Adapt the Score or Localize the Calibration? A Finite-Sample Comparison for Conditional Conformal Prediction

Anonymous authors
Paper under double-blind review

Abstract

Given a fixed nonconformity score S and target coverage $1-\alpha$, should extra labels transform the score before global calibration, or calibrate it locally? Under score regularity, we bound both routes' errors in conditional coverage and oracle-relative length, with high probability over calibration, given the training data. Subtracting an estimated conditional score quantile $q_S^{1-\alpha}(\cdot)$ before global calibration gives a three-term error bound, valid simultaneously wherever the score assumptions hold: the pointwise quantile-estimation error, its average over covariates, and a calibration fluctuation. We prove that this average term is needed: an error confined to one region causes over-coverage elsewhere, even with population calibration. For randomly localized conformal prediction (RLCP), we bound localization bias by the target quantile's variation alone; the conditional score distribution need not be continuous in the covariate. These bounds hold over the realized centre's neighbourhood and account for the smaller effective calibration sample. We apply the global bounds to local-polynomial and deep-network quantile estimators. Neural estimators can exploit additive or compositional structure, whereas our RLCP bound depends on neighbourhood data mass in the full covariate space. The bounds favour quantile learning when the fitting data can capture this structure; local calibration remains useful when the fit is poor but nearby data are plentiful. Simulations and eight real datasets illustrate the global route's gains; low-dimensional simulations show when localization is preferable.

1 Introduction

Conformal prediction guarantees finite-sample marginal coverage under exchangeability (Vovk et al., 2005; Shafer and Vovk, 2008; Lei et al., 2018). Even when marginal coverage is at the nominal level, under-coverage in some regions can be offset by over-coverage in others. Without distributional assumptions, informative prediction sets cannot guarantee conditional coverage in finite samples (Lei and Wasserman, 2014; Barber et al., 2021b). We study coverage and length errors relative to the conditional oracle at a fixed covariate value and sample size.

Fix a miscoverage level $\alpha \in (0,1)$. Split conformal prediction uses m independent calibration observations to estimate the *marginal* $(1-\alpha)$-quantile of a score $S(X,Y)$. Conditional coverage at x instead calls for its *conditional* quantile given $X=x$, written $q_S^{1-\alpha}(x)$: the conditional score-quantile map. Two approaches address this gap, termed score-based and calibration-based by Min et al. (2026). Start from a score S, for example $|y-\hat{\mu}(x)|$, possibly standardized (Lei et al., 2018); the predictor $\hat{\mu}$ may be a black box whose training data are no longer available. One can *adapt the score*: learn an approximation $\widehat{S}$ to a transformed score $\widetilde{S}(x,y) = T_x(S(x,y))$ whose conditional quantile is constant in x, then calibrate $\widehat{S}$ globally. The common oracle threshold gives conditional coverage at least $1-\alpha$. A simple choice is $\widetilde{S}(x,y) = S(x,y) - q_S^{1-\alpha}(x)$, estimated by $\widehat{S}(x,y) = S(x,y) - \widehat{q}_S^{1-\alpha}(x)$ with $\widehat{q}_S^{1-\alpha}(\cdot)$ fitted on a separate fold (Han et al., 2022; Plassier et al., 2025a). Score adaptation is

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.

established: examples include normalized residuals (Papadopoulos et al., 2008; Lei et al., 2018), conditional-CDF scores (Chernozhukov et al., 2021), input-conditional normalizing flows (Colombo, 2024), and rectification (Plassier et al., 2025a). Or one can *localize the calibration*: keep S and estimate $q_S^{1-\alpha}(x)$ by a kernel-weighted quantile of nearby calibration scores. Localized conformal prediction (Guan, 2023) and its randomized version RLCP (Hore and Barber, 2025) follow this route; RLCP draws an auxiliary centre to retain finite-sample marginal validity.

Both routes receive N additional labelled observations, independent of any data used to fit S. The transformed-score route splits them between learning and global calibration, $N = n + m$; RLCP uses all N to calibrate S locally. The initial fitting cost is shared; transformation fitting and conformal calibration use disjoint folds. The common oracle is $\{y : S(x, y) \le q_S^{1-\alpha}(x)\}$. Quantile subtraction gives explicit bounds without restricting the general approach to that transformation (Section 4).

For both routes, we bound conditional-coverage and length errors relative to this oracle, with high probability over calibration, conditional on training and, for RLCP, on the realized centre.

The target $\widetilde{S}$ only needs to align the chosen quantile across covariates, not the whole conditional score distribution. We call this property *adaptivity* (Definition 1). For its learned approximation $\widehat{S}$, write the pointwise error $\Delta_{\widetilde{S}}(x) = \sup_{y\in\mathbb{R}} |\widehat{S}(x, y) - \widetilde{S}(x, y)|$. For quantile subtraction, this equals $|\widehat{q}_S^{1-\alpha}(x) - q_S^{1-\alpha}(x)|$.

We make three contributions. **1** Global calibration combines the pointwise error $\Delta_{\widetilde{S}}(x)$, its covariate average, and a fluctuation of order $m^{-1/2}$ at fixed confidence. We prove that the averaged term is necessary: a score error confined to one region causes over-coverage of the same order outside that region, where the score is exact, even under population calibration (Proposition 3). **2** We bound the RLCP localization bias using only variation of $q_S^{1-\alpha}(\cdot)$ within the neighbourhood, not of the full conditional score distribution (Theorem 4). The bias can therefore vanish even when that distribution is discontinuous in the covariate (Remark 20); the remaining error comes from the effective local calibration sample. **3** Our comparison favours learning the score quantile before global calibration when the fitting budget is sufficient to exploit low-dimensional structure. For compositional maps, the coverage and length rate exponents depend on the dimensions and smoothness of the component functions (Proposition 6). Additive and single-index maps give one-dimensional exponents even when all covariates matter, with a factor polynomial in the dimension (Corollary 24). The RLCP bound instead depends on neighbourhood mass in the full covariate space (Proposition 8). Exploiting structure can therefore offset the labels spent on fitting; higher-order smoothness offers another advantage (Section 4). Local calibration remains useful when the fit is inaccurate but the quantile varies little over well-populated neighbourhoods. Local-polynomial fitting gives uniform guarantees; neural fitting gives root-mean-square guarantees over the covariate distribution, both with high probability over fitting and calibration (Propositions 6 and 7). These comparisons concern class-tuned upper bounds in their stated norms and domains.

The structured simulations favour quantile learning over ambient RLCP in root mean square coverage error, as do group diagnostics on most of eight real datasets; low-dimensional examples show when local calibration wins and how label allocation changes the comparison.

Related work. Min et al. (2026) analyze score-based and calibration-based methods through a common decomposition of conditional-coverage error. They give method-specific rates and a selection rule based on conditional coverage. We compare these routes under a shared score and additional label budget, and also control oracle-relative length. Laplante (2026) gives pointwise-plus-averaged bounds and high-probability oracle comparisons using a conditional-CDF transformation; our single-level condition does not require full distributional pivotality, and we give a lower bound showing why the averaged error is needed. Conrad et al. (2026) give high-probability RLCP coverage and oracle-relative length bounds over the realized neighbourhood; our fixed-score analysis controls localization bias through variation of the target quantile alone. Section A gives the full discussion.

Data and scores. Let (X, Y) take values in $\mathcal{X} \times \mathbb{R}$ with joint law $P_{XY} = P_X \otimes P_{Y|X}$, where $\mathcal{X}$ is a measurable space and $x \mapsto P_{Y|X=x}$ is a fixed regular conditional version, to which all conditional statements refer. From $n + m$ i.i.d. observations we form a training sample $\mathcal{D}_n^{\mathrm{tr}}$ of size n and a calibration sample $\mathcal{D}_m^{\mathrm{cal}} = \{(X_i, Y_i)\}_{i=1}^m$, independent of each other and of the test pair $(X, Y) \sim P_{XY}$. A score is a measurable map $S\colon \mathcal{X} \times \mathbb{R} \to \mathbb{R}$, not necessarily adaptive. If it was fitted on earlier data, we condition on that fit and take the training and calibration samples and test pair independent of those data. Its conditional and marginal distribution functions are

$$F^{(S|X)}(t \mid x) \coloneqq \mathbb{P}\big(S(x, Y) \le t \mid X = x\big), \qquad F^{(S)}(t) \coloneqq \mathbb{P}\big(S(X, Y) \le t\big).$$

For a nondecreasing right-continuous $H\colon \mathbb{R} \to [0, \infty)$ with $\lim_{t\to-\infty} H(t) = 0$ and $\tau > 0$, write $\mathrm{Q}(\tau; H) \coloneqq \inf\{t \in \mathbb{R} : H(t) \ge \tau\}$, with $\inf \varnothing = +\infty$. At the level $1-\alpha$, the conditional and marginal score quantiles are

$$q_S^{1-\alpha}(x) \coloneqq \mathrm{Q}(1-\alpha; F^{(S|X)}(\cdot \mid x)), \qquad \bar{q}_S \coloneqq \mathrm{Q}(1-\alpha; F^{(S)}). \tag{1}$$

Sublevel sets are $\mathcal{C}_S(x, t) \coloneqq \{y \in \mathbb{R} : S(x, y) \le t\}$; $|C|$ denotes the Lebesgue measure of a measurable $C \subset \mathbb{R}$. The set $\mathcal{C}_S(x, q_S^{1-\alpha}(x))$, thresholded at the conditional quantile, is the *conditional oracle* of S: it covers Y with conditional probability at least $1-\alpha$ at x, and exactly $1-\alpha$ when $F^{(S|X)}(\cdot \mid x)$ is continuous at $q_S^{1-\alpha}(x)$. The set $\mathcal{C}_S(x, \bar{q}_S)$, thresholded at the marginal quantile, is the *global oracle*: the set that split-conformal calibration targets. For a set-valued C, $\mathrm{Cov}_C(x) \coloneqq \int_{\mathbb{R}} \mathbb{1}_{C(x)}(y) P_{Y|X=x}(dy)$ is its conditional-coverage profile.

The pointwise score assumptions are imposed on a fixed measurable set $\mathcal{X}_0 \subseteq \mathcal{X}$ of full P_X-measure. Assumptions involving a quantile level are indexed by $(S, 1-\alpha)$; their constants and radii may depend on this pair, with no uniformity over coverage levels assumed. The following regularity conditions apply to any score S.

A 1 $(S, 1-\alpha)$**.** *There are constants $0 < \sigma_{\mathrm{low}} \le \sigma_{\mathrm{up}} < \infty$ and radii $r_-, r_+ > 0$ such that for every $x \in \mathcal{X}_0$: (a) $0 \le F^{(S|X)}(t' \mid x) - F^{(S|X)}(t \mid x) \le \sigma_{\mathrm{up}}(t' - t)$ for all $t \le t'$; (b) $q_S^{1-\alpha}(x)$ is finite, $F^{(S|X)}(q_S^{1-\alpha}(x) \mid x) - F^{(S|X)}(q_S^{1-\alpha}(x) - u \mid x) \ge \sigma_{\mathrm{low}} u$ for $0 \le u \le r_-$, and $F^{(S|X)}(q_S^{1-\alpha}(x) + v \mid x) - F^{(S|X)}(q_S^{1-\alpha}(x) \mid x) \ge \sigma_{\mathrm{low}} v$ for $0 \le v \le r_+$.*

In particular, $q_S^{1-\alpha}(x)$ may vary with x: this condition will be used for the initial score in both routes. For length bounds, we also require:

A 2 (S)**.** *For every $x \in \mathcal{X}_0$, the map $t \mapsto |\mathcal{C}_S(x, t)|$ is finite and L_{len}-Lipschitz on $\mathbb{R}$.*

2 Adaptive Scores and Global Calibration

The following definition formalizes the quantile alignment described in the introduction.

Definition 1 ($(1-\alpha)$-adaptive score)**.** *A score S is $(1-\alpha)$-*adaptive *(for P_{XY}) if $q_S^{1-\alpha}(x) = \bar{q}_S$ for P_X-almost every $x \in \mathcal{X}$.*

Adaptivity is a property of the pair (S, P_{XY}) at the level $1-\alpha$; a score can be adaptive at one level and not at another. The idea of adapting a score to obtain conditional coverage is established: examples include oracle scale-normalized residuals under a location–scale model (Papadopoulos et al., 2008; Lei et al., 2018), oracle CQR (Romano et al., 2019), and conditional-CDF scores (Chernozhukov et al., 2021). Learned score transformations include input-conditional normalizing flows (Colombo, 2024) and rectification (Plassier et al., 2025a); Section A discusses their respective assumptions and guarantees. We use this single-level condition to derive calibration bounds and compare global and local calibration.

For the bounds in this section, S denotes the adaptive reference score, written $\widetilde{S}$ in the introduction; $\widehat{S}$ denotes its learned approximation. In practice the score is produced by the training fold: $\widehat{S} = \widehat{S}_{\mathcal{D}_n^{\mathrm{tr}}}$; all statements about a learned score are made conditionally on a realization of $\mathcal{D}_n^{\mathrm{tr}}$: $\widehat{S}$ is then a fixed score and the high-probability events are over the calibration fold alone. For any reference score S, define the pointwise score error

$$\Delta_S(x) \coloneqq \sup_{y \in \mathbb{R}} \big|\widehat{S}(x, y) - S(x, y)\big| \in [0, +\infty], \tag{2}$$

assumed measurable, as when $\widehat{S}(x,\cdot)$ and $S(x,\cdot)$ are continuous, the case in all our examples. The subscript always identifies the reference score; the learned score being compared with it is specified in context. The bounds use $\mathbf{A}\,1(S,1-\alpha)$ and the pivot condition below; length bounds also require $\mathbf{A}\,2(S)$. The learned score $\widehat{S}$ need not be adaptive; it is the score used for calibration, and Δ_S measures its error relative to S.

Given $\widehat{S}$, compute the calibration scores $\widehat{S}_i \coloneqq \widehat{S}(X_i,Y_i)$, $i \le m$, their empirical CDF $\widehat{F}_m^{(\widehat{S})}(t) \coloneqq m^{-1}\sum_{i=1}^m \mathbb{1}_{\{\widehat{S}_i \le t\}}$, and

$$\beta_m \coloneqq \frac{\lceil (m+1)(1-\alpha)\rceil}{m}, \qquad \widehat{Q}_m \coloneqq \mathrm{Q}(\beta_m; \widehat{F}_m^{(\widehat{S})}), \qquad \hat{\mathcal{C}}_\alpha(x) \coloneqq \mathcal{C}_{\widehat{S}}(x, \widehat{Q}_m), \tag{3}$$

with $\hat{\mathcal{C}}_\alpha(x) = \mathbb{R}$ when $\widehat{Q}_m = +\infty$. Conditionally on $\mathcal{D}_n^{\mathrm{tr}}$ the calibration and test scores are exchangeable, so $\mathbb{P}(Y \in \hat{\mathcal{C}}_\alpha(X) \mid \mathcal{D}_n^{\mathrm{tr}}) \ge 1-\alpha$ almost surely (Vovk et al., 2005; Lei et al., 2018). In addition to $\mathbf{A}\,1(S,1-\alpha)$, we require the reference score S to have a common threshold with exact conditional coverage. The pivot may depend on $(S,1-\alpha)$.

A 3 $(S,1-\alpha)$**.** *There is $q^\star \in \mathbb{R}$ such that $F^{(S|X)}(q^\star \mid x) = 1-\alpha$ for every $x \in \mathcal{X}_0$.*

$\mathbf{A}3(S,1-\alpha)$ makes $q^\star$ a *pivot* for S, a common threshold at which every conditional score CDF attains level $1-\alpha$; with $\mathbf{A}\,1(S,1-\alpha)$, S is then $(1-\alpha)$-adaptive with $q_S^{1-\alpha}(x) = \bar{q}_S = q^\star$ on $\mathcal{X}_0$. The calibration step is described by the rank slack and the radius of a $(1-\delta)$ Dvoretzky–Kiefer–Wolfowitz (DKW) band for the calibration-score CDF (Dvoretzky et al., 1956; Massart, 1990),

$$s_m \coloneqq \beta_m - (1-\alpha) \in \big[(1-\alpha)/m, (2-\alpha)/m\big), \qquad \eta_m(\delta) \coloneqq \sqrt{\tfrac{\log(2/\delta)}{2m}}. \tag{4}$$

Theorem 2 (Global calibration of a learned score, pointwise)**.** *Let $\mathbf{A}\,1(S,1-\alpha)$ and $\mathbf{A}\,3(S,1-\alpha)$ hold, fix $\delta \in (0,1)$, and suppose that*

$$\sigma_{\mathrm{up}}\|\Delta_S\|_{L^1(P_X)} + \eta_m(\delta) + s_m \le \sigma_{\mathrm{low}}(r_- \wedge r_+). \tag{5}$$

Then there is an event $\mathcal{E}_m(\delta)$, measurable with respect to $\mathcal{D}_m^{\mathrm{cal}}$, with $\mathbb{P}(\mathcal{E}_m(\delta) \mid \mathcal{D}_n^{\mathrm{tr}}) \ge 1-\delta$, on which, for every $x \in \mathcal{X}_0$,

$$\big|\mathrm{Cov}_{\hat{\mathcal{C}}_\alpha}(x) - (1-\alpha)\big| \vee \big||\hat{\mathcal{C}}_\alpha(x)| - |\mathcal{C}_S(x,q^\star)|\big| \lesssim \sigma_{\mathrm{up}}\big(\Delta_S(x) + \|\Delta_S\|_{L^1(P_X)}\big) + \sqrt{\frac{\log(2/\delta)}{m}} + \frac{1}{m}, \tag{6}$$

the length part under $\mathbf{A}\,2(S)$; the constant hidden in $\lesssim$ depends only on the ratio $\sigma_{\mathrm{up}}/\sigma_{\mathrm{low}}$ for the coverage part and on $L_{\mathrm{len}}/\sigma_{\mathrm{low}}$ for the length part.

Condition (5) keeps the calibrated threshold within the growth window around the pivot on $\mathcal{E}_m(\delta)$; the complete statement with constants and proof is in Section B. Calibration contributes a fluctuation of order $m^{-1/2}$ and a rank correction of order m^{-1} at fixed confidence. The averaged score error is a different cost: it persists even with population calibration, where there is no sampling error. The following lower bound shows that the dependence on $\|\Delta_S\|_{L^1(P_X)}$ is necessary.

Proposition 3 (Propagation of score error)**.** *Let $\mathbf{A}\,1(S,1-\alpha)$ and $\mathbf{A}\,3(S,1-\alpha)$ hold. For a measurable $A \subset \mathcal{X}_0$ with $a \coloneqq P_X(A) \in (0,1)$, set $\widehat{S}(x,y) = S(x,y) + c\mathbb{1}_{\{x\in A\}}$, where $0 < c \le r_-$ and $(\sigma_{\mathrm{up}}/\sigma_{\mathrm{low}})ac \le r_+$. Using the population marginal quantile $\bar{q}_{\widehat{S}}$ as the global threshold gives, for every $x \in \mathcal{X}_0 \setminus A$,*

$$F^{(\widehat{S}|X)}(\bar{q}_{\widehat{S}} \mid x) - (1-\alpha) \ge \frac{\sigma_{\mathrm{low}}^2}{\sigma_{\mathrm{up}}}\|\Delta_S\|_{L^1(P_X)} = \frac{\sigma_{\mathrm{low}}^2}{\sigma_{\mathrm{up}}}ac, \qquad \Delta_S(x) = 0. \tag{7}$$

Population marginal coverage is exactly $1-\alpha$, yet the score error confined to A causes over-coverage at every $x \in \mathcal{X}_0 \setminus A$, where the score is exact. The threshold bounds and proof are in Section B.

3 Localized Calibration: RLCP with a Fixed Score

Throughout this section $\mathcal{X} \subset \mathbb{R}^d$, and $B(c,r)$ is the closed Euclidean ball. Localization uses a fixed kernel $\mathcal{K}\colon \mathbb{R}^d \to [0,\infty)$ satisfying the following throughout.

K 1 (Kernel). *The kernel $\mathcal{K}$ is bounded, symmetric, integrates to one, is supported in $B(0,1)$, and satisfies $\mathcal{K}(v) \geq \kappa \mathbb{1}_{\{\|v\| \leq \eta\}}$ for some $\kappa > 0$ and $\eta \in (0,1]$.*

For a centre $c \in \mathbb{R}^d$ and a bandwidth $h > 0$ put $w_{c,h}(u) \coloneqq \mathcal{K}((u-c)/h)$ and $Z_{c,h} \coloneqq \mathbb{E}[w_{c,h}(X)]$. When $Z_{c,h} > 0$, the localized design law $\Lambda_{c,h}(\mathrm{d}u) \coloneqq w_{c,h}(u)\, P_X(\mathrm{d}u)/Z_{c,h}$ is a probability measure supported in $B(c,h) \cap \mathcal{X}$. For a score S, its localized CDF is $F^{(S)}_{c,h}(t) \coloneqq \int_{\mathcal{X}} F^{(S|X)}(t \mid u)\, \Lambda_{c,h}(\mathrm{d}u)$, $t \in \mathbb{R}$. The local analogue of the calibration size is $m_{c,h} \coloneqq m\, Z_{c,h}/\|\mathcal{K}\|_\infty$.

P 1 (Design regularity). *P_X has a density p_X with $p_X \geq p_{\min}$ on $\mathcal{X}$, for some $p_{\min} > 0$. At the centre $\tilde{x} \in \mathbb{R}^d$ and bandwidth $h > 0$, $|\mathcal{X} \cap B(\tilde{x}, \eta h)| \geq c_{\mathrm{reg}} |B(\tilde{x}, \eta h)|$ for some $c_{\mathrm{reg}} \in (0,1]$.*

Under **K**1 and **P**1, $m_{\tilde{x},h} \geq c_K m p_{\min} h^d$, where $c_K > 0$ depends only on the kernel and support geometry (Remark 13). Randomly localized conformal prediction (RLCP), introduced by Hore and Barber (2025), calibrates around a randomized centre: given X, an auxiliary point $\widetilde{X}$ is drawn with density $h^{-d}\mathcal{K}((\cdot - X)/h)$, independently of $(\mathcal{D}^{\mathrm{tr}}_n, \mathcal{D}^{\mathrm{cal}}_m, Y)$; under **K** 1, the support of $\mathcal{K}$ gives $\|X - \widetilde{X}\| \leq h$. For the realized centre $\widetilde{X} = \tilde{x}$ and a candidate covariate x, set $D(x) \coloneqq \sum_{j \leq m} w_{\tilde{x},h}(X_j) + w_{\tilde{x},h}(x)$ and

$$\widetilde{w}_i(x) \coloneqq w_{\tilde{x},h}(X_i)/D(x) \ (i \leq m), \qquad \widetilde{w}_{m+1}(x) \coloneqq w_{\tilde{x},h}(x)/D(x), \tag{8}$$

and, with $S_i \coloneqq S(X_i, Y_i)$, let $\widehat{F}^{\mathrm{R}}_{x,\tilde{x}}(t) \coloneqq \sum_{i \leq m} \widetilde{w}_i(x) \mathbb{1}_{\{S_i \leq t\}}$, the test score being represented by the mass $\widetilde{w}_{m+1}(x)$ at $+\infty$. Put $\widehat{Q}^{\mathrm{R}}(x;\tilde{x}) \coloneqq \mathrm{Q}(1-\alpha; \widehat{F}^{\mathrm{R}}_{x,\tilde{x}})$, equal to $+\infty$ if $D(x) = 0$, and $\widehat{\mathcal{C}}^{\mathrm{R}}(x;\tilde{x}) \coloneqq \mathcal{C}_S(x, \widehat{Q}^{\mathrm{R}}(x;\tilde{x}))$. Conditionally on $\widetilde{X} = \tilde{x}$ the test covariate has law $\Lambda_{\tilde{x},h}$ while calibration keeps P_X: RLCP is weighted conformal prediction under known covariate shift (Tibshirani et al., 2019), and $\mathbb{P}(Y \in \widehat{\mathcal{C}}^{\mathrm{R}}(X; \widetilde{X})) \geq 1 - \alpha$ for every score (Hore and Barber, 2025, Theorem 1). We complement this marginal guarantee with pointwise statements conditional on the realized centre; since $\widetilde{X} \perp (\mathcal{D}^{\mathrm{cal}}_m, Y)$ given X and $\mathcal{D}^{\mathrm{cal}}_m \perp (X, Y)$, these are the fixed-centre statements evaluated at $\tilde{x} = \widetilde{X}$.

