DISTRIBUTIONAL BALANCING WEIGHTS FOR CONFORMAL CALIBRATION UNDER COVARIATE SHIFT

Anonymous authors
Paper under double-blind review

ABSTRACT

Conformal prediction provides distribution-free marginal coverage under exchangeability. Under covariate shift, covariate distributions change while the conditional response distribution remains unchanged. Weighted conformal prediction corrects this mismatch using known density ratios, which can be difficult to estimate. We propose conformal calibration with *distributional balancing weights*, learned to align source calibration and unlabeled target covariate distributions without using calibration or target responses. Characteristic function distance provides a common framework for kernel-based and energy-based balancing. We establish a finite-sample error bound for estimating the target score distribution, yielding approximate marginal coverage guarantees. The bound separates covariate imbalance, approximation of conditional score probabilities, target-sample uncertainty, and weight concentration. We connect this bound to weight optimization and develop a validation rule for selecting weights without responses. A spectral analysis shows that nearly uniform fitted weights can miss increasingly fine-scale shifts and yield persistent undercoverage despite bounded density ratios. Controlled synthetic and semi-synthetic experiments reproduce this failure and show that tuning bandwidth and regularization can restore near-nominal coverage. Together, these results highlight the importance of kernel choice and regularization beyond controlling weight concentration alone.

1 INTRODUCTION

Conformal prediction calibrates the errors of a fitted model to construct prediction sets with finite-sample marginal coverage under exchangeability (Papadopoulos et al., 2002; Vovk et al., 2005; Angelopoulos & Bates, 2023). When a model is deployed in a new population, its calibration sample may no longer represent the cases it encounters. Under covariate shift, the conditional response law is shared but the covariate frequencies change (Shimodaira, 2000). A target population can therefore contain a different mixture of easy and difficult predictions, requiring a different calibration threshold. We study how labeled source observations and unlabeled target covariates can be used to recover that threshold.

Weighted conformal prediction (WCP) provides an exact correction when the target-to-source density ratio is known (Tibshirani et al., 2019). With estimated ratios, calibration depends both on representing the shift and on controlling weight concentration. Clipping controls extreme ratios, while likelihood-ratio regularization learns an adjusted calibration threshold (Wang & Goel, 2026; Joshi et al., 2025). Representing the shift can be challenging even with substantial overlap: a change in covariate dependence can alter the frequency of high-error regions while leaving all univariate marginals unchanged and the true ratio uniformly bounded. A restricted ratio model may then produce nearly uniform weights while missing the shift. This motivates evaluating weights through the distributional balance they achieve, together with their concentration.

Our approach starts from the conformity score, a scalar measure of agreement between an observed response and a fitted prediction rule. Covariate shift preserves its conditional distribution at each covariate value but changes the population mixture. Recovering the target score distribution therefore requires balancing the covariate-dependent probabilities that scores fall below each threshold. A discrepancy that controls these probabilities can turn weighted source scores into an estimate of the target score distribution, whose quantiles provide deployment-specific calibration thresholds.

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.

Kernel mean matching (KMM) already provides a distribution-level correction (Huang et al., 2007), with rates governed by regression regularity and kernel capacity (Yu & Szepesvári, 2012), and KMM-CP connects kernel balance to conformal calibration (Laghuvarapu et al., 2026). We use *distributionally balanced conformal calibration* (DB-CC) as a common interface for these response-free weighting rules. Our focus is how geometry, regularization, and empirical fitting recover score-relevant shifts. We develop a calibration-error decomposition and a response-free selection guarantee. For bounded shifts with identical oracle-ratio distributions, we establish a coverage separation between kernels with different spectral resolution. This isolates insufficient spectral resolution from extreme-weight instability.

The choice of discrepancy determines which conditional score patterns balance controls effectively. We study characteristic function distance (CFD), whose frequency weighting encompasses translation-invariant kernel mean discrepancies and generalized energy distance (Sejdinovic et al., 2013; Santra et al., 2026). The same representation describes Gaussian, product-Laplace, Matérn, and energy balancing. We connect its function-space interpretation to actual target coverage: even with uniformly bounded ratios, fixed Gaussian and Matérn geometries can yield different limiting calibration behavior as score-relevant shifts become more oscillatory. Adapting spatial scales and regularization offers a way to resolve the missed structure.

The paper makes three contributions: **(i) Calibration error from distributional balance.** A uniform bound on score-distribution error separates achieved imbalance, approximation of conditional score probabilities, finite-target sampling, and effective sample size. It yields target-coverage bounds and tie-aware quantile consequences for general balance–transfer pairs. **(ii) Optimization and response-free selection.** An objective-value inequality links the bound to the fitted balancing program. Common-geometry validation gives a population-balance oracle inequality without a union-bound penalty in the number of fitted candidates, conditional on the fixed validation geometry. **(iii) End-to-end spectral resolution.** We derive a calibration rate and an explicit coverage separation for empirical balancing weights under fixed-bandwidth geometries. Controlled experiments evaluate the separation and adaptive recovery; bounded nonlinear-shift experiments compare DB-CC with interaction-aware and flexible ratio WCP.

Appendix C details the connections with other balancing and covariate-shift calibration methods.

2 Setup and Calibration Target

2.1 Covariate Shift and a Fixed Calibration Score

Let $X \in \mathbb{R}^d$ and Y denote covariates and response. The source supplies labeled training and calibration observations; the target is the deployment population. Their joint laws factor as $P_{\mathrm{src},X} \times P_{Y|X}$ and $P_{\mathrm{tar},X} \times P_{Y|X}$, with a shared conditional response law.

Let $\mathcal{H}$ contain the independent source training and independent algorithmic randomization. Conditional on $\mathcal{H}$, the score, preprocessing, candidate balancing configurations, and common validation discrepancy are fixed. Appendix B specifies the kernel implementation. The calibration observations and unlabeled target covariates are mutually independent samples, independent of $\mathcal{H}$:

$$(X_i, Y_i)_{i=1}^n \overset{\text{iid}}{\sim} P_{\mathrm{src},X} \times P_{Y|X}, \qquad Z_1, \dots, Z_m \overset{\text{iid}}{\sim} P_{\mathrm{tar},X}.$$

Tuning additionally uses $V_1, \dots, V_v \overset{\text{iid}}{\sim} P_{\mathrm{tar},X}$, independent of training, calibration, and balancing samples. An independent future pair (X_{n+1}, Y_{n+1}) follows the target joint law.

For a covariate value x, candidate response y, and cutoff t, a fixed real-valued score $S(x, y)$ defines nested prediction sets

$$C_t(x) = \{y : S(x, y) \le t\},$$

where larger scores indicate poorer agreement with the fitted rule. Absolute residuals, conformalized quantile-regression scores, and ordered discrete scores all fit this construction; Appendix A defines the three implementations used here.

For each score cutoff $s \in \mathbb{R}$, define

$$h_s(x) = P\{S(X, Y) \le s \mid X = x, \mathcal{H}\}, \qquad F_{\mathrm{tar}}(s) = E_{\mathrm{tar}}\{h_s(X) \mid \mathcal{H}\}.$$

We suppress the dependence on $\mathcal{H}$ below. The function h_s is shared across domains, and F_{tar} is the fitted score's target cumulative distribution function (CDF). DB-CC estimates this CDF from weighted source calibration scores, and its quantiles supply the cutoff for C_t.

2.2 Target-score calibration and weighted conformal prediction

When $r = dP_{\mathrm{tar},X}/dP_{\mathrm{src},X}$ is known, oracle WCP assigns masses proportional to $r(X_1), \ldots, r(X_n), r(x)$ to calibration scores and a candidate-specific maximal-score atom (Tibshirani et al., 2019). These masses yield finite-sample marginal target coverage of at least $1 - \alpha$ through weighted exchangeability. The unified framework of Barber & Tibshirani (2026) expresses this construction through a conditional law over data orderings. DB-CC instead uses global balancing weights to estimate F_{tar} for the deployment population. Its quantile rule is governed by a uniform score-CDF error bound, developed in Section 4.

3 Distributionally balanced conformal calibration

3.1 A general balancing-weight interface

For normalized weights $\pi = (\pi_1, \ldots, \pi_n)$, write

$$\widehat{P}_\pi = \sum_{i=1}^{n} \pi_i \delta_{X_i}, \qquad \widehat{Q}_m = \frac{1}{m} \sum_{j=1}^{m} \delta_{Z_j}, \qquad \mathrm{ESS}(\pi) = \frac{1}{\sum_{i=1}^{n} \pi_i^2}.$$

For a relative cap $B_w \geq 1$, let

$$\Delta_n(B_w) = \left\{ \pi \in \mathbb{R}^n : \pi_i \geq 0,\ \sum_i \pi_i = 1,\ \pi_i \leq B_w/n \right\}.$$

Let $D(P, Q)$ be a discrepancy between covariate distributions and let $R(\pi)$ be a convex dispersion penalty. A general distributional-balancing estimator takes the form

$$\widehat{\pi}_{D,\lambda} \in \mathrm{argmin}_{\pi \in \Delta_n(B_w)} \left\{ D^2(\widehat{P}_\pi, \widehat{Q}_m) + \lambda R(\pi) \right\}. \tag{1}$$

Finite-moment balancing compares selected feature means, while kernel and energy balancing compare richer classes of functions. Entropy or quadratic penalties control weight dispersion. A constrained version minimizes dispersion subject to a specified imbalance tolerance. Section 4 applies to nonnegative, normalized, response-free weights through the achieved discrepancy and effective sample size (ESS), so it also covers alternative solvers and constrained formulations.

3.2 Characteristic function distance balancing

For a probability law P on $\mathbb{R}^d$, let $\varphi_P(t) = E_P\{e^{\mathrm{i}t^\top X}\}$ denote its characteristic function. Given a nonnegative, even frequency weighting function ω, define

$$\mathrm{CFD}^2_\omega(P, Q) = \int_{\mathbb{R}^d} |\varphi_P(t) - \varphi_Q(t)|^2 \omega(t)\, dt, \tag{2}$$

whenever the integral is finite. For positive integrable spectra, write $\Omega_\omega = \int \omega(t)\, dt$ and normalize it to one when convenient. Generalized spectra include the energy weight, whose CFD is finite under a first-moment condition.

Our primary estimator uses a quadratic dispersion penalty whose scale is normalized across calibration sample sizes. For a tuning parameter $\lambda_0 \geq 0$, define

$$\widehat{\pi}_{\omega,\lambda_0} \in \mathrm{argmin}_{\pi \in \Delta_n(B_w)} \left\{ \mathrm{CFD}^2_\omega(\widehat{P}_\pi, \widehat{Q}_m) + \lambda_0 n \sum_{i=1}^{n} \pi_i^2 \right\}. \tag{3}$$

Because $n \sum_i \pi_i^2 = n/\mathrm{ESS}(\pi)$ equals one under uniform weights, λ_0 has a comparable interpretation across calibration sample sizes. The regularization and cap directly control concentration of the calibration masses.

For integrable ω, Bochner's theorem identifies CFD with the maximum mean discrepancy (MMD) of $k_\omega(x, x') = \int e^{\mathrm{i}t^\top(x-x')}\omega(t)\,dt$. Writing $K_{XX} = (k_\omega(X_i, X_j))$ and $K_{XZ} = (k_\omega(X_i, Z_j))$, the objective in (3), up to a constant, is

$$\pi^\top(K_{XX} + \lambda_0 nI)\pi - \frac{2}{m}\pi^\top K_{XZ}\mathbf{1}_m.$$

Thus the weights follow from a convex quadratic program. Energy balancing uses the analogous pairwise-distance objective, and Appendix C gives the correspondence.

3.3 Spectral geometry

Gaussian CFD has exponentially decaying spectral tails, whereas Product-Laplace, Matérn and generalized energy geometries have polynomial spectral tails. A slowly decaying spectrum puts more weight on high-frequency covariate discrepancies. Its reciprocal appears in the transfer norm, so the same choice lowers the relative cost of representing high-frequency conditional score probabilities. Section 4.4 quantifies this relationship. The spectrum determines which score patterns are effectively controlled by balancing.

3.4 Target score-distribution estimation

Given response-free weights $\widehat{\pi}$, define

$$\widehat{F}_{\widehat{\pi}}(s) = \sum_{i=1}^{n} \widehat{\pi}_i \mathbf{1}\{S_i \le s\}, \qquad S_i = S(X_i, Y_i), \tag{4}$$

and the generalized weighted quantile

$$\widehat{q}_\tau = \inf\{s : \widehat{F}_{\widehat{\pi}}(s) \ge \tau\}. \tag{5}$$

The resulting prediction set is

$$\widehat{C}_\tau(x) = \{y : S(x, y) \le \widehat{q}_\tau\}.$$

We use the direct weighted quantile at $\tau = 1-\alpha$ throughout. Unlike ordinary split CP, DB-CC does not apply an $(n+1)$ rank correction or include a candidate-specific test mass. Algorithm 1 summarizes the calibration procedure, including optional response-free selection in a common validation geometry.

Algorithm 1 Distributionally balanced conformal calibration

Require: Source labels; target covariates; score; level α; balancing configurations and cap $B_w \ge 1$; if tuning, ESS fraction $\rho \in (0, 1]$ and common discrepancy D_0.

1: Split source observations into independent training and calibration subsets.
2: Fix the score, preprocessing, configurations, and D_0 from training information.
3: If tuning, split target covariates into independent balancing $Z_{1:m}$ and validation $V_{1:v}$ samples; otherwise use all target covariates for balancing.
4: Fit response-free weights using calibration covariates and $Z_{1:m}$ only.
5: If tuning, screen fitted candidates for feasibility, numerical acceptance, and ESS $\ge \rho n$; add $\pi_i^{(0)} = 1/n$ separately as a uniform fallback.
6: If tuning, minimize $D_0(\widehat{P}_{\pi^{(a)}}, \widehat{Q}_v^{\mathrm{val}})$ with a fixed tie rule and retain the selected weights without refitting; otherwise use the prespecified feasible fit.
7: Compute scores and (4); take the direct quantile (5) at $\tau = 1 - \alpha$.
8: Return $\widehat{C}_{1-\alpha}(x)$; record imbalance, ESS, maximum weight, and fitting failures.

Common-geometry tuning. Training covariates fix the preprocessing and a finite kernel library. Candidate weights are fitted against $Z_{1:m}$ and screened by $\mathrm{ESS}/n \ge \rho$. A single, training-fixed validation kernel k_0 evaluates every candidate against $V_{1:v}$ and we retain the weights with smallest validation MMD. For $B_w \ge 1$ and $0 < \rho \le 1$, the uniform fallback ensures a nonempty candidate set; coverage still depends on the score-CDF error. The selected weights are used directly, without refitting on the validation sample. The nonlinear study uses the equal mixture of 15 unit-diagonal kernels at five bandwidths per family. All bandwidths and spectral families are therefore compared in one fixed geometry. Appendix B gives the complete rule and Appendix E proves its guarantee.

4 COVERAGE ANALYSIS FOR DISTRIBUTIONAL BALANCING

4.1 A GENERAL BALANCE-TO-SCORE TRANSFER PRINCIPLE

The finite-sample guarantees below are conditional on $\mathcal{H}$ while spectral-resolution probability limits are joint over training and calibration/target samples. Let

$$\mathcal{F}_{\mathrm{cal}} = \mathcal{H} \vee \sigma((X_i, Y_i)_{i=1}^n,\, Z_1, \ldots, Z_m)$$

include the fitted rule and calibration information. When validation is used, include $\sigma(V_1, \ldots, V_v)$ as well.

Let $D(P, Q)$ be a symmetric discrepancy between probability laws, and let $\|\cdot\|_D^\star$ be a seminorm on a class of real-valued test functions. We call $(D, \|\cdot\|_D^\star)$ a *balance–transfer pair* if

$$|E_P f(X) - E_Q f(X)| \le \|f\|_D^\star D(P, Q) \tag{6}$$

for every admissible P, Q, and f. Section 4.3 establishes this inequality for CFD through Fourier duality.

Assumption 1. *Assume that the source and target covariate laws are supported on a common compact set $\mathcal{X} \subset \mathbb{R}^d$, and that for every score cutoff $s \in \mathbb{R}$, there exists a globally defined admissible function g_s such that*

$$\|g_s\|_D^\star \le B_D, \qquad \sup_{x \in \mathcal{X}} |h_s(x) - g_s(x)| \le \varepsilon_D,$$

where B_D and ε_D do not depend on s. The discrepancy test functions are interpreted through their restrictions to $\mathcal{X}$ when expectations are taken under the source and target laws.

The pair (B_D, ε_D) trades approximation accuracy against representation cost in the transfer seminorm. For MMD, this cost is the norm in the reproducing kernel Hilbert space (RKHS). Population transfer is stated in Proposition 1 in Appendix D; Appendix D.3 gives the residual-score calculation and Appendix D.4 treats unbounded support.

4.2 FINITE-SAMPLE SCORE-DISTRIBUTION AND COVERAGE BOUNDS

For $\eta \in (0, 1)$ let $\mathfrak{r}_{m,D}(\eta)$ be any quantity fixed by $\mathcal{H}$ satisfying

$$P\left\{ D(\widehat{Q}_m, P_{\mathrm{tar}}) \le \mathfrak{r}_{m,D}(\eta) \mid \mathcal{H} \right\} \ge 1 - \eta. \tag{7}$$

The target radius measures the sampling error of the unlabeled target representation, and each discrepancy supplies its own concentration bound.

Theorem 1. *Assume that $(D, \|\cdot\|_D^\star)$ is a balance–transfer pair and that Assumption 1 holds. Let π be any nonnegative normalized weight vector measurable with respect to $\mathcal{H}$ and calibration, balancing, and optional validation covariates, but not calibration responses. Conditional on $\mathcal{H}$, with probability at least $1 - \delta$ jointly over calibration and unlabeled target samples,*

$$\sup_{s \in \mathbb{R}} |F_{\mathrm{tar}}(s) - \widehat{F}_\pi(s)| \le B_D D(\widehat{P}_\pi, \widehat{Q}_m) + 2\varepsilon_D + B_D \mathfrak{r}_{m,D}(\delta/2) + \frac{4 + \sqrt{\frac{1}{2}\log(2/\delta)}}{\sqrt{\mathrm{ESS}(\pi)}} =: \Delta_{\delta,D}(\pi). \tag{8}$$

The four terms separate achieved covariate imbalance, approximation of conditional score probabilities, finite-target sampling, and weighted calibration fluctuation. The last term increases as the weights concentrate. The bound applies to any response-free weighting algorithm, and the fitted weights affect the bound only through their achieved discrepancy and effective sample size.

Corollary 1. *Let $\widehat{q}_\tau$ be defined by (5). On the event in Theorem 1, an independent target pair satisfies*

$$P\left\{ Y_{n+1} \in \widehat{C}_\tau(X_{n+1}) \mid \mathcal{F}_{\mathrm{cal}} \right\} = F_{\mathrm{tar}}(\widehat{q}_\tau) \ge \tau - \Delta_{\delta,D}(\pi).$$

Hence the nominal choice $\tau = 1 - \alpha$ has target coverage at least $1 - \alpha - \Delta_{\delta,D}(\pi)$ on the same event. If a valid upper bound $\overline{\Delta} \ge \Delta_{\delta,D}(\pi)$ is available and $1 - \alpha + \overline{\Delta} \le 1$, using $\tau = 1 - \alpha + \overline{\Delta}$ yields target coverage at least $1 - \alpha$.

The same CDF event also controls overcoverage: the upper bound is $\tau + \Delta_{\delta,D}(\pi) + J_\pi(\widehat{q}_\tau)$, where J_π is the weighted score-CDF jump. Boundary randomization removes this jump contribution in expectation. Appendix D proves both statements.

Corollary 2. *Let $\lambda_0 > 0$, define*

$$\mathcal{J}_{D,\lambda_0}(\pi) = D^2(\widehat{P}_\pi, \widehat{Q}_m) + \lambda_0 n \sum_{i=1}^{n} \pi_i^2,$$

and let $\widehat{\pi}_{D,\lambda_0}$ minimize this objective over $\Delta_n(B_w)$. Writing $c_\delta = 4 + \sqrt{\frac{1}{2}\log(2/\delta)}$, one has

$$\begin{aligned}\Delta_{\delta,D}(\widehat{\pi}_{D,\lambda_0}) \le{}& 2\varepsilon_D + B_D \mathfrak{r}_{m,D}(\delta/2) \\ &+ \left(B_D^2 + \frac{c_\delta^2}{\lambda_0 n}\right)^{1/2} \left\{ \inf_{\pi \in \Delta_n(B_w)} \mathcal{J}_{D,\lambda_0}(\pi) \right\}^{1/2}. \end{aligned} \tag{9}$$

Taking $D = \mathrm{CFD}_\omega$ gives the corresponding objective-to-coverage bound for CFD balancing.

For any feasible comparator $\pi_0 \in \Delta_n(B_w)$, the infimum in (9) can be upper-bounded by $\mathcal{J}_{D,\lambda_0}(\pi_0)$. Appendix D gives the numerical near-minimizer version and asymptotic validity when the full error bound vanishes.

Guarantee for common-geometry tuning. Let $D_0 = \mathrm{MMD}_{k_0}$ and let $\mathcal{A}_\rho$ contain the fitted candidates satisfying the ESS floor. Conditional on $\mathcal{H}$ and calibration/balancing covariates, with probability at least $1 - \eta$ over validation,

$$D_0(\widehat{P}_{\pi^{(\widehat{a})}}, P_{\mathrm{tar}}) \le \min_{a \in \mathcal{A}_\rho} D_0(\widehat{P}_{\pi^{(a)}}, P_{\mathrm{tar}}) + 2\mathfrak{r}_{v,0}(\eta). \tag{10}$$

For unit-diagonal k_0, $\mathfrak{r}_{v,0}(\eta) = \{1 + \sqrt{2\log(1/\eta)}\}/\sqrt{v}$. This common-geometry oracle inequality needs no union bound over candidates, and Theorem 3 gives the corresponding score-distribution guarantee.

4.3 CFD SPECIALIZATION

Fourier Cauchy–Schwarz makes CFD a balance–transfer pair with the reciprocal-spectrum norm in (15), and for integrable spectra this is the RKHS–MMD inequality. Proposition 2 and Corollary 7 in Appendix D give the details and explicit target radius. For a normalized spectrum and fixed approximation envelope, the target-sampling contribution is $O(m^{-1/2})$.

4.4 SPECTRAL RESOLUTION AND TARGET COVERAGE

The reciprocal spectrum in (15) describes the cost of representing score variation. Appendix F gives the general exponential-versus-polynomial norm comparison. We now connect spectral resolution to the empirical optimizer and its actual target coverage, using a model in which overlap remains unchanged as spatial complexity increases.

Let the source angle θ be uniform on $[-\pi, \pi)$, with observed covariates $X(\theta) = (\cos\theta, \sin\theta)$ and target density ratio

$$r_K(\theta) = 1 + a\cos(K\theta), \qquad a = 0.8, \quad K \in \mathbb{N}.$$

Conditional on θ, draw $B \sim \mathrm{Bernoulli}\{(1+\cos(K\theta))/2\}$, independently draw $U \sim \mathrm{Unif}(0,1)$ and a symmetric sign ξ, and set $Y = \xi(U + 2B)$. Thus the residual magnitude is a mixture of $\mathrm{Unif}(0,1)$ and $\mathrm{Unif}(2,3)$. The source and target hard-component probabilities are 0.5 and 0.7. For every K, the ratio lies in $[0.2, 1.8]$ and has the same distribution. The marginal response and zero-centered score laws are also unchanged.

For the asymptotic comparison, let $K_n = \lceil \log n \rceil$, use n source calibration pairs, n independent target covariates, and an independent source-training mean $\widehat{c}_n$ based on $2n$ observations. For each of the fixed unit-length-scale Gaussian and Matérn-$3/2$ kernels, solve (3) over $\Delta_n(5)$ with

$$\lambda_n = c_\lambda \frac{\log n}{n}, \qquad c_\lambda > 0,$$

allowing an additive objective gap $t_n = O(n^{-1})$.

