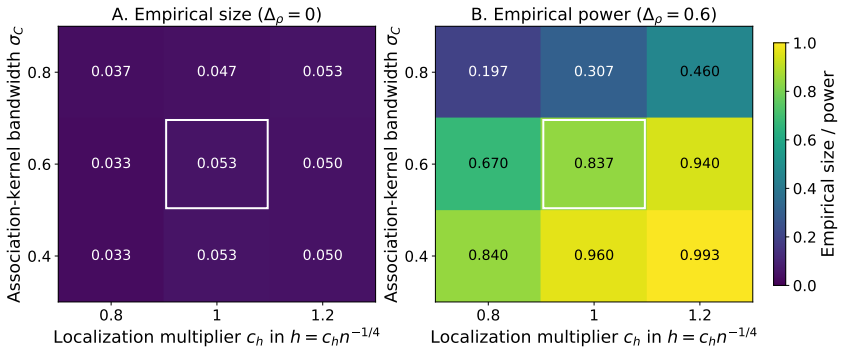

Figure 7: **Bandwidth sensitivity.** Empirical size under the exact null and empirical power under the fixed alternative across localization bandwidth multipliers and association-kernel bandwidths.

Implementation details. All first-stage estimators are fitted separately by population and cross-fitting fold. The oracle method samples from the true conditional distribution. The linear-Gaussian estimator uses a linear conditional mean and a constant residual covariance, while the kernel-based estimator constructs fitted conditional response distributions using local kernel weighting. CNF, GMMN, and Engression are fitted as conditional generators. Training configurations are held fixed across signal levels.

For each population, cross-fitting fold, and evaluation profile x_0, we draw M responses once from the fitted conditional distribution. These draws are used to approximate the fitted marginal CDFs, construct the fitted PITs, and compute the Monte Carlo terms entering the plug-in and influence-function corrections. The marginal CDFs entering the PIT are approximated by right-continuous empirical CDFs based on the M conditional Monte Carlo draws. For the Gaussian association kernel, the integral with respect to the uniform reference coordinate in (20) and the corresponding kernel derivatives are evaluated analytically. The multiplier bootstrap conditions on the fitted conditional draws and resamples the observation-level Gaussian multipliers. The asymptotic results in Section 4 are stated for the corresponding estimator with expectations under the fitted conditional distributions evaluated exactly. In the experiments, these expectations are approximated by Monte Carlo averages using the M conditional draws described above.

RKHS quadratic forms are evaluated from symmetrized Gram matrices. Random seeds for data generation, cross-fitting, first-stage fitting, conditional sampling, and multiplier-bootstrap draws are generated from independent child streams. Simulation summaries and figures are computed from saved replication-level results.

D.5 Additional Details for the Real-Data Applications

LLM response audit. We use frozen response banks from Qwen2.5-7B-Instruct on MMLU-Pro questions in Math, Physics, Chemistry, Engineering, and Biology. Qwen2.5-1.5B-Instruct serves as the auxiliary reference model used to construct the conditioning profile and the advice. We treat the observed responses as samples from underlying subject-specific response distributions and use the frozen banks to construct empirical conditional approximations. For a question q in subject j, let m_{qj} denote the reference-model margin, defined as the score of the correct answer minus the largest

score among the incorrect answers. We convert these margins to within-subject rank percentiles

$$u_{qj} = \frac{r_{qj} - 1/2}{N_j},$$

where r_{qj} is the ascending rank of m_{qj} and N_j is the number of questions in subject j.

The conditioning profile is drawn as

$$X \sim \mathrm{Unif}[0.1, 0.9].$$

In implementation, conditional on $X = x$, we approximate the subject-specific conditional question distribution by independently resampling from the observed benchmark questions. For question q in subject j, the empirical resampling weight is

$$\widehat{w}_{qj}(x) = \frac{\exp\{-(u_{qj} - x)^2/(2\tau_X^2)\}}{\sum_{m=1}^{N_j} \exp\{-(u_{mj} - x)^2/(2\tau_X^2)\}}, \qquad \tau_X = 0.10.$$

A question is drawn from the observed bank according to $\widehat{w}_{qj}(x)$. Hence, X indexes the local question-difficulty profile, with smaller values placing greater weight on questions having smaller reference-model margins. Larger values of X place greater weight on questions with larger reference-model margins and therefore correspond to easier question profiles.