For a fixed score S satisfying $\mathbf{A}\, 1(S, 1-\alpha)$, the target at x is the conditional oracle $\mathcal{C}_S(x, q^{1-\alpha}_S(x))$. The conditional quantile may vary with x. Local calibration combines score distributions from nearby covariates. Its bias is controlled by the largest difference between their conditional quantiles,

$$\omega_S(\tilde{x}, h) \coloneqq \sup\big\{ |q^{1-\alpha}_S(u) - q^{1-\alpha}_S(u')| : u, u' \in B(\tilde{x}, h) \cap \mathcal{X}_0 \big\} \in [0, \infty]. \tag{9}$$

Theorem 4 (RLCP with a fixed score). *Fix $h > 0$, a centre $\tilde{x}$ and $\delta \in (0,1)$. Assume **K 1** and **P 1**, **A** 1$(S, 1-\alpha)$ and $m p_{\min} h^d \geq C_{\mathrm{R}} \log(4/\delta)$, with C_{R} as in* (31). *With probability at least $1 - \delta$ over $\mathcal{D}^{\mathrm{cal}}_m$, simultaneously for every $x \in B(\tilde{x}, h) \cap \mathcal{X}_0$, $\widehat{Q}^{\mathrm{R}}(x;\tilde{x})$ is finite and*

$$\big|\mathrm{Cov}_{\widehat{\mathcal{C}}^{\mathrm{R}}(\cdot;\tilde{x})}(x) - (1-\alpha)\big| \vee \big||\widehat{\mathcal{C}}^{\mathrm{R}}(x;\tilde{x})| - |\mathcal{C}_S(x, q^{1-\alpha}_S(x))|\big| \lesssim \sigma_{\mathrm{up}} \omega_S(\tilde{x}, h) + \sqrt{\frac{\log(4/\delta)}{m p_{\min} h^d}}, \tag{10}$$

*the length bound requiring **A** 2(S). The implicit constants depend only on $(\sigma_{\mathrm{up}}/\sigma_{\mathrm{low}}, c_K)$ for coverage and $(L_{\mathrm{len}}/\sigma_{\mathrm{low}}, c_K)$ for length.*

The complete statement and proof are in Section C.3.

The bias bound uses only variation of the target quantile, not of the full conditional score distribution. $\mathbf{A}\, 1(S, 1-\alpha)$ imposes regularity in the score threshold, but no continuity in the covariate. In particular, the bias vanishes when $q^{1-\alpha}_S(\cdot)$ is constant throughout the neighbourhood, even if the conditional score CDF is discontinuous in x (Remark 20).

4 Comparison with a Common Score

Condition on any initial data $\mathcal{D}_0$ used to fit the finite measurable score S, which remains fixed and need not be adaptive. Both routes use N additional i.i.d. observations, independent of $\mathcal{D}_0$, and target the same conditional oracle

$$C_S^\circ(x) \coloneqq \{y : S(x,y) \le q_S^{1-\alpha}(x)\}. \tag{11}$$

The global route splits $N = n + m$ between learning a transformation and calibrating the transformed score; RLCP uses all N to calibrate the unchanged score locally. When can learning offset this loss of calibration data? We use quantile subtraction (Han et al., 2022; Plassier et al., 2025a) to make the comparison explicit. For a finite measurable estimate $\widehat{q}_S^{1-\alpha}(\cdot)$ of $q_S^{1-\alpha}(\cdot)$, fitted on $\mathcal{D}_n^{\mathrm{tr}}$ independently of $\mathcal{D}_m^{\mathrm{cal}}$, set

$$S^{\mathrm{ad}}(x,y) \coloneqq S(x,y) - q_S^{1-\alpha}(x), \qquad \widehat{S}^{\mathrm{ad}}(x,y) \coloneqq S(x,y) - \widehat{q}_S^{1-\alpha}(x). \tag{12}$$

Here S^{ad} is the adaptive reference and $\widehat{S}^{\mathrm{ad}}$ its approximation; S remains the initial score. Global calibration of $\widehat{S}^{\mathrm{ad}}$ gives

$$\hat{\mathcal{C}}_\alpha^{\mathrm{ad}}(x) = \{y : S(x,y) \le \widehat{q}_S^{1-\alpha}(x) + \widehat{Q}_m\}, \tag{13}$$

where $\widehat{Q}_m$ is the split-conformal threshold of the calibration scores $S(X_i,Y_i) - \widehat{q}_S^{1-\alpha}(X_i)$. RLCP instead thresholds S directly by its locally weighted quantile (8).

Proposition 5 (Quantile subtraction)**.** *Under* $\boldsymbol{A}$ *1*$(S, 1-\alpha)$*,* S^{ad} *is* $(1-\alpha)$*-adaptive with pivot* 0 *and the same score-regularity constants, and* $\mathcal{C}_{S^{\mathrm{ad}}}(x,0) = C_S^\circ(x)$ *on* $\mathcal{X}_0$*. If* $\boldsymbol{A}$ *2*(S) *holds, it transfers with the same* L_{len}*. Moreover,*

$$\Delta_{S^{\mathrm{ad}}}(x) = |\widehat{q}_S^{1-\alpha}(x) - q_S^{1-\alpha}(x)|. \tag{14}$$

Only the chosen quantile needs to be constant; the full conditional score law need not be. Theorem 2 applies with the quantile-estimation error (14); we give two ways to control it.

Conditions for the rates. For the isotropic rates below, condition on $\mathcal{D}_0$ and assume $\mathcal{X} = [0,1]^d$, $0 < p_{\min} \le p_X \le p_{\max} < \infty$, and $q_S^{1-\alpha}(\cdot) \in \mathcal{H}^{\beta_q}(\mathcal{X}, L_q)$, bounded by M, for a fixed $\beta_q > 0$. Assume also the CDF bounds (42): a global upper Lipschitz bound and a lower increment bound throughout a fixed-radius window around $q_S^{1-\alpha}(x)$. These imply $\mathbf{A}\,1(S, 1-\alpha)$; length conclusions additionally require $\mathbf{A}\,2(S)$. Section D.2 gives the full conditions; no bounded score support is required.

1. Global quantile regression with a neural network. On the training fold, form responses $Z_i = S(X_i, Y_i)$ and fit one function to all pairs (X_i, Z_i) by minimizing the pinball loss $\rho_\tau(u) = u(\tau - \mathbb{1}_{\{u<0\}})$ at $\tau = 1-\alpha$:

$$\widehat{f} \in \operatorname*{arg\,min}_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^{n} \rho_{1-\alpha}(Z_i - f(X_i)), \qquad \widehat{q}_S^{1-\alpha}(x) = T_M(\widehat{f}(x)), \tag{15}$$

where $T_M(z) = \max\{-M, \min\{M, z\}\}$ and $\mathcal{F}_n$ is a sparse ReLU class (Schmidt-Hieber, 2020; Madrid Padilla et al., 2022) tuned to the function class below (Section D.4). This is quantile regression on the *score*, not on Y. Networks can exploit structure that ambient localization does not use. Let $q_S^{1-\alpha}(\cdot) = g_J \circ \cdots \circ g_0$, where each scalar component of g_j depends on at most t_j of its inputs and is β_j-Hölder (Definition 23); the components and their active coordinates are not supplied to the estimator. Its exponent is $\phi = \min_j \beta_j^*/(2\beta_j^* + t_j)$, with $\beta_j^* = \beta_j \prod_{\ell>j}(\beta_\ell \wedge 1)$; an isotropic β_q-Hölder map is the case $J = 0$, $t_0 = d$, with $\phi = \beta_q/(2\beta_q + d)$.

Proposition 6 (Neural quantile fitting followed by global calibration)**.** *Under the rate conditions above, with the Hölder ball replaced by such a composition class, use* (15) *and* (13)*. For each* $\delta \in (0,1)$ *and sufficiently large* n, m*, with probability at least* $1-\delta$ *over both folds,*

$$\|\operatorname{Cov}_{\hat{\mathcal{C}}_\alpha^{\mathrm{ad}}} - (1-\alpha)\|_{L^2(P_X)} \ \vee \ \| |\hat{\mathcal{C}}_\alpha^{\mathrm{ad}}(\cdot)| - |C_S^\circ(\cdot)| \|_{L^2(P_X)} \lesssim n^{-\phi}(\log n)^{3/2} + \sqrt{\tfrac{\log(4/\delta)}{n \wedge m}}. \tag{16}$$

This is an $L^2(P_X)$ guarantee for an empirical risk minimizer, not a uniform-in-x rate or a guarantee for an arbitrary output of network training. Additive and single-index maps with univariate smoothness β_q give the exponent $\beta_q/(2\beta_q+1)$ even when all d covariates matter, with a factor polynomial in d (Corollary 24).

2. Local-polynomial quantile regression. Alternatively, at each x, fit a polynomial to nearby score responses using the same pinball loss (Chaudhuri, 1991b; Guerre and Sabbah, 2012). With degree $s_q = \lceil \beta_q \rceil - 1$, bandwidth $b > 0$ and multi-indices j, set

$$\widehat{\theta}(x) \in \arg\min_{\theta=(\theta_j)_{|j|\le s_q}} \sum_{i=1}^n \mathcal{K}\big(\tfrac{X_i-x}{b}\big)\, \rho_{1-\alpha}\Big(Z_i - \textstyle\sum_{|j|\le s_q} \theta_j \big(\tfrac{X_i-x}{b}\big)^j\Big), \tag{17}$$

and set $\widehat{q}_S^{1-\alpha}(x) = T_M\big(\widehat{\theta}_0(x)\big)$. The polynomial fit is local, but calibration of the transformed score is still *global*: the same $\widehat{Q}_m$ is used at every x.

Proposition 7 (Local-polynomial fitting followed by global calibration)**.** *Under the rate conditions above, let $\mathcal{K}$ satisfy **K** 1 and be Lipschitz. Use* (17) *with $b = (\log(16n/\delta)/n)^{1/(2\beta_q+d)}$, then* (13)*. For each $\delta \in (0,1)$ and sufficiently large n, m, with probability at least $1-\delta$ over both folds, simultaneously for every $x \in \mathcal{X}$,*

$$\big|\mathrm{Cov}_{\hat{\mathcal{C}}_\alpha^{\mathrm{ad}}}(x) - (1-\alpha)\big| \;\vee\; \big|\,|\hat{\mathcal{C}}_\alpha^{\mathrm{ad}}(x)| - |C_S^\circ(x)|\,\big| \lesssim \Big(\tfrac{\log(n/\delta)}{n}\Big)^{\beta_q/(2\beta_q+d)} + \sqrt{\tfrac{\log(4/\delta)}{m}}. \tag{18}$$

Both propositions transfer established estimation bounds through Theorem 2; proofs and sample-size conditions are in Sections D.4 and D.5.

3. Local calibration of the original score. RLCP fits no transformation and uses all N additional observations for calibration. The Hölder regularity assumed above gives $\omega_S(\tilde{x}, h) \lesssim h^{\bar{\beta}_q}$, with $\bar{\beta}_q = \min(\beta_q, 1)$: for $\beta_q > 1$, this uses only bounded first derivatives. Balancing this variation against the calibration fluctuation in Theorem 4 gives the following rate.

Proposition 8 (RLCP for the common score)**.** *Under the rate conditions above and **K** 1, fix $\delta \in (0,1)$ and set $h_\star = (\log(4/\delta)/(Np_{\min}))^{1/(2\bar{\beta}_q+d)}$. Run RLCP on S with $m = N$. Condition on a realized centre $\tilde{x}$ satisfying **P** 1 at $(\tilde{x}, h_\star)$. If $Np_{\min} \ge c_{\mathrm{R}} \log(4/\delta)$, with c_{R} as specified in Section D.6, then with probability at least $1-\delta$ over calibration, simultaneously for every $x \in B(\tilde{x}, h_\star) \cap \mathcal{X}$,*

$$\big|\mathrm{Cov}_{\hat{\mathcal{C}}^{\mathrm{R}}(\cdot;\tilde{x})}(x) - (1-\alpha)\big| \;\vee\; \big|\,|\widehat{\mathcal{C}}^{\mathrm{R}}(x;\tilde{x})| - |C_S^\circ(x)|\,\big| \lesssim \Big(\frac{\log(4/\delta)}{Np_{\min}}\Big)^{\bar{\beta}_q/(2\bar{\beta}_q+d)}. \tag{19}$$

The dependence on d here comes from fixed neighbourhoods in the original covariate space. The weights can instead be learned: Lin et al. (2021) fit a low-rank Mahalanobis distance on neural-network features and use the resulting Gaussian kernel to weight calibration residuals; Amoukou and Brunel (2023) fit a quantile regression forest to the scores and obtain weights from shared tree leaves. The cost and resulting rates of learning such weights are outside this comparison; Section 5 examines a learned coordinate empirically.

What the comparison shows. With $n \asymp m \asymp N$, splitting changes constants, not the estimation exponent; global calibration contributes a faster $N^{-1/2}$ fluctuation at fixed confidence. Learning can therefore offset its data cost by exploiting structure. For additive or single-index maps and $d > 1$, $\beta_q/(2\beta_q+1)$ exceeds the ambient RLCP benchmark $\bar{\beta}_q/(2\bar{\beta}_q+d)$, including when $\beta_q \le 1$. In isotropic Hölder classes, fitting has exponent $\beta_q/(2\beta_q+d)$, up to logarithms: it improves the displayed RLCP exponent for $\beta_q > 1$ and ties it otherwise. Local calibration remains attractive at finite budgets when fitting is inaccurate but neighbourhoods contain many observations and little quantile variation; it avoids both the fitting split and global propagation of fitting errors.

These are class-tuned upper bounds, not an RLCP lower bound or a universal ordering. Polynomial guarantees are uniform, neural guarantees are in $L^2(P_X)$, and RLCP is controlled within the realized centre's neighbourhood, not uniformly over centres. The direct pointwise comparison therefore uses Propositions 7 and 8. Both routes retain marginal validity without the estimation assumptions, with independent fitting and calibration; RLCP additionally uses its prescribed randomization with a fixed kernel and bandwidth.

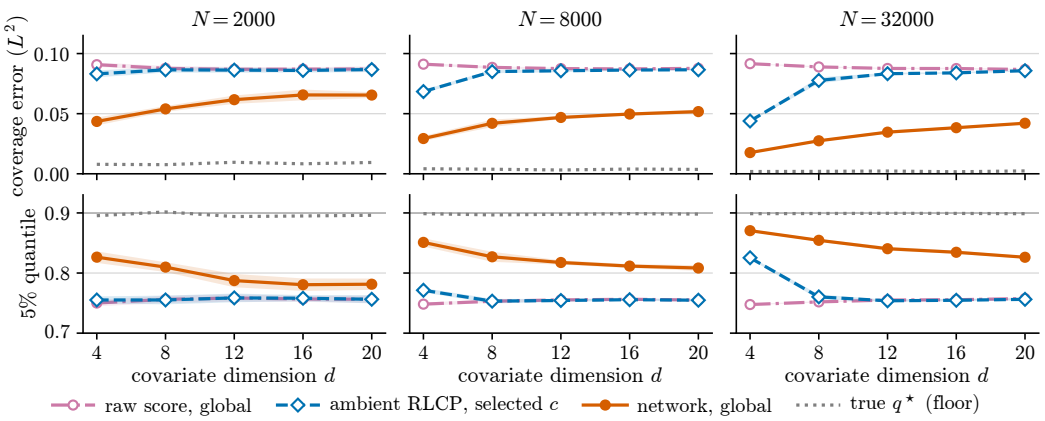

Figure 1: Law (20) against d, one column per budget. Top: root mean square coverage error over 2048 evaluation covariates, one RLCP centre per point; bottom: 5% quantile of conditional coverage over those points (line at 0.9; minima in Figure 3). Means and 95% t-intervals over twenty replications; selected RLCP returns at most 1.1% infinite sets on average, counted as covered.

5 Experiments

We examine the advantage of learning a structured score quantile before global calibration, compared with RLCP in the original covariates. Both routes receive the same score and budget N; $\alpha = 0.1$, with one RLCP centre per test point. Implementation details and low-dimensional comparisons are in Section F.

Synthetic data. For even d, write $x = (x_1, \dots, x_d)$, let $X \sim \mathrm{Unif}([0,1]^d)$ and $\varepsilon \sim N(0,1)$ be independent, and set $v_j = \sqrt{12}\,(x_j - 1/2)$, $J = d/2$, $z_\alpha = \Phi^{-1}(1-\alpha/2)$ and

$$T(x) = \frac{1}{\sqrt{J}} \sum_{j=1}^{J} \frac{2}{3}\Big(v_{2j-1} + v_{2j} + \frac{v_{2j-1}v_{2j}}{2}\Big), \quad q^\star(x) = 1 + \frac{\tanh T(x)}{2}, \quad Y = \frac{q^\star(X)\,\varepsilon}{z_\alpha}. \tag{20}$$

Every coordinate is active, $\mathrm{Var}\,T(X) = 1$ at every d, and the pairing is unknown to the learners. For $S(x,y) = |y|$, $q_S^{1-\alpha}(x) = q^\star(x)$; conditional coverage is computed exactly from the Gaussian CDF.

Procedures. We use $d \in \{4, 8, \dots, 20\}$, $N \in \{2000, 8000, 32000\}$ and twenty replications. The *network* estimates the score's 0.9-quantile on $N/2$ observations, including early stopping, then subtracts it and calibrates globally on the other half. *Ambient RLCP* uses a Euclidean-ball box kernel and selects $h = ch_\star$, $c \in \{1/4, 1/2, 1, 2, 4, 8\}$, on $N/8$ observations before calibration on $7N/8$. Selection minimizes a local-coverage criterion subject to at most 5% infinite validation sets (Section F.2). The references are global calibration of the *raw* score on all N, and subtraction of the true quantile before calibration on $N/2$.

Synthetic results (Figure 1). The network has the lowest root mean square coverage error among the plotted non-oracle procedures in all fifteen cells. At $N = 32000$, its root mean square error rises from 0.018 to 0.042 as d grows, against 0.044 to 0.086 for RLCP, which approaches the raw-score error. The lower tail agrees: the 5% quantile of the network's conditional coverage is 0.78–0.87, against 0.75–0.76 for the raw score and for RLCP beyond $d = 4$; the minimum over the points can still fall below the raw score's (Figure 3). The risk charges RLCP for the variability of its random centre (Section F.1).

Real-data diagnostics (Figure 2). All procedures use $S(x,y) = |y - \widehat{\mu}(x)|$ with the same separately fitted predictor; RLCP searches the wider grid $c = 2^k$, $-2 \le k \le 9$. On bike, facebook, community, MEPS and blog, mean group-coverage errors are 0.034–0.045 for the network, against 0.057–0.092 for RLCP and 0.073–0.145 for raw calibration, with higher minimum group coverage for the network on each. Gains are smaller on bio and concrete, where paired intervals still exclude zero, and absent on star, where raw calibration is already

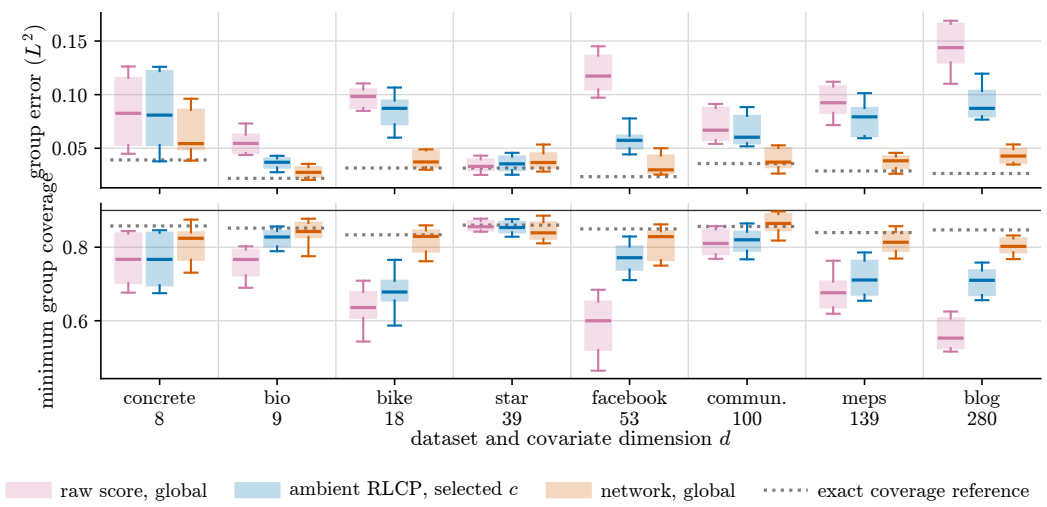

Figure 2: Eight datasets ordered by dimension. Top: root mean square group-coverage error on a common k-means partition; bottom: minimum group coverage. Boxes: quartiles over twenty splits; whiskers: 10th/90th percentiles; dotted: sampling reference under exact coverage. Infinite sets are counted as covered (at most 1.3% on average).

close to the sampling reference. These are group diagnostics, not pointwise guarantees or a controlled effect of dimension; Section F.3 gives the folds and preprocessing.

When does local calibration help? What fails in high dimension is ambient localization: on a one-index law, no bandwidth beats the raw score from $d = 10$ (Section F.4). In the low-dimensional allocation experiment, RLCP wins per centre draw for a rapidly varying quantile map at every tested fitting share, while more fitting labels reverse the comparison for smooth maps (Section F.6).

6 Conclusion

Conditional coverage with a given score S calls for its conditional quantile $q_S^{1-\alpha}(x)$ in place of the marginal one. Either the score is adapted so that a single global threshold becomes conditionally valid, or the calibration is localized so that the threshold varies with x; which route makes the better use of a fixed budget of additional labels, and under what regularity, is the question studied here.

With a common score, oracle and data budget, we identify two distinct errors: global bounds need an averaged estimation term in general, even under population calibration; RLCP bias is bounded through the target quantile alone, without continuity of the full conditional score distribution. Both routes admit high-probability coverage and length guarantees.

Quantile learning can exploit smoothness and composition. In our structured experiment, subtraction and global calibration have lower root mean square coverage error than ambient RLCP in every cell, and group diagnostics agree on most of eight real datasets. Local calibration remains useful when the fit is inaccurate and informative neighbourhoods contain enough data, as in the low-dimensional allocation experiment. These gains concern average error, not coverage at every covariate.

Limitations. Theoretical comparisons use class-tuned estimators and upper bounds in their stated norms and domains. Learned localization is outside the rate comparison. Conditional guarantees require regularity; marginal validity alone is distribution-free.

AI-Use Statement

Generative AI tools were used to assist with language editing, consistency checking, literature search, the conception of proofs, the synthetic data generators, the implementation of the methods, and critical review of the mathematical presentation. The authors retain full responsibility for all claims, proofs, citations, and conclusions.

Reproducibility Statement

Every theoretical result states its assumptions in the main text; complete statements and proofs are in Sections B to E. The synthetic laws have closed-form conditional distribution functions, so conditional coverage is computed exactly rather than estimated. Section F specifies, for every experiment, the data-generating law or dataset and its preprocessing, the estimators and architectures, the tuning grids and selection splits, the budgets, replication counts and evaluation designs, and the uncertainty summaries. The code, configuration files, run records and figure scripts that produce every reported number are included in the anonymized supplementary material.

Ethics Statement

We collected no data for this work, and none of our experiments involve human subjects. The concern behind the paper is itself an ethical one: a prediction set can cover the right fraction of the population and still be too short, systematically, for an identifiable subgroup. That gap is what our bounds and our experiments measure, under the assumptions we state; the group and worst-slab numbers we report are diagnostics, not proof of conditional coverage for any particular subgroup.

References

C. M. Achilles, Helen Pate Bain, Fred Bellott, Jayne Boyd-Zaharias, Jeremy Finn, John Folger, John Johnston, and Elizabeth Word. Tennessee's student teacher achievement ratio (STAR) project. Harvard Dataverse (Dataset), 2008. URL https://doi.org/10.7910/DVN/SIWH9F.

Agency for Healthcare Research and Quality. MEPS HC-181: 2015 full year consolidated data file. Medical Expenditure Panel Survey (Dataset), August 2017. URL https://meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-181. Panel 19 subset.

Salim I. Amoukou and Nicolas J.-B. Brunel. Adaptive conformal prediction by reweighting nonconformity score. *arXiv preprint arXiv:2303.12695*, 2023. URL https://arxiv.org/abs/2303.12695.

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.

Jean-Yves Audibert and Alexandre B. Tsybakov. Fast learning rates for plug-in classifiers. *The Annals of Statistics*, 35(2):608–633, 2007. doi: 10.1214/009053606000001217.

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. Predictive inference with the jackknife+. *The Annals of Statistics*, 49(1):486–507, 2021a. doi: 10.1214/20-AOS1965. URL https://arxiv.org/abs/1905.02928.

Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. The limits of distribution-free conditional predictive inference. *Information and Inference: A Journal of the IMA*, 10(2):455–482, 2021b. doi: 10.1093/imaiai/iaaa017. URL https://arxiv.org/abs/1903.04684.

Alexandre Belloni, Victor Chernozhukov, Denis Chetverikov, and Iván Fernández-Val. Conditional quantile processes based on series or many regressors. *Journal of Econometrics*, 213(1):4–29, 2019. doi: 10.1016/j.jeconom.2019.04.003.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013. ISBN 9780199535255. doi: 10.1093/acprof:oso/9780199535255.001.0001.