Theorem 2. *Under the preceding circular-shift model and asymptotic regime, let*

$$S_n(x, y) = |y - \widehat{c}_n|,$$

and calibrate at level 0.90*. For Matérn-*$3/2$ *balancing,*

$$\sup_s \left| \widehat{F}_{\widehat{\pi}_M}(s) - F_{\mathrm{tar},n}(s) \right| = O_p\left(\frac{(\log n)^{5/2}}{\sqrt{n}} \right), \qquad F_{\mathrm{tar},n}(\widehat{q}_{M,.90}) \xrightarrow{p} 0.90.$$

For Gaussian balancing,

$$\left\| \widehat{\pi}_G - \frac{1}{n}\mathbf{1} \right\|_1 \xrightarrow{p} 0, \qquad F_{\mathrm{tar},n}(\widehat{q}_{G,.90}) \xrightarrow{p} 0.86.$$

Here $F_{\mathrm{tar},n}$ *denotes the target-score CDF conditional on* $\widehat{c}_n$*. The same conclusions hold when the score center is fixed at zero.*

Proposition 3 in Appendix F.3 gives sufficient resolution conditions for general K_n, target size, and regularization. The restricted Gaussian eigenvalues decay factorially, whereas the Matérn eigenvalues have order K^{-4}. Both fixed-bandwidth kernels are consistent at fixed K as sample sizes grow and regularization and optimization error vanish. The theorem concerns a triangular array: both the shift and conditional response law vary with K_n, while their marginal score laws and overlap remain fixed.

The population optimizer makes the mechanism explicit. Its weight is $1 + b_{k,\lambda}\cos(K\theta)$, where $b_{k,\lambda} = a\mu_{k,K}/(\mu_{k,K} + \lambda)$. For zero-centered scores its target coverage is

$$C_{k,\lambda} = 0.90 - \frac{0.1a\lambda}{2\lambda + (2+a)\mu_{k,K}}. \tag{11}$$

Thus the ratio of mode strength to regularization determines the correction. Theorem 2 concerns increasing-frequency laws at fixed bandwidths. Bandwidth adaptation changes this ratio; the experiments below compare adaptive and fixed-bandwidth procedures.

5 Empirical evaluation

5.1 Bounded nonlinear shifts

Our primary study separates nonlinear shift structure from extreme importance weights. Source covariates are uniform on $[-1, 1]^5$, and the target density ratio is $r(x) = 1 + 0.8g(x)$, with

$$g(x) = \tanh(6x_1x_2) \quad \text{or} \quad g_k(x) = \cos(\pi k x_1)\cos(\pi k x_2), \quad k \in \{1, 2\}.$$

The first is an interaction shift and the second is a checkerboard shift. All preserve univariate marginals and satisfy $0.2 \le r \le 1.8$. Checkerboard frequency changes spatial complexity while preserving the entire oracle-ratio distribution. A shared nonlinear Gaussian response model has scale $\sigma(x) = 1.25 + 0.75g(x)$, so the target emphasizes regions with larger residual dispersion. We consider base sample sizes $n \in \{300, 1{,}000\}$. All tuning is response-free and uses independent common-geometry validation. We also include selected flexible WCP, which uses common-validation MMD to select among RBF Nyström and boosted-tree ratio estimators and clipping levels. WCP retains its candidate-specific test-point mass, which can yield more conservative thresholds. Appendix G.1 gives the full protocol and controls.

Table 1: Mean target coverage at nominal 0.90 over 200 paired replications per cell. Tables 7–8 report Monte Carlo SEs, interval lengths, and weight diagnostics.

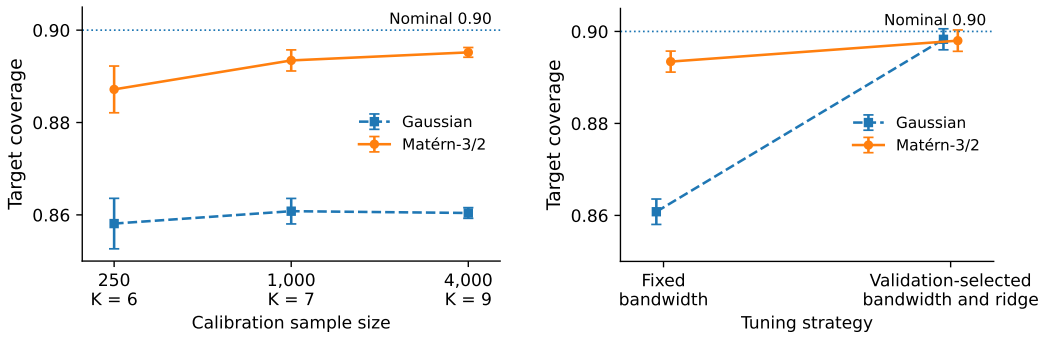
	Interaction		Checkerboard $k = 1$		Checkerboard $k = 2$	
Method	$n = 300$	1,000	$n = 300$	1,000	$n = 300$	1,000
Unweighted split CP	0.844	0.841	0.870	0.869	0.871	0.870
Linear WCP, cap 10	0.846	0.841	0.871	0.869	0.873	0.870
Quadratic WCP	0.888	0.893	0.871	0.869	0.873	0.870
Selected flexible WCP	0.877	0.889	0.872	0.874	0.872	0.871
Gaussian-only DB-CC	0.877	0.892	0.871	0.877	0.870	0.870
Selected DB-CC	0.879	0.891	0.871	0.878	0.870	0.871
Oracle WCP	0.903	0.901	0.902	0.900	0.903	0.901

Nonlinear balance and residual calibration error. Table 1 summarizes coverage. With the larger sample size ($n = 1{,}000$), the nonlinear procedures recover much of the interaction-induced miscalibration missed by linear WCP. Selected DB-CC raises coverage from 0.841 to 0.891 and substantially reduces profile imbalance, performing comparably to the interaction-aware quadratic and flexible WCP procedures. For the low-frequency ($k = 1$) checkerboard, DB-CC yields a more modest improvement in coverage and balance. In contrast, none of the learned procedures resolves the higher-frequency ($k = 2$) shift, although oracle WCP remains near nominal coverage. The nearly uniform fitted weights and persistent profile imbalance suggest that the tested kernel and ratio-model libraries have insufficient effective resolution for this more oscillatory structure. Appendix G.1 provides the complete diagnostics and paired comparisons.

5.2 Spectral resolution and adaptive recovery

We empirically evaluate the finite-sample behavior predicted by Theorem 2 using several (n, K) pairs from the increasing-frequency setting in Section 4.4. We also include a low-frequency control at $(n, K) = (1{,}000, 2)$ and evaluate response-free joint bandwidth–regularization selection at $(n, K) = (1{,}000, 7)$. Target coverage is computed analytically under the known target law. Full experimental details are given in Appendix G.2.

Figure 1a shows that fixed-bandwidth Gaussian coverage remains near 0.86, whereas Matérn coverage approaches 0.90. At $n = 4{,}000$, the Gaussian ESS fraction is 0.996, consistent with Theorem 2. Both kernels cover near 0.898 in the low-frequency control. Figure 1b shows that joint bandwidth–regularization selection gives essentially identical Gaussian and Matérn coverage at $(n, K) = (1{,}000, 7)$, removing the fixed-bandwidth gap in this setting.



(a) Increasing-frequency, fixed-bandwidth sequence. (b) Common-validation tuning at $n = 1{,}000$, $K = 7$.

Figure 1: **(a)** Fixed-bandwidth coverage across (n, K) pairs with unchanged oracle-weight distributions. **(b)** Common-validation bandwidth–ridge selection restores Gaussian coverage. Points show mean analytic target coverage for a learned center over 100 paired repetitions; bars are pointwise 95% Monte Carlo intervals (± 1.96 MCSE).

5.3 Semi-synthetic evaluation under covariate shift

We complement the synthetic experiments with artificial selection within the fixed 442-patient diabetes cohort of Efron et al. (2004), retaining observed outcomes. Across 300 paired assignment-and-partition repetitions at nominal 0.90, linear BMI log-odds selection gives coverage 0.865 for unweighted direct calibration, 0.887 for selected DB-CC, and 0.897 for linear WCP (MCSEs 0.003). Under bounded nonlinear BMI–blood-pressure selection, unweighted, selected DB-CC, and matched Gaussian balancing cover 0.913, 0.910, and 0.909 (MCSEs about 0.002). Paired absolute-gap contrasts do not establish an advantage over Gaussian balancing or flexible WCP. Coverage differences are modest within this cohort; generalization to independent cohorts remains untested. Appendix G.4 gives the protocol, lengths, and paired comparisons.

5.4 Recording-level Parkinson's evaluation

We further evaluate distributionally balanced calibration on the UCI Parkinson's Telemonitoring data, retaining the observed responses and 19 covariates. A nonlinear exponential tilt of the empirical covariate distribution induces the target shift. This recording-level analysis uses pooled calibration and target-balancing covariates for preprocessing, selects tuning parameters within each kernel family, and refits the weights using the full unlabeled target sample after selection. Table 2 summarizes the results, and Appendix G.5 provides the complete experimental protocol.

Table 2: Moderate-shift Parkinson's split-CP results at nominal 0.90 over 100 repetitions. Means are conditional on fitting success; failures are not replaced by uniform weights. Target-reference calibration uses separately resampled labeled records; its CDF error is omitted because self-comparison is zero by construction.

Method	Success	Coverage (MCSE)	Length	CDF error	ESS/n
Unweighted split CP	1.00	0.938 (0.002)	21.35	0.197	1.000
Target reference	1.00	0.899 (0.001)	18.12	–	–
Clipped ratio WCP	1.00	0.926 (0.002)	20.41	–	0.288
Finite-moment balancing	1.00	0.933 (0.002)	20.91	0.154	0.271
Energy balancing	1.00	0.903 (0.002)	18.53	0.148	0.069
Tuned Gaussian CFD	1.00	0.919 (0.002)	19.76	0.159	0.253
Tuned Product-Laplace CFD	0.99	0.917 (0.002)	19.65	0.150	0.244
Tuned Matérn-3/2 CFD	0.98	0.919 (0.002)	19.78	0.147	0.245

Unweighted calibration is conservative. Energy balancing attains mean coverage 0.903, while tuned CFD covers 0.917–0.919, with shorter intervals and mean coverage closer to nominal than clipped ratio WCP or finite-moment balancing. Mean coverage alone does not characterize stability or mean absolute coverage error. Energy's lower ESS indicates greater weight concentration. Appendix G.5 reports sensitivity to tuning and weight constraints, alternative scores, and tie diagnostics.

Because participants contribute repeated recordings, these results concern recording-level performance and do not establish calibration for previously unseen patients.

6 Discussion

DB-CC connects response-free distributional balancing to estimation of the target score CDF. Building on existing balancing objectives, our analysis characterizes how geometry, regularization, and empirical fitting affect calibration, separating achieved imbalance, score approximation, finite-target sampling, and weight concentration. The spectral analysis links the frequency content of score-relevant shifts to the resolution of the balancing kernel: a fixed geometry may miss a bounded high-frequency shift, yielding nearly uniform weights while leaving undercoverage unresolved. The adaptive experiment in Figure 1b also shows that the fixed-bandwidth separation is not an intrinsic ranking of kernel families: changing bandwidth and regularization can restore Gaussian calibration. The empirical results reinforce the importance of shift structure. In the smooth interaction setting, selected DB-CC performs comparably to interaction-aware quadratic WCP, whereas the more oscillatory checkerboard remains difficult for both balancing and flexible ratio estimation. Together, these results highlight the importance of adapting the balancing geometry to score-relevant shift structure.

Limitations and future work. Our coverage guarantees are approximate and depend on how well the balancing functions represent conditional score probabilities. This approximation error is generally unknown, so the theoretical bound cannot directly certify coverage from observed balance and ESS alone. Although response-free validation guides kernel and regularization selection, its performance depends on the candidate library and validation criterion; better covariate balance need not imply smaller coverage error. Developing balancing representations and selection criteria that better capture score-relevant variation is therefore an important direction. Dense kernel matrices and repeated fitting across candidates can also limit scalability, motivating computational approximations with explicit control of their effect on calibration.

AI Use Statement

Generative AI tools including Research Pro and Codex were used to assist us with the formulation and refinement of mathematical claims related to the main theorem, the development and verification of intermediate derivations and proof steps, and the implementation of parts of the proposed methodology. They were also used for software-code development and debugging, LaTeX formatting, and language editing. All AI-assisted mathematical arguments, proof steps, and implementations were independently checked by the authors. The authors made the final decisions regarding the theorem statements, assumptions, methodology, interpretation of results, and scientific conclusions, and independently verified the theoretical claims, experimental results, citations, and code used in the paper. The authors take full responsibility for the final content of the work.

Ethics Statement

This work develops statistical methodology for uncertainty quantification under covariate shift. The proposed guarantees concern marginal target coverage under the assumptions stated in the paper and should not be interpreted as ensuring conditional validity, robustness to arbitrary distribution shift, or safety in downstream deployment. In applications where prediction errors have substantial consequences, distributional assumptions and calibration performance should be assessed for the specific deployment population rather than relying on the method alone. The empirical analyses use existing benchmark datasets and semi-synthetic or simulated source–target shifts; the study does not involve collection of new human-subject data or interventions. We report limitations of the empirical evaluations and do not interpret them as evidence of clinical effectiveness or deployment readiness.

Reproducibility Statement

The accompanying archive provides annotated Python and R code, experiment configurations and seeds, software requirements, and reference results. The README maps experiments to tables and figures and gives commands for full simulations, smoke tests, and regeneration from saved results. Appendices A–G give the score constructions, tuning procedures, related methods, proofs, and experimental protocols.

References

Anastasios N. Angelopoulos and Stephen Bates. Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 16(4):494–591, 2023. doi: 10.1561/2200000101.

Rina Foygel Barber and Ryan J. Tibshirani. Unifying different theories of conformal prediction. *Electronic Journal of Statistics*, 20(1):1428–1474, 2026. doi: 10.1214/26-EJS2513.

Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2):816–845, 2023. doi: 10.1214/23-AOS2276.

Blake Bordelon, Abdulkadir Canatar, and Cengiz Pehlevan. Spectrum dependent learning curves in kernel regression and wide neural networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 1024–1034. PMLR, 2020. URL https://proceedings.mlr.press/v119/bordelon20a.html.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford, 2013. doi: 10.1093/acprof:oso/9780199535255.001.0001.

Andrea Caponnetto and Ernesto De Vito. Optimal rates for the regularized least-squares algorithm. *Foundations of Computational Mathematics*, 7(3):331–368, 2007. doi: 10.1007/s10208-006-0196-8.

Maxime Cauchois, Suyash Gupta, Alnur Ali, and John C. Duchi. Robust validation: Confident predictions even when distributions shift. *Journal of the American Statistical Association*, 119 (548):3033–3044, 2024. doi: 10.1080/01621459.2023.2298037.

Kwun Chuen Gary Chan, Sheung Chi Phillip Yam, and Zheng Zhang. Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(3):673–700, 2016. doi: 10.1111/rssb.12129.

Yihang Chen, Fanghui Liu, Taiji Suzuki, and Volkan Cevher. High-dimensional kernel methods under covariate shift: Data-dependent implicit regularization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 7081–7102. PMLR, 2024. URL https://proceedings.mlr.press/v235/chen24aa.html.

Kacper P. Chwialkowski, Aaditya Ramdas, Dino Sejdinovic, and Arthur Gretton. Fast two-sample testing with analytic representations of probability measures. In *Advances in Neural Information Processing Systems 28*, 2015. URL https://papers.nips.cc/paper_files/paper/2015/hash/b571ecea16a9824023ee1af16897a582-Abstract.html.

Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407–499, 2004. doi: 10.1214/009053604000000067.

Isaac Gibbs and Emmanuel J. Candès. Adaptive conformal inference under distribution shift. In *Advances in Neural Information Processing Systems 34*, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/0d441de75945e5acbc865406fc9a2559-Abstract.html.

Isaac Gibbs, John J. Cherian, and Emmanuel J. Candès. Conformal prediction with conditional guarantees. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(4):1100–1126, 2025. doi: 10.1093/jrsssb/qkaf008.

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

Leying Guan. Localized conformal prediction: A generalized inference framework for conformal prediction. *Biometrika*, 110(1):33–50, 2023. doi: 10.1093/biomet/asac040.

Jens Hainmueller. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46, 2012. doi: 10.1093/pan/mpr025.

Chad Hazlett. Kernel balancing: A flexible non-parametric weighting procedure for estimating causal effects. *Statistica Sinica*, 30(3):1155–1189, 2020. doi: 10.5705/ss.202017.0555.

Christopher R. Heathcote. The integrated squared error estimation of parameters. *Biometrika*, 64(2):255–264, 1977. doi: 10.1093/biomet/64.2.255.

Rohan Hore and Rina Foygel Barber. Conformal prediction with local weights: Randomization enables robust guarantees. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(2):549–578, 2025. doi: 10.1093/jrsssb/qkae103.

Jiayuan Huang, Alexander J. Smola, Arthur Gretton, Karsten M. Borgwardt, and Bernhard Schölkopf. Correcting sample selection bias by unlabeled data. In Bernhard Schölkopf, John Platt, and Thomas Hofmann (eds.), *Advances in Neural Information Processing Systems 19*, pp. 601–608. MIT Press, 2007.

Jared D. Huling and Simon Mak. Energy balancing of covariate distributions. *Journal of Causal Inference*, 12(1):20220029, 2024. doi: 10.1515/jci-2022-0029.

Sunay Joshi, Shayan Kiyani, George J. Pappas, Edgar Dobriban, and Hamed Hassani. Conformal inference under high-dimensional covariate shifts via likelihood-ratio regularization. In *Advances in Neural Information Processing Systems 38*, 2025.

Nathan Kallus. Generalized optimal matching methods for causal inference. *Journal of Machine Learning Research*, 21(62):1–54, 2020. URL https://jmlr.org/papers/v21/19-120.html.

Takafumi Kanamori, Shohei Hido, and Masashi Sugiyama. A least-squares approach to direct importance estimation. *Journal of Machine Learning Research*, 10(48):1391–1445, 2009. URL https://www.jmlr.org/papers/v10/kanamori09a.html.

Siddhartha Laghuvarapu, Rohan Deb, and Jimeng Sun. KMM-CP: Practical conformal prediction under covariate shift via selective kernel mean matching. In Emilija Perković and Daniel Malinsky (eds.), *Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence*, volume 337 of *Proceedings of Machine Learning Research*, pp. 3237–3261. PMLR, 17–21 Aug 2026.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113 (523):1094–1111, 2018. doi: 10.1080/01621459.2017.1307116.

Cheng Li, Siming Zheng, and Yuanyuan Lin. Generative conformal prediction under covariate shift. *Journal of Computational and Graphical Statistics*, 0(0):1–9, 2026. doi: 10.1080/10618600.2026.2703276.

Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning*, 10(1–2):1–141, 2017. doi: 10.1561/2200000060.

Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman. Inductive confidence machines for regression. In *Machine Learning: ECML 2002*, pp. 345–356. Springer, 2002. doi: 10.1007/3-540-36755-1_29.

Hongxiang Qiu, Edgar Dobriban, and Eric Tchetgen Tchetgen. Prediction sets adaptive to unknown covariate shift. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(5): 1680–1705, 2023. doi: 10.1093/jrsssb/qkad069.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems 20*, 2007. URL https://papers.nips.cc/paper/2007/hash/013a006f03dbc5392effeb8f18fda755-Abstract.html.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Yaniv Romano, Evan Patterson, and Emmanuel J. Candès. Conformalized quantile regression. In *Advances in Neural Information Processing Systems 32*, 2019.

Raphael Rossellini, Rina Foygel Barber, and Rebecca Willett. Integrating uncertainty awareness into conformalized quantile regression. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 1540–1548, 2024.

Diptanil Santra, Guanhua Chen, and Chan Park. Distributional balancing for causal inference: A unified framework via characteristic function distance. arXiv preprint arXiv:2601.15449, 2026. URL https://arxiv.org/abs/2601.15449.

Dino Sejdinovic, Bharath Sriperumbudur, Arthur Gretton, and Kenji Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics*, 41(5): 2263–2291, 2013. doi: 10.1214/13-AOS1140.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(12):371–421, 2008. URL https://www.jmlr.org/papers/v9/shafer08a.html.

Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2):227–244, 2000. doi: 10.1016/S0378-3758(00)00115-4.

Bharath K. Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert R. G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11(50):1517–1561, 2010. URL https://www.jmlr.org/papers/v11/sriperumbudur10a.html.

Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer, 2008. doi: 10.1007/978-0-387-77242-4.

Masashi Sugiyama, Shinichi Nakajima, Hisashi Kashima, Paul von Bünau, and Motoaki Kawanabe. Direct importance estimation with model selection and its application to covariate shift adaptation. In *Advances in Neural Information Processing Systems 20*, 2007. URL https://papers.nips.cc/paper_files/paper/2007/hash/be83ab3ecd0db773eb2dc1b0a17836a1-Abstract.html.

Ryan J. Tibshirani, Rina Foygel Barber, Emmanuel J. Candès, and Aaditya Ramdas. Conformal prediction under covariate shift. In *Advances in Neural Information Processing Systems 32*, 2019.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York, 1996. doi: 10.1007/978-1-4757-2545-2.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005. doi: 10.1007/b106715.

James Wang and Surbhi Goel. Weight clipping for robust conformal inference under unbounded covariate shifts. arXiv preprint arXiv:2605.02072, 2026. URL https://arxiv.org/abs/2605.02072.

Yixin Wang and José R. Zubizarreta. Minimal dispersion approximately balancing weights: Asymptotic properties and practical considerations. *Biometrika*, 107(1):93–105, 2020. doi: 10.1093/biomet/asz050.

Yachong Yang, Arun Kumar Kuchibhotla, and Eric Tchetgen Tchetgen. Doubly robust calibration of prediction sets under covariate shift. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(4):943–965, 2024. doi: 10.1093/jrsssb/qkae009.

Yaoliang Yu and Csaba Szepesvári. Analysis of kernel mean matching under covariate shift. In *Proceedings of the 29th International Conference on Machine Learning*, 2012. URL https://arxiv.org/abs/1206.4650.

José R. Zubizarreta. Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*, 110(511):910–922, 2015. doi: 10.1080/01621459.2015.1023805.

Supplement to

"Distributional Balancing Weights for Conformal Calibration under Covariate Shift"

This supplement gives score constructions, response-free tuning, related work, proofs, and experimental details. Notation and numbering continue from the main paper.

A Conformal-score instantiations

All predictors, quantile functions, and ensembles below are fitted on the training split. Conditional on that split, each score defines a fixed nested family $C_t(x) = \{y : S(x, y) \le t\}$. Calibration uses the weighted CDF and quantile in (4)–(5); inversion gives the corresponding prediction set. For general introductions, see Shafer & Vovk (2008); Angelopoulos & Bates (2023).

A.1 Split residual conformal prediction

For a fitted regression function $\widehat{\mu}$, the absolute-residual construction of Lei et al. (2018) uses

$$S^{\mathrm{split}}(x, y) = |y - \widehat{\mu}(x)|.$$

For a threshold $t \ge 0$, inversion gives

$$C_t^{\mathrm{split}}(x) = [\widehat{\mu}(x) - t, \widehat{\mu}(x) + t].$$

The calibration scores are $S_i^{\mathrm{split}} = |Y_i - \widehat{\mu}(X_i)|$, and distributionally balanced calibration uses their weighted generalized quantile in (5).

A.2 Conformalized quantile regression

Let $\widehat{q}_\ell$ and $\widehat{q}_u$ estimate lower and upper conditional quantiles, typically at levels $\alpha/2$ and $1 - \alpha/2$. The CQR score is

$$S^{\mathrm{CQR}}(x, y) = \max\{\widehat{q}_\ell(x) - y,\ y - \widehat{q}_u(x)\}.$$

The associated nested interval is

$$C_t^{\mathrm{CQR}}(x) = [\widehat{q}_\ell(x) - t, \widehat{q}_u(x) + t].$$

The score can be negative when the fitted quantile interval is conservative; the generalized weighted quantile remains well-defined. This is the standard split-CQR inversion of Romano et al. (2019).