For a selected question, let s_a denote the audited model's stored score for answer option a, let $a^\star$ be the correct option, and let $\mathcal{A}$ be the set of valid options. The response score is

$$Y_j = s_{a^\star} - \log\left(\sum_{a \in \mathcal{A} \setminus \{a^\star\}} e^{s_a}\right).$$

The five-dimensional response vector combines one subject-specific score from each of Math, Physics, Chemistry, Engineering, and Biology, yielding a five-channel representation for conditional auditing. A common coordinatewise standardization is fixed from the frozen response bank and applied to both populations.

Population 1 always uses the neutral response. In population 2, advice is activated with conditional probability

$$D \mid X = x \sim \mathrm{Bernoulli}\{\rho s(x)\}, \qquad s(x) = 0.5 + 0.5 S\left(\frac{x - 0.35}{0.30}\right),$$

and

$$S(t) = \begin{cases} 0, & t \le 0, \\ t^3(10 - 15t + 6t^2), & 0 < t < 1, \\ 1, & t \ge 1. \end{cases}$$

Thus, the local probability of receiving advice increases with both the routing intensity ρ and the conditioning profile x.

For the Physics analysis, advice is applied jointly to the Math and Physics coordinates when $D = 1$; the remaining coordinates use their neutral responses. The advice for each routed question is the auxiliary model's highest-scoring valid answer. This construction induces a conditional distributional change whose strength varies with x, allowing the Physics coordinate to be tested at different difficulty profiles.

For the repeated Physics experiment, we generate independent samples with $n_1 = n_2 = 500$ and evaluate

$$x_0 \in \{0.25, 0.75\}, \qquad \rho \in \{0, 0.05, 0.10, 0.20, 0.30\}.$$

Conditional response distributions are estimated using a conditional kernel density estimator under two-fold cross-fitting. Each setting uses $M = 2048$ conditional draws and $B = 1999$ multiplier-bootstrap draws over 100 independently resampled snapshots.

ImageNet-9 background sensitivity. We use the official ImageNet-9 Backgrounds Challenge data, which contain 4,050 foreground identities from nine semantic classes. Foreground identities are split within class into pilot, formal, and interpretation subsets with proportions 20%, 60%, and 20%, respectively. Only the formal subset is used for hypothesis testing. The $2{,}430$ formal foreground identities are further divided into two non-overlapping groups of equal size, stratified by true class and pilot-defined foreground-size bins.

For population 1, each selected foreground is evaluated using its Mixed-Same image, in which the background is sampled from the same class. For population 2, each selected foreground is evaluated using its Mixed-Rand image, in which the background is sampled without matching the foreground class. No foreground identity appears in both formal samples, yielding independent samples with

$$n_1 = n_2 = 1215.$$

For each image, let M_i denote the foreground mask after the deterministic resize and center crop used by CLIP. The conditioning variable is the fraction of foreground pixels in the resulting 224×224 input,

$$X_i^{\mathrm{raw}} = \frac{1}{224^2}\sum_{u,v} M_i(u,v).$$

Thus, X measures foreground coverage in the CLIP input. The pilot sample is used to standardize X. We evaluate the formal tests at the 5th, 35th, and 90th percentiles of the pilot foreground-coverage distribution, corresponding to raw coverages of 6.0%, 25.6%, and 69.3%.

The response is obtained from a frozen CLIP ViT-B/32 model. Let $f_{\mathrm{img}}(I_i)$ denote the normalized image embedding and $f_{\mathrm{text}}(t_j)$ the normalized text embedding for the prompt ``a photo of a {classname}'', where j indexes the nine ImageNet-9 classes. We define

$$Y_{ij}^{\mathrm{raw}} = f_{\mathrm{img}}(I_i)^\top f_{\mathrm{text}}(t_j), \qquad j = 1,\dots,9.$$

Hence $Y_i \in \mathbb{R}^9$ is the vector of class-specific CLIP image–text cosine similarities. Each response coordinate is standardized using the pooled pilot scores, and the same transformation is applied to both formal samples.