Sacha Braun, David Holzmüller, Michael I. Jordan, and Francis Bach. Conditional coverage diagnostics for conformal prediction. In *Forty-third International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2512.11779.

Lawrence D. Brown and Roger Purves. Measurable selections of extrema. *The Annals of Statistics*, 1(5):902–912, 1973. doi: 10.1214/aos/1176342510.

Krisztian Buza. Feedback prediction for blogs. In Myra Spiliopoulou, Lars Schmidt-Thieme, and Ruth Janning, editors, *Data Analysis, Machine Learning and Knowledge Discovery*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 145–152, Cham, 2014. Springer. doi: 10.1007/978-3-319-01595-8_16.

Probal Chaudhuri. Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis*, 39(2):246–269, 1991a. doi: 10.1016/0047-259X(91)90100-G.

Probal Chaudhuri. Nonparametric estimates of regression quantiles and their local Bahadur representation. *The Annals of Statistics*, 19(2):760–777, 1991b. doi: 10.1214/aos/1176348119.

Victor Chernozhukov, Kaspar Wüthrich, and Yinchu Zhu. Distributional conformal prediction. *Proceedings of the National Academy of Sciences*, 118(48):e2107794118, 2021. doi: 10.1073/pnas.2107794118. URL https://arxiv.org/abs/1909.07889.

Nicolò Colombo. Normalizing flows for conformal regression. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, volume 244 of *Proceedings of Machine Learning Research*, pages 881–893. PMLR, 2024. URL https://proceedings.mlr.press/v244/colombo24a.html.

Anton Conrad, Rustam Isaev, Denis Belomestny, Eric Moulines, and Sergey Samsonov. Beyond marginal validity: Finite-sample guarantees for localized conformal prediction. *arXiv preprint arXiv:2608.06206*, 2026. URL https://arxiv.org/abs/2608.06206.

Nicolas Deutschmann, Mattia Rigotti, and María Rodríguez Martínez. Adaptive conformal regression with Split-Jackknife+ scores. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=1fbTGC3BUD.

Nicolas Dewolf, Bernard De Baets, and Willem Waegeman. Conditional validity of heteroskedastic conformal regression. *Information and Inference: A Journal of the IMA*, 14(2):iaaf013, 2025. doi: 10.1093/imaiai/iaaf013. URL https://arxiv.org/abs/2309.08313.

Biprateep Dey, David Zhao, Brett H. Andrews, Jeffrey A. Newman, Rafael Izbicki, and Ann B. Lee. Towards instance-wise calibration: Local amortized diagnostics and reshaping of conditional densities (LADaR). *Machine Learning: Science and Technology*, 6(4):045058, 2025. doi: 10.1088/2632-2153/ae1f05. URL https://arxiv.org/abs/2205.14568v8.

Guneet S. Dhillon, George Deligiannidis, and Tom Rainforth. On the expected size of conformal prediction sets. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 1549–1557. PMLR, 2024. URL https://proceedings.mlr.press/v238/dhillon24a.html.

Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 27(3):642–669, 1956. doi: 10.1214/aoms/1177728174.

Paul Embrechts and Marius Hofert. A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3):423–432, 2013. doi: 10.1007/s00186-013-0436-7.

Hadi Fanaee-T and João Gama. Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence*, 2(2–3):113–127, 2014. doi: 10.1007/s13748-013-0040-3.

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.

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

Emmanuel Guerre and Camille Sabbah. Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory*, 28(1):87–129, 2012. doi: 10.1017/S0266466611000132.

Xing Han, Ziyang Tang, Joydeep Ghosh, and Qiang Liu. Split localized conformal prediction. *arXiv preprint arXiv:2206.13092*, 2022. URL https://arxiv.org/abs/2206.13092.

Xuming He and Peide Shi. Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *Journal of Nonparametric Statistics*, 3(3–4):299–308, 1994. doi: 10.1080/10485259408832589.

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.

Rafael Izbicki, Gilson Shimizu, and Rafael B. Stern. CD-split and HPD-split: Efficient conformal regions in high dimensions. *Journal of Machine Learning Research*, 23(87):1–32, 2022. URL https://jmlr.org/papers/v23/20-797.html.

Christopher Jung, Georgy Noarov, Ramya Ramalingam, and Aaron Roth. Batch multivalid conformal prediction. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Dk7QQp8jHEo.

Danijel Kivaranovic, Kory D. Johnson, and Hannes Leeb. Adaptive, distribution-free prediction intervals for deep networks. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 4346–4356. PMLR, 2020. URL https://proceedings.mlr.press/v108/kivaranovic20a.html.

Shayan Kiyani, George J. Pappas, and Hamed Hassani. Length optimization in conformal prediction. In *Advances in Neural Information Processing Systems*, volume 37. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/b41907dd4df5c60f86216b73fe0c7465-Paper-Conference.pdf.

Roger Koenker, Pin Ng, and Stephen Portnoy. Quantile smoothing splines. *Biometrika*, 81 (4):673–680, 1994. doi: 10.1093/biomet/81.4.673.

Félix Laplante. A post-processing conformal prediction approach for conditional coverage via pivotal scores. *arXiv preprint arXiv:2605.25852*, 2026.

Batiste Le Bars and Pierre Humbert. On volume minimization in conformal regression. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 3030–3053. PMLR, 2025. URL https://proceedings.mlr.press/v267/bars25a.html.

Jing Lei and Larry Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76 (1):71–96, 2014. doi: 10.1111/rssb.12021.

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.

Youjuan Li, Yufeng Liu, and Ji Zhu. Quantile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association*, 102(477):255–268, 2007. doi: 10.1198/016214506000001086.

Zhen Lin, Shubhendu Trivedi, and Jimeng Sun. Locally valid and discriminative prediction intervals for deep learning models. In *Advances in Neural Information Processing Systems*, volume 34, pages 8378–8391. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/46c7cb50b373877fb2f8d5c4517bb969-Abstract.html.

Oscar Hernan Madrid Padilla, Wesley Tansey, and Yanzhen Chen. Quantile regression with ReLU networks: Estimators and minimax rates. *Journal of Machine Learning Research*, 23(247):1–42, 2022. URL http://jmlr.org/papers/v23/21-0309.html.

Pascal Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, 18(3):1269–1283, 1990. doi: 10.1214/aop/1176990746.

Yinjie Min, Liuhua Peng, and Changliang Zou. A unified theory of conditional coverage in conformal prediction with applications. *arXiv preprint arXiv:2605.11602*, 2026.

Harris Papadopoulos, Alex Gammerman, and Vladimir Vovk. Normalized nonconformity measures for regression conformal prediction. In *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications (AIA 2008)*, pages 64–69. ACTA Press, 2008.

Vincent Plassier, Alexander Fishkov, Victor Dheur, Mohsen Guizani, Souhaib Ben Taieb, Maxim Panov, and Eric Moulines. Rectifying conformity scores for better conditional coverage. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 49459–49492. PMLR, 2025a. URL https://proceedings.mlr.press/v267/plassier25a.html.

Vincent Plassier, Alexander Fishkov, Mohsen Guizani, Maxim Panov, and Eric Moulines. Probabilistic conformal prediction with approximate conditional validity. In *International Conference on Learning Representations*, 2025b. URL https://openreview.net/forum?id=Nfd7z9d6Bb.

Prashant Rana. Physicochemical properties of protein tertiary structure. UCI Machine Learning Repository (Dataset), 2013. URL https://archive.ics.uci.edu/dataset/265/physicochemical+properties+of+protein+tertiary+structure.

Michael Redmond. Communities and crime. UCI Machine Learning Repository (Dataset), 2002. URL https://archive.ics.uci.edu/dataset/183/communities+and+crime.

Markus Reiß, Yves Rozenholc, and Charles-André Cuenod. Pointwise adaptive estimation for robust and quantile regression. *arXiv preprint arXiv:0904.0543*, 2009. URL https://arxiv.org/abs/0904.0543.

Yaniv Romano, Evan Patterson, and Emmanuel J. Candès. Conformalized quantile regression. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/hash/5103c3584b063c431bd1268e9b5e76fb-Abstract.html.

Johannes Schmidt-Hieber. Nonparametric regression using deep neural networks with ReLU activation function. *The Annals of Statistics*, 48(4):1875–1897, 2020. doi: 10.1214/19-AOS1875.

Johannes Schmidt-Hieber and Don Vu. Correction to "Nonparametric regression using deep neural networks with ReLU activation function". *The Annals of Statistics*, 52(1):413–414, 2024. doi: 10.1214/24-AOS2351.

Matteo Sesia and Emmanuel J. Candès. A comparison of some conformal quantile regression methods. *Stat*, 9:e261, 2020. doi: 10.1002/sta4.261. URL https://arxiv.org/abs/1909.05433.

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

Guohao Shen, Yuling Jiao, Yuanyuan Lin, Joel L. Horowitz, and Jian Huang. Deep quantile regression: Mitigating the curse of dimensionality through composition. *arXiv preprint arXiv:2107.04907*, 2021.

Guohao Shen, Yuling Jiao, Yuanyuan Lin, Joel L. Horowitz, and Jian Huang. Nonparametric estimation of non-crossing quantile regression process with deep ReQU neural networks. *Journal of Machine Learning Research*, 25(88):1–75, 2024. URL http://jmlr.org/papers/v25/22-0488.html.

Kamaljot Singh. Facebook comment volume. UCI Machine Learning Repository (Dataset), 2016. URL https://archive.ics.uci.edu/dataset/363/facebook+comment+volume+dataset.

Kamaljot Singh, Ranjeet Kaur Sandhu, and Dinesh Kumar. Comment volume prediction using neural networks and decision trees. In *2015 17th UKSim-AMSS International Conference on Modelling and Simulation*, Cambridge, United Kingdom, 2015.

Vladimir Spokoiny, Weining Wang, and Wolfgang Karl Härdle. Local quantile regression. *Journal of Statistical Planning and Inference*, 143(7):1109–1129, 2013. doi: 10.1016/j.jspi.2013.03.008. URL https://arxiv.org/abs/1208.5384.

Ingo Steinwart and Andreas Christmann. Estimating conditional quantiles with the help of the pinball loss. *Bernoulli*, 17(1):211–225, 2011. doi: 10.3150/10-BEJ267. URL https://arxiv.org/abs/1102.2101.

Ichiro Takeuchi, Quoc V. Le, Timothy D. Sears, and Alexander J. Smola. Nonparametric quantile estimation. *Journal of Machine Learning Research*, 7(45):1231–1264, 2006. URL https://jmlr.org/papers/v7/takeuchi06a.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*, volume 32, pages 2526–2536. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/hash/8fb21ee7a2207526da55a679f0332de2-Abstract.html.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, 1996.

Vladimir Vovk. Conditional validity of inductive conformal predictors. In *Proceedings of the Asian Conference on Machine Learning*, volume 25 of *Proceedings of Machine Learning Research*, pages 475–490. PMLR, 2012. URL https://proceedings.mlr.press/v25/vovk12.html.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.

Yunzhen Yao, Lie He, and Michael Gastpar. Non-asymptotic analysis of efficiency in conformalized regression. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2510.07093.

I-Cheng Yeh. Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete Research*, 28(12):1797–1808, 1998. doi: 10.1016/S0008-8846(98)00165-3.

Keming Yu and M. C. Jones. Local linear quantile regression. *Journal of the American Statistical Association*, 93(441):228–237, 1998. doi: 10.1080/01621459.1998.10474104.

Zheng Zhou, Xiangfei Zhang, Chongguang Tao, and Yuhong Yang. Conformal prediction assessment: A framework for conditional coverage evaluation and selection. *arXiv preprint arXiv:2603.27189*, 2026. URL https://arxiv.org/abs/2603.27189.

Contents

1 Introduction **1**

2 Adaptive Scores and Global Calibration **3**

3 Localized Calibration: RLCP with a Fixed Score **5**

4 Comparison with a Common Score **6**

5 Experiments **8**

6 Conclusion **9**

A Extended Related Work **18**

B Complete statement and proof of Theorem 2 **20**

C Complete statements and proofs for Section 3 **23**

C.1 Three deterministic lemmas 24

C.2 The localized calibration event 24

C.3 Complete statement and proof of Theorem 4 26

C.4 Quantile regularity and the RLCP rate 27

C.5 Relation to conditional-CDF regularity 28

D Proofs for the Common-Score Comparison **29**

D.1 Complete statement and proof of Proposition 5 29

D.2 Conditions for the common-score rates 30

D.3 Isotropic Hölder score quantiles 30

D.4 Compositional score quantiles 31

D.5 Complete statement and proof of Proposition 7 33

D.6 Complete statement and proof of Proposition 8 34

D.7 Verifying the density window for a residual score 35

E Quantile-Estimation Tools **35**

E.1 Quantile-estimation assumptions and statements 35

E.2 Proof of Proposition 25 36

F Experimental details **37**

F.1 Evaluation and implementation conventions 37

F.2 Compositional score quantile in growing dimension 38

F.3 Real data read on groups 40

F.4 Dimension on a simple synthetic law 43

F.5 Low dimension with a roughness dial 43
F.6 Budget allocation with a fixed score 44
F.7 Population mechanisms 47

A Extended Related Work

Marginal validity and the meaning of conditioning. Split conformal prediction gives finite-sample marginal coverage under exchangeability (Vovk et al., 2005; Shafer and Vovk, 2008; Lei et al., 2018; Angelopoulos and Bates, 2023). Jackknife+ extends distribution-free guarantees to procedures that reuse observations for fitting, with coverage at least $1 - 2\alpha$ in general (Barber et al., 2021a). Exact distribution-free conditional coverage at continuously distributed covariates is impossible without essentially uninformative prediction sets (Lei and Wasserman, 2014; Barber et al., 2021b). Conditioning on the test covariate must also be distinguished from conditioning on the sample that constructs the predictor (Vovk, 2012). A high-probability guarantee for the marginal coverage of a realized predictor does not by itself give pointwise conditional coverage. Our conditional-error bounds require score and design regularity.

Established constructions of adaptive scores. Normalized nonconformity measures adjust prediction widths to the difficulty of individual examples (Papadopoulos et al., 2008). Locally scaled residuals (Lei et al., 2018) give a covariate-independent oracle score law under a location–scale model with covariate-independent standardized noise. Such assumptions, rather than heteroskedasticity alone, explain when normalization yields conditional validity (Dewolf et al., 2025). CQR constructs scores from fitted conditional quantiles (Romano et al., 2019); related deep-network constructions and comparisons are given by Kivaranovic et al. (2020); Sesia and Candès (2020). Distributional conformal prediction uses conditional ranks, whose oracle distribution is uniform under continuity (Chernozhukov et al., 2021). CD-split and HPD-split target highest-density regions and establish local or asymptotic conditional validity under their assumptions (Izbicki et al., 2022). The conditional-density value itself need not have a covariate-independent distribution; CD-split uses a partition and HPD-split a rank transformation to obtain their respective guarantees.

Transforming an existing score. A direct predecessor of our additive example is split localized conformal prediction (Han et al., 2022). It estimates a conditional score quantile by kernel smoothing on a fitting sample, subtracts it from the original score, and applies global split calibration. Its analysis establishes marginal validity and asymptotic conditional coverage, and extends the construction to monotone transformations. Thus, despite its name, this method fits our transformed-score route rather than RLCP on the unchanged score. Colombo (2024) learns input-conditional normalizing flows to transform regression errors towards a common distribution without retraining the point predictor. Plassier et al. (2025a) state single-level quantile alignment as the criterion for conditional coverage and develop a rectification family that includes quantile subtraction and positive rescaling. Their coverage analysis involves conditional-CDF errors and marginal transformed-score calibration terms; their supplement also gives two-sided bounds. Deutschmann et al. (2024) study local score rescaling and a split/jackknife+ construction that reuses calibration observations, relating conditional coverage to score–covariate dependence. Related conditional-distribution post-processing includes Cal-PIT (Dey et al., 2025), which learns a local probability–probability map to diagnose and reshape a predictive distribution. CP^2 (Plassier et al., 2025b) instead calibrates candidate sets constructed from an estimated response law; its conditional under-coverage bound combines total-variation error in that law with a calibration correction. Our analysis retains separate fitting and calibration folds and uses quantile subtraction as an established, particularly transparent example.

Full pivotality, quantile alignment, and exact coverage. A score law independent of the covariate aligns every conditional quantile. This principle underlies distributional scores (Chernozhukov et al., 2021) and the pivotal-score approach of Laplante (2026). Our definition asks only that the conditional quantile be constant at the chosen level. It then gives a common oracle threshold with conditional coverage at least $1 - \alpha$; exact equality additionally requires continuity at that quantile. For general CDFs, a common threshold at which each CDF equals $1 - \alpha$ need not align their left quantiles because of plateaus, while aligned quantiles can overshoot the level at atoms. Under our continuity and local growth assumptions these difficulties disappear. Single-level adaptivity is therefore weaker than distributional pivotality.

Localized and weighted calibration. Weighted conformal prediction gives marginal validity under a specified covariate shift when the covariate density ratio is known (Tibshirani et al., 2019). Arbitrary test-centred kernel weights do not automatically inherit this guarantee. Localized conformal prediction (Guan, 2023) uses nearby scores and a data-dependent calibration-level correction. RLCP (Hore and Barber, 2025) draws an auxiliary centre and weights calibration and test observations relative to that centre, retaining finite-sample marginal validity and providing relaxed local and covariate-shift guarantees. The partial-information framework of Barber and Tibshirani (2026) connects these constructions to broader weighted-conformal arguments. The marginal RLCP guarantee averages over its random centre; our fixed-score conditional-error event instead holds over the realized centre's bandwidth neighbourhood. This distinction matters when interpreting a local guarantee, especially near boundaries or regions of low design density.

Groupwise, multivalid, and functional conditional guarantees. Mondrian-type methods calibrate within a partition and guarantee coverage conditional on category membership (Vovk et al., 2005; Vovk, 2012). Batch multivalid conformal prediction learns thresholds with approximate coverage conditional on membership in specified overlapping groups and on the selected threshold (Jung et al., 2023). Its PAC-style guarantees hold with high probability over calibration under score-distribution regularity. Conditional calibration (Gibbs et al., 2025) uses functional quantile regression to target coverage moments indexed by covariate functions. Its randomized version gives exact finite-sample guarantees for finite-dimensional classes, including overlapping groups, and the analysis controls errors for richer classes. These targets differ from accuracy at every covariate and from our comparison with a fixed score-induced oracle.

Unified finite-sample analyses. Min et al. (2026) decompose conditional miscoverage into score-estimation, calibration, and intrinsic conditional-mismatch terms, give method-specific rates, and propose a conditional-coverage-oriented selection rule. Their fixed-score localized specialization gives a Lipschitz bias–variance rate; their learned-score specialization (CQR) charges training through a tail bound for a uniform-in-x quantile-estimation error and averages over fitting randomness. The coverage probabilities in these comparisons average over calibration and, for randomized localization, over its auxiliary randomness. Their supplement also implements conditional score-quantile subtraction followed by global calibration. Our comparison concerns realized coverage and oracle-relative length under the same additional label budget, with the global and local error terms kept explicit.

Recent bounds for post-processed scores. Laplante (2026) estimates the conditional CDF of a fixed base score and applies a probability integral transform. Its coverage bounds have a pointwise-plus-averaged structure, expressed through conditional-CDF estimation error in Kolmogorov–Smirnov distance. Its Lemma 3 gives simultaneous-in-x, high-probability oracle-set inclusions and associated volume bounds; its Theorem 3 controls probability-weighted symmetric difference by an averaged CDF error plus a finite-sample calibration remainder. Our bounds use approximation to a single-level adaptive score, control absolute Lebesgue-length error under explicit sublevel-set regularity, and place the result alongside local calibration of the same initial score. Proposition 3 gives a matching-order lower bound: score error confined to one region causes coverage error elsewhere, even under population calibration.

Recent bounds for localized scores. Conrad et al. (2026) prove finite-sample, high-probability RLCP coverage and oracle-relative length bounds over the realized localization neighbourhood, for fixed scores and for data-split learned scores approaching a pivotal target. Their fixed-score bias uses Hölder continuity of the conditional score CDF in the covariate, uniformly over thresholds. For fixed scores, we control bias through variation of the conditional quantile alone. Under local growth, the CDF condition implies the corresponding quantile condition on sufficiently small neighbourhoods (Remark 19); Remark 20 shows that the quantile condition allows discontinuous conditional score CDFs. The contribution is this refinement and its use in the common-budget comparison.

Efficiency and the choice of oracle. Dhillon et al. (2024) characterize expected split-conformal set size and provide empirical estimates and confidence bounds. Kiyani et al. (2024) optimize length subject to coverage constraints over specified covariate-shift classes, including group coverage. Le Bars and Humbert (2025) derive finite-sample excess-volume bounds and learn predictors to reduce interval length. Yao et al. (2026) bound expected oracle-relative length deviation for conformalized quantile and median regression, tracking training size, calibration size and miscoverage level; their CQR theory assumes well-specified linear conditional quantiles trained by stochastic gradient descent. Our length error measures deviation from the conditional oracle of the fixed initial score, not excess length relative to the shortest valid region. For a residual score with a supplied centre, this oracle need not be equal-tailed or minimum-length. This is why Section 4 fixes the same score and oracle for both routes; a CQR-versus-residual comparison would generally involve different oracle targets.

Conditional quantile estimation at arbitrary smoothness. Higher-order conditional-quantile estimation is classical: local polynomials (Chaudhuri, 1991b;a; Yu and Jones, 1998; Guerre and Sabbah, 2012), splines and series (Koenker et al., 1994; He and Shi, 1994; Belloni et al., 2019), and kernel/RKHS methods (Takeuchi et al., 2006; Li et al., 2007; Steinwart and Christmann, 2011) provide several alternatives. Local-polynomial fits exploit derivatives through their polynomial basis, so the first-order variation in our RLCP bound is specific to local-constant calibration. Deep ReLU quantile estimators attain rates over Hölder, Besov and hierarchical-composition classes (Madrid Padilla et al., 2022); Shen et al. (2021) also study composition structure. Shen et al. (2024) estimate the full non-crossing quantile process with ReQU networks, a different estimation problem from a fixed quantile level. Section E records the needed local-polynomial uniform and deep-network $L^2(P_X)$ instantiations, drawing on Guerre and Sabbah (2012); Madrid Padilla et al. (2022); Schmidt-Hieber (2020).

Adaptation to unknown smoothness. Exploiting a specified Hölder class is distinct from adapting to unknown regularity. Reiß et al. (2009) study Lepski-type sequential selection for robust and quantile regression, with pointwise adaptation and nonasymptotic risk bounds. Spokoiny et al. (2013) select local quantile-regression bandwidths through likelihood-ratio tests and obtain finite-sample oracle risk guarantees. Proposition 7 requires a uniform high-probability statement that these results do not provide. Our benchmark polynomial degrees and bandwidths depend on smoothness; replacing them by adaptive choices requires checking the learner's assumptions, norm and probability event.

Assessing conditional coverage empirically. Marginal coverage and average length alone cannot assess the conditional errors studied here. Braun et al. (2026) formulate diagnostics as classification: at the population level, suitable excess risks lower-bound integrated conditional-miscoverage measures. Zhou et al. (2026) estimate instance-level coverage probabilities and use them for conditional-validity assessment and method selection. These approaches complement synthetic experiments with known conditional laws, where pointwise errors and the common oracle can be evaluated directly.

Scope of the comparison. We compare established constructions at a common initial score, conditional oracle and additional data budget. The global route fits a transformation and then calibrates; RLCP uses the additional labels for local calibration of the unchanged score. The explicit quantile-subtraction example links both bounds to the same conditional-quantile smoothness. When the exponents coincide, the allocation of labels, design mass, error propagation and problem-dependent constants remain central.

B Complete statement and proof of Theorem 2

Conditioning on $\mathcal{D}_n^{\mathrm{tr}}$, we first bound the marginal CDF error by $\sigma_{\mathrm{up}}\|\Delta_S\|_{L^1(P_X)}$ using pointwise score comparisons. The DKW inequality adds the calibration fluctuation, and local growth at the pivot converts the CDF error into a threshold bound. Pointwise set inclusions then give coverage and length bounds.