A.3 Positional version of uncertainty-aware CQR

The positional formulation uses a nested sequence of ensemble-derived intervals

$$\mathcal{I}_0(x) \subseteq \mathcal{I}_1(x) \subseteq \cdots \subseteq \mathcal{I}_{B+1}(x),$$

where $\mathcal{I}_{B+1}(x)$ is the designated maximal set. Define the discrete score

$$S^{\mathrm{UACQR-P}}(x, y) = \min\{t \in \{0, \ldots, B + 1\} : y \in \mathcal{I}_t(x)\}.$$

For a weighted quantile $\widehat{q}_\tau \in \{0, \ldots, B + 1\}$, the prediction set is

$$C_{\widehat{q}_\tau}^{\mathrm{UACQR-P}}(x) = \mathcal{I}_{\widehat{q}_\tau}(x).$$

The ordered ensemble construction and its uncertainty interpretation follow Rossellini et al. (2024). The positional score takes discrete values, so the tie-aware results in Appendix D apply directly.

B RESPONSE-FREE TUNING

B.1 TRAINING-FIXED REPRESENTATION

Let $\mathcal{H}$ denote the information generated by the independent source training sample and any algorithmic randomization. The kernel families and tuning grids are prespecified. Within each replication, the source training covariates determine the standardization and base bandwidth scales, thereby fixing the candidate kernels and their equal-mixture validation kernel k_0. Conditional on $\mathcal{H}$, these objects and the fitted score are fixed.

For normalized candidate kernels $k_1, \dots, k_J$ with $k_j(x, x) = 1$, set

$$k_0(x, x') = J^{-1} \sum_{j=1}^{J} k_j(x, x'), \qquad D_0 = \mathrm{MMD}_{k_0}.$$

All kernels are evaluated after the same training-fixed preprocessing. The diagonal of k_0 equals one, and $D_0^2 = J^{-1} \sum_j D_j^2$. The common kernel evaluates weights regardless of which discrepancy generated them. Experiment-specific libraries and grids are specified in Appendix G.

B.2 CANDIDATE FITTING AND COMMON-GEOMETRY SELECTION

Split unlabeled target observations into $Z_{1:m}$ for fitting weights and $V_{1:v}$ for validation. Let $\mathcal{A}_{\mathrm{fit}}$ index configurations fitted using only calibration covariates and $Z_{1:m}$. For $B_w \geq 1$ and $0 < \rho \leq 1$, define $\pi_i^{(0)} = 1/n$ and

$$\mathcal{A}_\rho = \{0\} \cup \left\{ a \in \mathcal{A}_{\mathrm{fit}} : \pi^{(a)} \in \Delta_n(B_w),\ \mathrm{ESS}(\pi^{(a)}) \geq \rho n,\ G_a(\pi^{(a)}) \leq \varepsilon_{\mathrm{opt}} \right\}.$$

Here G_a is the numerical optimality certificate; for a differentiable convex objective $\mathcal{J}_a$, the Frank–Wolfe gap is

$$G_a(\pi) = \max_{u \in \Delta_n(B_w)} \nabla \mathcal{J}_a(\pi)^\top (\pi - u).$$

The uniform fallback is added separately and need not satisfy an optimizer-gap test. With $\widehat{Q}_v^{\mathrm{val}} = v^{-1} \sum_j \delta_{V_j}$, select

$$\widehat{a} \in \arg\min_{a \in \mathcal{A}_\rho} D_0(\widehat{P}_{\pi^{(a)}}, \widehat{Q}_v^{\mathrm{val}}). \tag{12}$$

Ties are broken by a prespecified deterministic rule, and the selected weights are retained without refitting on the validation sample. Uniform weights are feasible and have ESS n, so $\mathcal{A}_\rho$ is nonempty. The selection rule is therefore always defined, although uniform calibration can remain inaccurate under shift. The same validation criterion and admissibility rules can also be used to evaluate energy and finite-moment candidates as separate comparators.

B.3 SELECTION GUARANTEE

Theorem 3 gives a population-balance oracle inequality in the fixed common geometry. A single validation mean-embedding event controls all fitted candidates, avoiding a candidate-wise union penalty. The score-CDF guarantee additionally depends on the common geometry's approximation envelope and the selected weights' ESS; it applies to the retained weights, without refitting.

C DISTRIBUTIONAL BALANCING CRITERIA AND RELATED CONFORMAL METHODS

C.1 HOW CFD REPRESENTS COMMON FULL-DISTRIBUTION DISCREPANCIES

Integrated characteristic-function criteria have a long history in estimation (Heathcote, 1977); smoothed characteristic functions also yield distribution representations for two-sample testing (Chwialkowski et al., 2015). Kernel mean embeddings represent distributions in an RKHS (Muandet et al., 2017). Characteristic kernels distinguish probability laws, but the magnitude of MMD

depends on which differences the kernel emphasizes (Sriperumbudur et al., 2010; Gretton et al., 2012).

The Bochner representation and quadratic program in Section 3.2 identify Gaussian, product-Laplace, and Matérn mean balancing with integrable CFD after matching normalization, constraints, and regularization. Energy distance instead uses a generalized singular spectral weight and can be evaluated through pairwise distances (Sejdinovic et al., 2013; Santra et al., 2026). Finite-moment and entropy balancing fit the general interface but are not generally CFD instances.

Moment and functional balancing. Entropy balancing matches specified covariate moments (Hainmueller, 2012); empirical calibration weighting uses growing collections of balancing functions for causal estimation (Chan et al., 2016). Stable and approximately balancing weights make the balance–dispersion trade-off explicit (Zubizarreta, 2015; Wang & Zubizarreta, 2020). Kernel balancing controls imbalance over richer function spaces (Hazlett, 2020), and generalized optimal matching formulates balance–variance trade-offs through functional-analytic objectives (Kallus, 2020). Energy balancing directly targets covariate distributional discrepancy (Huling & Mak, 2024). These works motivate the general weighting interface, and DB-CC applies it to the family of conditional score probabilities rather than an outcome mean.

C.2 Relation to covariate-shift conformal methods

Laghuvarapu et al. (2026) analyze kernel mean matching for conformal calibration through MMD control of conditional score probabilities. Their RKHS-based argument accommodates a characteristic kernel, and their implementation uses a Gaussian kernel. They also introduce target selection for limited-overlap settings. Our analysis works with general balance–transfer pairs and separates achieved empirical imbalance from finite-target uncertainty. It further provides an objective-value bound, score-tie consequences, and a comparison of spectral transfer costs. CQR and UACQR supply regression-score applications of this calibration interface.

Barber & Tibshirani (2026) organize conformal methods through conditional randomization laws over data orderings and quantify misspecification through total variation. DB-CC instead controls the marginal target score CDF through expectations of conditional score probabilities. Its discrepancy-level analysis is tailored to these threshold events.

Density-ratio estimation and direct calibration. Importance weighting under covariate shift predates conformal prediction (Shimodaira, 2000). Direct ratio estimators include Kullback–Leibler importance estimation (Sugiyama et al., 2007) and least-squares importance fitting (Kanamori et al., 2009). WCP inserts a known or estimated ratio into a candidate-specific rank construction (Tibshirani et al., 2019). The clipped method of Wang & Goel (2026) combines clipped least-squares fitting with a clipping-bias correction. LR-QR instead learns a threshold through regularized quantile regression without directly estimating the likelihood ratio (Joshi et al., 2025). Generative methods use conditional modeling (Li et al., 2026), and semiparametric or doubly robust methods combine conditional-score nuisance models with shift corrections (Qiu et al., 2023; Yang et al., 2024).

Other coverage targets and departures from exchangeability. Localized and randomly localized conformal methods adapt calibration to a test covariate (Guan, 2023; Hore & Barber, 2025). Gibbs et al. (2025) formulate conditional coverage through simultaneous guarantees over a class of covariate shifts. Nonexchangeable conformal prediction controls coverage degradation through departures from exchangeability (Barber et al., 2023); adaptive conformal inference targets long-run coverage with sequential response feedback (Gibbs & Candès, 2021). Robust validation calibrates for distributional uncertainty sets defined by divergence constraints (Cauchois et al., 2024). DB-CC instead uses an unlabeled sample to target one deployment population under a shared conditional response law. These different targets explain why the methods provide complementary rather than interchangeable guarantees.

Table 3 summarizes the conformal-method relationships.

Table 3: Covariate-shift calibration strategies. Oracle coverage refers to at least nominal finite-sample marginal coverage under the stated assumptions; learned corrections have error-dependent or asymptotic guarantees.

Method	Shift correction	Guarantee	Scope
Oracle WCP	Uses true likelihood ratios at calibration and candidate covariates in a weighted-rank rule.	At least nominal finite-sample marginal target coverage under weighted exchangeability.	Generic scores; pointwise ratio and candidate-test mass.
DB-CC (this paper)	Learns global balancing weights and calibrates their weighted empirical score CDF.	High-probability score-CDF and coverage-error bounds for balance–transfer pairs.	General nested scores and discrepancies; CFD spectral analysis.
KMM-CP	Balances RKHS means and uses weighted calibration-score quantiles.	Coverage-error analysis through MMD, approximation error, and ESS.	Split conformal; characteristic-kernel theory and Gaussian implementation (Laghuvarapu et al., 2026).
Selective KMM-CP	Learns source weights and target selection, then refits KMM on retained target points.	Coverage concerns the selected target population under the method's assumptions.	Explicit selection or abstention for limited overlap.
Estimated or clipped WCP	Fits or regularizes a pointwise ratio, optionally clips it, and substitutes it into WCP.	Depends on ratio-estimation error and, where applicable, clipping bias.	Generic scores; candidate-specific weights.
Semiparametric and doubly robust calibration	Combines conditional-score or outcome models with ratio or influence-function corrections.	Asymptotic, one-step, or doubly robust guarantees under nuisance-rate conditions.	Targets score quantiles or coverage using shift and conditional-response information.

C.3 Spectral learning and computational approximation

Kernel regularization rates depend on function regularity and effective dimension (Caponnetto & De Vito, 2007). Bordelon et al. (2020) characterize learning of spectral modes in kernel regression, while Yu & Szepesvári (2012) analyze KMM under covariate shift and Chen et al. (2024) study shift and implicit regularization in high-dimensional kernel methods. At the distribution level, Sriperumbudur et al. (2010) show that high-frequency differences can have small kernel embedding distance. These analyses explain how kernels resolve distributional differences and target functions.

Theorem 2 connects spectral resolution to capped, nonnegative empirical balancing weights and target coverage. Under one regularization schedule and bounded shifts with unchanged oracle-weight distributions, it gives a vanishing score-CDF error rate for one fixed geometry and a nonzero limiting coverage gap for another. The adaptive control examines recovery through joint scale and regularization selection.

Computational approximation presents a related design choice. Random Fourier features replace translation-invariant kernels by finite feature maps (Rahimi & Recht, 2007). For calibration, the relevant requirement is to control the resulting discrepancy or objective error: the near-minimizer analysis in Appendix F.3 supplies one route. The circle experiment uses Fourier truncation; extending the analysis to general random-feature approximations remains a direction for future work.

D Proofs for the distributional-balancing coverage results

Throughout, we condition on $\mathcal{H}$ from Section 2, which fixes the score, preprocessing, candidate geometries, and independent algorithmic randomness. Further conditioning on all calibration, balancing, and optional validation covariates fixes any response-free selected weights while preserving

conditional independence of calibration responses. Write

$$\widehat{Q}_m f = m^{-1} \sum_{j=1}^{m} f(Z_j), \qquad \widehat{P}_\pi f = \sum_{i=1}^{n} \pi_i f(X_i).$$

The empirical-process proof uses standard conditional symmetrization, Doob's maximal inequality, and bounded-difference concentration (van der Vaart & Wellner, 1996; Boucheron et al., 2013; Laghuvarapu et al., 2026). The CFD transfer step is a Fourier Cauchy–Schwarz argument. The details below yield explicit constants and establish the result for generic normalized response-free weights, including settings with tied scores.

D.1 Population target score-distribution transfer

Proposition 1. *Under Assumption 1, for every reweighted source covariate law P^{π}_{src} supported on $\mathcal{X}$,*

$$\sup_{s\in\mathbb{R}} \left| F_{\mathrm{tar}}(s) - E_{P^{\pi}_{\mathrm{src}}}\{h_s(X)\} \right| \le B_D D(P^{\pi}_{\mathrm{src}}, P_{\mathrm{tar}}) + 2\varepsilon_D.$$

Proof. For each s, let g_s satisfy Assumption 1. Since both laws are supported on $\mathcal{X}$, the balance–transfer inequality (6) gives

$$\begin{aligned} |P_{\mathrm{tar}} h_s - P^{\pi}_{\mathrm{src}} h_s| &\le P_{\mathrm{tar}}|h_s - g_s| + P^{\pi}_{\mathrm{src}}|h_s - g_s| + |P_{\mathrm{tar}} g_s - P^{\pi}_{\mathrm{src}} g_s| \\ &\le 2\varepsilon_D + B_D D(P^{\pi}_{\mathrm{src}}, P_{\mathrm{tar}}). \end{aligned}$$

Taking the supremum over s proves the claim. □

D.2 A weighted empirical-process bound for the score distribution

Lemma 1. *Let $\mathcal{C}$ be a σ-field such that $X_{1:n}$ and the nonnegative weights $\pi_{1:n}$ are $\mathcal{C}$-measurable, with $\sum_{i=1}^{n} \pi_i = 1$. Suppose that, conditional on $\mathcal{C}$, the scores $S_1, \dots, S_n$ are independent and satisfy*

$$P(S_i \le s \mid \mathcal{C}) = h_s(X_i), \qquad i = 1, \dots, n, \quad s \in \mathbb{R}.$$

Then, for every $\eta \in (0,1)$, with conditional probability at least $1-\eta$,

$$\sup_{s\in\mathbb{R}} \left| \sum_{i=1}^{n} \pi_i \{ \mathbf{1}(S_i \le s) - h_s(X_i) \} \right| \le \left\{ 4 + \sqrt{\frac{1}{2}\log(1/\eta)} \right\} \sqrt{\sum_{i=1}^{n} \pi_i^2}.$$

The result allows tied scores.

Proof. Write

$$A_i(s) = \mathbf{1}(S_i \le s), \qquad G = \sup_{s\in\mathbb{R}} \left| \sum_{i=1}^{n} \pi_i \{ A_i(s) - h_s(X_i) \} \right|.$$

Since $s \mapsto A_i(s) - h_s(X_i)$ is right-continuous,

$$\sup_{s\in\mathbb{R}} \left| \sum_i \pi_i \{ A_i(s) - h_s(X_i) \} \right| = \sup_{s\in\mathbb{Q}} \left| \sum_i \pi_i \{ A_i(s) - h_s(X_i) \} \right|,$$

so G is measurable.

Let $S'_1, \dots, S'_n$ be a conditionally independent copy of $S_1, \dots, S_n$ given $\mathcal{C}$, and set $A'_i(s) = \mathbf{1}(S'_i \le s)$. Then

$$h_s(X_i) = E\{A'_i(s) \mid \mathcal{C}\},$$

and conditional Jensen's inequality gives

$$\begin{aligned} E(G \mid \mathcal{C}) &= E\left[\sup_s \left| E\left\{ \sum_i \pi_i (A_i(s) - A'_i(s)) \,\middle|\, \mathcal{C}, S_{1:n} \right\} \right| \,\middle|\, \mathcal{C} \right] \\ &\le E\left[\sup_s \left| \sum_i \pi_i \{ A_i(s) - A'_i(s) \} \right| \,\middle|\, \mathcal{C} \right]. \end{aligned}$$

Let $\epsilon_1, \ldots, \epsilon_n$ be independent Rademacher signs, independent of all scores conditional on $\mathcal{C}$. Conditional exchangeability of each pair (S_i, S_i') implies

$$\big(A_i(\cdot) - A_i'(\cdot)\big)_{i=1}^n \overset{d}{=} \big(\epsilon_i\{A_i(\cdot) - A_i'(\cdot)\}\big)_{i=1}^n \quad \text{given } \mathcal{C}.$$

Consequently,

$$\begin{aligned} E(G \mid \mathcal{C}) &\le E\left[\sup_s \left|\sum_i \pi_i \epsilon_i \{A_i(s) - A_i'(s)\}\right| \,\middle|\, \mathcal{C}\right] \\ &\le E\left[\sup_s \left|\sum_i \pi_i \epsilon_i A_i(s)\right| \,\middle|\, \mathcal{C}\right] + E\left[\sup_s \left|\sum_i \pi_i \epsilon_i A_i'(s)\right| \,\middle|\, \mathcal{C}\right] \\ &= 2E\left[\sup_s \left|\sum_i \pi_i \epsilon_i A_i(s)\right| \,\middle|\, \mathcal{C}\right]. \end{aligned}$$

Further condition on the realized scores and choose a permutation σ such that

$$S_{\sigma(1)} \le \cdots \le S_{\sigma(n)}.$$

For every s, the threshold set $\{i : S_i \le s\}$ is one of the attainable prefixes of this ordering. Ties can only remove prefixes from the collection. Define

$$M_0 = 0, \qquad M_k = \sum_{j=1}^k \pi_{\sigma(j)} \epsilon_{\sigma(j)}, \qquad k = 1, \ldots, n.$$

Conditional on $(\mathcal{C}, S_{1:n})$, $(M_k)_{k=0}^n$ is a square-integrable martingale, and

$$\sup_s \left|\sum_i \pi_i \epsilon_i A_i(s)\right| \le \max_{0 \le k \le n} |M_k|.$$

Doob's L^2 maximal inequality and Cauchy–Schwarz yield

$$\begin{aligned} E_\epsilon \max_{0 \le k \le n} |M_k| &\le \left\{E_\epsilon \max_{0 \le k \le n} |M_k|^2\right\}^{1/2} \\ &\le 2\{E_\epsilon M_n^2\}^{1/2} \\ &= 2\left\{\sum_{j=1}^n \pi_{\sigma(j)}^2\right\}^{1/2} = 2\sqrt{\sum_{i=1}^n \pi_i^2}. \end{aligned}$$

Substitution into the symmetrization bound gives

$$E(G \mid \mathcal{C}) \le 4\sqrt{\sum_{i=1}^n \pi_i^2}.$$

It remains to control deviations around this conditional expectation. If only S_i is replaced by $\widetilde{S}_i$, then

$$\begin{aligned} |G(S_{1:n}) - G(S_1, \ldots, \widetilde{S}_i, \ldots, S_n)| &\le \sup_s \pi_i \left|\mathbf{1}(S_i \le s) - \mathbf{1}(\widetilde{S}_i \le s)\right| \\ &\le \pi_i. \end{aligned}$$

Thus the bounded-difference constants are $c_i = \pi_i$. Conditional McDiarmid's inequality gives, for every $t > 0$,

$$P\left\{G - E(G \mid \mathcal{C}) \ge t \,|\, \mathcal{C}\right\} \le \exp\left\{-\frac{2t^2}{\sum_i \pi_i^2}\right\}.$$

Taking

$$t = \sqrt{\frac{1}{2}\log(1/\eta) \sum_{i=1}^n \pi_i^2}$$

and combining this bound with the conditional expectation bound proves the result. □

D.3 Interpretation of the score-approximation condition

Assumption 1 controls the spatial regularity of the conditional score distribution, uniformly over the score cutoff. For the absolute residual score

$$S(x, y) = |y - \widehat{\mu}(x)|,$$

consider the model

$$Y = \mu(X) + \sigma(X)\varepsilon, \qquad \sigma(x) > 0, \qquad \varepsilon \perp X,$$

and write

$$b(x) := \mu(x) - \widehat{\mu}(x).$$

Then $h_s(x) = 0$ for $s < 0$, while for $s \geq 0$,

$$h_s(x) = F_\varepsilon\left(\frac{s - b(x)}{\sigma(x)}\right) - F_\varepsilon\left(\left(\frac{-s - b(x)}{\sigma(x)}\right)^-\right).$$

We next give a sufficient condition for spatial regularity uniform in s. Suppose that ε has density f_ε and

$$M_0 := \sup_u f_\varepsilon(u) < \infty, \qquad M_1 := \sup_u |u| f_\varepsilon(u) < \infty, \qquad \inf_x \sigma(x) \geq \sigma_{\min} > 0.$$

Define

$$H_t(b, \ell) := F_\varepsilon\big((t - b)e^{-\ell}\big), \qquad \ell = \log \sigma.$$

Its derivatives satisfy

$$|\partial_b H_t(b, \ell)| = e^{-\ell} f_\varepsilon\big((t - b)e^{-\ell}\big) \leq \frac{M_0}{\sigma_{\min}},$$

and

$$|\partial_\ell H_t(b, \ell)| = \big|(t - b)e^{-\ell} f_\varepsilon\big((t - b)e^{-\ell}\big)\big| \leq M_1.$$

Since

$$h_s(x) = H_s(b(x), \log \sigma(x)) - H_{-s}(b(x), \log \sigma(x)),$$

the mean-value inequality gives

$$\sup_{s \in \mathbb{R}} |h_s(x) - h_s(y)| \leq \frac{2M_0}{\sigma_{\min}} |b(x) - b(y)| + 2M_1 |\log \sigma(x) - \log \sigma(y)|.$$

In particular, suppose that b and $\log \sigma$ are κ-Hölder continuous: for some $0 < \kappa \leq 1$,

$$|b(x) - b(y)| \leq L_b \|x - y\|^\kappa, \qquad |\log \sigma(x) - \log \sigma(y)| \leq L_\sigma \|x - y\|^\kappa$$

The case $\kappa = 1$ corresponds to Lipschitz continuity. Then

$$\sup_s |h_s(x) - h_s(y)| \leq L_h \|x - y\|^\kappa, \qquad L_h := \frac{2M_0 L_b}{\sigma_{\min}} + 2M_1 L_\sigma.$$

To connect this modulus to Assumption 1, suppose that $h_s|_{\mathcal{X}}$ admits a globally defined extension f_s such that

$$f_s = h_s \text{ on } \mathcal{X}, \qquad \sup_s \|f_s\|_2 \leq M_2, \qquad |f_s(x) - f_s(y)| \leq L_h \|x - y\|^\kappa.$$

Let $\eta \in \mathcal{S}(\mathbb{R}^d)$ be a real-valued Schwartz mollifier satisfying $\int \eta(u)\, du = 1$, and define

$$g_{s,\epsilon} := \eta_\epsilon * f_s, \qquad \eta_\epsilon(x) := \epsilon^{-d} \eta(x/\epsilon).$$

For $x \in \mathcal{X}$,

$$\begin{aligned} |g_{s,\epsilon}(x) - h_s(x)| &= |g_{s,\epsilon}(x) - f_s(x)| \\ &= \left| \int \eta(u) \{f_s(x - \epsilon u) - f_s(x)\}\, du \right| \\ &\leq L_h \epsilon^\kappa \int \|u\|^\kappa |\eta(u)|\, du. \end{aligned}$$

Therefore,

$$\sup_s \sup_{x\in\mathcal{X}} |g_{s,\epsilon}(x) - h_s(x)| \le C_{\eta,\kappa} L_h \epsilon^\kappa =: \varepsilon_{\omega,\epsilon}.$$

If the CFD spectrum satisfies

$$\omega(t)^{-1} \lesssim (1 + \|t\|^2)^\beta,$$

Theorem 5 further gives

$$\sup_s \|g_{s,\epsilon}\|_\omega^\star \lesssim M_2 \epsilon^{-\beta} =: B_{\omega,\epsilon}.$$

Hence the approximation–transfer contribution obeys

$$B_{\omega,\epsilon} D(P,Q) + 2\varepsilon_{\omega,\epsilon} \lesssim M_2 \epsilon^{-\beta} D(P,Q) + L_h \epsilon^\kappa.$$

Thus finer smoothing decreases the uniform approximation error but increases the spectral representation cost. For exponentially decaying spectra, sharp high-frequency approximants can incur a substantially larger cost, as quantified by Theorem 5.