The formal analysis uses two-fold cross-fitting with a conditional RealNVP estimator, $M = 1024$ conditional draws, $B = 5000$ multiplier-bootstrap draws, $\sigma_F = 1$, and localization bandwidth $h = 0.8n^{-1/4}$. For the association kernel, the median-distance rule for independent $\mathrm{Unif}([0,1]^9)$ draws gives $\sigma_C = 1.204$.

At each foreground-coverage profile, LTCC tests which semantic-score coordinates contribute to the conditional distributional discrepancy between the Mixed-Same and Mixed-Rand populations.

E Competing Coordinate-Localization Methods

We compare LTCC with two complementary notions of response-coordinate relevance. Marginal Local-CMMD targets coordinatewise conditional marginal discrepancies, whereas Zipper-LOCO targets the incremental predictive value of a coordinate for distinguishing the two sample groups. Together, these comparisons distinguish the coordinate contribution targeted by LTCC from marginal-only and prediction-based notions.

Marginal Local-CMMD. We apply the Local-CMMD test of Yan et al. (2022) separately to each scalar response coordinate. For coordinate j, it tests

$$H_{0,j}^{\mathrm{marg}}(x_0) : F_j^{(1)}(\cdot \mid x_0) = F_j^{(2)}(\cdot \mid x_0).$$

Writing

$$\mu_{\ell,j}(x_0) = \mathbb{E}\Big[k_\gamma(Y_j^{(\ell)},\cdot)\,\Big|\,X^{(\ell)} = x_0\Big], \qquad \ell = 1,2,$$

the corresponding population discrepancy is

$$\mathrm{CMMD}_j^2(x_0) = \big\|\mu_{1,j}(x_0) - \mu_{2,j}(x_0)\big\|_{\mathcal{H}_{k_\gamma}}^2.$$

We use the Gaussian response kernel $k_\gamma(y, y') = \exp\{-(y - y')^2/(2\gamma^2)\}$, with γ selected by the pooled median heuristic, and Gaussian-kernel smoothing over X. The two smoothing bandwidths are selected separately by leave-one-out cross-validation for the conditional kernel mean estimator. We use the conditional U-statistic construction and pooled local bootstrap of Yan et al. (2022). Thus, this baseline isolates conditional marginal changes in Y_j, whereas LTCC additionally captures changes in its marginal-standardized association with Y_{-j}.

Zipper-LOCO. We recast the two-sample problem as binary classification, with group label $G = 0$ for observations from group 1 and $G = 1$ for observations from group 2. For coordinate j, a full classifier predicts G from Y, while a reduced classifier predicts G from Y_{-j}. Using the negative Brier score

$$C(f) = -\mathbb{E}\big[\{G - f(Y)\}^2\big]$$

as the predictiveness criterion, the resulting leave-one-coordinate-out target is

$$\psi_j = C(f_{\text{full}}) - C(f_{-j}),$$

and we use Zipper (Chen et al., 2024) to test $H_{0,j}^{\text{LOCO}} : \psi_j = 0$ against $\psi_j > 0$.

We fit standardized L_2-regularized logistic regressions with inverse regularization strength 10 using stratified five-fold cross-fitting. Within each held-out fold, Zipper evaluates the full and reduced predictors on two partially overlapping subsets; the slider parameter τ controls their overlap and is selected using the rule of Chen et al. (2024) with $n_0 = 50$. This device avoids the null degeneracy that arises when both predictors are evaluated on exactly the same observations. In the redundancy experiment of Figure 4, the response distribution does not vary with X, so Zipper-LOCO measures the incremental predictive value of Y_j for group discrimination. LTCC instead measures the contribution of Y_j to the joint-distribution discrepancy.