Theorem 2 is the simplified form of the following statement, which also controls the calibrated threshold and makes the constants explicit. Set

$$\rho_m(\delta) \coloneqq \frac{\sigma_{\mathrm{up}}\|\Delta_S\|_{L^1(P_X)} + \eta_m(\delta) + s_m}{\sigma_{\mathrm{low}}}. \tag{21}$$

Theorem 2 (complete version). *Let* $\boldsymbol{A}\,1(S, 1-\alpha)$ *and* $\boldsymbol{A}\,3(S, 1-\alpha)$ *hold, and condition throughout on a realization of* $\mathcal{D}_n^{\mathrm{tr}}$. *Fix* $\delta \in (0,1)$ *and suppose that*

$$\sigma_{\mathrm{up}}\|\Delta_S\|_{L^1(P_X)} + \eta_m(\delta) \le \sigma_{\mathrm{low}} r_-, \qquad \sigma_{\mathrm{up}}\|\Delta_S\|_{L^1(P_X)} + \eta_m(\delta) + s_m \le \sigma_{\mathrm{low}} r_+. \tag{22}$$

Then there is an event $\mathcal{E}_m(\delta)$, *measurable with respect to* $\mathcal{D}_m^{\mathrm{cal}}$, *with* $\mathbb{P}(\mathcal{E}_m(\delta) \mid \mathcal{D}_n^{\mathrm{tr}}) \ge 1-\delta$, *on which the following hold.*

(i) (Threshold) $\widehat{Q}_m$ *is finite and* $-\dfrac{\sigma_{\mathrm{up}}\|\Delta_S\|_{L^1(P_X)} + \eta_m(\delta)}{\sigma_{\mathrm{low}}} \le \widehat{Q}_m - q^\star \le \dfrac{\sigma_{\mathrm{up}}\|\Delta_S\|_{L^1(P_X)} + \eta_m(\delta) + s_m}{\sigma_{\mathrm{low}}}$; *in particular* $|\widehat{Q}_m - q^\star| \le \rho_m(\delta)$.

(ii) (Coverage) For every $x \in \mathcal{X}_0$, $\big|\mathrm{Cov}_{\hat{\mathcal{C}}_\alpha}(x) - (1-\alpha)\big| \le \sigma_{\mathrm{up}}\Delta_S(x) + \sigma_{\mathrm{up}}\rho_m(\delta)$.

(iii) (Length) If moreover $\boldsymbol{A}\,2(S)$ *holds, then for every* $x \in \mathcal{X}_0$, $\big|\,|\hat{\mathcal{C}}_\alpha(x)| - |\mathcal{C}_S(x, q^\star)|\,\big| \le L_{\mathrm{len}}\big(\Delta_S(x) + \rho_m(\delta)\big)$.

Proof of Theorem 2, simplified form, from the complete version. Condition (5) implies both inequalities in (22). By (4), $\eta_m(\delta) \le \sqrt{\log(2/\delta)/m}$ and $s_m < 2/m$; hence $\rho_m(\delta) \le (\sigma_{\mathrm{up}}/\sigma_{\mathrm{low}})\|\Delta_S\|_{L^1(P_X)} + \sigma_{\mathrm{low}}^{-1}\sqrt{\log(2/\delta)/m} + 2/(\sigma_{\mathrm{low}} m)$. Inserting this bound into (ii) and (iii), and using $\sigma_{\mathrm{low}} \le \sigma_{\mathrm{up}}$ to write $\Delta_S(x) \le (\sigma_{\mathrm{up}}/\sigma_{\mathrm{low}})\Delta_S(x)$ in the leading terms, gives the coverage part of (6) with constant $2\sigma_{\mathrm{up}}/\sigma_{\mathrm{low}}$ and its length part with constant $2L_{\mathrm{len}}/\sigma_{\mathrm{low}}$. □

The following lemmas establish the score comparisons and identify the pivot needed for the complete proof.

Lemma 9. *For every* x *with* $\Delta_S(x) < \infty$ *and every* $t \in \mathbb{R}$,

$$\mathcal{C}_S(x, t - \Delta_S(x)) \subset \mathcal{C}_{\widehat{S}}(x, t) \subset \mathcal{C}_S(x, t + \Delta_S(x)),$$
$$\textit{hence} \qquad F^{(S|X)}(t - \Delta_S(x) \mid x) \le F^{(\widehat{S}|X)}(t \mid x) \le F^{(S|X)}(t + \Delta_S(x) \mid x).$$

Proof. By definition of Δ_S, $S(x,y) - \Delta_S(x) \le \widehat{S}(x,y) \le S(x,y) + \Delta_S(x)$ for all y; the inclusions follow, and the CDF inequalities are their $P_{Y|X=x}$-masses. □

Lemma 10. *Under* $\boldsymbol{A}\,1(S, 1-\alpha)(a)$, $\sup_{t\in\mathbb{R}} |F^{(\widehat{S})}(t) - F^{(S)}(t)| \le \sigma_{\mathrm{up}}\|\Delta_S\|_{L^1(P_X)}$.

Proof. If $\|\Delta_S\|_{L^1(P_X)} = \infty$ there is nothing to prove; otherwise $\Delta_S < \infty$ a.e. For $x \in \mathcal{X}_0$ with $\Delta_S(x) < \infty$, Lemma 9 and $\mathbf{A}\,1(S, 1-\alpha)$(a) give $F^{(\widehat{S}|X)}(t \mid x) \le F^{(S|X)}(t + \Delta_S(x) \mid x) \le F^{(S|X)}(t \mid x) + \sigma_{\mathrm{up}}\Delta_S(x)$ and, symmetrically, $F^{(\widehat{S}|X)}(t \mid x) \ge F^{(S|X)}(t \mid x) - \sigma_{\mathrm{up}}\Delta_S(x)$. Integrating against P_X over the P_X-full set $\mathcal{X}_0 \cap \{\Delta_S < \infty\}$ gives $|F^{(\widehat{S})}(t) - F^{(S)}(t)| \le \sigma_{\mathrm{up}}\|\Delta_S\|_{L^1(P_X)}$ for every t. □

Lemma 11. *Under* $\boldsymbol{A}\,1(S, 1-\alpha)$ *and* $\boldsymbol{A}\,3(S, 1-\alpha)$, $q_S(x) = q^\star$ *for every* $x \in \mathcal{X}_0$ *and* $\bar{q}_S = q^\star$; *in particular* S *is* $(1-\alpha)$*-adaptive in the sense of Definition 1.*

Proof. Fix $x \in \mathcal{X}_0$. $F^{(S|X)}(q^\star \mid x) = 1-\alpha$ gives $q_S(x) \le q^\star$; if $q_S(x) < q^\star$, $\mathbf{A}\,1(S, 1-\alpha)$(b) with $v = (q^\star - q_S(x)) \wedge r_+ > 0$, the right-continuity of $F^{(S|X)}(\cdot \mid x)$ at $q_S(x)$, and monotonicity give $F^{(S|X)}(q^\star \mid x) \ge F^{(S|X)}(q_S(x)+v \mid x) \ge F^{(S|X)}(q_S(x) \mid x) + \sigma_{\mathrm{low}} v \ge 1-\alpha+\sigma_{\mathrm{low}} v > 1-\alpha$, a contradiction. Hence $q_S(x) = q^\star$ on $\mathcal{X}_0$, and $\mathbf{A}\,1(S, 1-\alpha)$(b) reads $F^{(S|X)}(q^\star - u \mid x) \le 1 - \alpha - \sigma_{\mathrm{low}} u$ for $0 \le u \le r_-$ on $\mathcal{X}_0$. Integrating this and $\mathbf{A}\,3(S, 1-\alpha)$ over $\mathcal{X}_0$ gives $F^{(S)}(q^\star) = 1-\alpha$ and $F^{(S)}(q^\star - u) \le 1 - \alpha - \sigma_{\mathrm{low}} u$ for $0 \le u \le r_-$; with monotonicity for $t \le q^\star - r_-$, $F^{(S)}(t) < 1-\alpha$ for every $t < q^\star$, so $\bar{q}_S = q^\star$. □

Proof of Theorem 2, complete version. Event. Conditionally on $\mathcal{D}_n^{\mathrm{tr}}$, the calibration scores $\widehat{S}_1, \dots, \widehat{S}_m$ are i.i.d. with CDF $F^{(\widehat{S})}$. Let $\mathcal{E}_m(\delta) \coloneqq \{\sup_t |\widehat{F}_m^{(\widehat{S})}(t) - F^{(\widehat{S})}(t)| \le \eta_m(\delta)\}$. By the Dvoretzky–Kiefer–Wolfowitz inequality with Massart's constant (Dvoretzky et al., 1956; Massart, 1990), $\mathbb{P}(\mathcal{E}_m(\delta)^c \mid \mathcal{D}_n^{\mathrm{tr}}) \le 2\exp(-2m\,\eta_m(\delta)^2) = \delta$. *(i) Threshold.* On $\mathcal{E}_m(\delta)$, Lemma 10 gives $\sup_t |\widehat{F}_m^{(\widehat{S})}(t) - F^{(S)}(t)| \le \zeta \coloneqq \sigma_{\mathrm{up}}\|\Delta_S\|_{L^1(P_X)} + \eta_m(\delta)$. Integrating $\mathbf{A}\,3(S, 1-\alpha)$ and $\mathbf{A}\,1(S, 1-\alpha)$(b), with $q_S(x) \equiv q^\star$ by Lemma 11, gives $F^{(S)}(q^\star) = 1-\alpha$ and the same lower growth bounds for the marginal CDF around $q^\star$. Set $t_- = q^\star - \zeta/\sigma_{\mathrm{low}}$ and $t_+ = q^\star + (\zeta + s_m)/\sigma_{\mathrm{low}}$, which lie in the respective growth windows by (22). Since $s_m > 0$ and $\beta_m = 1 - \alpha + s_m$,

$$\widehat{F}_m^{(\widehat{S})}(t_-) \le F^{(S)}(t_-) + \zeta \le 1 - \alpha < \beta_m,$$
$$\widehat{F}_m^{(\widehat{S})}(t_+) \ge F^{(S)}(t_+) - \zeta \ge 1 - \alpha + s_m = \beta_m.$$

The defining inequalities for the generalized inverse (Embrechts and Hofert, 2013, Proposition 2.3(5)) therefore give $t_- \le \widehat{Q}_m \le t_+$. In particular, $\widehat{Q}_m$ is finite, proving (i).

(ii) Coverage. Fix $x \in \mathcal{X}_0$; if $\Delta_S(x) = \infty$ the bound is trivial, so assume $\Delta_S(x) < \infty$. By Lemma 9 at $t = \widehat{Q}_m$, $F^{(S|X)}(\widehat{Q}_m - \Delta_S(x) \mid x) \le \mathrm{Cov}_{\hat{\mathcal{C}}_\alpha}(x) \le F^{(S|X)}(\widehat{Q}_m + \Delta_S(x) \mid x)$. By $\mathbf{A}\,3(S, 1-\alpha)$ and $\mathbf{A}\,1(S, 1-\alpha)$(a), $|F^{(S|X)}(t \mid x) - (1-\alpha)| \le \sigma_{\mathrm{up}}|t - q^\star|$ for every t, and $|\widehat{Q}_m \pm \Delta_S(x) - q^\star| \le |\widehat{Q}_m - q^\star| + \Delta_S(x) \le \rho_m(\delta) + \Delta_S(x)$ by (i). Hence $|\,\mathrm{Cov}_{\hat{\mathcal{C}}_\alpha}(x) - (1-\alpha)| \le \sigma_{\mathrm{up}}(\Delta_S(x) + \rho_m(\delta))$.

(iii) Length. With the same x — the case $\Delta_S(x) = \infty$ again being trivial — the inclusions of Lemma 9 at $t = \widehat{Q}_m$ give $|\mathcal{C}_S(x, \widehat{Q}_m - \Delta_S(x))| \le |\hat{\mathcal{C}}_\alpha(x)| \le |\mathcal{C}_S(x, \widehat{Q}_m + \Delta_S(x))|$, and by $\mathbf{A}\,2(S)$ both outer terms are within $L_{\mathrm{len}}(|\widehat{Q}_m - q^\star| + \Delta_S(x)) \le L_{\mathrm{len}}(\rho_m(\delta) + \Delta_S(x))$ of $|\mathcal{C}_S(x, q^\star)|$. □

Remark 12. *For $c > 0$, replacing $(\widehat{S}, S)$ by $(c\widehat{S}, cS)$ multiplies Δ_S, $q^\star$, $\widehat{Q}_m$, $\rho_m(\delta)$ and the growth radii by c, while dividing σ_{up}, σ_{low} and L_{len} by c. The prediction sets, admissibility conditions and coverage and length bounds are unchanged.*

Proposition 3 (complete version). *The term $\|\Delta_S\|_{L^1(P_X)}$ in Theorem 2 cannot be improved to the local error $\Delta_S(x)$, even when the population score distribution is known. Let $\mathbf{A}\,1(S, 1-\alpha)$ and $\mathbf{A}\,3(S, 1-\alpha)$ hold, fix a measurable $A \subset \mathcal{X}_0$ with $a \coloneqq P_X(A) \in (0,1)$ and $c \in (0, r_-]$ with $(\sigma_{\mathrm{up}}/\sigma_{\mathrm{low}})ac \le r_+$, and let $\widehat{S}(x,y) \coloneqq S(x,y) + c\mathbb{1}_{\{x \in A\}}$, so that $\Delta_S = c\mathbb{1}_{\{A\}}$ and $\|\Delta_S\|_{L^1(P_X)} = ac$. Then the population global threshold $\bar{q}_{\widehat{S}} = \mathrm{Q}(1-\alpha; F^{(\widehat{S})})$ satisfies*

$$(\sigma_{\mathrm{low}}/\sigma_{\mathrm{up}})\|\Delta_S\|_{L^1(P_X)} \le \bar{q}_{\widehat{S}} - q^\star \le (\sigma_{\mathrm{up}}/\sigma_{\mathrm{low}})\|\Delta_S\|_{L^1(P_X)},$$

and for every $x \in \mathcal{X}_0 \setminus A$ — where $\Delta_S(x) = 0$ — the conditional coverage of $\mathcal{C}_{\widehat{S}}(x, \bar{q}_{\widehat{S}})$ exceeds $1-\alpha$ by at least $(\sigma_{\mathrm{low}}^2/\sigma_{\mathrm{up}})\|\Delta_S\|_{L^1(P_X)}$.

Proof of Proposition 3. Recall $q_S(x) \equiv q^\star$ on $\mathcal{X}_0$ with $F^{(S|X)}(q^\star \mid x) = 1-\alpha$ (Lemma 11). Since $\widehat{S} = S$ off A and $\widehat{S} = S + c$ on A,

$$F^{(\widehat{S})}(t) = \int_{A^c} F^{(S|X)}(t \mid u)\, P_X(\mathrm{d}u) + \int_A F^{(S|X)}(t - c \mid u)\, P_X(\mathrm{d}u).$$

Since $\widehat{S}$ is real-valued, this is a proper CDF; as $1-\alpha \in (0,1)$, its quantile $\bar{q}_{\widehat{S}}$ is finite. The CDF is σ_{up}-Lipschitz, hence continuous, so $F^{(\widehat{S})}(\bar{q}_{\widehat{S}}) = 1-\alpha$. Since $c \le r_-$, lower growth gives $F^{(\widehat{S})}(q^\star) \le 1 - \alpha - a\sigma_{\mathrm{low}}c$. Consequently $\bar{q}_{\widehat{S}} > q^\star$ and

$$a\sigma_{\mathrm{low}}c \le F^{(\widehat{S})}(\bar{q}_{\widehat{S}}) - F^{(\widehat{S})}(q^\star) \le \sigma_{\mathrm{up}}(\bar{q}_{\widehat{S}} - q^\star),$$

which proves the lower threshold bound. For $0 \le v \le r_+$, Lipschitz continuity and lower growth also give $F^{(\widehat{S})}(q^\star + v) \ge F^{(S)}(q^\star + v) - a\sigma_{\mathrm{up}}c \ge 1 - \alpha + \sigma_{\mathrm{low}}v - a\sigma_{\mathrm{up}}c$. Taking $v = (\sigma_{\mathrm{up}}/\sigma_{\mathrm{low}})ac \le r_+$ proves the upper threshold bound. Finally, for $x \in \mathcal{X}_0 \setminus A$, $\widehat{S}(x, \cdot) = S(x, \cdot)$ and lower growth yields

$$F^{(\widehat{S}|X)}(\bar{q}_{\widehat{S}} \mid x) - (1-\alpha) \ge \sigma_{\mathrm{low}}(\bar{q}_{\widehat{S}} - q^\star) \ge (\sigma_{\mathrm{low}}^2/\sigma_{\mathrm{up}})ac.$$

□

C Complete statements and proofs for Section 3

For a score S and centre c with $Z_{c,h} > 0$, recall the localized score CDF

$$F^{(S)}_{c,h}(t) \coloneqq \int_{\mathcal{X}} F^{(S|X)}(t \mid u)\, \Lambda_{c,h}(\mathrm{d}u), \qquad t \in \mathbb{R}. \tag{23}$$

Throughout, $\tilde{x} \in \mathbb{R}^d$ is a fixed centre with $Z_{\tilde{x},h} > 0$, $h > 0$ is fixed, and $m_{\tilde{x},h} = mZ_{\tilde{x},h}/\|\mathcal{K}\|_\infty$ is the effective local calibration size of Section 3; recall that the weights $w_{\tilde{x},h}(X_i)$ take values in $[0, \|\mathcal{K}\|_\infty]$. For a fixed measurable score S with calibration scores $S_i = S(X_i, Y_i)$, set

$$B_m \coloneqq \frac{1}{m}\sum_{i=1}^{m} w_{\tilde{x},h}(X_i), \qquad A_m(t) \coloneqq \frac{1}{m}\sum_{i=1}^{m} w_{\tilde{x},h}(X_i)\mathbb{1}_{\{S_i \le t\}}, \tag{24}$$

and $A(t) \coloneqq \mathbb{E}[w_{\tilde{x},h}(X)\mathbb{1}_{\{S(X,Y)\le t\}}]$, so that $A(t) = Z_{\tilde{x},h}\, F^{(S)}_{\tilde{x},h}(t)$ by conditioning on X and (23); when $B_m > 0$, $\overline{F}^{(S)}_{m,\tilde{x},h}(t) = A_m(t)/B_m$ is the kernel-weighted empirical CDF of the calibration scores. For $x \in \mathbb{R}^d$ write $\rho_x \coloneqq \widetilde{w}_{m+1}(x) = w_{\tilde{x},h}(x)/(mB_m + w_{\tilde{x},h}(x))$ for the test weight, with the convention $\rho_x \coloneqq 0$ when $D(x) = 0$. Two quantities describe the localized calibration step, in the roles played by $\eta_m(\delta)$ and s_m in (4): the Kolmogorov radius of $\overline{F}^{(S)}_{m,\tilde{x},h}$ around $F^{(S)}_{\tilde{x},h}(\cdot)$ and the level shift caused by the test mass,

$$\eta_{m,h}(\delta) \coloneqq 23\sqrt{\frac{\log(4/\delta)}{m_{\tilde{x},h}}}, \qquad s_{m,h} \coloneqq \frac{4}{m_{\tilde{x},h}}, \tag{25}$$

both controlled on a single event (Proposition 17) as soon as

$$m_{\tilde{x},h} \ge 64\log(4/\delta) \vee 4/\alpha. \tag{26}$$

Remark 13. *The constant c_{reg} in **P** 1 is a lower bound on the fraction of the volume of $B(\tilde{x}, \eta h)$ lying in $\mathcal{X}$. It describes the geometry of the covariate support, not the score or coverage level. This is a local form of the regular-support condition used in nonparametric estimation; see Audibert and Tsybakov (2007, equation (2.1)). One may take $c_{\mathrm{reg}} = 1$ when this ball lies inside $\mathcal{X}$. For $\mathcal{X} = [0,1]^d$, $\tilde{x} \in \mathcal{X}$ and $\eta h \le 1/2$, $c_{\mathrm{reg}} = 2^{-d}$ works: at least one orthant of the ball lies in the cube, with equality at a corner.*

The constant enters the effective sample size through a direct integration bound, with $v_d \coloneqq |B(0,1)|$:

$$\begin{aligned} Z_{\tilde{x},h} &= \int_{\mathcal{X}} \mathcal{K}((u-\tilde{x})/h) p_X(u)\, \mathrm{d}u \\ &\ge \kappa p_{\min} |\mathcal{X} \cap B(\tilde{x}, \eta h)| \ge \kappa p_{\min} c_{\mathrm{reg}} v_d (\eta h)^d. \end{aligned}$$

The first inequality uses the kernel lower bound on the smaller ball and the density lower bound; the second uses its volume fraction. Multiplying by $m/\|\mathcal{K}\|_\infty$ gives

$$m_{\tilde{x},h} \ge c_K\, m p_{\min} h^d, \qquad c_K \coloneqq \kappa c_{\mathrm{reg}} v_d \eta^d / \|\mathcal{K}\|_\infty. \tag{27}$$

*For a bound uniform over centres or bandwidths, the same positive c_{reg} must work throughout. An RLCP centre can lie outside $\mathcal{X}$, so the cube example does not apply automatically; $Z_{\tilde{x},h} > 0$ alone does not guarantee this overlap condition. The complete bounds below use $m_{\tilde{x},h}$ directly and do not require **P** 1.*

The proof of Theorem 4 proceeds as follows: on the event $\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)$, Bernstein and Bousquet bounds give $\sup_t |\overline{F}^{(S)}_{m,\tilde{x},h}(t) - F^{(S)}_{\tilde{x},h}(t)| \le \eta_{m,h}(\delta)$ and a test weight $\widetilde{w}_{m+1}(x) \lesssim m^{-1}_{\tilde{x},h}$ uniformly in x; for a fixed score, $\mathbf{A}\,1(S, 1-\alpha)$(b) integrated against $\Lambda_{\tilde{x},h}$ shows that $F^{(S)}_{\tilde{x},h}(\cdot)$ rises at rate σ_{low} outside $[q_S(x) - \omega_S(\tilde{x}, h), q_S(x) + \omega_S(\tilde{x}, h)]$, and an inversion lemma places $\widehat{Q}^{\mathrm{R}}(x; \tilde{x})$ within $(\eta_{m,h}(\delta) + s_{m,h})/\sigma_{\mathrm{low}}$ of that interval; $\mathbf{A}\,1(S, 1-\alpha)$(a) and $\mathbf{A}\,2(S)$ then convert the threshold error into the coverage and length bounds.

C.1 Three deterministic lemmas

Lemma 14. *Fix $x \in \mathbb{R}^d$ and suppose $B_m > 0$. Then $\widehat{F}^{\mathrm{R}}_{x,\tilde{x}}(t) = (1-\rho_x)\overline{F}^{(S)}_{m,\tilde{x},h}(t)$ for every $t \in \mathbb{R}$. If moreover $\rho_x < \alpha$, then $p_x := (1-\alpha)/(1-\rho_x) \in [1-\alpha, 1)$, the threshold $\widehat{Q}^{\mathrm{R}}(x;\tilde{x})$ is finite, and*

$$\widehat{Q}^{\mathrm{R}}(x;\tilde{x}) = \mathrm{Q}(p_x; \overline{F}^{(S)}_{m,\tilde{x},h}) \in \{S_i : w_{\tilde{x},h}(X_i) > 0\}, \qquad 0 \le p_x - (1-\alpha) = \frac{(1-\alpha)\rho_x}{1-\rho_x}. \tag{28}$$

Proof. By (8), $\widehat{F}^{\mathrm{R}}_{x,\tilde{x}}(t) = \sum_i w_{\tilde{x},h}(X_i)\mathbb{1}_{\{S_i \le t\}}/D(x) = (mB_m/D(x))\overline{F}^{(S)}_{m,\tilde{x},h}(t) = (1-\rho_x)\overline{F}^{(S)}_{m,\tilde{x},h}(t)$. Hence $\{t : \widehat{F}^{\mathrm{R}}_{x,\tilde{x}}(t) \ge 1-\alpha\} = \{t : \overline{F}^{(S)}_{m,\tilde{x},h}(t) \ge p_x\}$ and the two left quantiles coincide. Since $\overline{F}^{(S)}_{m,\tilde{x},h}$ is a step CDF supported on $\{S_i : w_{\tilde{x},h}(X_i) > 0\}$ with $\overline{F}^{(S)}_{m,\tilde{x},h}(\max\{S_i : w_{\tilde{x},h}(X_i) > 0\}) = 1 > p_x$, the set is nonempty and its infimum is attained at one of these finitely many points. □

Lemma 15. *Let $\alpha \in (0,1)$, $\sigma_{\mathrm{low}}, r_-, r_+ > 0$, $a \le b$, and let $G\colon \mathbb{R} \to [0,1]$ be nondecreasing with*

$$G(b+v) \ge 1-\alpha+\sigma_{\mathrm{low}}v \quad (0 \le v \le r_+), \qquad G(a-v) \le 1-\alpha-\sigma_{\mathrm{low}}v \quad (0 \le v \le r_-).$$

Let $H\colon \mathbb{R} \to [0,\infty)$ be nondecreasing and right-continuous with $\lim_{t\to-\infty} H(t) = 0$ and $\sup_t |H(t) - G(t)| \le \zeta$, let $p \in [1-\alpha, 1-\alpha+\bar{s}]$ with $\bar{s} > 0$, and suppose $e := \zeta + \bar{s} \le \sigma_{\mathrm{low}}(r_- \wedge r_+)$. Then $\mathrm{Q}(p;H)$ is finite and $a - e/\sigma_{\mathrm{low}} \le \mathrm{Q}(p;H) \le b + e/\sigma_{\mathrm{low}}$.