D.4 Extension to unbounded covariate support

Remark 1. The compact-support condition can be replaced by either of the following localization conditions.

Cutoff approximation. Let $\chi_R : \mathbb{R}^d \to [0,1]$ be a measurable cutoff, fixed under the conditioning convention, and suppose that there exist globally admissible functions $g_{s,R}$ such that

$$\sup_s \|g_{s,R}\|_D^\star \le B_{D,R}, \qquad \sup_s \sup_{x\in\mathbb{R}^d} |\chi_R(x) h_s(x) - g_{s,R}(x)| \le \varepsilon_{D,R}.$$

For every s,

$$\begin{aligned} P_{\mathrm{tar}} h_s - \widehat{P}_\pi h_s = P_{\mathrm{tar}}(\chi_R h_s - g_{s,R}) + \{P_{\mathrm{tar}} g_{s,R} - \widehat{P}_\pi g_{s,R}\} \\ + \widehat{P}_\pi(g_{s,R} - \chi_R h_s) + R_{s,R}, \end{aligned}$$

where the signed remainder is

$$R_{s,R} := P_{\mathrm{tar}}\{(1-\chi_R)h_s\} - \widehat{P}_\pi\{(1-\chi_R)h_s\}.$$

Since $0 \le h_s \le 1$,

$$|R_{s,R}| \le P_{\mathrm{tar}}(1-\chi_R) + \widehat{P}_\pi(1-\chi_R) =: T_R.$$

The finite-sample decomposition therefore acquires the additional tail term T_R. For the population result, replace $\widehat{P}_\pi$ in T_R by P_{src}^π and omit the sampling terms.

Local approximation with a global envelope. Alternatively, let $\mathcal{X}_R \subseteq \mathbb{R}^d$ and suppose

$$\sup_s \|g_{s,R}\|_D^\star \le B_{D,R}, \qquad \sup_s \sup_{x\in\mathcal{X}_R} |h_s(x) - g_{s,R}(x)| \le \varepsilon_{D,R}, \qquad \sup_s \sup_{x\in\mathbb{R}^d} |g_{s,R}(x)| \le G_{D,R}.$$

Then, for any probability measure P,

$$\begin{aligned} |P(h_s - g_{s,R})| &\le P|h_s - g_{s,R}| \\ &\le \varepsilon_{D,R} + (1 + G_{D,R}) P(\mathcal{X}_R^c). \end{aligned}$$

Thus the additional finite-sample tail term is at most

$$(1 + G_{D,R}) \left\{ P_{\mathrm{tar}}(\mathcal{X}_R^c) + \widehat{P}_\pi(\mathcal{X}_R^c) \right\}.$$

The cutoff or global envelope controls the contribution of the approximants outside the localization region.

D.5 Proof of Theorem 1

Proof. Throughout the proof, condition on $\mathcal{H}$. For the response-process step, additionally use

$$\mathcal{C} = \mathcal{H} \vee \sigma(X_{1:n}, Z_{1:m}, V_{1:v}),$$

omitting $V_{1:v}$ when no validation sample is used. This fixes the response-free weights while leaving calibration responses conditionally independent. For each $s \in \mathbb{R}$, choose g_s as in Assumption 1. Since

$$F_{\mathrm{tar}}(s) = P_{\mathrm{tar}} h_s, \qquad \widehat{F}_\pi(s) = \sum_i \pi_i \mathbf{1}(S_i \le s),$$

adding and subtracting g_s, $\widehat{Q}_m g_s$, and $\widehat{P}_\pi h_s$ gives

$$\begin{aligned} F_{\mathrm{tar}}(s) - \widehat{F}_\pi(s) = P_{\mathrm{tar}}(h_s - g_s) &+ \left(P_{\mathrm{tar}} g_s - \widehat{Q}_m g_s\right) \\ &+ \left(\widehat{Q}_m g_s - \widehat{P}_\pi g_s\right) + \widehat{P}_\pi(g_s - h_s) \\ &+ \sum_i \pi_i \{h_s(X_i) - \mathbf{1}(S_i \le s)\}. \end{aligned}$$

Hence

$$\begin{aligned} \left|F_{\mathrm{tar}}(s) - \widehat{F}_\pi(s)\right| \le P_{\mathrm{tar}}|h_s - g_s| &+ \left|P_{\mathrm{tar}} g_s - \widehat{Q}_m g_s\right| \\ &+ \left|\widehat{Q}_m g_s - \widehat{P}_\pi g_s\right| + \widehat{P}_\pi |g_s - h_s| \\ &+ \left|\sum_i \pi_i \{\mathbf{1}(S_i \le s) - h_s(X_i)\}\right|. \end{aligned}$$

Because P_{tar} and $\widehat{P}_\pi$ are supported on $\mathcal{X}$, Assumption 1 gives

$$P_{\mathrm{tar}}|h_s - g_s| \le \varepsilon_D, \qquad \widehat{P}_\pi |g_s - h_s| \le \varepsilon_D.$$

The balance–transfer inequality and the symmetry of D give

$$\begin{aligned} \left|P_{\mathrm{tar}} g_s - \widehat{Q}_m g_s\right| &\le \|g_s\|_D^\star D(\widehat{Q}_m, P_{\mathrm{tar}}) \le B_D D(\widehat{Q}_m, P_{\mathrm{tar}}), \\ \left|\widehat{Q}_m g_s - \widehat{P}_\pi g_s\right| &\le \|g_s\|_D^\star D(\widehat{P}_\pi, \widehat{Q}_m) \le B_D D(\widehat{P}_\pi, \widehat{Q}_m). \end{aligned}$$

Define

$$G_\pi = \sup_{s \in \mathbb{R}} \left|\sum_i \pi_i \{\mathbf{1}(S_i \le s) - h_s(X_i)\}\right|.$$

Since all preceding bounds are uniform in s, we obtain the pathwise inequality

$$\sup_{s \in \mathbb{R}} \left|F_{\mathrm{tar}}(s) - \widehat{F}_\pi(s)\right| \le B_D D(\widehat{P}_\pi, \widehat{Q}_m) + 2\varepsilon_D + B_D D(\widehat{Q}_m, P_{\mathrm{tar}}) + G_\pi. \tag{13}$$

Let

$$c_\delta = 4 + \sqrt{\frac{1}{2}\log(2/\delta)}$$

and define

$$E_T = \left\{D(\widehat{Q}_m, P_{\mathrm{tar}}) \le r_{m,D}(\delta/2)\right\}, E_S = \left\{G_\pi \le c_\delta \sqrt{\sum_i \pi_i^2}\right\}.$$

By the target-radius condition (7),

$$P(E_T^c \mid \mathcal{H}) \le \frac{\delta}{2}.$$

Now conditional on $\mathcal{C}$, the response-free weights π are fixed, while the calibration scores remain independent and satisfy

$$P(S_i \le s \mid \mathcal{C}) = h_s(X_i).$$

Lemma 1, applied conditionally on $\mathcal{C}$ with $\eta = \delta/2$, therefore gives

$$P(E_S^c \mid \mathcal{C}) \le \frac{\delta}{2} \qquad \text{almost surely.}$$

Since $\mathcal{H} \subseteq \mathcal{C}$, the tower property yields

$$P(E_S^c \mid \mathcal{H}) = E\left[P(E_S^c \mid \mathcal{C}) \mid \mathcal{H}\right] \le \frac{\delta}{2}.$$

A conditional union bound gives

$$P\big((E_T \cap E_S)^c \mid \mathcal{H}\big) \le P(E_T^c \mid \mathcal{H}) + P(E_S^c \mid \mathcal{H}) \le \delta.$$

Thus, with conditional probability at least $1-\delta$, both events hold. Substituting their bounds into (13) and using

$$\sqrt{\sum_i \pi_i^2} = \frac{1}{\sqrt{\mathrm{ESS}(\pi)}}$$

gives

$$\sup_{s\in\mathbb{R}} \left| F_{\mathrm{tar}}(s) - \widehat{F}_\pi(s) \right| \le B_D D(\widehat{P}_\pi, \widehat{Q}_m) + 2\varepsilon_D + B_D r_{m,D}(\delta/2) + \frac{c_\delta}{\sqrt{\mathrm{ESS}(\pi)}},$$

which is exactly (8). □

D.6 Proof of Corollary 1

Proof. By definition of the generalized quantile, $\widehat{F}_\pi(\widehat{q}_\tau) \ge \tau$. On the event in Theorem 1,

$$F_{\mathrm{tar}}(\widehat{q}_\tau) \ge \widehat{F}_\pi(\widehat{q}_\tau) - \Delta_{\delta,D}(\pi) \ge \tau - \Delta_{\delta,D}(\pi).$$

Conditional on $\mathcal{F}_{\mathrm{cal}}$, the future target pair is independent and $\widehat{q}_\tau$ is fixed. Therefore

$$P\{Y_{n+1} \in \widehat{C}_\tau(X_{n+1}) \mid \mathcal{F}_{\mathrm{cal}}\} = F_{\mathrm{tar}}(\widehat{q}_\tau),$$

which proves the first claim. The inflated-level statement follows by substituting $\tau = 1 - \alpha + \overline{\Delta}$ and using $\overline{\Delta} \ge \Delta_{\delta,D}(\pi)$. □

D.7 Proof of Corollary 2

Proof. Let $c_\delta = 4 + \sqrt{\frac{1}{2}\log(2/\delta)}$. For any $\pi \in \Delta_n(B_w)$, Cauchy–Schwarz gives

$$\begin{aligned}
&B_D D(\widehat{P}_\pi, \widehat{Q}_m) + \frac{c_\delta}{\sqrt{\mathrm{ESS}(\pi)}} \\
&\quad = B_D D(\widehat{P}_\pi, \widehat{Q}_m) + \frac{c_\delta}{\sqrt{\lambda_0 n}} \left(\lambda_0 n \sum_i \pi_i^2\right)^{1/2} \\
&\quad \le \left(B_D^2 + \frac{c_\delta^2}{\lambda_0 n}\right)^{1/2} \left\{ D^2(\widehat{P}_\pi, \widehat{Q}_m) + \lambda_0 n \sum_i \pi_i^2 \right\}^{1/2}.
\end{aligned}$$

Apply this inequality to $\widehat{\pi}_{D,\lambda_0}$ and use its optimality over $\Delta_n(B_w)$. Adding the approximation and target-radius terms in (8) proves (9). □

Numerical fits. If a feasible numerical solution $\widetilde{\pi}$ satisfies $\mathcal{J}_{D,\lambda_0}(\widetilde{\pi}) \le \inf_{\Delta_n(B_w)} \mathcal{J}_{D,\lambda_0} + \xi$, the same proof replaces the infimum in (9) by the infimum plus ξ. For a differentiable convex objective, the computable gap $G(\widetilde{\pi}) = \max_{u\in\Delta_n(B_w)} \nabla\mathcal{J}(\widetilde{\pi})^\top(\widetilde{\pi} - u)$ bounds this optimization error. The independent-unit experiments require $G \le 10^{-7}$.

Comparator-based asymptotic validity. We provide conditions under which the objective-value bound vanishes. Consider a sequence of problems indexed by n, let $\delta_n \downarrow 0$, and write

$$c_n := 4 + \sqrt{\frac{1}{2}\log(2/\delta_n)}, \qquad A_n := \left(B_{D_n}^2 + \frac{c_n^2}{n\lambda_n}\right)^{1/2}.$$

Define $\mathcal{J}_n(\pi) := D_n^2(\widehat{P}_\pi, \widehat{Q}_{m_n}) + \lambda_n n \sum_{i=1}^n \pi_i^2$. Suppose that a feasible numerical solution $\widetilde{\pi}_n$ satisfies

$$\mathcal{J}_n(\widetilde{\pi}_n) \le \inf_{\pi \in \Delta_n(B_w)} \mathcal{J}_n(\pi) + \xi_n.$$

Then, for every feasible comparator $\pi_n^\circ \in \Delta_n(B_w)$, the numerical version of Corollary 2 gives

$$\begin{aligned}\Delta_{\delta_n, D_n}(\widetilde{\pi}_n) \le{}& 2\varepsilon_{D_n} + B_{D_n}\mathfrak{r}_{m_n, D_n}(\delta_n/2) \\ &+ A_n\left\{\mathcal{J}_n(\pi_n^\circ) + \xi_n\right\}^{1/2}. \end{aligned} \tag{14}$$

Consequently, it is sufficient that $\varepsilon_{D_n} \to 0, B_{D_n}\mathfrak{r}_{m_n,D_n}(\delta_n/2) \to 0, A_n\left\{\mathcal{J}_n(\pi_n^\circ) + \xi_n\right\}^{1/2} \xrightarrow{p} 0$.

A more primitive sufficient set of conditions is

$$B_{D_n} = O(1), \qquad \varepsilon_{D_n} \to 0, \qquad B_{D_n}\mathfrak{r}_{m_n,D_n}(\delta_n/2) \to 0,$$

$$\lambda_n \to 0, \qquad \frac{n\lambda_n}{\log(2/\delta_n)} \to \infty,$$

$$D_n(\widehat{P}_{\pi_n^\circ}, \widehat{Q}_{m_n}) = o_p(1), \qquad n\sum_{i=1}^n (\pi_{n,i}^\circ)^2 = O_p(1), \qquad \xi_n = o_p(1).$$

Indeed, $c_n^2 = O\{\log(2/\delta_n)\}$ implies $A_n^2 = O(1)$, while

$$\mathcal{J}_n(\pi_n^\circ) = D_n^2(\widehat{P}_{\pi_n^\circ}, \widehat{Q}_{m_n}) + \lambda_n n \sum_i (\pi_{n,i}^\circ)^2 = o_p(1) + \lambda_n O_p(1) = o_p(1).$$

Equation (14) and Corollary 6 then give asymptotic target validity. Exact optimization corresponds to $\xi_n = 0$.

D.8 Additional consequences of Theorem 1

For a weighted empirical CDF, write

$$\widehat{F}_\pi(q^-) = \lim_{u \uparrow q} \widehat{F}_\pi(u), \qquad J_\pi(q) = \widehat{F}_\pi(q) - \widehat{F}_\pi(q^-).$$

Corollary 3. *On the event in Theorem 1,*

$$\tau - \Delta_{\delta,D}(\pi) \le P\{Y_{n+1} \in \widehat{C}_\tau(X_{n+1}) \mid \mathcal{F}_{\mathrm{cal}}\} \le \tau + \Delta_{\delta,D}(\pi) + J_\pi(\widehat{q}_\tau).$$

If the calibration scores are distinct, $J_\pi(\widehat{q}_\tau) \le \max_i \pi_i$; under the relative cap this is at most B_w/n.

Proof. The lower bound is Corollary 1. By definition of the generalized quantile, $\widehat{F}_\pi(\widehat{q}_\tau^-) < \tau \le \widehat{F}_\pi(\widehat{q}_\tau)$, and hence $\widehat{F}_\pi(\widehat{q}_\tau) < \tau + J_\pi(\widehat{q}_\tau)$. The uniform event in Theorem 1 then yields the upper bound. With distinct calibration scores, each jump equals the weight of a single calibration observation. □

Corollary 4. *Let $0 < \tau \le 1$, and define*

$$\gamma_\tau := \frac{\tau - \widehat{F}_\pi(\widehat{q}_\tau^-)}{J_\pi(\widehat{q}_\tau)} \in (0, 1].$$

Let $U \sim \mathrm{Unif}(0,1)$ be independent of $\mathcal{F}_{\mathrm{cal}}$ and the future target observation, and define

$$\widehat{C}_\tau^U(x) := \{y : S(x,y) < \widehat{q}_\tau\} \cup \{y : S(x,y) = \widehat{q}_\tau,\ U \le \gamma_\tau\}.$$

On the event in Theorem 1,

$$\left| P\left\{Y_{n+1} \in \widehat{C}_\tau^U(X_{n+1}) \,\middle|\, \mathcal{F}_{\mathrm{cal}}\right\} - \tau \right| \le \Delta_{\delta,D}(\pi),$$

where the conditional probability averages over U.

Proof. On the event of Theorem 1, write $q = \widehat{q}_\tau$, $\gamma = \gamma_\tau$, and $\Delta = \Delta_{\delta,D}(\pi)$. The uniform CDF bound also holds at left limits, so both $|F_{\mathrm{tar}}(q) - \widehat{F}_\pi(q)|$ and $|F_{\mathrm{tar}}(q^-) - \widehat{F}_\pi(q^-)|$ are at most Δ. Conditional on $\mathcal{F}_{\mathrm{cal}}$, independence of U and the future target score gives acceptance probability

$$A_{\mathrm{tar}} = (1-\gamma)F_{\mathrm{tar}}(q^-) + \gamma F_{\mathrm{tar}}(q).$$

By the definition of γ, the corresponding empirical mixture equals

$$(1-\gamma)\widehat{F}_\pi(q^-) + \gamma\widehat{F}_\pi(q) = \widehat{F}_\pi(q^-) + \gamma J_\pi(q) = \tau.$$

Subtracting these mixtures yields $|A_{\mathrm{tar}} - \tau| \le (1-\gamma)\Delta + \gamma\Delta = \Delta$. □

Corollary 5. *If $\pi_i \le B_w/n$ and $\sum_i \pi_i = 1$, then*

$$\mathrm{ESS}(\pi) = \frac{1}{\sum_i \pi_i^2} \ge \frac{n}{B_w}.$$

Consequently, the weighted calibration-fluctuation term in Theorem 1 is at most

$$\left\{4 + \sqrt{\tfrac{1}{2}\log(2/\delta)}\right\}\sqrt{\frac{B_w}{n}}.$$

Proof. Because $\sum_i \pi_i = 1$,

$$\sum_i \pi_i^2 \le \Big(\max_i \pi_i\Big)\sum_i \pi_i \le \frac{B_w}{n}.$$

Taking reciprocals gives $\mathrm{ESS}(\pi) \ge n/B_w$. Substitution into the final term of (8) proves the stated fluctuation bound. □

Corollary 6. *Consider a sequence of distributionally balanced calibration procedures satisfying the conditions of Theorem 1. Let $\delta_n \downarrow 0$ and suppose that*

$$\Delta_{\delta_n, D_n}(\widehat{\pi}_n) \xrightarrow{p} 0.$$

For the prediction set calibrated at level $1-\alpha$, let

$$C_n := P\Big\{Y_{n+1} \in \widehat{C}_{1-\alpha,n}(X_{n+1}) \,\Big|\, \mathcal{F}_{\mathrm{cal},n}\Big\}.$$

Then

$$[1-\alpha-C_n]_+ = o_p(1), \qquad \liminf_{n\to\infty} E(C_n) \ge 1-\alpha.$$

Proof. Write $\Delta_n = \Delta_{\delta_n,D_n}(\widehat{\pi}_n)$ and $L_n = [1-\alpha-C_n]_+$. The event E_n in Theorem 1 satisfies $P(E_n^c) \le \delta_n$ by the tower property, and Corollary 1 gives $L_n \le \Delta_n$ on E_n. Thus, for every $\varepsilon > 0$,

$$P(L_n > \varepsilon) \le P(\Delta_n > \varepsilon) + \delta_n \longrightarrow 0.$$

Because $0 \le L_n \le 1$, also $EL_n \le \varepsilon + P(L_n > \varepsilon) \to 0$ after taking the upper limit and then $\varepsilon \downarrow 0$. Finally, $EC_n \ge 1-\alpha-EL_n$, proving the unconditional coverage bound. □

D.9 CFD DUALITY AND ITS TRANSFER NORM

We use the unitary Fourier transform

$$\widehat{f}(t) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} f(x)e^{-it^\top x}\,dx$$

and, for globally defined functions for which the integral is finite, set

$$\|f\|_\omega^\star = (2\pi)^{-d/2}\left\{\int_{\mathbb{R}^d}\frac{|\widehat{f}(t)|^2}{\omega(t)}\,dt\right\}^{1/2}. \tag{15}$$

For an integrable positive spectrum this is the corresponding RKHS norm under the stated Fourier convention while for energy CFD it is a homogeneous seminorm. The approximants in Assumption 1 are globally defined and evaluated on the covariate support.

Proposition 2. *Let P and Q be probability laws for which (2) is finite. If f admits absolutely convergent Fourier inversion and $\|f\|_\omega^\star < \infty$, then*

$$|E_P f(X) - E_Q f(X)| \le \|f\|_\omega^\star \mathrm{CFD}_\omega(P, Q).$$

Thus $(\mathrm{CFD}_\omega, \|\cdot\|_\omega^\star)$ is a balance–transfer pair. Theorem 1 applies with $D = \mathrm{CFD}_\omega$, $B_D = B_\omega$, and $\varepsilon_D = \varepsilon_\omega$.

Proof. Let $\nu = P - Q$. Under the stated Fourier-inversion and integrability conditions,

$$\int_{\mathbb{R}^d} f(x)\, d\nu(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{f}(t)\{\varphi_P(t) - \varphi_Q(t)\}\, dt.$$

Inserting $\omega(t)^{-1/2}\omega(t)^{1/2}$ and applying Cauchy–Schwarz gives

$$\begin{aligned}\left|\int f\, d\nu\right| &\le (2\pi)^{-d/2} \left\{\int \frac{|\widehat{f}(t)|^2}{\omega(t)}\, dt\right\}^{1/2} \left\{\int |\varphi_P(t) - \varphi_Q(t)|^2 \omega(t)\, dt\right\}^{1/2} \\ &= \|f\|_\omega^\star \mathrm{CFD}_\omega(P, Q).\end{aligned}$$

The stated integrability assumptions justify the interchange of integrals and the use of Fourier inversion. □

D.10 Target empirical-CFD concentration for a finite spectral measure

Lemma 2. *Under the conditioning convention stated above, suppose that*

$$\Omega_\omega := \int_{\mathbb{R}^d} \omega(t)\, dt < \infty.$$

Then, for every $\eta \in (0, 1)$,

$$P\left\{\mathrm{CFD}_\omega(\widehat{Q}_m, P_{\mathrm{tar}}) \le \sqrt{\frac{\Omega_\omega}{m}}\left\{1 + \sqrt{2\log(1/\eta)}\right\} \,\middle|\, \mathcal{H}\right\} \ge 1 - \eta.$$

Proof. Condition on $\mathcal{H}$. The case $\Omega_\omega = 0$ is immediate. Otherwise use the feature map $\Phi_\omega(x)(t) = e^{\mathrm{i}t^\top x}$ in $\mathcal{L}_\omega = L^2(\mathbb{R}^d, \omega(t)\, dt; \mathbb{C})$. Its squared norm is Ω_ω, and dominated convergence makes it continuous, hence measurable. Its Hilbert-space expectation $\mu_\omega = E\Phi_\omega(Z)$ therefore exists. For

$$T = \left\| m^{-1} \sum_{j=1}^m \Phi_\omega(Z_j) - \mu_\omega \right\|_{\mathcal{L}_\omega} = \mathrm{CFD}_\omega(\widehat{Q}_m, P_{\mathrm{tar}}),$$

independence and centering give

$$ET \le (ET^2)^{1/2} = \left\{\frac{\Omega_\omega - \|\mu_\omega\|_{\mathcal{L}_\omega}^2}{m}\right\}^{1/2} \le \sqrt{\frac{\Omega_\omega}{m}}.$$

Replacing one observation changes T by at most $2\sqrt{\Omega_\omega}/m$. McDiarmid's inequality therefore gives

$$P(T - ET \ge u \mid \mathcal{H}) \le \exp\{-mu^2/(2\Omega_\omega)\}.$$

Set $u = \sqrt{2\Omega_\omega \log(1/\eta)/m}$ and combine the two bounds. □

Corollary 7. *If $\Omega_\omega = \int \omega(t)\, dt < \infty$, then (7) holds with*

$$\mathfrak{r}_{m,\omega}(\eta) = \sqrt{\frac{\Omega_\omega}{m}}\{1 + \sqrt{2\log(1/\eta)}\}.$$

For a normalized frequency density, the target-sampling contribution to Theorem 1 is $B_\omega\{1 + \sqrt{2\log(2/\delta)}\}/\sqrt{m}$.

Proof. Use Lemma 2 in the target-radius condition (7). If $\Omega_\omega = 1$, substitute $\eta = \delta/2$ in Theorem 1. □

E RESPONSE-FREE SELECTION AND COMMON-GEOMETRY VALIDATION

E.1 COMMON VALIDATION AND ITS SCORE-DISTRIBUTION GUARANTEE

Condition on independent training information $\mathcal{H}$, which fixes the score and a measurable positive-semidefinite kernel k_0 with separable RKHS satisfying $k_0(x,x) \le \kappa_0^2$. Write $D_0 = \mathrm{MMD}_{k_0}$ and $\mu_0(P)$ for its kernel mean embedding. Apply (12) to the response-free candidate set $\mathcal{A}_\rho$, including the uniform weights, using an independent target validation sample of size v.

Theorem 3. *Write $Q = P_{\mathrm{tar},X}$, and define*

$$r_{v,0}(\eta) := \frac{\kappa_0}{\sqrt{v}} \left\{ 1 + \sqrt{2\log(1/\eta)} \right\}.$$

Conditional on $\mathcal{H}$ and the calibration and balancing covariates, with probability at least $1-\eta$ over the independent validation covariates,

$$D_0\Big(\widehat{P}_{\pi^{(\widehat{a})}}, Q\Big) \le \min_{a \in \mathcal{A}_\rho} D_0\Big(\widehat{P}_{\pi^{(a)}}, Q\Big) + 2r_{v,0}(\eta). \tag{16}$$

Suppose in addition that Assumption 1 holds for $(D_0, \|\cdot\|_{\mathcal{H}_0})$ with constants (B_0, ε_0). Then, conditional on $\mathcal{H}$, with probability at least $1-\delta$ jointly over the calibration, balancing, and validation observations,

$$\sup_s \left| F_{\mathrm{tar}}(s) - \widehat{F}_{\pi^{(\widehat{a})}}(s) \right| \le B_0 D_0\Big(\widehat{P}_{\pi^{(\widehat{a})}}, \widehat{Q}_v^{\mathrm{val}}\Big) + 2\varepsilon_0 + B_0 r_{v,0}(\delta/2) + \frac{c_\delta}{\sqrt{\mathrm{ESS}(\pi^{(\widehat{a})})}} \tag{17}$$

$$\le B_0 \min_{a \in \mathcal{A}_\rho} D_0\Big(\widehat{P}_{\pi^{(a)}}, Q\Big) + 2\varepsilon_0 + 2B_0 r_{v,0}(\delta/2) + \frac{c_\delta}{\sqrt{\rho n}}, \tag{18}$$

where

$$c_\delta = 4 + \sqrt{\frac{1}{2}\log(2/\delta)}.$$

Proof. Condition first on $\mathcal{H}$ and the calibration and balancing covariates. The validation observations remain iid from Q. Let ϕ_0 be the feature map of k_0, $\widehat{\mu}_v = v^{-1}\sum_j \phi_0(V_j)$, $\mu = \mu_0(Q)$, and $T_v = \|\widehat{\mu}_v - \mu\|_{\mathcal{H}_0}$. The bounded-Hilbert-space calculation in Lemma 2, now using $\|\phi_0(x)\| \le \kappa_0$, gives conditional mean at most $\kappa_0/\sqrt{v}$ and change at most $2\kappa_0/v$ when one validation observation is replaced. McDiarmid's inequality therefore gives an event $\mathcal{E}_v(\eta) = \{T_v \le r_{v,0}(\eta)\}$ with conditional probability at least $1-\eta$.

On this event, the reverse triangle inequality gives, simultaneously for every probability measure P,

$$\left| D_0(P, \widehat{Q}_v^{\mathrm{val}}) - D_0(P, Q) \right| \le \|\widehat{\mu}_v - \mu\|_{\mathcal{H}_0} \le r_{v,0}(\eta). \tag{19}$$

Since the admissible set is nonempty, the selection rule implies

$$\begin{aligned} D_0(\widehat{P}_{\pi^{(\widehat{a})}}, Q) &\le D_0(\widehat{P}_{\pi^{(\widehat{a})}}, \widehat{Q}_v^{\mathrm{val}}) + r_{v,0}(\eta) \\ &= \min_{a \in \mathcal{A}_\rho} D_0(\widehat{P}_{\pi^{(a)}}, \widehat{Q}_v^{\mathrm{val}}) + r_{v,0}(\eta) \\ &\le \min_{a \in \mathcal{A}_\rho} D_0(\widehat{P}_{\pi^{(a)}}, Q) + 2r_{v,0}(\eta), \end{aligned}$$

proving (16).

For the score-CDF bound, condition on $\mathcal{G} = \sigma(\mathcal{H}, X_{1:n}, Z_{1:m}, V_{1:v})$. The selected response-free weights $\widehat{\pi} = \pi^{(\widehat{a})}$ are fixed, and calibration scores remain independent with conditional CDFs $h_s(X_i)$. Lemma 1 gives an event $\mathcal{E}_S$, with conditional failure probability at most $\delta/2$, on which

$$\sup_s |\widehat{F}_{\widehat{\pi}}(s) - \widehat{P}_{\widehat{\pi}} h_s| \le c_\delta / \sqrt{\mathrm{ESS}(\widehat{\pi})}.$$

Use the pathwise decomposition (13) with D_0, (B_0, ε_0), and $\widehat{Q}_v^{\mathrm{val}}$ in place of the balancing empirical law. On $\mathcal{E}_v(\delta/2) \cap \mathcal{E}_S$, it yields (17). The tower property and a union bound give conditional probability at least $1-\delta$ given $\mathcal{H}$. Finally, (19) and selection give

$$D_0(\widehat{P}_{\widehat{\pi}}, \widehat{Q}_v^{\mathrm{val}}) \le \min_{a \in \mathcal{A}_\rho} D_0(\widehat{P}_{\pi^{(a)}}, Q) + r_{v,0}(\delta/2).$$

Substitute this and $\mathrm{ESS}(\widehat{\pi}) \ge \rho n$ into (17) to obtain (18). □

Mixture geometry and score approximation. Suppose

$$k_0 = \sum_{j=1}^{J} w_j k_j, \qquad w_j > 0, \qquad \sum_{j=1}^{J} w_j = 1.$$

Then

$$D_0^2(P,Q) = \sum_{j=1}^{J} w_j D_j^2(P,Q), \qquad D_j(P,Q) \le \frac{D_0(P,Q)}{\sqrt{w_j}}.$$

Hence an approximation satisfying

$$\sup_s \|g_s\|_{\mathcal{H}_j} \le B_j, \qquad \sup_s \sup_{x\in\mathcal{X}} |g_s(x) - h_s(x)| \le \varepsilon_j$$

also satisfies

$$|Pg_s - Qg_s| \le B_j D_j(P,Q) \le \frac{B_j}{\sqrt{w_j}} D_0(P,Q),$$

with the same approximation error ε_j. For the equal mixture $w_j = 1/J$, the resulting multiplier is $\sqrt{J}$: it is $\sqrt{15}$ for the nonlinear and diabetes libraries and $\sqrt{8}$ for the spectral adaptive library. Thus the validation concentration event has no explicit $|\mathcal{A}_\rho|$-penalty, but changing the common mixture may change the approximation envelope B_0. The oracle inequality (16) concerns balance in the $\mathcal{H}$-fixed common geometry, whereas the calibration bound also depends on (B_0, ε_0).

Remark 2 (A general common validation discrepancy). The population-balance argument does not require a kernel: if a training-fixed pseudometric D_0 satisfies $D_0(\widehat{Q}_v^{\mathrm{val}}, Q) \le r_v$ on an event, its reverse triangle inequality gives

$$\sup_P \left| D_0(P, \widehat{Q}_v^{\mathrm{val}}) - D_0(P,Q) \right| \le r_v.$$

The same selection rule then has population discrepancy at most the best fitted candidate's discrepancy plus $2r_v$, on that event. A score-CDF conclusion additionally requires a balance–transfer pair and the stated score-approximation assumptions. The explicit $v^{-1/2}$ radius in Theorem 3 follows from bounded-kernel MMD concentration; other discrepancies require their own concentration bounds.

E.2 Selection analyzed in each candidate's own geometry

The following complementary result applies when the selected candidate is analyzed with its own discrepancy. Its target-radius term uses a finite-library union bound.

Theorem 4. *Conditional on the independent training information, let $\{(D_\ell, \|\cdot\|_\ell^\star) : \ell = 1, \dots, L\}$ be a fixed finite library of balance–transfer pairs. For each ℓ, let $\pi^{(\ell)}$ be a nonnegative normalized weight vector measurable with respect to the calibration and balancing target covariates. Suppose that, on the common compact support $\mathcal{X}$,*

$$\sup_s \|g_{s,\ell}\|_\ell^\star \le B_\ell, \qquad \sup_s \sup_{x\in\mathcal{X}} |h_s(x) - g_{s,\ell}(x)| \le \varepsilon_\ell.$$

Let $\widehat{\ell}$ be any response-free selection rule measurable with respect to $\mathcal{H} \vee \sigma(X_{1:n}, Z_{1:m}, V_{1:v})$, where the optional validation sample is independent as in Section 2. Candidate weights remain fitted from calibration and balancing covariates only. Then, for every $\delta \in (0,1)$, conditional on the training information, with probability at least $1-\delta$,

$$\sup_s \left| F_{\mathrm{tar}}(s) - \widehat{F}_{\pi^{(\widehat{\ell})}}(s) \right| \le B_{\widehat{\ell}} D_{\widehat{\ell}} \left(\widehat{P}_{\pi^{(\widehat{\ell})}}, \widehat{Q}_m \right) + 2\varepsilon_{\widehat{\ell}} + B_{\widehat{\ell}} \mathfrak{r}_{m, D_{\widehat{\ell}}} \left(\frac{\delta}{2L} \right) + \frac{4 + \sqrt{\frac{1}{2}\log(2/\delta)}}{\sqrt{\mathrm{ESS}(\pi^{(\widehat{\ell})})}}.$$

If every candidate is an integrable CFD with normalized spectral density, then

$$B_{\widehat{\ell}} \mathfrak{r}_{m, D_{\widehat{\ell}}} \left(\frac{\delta}{2L} \right) \le \frac{B_{\widehat{\ell}}}{\sqrt{m}} \left\{ 1 + \sqrt{2\log(2L/\delta)} \right\}.$$

Proof. Apply the target-radius condition (7) to each D_ℓ at failure probability $\delta/(2L)$. A union bound gives an event $\mathcal{E}_T$, with conditional probability at least $1-\delta/2$ given $\mathcal{H}$, on which

$$D_\ell(\widehat{Q}_m, P_{\mathrm{tar}}) \le \mathfrak{r}_{m,D_\ell}\{\delta/(2L)\}, \qquad \ell = 1, \dots, L.$$

This includes the data-selected index $\widehat{\ell}$.

Condition on $\mathcal{C} = \mathcal{H} \vee \sigma(X_{1:n}, Z_{1:m}, V_{1:v})$, omitting unused validation covariates. Response-free selection fixes $\widehat{\ell}$ and $\pi^{(\widehat{\ell})}$ while preserving conditional independence of calibration responses. Lemma 1, applied once at failure probability $\delta/2$, bounds the selected score-process error by

$$\frac{4 + \sqrt{\frac{1}{2}\log(2/\delta)}}{\sqrt{\mathrm{ESS}(\pi^{(\widehat{\ell})})}}.$$

The tower property and a union bound combine this event with $\mathcal{E}_T$ at conditional probability at least $1-\delta$. Substitution into (13) for the selected balance–transfer pair proves the first bound. Corollary 7, with $\Omega_\omega = 1$ and $\eta = \delta/(2L)$, gives the explicit CFD radius. □

The theorem applies to held-out imbalance, ESS constraints, caps, and other covariate-only selection rules. Each selected discrepancy carries its own approximation envelope and error, as shown explicitly in the bound.

F Spectral resolution and Fourier analysis

This appendix proves the calibration-level result in Section 4.4, then gives the general spectral norm comparison and examples. Kernel length scales remain fixed in the empirical-optimizer theorem.

F.1 Circular shift model and restricted kernel spectra

Write P for the uniform angular law, Q_K for the target angular law with density $r_K = 1 + a g_K$ relative to P, and $g_K(\theta) = \cos(K\theta)$. We set $a = 0.8$ in the coverage calculations. The kernels are the ordinary kernels on $\mathbb{R}^2$,

$$k_{G,\ell}(x, x') = e^{-\|x-x'\|_2^2/(2\ell^2)}, \qquad k_{M,\ell}(x, x') = (1 + \sqrt{3}\|x - x'\|_2/\ell)e^{-\sqrt{3}\|x-x'\|_2/\ell},$$

restricted to $X(\theta) = (\cos\theta, \sin\theta)$. Each has diagonal one. The main theorem uses $\ell = 1$, and the proof below also applies to any fixed positive length scale for each family.

Overlap and score identities. Orthogonality gives

$$P g_K = 0, \quad P g_K^2 = \tfrac{1}{2}, \quad Q_K g_K = \tfrac{a}{2}, \quad P r_K^2 = 1 + \tfrac{a^2}{2}, \quad \mathrm{TV}(P, Q_K) = \tfrac{a}{\pi}.$$

Since $K\theta$ modulo 2π is uniform, the distributions of g_K and r_K under P do not depend on the positive integer K. Under Q_K, they are the corresponding size-biased distributions and are also invariant in K. For $K \ge 3$, all first- and second-order Cartesian covariate moments agree between the two laws, by trigonometric orthogonality. The response law is shared between domains and the conditional mean is zero.

For a fixed center c, let $A_j(s; c)$ denote the distribution function of $|\xi(U + 2j) - c|$, for $j \in \{0, 1\}$. Then

$$h_{s,c}(\theta) = \frac{A_0(s; c) + A_1(s; c)}{2} + \frac{A_1(s; c) - A_0(s; c)}{2} g_K(\theta). \tag{20}$$

The coefficient of g_K has absolute value at most $1/2$, uniformly in s and c. The source hard-component probability is $1/2$, whereas its target value is $p_a = 1/2 + a/4$. Therefore

$$F_{P,c}(s) = \tfrac{1}{2}A_0(s; c) + \tfrac{1}{2}A_1(s; c), \qquad F_{Q,c}(s) = (1 - p_a)A_0(s; c) + p_a A_1(s; c).$$

For each fixed c, both laws are continuous and independent of K. At $c = 0$, A_0 and A_1 are the CDFs of $\mathrm{Unif}(0, 1)$ and $\mathrm{Unif}(2, 3)$. For $0 < \alpha < 1/2$, the source $(1-\alpha)$ quantile is $3 - 2\alpha$ and its target coverage is $1 - \alpha - a\alpha/2$. At $a = 0.8$ and $\alpha = 0.1$ these are 2.8 and 0.86, and the target quantile is $3 - 0.1/0.7$.

Lemma 3. *Let P be the uniform probability measure on the circle*

$$X(\theta) = (\cos\theta, \sin\theta), \qquad \theta \in [-\pi, \pi),$$

and let

$$(T_k f)(\theta) := \int k\{X(\theta), X(\varphi)\} f(\varphi)\, dP(\varphi).$$

For a radial kernel, define its restriction

$$\kappa_k(t) := k\{X(t), X(0)\}, \qquad \mu_{k,j} := \frac{1}{\pi}\int_0^\pi \kappa_k(t)\cos(jt)\, dt, \qquad j \geq 0.$$

Then 1 is an eigenfunction of T_k with eigenvalue $\mu_{k,0}$, while, for every $j \geq 1$, $\cos(j\theta)$ and $\sin(j\theta)$ are eigenfunctions with the common eigenvalue $\mu_{k,j}$.

For the Gaussian and Matérn-$3/2$ kernels with fixed length scale $\ell > 0$,

$$\mu_{G,j} = e^{-z} I_j(z), \qquad z = \ell^{-2},$$

and

$$\mu_{M,j} \sim \frac{6\sqrt{3}}{\pi \ell^3}\, j^{-4}, \qquad j \to \infty,$$

where

$$I_j(z) = \sum_{q=0}^{\infty} \frac{(z/2)^{2q+j}}{q!(q+j)!}.$$

Moreover,

$$\mu_{G,j} > 0, \qquad \mu_{M,j} > 0, \qquad j \geq 0.$$

Proof. Radiality and rotational invariance give

$$k\{X(\theta), X(\varphi)\} = \kappa_k(\theta - \varphi), \qquad \kappa_k(-t) = \kappa_k(t).$$

Hence, for $j \geq 1$,

$$\begin{aligned} T_k\{\cos(j\cdot)\}(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \kappa_k(t)\cos\{j(\theta - t)\}\, dt \\ &= \cos(j\theta)\frac{1}{2\pi}\int_{-\pi}^{\pi} \kappa_k(t)\cos(jt)\, dt \\ &\quad + \sin(j\theta)\frac{1}{2\pi}\int_{-\pi}^{\pi} \kappa_k(t)\sin(jt)\, dt \\ &= \mu_{k,j}\cos(j\theta). \end{aligned}$$

The sine integral vanishes because κ_k is even. The same calculation gives

$$T_k\{\sin(j\cdot)\}(\theta) = \mu_{k,j}\sin(j\theta),$$

and the case $j = 0$ follows similarly.

For the Gaussian kernel,

$$\|X(t) - X(0)\|^2 = 2 - 2\cos t,$$

so, with $z = \ell^{-2}$,

$$\kappa_G(t) = \exp\left\{-\frac{1 - \cos t}{\ell^2}\right\} = e^{-z}e^{z\cos t}.$$

Using

$$e^{z\cos t} = I_0(z) + 2\sum_{j=1}^{\infty} I_j(z)\cos(jt),$$

orthogonality of the cosine modes yields

$$\mu_{G,j} = \frac{e^{-z}}{\pi}\int_0^\pi e^{z\cos t}\cos(jt)\, dt = e^{-z}I_j(z).$$

Furthermore,

$$I_j(z) = \Big(\frac{z}{2}\Big)^j \sum_{q=0}^{\infty} \frac{(z^2/4)^q}{q!(q+j)!} \\ \le \frac{(z/2)^j}{j!} \sum_{q=0}^{\infty} \frac{(z^2/4)^q}{q!} = e^{z^2/4} \frac{(z/2)^j}{j!},$$

and therefore

$$0 < \mu_{G,j} \le e^{-z+z^2/4} \frac{(z/2)^j}{j!}.$$

For the Matérn-$3/2$ kernel, put

$$b := \frac{\sqrt{3}}{\ell}, \qquad d(t) := \|X(t) - X(0)\| = 2\sin(t/2), \qquad t \in [0, \pi],$$

and write

$$f(t) := \kappa_M(t) = \{1 + bd(t)\}e^{-bd(t)}.$$

Since

$$d(t) = t - \frac{t^3}{24} + O(t^5)$$

and

$$(1+u)e^{-u} = 1 - \frac{u^2}{2} + \frac{u^3}{3} + O(u^4),$$

we have, as $t \downarrow 0$,

$$f(t) = 1 - \frac{b^2t^2}{2} + \frac{b^3t^3}{3} + O(t^4).$$

Consequently,

$$f'(0) = 0, \qquad f'''(0) = 2b^3.$$

At the other endpoint,

$$d'(\pi) = d'''(\pi) = 0,$$

and the chain rule gives

$$f'(\pi) = f'''(\pi) = 0.$$

For $j \ge 1$, four integrations by parts now give

$$\begin{aligned} \int_0^\pi f(t)\cos(jt)\,dt &= -\frac{1}{j}\int_0^\pi f'(t)\sin(jt)\,dt \\ &= -\frac{1}{j^2}\int_0^\pi f''(t)\cos(jt)\,dt \\ &= \frac{1}{j^3}\int_0^\pi f'''(t)\sin(jt)\,dt \\ &= \frac{f'''(0) - (-1)^j f'''(\pi) + \int_0^\pi f^{(4)}(t)\cos(jt)\,dt}{j^4} \\ &= \frac{2b^3 + \int_0^\pi f^{(4)}(t)\cos(jt)\,dt}{j^4}. \end{aligned}$$

Since $f^{(4)} \in L^1([0,\pi])$, the Riemann–Lebesgue lemma implies

$$\int_0^\pi f^{(4)}(t)\cos(jt)\,dt = o(1).$$

Thus

$$\mu_{M,j} = \frac{1}{\pi}\int_0^\pi f(t)\cos(jt)\,dt \sim \frac{2b^3}{\pi} j^{-4} = \frac{6\sqrt{3}}{\pi\ell^3} j^{-4}.$$

It remains to verify strict positivity for the Matérn modes. For $j \geq 1$, let

$$g_j\{X(\theta)\} := \cos(j\theta), \qquad d\nu_j := g_j \, dP.$$

Then

$$\left\| \int k_M(\,\cdot\,, x)\, d\nu_j(x) \right\|_{\mathcal{H}_M}^2 = \iint k_M(x, x')\, d\nu_j(x)\, d\nu_j(x')$$
$$= \langle g_j, T_M g_j \rangle_{L^2(P)} = \mu_{M,j} \|g_j\|_{L^2(P)}^2 = \frac{\mu_{M,j}}{2}.$$

The Matérn spectral density is strictly positive on $\mathbb{R}^2$. Hence the squared norm above can vanish only if the Fourier transform of ν_j vanishes identically. Fourier uniqueness would then imply $\nu_j = 0$, contradicting $g_j \not\equiv 0$. Therefore

$$\mu_{M,j} > 0, \qquad j \geq 1.$$

Finally,

$$\mu_{M,0} = \frac{1}{\pi} \int_0^\pi \kappa_M(t)\, dt > 0.$$

Together with the positive Bessel-series representation for $\mu_{G,j}$, this proves the claim. □

For any two probability measures on the circle, their difference ν has zero total mass. The kernel Fourier expansion gives

$$D_k^2(P_1, P_2) = 2 \sum_{j \geq 1} \mu_{k,j} \{(\nu \cos(j\theta))^2 + (\nu \sin(j\theta))^2\}. \tag{21}$$

The expansion is absolutely convergent: nonnegative Fourier coefficients sum to the diagonal value one. In particular, with $\widehat{z} = \sum_i \pi_i g_K(\theta_i)$,

$$D_k^2(\widehat{P}_\pi, Q_K) \geq 2\mu_{k,K}(\widehat{z} - a/2)^2, \qquad D_k^2(P, Q_K) = a^2 \mu_{k,K}/2. \tag{22}$$

F.2 Population regularization and coverage

For $\lambda > 0$, consider

$$\mathcal{J}_{k,\lambda}(w) := D_k^2(wP, Q_K) + \lambda P w^2, \qquad Pw = 1, \qquad 0 \leq w \leq 5.$$

Write the $L^2(P)$ Fourier expansion as

$$w(\theta) = 1 + \sum_{j \geq 1} \{b_j \cos(j\theta) + d_j \sin(j\theta)\},$$

where

$$b_j = 2P\{w \cos(j\theta)\}, \qquad d_j = 2P\{w \sin(j\theta)\}.$$

Since $dQ_K/dP = 1 + a\cos(K\theta)$, the signed measure $\nu = wP - Q_K$ satisfies

$$\nu \cos(j\theta) = \frac{1}{2}\{b_j - a\mathbf{1}_{j=K}\}, \qquad \nu \sin(j\theta) = \frac{1}{2} d_j.$$

Therefore, (21) and Parseval's identity give

$$D_k^2(wP, Q_K) = \frac{1}{2} \sum_{j \geq 1} \mu_{k,j} \left\{ (b_j - a\mathbf{1}_{j=K})^2 + d_j^2 \right\},$$
$$Pw^2 = 1 + \frac{1}{2} \sum_{j \geq 1} (b_j^2 + d_j^2),$$

and hence

$$\mathcal{J}_{k,\lambda}(w) = \lambda + \frac{1}{2} \sum_{j \geq 1} \left[\mu_{k,j} \left\{ (b_j - a\mathbf{1}_{j=K})^2 + d_j^2 \right\} + \lambda(b_j^2 + d_j^2) \right].$$

First minimize this quadratic objective under $Pw = 1$, temporarily dropping the pointwise constraints. For $j \neq K$,

$$b_j = d_j = 0,$$

while the K-th cosine coefficient minimizes

$$\mu_{k,K}(b_K - a)^2 + \lambda b_K^2.$$

Thus

$$(\mu_{k,K} + \lambda)b_K = a\mu_{k,K}, \qquad b_K = b_{k,\lambda} := \frac{a\mu_{k,K}}{\mu_{k,K} + \lambda}.$$

The resulting weight is

$$w_{k,\lambda}(\theta) = 1 + b_{k,\lambda}\cos(K\theta).$$

Because

$$0 < b_{k,\lambda} < a < 1,$$

it satisfies

$$0 < 1 - b_{k,\lambda} \leq w_{k,\lambda}(\theta) \leq 1 + b_{k,\lambda} < 2 < 5.$$

Hence it is feasible and is also the unique minimizer of the constrained population problem.