Proof. Put $t_+ := b + e/\sigma_{\mathrm{low}}$; since $e/\sigma_{\mathrm{low}} \le r_+$, $G(t_+) \ge 1-\alpha+e$ and $H(t_+) \ge G(t_+) - \zeta \ge 1-\alpha+\bar{s} \ge p$, so the superlevel set $\{t : H(t) \ge p\}$ is nonempty and $\mathrm{Q}(p;H) \le t_+$. Let $t < a - e/\sigma_{\mathrm{low}}$ and $v := a - t > e/\sigma_{\mathrm{low}}$. If $v \le r_-$, then $G(t) \le 1-\alpha-\sigma_{\mathrm{low}}v < 1-\alpha-e$; if $v > r_-$, monotonicity gives $G(t) \le G(a-r_-) \le 1-\alpha-\sigma_{\mathrm{low}}r_- \le 1-\alpha-e$. In both cases $H(t) \le G(t) + \zeta \le 1-\alpha-\bar{s} < 1-\alpha \le p$, so t is not in the superlevel set. Hence $\mathrm{Q}(p;H) \ge a - e/\sigma_{\mathrm{low}}$. □

Lemma 16. *Let **K** 1 hold and $Z_{\tilde{x},h} > 0$. Under **A** 1$(S, 1-\alpha)$, for every $x \in B(\tilde{x},h) \cap \mathcal{X}_0$ with $\omega_S(\tilde{x},h) < \infty$, the function $G := F^{(S)}_{\tilde{x},h}(\cdot)$ satisfies the hypotheses of Lemma 15 with $a = q_S(x) - \omega_S(\tilde{x},h)$ and $b = q_S(x) + \omega_S(\tilde{x},h)$.*

Proof. Since $\Lambda_{\tilde{x},h} \ll P_X$ and $\Lambda_{\tilde{x},h}$ is supported in $B(\tilde{x},h)$, $\Lambda_{\tilde{x},h}$-almost every u lies in $B(\tilde{x},h) \cap \mathcal{X}_0$. Fix $u \in \mathcal{X}_0$. By **A** 1$(S,1-\alpha)$(a), $F^{(S|X)}(\cdot \mid u)$ is continuous; as $F^{(S|X)}(t \mid u) < 1-\alpha$ for $t < q_S(u)$ and $F^{(S|X)}(q_S(u) \mid u) \ge 1-\alpha$, continuity gives $F^{(S|X)}(q_S(u) \mid u) = 1-\alpha$. Hence (b) reads $F^{(S|X)}(q_S(u)+v \mid u) \ge 1-\alpha+\sigma_{\mathrm{low}}v$ for $0 \le v \le r_+$ and $F^{(S|X)}(q_S(u)-v \mid u) \le 1-\alpha-\sigma_{\mathrm{low}}v$ for $0 \le v \le r_-$. For $u \in B(\tilde{x},h) \cap \mathcal{X}_0$, $|q_S(u) - q_S(x)| \le \omega_S(\tilde{x},h)$ by (9), so $q_S(x) + \omega_S(\tilde{x},h) + v \ge q_S(u) + v$ and $q_S(x) - \omega_S(\tilde{x},h) - v \le q_S(u) - v$; monotonicity of $F^{(S|X)}(\cdot \mid u)$ gives $F^{(S|X)}(q_S(x) + \omega_S(\tilde{x},h) + v \mid u) \ge 1-\alpha+\sigma_{\mathrm{low}}v$ and $F^{(S|X)}(q_S(x) - \omega_S(\tilde{x},h) - v \mid u) \le 1-\alpha-\sigma_{\mathrm{low}}v$ for such u. Integrating against $\Lambda_{\tilde{x},h}$ in (23) proves the claim. □

C.2 The localized calibration event

Proposition 17 (Localized calibration event)**.** *Let **K** 1 hold, let S be a fixed measurable score, and fix $\delta \in (0,1)$. Define, with the notation of* (24)*,*

$$\begin{aligned}\mathcal{E}^{\mathrm{R}}_{m,h}(\delta) := &\Big\{|B_m - Z_{\tilde{x},h}| \le Z_{\tilde{x},h}\Big(\sqrt{\tfrac{2\log(4/\delta)}{m_{\tilde{x},h}}} + \tfrac{\log(4/\delta)}{3m_{\tilde{x},h}}\Big)\Big\} \\ &\cap \Big\{\sup_{t\in\mathbb{R}} |A_m(t) - A(t)| \le Z_{\tilde{x},h}\Big(\tfrac{8}{\sqrt{m_{\tilde{x},h}}} + \sqrt{\tfrac{2\log(4/\delta)}{m_{\tilde{x},h}}} + \tfrac{4\log(4/\delta)}{3m_{\tilde{x},h}}\Big)\Big\}.\end{aligned} \tag{29}$$

Then $\mathbb{P}(\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)) \ge 1-\delta$. If moreover (26) *holds, then on $\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)$:*

(i) $B_m \geq Z_{\tilde{x},h}/2 > 0$;
(ii) $\rho_x \leq 2/m_{\tilde{x},h} \leq \alpha/2$ *for every* $x \in \mathbb{R}^d$;
(iii) $\sup_{t\in\mathbb{R}} |\overline{F}^{(S)}_{m,\tilde{x},h}(t) - F^{(S)}_{\tilde{x},h}(t)| \leq \eta_{m,h}(\delta)$;
(iv) for every $x \in \mathbb{R}^d$, $\widehat{Q}^{\mathrm{R}}(x;\tilde{x}) = \mathrm{Q}(p_x; \overline{F}^{(S)}_{m,\tilde{x},h})$ *is finite, with* p_x *as in Lemma 14 and* $1-\alpha \leq p_x \leq 1-\alpha+s_{m,h}$.

Proof. Kernel mass. The weights $w_{\tilde{x},h}(X_i)$ are i.i.d. with values in $[0, \|\mathcal{K}\|_\infty]$, $\mathbb{E}w_{\tilde{x},h}(X_1) = Z_{\tilde{x},h}$ and $\mathbb{E}w_{\tilde{x},h}(X_1)^2 \leq \|\mathcal{K}\|_\infty Z_{\tilde{x},h}$; both $w_{\tilde{x},h}(X_1) - Z_{\tilde{x},h}$ and $Z_{\tilde{x},h} - w_{\tilde{x},h}(X_1)$ are bounded above by $\|\mathcal{K}\|_\infty$. Bernstein's inequality (Boucheron et al., 2013, Theorem 2.10), applied to $\sum_i (w_{\tilde{x},h}(X_i) - Z_{\tilde{x},h})$ and to its negative with variance proxy $m\|\mathcal{K}\|_\infty Z_{\tilde{x},h}$ and scale $\|\mathcal{K}\|_\infty/3$, gives for each sign

$$\mathbb{P}\Big(\pm\sum_{i=1}^m \big(w_{\tilde{x},h}(X_i) - Z_{\tilde{x},h}\big) \geq \sqrt{2m\|\mathcal{K}\|_\infty Z_{\tilde{x},h}\log(4/\delta)} + \tfrac{1}{3}\|\mathcal{K}\|_\infty \log(4/\delta)\Big) \leq \delta/4.$$

Dividing by m and using $\|\mathcal{K}\|_\infty/(mZ_{\tilde{x},h}) = 1/m_{\tilde{x},h}$, the first event in (29) fails with probability at most $\delta/2$.

Weighted empirical process. Both A_m and A are right-continuous in t, so the supremum over $t \in \mathbb{R}$ equals the supremum over $t \in \mathbb{Q}$. By symmetrization (van der Vaart and Wellner, 1996, Lemma 2.3.1) with i.i.d. Rademacher signs $\epsilon_{1:m}$ independent of the data, $\mathbb{E}\sup_t |A_m(t) - A(t)| \leq (2/m)\mathbb{E}\sup_t |\sum_i \epsilon_i w_{\tilde{x},h}(X_i)\mathbb{1}_{\{S_i\leq t\}}|$. Conditionally on the data, fix an ordering $S_{\pi(1)} \leq \cdots \leq S_{\pi(m)}$; every set $\{i : S_i \leq t\}$ is a prefix $\{\pi(1),\ldots,\pi(k)\}$ of this ordering (with tied scores, not every prefix occurs), so the inner supremum is at most $\max_{0\leq k\leq m} |\sum_{j\leq k} \epsilon_{\pi(j)} w_{\tilde{x},h}(X_{\pi(j)})|$, the maximal absolute value of a martingale, and Doob's L^2 inequality bounds its conditional expectation by $2(\sum_i w_{\tilde{x},h}(X_i)^2)^{1/2}$. Since $\mathbb{E}\sum_i w_{\tilde{x},h}(X_i)^2 \leq m\|\mathcal{K}\|_\infty Z_{\tilde{x},h}$, Jensen's inequality yields

$$\mathbb{E}\sup_t |A_m(t) - A(t)| \leq 4\sqrt{\|\mathcal{K}\|_\infty Z_{\tilde{x},h}/m} = 4Z_{\tilde{x},h}/\sqrt{m_{\tilde{x},h}}. \tag{30}$$

Next apply Bousquet's inequality (Boucheron et al., 2013, Theorem 12.5) to the countable class of functions $(u,y) \mapsto \pm\big(w_{\tilde{x},h}(u)\mathbb{1}_{\{S(u,y)\leq t\}} - A(t)\big)/\|\mathcal{K}\|_\infty$, $t \in \mathbb{Q}$, which are centred, take values in $[-1,1]$, and have variance at most $\mathbb{E}[w_{\tilde{x},h}(X)^2]/\|\mathcal{K}\|_\infty^2 \leq Z_{\tilde{x},h}/\|\mathcal{K}\|_\infty$. With $\mathcal{Z} := m\sup_t |A_m(t) - A(t)|/\|\mathcal{K}\|_\infty$ and $v := 2\mathbb{E}\mathcal{Z} + mZ_{\tilde{x},h}/\|\mathcal{K}\|_\infty$, it gives $\mathbb{P}\big(\mathcal{Z} \geq \mathbb{E}\mathcal{Z} + \sqrt{2v\log(4/\delta)} + \frac{1}{3}\log(4/\delta)\big) \leq \delta/4$. Using $\sqrt{2v\log(4/\delta)} \leq 2\sqrt{\log(4/\delta)\,\mathbb{E}\mathcal{Z}} + \sqrt{2mZ_{\tilde{x},h}\log(4/\delta)/\|\mathcal{K}\|_\infty} \leq \mathbb{E}\mathcal{Z} + \log(4/\delta) + \sqrt{2mZ_{\tilde{x},h}\log(4/\delta)/\|\mathcal{K}\|_\infty}$, then multiplying by $\|\mathcal{K}\|_\infty/m$ and inserting (30), we get with probability at least $1-\delta/4$

$$\begin{aligned}\sup_t |A_m(t) - A(t)| &\leq 2\mathbb{E}\sup_t |A_m(t) - A(t)| + \sqrt{\frac{2\|\mathcal{K}\|_\infty Z_{\tilde{x},h}\log(4/\delta)}{m}} + \frac{4\|\mathcal{K}\|_\infty\log(4/\delta)}{3m} \\ &\leq Z_{\tilde{x},h}\Big(\frac{8}{\sqrt{m_{\tilde{x},h}}} + \sqrt{\frac{2\log(4/\delta)}{m_{\tilde{x},h}}} + \frac{4\log(4/\delta)}{3m_{\tilde{x},h}}\Big),\end{aligned}$$

which is the second event in (29). A union bound gives $\mathbb{P}(\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)) \geq 1 - 3\delta/4 \geq 1-\delta$.

Consequences. Assume (26), i.e. $m_{\tilde{x},h} \geq 64\log(4/\delta) \vee 4/\alpha$, and work on $\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)$; note $\log(4/\delta) \geq \log 4 > 1$ and $\sqrt{\log(4/\delta)/m_{\tilde{x},h}} \leq 1/8$. (i) $\sqrt{2\log(4/\delta)/m_{\tilde{x},h}} + \log(4/\delta)/(3m_{\tilde{x},h}) \leq \sqrt{2/64} + 1/192 < 1/2$, so $B_m \geq Z_{\tilde{x},h}/2$. (ii) $\rho_x \leq w_{\tilde{x},h}(x)/(mB_m) \leq \|\mathcal{K}\|_\infty/(mZ_{\tilde{x},h}/2) = 2/m_{\tilde{x},h} \leq \alpha/2$. (iii) Writing $\overline{F}^{(S)}_{m,\tilde{x},h} - F^{(S)}_{\tilde{x},h}(\cdot) = (A_m - A)/B_m + (A/Z_{\tilde{x},h})(Z_{\tilde{x},h} - B_m)/B_m$

and using $A(t)/Z_{\tilde{x},h} = F^{(S)}_{\tilde{x},h}(t) \le 1$ and (i),

$$\begin{aligned}\sup_t |\overline{F}^{(S)}_{m,\tilde{x},h}(t) - F^{(S)}_{\tilde{x},h}(t)| &\le \frac{2}{Z_{\tilde{x},h}}\Big(\sup_t |A_m(t) - A(t)| + |B_m - Z_{\tilde{x},h}|\Big)\\ &\le \frac{16}{\sqrt{m_{\tilde{x},h}}} + 4\sqrt{\frac{2\log(4/\delta)}{m_{\tilde{x},h}}} + \frac{10\log(4/\delta)}{3m_{\tilde{x},h}}\\ &\le \big(16 + 4\sqrt{2} + \tfrac{10}{24}\big)\sqrt{\frac{\log(4/\delta)}{m_{\tilde{x},h}}} \le 23\sqrt{\frac{\log(4/\delta)}{m_{\tilde{x},h}}} = \eta_{m,h}(\delta),\end{aligned}$$

where $16/\sqrt{m_{\tilde{x},h}} \le 16\sqrt{\log(4/\delta)/m_{\tilde{x},h}}$ because $\log(4/\delta) > 1$, and $\log(4/\delta)/m_{\tilde{x},h} \le \frac{1}{8}\sqrt{\log(4/\delta)/m_{\tilde{x},h}}$. (iv) By (i) and (ii), Lemma 14 applies with $\rho_x \le \alpha/2 < \alpha$; since $\rho_x \le 1/2$, $p_x - (1-\alpha) \le 2(1-\alpha)\rho_x \le 2\rho_x \le 4/m_{\tilde{x},h} = s_{m,h}$.

□

C.3 Complete statement and proof of Theorem 4

The complete theorem below does not use **P** 1: it is stated for any centre with $Z_{\tilde{x},h} > 0$, in terms of $m_{\tilde{x},h}$, under the size condition (26) and an admissibility condition on $\eta_{m,h}(\delta) + s_{m,h}$. The simplified form of Theorem 4 follows through (27) with the sample-size constant

$$C_{\mathrm{R}} := \frac{1}{c_K}\max\Big\{64,\ \frac{4}{\alpha},\ \frac{2304}{\sigma^2_{\mathrm{low}}(r_- \wedge r_+)^2}\Big\}. \tag{31}$$

The sample-size condition is

$$mp_{\min}h^d \ge C_{\mathrm{R}}\log(4/\delta). \tag{32}$$

Under **K** 1 and **P** 1 and (32), (27) gives $m_{\tilde{x},h} \ge c_K mp_{\min}h^d \ge \max\{64, 4/\alpha, 2304/(\sigma_{\mathrm{low}}(r_- \wedge r_+))^2\}\log(4/\delta)$; since $\log(4/\delta) > 1$ this implies (26), and it implies $\sqrt{\log(4/\delta)/m_{\tilde{x},h}} \le \sigma_{\mathrm{low}}(r_- \wedge r_+)/48$. Moreover, under (26), $s_{m,h} = 4/m_{\tilde{x},h} \le \frac{1}{2}\sqrt{\log(4/\delta)/m_{\tilde{x},h}}$ because $\sqrt{m_{\tilde{x},h}\log(4/\delta)} \ge 8\log(4/\delta) \ge 8$; combining with (27),

$$\begin{aligned}\eta_{m,h}(\delta) &\le \frac{23}{\sqrt{c_K}}\sqrt{\frac{\log(4/\delta)}{mp_{\min}h^d}}, \qquad s_{m,h} \le \frac{4}{c_K mp_{\min}h^d},\\ \eta_{m,h}(\delta) + s_{m,h} &\le 24\sqrt{\frac{\log(4/\delta)}{m_{\tilde{x},h}}} \le \tfrac{1}{2}\sigma_{\mathrm{low}}(r_- \wedge r_+).\end{aligned} \tag{33}$$

Theorem 4 (complete version). *Let **K** 1 hold, fix $h > 0$, a centre $\tilde{x}$ with $Z_{\tilde{x},h} > 0$, and $\delta \in (0,1)$, suppose that (26) holds, and let $\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)$ be the event (29), which has $\mathbb{P}(\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)) \ge 1-\delta$. If **A** 1$(S, 1-\alpha)$ holds and*

$$\eta_{m,h}(\delta) + s_{m,h} \le \sigma_{\mathrm{low}}(r_- \wedge r_+), \tag{34}$$

then on $\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)$ the following hold simultaneously for every $x \in B(\tilde{x}, h) \cap \mathcal{X}_0$.

(i) (Threshold) $\widehat{Q}^{\mathrm{R}}(x;\tilde{x})$ is finite and $\big|\widehat{Q}^{\mathrm{R}}(x;\tilde{x}) - q_S(x)\big| \le \omega_S(\tilde{x},h) + \dfrac{\eta_{m,h}(\delta) + s_{m,h}}{\sigma_{\mathrm{low}}}$.

(ii) (Coverage) $\big|\mathrm{Cov}_{\widehat{\mathcal{C}}^{\mathrm{R}}(\cdot;\tilde{x})}(x) - (1-\alpha)\big| \le \sigma_{\mathrm{up}}\Big(\omega_S(\tilde{x},h) + \dfrac{\eta_{m,h}(\delta) + s_{m,h}}{\sigma_{\mathrm{low}}}\Big)$.

*(iii) (Length) If moreover **A** 2(S) holds, then $\big|\,|\widehat{\mathcal{C}}^{\mathrm{R}}(x;\tilde{x})| - |\mathcal{C}_S(x, q_S(x))|\,\big| \le L_{\mathrm{len}}\Big(\omega_S(\tilde{x},h) + \dfrac{\eta_{m,h}(\delta) + s_{m,h}}{\sigma_{\mathrm{low}}}\Big)$.*

Proof of Theorem 4, simplified form, from the complete version. Under **P** 1, $Z_{\tilde{x},h} > 0$ by (27), and (32) implies (26) and, by (33), (34); so the complete version applies. By (33) and (27), $\eta_{m,h}(\delta) + s_{m,h} \le (24/\sqrt{c_K})\sqrt{\log(4/\delta)/(mp_{\min}h^d)}$: the rank correction is absorbed

into the fluctuation term. Inserting this into (ii) and (iii), and using $\sigma_{\mathrm{low}} \le \sigma_{\mathrm{up}}$, gives (10) with constants $(\sigma_{\mathrm{up}}/\sigma_{\mathrm{low}}) \max\{1, 24/\sqrt{c_K}\}$ for coverage and $(L_{\mathrm{len}}/\sigma_{\mathrm{low}}) \max\{1, 24/\sqrt{c_K}\}$ for length, on the event $\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)$. □

Proof of Theorem 4, complete version. Fix $x \in B(\tilde{x}, h) \cap \mathcal{X}_0$. If $\omega_S(\tilde{x}, h) = \infty$ the bounds are trivial once $\widehat{Q}^{\mathrm{R}}(x; \tilde{x})$ is finite, which Proposition 17(iv) gives; so assume $\omega_S(\tilde{x}, h) < \infty$. On $\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)$, Proposition 17(iii)–(iv) give $\widehat{Q}^{\mathrm{R}}(x; \tilde{x}) = \mathrm{Q}(p_x; \overline{F}^{(S)}_{m,\tilde{x},h})$ with $p_x \in [1-\alpha, 1-\alpha+s_{m,h}]$ and $\sup_t |\overline{F}^{(S)}_{m,\tilde{x},h}(t) - F^{(S)}_{\tilde{x},h}(t)| \le \eta_{m,h}(\delta)$. By Lemma 16, $G = F^{(S)}_{\tilde{x},h}(\cdot)$ satisfies the anchor hypotheses of Lemma 15 with $[a, b] = [q_S(x) - \omega_S(\tilde{x}, h), q_S(x) + \omega_S(\tilde{x}, h)]$; the lemma applies with $H = \overline{F}^{(S)}_{m,\tilde{x},h}$, $\zeta = \eta_{m,h}(\delta)$, $\bar{s} = s_{m,h} > 0$, $p = p_x$, and $e = \eta_{m,h}(\delta) + s_{m,h} \le \sigma_{\mathrm{low}}(r_- \wedge r_+)$ by (34). This proves (i).

(ii) Since $\widehat{\mathcal{C}}^{\mathrm{R}}(x; \tilde{x}) = \mathcal{C}_S(x, \widehat{Q}^{\mathrm{R}}(x; \tilde{x}))$ and $\widehat{Q}^{\mathrm{R}}(x; \tilde{x})$ is a function of $(\mathcal{D}^{\mathrm{cal}}_m, \tilde{x}, x)$, $\mathrm{Cov}_{\widehat{\mathcal{C}}^{\mathrm{R}}(\cdot;\tilde{x})}(x) = F^{(S|X)}(\widehat{Q}^{\mathrm{R}}(x; \tilde{x}) \mid x)$. By $\mathbf{A}\,1(S, 1-\alpha)$(a) and $F^{(S|X)}(q_S(x) \mid x) = 1 - \alpha$ (continuity, as in the proof of Lemma 16), $|F^{(S|X)}(\widehat{Q}^{\mathrm{R}}(x; \tilde{x}) \mid x) - (1-\alpha)| \le \sigma_{\mathrm{up}} |\widehat{Q}^{\mathrm{R}}(x; \tilde{x}) - q_S(x)|$, and (i) concludes.

(iii) $\mathbf{A}2(S)$ gives $\big||\mathcal{C}_S(x, \widehat{Q}^{\mathrm{R}}(x; \tilde{x}))| - |\mathcal{C}_S(x, q_S(x))|\big| \le L_{\mathrm{len}} |\widehat{Q}^{\mathrm{R}}(x; \tilde{x}) - q_S(x)|$, and (i) concludes. □

C.4 Quantile regularity and the RLCP rate

To obtain an explicit bandwidth rate from Theorem 4, we impose the following condition on the target quantile.

A 4 $(S, 1-\alpha, \beta_q)$**.** *The exponent satisfies $\beta_q \in (0, 1]$, and for some $L_q > 0$, $|q_S^{1-\alpha}(u) - q_S^{1-\alpha}(u')| \le L_q \|u - u'\|^{\beta_q}$ for all $u, u' \in \mathcal{X}_0$.*

Under $\mathbf{A}\,4(S, 1-\alpha, \beta_q)$, $\omega_S(\tilde{x}, h) \le L_q (2h)^{\beta_q}$. This increment condition is restricted to $\beta_q \le 1$; for higher-order Hölder classes, Section D.6 uses the resulting first-order variation bound.