We next compute its induced score quantile. Under the weighted source law $w_{k,\lambda}P$, the hard-component probability is

$$\begin{aligned} p_{k,\lambda} &:= (w_{k,\lambda}P)(B = 1) \\ &= P\left[\{1 + b_{k,\lambda}g_K\}\frac{1 + g_K}{2}\right] \\ &= \frac{1}{2} + \frac{b_{k,\lambda}}{4}, \end{aligned}$$

where $Pg_K = 0$ and $Pg_K^2 = 1/2$. For zero-centered scores and $q \in [2, 3]$,

$$F_{w_{k,\lambda}P,0}(q) = (1 - p_{k,\lambda}) + p_{k,\lambda}(q - 2) = 1 - p_{k,\lambda}(3 - q).$$

Since $p_{k,\lambda} \geq 1/2$, for $0 < \alpha < 1/2$ the $(1 - \alpha)$-quantile lies in $[2, 3]$ and equals

$$q_{k,\lambda}(1 - \alpha) = 3 - \frac{\alpha}{p_{k,\lambda}} = 3 - \frac{\alpha}{1/2 + b_{k,\lambda}/4}.$$

Under Q_K, the hard-component probability is

$$p_a = \frac{1}{2} + \frac{a}{4}.$$

Evaluating the preceding threshold under the target law gives

$$\begin{aligned} F_{Q_K,0}\{q_{k,\lambda}(1 - \alpha)\} &= 1 - p_a\,[3 - q_{k,\lambda}(1 - \alpha)] \\ &= 1 - \alpha\frac{p_a}{p_{k,\lambda}} \\ &= 1 - \alpha - \frac{\alpha(a - b_{k,\lambda})}{2 + b_{k,\lambda}}. \end{aligned}$$

Finally,

$$a - b_{k,\lambda} = \frac{a\lambda}{\mu_{k,K} + \lambda}, \qquad 2 + b_{k,\lambda} = \frac{2\lambda + (2 + a)\mu_{k,K}}{\mu_{k,K} + \lambda},$$

so

$$F_{Q_K,0}\{q_{k,\lambda}(1 - \alpha)\} = 1 - \alpha - \frac{\alpha a\lambda}{2\lambda + (2 + a)\mu_{k,K}},$$

which proves (11).

The population calculation shows how the relative magnitudes of the eigenvalue $\mu_{k,K}$ and regularization level λ determine the coverage correction.

F.3 General empirical resolution conditions

Proposition 3. *Consider the circular model of Appendix F.1 with $a = 0.8$. Let K_n be any deterministic sequence of positive integers, and let $n \to \infty$ source calibration pairs and $m = m_n \to \infty$ target covariates be mutually independent and independent of the training information supplying $\widehat{c}_n$.*

Fix a continuous positive-semidefinite radial kernel k with $k(x, x) = 1$ on the circle, and write $\mu_n := \mu_{k,K_n}$. For deterministic $\lambda_n > 0$ and $t_n \geq 0$, suppose that the response-free weights $\widehat{\pi}_k \in \Delta_n(5)$ satisfy

$$\mathcal{J}_{k,\lambda_n}(\widehat{\pi}_k) \leq \inf_{\pi \in \Delta_n(5)} \mathcal{J}_{k,\lambda_n}(\pi) + t_n, \qquad \mathcal{J}_{k,\lambda_n}(\pi) := D_k^2(\widehat{P}_\pi, \widehat{Q}_m) + \lambda_n n \sum_{i=1}^n \pi_i^2.$$

Set $a_n := n^{-1} + m^{-1} + t_n$. Let $F_{Q,\widehat{c}_n}$ be the target CDF of $|Y - \widehat{c}_n|$, conditional on training, and let $\widehat{q}_{k,.90}$ be the weighted empirical 0.90 quantile.

(i) If $\mu_n > 0$, then

$$\left\| \widehat{F}_{\widehat{\pi}_k} - F_{Q,\widehat{c}_n} \right\|_\infty = O_p\left\{ \sqrt{\frac{a_n + \lambda_n}{\mu_n}} + n^{-1/2} \right\}. \tag{23}$$

Consequently, if $(a_n + \lambda_n)/\mu_n \longrightarrow 0$, then $\quad F_{Q,\widehat{c}_n}(\widehat{q}_{k,.90}) \xrightarrow{p} 0.90$.

(ii) If$(\mu_n + a_n)/\lambda_n \longrightarrow 0$, then

$$\left\| \widehat{\pi}_k - \frac{1}{n}\mathbf{1} \right\|_1 = O_p\left\{ \sqrt{\frac{\mu_n + a_n}{\lambda_n}} \right\} \xrightarrow{p} 0.$$

If additionally $\widehat{c}_n \xrightarrow{p} 0$, then $F_{Q,\widehat{c}_n}(\widehat{q}_{k,.90}) \xrightarrow{p} 0.86$.

(iii) If $K_n \equiv K$ for a fixed positive integer K, then both the Gaussian and radial Matérn-$3/2$ kernels, at any fixed positive length scale, satisfy the conclusion of part (i) whenever $\lambda_n \to 0, t_n \to 0$.

Proof. Write $r := r_{K_n}, \bar{r}_n := \frac{1}{n}\sum_{i=1}^n r(\theta_i)$. Since $0.2 \leq r \leq 1.8, Pr = 1$, we have

$$\bar{r}_n = 1 + O_p(n^{-1/2}).$$

Define the feasible oracle comparator by

$$\pi_i^r := \begin{cases} \dfrac{r(\theta_i)}{n\bar{r}_n}, & \bar{r}_n \geq 1/2, \\ \dfrac{1}{n}, & \bar{r}_n < 1/2. \end{cases}$$

On $\{\bar{r}_n \geq 1/2\}$,

$$0 \leq \pi_i^r \leq \frac{1.8}{n/2} = \frac{3.6}{n} < \frac{5}{n},$$

and

$$n\sum_{i=1}^n (\pi_i^r)^2 = \frac{n^{-1}\sum_{i=1}^n r^2(\theta_i)}{\bar{r}_n^2} = O_p(1).$$

The same bound is immediate on the complementary event.

Let ϕ_k be a feature map for k. Since $k(x, x) = 1$,

$$\|\phi_k(x)\|_{\mathcal{H}_k} = 1.$$

Moreover,

$$E_P\{r(\theta)\phi_k(X(\theta))\} = E_{Q_{K_n}}\phi_k(X).$$

Hilbert-space second-moment bounds therefore give

$$\left\| \frac{1}{n} \sum_{i=1}^{n} r(\theta_i)\phi_k(X_i) - E_{Q_{K_n}} \phi_k(X) \right\|_{\mathcal{H}_k} = O_p(n^{-1/2}),$$

and hence

$$D_k(\widehat{P}_{\pi^r}, Q_{K_n}) = O_p(n^{-1/2}).$$

Similarly,

$$D_k(\widehat{Q}_m, Q_{K_n}) = O_p(m^{-1/2}).$$

It follows that

$$D_k(\widehat{P}_{\pi^r}, \widehat{Q}_m) \le D_k(\widehat{P}_{\pi^r}, Q_{K_n}) + D_k(\widehat{Q}_m, Q_{K_n}) = O_p(n^{-1/2} + m^{-1/2}).$$

By approximate optimality,

$$\begin{aligned} \mathcal{J}_{k,\lambda_n}(\widehat{\pi}_k) &\le \mathcal{J}_{k,\lambda_n}(\pi^r) + t_n \\ &= D_k^2(\widehat{P}_{\pi^r}, \widehat{Q}_m) + \lambda_n n \sum_i (\pi_i^r)^2 + t_n \\ &= O_p(n^{-1} + m^{-1} + \lambda_n + t_n) \\ &= O_p(a_n + \lambda_n). \end{aligned}$$

Since the regularization term is nonnegative,

$$D_k^2(\widehat{P}_{\widehat{\pi}_k}, \widehat{Q}_m) = O_p(a_n + \lambda_n).$$

Together with

$$D_k(\widehat{Q}_m, Q_{K_n}) = O_p(m^{-1/2}),$$

the triangle inequality gives

$$D_k^2(\widehat{P}_{\widehat{\pi}_k}, Q_{K_n}) = O_p(a_n + \lambda_n). \tag{24}$$

We now convert this discrepancy bound into a score-CDF bound. Set

$$\widehat{z}_n := \sum_{i=1}^{n} \widehat{\pi}_{k,i} \cos(K_n \theta_i).$$

Writing

$$\beta_{s,c} := \frac{A_1(s;c) - A_0(s;c)}{2}, \qquad |\beta_{s,c}| \le \frac{1}{2},$$

the affine score identity (20) gives

$$\sum_i \widehat{\pi}_{k,i} h_{s,c}(\theta_i) - F_{Q,c}(s) = \beta_{s,c} \left(\widehat{z}_n - \frac{a}{2} \right).$$

By the mode inequality (22),

$$2\mu_n \left(\widehat{z}_n - \frac{a}{2} \right)^2 \le D_k^2(\widehat{P}_{\widehat{\pi}_k}, Q_{K_n}).$$

Consequently,

$$\begin{aligned} \sup_s \left| \sum_i \widehat{\pi}_{k,i} h_{s,\widehat{c}_n}(\theta_i) - F_{Q,\widehat{c}_n}(s) \right| &\le \frac{1}{2} \left| \widehat{z}_n - \frac{a}{2} \right| \\ &\le \frac{D_k(\widehat{P}_{\widehat{\pi}_k}, Q_{K_n})}{2\sqrt{2\mu_n}} \\ &= O_p \left\{ \sqrt{\frac{a_n + \lambda_n}{\mu_n}} \right\}. \end{aligned} \tag{25}$$

Conditional on the training information and all covariates, the weights are fixed and the calibration responses remain independent. Since

$$\sum_i \widehat{\pi}_{k,i}^2 \le \left(\max_i \widehat{\pi}_{k,i}\right) \sum_i \widehat{\pi}_{k,i} \le \frac{5}{n},$$

Lemma 1 gives

$$\left\| \widehat{F}_{\widehat{\pi}_k} - \sum_i \widehat{\pi}_{k,i} h_{\cdot,\widehat{c}_n}(\theta_i) \right\|_\infty = O_p(n^{-1/2}).$$

Combining this with (25) proves (23).

For completeness, let

$$\Delta_n := \left\| \widehat{F}_{\widehat{\pi}_k} - F_{Q,\widehat{c}_n} \right\|_\infty .$$

Since $F_{Q,\widehat{c}_n}$ is continuous and

$$\widehat{F}_{\widehat{\pi}_k}(\widehat{q}_{k,\tau}^{-}) \le \tau \le \widehat{F}_{\widehat{\pi}_k}(\widehat{q}_{k,\tau}),$$

we have

$$F_{Q,\widehat{c}_n}(\widehat{q}_{k,\tau}) \ge \widehat{F}_{\widehat{\pi}_k}(\widehat{q}_{k,\tau}) - \Delta_n \ge \tau - \Delta_n$$

and

$$F_{Q,\widehat{c}_n}(\widehat{q}_{k,\tau}) = F_{Q,\widehat{c}_n}(\widehat{q}_{k,\tau}^{-}) \le \widehat{F}_{\widehat{\pi}_k}(\widehat{q}_{k,\tau}^{-}) + \Delta_n \le \tau + \Delta_n.$$

Thus

$$|F_{Q,\widehat{c}_n}(\widehat{q}_{k,\tau}) - \tau| \le \Delta_n,$$

which proves the coverage conclusion in part (i).

For part (ii), let $u_i = 1/n$. Since $u \in \Delta_n(5)$, approximate optimality gives

$$\begin{aligned} D_k^2(\widehat{P}_{\widehat{\pi}_k}, \widehat{Q}_m) + \lambda_n n \sum_i \widehat{\pi}_{k,i}^2 \\ \le D_k^2(\widehat{P}_u, \widehat{Q}_m) + \lambda_n + t_n. \end{aligned} \tag{26}$$

Dropping the first term on the left yields

$$\lambda_n \left(n \sum_i \widehat{\pi}_{k,i}^2 - 1 \right) \le D_k^2(\widehat{P}_u, \widehat{Q}_m) + t_n. \tag{27}$$

The triangle inequality and $D_k^2(P, Q_{K_n}) = a^2 \mu_n / 2$ give

$$\begin{aligned} D_k(\widehat{P}_u, \widehat{Q}_m) &\le D_k(\widehat{P}_u, P) + D_k(P, Q_{K_n}) + D_k(Q_{K_n}, \widehat{Q}_m) \\ &= O_p\left(n^{-1/2} + \sqrt{\mu_n} + m^{-1/2} \right). \end{aligned}$$

Therefore,

$$D_k^2(\widehat{P}_u, \widehat{Q}_m) + t_n = O_p(\mu_n + a_n).$$

By normalization and Cauchy–Schwarz,

$$\begin{aligned} \|\widehat{\pi}_k - u\|_1^2 &\le n \sum_i \left(\widehat{\pi}_{k,i} - \frac{1}{n} \right)^2 \\ &= n \sum_i \widehat{\pi}_{k,i}^2 - 1. \end{aligned}$$

Combining this identity with (27) gives

$$\left\| \widehat{\pi}_k - \frac{1}{n}\mathbf{1} \right\|_1 = O_p\left\{ \sqrt{\frac{\mu_n + a_n}{\lambda_n}} \right\}.$$

Let $\widehat{F}_u$ be the unweighted empirical score CDF. Then

$$\left\|\widehat{F}_{\widehat{\pi}_k} - \widehat{F}_u\right\|_\infty \le \|\widehat{\pi}_k - u\|_1 = o_p(1),$$

while conditional empirical-process convergence gives

$$\left\|\widehat{F}_u - F_{P,\widehat{c}_n}\right\|_\infty = O_p(n^{-1/2}).$$

If $\widehat{c}_n \xrightarrow{p} 0$, then

$$\big||Y - \widehat{c}_n| - |Y|\big| \le |\widehat{c}_n|,$$

and uniform continuity of the zero-centered source and target CDFs implies

$$\|F_{P,\widehat{c}_n} - F_{P,0}\|_\infty + \|F_{Q,\widehat{c}_n} - F_{Q,0}\|_\infty = o_p(1).$$

Hence

$$\left\|\widehat{F}_{\widehat{\pi}_k} - F_{P,0}\right\|_\infty = o_p(1).$$

Since $F_{P,0}$ has a unique 0.90 quantile at 2.8 and is strictly increasing in a neighborhood of 2.8,

$$\widehat{q}_{k,.90} \xrightarrow{p} 2.8.$$

Continuity of $F_{Q,0}$ then gives

$$F_{Q,\widehat{c}_n}(\widehat{q}_{k,.90}) \xrightarrow{p} F_{Q,0}(2.8) = 0.86.$$

For the independent source-training mean based on $2n$ observations,

$$E(\widehat{c}_n) = 0, \qquad \mathrm{Var}(\widehat{c}_n) = \frac{\mathrm{Var}(Y)}{2n} \le \frac{9}{2n},$$

so $\widehat{c}_n = O_p(n^{-1/2})$. The fixed zero center satisfies the same condition trivially.

Finally, if $K_n \equiv K$, Lemma 3 gives $\mu_{G,K} > 0, \mu_{M,K} > 0$. Thus, when $n, m \to \infty$, $t_n \to 0$, and $\lambda_n \to 0$,

$$\frac{a_n + \lambda_n}{\mu_{k,K}} \longrightarrow 0,$$

and part (iii) follows from part (i). □

F.4 Proof of Theorem 2

Under the regime of Theorem 2,

$$m = n, \qquad K_n = \lceil \log n \rceil, \qquad \lambda_n = c_\lambda \frac{\log n}{n}, \qquad t_n = O(n^{-1}),$$

and hence $a_n = 2n^{-1} + t_n = O(n^{-1})$.

For the Matérn-$3/2$ kernel, Lemma 3 gives $\mu_{M,K_n} \asymp K_n^{-4} \asymp (\log n)^{-4}$. Therefore,

$$\frac{a_n + \lambda_n}{\mu_{M,K_n}} = O\left\{\frac{\log(n)/n}{(\log n)^{-4}}\right\} = O\left\{\frac{(\log n)^5}{n}\right\} \longrightarrow 0.$$

Proposition 3(i) now yields

$$\begin{aligned}\left\|\widehat{F}_{\widehat{\pi}_M} - F_{Q,\widehat{c}_n}\right\|_\infty &= O_p\left\{\sqrt{\frac{a_n + \lambda_n}{\mu_{M,K_n}}} + n^{-1/2}\right\} \\ &= O_p\left\{\frac{(\log n)^{5/2}}{\sqrt{n}}\right\},\end{aligned}$$

and consequently

$$F_{Q,\widehat{c}_n}(\widehat{q}_{M,.90}) \xrightarrow{p} 0.90.$$

For the Gaussian kernel, the Bessel bound in Lemma 3 and $j! \geq (j/e)^j$ imply

$$\mu_{G,j} \leq C\frac{(1/2)^j}{j!} \leq C\left(\frac{e}{2j}\right)^j.$$

Thus, since $K_n \asymp \log n$,

$$\log\{n\mu_{G,K_n}\} \leq \log n - K_n \log K_n + O(K_n) \longrightarrow -\infty.$$

In particular, $n\mu_{G,K_n} \to 0$, and hence

$$\frac{\mu_{G,K_n} + a_n}{\lambda_n} = O\left\{\frac{n\mu_{G,K_n}}{\log n} + \frac{1}{\log n}\right\} \longrightarrow 0.$$

Proposition 3(ii) therefore gives

$$\left\|\widehat{\pi}_G - \frac{1}{n}\mathbf{1}\right\|_1 \xrightarrow{p} 0.$$

The independent source-training mean satisfies $\widehat{c}_n = O_p(n^{-1/2})$, so

$$F_{Q,\widehat{c}_n}(\widehat{q}_{G,.90}) \xrightarrow{p} 0.86.$$

The same conclusion holds when the score center is fixed at zero.

ESS screening. For Gaussian balancing, the preceding argument and $\widehat{\pi}_G \in \Delta_n(5)$ give

$$0 \leq n\sum_i \widehat{\pi}_{G,i}^2 - 1 = n\sum_i \left(\widehat{\pi}_{G,i} - \frac{1}{n}\right)^2 \leq 4\left\|\widehat{\pi}_G - \frac{1}{n}\mathbf{1}\right\|_1 = o_p(1).$$

Since

$$\frac{\mathrm{ESS}(\pi)}{n} = \frac{1}{n\sum_i \pi_i^2},$$

it follows that $\mathrm{ESS}(\widehat{\pi}_G)/n \xrightarrow{p} 1$.

For the normalized oracle comparator π^r from Proposition 3,

$$n\sum_i (\pi_i^r)^2 = \frac{n^{-1}\sum_i r^2(\theta_i)}{\bar{r}_n^2} = Pr^2 + O_p(n^{-1/2}) = 1 + \frac{a^2}{2} + O_p(n^{-1/2}).$$

For Matérn balancing, comparison with π^r gives

$$\begin{aligned} n\sum_i \widehat{\pi}_{M,i}^2 &\leq n\sum_i (\pi_i^r)^2 + \frac{D_M^2(\widehat{P}_{\pi^r}, \widehat{Q}_n) + t_n}{\lambda_n} \\ &= 1 + \frac{a^2}{2} + O_p\left(\frac{n^{-1}}{\log(n)/n}\right) + O_p(n^{-1/2}) \\ &= 1 + \frac{a^2}{2} + o_p(1). \end{aligned}$$

Consequently,

$$\frac{\mathrm{ESS}(\widehat{\pi}_M)}{n} \geq \frac{1}{1 + a^2/2} + o_p(1).$$

For $a = 0.8$, the limiting lower bound is $1/1.32 \approx 0.758 > 0.25$. Hence the implementation's ESS floor of $0.25n$ is asymptotically inactive for both procedures.

Numerical kernel approximation. Let $\widetilde{k}$ be a computational approximation satisfying

$$\sup_{x,x'} \left|\widetilde{k}(x,x') - k(x,x')\right| \leq \varepsilon_k.$$

For any probability measures R and S, writing $\Delta k = \widetilde{k} - k$,

$$\left|\widetilde{D}_k^2(R,S) - D_k^2(R,S)\right| = \left|\iint \Delta k(x,x')\, d(R-S)(x)\, d(R-S)(x')\right| \\ \leq \varepsilon_k \|R-S\|_{\mathrm{TV}}^2 \leq 4\varepsilon_k.$$

The same bound therefore holds uniformly over normalized nonnegative weights.

Let $\widetilde{\mathcal{J}}$ and $\mathcal{J}$ denote the approximate- and original-kernel objectives, with the same regularization term. If a feasible $\widehat{\pi}$ satisfies

$$\widetilde{\mathcal{J}}(\widehat{\pi}) \leq \inf_{\pi \in \Delta_n(5)} \widetilde{\mathcal{J}}(\pi) + \widetilde{t}_n,$$

then

$$\begin{aligned} \mathcal{J}(\widehat{\pi}) &\leq \widetilde{\mathcal{J}}(\widehat{\pi}) + 4\varepsilon_k \\ &\leq \inf_{\pi \in \Delta_n(5)} \widetilde{\mathcal{J}}(\pi) + \widetilde{t}_n + 4\varepsilon_k \\ &\leq \inf_{\pi \in \Delta_n(5)} \mathcal{J}(\pi) + \widetilde{t}_n + 8\varepsilon_k. \end{aligned}$$

Thus $\widehat{\pi}$ is a $(\widetilde{t}_n + 8\varepsilon_k)$-near-minimizer of the original objective.

For a positive Fourier truncation, the absolute truncation error is bounded by the sum of the omitted nonnegative Fourier coefficients, which equals the omitted diagonal kernel mass. This quantity may therefore be used as ε_k. The approximation used at $n = 4{,}000$ is consequently covered by the theorem through a quantified original-objective gap.

F.5 General spectral-tail regularity

Constants in this subsection may depend on dimension, bandwidth, the spectral family, and the Fourier convention, but not on ε.

Theorem 5. *(i) Exponential spectral tails. Let $f_\epsilon \in L^2(\mathbb{R}^d)$, and suppose that for some $0 < a < b < \infty$, $c_f > 0$, and $r \geq 0$,*

$$\int_{\{a/\epsilon \leq \|t\|_2 \leq b/\epsilon\}} |\widehat{f}_\epsilon(t)|^2\, dt \geq c_f \epsilon^r$$

for all sufficiently small ϵ. If, for sufficiently large $\|t\|_2$, $\omega(t) \leq C_0 \exp\{-c_0 \|t\|_2^\nu\}$, $\nu > 0$, then

$$\|f_\epsilon\|_\omega^{\star 2} \geq C \epsilon^r \exp\{c_0 a^\nu \epsilon^{-\nu}\}.$$

(ii) Polynomial spectral tails. Suppose that $\omega(t)^{-1} \leq C_1(1 + \|t\|_2^2)^\beta$ for some $\beta > 0$. Let $f \in L^2(\mathbb{R}^d)$, $\eta \in \mathcal{S}(\mathbb{R}^d)$, and $f_\epsilon = \eta_\epsilon * f, \eta_\epsilon(x) = \epsilon^{-d}\eta(x/\epsilon)$. Then, for $0 < \epsilon \leq 1$,

$$\|f_\epsilon\|_\omega^{\star 2} \leq C' \epsilon^{-2\beta} \|f\|_2^2.$$

The costs in Theorem 5 enter the transfer bound through B_ω. The kernel-specific exponents and a smoothed-interface example are given below.

Proof. For part (i), the exponential-tail condition implies

$$\omega(t)^{-1} \geq C_0^{-1} \exp\{c_0 \|t\|_2^\nu\}$$

for sufficiently large $\|t\|_2$. Restrict the integral in (15) to the annulus $A_\varepsilon = \{a/\varepsilon \leq \|t\|_2 \leq b/\varepsilon\}$. For all sufficiently small ε,

$$\begin{aligned} \|f_\varepsilon\|_\omega^{\star\, 2} &\geq (2\pi)^{-d} C_0^{-1} \exp\{c_0 a^\nu \varepsilon^{-\nu}\} \int_{A_\varepsilon} |\widehat{f}_\varepsilon(t)|^2\, dt \\ &\geq C \varepsilon^r \exp\{c_0 a^\nu \varepsilon^{-\nu}\}. \end{aligned}$$

For part (ii), $\widehat{f}_\varepsilon(t) = (2\pi)^{d/2} \widehat{\eta}(\varepsilon t) \widehat{f}(t)$. For every $N > 0$, the Schwartz property gives $|\widehat{\eta}(u)| \leq C_N (1 + \|u\|_2)^{-N}$. Choose $N > \beta$. Then, for $0 < \varepsilon \leq 1$,

$$\sup_{t \in \mathbb{R}^d} (1 + \|t\|_2^2)^\beta |\widehat{\eta}(\varepsilon t)|^2 \leq C \varepsilon^{-2\beta}.$$

Consequently, using Plancherel's theorem,

$$\|f_\varepsilon\|_\omega^{\star\,2} \le C \int |\widehat{f}(t)|^2 |\widehat{\eta}(\varepsilon t)|^2 (1 + \|t\|_2^2)^\beta \, dt$$
$$\le C' \varepsilon^{-2\beta} \|f\|_2^2.$$

□

F.6 High-frequency mass of a smoothed interface

Lemma 4. *Let*

$$g(x) = \mathbf{1}_{[0,1]}(x_1)\psi(x'), \qquad x' = (x_2, \dots, x_d),$$

where $\psi \in C_c^\infty(\mathbb{R}^{d-1})$ is nonzero. Let $\eta \in \mathcal{S}(\mathbb{R}^d)$ satisfy $\int \eta(x)\,dx = 1$, and put $g_\varepsilon = \eta_\varepsilon * g$, $\eta_\varepsilon(x) = \varepsilon^{-d}\eta(x/\varepsilon)$. There exist $0 < a < b < \infty$, a point $t_0' \in \mathbb{R}^{d-1}$, $\delta > 0$, and $c > 0$ such that

$$E_\varepsilon = [a/\varepsilon, b/\varepsilon] \times B_{d-1}(t_0', \delta)$$

satisfies

$$\int_{E_\varepsilon} |\widehat{g}_\varepsilon(t)|^2 \, dt \ge c\varepsilon$$

for all sufficiently small ε.

Proof. Continuity of $\widehat{\eta}$ and $\widehat{\eta}(0) = (2\pi)^{-d/2}$ imply that there are $\rho, c_\eta > 0$ such that $|\widehat{\eta}(u)| \ge c_\eta$ whenever $\|u\|_2 \le \rho$. Since $\psi \ne 0$, its Fourier transform is not identically zero. Choose t_0' with $\widehat{\psi}(t_0') \ne 0$, and then choose $\delta, c_\psi > 0$ such that $|\widehat{\psi}(t')| \ge c_\psi$ on $B_{d-1}(t_0', \delta)$.

Choose $0 < a < b < \rho/2$. If $M = \sup_{t' \in B(t_0', \delta)} \|t'\|_2$, then for $t \in E_\varepsilon$,

$$\|\varepsilon t\|_2^2 \le b^2 + \varepsilon^2 M^2 < \rho^2$$

for all sufficiently small ε. Hence $|\widehat{\eta}(\varepsilon t)| \ge c_\eta$ on E_ε. Up to a positive normalization constant,

$$\widehat{g}(t_1, t') = \frac{1 - e^{-\mathrm{i}t_1}}{\mathrm{i}t_1} \widehat{\psi}(t').$$

Therefore, on E_ε,

$$|\widehat{g}_\varepsilon(t)|^2 \ge c_0 \frac{\sin^2(t_1/2)}{t_1^2}.$$

It remains to lower-bound

$$I_\varepsilon = \int_{a/\varepsilon}^{b/\varepsilon} \frac{\sin^2(t/2)}{t^2} \, dt.$$

Since $\sin^2(t/2) = (1 - \cos t)/2$,

$$I_\varepsilon = \frac{\varepsilon}{2}\left(\frac{1}{a} - \frac{1}{b}\right) - \frac{1}{2}\int_{a/\varepsilon}^{b/\varepsilon} \frac{\cos t}{t^2} \, dt.$$

For $A = a/\varepsilon$ and $B = b/\varepsilon$, integration by parts gives

$$\int_A^B \frac{\cos t}{t^2} \, dt = \left[\frac{\sin t}{t^2}\right]_A^B + 2\int_A^B \frac{\sin t}{t^3} \, dt,$$

whose absolute value is at most $3/A^2 = 3\varepsilon^2/a^2$. Thus

$$I_\varepsilon \ge \frac{\varepsilon}{4}\left(\frac{1}{a} - \frac{1}{b}\right)$$

for all sufficiently small ε, proving the lemma. □

For small ε, E_ε is contained in a fixed-multiple annulus of radius $1/\varepsilon$. Lemma 4 therefore verifies the high-frequency-mass condition in Theorem 5 with $r = 1$. The smoothed interface is a concrete family of sharp score approximants. A Gaussian spectrum yields

$$\|g_\varepsilon\|_{\omega_G}^{\star\,2} \gtrsim \varepsilon \exp(c/\varepsilon^2),$$

whereas the polynomial reciprocal spectra covered by part (ii) yield only polynomial-order upper bounds for the same mollified interface.

F.7 Gaussian, product-Laplace, Matérn, and energy spectra

Up to constants and bandwidth rescaling, the spectra used in the paper have the following forms.

$$\begin{aligned}
\text{Gaussian:} \quad & \omega_G(t) \asymp \exp(-c\|t\|_2^2), \\
\text{product-Laplace:} \quad & \omega_L(t) \asymp \prod_{j=1}^{d} (1 + c_j t_j^2)^{-1}, \\
\text{Matérn:} \quad & \omega_M(t) \asymp (a^2 + \|t\|_2^2)^{-(\nu_M + d/2)}, \\
\text{energy:} \quad & \omega_E(t) \asymp \|t\|_2^{-(d+1)}.
\end{aligned}$$

The Gaussian case satisfies the exponential-tail condition with exponent two. For product-Laplace,

$$\omega_L(t)^{-1} \lesssim (1 + \|t\|_2^2)^d,$$

so $\beta = d$. For Matérn, $\beta = \nu_M + d/2$. For energy,

$$\omega_E(t)^{-1} = \|t\|_2^{d+1} \lesssim (1 + \|t\|_2^2)^{(d+1)/2},$$

so the generalized dual seminorm has $\beta = (d+1)/2$.

These calculations determine the spectral contribution to the dual envelope B_ω. The complete error bound combines that envelope with achieved imbalance, target sampling, and ESS.

Native-space interpretation. For integrable spectra, CFD is the MMD associated with the translation-invariant kernel having spectral density ω, and the corresponding Fourier dual norm agrees with the native RKHS norm up to the adopted Fourier normalization. After restriction to the compact covariate domain $\mathcal{X}$, the representative spectra admit the regularity interpretations summarized in Table 4. These interpretations are standard for translation-invariant kernels and Sobolev/native spaces (Santra et al., 2026; Steinwart & Christmann, 2008; Rasmussen & Williams, 2006).

Table 4: Native-space interpretation of representative CFD geometries. Descriptions are up to norm equivalence, bandwidth rescaling, and restriction to the compact covariate domain.

Geometry	Native-space interpretation
Gaussian	Gaussian RKHS with very strong smoothness; on standard compact domains it is embedded in $W^{s,2}(\mathcal{X})$ for every finite $s > 0$.
Product-Laplace	Mixed tensor-product Sobolev regularity, $\bigotimes_{j=1}^{d} W^{1,2}(\mathcal{X}_j)$ on product domains, controlling mixed derivatives across coordinates.
Matérn	Isotropic Sobolev-type regularity; for the spectrum above, the native norm is equivalent on standard compact domains to a $W^{\nu_M + d/2, 2}(\mathcal{X})$ norm.
Energy	Generalized homogeneous Sobolev-type regularity of order $(d+1)/2$, with the seminorm induced by the singular energy spectrum.

G Additional experimental details and results

We examine bounded nonlinear shifts, increasing-frequency shifts, and the $m^{-1/2}$ scaling of target-sampling error. The Parkinson's recording-level study also evaluates alternative conformity scores. Table 5 summarizes the sampling and selection protocols; the study-specific details follow.

Table 5: Experimental protocols. Balancing and validation samples contain covariates only. "No refit" means that selection retains the fitted candidate weights. Tuning grids and exact sample sizes are specified in the corresponding subsections.

Study	Sampling unit and repetitions	Preprocessing and selection	Calibration/evaluation
Bounded nonlinear	Independent synthetic observations; 200 per cell	Training-fixed; 15-kernel common validation; no refit	Residual scores; independent test sample
Spectral resolution	Independent points on the circle; 100 per cell	Raw Cartesian coordinates; fixed kernels or 8-kernel common validation; no refit	Learned center; analytic target coverage
Target sampling	Independent Gaussian observations; 100 or 500 per cell	Fixed bandwidths and oracle predictor	CFD or score-CDF discrepancy
Diabetes	442 fixed patients; 300 assignments/partitions per setting	Training-fixed; 15-kernel common validation; no refit	Residual scores; disjoint patient roles
Parkinson's recordings	100 split-CP and CQR; 50 UACQR-P per shift	Pooled preprocessing; within-family validation; refit	Recording-level evaluation; 50-partition subject-disjoint sensitivity

G.1 Bounded nonlinear-shift experiments

G.1.1 Design and estimands

The source law is uniform on $[-1,1]^5$. Let $g_A(x) = \tanh(6x_1x_2)$ for the interaction case and $g_k(x) = \cos(\pi k x_1)\cos(\pi k x_2)$ for checkerboards, $k \in \{1,2\}$. The target density is $p_{\mathrm{tar}} = (1+0.8g)p_{\mathrm{src}}$. Symmetry gives $E_{\mathrm{src}}g = 0$ and unchanged univariate marginals. For integer checkerboard frequency, the distribution of each cosine under a uniform covariate is unchanged, so the joint distribution of the oracle ratio and conditional scale is unchanged across k. In particular,

$$E_{\mathrm{src}}r^2 = 1 + 0.8^2/4 = 1.16, \qquad E_{\mathrm{tar}}g_k = 0.8/4 = 0.2, \qquad \mathrm{TV}(P_{\mathrm{src}}, P_{\mathrm{tar}}) = \frac{1.6}{\pi^2}.$$

First and second covariate moments also agree across domains for the checkerboards. The shared response law is

$$Y = f(X) + \{1.25 + 0.75g(X)\}E, \qquad E \sim N(0,1), \quad E \perp X,$$
$$f(x) = 2\sin(x_1) + \tfrac{1}{2}x_2^2 - \tfrac{1}{2}x_3.$$

The predictor is fitted using independent source-training data.

For each shift and $n \in \{300, 1{,}000\}$, we generate $2n$ source training pairs, n source calibration pairs, n target balancing covariates, n target validation covariates, and 20,000 independent target test pairs. There are 200 independent replications per cell and all methods share the data and predictor within a replication.

Let $\widehat{C}_r$ denote empirical target coverage in replication r. We report $R^{-1}\sum_r \widehat{C}_r$ and $R^{-1}\sum_r |\widehat{C}_r - \tau|$ separately. Paired differences use methods evaluated on the same data. Interval lengths and maximal-set rates accompany coverage. Profile imbalance is

$$\left| \sum_i \pi_i g(X_i) - E_{\mathrm{tar}} g(X) \right|,$$

where g is available only to the evaluator. Score-CDF discrepancy compares the normalized weighted calibration-score CDF with the independent test-score CDF.

G.1.2 Shared nonlinear and diabetes selection protocol

The shared fitting and selection settings for the nonlinear and diabetes studies are summarized below. For DB-CC, after standardization, the kernel families are

$$k_{\mathrm{G}}(x,z) = \exp\left\{-\frac{\|x-z\|_2^2}{2h^2}\right\}, k_{\mathrm{PL}}(x,z) = \exp\left\{-\frac{\|x-z\|_1}{h}\right\},$$
$$k_{\mathrm{M}}(x,z) = \left(1 + \frac{\sqrt{3}\|x-z\|_2}{h}\right)\exp\left\{-\frac{\sqrt{3}\|x-z\|_2}{h}\right\}.$$

Table 6: Candidate libraries and tuning grids shared by the nonlinear and diabetes experiments. All candidate selection is response-free.

Component	Candidates and settings
Prediction model	Random forest with 100 trees, minimum leaf size 5, and all covariates available at each split.
DB-CC kernel families	Gaussian, product-Laplace, and Matérn-$3/2$.
Bandwidth construction	Median Euclidean distance for the Gaussian and Matérn families and median ℓ_1 distance for the product-Laplace family, computed among the first at most 300 source-training observations, with multipliers $\{0.125, 0.25, 0.5, 1, 2\}$.
DB-CC regularization	$\lambda_0 \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$.
DB-CC admissibility	Relative weight cap $B_w = 5$ and $\mathrm{ESS}/n \geq 0.25$; uniform weights are included as a candidate.
Validation geometry	Equal mixture of the 15 training-fixed kernel geometries.
Selected multiscale DB-CC	Minimum common-validation MMD over the 60 kernel–ridge candidates and uniform weights.
Matched Gaussian-only DB-CC	Same bandwidth and ridge grids, weight cap, ESS rule, uniform candidate, and validation criterion, restricted to Gaussian kernels.
Additional balancing comparators	Energy balancing and marginal and full degree-two moment balancing in the synthetic experiments.
Logistic ratio models	Linear and full degree-two features, with inverse penalty $C \in \{0.01, 0.1, 1, 10, 100\}$.
RBF ratio model	Nyström logistic regression with at most 256 training-derived landmarks and bandwidth multipliers $\{0.125, 0.25, 0.5, 1, 2\}$.
Boosted-tree ratio model	Histogram gradient boosting with domain-only sigmoid probability calibration; maximum leaves $\{7, 15, 31\}$, minimum leaf sizes $\{10, 20\}$, and ℓ_2 regularization $\{0, 1\}$; learning rate 0.05 and 300 iterations without early stopping.
Ratio clipping	Caps $\{2, 5, 10, \infty\}$, selected by common-validation MMD unless otherwise specified.
Selected flexible WCP	Eight candidates formed by crossing the RBF and boosted-tree ratio estimators with the four clipping levels, selected by common-validation MMD.
Optimization certificate	Balancing QPs are solved by ADMM and accepted only when the Frank–Wolfe optimality gap is at most 10^{-7}.

Ratio models are fitted using source-training and target-balancing covariates. Their internal hyperparameters are selected by three-fold stratified domain-label log loss. For estimated target-domain probability $\widehat{e}(x)$, with N_s source and N_t target classifier-training observations, the density-ratio estimate is

$$\widehat{r}(x) = \frac{N_s}{N_t} \frac{\widehat{e}(x)}{1 - \widehat{e}(x)}.$$

The synthetic linear cap-10 baseline remains fixed at cap 10, whereas the diabetes linear and quadratic WCP comparators select their caps.

WCP evaluates calibration and test ratios on the same numerical scale and retains its candidate-specific test-point mass. These clipped WCP comparators are distinct from the bias-corrected CWCP method of Wang & Goel (2026).

G.1.3 Coverage, width, and paired comparisons

Tables 7 and 8 summarize 200 replications; parentheses give Monte Carlo SEs, and all prediction sets are finite. Selected DB-CC improves profile balance most under the interaction shift and less under the low-frequency checkerboard. At $k = 2, n = 1{,}000$, its profile imbalance is 0.192, versus 0.200 unweighted and 0.013 with oracle weights, and none of the learned procedures resolves the undercoverage. This pattern is consistent with insufficient effective resolution of the tested kernel and ratio-model libraries.

The paired absolute-gap contrasts vary across settings (Tables 9 and 10). Relative to flexible WCP, DB-CC has smaller mean gaps in the interaction and $k = 1$ cells at $n = 1{,}000$, but larger gaps

for both checkerboards at $n = 300$. Differences from matched Gaussian-only selection are small; pointwise intervals exclude zero only for the two $n = 1{,}000$ checkerboard cells. These comparisons do not establish a consistent cross-family advantage.

A same-ratio diagnostic further shows that WCP's candidate-specific test-point mass increases coverage by 0.0010–0.0046, thereby reducing the coverage gap in these undercovered settings. Thus, coverage differences between WCP and global weighted-score calibration need not arise solely from ratio estimation.

Table 7: Bounded nonlinear shifts at $n = 300$ and nominal coverage 0.90.

Method	Coverage (SE)	Length	Gap (SE)	ESS$/n$	Profile
Interaction					
Unweighted split CP	0.8441 (0.0020)	4.85	0.0562 (0.0019)	1.000	0.450
Linear WCP, cap 10	0.8463 (0.0020)	4.88	0.0541 (0.0019)	0.989	0.450
Quadratic WCP	0.8882 (0.0019)	5.54	0.0225 (0.0013)	0.797	0.142
Selected flexible WCP	0.8775 (0.0020)	5.36	0.0280 (0.0016)	0.859	0.238
Gaussian-only DB-CC	0.8773 (0.0020)	5.37	0.0285 (0.0016)	0.815	0.203
Selected DB-CC	0.8787 (0.0020)	5.39	0.0276 (0.0016)	0.786	0.192
Oracle WCP	0.9025 (0.0017)	5.81	0.0195 (0.0009)	0.735	0.030
Checkerboard $k = 1$					
Unweighted split CP	0.8696 (0.0015)	4.50	0.0310 (0.0014)	1.000	0.198
Linear WCP, cap 10	0.8709 (0.0015)	4.52	0.0300 (0.0014)	0.988	0.199
Quadratic WCP	0.8708 (0.0016)	4.52	0.0300 (0.0015)	0.955	0.198
Selected flexible WCP	0.8717 (0.0015)	4.53	0.0291 (0.0015)	0.987	0.192
Gaussian-only DB-CC	0.8706 (0.0016)	4.51	0.0304 (0.0015)	0.943	0.179
Selected DB-CC	0.8707 (0.0016)	4.52	0.0304 (0.0014)	0.944	0.178
Oracle WCP	0.9020 (0.0014)	4.99	0.0163 (0.0008)	0.862	0.021
Checkerboard $k = 2$					
Unweighted split CP	0.8713 (0.0014)	4.52	0.0295 (0.0013)	1.000	0.197
Linear WCP, cap 10	0.8727 (0.0014)	4.53	0.0283 (0.0013)	0.989	0.197
Quadratic WCP	0.8726 (0.0014)	4.54	0.0283 (0.0013)	0.957	0.197
Selected flexible WCP	0.8724 (0.0014)	4.53	0.0286 (0.0013)	0.990	0.196
Gaussian-only DB-CC	0.8696 (0.0015)	4.50	0.0315 (0.0014)	0.945	0.195
Selected DB-CC	0.8697 (0.0015)	4.50	0.0314 (0.0014)	0.946	0.195
Oracle WCP	0.9025 (0.0014)	4.99	0.0155 (0.0009)	0.863	0.022

Table 8: Bounded nonlinear shifts at $n = 1{,}000$ and nominal coverage 0.90.

Method	Coverage (SE)	Length	Gap (SE)	ESS$/n$	Profile
Interaction					
Unweighted split CP	0.8409 (0.0010)	4.76	0.0591 (0.0010)	1.000	0.452
Linear WCP, cap 10	0.8415 (0.0010)	4.77	0.0585 (0.0010)	0.994	0.452
Quadratic WCP	0.8926 (0.0009)	5.56	0.0123 (0.0006)	0.772	0.087
Selected flexible WCP	0.8887 (0.0010)	5.49	0.0149 (0.0007)	0.793	0.125
Gaussian-only DB-CC	0.8917 (0.0009)	5.54	0.0130 (0.0006)	0.683	0.080
Selected DB-CC	0.8909 (0.0009)	5.53	0.0130 (0.0007)	0.715	0.089
Oracle WCP	0.9007 (0.0008)	5.71	0.0093 (0.0005)	0.734	0.016
Checkerboard $k = 1$					
Unweighted split CP	0.8686 (0.0008)	4.45	0.0314 (0.0008)	1.000	0.201
Linear WCP, cap 10	0.8690 (0.0008)	4.45	0.0310 (0.0008)	0.994	0.201
Quadratic WCP	0.8692 (0.0008)	4.45	0.0308 (0.0008)	0.977	0.201
Selected flexible WCP	0.8742 (0.0009)	4.52	0.0258 (0.0008)	0.979	0.170
Gaussian-only DB-CC	0.8770 (0.0009)	4.56	0.0232 (0.0009)	0.909	0.149
Selected DB-CC	0.8777 (0.0009)	4.57	0.0224 (0.0009)	0.902	0.143
Oracle WCP	0.9004 (0.0008)	4.91	0.0093 (0.0005)	0.862	0.012
Checkerboard $k = 2$					
Unweighted split CP	0.8700 (0.0008)	4.45	0.0300 (0.0008)	1.000	0.200
Linear WCP, cap 10	0.8704 (0.0008)	4.46	0.0296 (0.0008)	0.994	0.200
Quadratic WCP	0.8704 (0.0008)	4.46	0.0296 (0.0008)	0.977	0.200
Selected flexible WCP	0.8708 (0.0008)	4.46	0.0292 (0.0008)	0.995	0.198
Gaussian-only DB-CC	0.8702 (0.0008)	4.45	0.0298 (0.0008)	0.952	0.195
Selected DB-CC	0.8710 (0.0008)	4.46	0.0290 (0.0008)	0.941	0.192
Oracle WCP	0.9008 (0.0007)	4.90	0.0080 (0.0004)	0.862	0.013

Table 9: Selected DB-CC minus selected flexible WCP at nominal 0.90. Coverage-gap contrasts use within-replication pairing; intervals are pointwise 95% Monte Carlo intervals, not multiplicity-adjusted. CDF and profile columns are differences in the mean diagnostics. All differences are multiplied by 100; negative values favor DB-CC for the error metrics, while a negative ESS difference means more concentrated weights.

Shift	n	Gap difference [95% interval]	CDF	Profile	ESS$/n$
Interaction	300	-0.035 [-0.221, 0.150]	-0.956	-4.615	-7.257
Interaction	1000	-0.192 [-0.295, -0.088]	-0.723	-3.571	-7.768
Checkerboard $k = 1$	300	0.135 [0.035, 0.235]	-0.243	-1.416	-4.385
Checkerboard $k = 1$	1000	-0.343 [-0.428, -0.257]	-0.569	-2.781	-7.706
Checkerboard $k = 2$	300	0.282 [0.179, 0.384]	-0.035	-0.178	-4.381
Checkerboard $k = 2$	1000	-0.018 [-0.067, 0.032]	-0.136	-0.686	-5.450

Table 10: Selected multiscale DB-CC minus matched Gaussian-only DB-CC: within-replication absolute coverage-gap differences at nominal 0.90, over 200 paired replications. Both selectors share bandwidths, ridge values, cap, ESS rule, uniform candidate, and common validation criterion. Differences and pointwise 95% Monte Carlo intervals are multiplied by 100; negative values favor multiscale selection. Intervals are not adjusted for multiple comparisons.