Set

$$c_{\mathrm{R}} \coloneqq 1 \vee C_{\mathrm{R}}^{(2\beta_q + d)/(2\beta_q)}. \tag{35}$$

Corollary 18 (Rate of RLCP with a fixed score)**.** *Fix a centre $\tilde{x}$, $\delta \in (0, 1)$, and $\beta_q \in (0, 1]$, let **K** 1, **A** 1$(S, 1-\alpha)$, and **A** 4$(S, 1-\alpha, \beta_q)$ hold, and let **P** 1 hold at bandwidth $h = h_\star$, where*

$$h_\star \coloneqq \Big(\frac{\log(4/\delta)}{m p_{\min}}\Big)^{1/(2\beta_q + d)}. \tag{36}$$

If $m p_{\min} \ge c_{\mathrm{R}} \log(4/\delta)$, with c_{R} as in (35), then Theorem 4 applies at $h = h_\star$ and, on the event $\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)$, for every $x \in B(\tilde{x}, h_\star) \cap \mathcal{X}_0$,

$$\big|\mathrm{Cov}_{\widehat{\mathcal{C}}^{\mathrm{R}}(\cdot;\tilde{x})}(x) - (1-\alpha)\big| \;\vee\; \big|\,|\widehat{\mathcal{C}}^{\mathrm{R}}(x; \tilde{x})| - |\mathcal{C}_S(x, q_S(x))|\,\big| \lesssim \Big(\frac{\log(4/\delta)}{m p_{\min}}\Big)^{\beta_q/(2\beta_q + d)}, \tag{37}$$

*the length part under **A** 2(S), with constants depending only on $(\sigma_{\mathrm{up}}, \sigma_{\mathrm{low}}, L_{\mathrm{len}}, c_K, L_q)$.*

Proof of Corollary 18. By definition of $h_\star$,

$$m p_{\min}\, h_\star^d = \log(4/\delta) \Big(\frac{m p_{\min}}{\log(4/\delta)}\Big)^{2\beta_q/(2\beta_q + d)}, \qquad \sqrt{\frac{\log(4/\delta)}{m p_{\min}\, h_\star^d}} = \Big(\frac{\log(4/\delta)}{m p_{\min}}\Big)^{\beta_q/(2\beta_q + d)} = h_\star^{\beta_q}. \tag{38}$$

By (35), $m p_{\min} \ge c_{\mathrm{R}} \log(4/\delta)$ implies $(m p_{\min}/\log(4/\delta))^{2\beta_q/(2\beta_q + d)} \ge C_{\mathrm{R}}$, i.e. (32) at $h = h_\star$ by (38). Hence Theorem 4 applies at $h = h_\star$, and since $\|u - u'\| \le 2h_\star$ for all $u, u' \in B(\tilde{x}, h_\star) \cap \mathcal{X}_0$, $\mathbf{A}\,4(S, 1-\alpha, \beta_q)$ gives $\omega_S(\tilde{x}, h_\star) \le L_q (2h_\star)^{\beta_q} = 2^{\beta_q} L_q h_\star^{\beta_q}$. Inserting this and (38) into (10) bounds both parts of its left-hand side by a constant multiple of $(\sigma_{\mathrm{up}} 2^{\beta_q} L_q + 1) h_\star^{\beta_q} = (\sigma_{\mathrm{up}} 2^{\beta_q} L_q + 1) \big(\log(4/\delta)/(m p_{\min})\big)^{\beta_q/(2\beta_q + d)}$. □

C.5 Relation to conditional-CDF regularity

Remark 19 (Hölder continuity of the conditional score CDF in the covariate)**.** *Consider Hölder continuity in the covariate of the conditional score CDF, uniformly over the threshold: for all $u, u' \in \mathcal{X}_0$,*

$$\sup_{t\in\mathbb{R}} |F^{(S|X)}(t \mid u) - F^{(S|X)}(t \mid u')| \le L_F \|u - u'\|^{\beta_F} \quad \text{for some } \beta_F \in (0,1],\ L_F > 0. \tag{39}$$

Thus nearby covariates have similar conditional score distributions at every threshold t; the Hölder regularity is in the covariate, not in t. This is the localization condition of Guan (2023) with dissimilarity $\|u - u'\|^{\beta_F}$. It controls the whole conditional score distribution, whereas $\boldsymbol{A}$ 4$(S, 1-\alpha, \beta_q)$ only controls its $(1-\alpha)$-quantile. To turn this distributional condition into a quantile bound, we also need growth near the target quantile. If $\boldsymbol{A}$ 1$(S, 1-\alpha)$ holds and $L_F(2h)^{\beta_F} \le \frac{1}{2}\sigma_{\mathrm{low}}(r_- \wedge r_+)$, then

$$\omega_S(\tilde{x}, h) \le L_F(2h)^{\beta_F}/\sigma_{\mathrm{low}}, \tag{40}$$

so the bounds of Theorem 4 hold with $\omega_S(\tilde{x}, h)$ replaced by $L_F(2h)^{\beta_F}/\sigma_{\mathrm{low}}$; Corollary 18 holds with (β_q, L_q) replaced by $(\beta_F, L_F/\sigma_{\mathrm{low}})$ provided c_{R} in (35) is enlarged to also dominate $\big(2^{\beta_F+1}L_F/(\sigma_{\mathrm{low}}(r_- \wedge r_+))\big)^{(2\beta_F+d)/\beta_F}$, which guarantees the bandwidth restriction at $h = h_\star$. The converse fails, even for a constant quantile and a discontinuous conditional CDF (Remark 20).

Proof of (40). Fix $u, u' \in B(\tilde{x}, h) \cap \mathcal{X}_0$, so that $\|u - u'\| \le 2h$. Put $\zeta := L_F(2h)^{\beta_F} \ge L_F\|u - u'\|^{\beta_F}$ and $r := r_- \wedge r_+$, so that $\zeta/\sigma_{\mathrm{low}} \le r/2$ by assumption. As in the proof of Lemma 16, $\mathbf{A}\,1(S, 1-\alpha)$ gives $F^{(S|X)}(q_S(u') + v \mid u') \ge 1 - \alpha + \sigma_{\mathrm{low}} v$ for $0 \le v \le r_+$ and $F^{(S|X)}(q_S(u') - v \mid u') \le 1 - \alpha - \sigma_{\mathrm{low}} v$ for $0 \le v \le r_-$. Taking $v = \zeta/\sigma_{\mathrm{low}} \le r_+$ in the first inequality and using (39), $F^{(S|X)}(q_S(u') + \zeta/\sigma_{\mathrm{low}} \mid u) \ge 1 - \alpha + \zeta - \zeta = 1 - \alpha$, whence $q_S(u) \le q_S(u') + \zeta/\sigma_{\mathrm{low}}$. Next let $t < q_S(u') - \zeta/\sigma_{\mathrm{low}}$ and $v := q_S(u') - t > \zeta/\sigma_{\mathrm{low}}$. If $v \le r_-$, then $F^{(S|X)}(t \mid u) \le F^{(S|X)}(t \mid u') + \zeta \le 1 - \alpha - \sigma_{\mathrm{low}} v + \zeta < 1 - \alpha$; if $v > r_-$, monotonicity and $\zeta \le \sigma_{\mathrm{low}} r_-/2$ give $F^{(S|X)}(t \mid u) \le F^{(S|X)}(q_S(u') - r_- \mid u') + \zeta \le 1 - \alpha - \sigma_{\mathrm{low}} r_-/2 < 1 - \alpha$. Hence $t < q_S(u)$ for every such t, i.e. $q_S(u) \ge q_S(u') - \zeta/\sigma_{\mathrm{low}}$. Exchanging u and u' and taking the supremum gives $\omega_S(\tilde{x}, h) \le \zeta/\sigma_{\mathrm{low}}$. □

Remark 20. *Let $X \sim \mathrm{Unif}[-1, 1]$, put $p = 1 - \alpha$, and take $S(x, y) = |y|$, with*

$$Y \mid X = x \sim \begin{cases} p\,\mathrm{Unif}[0,1] + \alpha\,\mathrm{Unif}[1,2], & x < 0, \\ p\,\mathrm{Unif}[1/2,1] + \alpha\,\mathrm{Unif}[1,2], & x \ge 0. \end{cases}$$

Then $q_S^{1-\alpha}(x) = 1$ for every x. The score densities are bounded by $\max\{2p, \alpha\}$; for $0 \le u \le 1/2$ and $0 \le v \le 1$, $F^{(S|X)}(1 \mid x) - F^{(S|X)}(1 - u \mid x) \ge pu$ and $F^{(S|X)}(1 + v \mid x) - F^{(S|X)}(1 \mid x) = \alpha v$. Thus $\boldsymbol{A}$ 1$(S, 1-\alpha)$ holds with $\sigma_{\mathrm{up}} = \max\{2p, \alpha\}$, $\sigma_{\mathrm{low}} = \min\{p, \alpha\}$, $r_- = 1/2$ and $r_+ = 1$. Also $|\mathcal{C}_S(x, t)| = 2\max\{t, 0\}$, so $\boldsymbol{A}$ 2(S) holds with $L_{\mathrm{len}} = 2$. Yet $F^{(S|X)}(1/2 \mid x)$ jumps from $p/2$ to zero at $x = 0$, violating (39) for every positive exponent. Every mixture of these conditional score laws has CDF equal to p at 1 and strictly below p for $t < 1$, by the uniform lower-growth bound and monotonicity. Hence every local mixture retains quantile 1 and $\omega_S(\tilde{x}, h) = 0$, including in neighbourhoods crossing zero.

The next proposition gives a direct coverage bound under (39). It retains the upper CDF regularity of $\mathbf{A}\,1(S, 1-\alpha)$(a), but needs neither the lower-growth condition in part (b) nor the admissibility condition (34). Unlike Theorem 4, it does not bound the threshold or length errors.

Proposition 21 (Coverage under Hölder continuity of the conditional score CDF in the covariate)**.** *Let $\boldsymbol{K}$1 hold, let S be a fixed measurable score, fix $h > 0$, a centre $\tilde{x}$ with $Z_{\tilde{x},h} > 0$, and $\delta \in (0, 1)$, and suppose that (26) holds. If $\boldsymbol{A}$ 1$(S, 1-\alpha)$(a) and the CDF continuity condition (39) hold — neither $\boldsymbol{A}$ 1$(S, 1-\alpha)$(b) nor (34) being assumed — then on the event $\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)$ of (29), simultaneously for every $x \in B(\tilde{x}, h) \cap \mathcal{X}_0$, $\widehat{Q}^{\mathrm{R}}(x; \tilde{x})$ is finite and*

$$\big|\mathrm{Cov}_{\widehat{\mathcal{C}}^{\mathrm{R}}(\cdot;\tilde{x})}(x) - (1-\alpha)\big| \le L_F(2h)^{\beta_F} + \eta_{m,h}(\delta) + s_{m,h}. \tag{41}$$

Proof of Proposition 21. Fix $x \in B(\tilde{x}, h) \cap \mathcal{X}_0$ and work on $\mathcal{E}^{\mathrm{R}}_{m,h}(\delta)$. Only Proposition 17(iii)–(iv) are used, which require (26) alone; in particular $\widehat{Q}^{\mathrm{R}}(x; \tilde{x})$ is finite, and, as in part (ii) of the proof of the complete version of Theorem 4, $\mathrm{Cov}_{\widehat{\mathcal{C}}^{\mathrm{R}}(\cdot;\tilde{x})}(x) = F^{(S|X)}(\widehat{Q}^{\mathrm{R}}(x; \tilde{x}) \mid x)$. Decompose $F^{(S|X)}(\widehat{Q}^{\mathrm{R}}(x; \tilde{x}) \mid x) - (1-\alpha) = \big(F^{(S|X)}(\widehat{Q}^{\mathrm{R}}(x; \tilde{x}) \mid x) - F^{(S)}_{\tilde{x},h}(\widehat{Q}^{\mathrm{R}}(x; \tilde{x}))\big) + \big(F^{(S)}_{\tilde{x},h}(\widehat{Q}^{\mathrm{R}}(x; \tilde{x})) - p_x\big) + \big(p_x - (1-\alpha)\big)$. For the first term, $\Lambda_{\tilde{x},h}$-almost every u lies in $B(\tilde{x}, h) \cap \mathcal{X}_0$, so $\|u - x\| \le 2h$ and (39) gives $|F^{(S|X)}(t \mid x) - F^{(S)}_{\tilde{x},h}(t)| \le \int |F^{(S|X)}(t \mid x) - F^{(S|X)}(t \mid u)| \, \Lambda_{\tilde{x},h}(\mathrm{d}u) \le L_F (2h)^{\beta_F}$ for every t. For the second, $\overline{F}^{(S)}_{m,\tilde{x},h}(\widehat{Q}^{\mathrm{R}}(x; \tilde{x})) \ge p_x$ and $\overline{F}^{(S)}_{m,\tilde{x},h}(t) < p_x$ for $t < \widehat{Q}^{\mathrm{R}}(x; \tilde{x})$ by definition of the left quantile, so Proposition 17(iii) gives $F^{(S)}_{\tilde{x},h}(\widehat{Q}^{\mathrm{R}}(x; \tilde{x})) \ge p_x - \eta_{m,h}(\delta)$ and $F^{(S)}_{\tilde{x},h}(t) < p_x + \eta_{m,h}(\delta)$ for $t < \widehat{Q}^{\mathrm{R}}(x; \tilde{x})$; since $F^{(S)}_{\tilde{x},h}(\cdot)$ is σ_{up}-Lipschitz by $\mathbf{A}\,1(S, 1-\alpha)$(a) integrated against $\Lambda_{\tilde{x},h}$, letting $t \uparrow \widehat{Q}^{\mathrm{R}}(x; \tilde{x})$ gives $|F^{(S)}_{\tilde{x},h}(\widehat{Q}^{\mathrm{R}}(x; \tilde{x})) - p_x| \le \eta_{m,h}(\delta)$. The third term lies in $[0, s_{m,h}]$ by Proposition 17(iv). □

D Proofs for the Common-Score Comparison

Throughout this appendix, we omit the fixed coverage level $1-\alpha$ from the quantile notation for brevity.

D.1 Complete statement and proof of Proposition 5

Proposition 5 (complete version). *Let $\boldsymbol{A}$ 1$(S, 1-\alpha)$ hold, conditional on the initial fit. Take a finite measurable version of $q_S(\cdot)$ on $\mathcal{X}$, with arbitrary finite values outside the full-measure set $\mathcal{X}_0$ if necessary. Then:*

(i) For every $x \in \mathcal{X}_0$ and $t \in \mathbb{R}$,

$$F^{(S^{\mathrm{ad}}|X)}(t \mid x) = F^{(S|X)}(t + q_S(x) \mid x), \qquad F^{(S|X)}(q_S(x) \mid x) = 1 - \alpha.$$

(ii) $\boldsymbol{A}$ 3$(S^{\mathrm{ad}}, 1-\alpha)$ holds with $q^\star = 0$, and $\boldsymbol{A}$ 1$(S^{\mathrm{ad}}, 1-\alpha)$ holds with the same $(\sigma_{\mathrm{up}}, \sigma_{\mathrm{low}}, r_-, r_+)$. Its conditional $(1-\alpha)$-quantile is zero on $\mathcal{X}_0$, and its marginal $(1-\alpha)$-quantile is zero.

(iii) If $\boldsymbol{A}$ 2(S) holds, it also holds for S^{ad} with the same L_{len}, and $\mathcal{C}_{S^{\mathrm{ad}}}(x, 0) = C^\circ_S(x)$ for $x \in \mathcal{X}_0$.

(iv) The error $\Delta_{S^{\mathrm{ad}}}(x)$ of $\widehat{S}^{\mathrm{ad}}$ relative to S^{ad} is given by (14), and is measurable whenever $\widehat{q}_S(\cdot)$ is.

The compact statement in the main text collects items (ii)–(iv).

Proof. Condition on $\mathcal{D}_0$ and fix $x \in \mathcal{X}_0$. The quantile $q_S(x)$ is finite by $\mathbf{A}\,1(S, 1-\alpha)$(b). Its measurability follows from that of the conditional CDF, using rational thresholds to characterize the left quantile. The identity $\{y : S^{\mathrm{ad}}(x, y) \le t\} = \{y : S(x, y) \le t + q_S(x)\}$ gives (i)'s CDF formula. By $\mathbf{A}\,1(S, 1-\alpha)$(a), this conditional CDF is continuous. It equals $1-\alpha$ at its finite left quantile: the value there is at least $1-\alpha$ by right-continuity and the definition of the left quantile, and a strictly larger value would, by continuity, give a value larger than $1-\alpha$ at a smaller threshold, contradicting minimality.

Translation of a CDF translates its left quantile, so $q_{S^{\mathrm{ad}}}(x) = 0$. Its upper Lipschitz bound is unchanged, and its lower increments at zero are exactly those of the original CDF at $q_S(x)$. Thus $\mathbf{A}\,1(S^{\mathrm{ad}}, 1-\alpha)$ holds at the original constants and $\mathbf{A}\,3(S^{\mathrm{ad}}, 1-\alpha)$ holds at zero. Lemma 11 then identifies the marginal quantile as zero, proving (ii). The sublevel-set identity $\mathcal{C}_{S^{\mathrm{ad}}}(x, t) = \mathcal{C}_S(x, t + q_S(x))$ gives both claims in (iii), since translation preserves the length Lipschitz constant. Finally, $\widehat{S}^{\mathrm{ad}}(x, y) - S^{\mathrm{ad}}(x, y) = q_S(x) - \widehat{q}_S(x)$ does not depend on y, which proves (iv), including measurability. □

D.2 Conditions for the common-score rates

Condition throughout on the initial sample $\mathcal{D}_0$ that fixes S; all remaining observations are i.i.d. and independent of $\mathcal{D}_0$. Fix $\alpha \in (0,1)$, let $\mathcal{X} = [0,1]^d$, and assume $0 < p_{\min} \le p_X \le p_{\max} < \infty$. Suppose that $q_S(\cdot) \in \mathcal{H}^{\beta_q}(\mathcal{X}, L_q)$ for a fixed $\beta_q > 0$ and that $|q_S(x)| \le M$. Here $\mathcal{H}^{\beta_q}$ is the Hölder ball with derivatives through order $s_q = \lceil \beta_q \rceil - 1$ bounded by L_q and derivatives of order s_q Hölder of exponent $\beta_q - s_q$ with constant L_q; this convention includes integer smoothness. For every $x \in \mathcal{X}$, assume

$$\begin{aligned} 0 \le F^{(S|X)}(v \mid x) - F^{(S|X)}(u \mid x) \le \sigma_{\mathrm{up}}(v-u) \quad (u \le v), \\ F^{(S|X)}(v \mid x) - F^{(S|X)}(u \mid x) \ge \sigma_{\mathrm{low}}(v-u) \quad \big(q_S(x) - r_0 \le u \le v \le q_S(x) + r_0\big), \end{aligned} \tag{42}$$

for $\sigma_{\mathrm{up}} \ge \sigma_{\mathrm{low}} > 0$ and $r_0 > 0$. The lower bound holds between every pair of thresholds in the window, as used by the local-polynomial bound, and not only for increments from $q_S(x)$. Assume $\mathbf{A}\,2(S)$ only for the length conclusions. The score response $Z = S(X,Y)$ need not have bounded support.

The first line of (42) makes every conditional score CDF continuous; since $|q_S(x)| \le M$ makes the left quantile finite, its value there is at least $1-\alpha$ by right-continuity, and continuity rules out a strictly larger value, so $F^{(S|X)}(q_S(x) \mid x) = 1-\alpha$. The second line, evaluated with one endpoint at $q_S(x)$, gives the two lower-growth inequalities. Thus $\mathbf{A}\,1(S, 1-\alpha)$ holds with $\mathcal{X}_0 = \mathcal{X}$ and constants $(\sigma_{\mathrm{up}}, \sigma_{\mathrm{low}}, r_0, r_0)$. By Proposition 5, the same constants hold for S^{ad}, with pivot zero and oracle set C_S°.

D.3 Isotropic Hölder score quantiles

The isotropic case $J = 0$, $t_0 = d$, $\beta_0 = \beta_q$ of Proposition 6 is recorded separately, with the explicit architecture and finite-confidence estimation bound of Proposition 26 (Section E.1).

Proposition 22 (Isotropic Hölder score quantiles)**.** *Let the conditions of Section D.2 hold and fix $\delta \in (0,1)$. Split $N = n+m$, use the architecture of Proposition 26 on $\mathcal{D}_n^{\mathrm{tr}}$, and calibrate $\widehat{S}^{\mathrm{ad}}$ globally on the independent fold $\mathcal{D}_m^{\mathrm{cal}}$. Let $C_{\mathrm{NN}}, n_{\mathrm{NN}}$ be the constants in Proposition 26, and put*

$$\varepsilon_{\mathrm{NN}} := C_{\mathrm{NN}} \left\{ n^{-\beta_q/(2\beta_q+d)} (\log n)^{3/2} + \sqrt{\frac{\log(4/\delta)}{n}} \right\}.$$

If

$$n \ge n_{\mathrm{NN}} \vee 2, \qquad m \ge 4, \qquad \sigma_{\mathrm{up}} \varepsilon_{\mathrm{NN}} + \eta_m(\delta/2) + s_m \le \sigma_{\mathrm{low}} r_0,$$

then, with probability at least $1-\delta$ over both folds,

$$\| \operatorname{Cov}_{\hat{\mathcal{C}}_\alpha^{\mathrm{ad}}} - (1-\alpha) \|_{L^2(P_X)} \ \vee \ \| |\hat{\mathcal{C}}_\alpha^{\mathrm{ad}}(\cdot)| - |C_S^\circ(\cdot)| \|_{L^2(P_X)} \lesssim n^{-\beta_q/(2\beta_q+d)} (\log n)^{3/2} + \sqrt{\tfrac{\log(4/\delta)}{n \wedge m}}. \tag{43}$$

Both norms are $L^2(P_X)$ norms; this is not a uniform pointwise rate. The constants depend only on $(\alpha, M, \sigma_{\mathrm{low}}, \sigma_{\mathrm{up}}, r_0, \beta_q, L_q, d, p_{\min}, p_{\max}, F)$ and the fixed architecture constants, also on L_{len} for length. For every fixed δ, the sample-size conditions hold for sufficiently large n, m.

Proof. Proposition 26 bounds the $L^2(P_X)$ error of (15) for the conditional $(1-\alpha)$-quantile of $Z = S(X,Y)$ under the rate conditions above. Applied at confidence $\delta/2$, so that $\log(2/(\delta/2)) = \log(4/\delta)$, it gives a training event $\mathcal{T}$ with $\mathbb{P}(\mathcal{T} \mid \mathcal{D}_0) \ge 1 - \delta/2$ on which

$$\|\Delta_{S^{\mathrm{ad}}}\|_{L^2(P_X)} = \|\widehat{q}_S(\cdot) - q_S(\cdot)\|_{L^2(P_X)} \le \varepsilon_{\mathrm{NN}},$$

the equality by (14).

On $\mathcal{T}$, $\|\Delta_{S^{\mathrm{ad}}}\|_{L^1(P_X)} \le \|\Delta_{S^{\mathrm{ad}}}\|_{L^2(P_X)} \le \varepsilon_{\mathrm{NN}}$, since P_X is a probability measure. The sample-size condition therefore allows the complete version of Theorem 2 to be applied directly to $(S^{\mathrm{ad}}, \widehat{S}^{\mathrm{ad}})$ at confidence $\delta/2$. On the resulting calibration event,

$$\rho_m(\delta/2) \le \frac{\sigma_{\mathrm{up}} \varepsilon_{\mathrm{NN}} + \eta_m(\delta/2) + s_m}{\sigma_{\mathrm{low}}}.$$

Taking $L^2(P_X)$ norms in the theorem's coverage and length bounds and using Minkowski's inequality gives, respectively, bounds by σ_{up} and L_{len} times

$$\left(1+\frac{\sigma_{\mathrm{up}}}{\sigma_{\mathrm{low}}}\right)\varepsilon_{\mathrm{NN}}+\frac{\eta_m(\delta/2)+s_m}{\sigma_{\mathrm{low}}}.$$

Here the oracle is $\mathcal{C}_{S^{\mathrm{aa}}}(x,0)=C_S^{\circ}(x)$ by Proposition 5. The intersection of training and calibration events has probability at least $1-\delta$. Finally, $s_m<2/m\le\sqrt{\log(4/\delta)/m}$ for $m\ge 4$. Both confidence terms are bounded by $\sqrt{\log(4/\delta)/(n\wedge m)}$, giving (43) after absorbing a factor of two into the constant. □

D.4 Compositional score quantiles

The quantile-estimation result of Madrid Padilla et al. (2022, Theorem 6) covers the composition classes of Schmidt-Hieber (2020), not only isotropic Hölder balls. We define these classes and give their consequences for global calibration.

Definition 23. *Fix a nonnegative integer J, positive integer dimensions $\mathbf{d}=(d_0,\dots,d_{J+1})$ with $d_0=d$ and $d_{J+1}=1$, integers $1\le t_j\le d_j$, smoothness indices $\beta_j>0$, and $K\ge 1$. The class $\mathcal{H}_{\mathrm{comp}}(J,\mathbf{d},\mathbf{t},\boldsymbol{\beta},K)$ consists of functions $g_J\circ\cdots\circ g_0$, where*

$$g_j:[a_j,b_j]^{d_j}\longrightarrow[a_{j+1},b_{j+1}]^{d_{j+1}},\qquad 0\le j\le J.$$

Here $[a_0,b_0]=[0,1]$ and $-K\le a_j<b_j\le K$ for $0\le j\le J+1$. Each scalar component of g_j depends on at most t_j input coordinates and, as a function of those coordinates, belongs to a Hölder ball of smoothness β_j and radius K. We use the convention of Section D.2: derivatives through order $\lceil\beta_j\rceil-1$ are bounded by K, and the highest-order derivatives are Hölder with exponent $\beta_j-\lceil\beta_j\rceil+1$ and constant K. The component functions and their coordinate subsets are not specified to the estimator.

The composition depth J is distinct from the network depth L. Define the effective smoothness indices and the estimation exponent by

$$\beta_j^*\coloneqq\beta_j\prod_{\ell=j+1}^{J}(\beta_\ell\wedge 1),\qquad \phi\coloneqq\min_{0\le j\le J}\frac{\beta_j^*}{2\beta_j^*+t_j}\in(0,1/2),\tag{44}$$

with an empty product equal to one. The factors from later layers account for propagation of approximation error through the composition.