Shift	n	Gap difference	95% MC interval
Interaction	300	-0.08700	[-0.17425, 0.00025]
Interaction	1,000	-0.00195	[-0.04824, 0.04434]
Checkerboard $k = 1$	300	-0.00383	[-0.06272, 0.05507]
Checkerboard $k = 1$	1,000	-0.08217	[-0.14478, -0.01957]
Checkerboard $k = 2$	300	-0.01017	[-0.06002, 0.03967]
Checkerboard $k = 2$	1,000	-0.07730	[-0.11119, -0.04341]

G.1.4 NO-SHIFT AND LINEAR-SHIFT CONTROLS

The no-shift control retains the interaction response law. The linear-shift control uses $r(x) = e^{x_1}/\sinh(1)$ on $[-1,1]^5$ and

$$Y = 2\sin(X_1) + \tfrac{1}{2}X_2^2 - \tfrac{1}{2}X_3 + \{0.5 + 1.5\,\mathrm{logit}^{-1}(X_1/0.1)\}E, \quad E \sim N(0,1), \quad E \perp X.$$

Both controls follow the $n = 300$ protocol described above. As shown in Table 11, selected DB-CC maintains near-nominal coverage under no shift and substantially reduces undercoverage under the linear shift.

Table 11: Controls using the same nonlinear-study protocol at $n = 300$ and nominal 0.90, each with 200 independently seeded replications. Entries are coverage (Monte Carlo SE).

Method	No shift	Linear log ratio, $\beta = 1$
Unweighted split CP	0.9017 (0.0012)	0.8501 (0.0018)
Linear WCP, cap 10	0.9027 (0.0012)	0.9005 (0.0015)
Quadratic WCP	0.9022 (0.0013)	0.8911 (0.0016)
Selected flexible WCP	0.9028 (0.0012)	0.8993 (0.0016)
Selected DB-CC	0.8997 (0.0012)	0.8938 (0.0015)
Oracle WCP	0.9017 (0.0012)	0.9024 (0.0015)

G.2 SPECTRAL-RESOLUTION EXPERIMENT

Design and evaluation. We use the circle and mixture-response model of Theorem 2, with $a = 0.8$, $\alpha = 0.1$, and $K_n = \lceil \log n \rceil$. The primary settings are $(n, K) = (250, 6), (1{,}000, 7), (4{,}000, 9)$, with $2n$ source-training observations and n target-balancing covariates, and a low-frequency control uses $(n, K) = (1{,}000, 2)$. The fitted center is the independent source-training mean, and the kernels use raw Cartesian coordinates.

Target coverage is evaluated analytically. For $s \geq 0$ and a fitted center c, let

$$A_j(s;c) = \tfrac{1}{2}\big|[2j, 2j+1] \cap [c-s, c+s]\big| + \tfrac{1}{2}\big|[-2j-1, -2j] \cap [c-s, c+s]\big|,$$

where $|\cdot|$ denotes interval length. A global threshold q then has exact target coverage

$$0.3A_0(q;c) + 0.7A_1(q;c).$$

Optimization and adaptive control. Primary kernels use unit length scale, a relative weight cap of five, and $\lambda_n = 0.05\log(n)/n$. The solver targets an additive optimality gap of $\min\{10^{-7}, 10^{-3}/n\}$ in the original objective, and all fits impose an ESS floor of $0.25n$. At $n = 4{,}000$, we use positive Fourier truncations and include the omitted-kernel-mass correction from Appendix F.4 in the optimality certificate.

The adaptive control reuses the $n = 1{,}000, K = 7$ datasets and adds $v = 1{,}000$ independent target-validation covariates. Each family jointly selects length scale and regularization over $h \in \{0.25, 0.5, 1, 2\}$ and $c_\lambda \in \{0.01, 0.05, 0.25\}$ in $\lambda_n = c_\lambda \log(n)/n$. The equal mixture of eight family–bandwidth kernels evaluates all 24 fitted candidates and the uniform candidate.

Table 12: Spectral-resolution experiment at nominal coverage 0.90. The final two rows use independent-validation selection. $\widehat{z}$ is the weighted cosine moment, whose target value is 0.4, and $C_{k,\lambda}$ denotes the analytic zero-center population coverage.

n, K	Kernel	Coverage (MCSE)	Gap	ESS$/n$	$\widehat{z}$	$C_{k,\lambda}$
250, 6	Gaussian	0.8581 (0.0028)	0.0426	0.952	-0.0017	0.8604
250, 6	Matérn-$3/2$	0.8872 (0.0026)	0.0214	0.807	0.2616	0.8897
1,000, 7	Gaussian	0.8608 (0.0014)	0.0392	0.986	0.0011	0.8601
1,000, 7	Matérn-$3/2$	0.8934 (0.0012)	0.0108	0.815	0.3128	0.8935
4,000, 9	Gaussian	0.8604 (0.0006)	0.0396	0.996	-0.0015	0.8600
4,000, 9	Matérn-$3/2$	0.8952 (0.0005)	0.0061	0.813	0.3291	0.8947
1,000, 7	Gaussian, selected	0.8983 (0.0012)	0.0094	0.750	0.3788	–
1,000, 7	Matérn-$3/2$, selected	0.8980 (0.0012)	0.0096	0.738	0.3762	–

Calibration and controls. Along the increasing-frequency sequence, the Matérn–Gaussian coverage gap widens as Matérn increasingly recovers the target cosine moment 0.4, while the Gaussian moment remains near zero (Table 12). Both kernels achieve near-nominal coverage in the low-frequency control; joint bandwidth–regularization selection also restores Gaussian coverage at $(n, K) = (1{,}000, 7)$. Thus the observed separation concerns fixed geometry and can be mitigated by adaptive tuning.

G.3 Target-sampling error and score heterogeneity

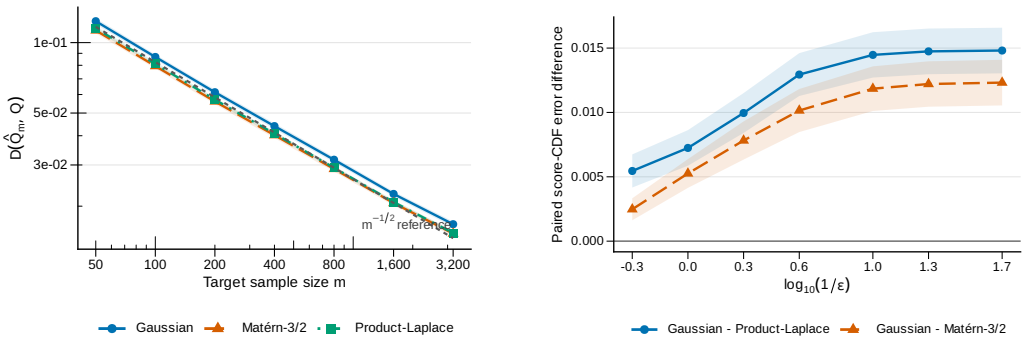
Figure 2 uses $X_{\mathrm{src}} \sim N(0, I_5)$ and $X_{\mathrm{tar}} \sim N(\mu e_1, I_5)$, with known ratio $r(x) = \exp(\mu x_1 - \mu^2/2)$ and shared response law

$$Y = f(X) + \sigma_\epsilon(X)E, \quad E \sim N(0,1), \quad E \perp X,$$
$$f(x) = 2\sin(x_1) + \tfrac{1}{2}x_2^2 - \tfrac{1}{2}x_3, \qquad \sigma_\epsilon(x) = 0.5 + 1.5\,\mathrm{logit}^{-1}(x_1/\epsilon).$$

The oracle residual score $S(X,Y) = |Y - f(X)|$ gives $h_s(x) = 2\Phi(s/\sigma_\epsilon(x)) - 1$ for $s \geq 0$, and zero otherwise. Both experiments use $d = 5$, $\mu = 2$, $n_{\mathrm{cal}} = 500$, $n_{\mathrm{test}} = 10{,}000$, and an independent target reference of size 20,000. Gaussian, product-Laplace, and Matérn-$3/2$ bandwidths are fixed at $\sqrt{d}$, d, and $\sqrt{d}$, respectively.

In panel (a), $\epsilon = 0.25$ and $m \in \{50, 100, 200, 400, 800, 1600, 3200\}$. The discrepancy between the m-sample target empirical distribution and the independent reference approximates the population-target discrepancy in Corollary 7. The fitted log–log slopes are -0.482, -0.482, and -0.486 for Gaussian, Matérn-$3/2$, and product-Laplace CFD, respectively.

In panel (b), $m = 500$ and $\epsilon \in \{2, 1, 0.5, 0.25, 0.1, 0.05, 0.02\}$. The CFD objective uses $\lambda_0 = 10^{-3}$ without a weight cap, and calibration uses $\tau = 0.90$. A fine threshold grid $\mathcal{G}$ approximates score-CDF error by $\max_{s \in \mathcal{G}} |\widehat{F}_{\mathrm{ref}}(s) - \widehat{F}_\pi(s)|$. Decreasing ϵ sharpens the conditional scale transition. Together, the two experiments examine the target-sampling and score-heterogeneity effects characterized by our theory.



(a) Finite-target sampling error.

(b) Spectral geometry and score heterogeneity.

Figure 2: Controlled experiments with fixed bandwidths. **(a)** Empirical target CFD follows an approximately $m^{-1/2}$ rate. **(b)** Sharpening the conditional scale transition increases the score-CDF error of Gaussian relative to the polynomial-tail kernels. Shaded bands are pointwise 95% t-based Monte Carlo confidence intervals for the mean paired differences, using 494 complete repetitions per setting.

G.4 Patient-level diabetes under linear and nonlinear assignment

The raw-feature diabetes data associated with Efron et al. (2004) contain 442 patients with ten baseline covariates and one observed disease-progression outcome. We retain these outcomes and assign patients to the target domain using either

$$e_{\mathrm{lin}}(x) = \mathrm{logit}^{-1}\{0.8(x_{\mathrm{BMI}} - 26)/5\}, \qquad e_{\mathrm{nonlin}}(x) = 0.5 + 0.3\tanh(z_{\mathrm{BMI}})\tanh(z_{\mathrm{BP}}).$$

Here $z_{\mathrm{BMI}} = (x_{\mathrm{BMI}} - 26.3757918552)/4.4181215606$ and $z_{\mathrm{BP}} = (x_{\mathrm{BP}} - 94.6470135747)/13.8312834198$. These full-cohort covariate means and sample standard deviations define the artificial assignment mechanism and are fixed before assignment; they are distinct from training-only method preprocessing. Each of 300 paired repetitions independently assigns patients using the stated probabilities, then partitions source patients into random-priority training and calibration halves and target patients into balancing, validation, and test thirds, without replacement (any remaining source and target patients enter calibration and test, respectively). Prediction and response-free selection follow Appendix G.1.2. Unweighted direct calibration uses the global empirical score quantile without a test atom or an $(n+1)$ rank correction; WCP retains candidate-specific test masses. Table 13 reports mean coverage, finite length, maximal-set rate, and ESS.

Under nonlinear selection, the paired absolute-gap differences in Table 14 are -0.018 percentage points for DB-CC minus Gaussian balancing (95% Monte Carlo interval $[-0.106, 0.070]$) and -0.121 points for DB-CC minus flexible WCP ($[-0.307, 0.066]$). Both intervals include zero and reflect assignment-and-partition variation within this fixed cohort, not uncertainty across new clinical cohorts.

Table 13: Diabetes results at nominal coverage 0.90 over 300 assignment-and-partition repetitions per setting.

Method	Coverage (MCSE)	Finite length	Maximal (%)	ESS/n
Linear log-odds assignment				
Unweighted direct	0.865 (0.003)	186.3	0.000	1.000
Linear logistic WCP	0.897 (0.003)	202.9	0.000	0.802
Quadratic logistic WCP	0.892 (0.003)	199.8	0.021	0.817
Selected flexible WCP	0.896 (0.003)	201.8	0.000	0.791
Matched Gaussian DB-CC	0.887 (0.003)	198.0	0.000	0.761
Selected multiscale DB-CC	0.887 (0.003)	197.6	0.000	0.774
Oracle WCP	0.915 (0.003)	213.1	0.357	0.618
Nonlinear BMI–blood-pressure assignment				
Unweighted direct	0.913 (0.002)	199.5	0.000	1.000
Linear logistic WCP	0.919 (0.002)	204.3	0.000	0.968
Quadratic logistic WCP	0.914 (0.002)	201.2	0.000	0.866
Selected flexible WCP	0.920 (0.002)	205.0	0.000	0.959
Matched Gaussian DB-CC	0.909 (0.002)	197.5	0.000	0.909
Selected multiscale DB-CC	0.910 (0.002)	198.2	0.000	0.920
Oracle WCP	0.907 (0.002)	197.5	0.000	0.783

Table 14: Selected multiscale DB-CC minus the indicated comparator on diabetes, using 300 paired repetitions per setting. Coverage and mean within-repetition absolute-gap differences, including pointwise 95% Monte Carlo intervals, are multiplied by 100. Length differences are in outcome units. Negative gap differences favor DB-CC. These Monte Carlo intervals are not adjusted for multiple comparisons.

Setting	Comparator	Coverage diff.	Gap diff. [95% MC interval]	Length diff.
Linear	Flexible WCP	-0.833	0.308 [0.064, 0.551]	-4.27
Linear	Gaussian DB-CC	0.009	-0.110 [-0.231, 0.011]	-0.41
Linear	Unweighted direct	2.244	-0.714 [-1.027, -0.401]	11.22
Nonlinear	Flexible WCP	-1.031	-0.121 [-0.307, 0.066]	-6.86
Nonlinear	Gaussian DB-CC	0.119	-0.018 [-0.106, 0.070]	0.62
Nonlinear	Unweighted direct	-0.280	0.071 [-0.059, 0.201]	-1.33

G.5 Recording-level Parkinson's evaluation and subject-disjoint sensitivity

These experiments evaluate shifted recordings from a fixed cohort. The Parkinson's Telemonitoring data contain 5,875 recordings from 42 participants, with participant membership identified by the

`subject#` field. Recording-level splits can share participants across roles and therefore do not establish new-patient performance.

The dataset is available from the UCI Machine Learning Repository. We retain 19 covariates and the observed response, and sample target records with probabilities proportional to $\exp(\gamma_1 Z_1 + \gamma_2 Z_2 Z_3)$. Mild and moderate shifts use $(\gamma_1, \gamma_2) = (0.75, 0.50)$ and $(1.25, 0.75)$. Score definitions and set inversions are in Appendix A.

Each replication draws 2,937 source records without replacement, divided into 1,468 training and 1,469 calibration records. From the complementary pool of 2,938 records, it separately resamples 734 unlabeled target records, 734 shifted test records, and 5,000 labeled reference records with replacement under the normalized tilt. These draws are independent conditional on the pool, but can repeat the same records. The full-cohort covariate scaler defines the artificial tilt, separately from method preprocessing, which pools calibration and target-balancing covariates. The implementation uses `ranger` for prediction.

Within each kernel family, 35% of target records are held out for validation. Candidates use bandwidth multipliers $\{0.5, 1, 2\}$, ridge levels $\{10^{-4}, 10^{-3}, 10^{-2}\}$, and caps $\{5, 10, 20\}$, subject to $\mathrm{ESS}/n \geq 0.20$. Selection minimizes that candidate's validation CFD plus $0.05/\sqrt{\mathrm{ESS}}$, then refits on all 734 unlabeled target records. Energy uses its standard full-distribution objective. Split-CP and CQR use 100 repetitions each, and UACQR-P uses 50 per shift. Summaries are conditional on successful fitting, with success rates reported separately; failed fits are not replaced by the uniform fallback from Appendix B.

The family-specific tuning and refitting used here differ from the common-geometry selection rule in Algorithm 1, which retains the selected weights without refitting. Accordingly, the common-validation guarantee in Theorem 3 does not directly apply to this implementation.

The reference-score CDF is the comparison distribution for reported score-CDF errors. Its own error is therefore zero by construction and is omitted in Table 2; reference-quantile coverage is evaluated on the separate test resample.

Additional nominal levels. Table 15 reports the principal split-CP comparison across five nominal coverage levels. The qualitative ordering is stable: unweighted split CP, clipped ratio WCP, and finite-moment balancing remain conservative, while energy balancing generally tracks the target-reference benchmark most closely. The three tuned CFD geometries provide similar intermediate corrections across all five levels.

Table 15: Moderate-shift split-CP coverage across nominal levels, conditional on fitting success over 100 repetitions. Known-tilt WCP includes its MCSEs and is a diagnostic, not an exact-ratio oracle under this split-cohort resampling design.

Method	0.90	0.80	0.70	0.60	0.50
Unweighted split CP	0.938	0.865	0.794	0.717	0.634
Target reference	0.899	0.800	0.699	0.599	0.502
Known-tilt WCP	0.894 (0.005)	0.788 (0.009)	0.681 (0.013)	0.575 (0.017)	0.484 (0.019)
Clipped ratio WCP	0.926	0.848	0.775	0.698	0.617
Finite-moment balancing	0.933	0.853	0.776	0.691	0.597
Energy balancing	0.903	0.803	0.712	0.617	0.521
Tuned Gaussian CFD	0.919	0.832	0.748	0.659	0.568
Tuned Product-Laplace CFD	0.917	0.830	0.744	0.655	0.562
Tuned Matérn-3/2 CFD	0.919	0.831	0.744	0.653	0.559

Interpretation of known-tilt weights. Known-tilt WCP evaluates the imposed exponential tilt at calibration and test covariates using one common numerical offset and includes the WCP test-point atom. Source calibration and target resampling use complementary finite pools, and target records are sampled with replacement. Under these conditional sampling laws, the supplied tilt is not the exact Radon–Nikodym derivative. The resulting coverage estimates, with their Monte Carlo uncertainty, therefore describe known-tilt weighting in this finite-pool experiment rather than the iid exact-ratio construction in Section 2.

Additional conformity scores. Table 16 shows the weighting interface applied to CQR and UACQR-P. Tuned CFD moves moderate-shift CQR coverage toward 0.90, compared with 0.806 unweighted. The UACQR-P procedures in the table are close to nominal under mild shift and somewhat conservative under moderate shift.

Table 16: Recording-level coverage (MCSE) at nominal 0.90, conditional on fitting success. CQR has 98/100 successful Gaussian fits and 100/100 for the other displayed methods; UACQR-P has 50/50 per method and shift. No uniform fallback replaces failed fits.

Method	CQR moderate	UACQR-P mild	UACQR-P moderate
Unweighted	0.806 (0.011)	0.911 (0.002)	0.925 (0.003)
Target reference	0.953 (0.004)	0.903 (0.002)	0.903 (0.002)
Energy balancing	0.948 (0.007)	0.903 (0.003)	0.920 (0.004)
Tuned Gaussian CFD	0.895 (0.011)	0.905 (0.002)	0.925 (0.003)
Tuned Product-Laplace CFD	0.904 (0.011)	0.906 (0.003)	0.927 (0.003)
Tuned Matérn-3/2 CFD	0.911 (0.010)	0.906 (0.002)	0.929 (0.003)

Sensitivity to tuning and weight constraints. Table 17 compares weakly regularized, uncapped fits ($\lambda_0 = 10^{-7}$) with joint response-free bandwidth/ridge/cap selection. The weak fits have mean coverage closer to nominal and shorter intervals, but lower ESS and maximum relative weights well above 20. Tuning with caps in $\{5, 10, 20\}$ yields higher ESS, longer intervals, and more conservative mean coverage. Because bandwidth, regularization, and weight constraints change together, this comparison does not isolate the effect of regularization or spectral family.

Table 17: Joint tuning and weight-constraint sensitivity for moderate-shift split CP at nominal 0.90. Weak fits use $\lambda_0 = 10^{-7}$ without a cap; tuned fits jointly select bandwidth, ridge, and cap. Means are conditional on fitting success over 100 repetitions.

Configuration	Success	Coverage (MCSE)	Length	CDF error	ESS/n	Max. relative weight
Weak Gaussian CFD	0.97	0.906 (0.003)	18.77	0.163	0.042	155.4
Weak Product-Laplace CFD	1.00	0.905 (0.002)	18.61	0.143	0.081	98.0
Weak Matérn-3/2 CFD	0.99	0.904 (0.002)	18.53	0.154	0.055	139.7
Tuned Gaussian CFD	1.00	0.919 (0.002)	19.76	0.159	0.253	18.5
Tuned Product-Laplace CFD	0.99	0.917 (0.002)	19.65	0.150	0.244	17.8
Tuned Matérn-3/2 CFD	0.98	0.919 (0.002)	19.78	0.147	0.245	17.5

Score jumps. The implementation uses the signed CQR score $\max\{\widehat{q}_\ell(x) - y,\ y - \widehat{q}_u(x)\}$, without truncation at zero, matching Appendix A. Reference records are resampled with replacement from a fixed finite pool, so their empirical score distribution is discrete and repeated records can produce exact score ties. These empirical jumps do not imply atoms in the underlying population score distribution. Exact equality is used for the primary tie calculation; a $10^{-8}\max(1, \max_i |S_i|)$ tolerance check gives a nearly identical mean jump.

Table 18 summarizes target-reference CQR over nominal levels 0.5–0.9. Deterministic calibration exhibits substantial empirical jumps and overcoverage. Randomized coverage is computed analytically by averaging over the boundary-set mixture. With the weighting rule unchanged, the absolute gap decreases from 0.1372 to 0.0121, illustrating the effect of boundary handling in Corollary 4.

Table 18: Target-reference CQR tie diagnostics under moderate shift, averaged over nominal coverage levels 0.5–0.9 and 100 repetitions.

Procedure	Mean CDF jump	Mean overcoverage	Deterministic abs. gap	Randomized abs. gap
CQR / moderate	0.2153	0.1344	0.1372	0.0121

Subject-disjoint sensitivity. Using UCI participant identifiers, we partition the 42 participants into disjoint training, calibration, balancing, reference, and test groups of sizes $9/8/8/8/9$ in each of 50 repetitions. The moderate tilt is unchanged. Balancing and reference samples draw 500

and 5,000 records, respectively, with replacement from their separate subject pools. Test coverage is evaluated over test-pool records with normalized tilt weights. The primary estimand is shift-weighted row coverage; the equal-subject estimand first averages within each test participant and then equally across participants. In this analysis, stable CFD uses fixed $\lambda_0 = 10^{-3}$ and cap 10 rather than within-family tuning.

Coverage varies substantially across the 50 participant partitions (Table 19). Target-reference calibration attains mean row coverage 0.862 (MCSE 0.028), below the nominal level. The variation is consistent with sensitivity to between-participant heterogeneity. With only 42 participants in one cohort, these comparisons provide limited precision for ranking kernels and do not assess generalization to independent cohorts.

Table 19: Subject-disjoint sensitivity at nominal 0.90. Coverage entries are mean (MCSE) across successful partitions: stable Gaussian succeeds in 49/50 and all other methods in 50/50. Row and equal-subject coverage are distinct estimands; the absolute gap uses row coverage within each partition.

Method	Success	Row coverage	Equal-subject coverage	Mean abs. gap
Unweighted split CP	1.00	0.871 (0.026)	0.864 (0.020)	0.120
Target reference	1.00	0.862 (0.028)	0.834 (0.025)	0.124
Energy balancing	1.00	0.837 (0.031)	0.842 (0.023)	0.145
Stable Gaussian CFD	0.98	0.845 (0.029)	0.843 (0.023)	0.136
Stable product-Laplace CFD	1.00	0.851 (0.028)	0.847 (0.022)	0.128
Stable Matérn-$3/2$ CFD	1.00	0.857 (0.028)	0.851 (0.022)	0.129
Finite-moment balancing	1.00	0.830 (0.030)	0.811 (0.027)	0.145
Clipped estimated-ratio WCP	1.00	0.874 (0.026)	0.860 (0.021)	0.113
Estimated-ratio direct	1.00	0.816 (0.035)	0.806 (0.030)	0.158
Known-tilt direct	1.00	0.787 (0.036)	0.775 (0.032)	0.182
Known-tilt WCP	1.00	0.791 (0.036)	0.777 (0.032)	0.180