Use the bounded sparse ReLU class $\mathcal{G}(L,\mathbf{p},s,F)$ described before Proposition 26, with $p_0=d$, $p_{L+1}=1$ and fixed $F\ge\max\{1,M,K\}$. For suitable fixed constants $A_L,A_p,A_s>0$ depending only on the composition parameters, set

$$L=\lceil A_L\log_2 n\rceil,\qquad p_j=\lceil A_p n^{1-2\phi}\rceil\ (1\le j\le L),\qquad s=\lceil A_s n^{1-2\phi}\log n\rceil.\tag{45}$$

The constants are chosen large enough for the composition approximation, using the corrected depth requirement of Schmidt-Hieber and Vu (2024); in particular,

$$A_L>\sum_{j=0}^{J}\frac{\beta_j+t_j}{2\beta_j^*+t_j}\log_2(4t_j\vee 4\beta_j).$$

The class is deterministic; fit a measurable global minimizer of the empirical pinball loss on $Z_i=S(X_i,Y_i)$ and truncate to $[-M,M]$, as in (15).

Proposition 6 (complete version). *Retain the conditions of Section D.2, except replace the isotropic Hölder-ball assumption by $q_S(\cdot)\in\mathcal{H}_{\mathrm{comp}}(J,\mathbf{d},\mathbf{t},\boldsymbol{\beta},K)$, still with $|q_S(x)|\le M$. Use (45) for quantile fitting and (13) for global calibration on an independent fold. There are constants $C_{\mathrm{comp}}<\infty$ and $n_{\mathrm{comp}}\in\mathbb{N}$ such that, for $n\ge n_{\mathrm{comp}}\vee 2$ and $\delta\in(0,1)$, putting*

$$\varepsilon_{\mathrm{comp}}(\delta)\coloneqq C_{\mathrm{comp}}\left\{n^{-\phi}(\log n)^{3/2}+\sqrt{\frac{\log(4/\delta)}{n}}\right\},$$

we have $\|\widehat{q}_S(\cdot) - q_S(\cdot)\|_{L^2(P_X)} \le \varepsilon_{\mathrm{comp}}(\delta)$ *with probability at least* $1 - \delta/2$ *over training, conditional on* $\mathcal{D}_0$*. If also* $m \ge 4$ *and*

$$\sigma_{\mathrm{up}}\varepsilon_{\mathrm{comp}}(\delta) + \eta_m(\delta/2) + s_m \le \sigma_{\mathrm{low}} r_0, \tag{46}$$

then (16) *holds with probability at least* $1 - \delta$ *over both folds, conditional on* $\mathcal{D}_0$*. The length conclusion requires* ***A*** *2*(S)*. The constants depend only on the composition and architecture parameters, on* $(\alpha, M, \sigma_{\mathrm{low}}, \sigma_{\mathrm{up}}, r_0, p_{\min}, p_{\max}, F)$*, and also on* L_{len} *for length. For fixed* δ*, the sample-size conditions hold for sufficiently large* n, m*.*

Proof. The corrected composition approximation of Schmidt-Hieber (2020); Schmidt-Hieber and Vu (2024) gives $\inf_{f\in\mathcal{G}} \|f - q_S(\cdot)\|_\infty \lesssim n^{-\phi}$. The sparse-network covering bound (Schmidt-Hieber, 2020, Lemma 5) gives $\log \mathcal{N}(n^{-1}, \mathcal{G}, \|\cdot\|_\infty) \lesssim n^{1-2\phi}(\log n)^3$ for (45). Here $\mathcal{N}$ is the sup-norm covering number. For bounded network outputs, (42) makes the expected pinball-loss difference comparable to $\|f - q_S(\cdot)\|^2_{L^2(P_X)}$, with lower comparison constant $(\sigma_{\mathrm{low}}/2)\min\{1, r_0/(F+M)\}$ and upper constant $\sigma_{\mathrm{up}}/2$; its variance is at most this squared norm. Bernstein's inequality on a n^{-1}-net (an external n^{-1}-cover, as in Schmidt-Hieber, 2020, Lemma 5, yields an internal $2n^{-1}$-net, which only changes constants), the Lipschitz property of the loss and empirical minimality therefore give, for any deterministic class $\mathcal{G}$ of networks bounded by F, with training probability at least $1 - \delta/2$,

$$\|\widehat{q}_S(\cdot) - q_S(\cdot)\|^2_{L^2(P_X)} \lesssim \inf_{f\in\mathcal{G}} \|f - q_S(\cdot)\|^2_\infty + \frac{\log \mathcal{N}(n^{-1}, \mathcal{G}, \|\cdot\|_\infty) + \log(4/\delta)}{n}, \tag{47}$$

with implicit constant depending only on $(\sigma_{\mathrm{low}}, \sigma_{\mathrm{up}}, r_0, F, M)$. For (45), the right-hand side is $\lesssim n^{-2\phi}(\log n)^3 + \log(4/\delta)/n$. Loss differences are bounded even for unbounded Z; truncation cannot increase the error. This gives the finite-confidence form of the quantile-estimation rate in Madrid Padilla et al. (2022, Theorem 6), whose displayed statement is asymptotic in probability. On this event, $\|\Delta_{S^{\mathrm{ad}}}\|_{L^1(P_X)} \le \|\Delta_{S^{\mathrm{ad}}}\|_{L^2(P_X)} \le \varepsilon_{\mathrm{comp}}(\delta)$ by (14). Condition (46) permits the same application of Theorem 2 at confidence $\delta/2$ as in the proof of Proposition 22; the union bound and the same norm inequalities give (16). □

Corollary 24 (Additive and single-index score quantiles)**.** *Retain the distributional conditions and the bound* $|q_S(x)| \le M$ *above, and let* $d \le n$*. Fix* $\beta > 0$ *and* $R \ge 1$*, and suppose either*

$$q_S(x) = \sum_{j=1}^{d} f_j(x_j), \qquad f_j \in \mathcal{H}^\beta([0,1], R),$$

or

$$q_S(x) = g(\theta^\top x), \qquad \|\theta\|_1 \le R, \qquad g \in \mathcal{H}^\beta([-R, R], R).$$

Put

$$r_{\mathrm{add}} \coloneqq d^{(\beta+1)/(2\beta+1)}\, n^{-\beta/(2\beta+1)} (\log n)^{3/2}, \qquad r_{\mathrm{si}} \coloneqq n^{-\beta/(2\beta+1)} (\log n)^{3/2} + \sqrt{\frac{d(\log n)^3}{n}}.$$

There are deterministic sparse ReLU classes, depending only on (β, R, M, d, n) *and with outputs bounded by* $F = M + 1$*, and a threshold* $n_0 \ge 3$ *growing polynomially in* d*, such that for* $n \ge n_0$ *the fitted quantile satisfies, with training probability at least* $1 - \delta/2$*,*

$$\|\widehat{q}_S(\cdot) - q_S(\cdot)\|_{L^2(P_X)} \lesssim r + \sqrt{\frac{\log(4/\delta)}{n}},$$

with $r = r_{\mathrm{add}}$ *for additive and* $r = r_{\mathrm{si}}$ *for single-index maps. Global calibration then gives the conclusions of Proposition 6 with* $n^{-\phi}(\log n)^{3/2}$ *replaced by* r*. The implicit constants depend only on* $(\beta, R, \alpha, M, \sigma_{\mathrm{low}}, \sigma_{\mathrm{up}}, r_0)$*, polynomially on* M*, and also on* L_{len} *for length; for fixed* M *they do not depend on* d*. In the additive model* M *may have to be as large as* dR*, so the dependence on* d *is polynomial in all cases. The functions* f_j*,* g *and the index* θ *are unknown to the estimator.*

Proof. Embedding the two models in Definition 23 treats the sum or the index $\theta^\top x$ as a d-variate function to be approximated, which makes the constants exponential in d. Both maps are instead computed exactly by linear layers, so only univariate functions need approximating; we then apply (47). Networks are those of Schmidt-Hieber (2020), $f(x) = W_L \sigma_{v_L} W_{L-1} \cdots W_1 \sigma_{v_1} W_0 x$ with $\sigma_v(y) = \sigma(y - v)$ and parameters bounded by one in absolute value; a network is embedded in a deeper or wider class by identity layers on its nonnegative hidden units and zero weights.

Univariate approximation. Let h be β-Hölder on $[0, 1]$ with norm at most K in the convention of Schmidt-Hieber (2020). By their Theorem 5 with input dimension one, for every $\eta \in (0, 1]$ there is a network $\widetilde{h}$ with depth at most $c \log(2/\eta)$, hidden widths at most $c\eta^{-1/\beta}$, at most $c\eta^{-1/\beta} \log(2/\eta)$ nonzero parameters and $\|\widetilde{h} - h\|_\infty \le \eta$, where c depends only on (β, K).

Additive maps. Each f_j has norm at most $(\lceil \beta \rceil + 1)R$ in that convention. Approximate every f_j with $\eta = \varepsilon/d$ and a common depth, and place the d networks in parallel: the first layer feeds x_j to the jth block, the hidden layers are block diagonal, and the output layer adds the block outputs with their own output weights. Parameters remain bounded by one and the sum is exact, so the network has sup-norm error at most ε, depth $\lesssim \log(2d/\varepsilon)$, hidden widths $\lesssim d(d/\varepsilon)^{1/\beta}$ and $\lesssim d(d/\varepsilon)^{1/\beta} \log(2d/\varepsilon)$ nonzero parameters.

Single-index maps. Since $|\theta^\top x| \le \|\theta\|_1 \le R$ on $[0, 1]^d$, one hidden unit with weights $\theta_j/(2R)$ and shift $v = -1/2$ computes $\sigma(\theta^\top x/(2R) + 1/2) = u \in [0, 1]$ exactly. The map $u \mapsto g(R(2u-1))$ is β-Hölder on $[0, 1]$ with norm depending only on (β, R). Approximating it with $\eta = \varepsilon$ after this unit gives error at most ε, since the index is computed without error, with depth $\lesssim \log(2/\varepsilon)$, widths $\lesssim d + \varepsilon^{-1/\beta}$ (the input has d coordinates) and $\lesssim d + \varepsilon^{-1/\beta} \log(2/\varepsilon)$ nonzero parameters.

Rates. In both cases the approximating network is bounded by $M + \varepsilon \le F$ for $\varepsilon \le 1$. By Schmidt-Hieber (2020, Lemma 5), $\log \mathcal{N}(n^{-1}, \mathcal{G}, \|\cdot\|_\infty) \le (s + 1) \log\{2n(L + 1)V^2\}$, with s nonzero parameters, depth L and $V = \prod_{\ell=0}^{L+1}(p_\ell + 1)$. For $d \le n$ and $n^{-1} \le \varepsilon \le 1$, every logarithm above is $\lesssim \log n$ and $\log V \lesssim (\log n)^2$, so the entropy is $\lesssim d^{1+1/\beta}\varepsilon^{-1/\beta}(\log n)^3$ for additive maps and $\lesssim (d + \varepsilon^{-1/\beta})(\log n)^3$ for single-index maps. The choices $\varepsilon = d^{(\beta+1)/(2\beta+1)}\{(\log n)^3/n\}^{\beta/(2\beta+1)}$ and $\varepsilon = \{(\log n)^3/n\}^{\beta/(2\beta+1)}$ balance ε^2 against the entropy term of (47); since $(\log n)^{3\beta/(2\beta+1)} \le (\log n)^{3/2}$, this gives the stated bounds, whose constants involve d only through r. Both choices lie in $[n^{-1}, 1]$ once n exceeds a power of d; the sample-size condition (46), with $\varepsilon_{\text{comp}}$ replaced by the new bound, holds once $n \ge C_\delta d^a$ and $m \ge C_\delta$ for some $a > 0$ and C_δ depending on δ and the constants above. The calibration step is that of Proposition 6. The constants of (47) grow polynomially with M through $F = M + 1$ and the curvature constant; neither $p_{\min}$ nor $p_{\max}$ enters. □

These results use the known quantile-estimation theory; the calibration step transfers its structured rates to coverage and oracle-relative length. The single-index constants are free of d because of the normalization $\|\theta\|_1 \le R$; with $\|\theta\|_2 = 1$ instead, the index ranges over $[-\sqrt{d}, \sqrt{d}]$ and they become polynomial in d. The architecture is tuned to the class parameters, not to the unknown component functions, coordinate subsets or index; no adaptation to unknown smoothness is asserted. The comparison is with isotropic localization in the original covariate space. It does not cover learned distances or forest-based calibration weights (Lin et al., 2021; Amoukou and Brunel, 2023), or provide a lower bound for RLCP.

D.5 Complete statement and proof of Proposition 7

Proposition 7 (complete version). *Let the conditions of Section D.2 hold and fix $\delta \in (0, 1)$. Let $\mathcal{K}$ satisfy **K** 1 and be $L_{\mathcal{K}}$-Lipschitz on $\mathbb{R}^d$. Split $N = n + m$, fit (17) with degree $s_q = \lceil \beta_q \rceil - 1$ and deterministic bandwidth*

$$b = \left(\frac{\log(16n/\delta)}{n} \right)^{1/(2\beta_q + d)},$$

and calibrate $\widehat{S}^{\mathrm{ad}}$ *globally on the independent fold* $\mathcal{D}_m^{\mathrm{cal}}$*. Let* $C_{\mathrm{LP}}, n_{\mathrm{LP}}$ *be the constants in Proposition 25, and put*

$$\varepsilon_{\mathrm{LP}} \coloneqq 2C_{\mathrm{LP}} \left(\frac{\log(16n/\delta)}{n} \right)^{\beta_q/(2\beta_q+d)}.$$

If $n \geq n_{\mathrm{LP}} \vee 2$*,* $m \geq 4$*, and*

$$\sigma_{\mathrm{up}}\varepsilon_{\mathrm{LP}} + \eta_m(\delta/2) + s_m \leq \sigma_{\mathrm{low}} r_0,$$

then (18) *holds with probability at least* $1-\delta$ *over both folds, simultaneously for every* $x \in \mathcal{X}$*. The constants depend only on* $(\alpha, M, \sigma_{\mathrm{low}}, \sigma_{\mathrm{up}}, r_0, \beta_q, L_q, d, p_{\min}, p_{\max}, \|\mathcal{K}\|_\infty, \kappa, \eta, L_{\mathcal{K}})$*, also on* L_{len} *for length. For each fixed* δ*, the sample-size conditions hold when* n, m *are sufficiently large.*

Proof. Proposition 25 bounds the error of (17) for the conditional $(1-\alpha)$-quantile of $Z = S(X,Y)$ under the rate conditions above. Its polynomial degree is $s_q = \lceil \beta_q \rceil - 1$, including when β_q is an integer. At confidence $\delta/2$, with the stated bandwidth, there is a training event $\mathcal{T}$ of probability at least $1-\delta/2$ on which

$$\sup_{x \in \mathcal{X}} \Delta_{S^{\mathrm{ad}}}(x) \leq \varepsilon_{\mathrm{LP}}.$$

Truncation cannot increase the error because $|q_S(x)| \leq M$ on $\mathcal{X}$. For each training realization in $\mathcal{T}$, $\|\Delta_{S^{\mathrm{ad}}}\|_{L^1(P_X)} \leq \varepsilon_{\mathrm{LP}}$; the assumed sample-size condition therefore implies (5) for S^{ad} at confidence $\delta/2$. Apply the complete version of Theorem 2 directly to $(S^{\mathrm{ad}}, \widehat{S}^{\mathrm{ad}})$. It gives an event of conditional calibration probability at least $1-\delta/2$, on which its radius (21), evaluated at this reference score, satisfies

$$\Delta_{S^{\mathrm{ad}}}(x) + \rho_m(\delta/2) \leq \left(1 + \frac{\sigma_{\mathrm{up}}}{\sigma_{\mathrm{low}}}\right) \varepsilon_{\mathrm{LP}} + \frac{\eta_m(\delta/2) + s_m}{\sigma_{\mathrm{low}}} \quad \text{for every } x \in \mathcal{X}.$$

The intersection with $\mathcal{T}$ has probability at least $1-\delta$ over both folds. The theorem bounds coverage error by σ_{up} times this display and length error by L_{len} times it; use $\mathcal{C}_{S^{\mathrm{ad}}}(x,0) = C_S^\circ(x)$. Since $s_m < 2/m \leq \sqrt{\log(4/\delta)/m}$ for $m \geq 4$, these bounds prove (18). The claim for sufficiently large samples follows because $\varepsilon_{\mathrm{LP}}$ and the calibration terms tend to zero. □

D.6 Complete statement and proof of Proposition 8

Proposition 8 (complete version). *Let the conditions of Section D.2 and* ***K*** *1 hold, fix* $\delta \in (0,1)$*, use all* N *observations for calibration, and put*

$$\bar{\beta}_q \coloneqq \min(\beta_q, 1), \qquad L_{\mathrm{loc}} \coloneqq \begin{cases} L_q, & \beta_q \leq 1, \\ \sqrt{d}\, L_q, & \beta_q > 1, \end{cases} \qquad h_\star = \left(\frac{\log(4/\delta)}{N p_{\min}} \right)^{1/(2\bar{\beta}_q + d)}.$$

Fix a centre $\tilde{x}$ *satisfying* ***P*** *1 at* $(\tilde{x}, h_\star)$*. Let* c_{R} *be* (35) *evaluated at exponent* $\bar{\beta}_q$ *and score constants* $(\sigma_{\mathrm{up}}, \sigma_{\mathrm{low}}, r_-, r_+) = (\sigma_{\mathrm{up}}, \sigma_{\mathrm{low}}, r_0, r_0)$*. If* $N p_{\min} \geq c_{\mathrm{R}} \log(4/\delta)$*, then, conditionally on this realized centre,* (19) *holds with probability at least* $1-\delta$ *over calibration, simultaneously for every* $x \in B(\tilde{x}, h_\star) \cap \mathcal{X}$*. The rate constants depend only on* $(\sigma_{\mathrm{up}}, \sigma_{\mathrm{low}}, L_{\mathrm{loc}}, c_K)$*, also on* L_{len} *for length; the sample-size constant additionally depends on* $(\alpha, r_0, \bar{\beta}_q, d)$*.*

Proof. For $\beta_q \leq 1$, the Hölder ball directly gives $|q_S(u) - q_S(u')| \leq L_q \|u - u'\|^{\beta_q}$. For $\beta_q > 1$, each first partial derivative is bounded by L_q; convexity of $[0,1]^d$ and integration along the segment from u to u' give $|q_S(u) - q_S(u')| \leq \sqrt{d}\, L_q \|u - u'\|$. Hence $\mathbf{A}\,4(S, 1-\alpha, \bar{\beta}_q)$ holds with constant L_{loc}. Apply Corollary 18 with $m = N$, exponent $\bar{\beta}_q$, this Hölder constant, and the score regularity just verified. Its bandwidth is precisely the stated $h_\star$, its sample-size condition is the stated one with c_{R} evaluated at $\bar{\beta}_q$, and its conclusion is (19). The event is for the fixed centre and is simultaneous only within its neighbourhood. □

D.7 Verifying the density window for a residual score

The conditions of Section D.2 allow a score support that moves with x. For $S(x,y) = |y - m(x)|/a(x)$ with $a(x) > 0$, a conditional response density $f_{Y|x}$ induces, at $z > 0$,

$$g_x(z) = a(x)\{f_{Y|x}(m(x) + a(x)z) + f_{Y|x}(m(x) - a(x)z)\}.$$

Suppose that $g_x(z) \le \sigma_{\mathrm{up}}$ for all x and $z > 0$, and that $g_x(z) \ge \sigma_{\mathrm{low}}$ whenever $z > 0$ and $|z - q_S(x)| \le r_g$, for a uniform radius $r_g > 0$. Since $F^{(S|X)}(0 \mid x) = 0$, the upper bound gives $1 - \alpha = F^{(S|X)}(q_S(x) \mid x) \le \sigma_{\mathrm{up}} q_S(x)$, so any $0 < r_0 \le \min\{r_g, (1-\alpha)/(2\sigma_{\mathrm{up}})\}$ keeps the growth window strictly above zero and verifies (42), even though the score support changes with x. Also $|\mathcal{C}_S(x,t)| = 2a(x)t_+$, so $L_{\mathrm{len}} = 2\sup_x a(x)$ when this supremum is finite; the unstandardized residual has $L_{\mathrm{len}} = 2$. Smoothness of $q_S(\cdot)$ must still be checked for the frozen centre and scale; it is not automatic from their having been fitted.

E Quantile-Estimation Tools

E.1 Quantile-estimation assumptions and statements

Fix $\alpha \in (0,1)$ and condition on the initial sample $\mathcal{D}_0$ that fixes S. The training responses are $Z_i = S(X_i, Y_i)$, and the target is the single conditional score quantile $q_S^{1-\alpha}(x)$. This appendix records the two estimation bounds behind Propositions 7 and 22: a uniform local-polynomial bound, proved in Section E.2, and a neural-network $L^2(P_X)$ bound, Proposition 26, extracted from its source at finite confidence.

Use the conditions of Section D.2: $\mathcal{X} = [0,1]^d$, design density bounded between $p_{\min}$ and $p_{\max}$, and $q_S^{1-\alpha}(\cdot) \in \mathcal{H}^{\beta_q}(\mathcal{X}, L_q)$ with $\beta_q > 0$ and $|q_S^{1-\alpha}(x)| \le M$. The conditional CDF of Z satisfies the upper Lipschitz and local lower increment bounds (42), with constants $(\sigma_{\mathrm{up}}, \sigma_{\mathrm{low}}, r_0)$. The bound M applies to the *quantile*, not to Z: neither bounded score support nor a score moment is assumed. Length regularity is not needed for quantile estimation.

Set $\tau = 1 - \alpha$ and $\rho_\tau(u) = u(\tau - \mathbb{1}_{\{u<0\}})$. For degree $s = \lceil \beta_q \rceil - 1$, use the local-polynomial fit (17) on (X_i, Z_i) with deterministic bandwidth $b = b_n > 0$ and its truncated intercept

$$\widehat{q}_S^{1-\alpha}(x) = T_M\big(\widehat{\theta}_0(x)\big), \qquad T_M(z) = \max\{-M, \min\{M, z\}\}.$$

The bandwidth may depend on (n, δ) and the model constants, but not on the observations. At every x, the empirical pinball criterion is finite, nonnegative and polyhedral in the polynomial coefficients, hence attains its minimum. It is continuous in those coefficients and Borel in $(x, \mathcal{D}_n^{\mathrm{tr}})$; fix a jointly Borel-measurable selection $\widehat{\theta}(x)$ of minimizers (Brown and Purves, 1973). The bound below holds for every such selection.

Proposition 25 (Uniform local-polynomial score-quantile bound)**.** *Under the conditions above, let $\mathcal{K}$ satisfy **K** 1 and be $L_{\mathcal{K}}$-Lipschitz on $\mathbb{R}^d$. There are $C_{\mathrm{LP}} < \infty$ and $n_{\mathrm{LP}} \in \mathbb{N}$, depending only on*

$$(M, \beta_q, L_q, d, p_{\min}, p_{\max}, \|\mathcal{K}\|_\infty, \kappa, \eta, L_{\mathcal{K}}, \sigma_{\mathrm{up}}, \sigma_{\mathrm{low}}, r_0),$$

such that for every $n \ge n_{\mathrm{LP}}$, $\delta \in (0,1)$ and deterministic $b_n > 0$, there is a training event $\mathcal{E}^{\mathrm{LP}}_{n,b_n,\delta}$ with $\mathbb{P}(\mathcal{E}^{\mathrm{LP}}_{n,b_n,\delta} \mid \mathcal{D}_0) \ge 1 - \delta$ on which

$$\sup_{x \in \mathcal{X}} \big|\widehat{q}_S^{1-\alpha}(x) - q_S^{1-\alpha}(x)\big| \le C_{\mathrm{LP}} \left\{ b_n^{\beta_q} + \sqrt{\frac{\log(8n/\delta)}{n b_n^d}} \right\}, \tag{48}$$

simultaneously for every choice of minimizers in (17)*.*

The Lipschitz requirement excludes the box kernel from this estimation result, but not from the RLCP results. The degree convention covers every fixed $\beta_q > 0$, including integer smoothness.

We claim no novelty for this quantile estimator or its rate. The proof in Section E.2 records the finite-sample, boundary-inclusive form needed here; compare the classical pointwise

Bahadur representation of Chaudhuri (1991b) and the uniform representation of Guerre and Sabbah (2012), stated on compact inner subsets of the support.

The second estimator is the sparse ReLU quantile regression (15). Use the class $\mathcal{G}(L, \mathbf{p}, s, F)$ of Madrid Padilla et al. (2022, Definition 3 and Theorem 6), specialized to one d-variate β_q-Hölder component. Its weights and biases have absolute value at most one, at most s parameters are nonzero, and every network is uniformly bounded by F. Choose, up to integer rounding,

$$L \asymp \log n, \qquad p_j \asymp n^{d/(2\beta_q+d)} \quad \text{(each hidden layer)}, \qquad s \asymp n^{d/(2\beta_q+d)} \log n,$$

with the fixed multiplicative constants prescribed by that theorem, in particular a sufficiently large depth constant. Take a fixed $F \geq \max\{1, M, K\}$, where K bounds the Hölder norm in the convention of that source; K depends only on (β_q, d, L_q). The class is fixed independently of the observations. Choose a measurable global empirical minimizer and truncate its output to $[-M, M]$ as in (15). This specifies a statistical estimator; no guarantee for an arbitrary numerical optimizer is assumed.

Proposition 26 (Neural score-quantile bound in $L^2(P_X)$)**.** *Under the conditions above, use the architecture just described on the n training pairs (X_i, Z_i). There are $C_{\mathrm{NN}} < \infty$ and $n_{\mathrm{NN}} \in \mathbb{N}$, depending only on $(M, \beta_q, L_q, d, p_{\min}, p_{\max}, \sigma_{\mathrm{up}}, \sigma_{\mathrm{low}}, r_0, F)$ and the fixed architecture constants, such that for every $n \geq n_{\mathrm{NN}}$ and $\delta \in (0,1)$, there is a training event $\mathcal{T}$ with $\mathbb{P}(\mathcal{T} \mid \mathcal{D}_0) \geq 1 - \delta$ on which*

$$\|\widehat{q}_S^{1-\alpha}(\cdot) - q_S^{1-\alpha}(\cdot)\|_{L^2(P_X)} \leq C_{\mathrm{NN}} \left\{ n^{-\beta_q/(2\beta_q+d)} (\log n)^{3/2} + \sqrt{\frac{\log(2/\delta)}{n}} \right\}. \tag{49}$$

This is an $L^2(P_X)$ bound, not a uniform one.

Reduction to the cited result. Apply Madrid Padilla et al. (2022, Theorem 6) to $Z = S(X, Y)$ at $\tau = 1 - \alpha$. Condition (42) supplies its local growth and bounded-density assumptions; dependence on τ is through $(\sigma_{\mathrm{up}}, \sigma_{\mathrm{low}}, r_0)$. The approximation covers every $\beta_q > 0$ (Schmidt-Hieber, 2020, Theorem 5), and no score moment is needed since $|\rho_\tau(Z-f) - \rho_\tau(Z-g)| \leq |f-g|$. For finite confidence, use the peeling argument of their Appendix C.3 with the bounds in Appendix D (display (31) and Lemmas 23–25): there are $O(\log n)$ shell events, failure at most $Ce^{-\gamma}$ per shell, and additional terms $O(rF^2\sqrt{\gamma/n} + F^2\gamma/n)$ at radius r. Take $\gamma = \log(C \log n/\delta)$ and absorb the mixed term into the squared radius and γ/n. After taking square roots, $\sqrt{\log\log n/n}$ is absorbed by the leading rate in (49), since $\beta_q/(2\beta_q + d) < 1/2$, above a threshold independent of δ. Truncation to $[-M, M]$ cannot increase the error. □

E.2 Proof of Proposition 25

Proof. Condition on $\mathcal{D}_0$ and write $b = b_n$,

$$\lambda = \sqrt{\frac{\log(8n/\delta)}{nb^d}}, \qquad a = b^{\beta_q} + \lambda.$$

Constants $c, C > 0$ may change between occurrences and depend only on the parameters listed in the proposition. By truncation the error is at most $2M$, so it suffices to treat $a \leq a_0$ for a sufficiently small fixed $a_0 \leq 1$. This implies $n^{-1/d} \leq b \leq 1$ and $\lambda \leq 1$.

Let $s = \lceil \beta_q \rceil - 1$, $\gamma = \beta_q - s \in (0, 1]$, and $\psi(u) = (u^j)_{|j| \leq s} \in \mathbb{R}^P$. The scaled Taylor coefficients

$$\theta^\star(x) = \left(\frac{b^{|j|} \partial^j q_S^{1-\alpha}(x)}{j!} \right)_{|j| \leq s}$$

are uniformly bounded and γ-Hölder in x, with constants independent of $b \leq 1$. Taylor's formula gives, uniformly for $\|X - x\| \leq b$,

$$\left| \theta^\star(x)^\top \psi((X-x)/b) - q_S^{1-\alpha}(X) \right| \leq C b^{\beta_q}.$$

The local Gram matrices satisfy

$$G_x = b^{-d} \mathbb{E}\left[\mathcal{K}((X-x)/b) \psi((X-x)/b) \psi((X-x)/b)^\top \right] \succeq c I_P \quad (x \in \mathcal{X}).$$

Indeed, $(\mathcal{X}-x)/b \cap B(0,\eta)$ contains an inward reflection of $[0,\eta/(2\sqrt{d})]^d$. On that cube $\mathcal{K} \geq \kappa$ and $p_X \geq p_{\min}$; its polynomial Gram matrix is positive definite, and coordinate reflections preserve its eigenvalues.

For $u_x = (X-x)/b$, define the centered loss

$$\ell_{x,v}(X,Z) = \mathcal{K}(u_x)\left[\rho_\tau\big(Z-(\theta^\star(x)+v)^\top\psi(u_x)\big)-\rho_\tau\big(Z-\theta^\star(x)^\top\psi(u_x)\big)\right],$$
$$\widehat{M}_x(v) = \frac{1}{nb^d}\sum_{i=1}^n \ell_{x,v}(X_i,Z_i),$$

and $M_x(v) = \mathbb{E}[\widehat{M}_x(v)]$. The loss differences are bounded by $|v^\top\psi(u_x)|$ on the kernel support, so no score moment is needed. Integrating the conditional CDF increments in (42), using the Taylor remainder and $G_x \succeq cI_P$, yields

$$M_x(v) \geq c\|v\|^2 - Cb^{\beta_q}\|v\| \tag{50}$$

uniformly in x, provided $Cb^{\beta_q} \leq r_0/2$ and $\|v\| \leq r_c \coloneqq r_0/(2\sup_{\|u\|\leq 1}\|\psi(u)\|)$.

Fix a deterministic $r \in (0, r_c]$. For $\|v\| = r$, $|b^{-d}\ell_{x,v}| \leq Crb^{-d}$ and $\mathbb{E}[(b^{-d}\ell_{x,v})^2] \leq Cr^2b^{-d}$. Apply Bernstein's inequality on an ε-net of $\mathcal{X}\times\{\|v\|=r\}$, where $\varepsilon = n^{-(d+2)/\gamma}$; its log-cardinality is at most $C\log n$. The Lipschitz properties of $\mathcal{K}$ and ρ_τ, and the γ-Hölder modulus of $\theta^\star$, give, whenever $\|x-x'\| \vee \|v-v'\| \leq \varepsilon$,

$$b^{-d}|\ell_{x,v}(X,Z) - \ell_{x',v'}(X,Z)| \leq Cb^{-d}(\varepsilon^\gamma + \varepsilon/b) \leq C/n.$$

Here $\varepsilon \leq b$; if either kernel weight is nonzero, both polynomial arguments lie in $B(0,2)$. The same bound holds after averaging or taking expectations. A union bound and this transfer therefore give a measurable event of probability at least $1-\delta$ on which

$$\sup_{x\in\mathcal{X},\,\|v\|=r} |\widehat{M}_x(v) - M_x(v)| \leq Cr(\lambda+\lambda^2) + C/n \leq Cr\lambda + C/n. \tag{51}$$

Choose $r = Aa$ with A sufficiently large, then a_0 sufficiently small that the curvature conditions hold. On the event (51), (50) implies

$$\inf_{x\in\mathcal{X},\,\|v\|=r} \widehat{M}_x(v) \geq cr^2 - Car - C/n > 0$$

for all sufficiently large n, since $r^2 \geq A^2\log(8n)/n$. Convexity and $\widehat{M}_x(0) = 0$ then place every minimizer within distance r of $\theta^\star(x)$, simultaneously in x. Its intercept error is at most r, and truncation cannot increase it. Enlarging C_{LP} to cover $a > a_0$ proves (48). □

F Experimental details

F.1 Evaluation and implementation conventions

All experiments use $\alpha = 0.1$ and condition, within a replication, on the data used for fitting, tuning and calibration and on any algorithmic randomness. For the prediction set $C(x;\tilde{x})$, let

$$e(x,\tilde{x}) \coloneqq \mathrm{Cov}_{C(\cdot;\tilde{x})}(x) - (1-\alpha).$$

With $X \sim P_X$ fresh and, for RLCP, the single auxiliary centre $\tilde{X}$ the procedure draws for it, the synthetic risk is

$$R \coloneqq \big\{\mathbb{E}\big[e(X,\tilde{X})^2\big]\big\}^{1/2},$$

the root mean square error of the set actually returned, with the single centre the procedure draws for each covariate: it charges RLCP for the variability of its centre draw as well as for its bias, and it is the root mean square coverage error of Section 5. It is estimated with one realized centre per evaluation covariate in every experiment; global calibration has no centre. The centre-averaged risk appears in Section F.7 only.

Evaluation. Synthetic coverage is computed exactly as $F_x(\text{upper}) - F_x(\text{lower})$ from the known conditional CDF. Mean risks average replication-level root mean squares. Minima over evaluation points are finite-design diagnostics, not uniform guarantees.

Calibration. Global calibration uses the $\lceil (m+1)(1-\alpha) \rceil$-th order statistic of the m calibration scores, $+\infty$ if the rank exceeds m. RLCP uses the box kernel $\mathcal{K}(v) = \mathbb{1}_{\{\|v\| \le 1\}} / |B(0,1)|$, which satisfies **K** 1 with $\kappa = \|\mathcal{K}\|_\infty$ and $\eta = 1$, so that $c_K = c_{\text{reg}} |B(0,1)|$ and the centre $\tilde{x} = x + hV$ of Section 3 has V uniform on the unit ball, not clipped to the support, the same centre for calibration and test weights, the test mass at $+\infty$, and $m_{\text{eff}} = (\sum_i K_i + K_{\text{test}}) / \|K\|_\infty$; an unattainable weighted level gives an infinite set. Replacing the box kernel by the Epanechnikov kernel $K(v) = (1 - \|v\|^2)_+$ on the cells of Section F.5 moves R by at most 0.003 on average (rule, best and selected bandwidths; 95% intervals within ± 0.004) and changes no ordering with the score route. Infinite sets have coverage one (α^2 in the squared error), empty sets coverage zero; both stay in coverage summaries, while length summaries condition on finiteness and report the infinite share.

Folds, budgets and uncertainty. Within a replication all procedures receive the same N observations, and each splits them between fitting, bandwidth selection and calibration as stated in its description; the parts always sum to N. Budgets are independently seeded unless nested prefixes are declared; intervals are 95% Student t-intervals over replications or splits unless stated otherwise.

Quantile fitting and bandwidth selection. The network route fits the score responses $Z_i = S(X_i, Y_i)$ at level 0.9 with early stopping, then calibrates $S - \widehat{q}_S^{0.9}(\cdot)$ on a disjoint fold; RLCP selects its bandwidth on a pilot/validation split. Both are specified in Section F.2 and reused in Sections F.3 and F.5; the allocation experiment of Section F.6 uses cross-validated local-polynomial fits, specified there.

F.2 Compositional score quantile in growing dimension

This subsection completes the synthetic experiment of Section 5.

Data-generating model. Each bivariate block $\frac{2}{3}(v + v' + vv'/2)$ of (20) has mean zero and variance $\frac{4}{9}(1 + 1 + \frac{1}{4}) = 1$, because v and v' are independent, centred and standardized; the factor $J^{-1/2}$ therefore keeps $\operatorname{Var} T(X) = 1$ at every d, while the law of $T(X)$, hence of $q^\star(X)$, still depends on d. For $S(x, y) = |y|$, $q_S^{1-\alpha}(x) = q^\star(x)$; since $1/2 < q^\star < 3/2$, the conditional score density is bounded above and away from zero near $q^\star(x)$ with constants independent of x and d, and the unbounded score is allowed by Section D.2. The quantile map is thus a composition of three layers: a quadratic of each coordinate pair, their normalized sum, and the smooth univariate map $t \mapsto 1 + \tanh(t)/2$. This is the compositional structure assumed in Proposition 6. Since $q^\star$ is infinitely differentiable, isotropic theory (Proposition 22) also allows exponents close to $1/2$, with constants that grow quickly with d; the experiment illustrates the gain from a smooth, structured map and does not separate composition from smoothness.

Procedures and folds. Twenty replications; within each, procedures share N observations and 2048 fresh uniform evaluation covariates, independent across replications and budgets, with one centre per point.

Neural quantile fitting. The first $\lfloor N/2 \rfloor$ observations are split 80:20 between training and early stopping; the rest calibrates $S - \widehat{q}_S^{0.9}(\cdot)$ globally. Two 64-unit ReLU layers and a linear output, pinball loss at 0.9, Adam (learning rate 10^{-3}, weight decay 10^{-4}, batch 256); input and score standardization and the output bound $M = 1.2 \max_i |Z_i^{\text{std}}|$ come from the training observations alone, and standardized outputs are clipped to $[-M, M]$. Training stops after 20 epochs without a validation improvement above 10^{-6}, between 20 and 200 epochs, and the selected weights are restored.

RLCP bandwidth selection. RLCP reserves the first $n_{\text{sel}} = \lfloor N/8 \rfloor$ observations, $n_{\text{p}} = \text{round}(2n_{\text{sel}}/3)$ for pilot calibration and the rest for validation. Candidates $h = c h_\star(n_{\text{p}})$, with $h_\star(m) = (\log 40/m)^{1/(2+d)}$ and $c \in \{1/4, 1/2, 1, 2, 4, 8\}$, share one unit-ball draw per validation point. For each, coverage indicators (infinite set = covered) are averaged over each point's $\min\{50, n_{\text{v}}\}$ nearest validation neighbours, itself included, n_{v} being the validation

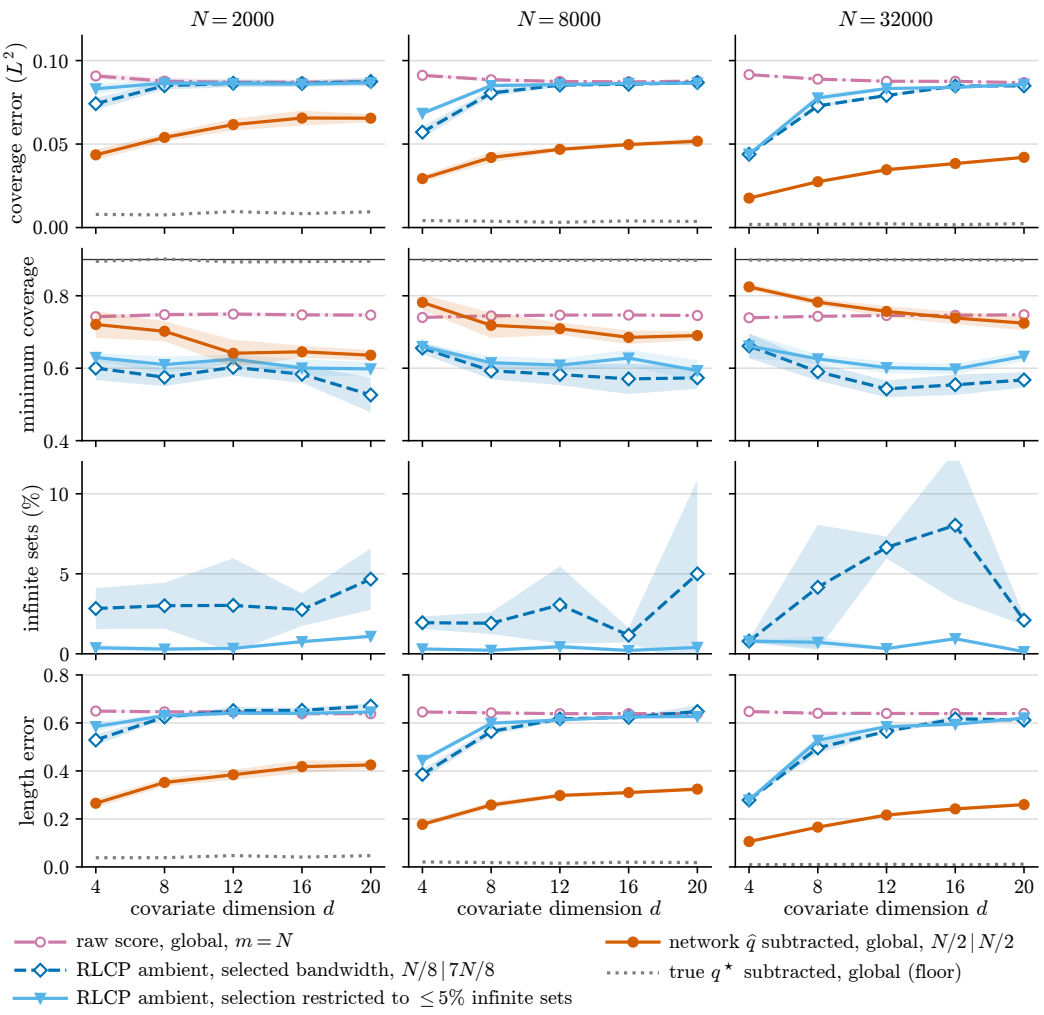

Figure 3: Figure 1 with the minimum evaluated coverage in place of its 5% quantile, the unrestricted selector added, and two more rows: the share of infinite sets, and the root mean square over finite sets of the returned length minus the oracle length $2q^\star(x)$. Means and 95% t-intervals over twenty replications; the length row excludes infinite sets.

size, and the RMS deviation of these averages from 0.9 is minimized among candidates with at most 5% infinite validation sets; if none qualifies the infinite fraction is minimized; ties favour the largest c. Calibration is then redone on the remaining $N - n_{\text{sel}}$ observations at $ch_\star(N - n_{\text{sel}})$, with fresh centre draws.

Bandwidth choice and infinite sets (Figure 3). At $c = 1$ on all N observations, the mean infinite share is 0.7–5.3% at $d = 4$, 28–61% at $d = 8$, 95–99.7% at $d = 12$ and 100% at $d = 16, 20$, where $R = \alpha$; hence the body plots a selected multiple. The restricted selector returns 0.1–1.1% infinite sets on average, against 0.8–8.0% unrestricted (up to 58% in one replication); the restriction moves mean R by less than 0.012, raises the minimum coverage by up to 0.07, and changes no ordering. The network route has lower mean R than either selector in all fifteen configurations.

Length error. The network route has the lower mean RMS length error in all fifteen configurations: at $N = 32000$, 0.11–0.26 across d against 0.28–0.62 for RLCP, with the raw score at 0.64–0.65 and the oracle below 0.05.

F.3 Real data read on groups

This subsection completes the real-data experiment of Section 5.

Data and folds. Eight datasets of the conformalized quantile regression benchmark (Romano et al., 2019), one per dimension: concrete ($d = 8$; Yeh, 1998), bio (9; Rana, 2013), bike (18; Fanaee-T and Gama, 2014), star (39; Achilles et al., 2008), facebook variant 1 (53; Singh et al., 2015; Singh, 2016), community (100; Redmond, 2002), MEPS panel 19 (139; Agency for Healthcare Research and Quality, 2017) and blog (280; Buza, 2014); rows are subsampled to 32000 where larger, non-finite rows dropped, and the count-like responses (bike, blog, facebook, MEPS) transformed by $y \mapsto \log(1 + \max\{y, 0\})$ for every procedure. Each of twenty splits permutes the rows into folds A (20%), B (20%), C (40%) and test (20%); $\widehat{\mu}$ is fitted on A by the architecture above (200 epochs, squared loss) and covariates are standardized on A alone, so $\widehat{\mu}$, the score and the RLCP metric are fixed before any of the N labels of $B \cup C$ is seen. Budgets: $N = 19200$ on bio, facebook and blog, 9471 on MEPS, 6531 on bike, 1296 on star, 1195 on community, 618 on concrete. The procedures follow Section F.2 with $c = 2^k$, $-2 \le k \le 9$; the selector's validation fold holds 26 points on concrete and 50 on community, where its criterion reduces to the marginal deviation, and at least 54 elsewhere.

Groups and floor. Groups are the cells of a k-means partition with $\min\{20, \max(4, \lfloor n_{\text{test}}/80 \rfloor)\}$ cells, fitted on the standardized covariates of A and shared by all procedures; test groups below 20 points are left out. Figure 2 reads the root mean square over groups of group coverage minus $1 - \alpha$ and the smallest group coverage, infinite sets counted as covered, one centre per test point. The floor is the expected reading under independent coverage with probability $1 - \alpha$ at the realized group sizes (2000 binomial draws), 0.022–0.039 and 0.834–0.860 across datasets: a sampling reference, not a lower bound. These group readings estimate no conditional coverage at x, and another partition could rank the procedures differently. Paired over the twenty splits, the network's group error is below that of selected RLCP by 0.010 ± 0.007 on bio and 0.020 ± 0.013 on concrete, whose test fold holds 206 points, and not on star ($+0.003 \pm 0.006$).

The restriction on infinite sets. Table 1 compares the unrestricted selector with the restriction to constants returning at most 5% infinite validation sets (else the smallest infinite share, then the largest constant). Unrestricted, 2.2–24.0% of the test sets are infinite; restricted, 0.1–1.3%, with median effective sizes of 42–97% of the calibration fold: broad neighbourhoods. Mean marginal coverage of the three procedures lies in 0.888–0.905 on every dataset.

Table 1: Ambient RLCP with the local-coverage selector, unrestricted and restricted to at most 5% infinite validation sets; means over twenty splits. Reading: root mean square over groups of group coverage minus 0.9, infinite sets counted as covered; m_{eff}: median effective local size; m: calibration size.

			unrestricted			restricted		
dataset	d	m	infinite	reading	m_{eff}	infinite	reading	m_{eff}
concrete	8	541	5.8%	0.076	385	0.1%	0.084	520
bio	9	16800	2.2%	0.040	4778	0.5%	0.039	7092
bike	18	5715	5.3%	0.051	752	0.2%	0.084	3990
star	39	1134	5.9%	0.036	729	0.2%	0.036	1084
facebook	53	16800	7.2%	0.050	2522	0.8%	0.057	10496
community	100	1046	6.1%	0.062	630	0.2%	0.068	1014
MEPS	139	8288	9.5%	0.049	1907	0.6%	0.079	6097
blog	280	16800	24.0%	0.059	1616	1.3%	0.092	12717

Excess-risk and worst-slab diagnostics. Table 2 adds two diagnostics computed on the test fold from the covariates and the coverage indicators $c_i = \mathbf{1}\{Y_i \in C(X_i)\}$. ERT (Braun et al., 2026) fits a classifier $\widehat{p}$ of c_i on X_i by five-fold cross-fitting (gradient boosting, 500 trees, learning rate 0.05, 31 leaves, at least 40 observations per leaf, fixed seed, probabilities not clipped) and reports the cross-fitted mean of $\operatorname{sign}\{\widehat{p}(X_i) - (1 - \alpha)\}\,\{c_i - (1 - \alpha)\}$, an estimate of $\mathbb{E}|\mathbb{P}\{Y \in C(X) \mid X\} - (1 - \alpha)|$ that is exact when $\widehat{p}$ has the right sign and

attenuated otherwise, so a value near zero is not exact conditional coverage. WSC is the smallest empirical coverage over slabs $\{a \le v^\top x \le b\}$ holding at least 10% of the test points, minimized over 1000 random unit directions v; searched and evaluated on the same points, it is pessimistic, more so on the small datasets, and it counts infinite sets as covered. Both, like the group readings, depend on the covariates alone and so average the RLCP centre draw inside the conditional coverage, whereas R squares the error of the returned set, centre included. The network route has the lowest ERT on all eight datasets, with paired differences excluding zero on seven (star excepted, where the raw score is already near the floor), the largest gaps on blog (0.033 against 0.073 raw and 0.065 RLCP), facebook and community; it has the highest WSC on seven (star excepted), and RLCP improves WSC over the raw score on seven (star excepted, 0.681 against 0.683), by less than the network.

Table 2: Real data, means over the splits: excess-risk diagnostic (ERT, L^1 miscoverage, lower is better), worst-slab coverage (WSC, closer to 0.9 is better), marginal coverage, mean finite half-width, share of infinite sets and median effective local size. ΔERT: network route minus the arm, with a 95% t-interval over the splits. Bold: lowest ERT and highest WSC per dataset.

dataset	procedure	ERT	WSC	marg.	half-w.	%inf	m_{eff}	ΔERT
concrete ($d = 8$)	raw score, global	0.054	0.463	0.891	11.85	0.0	–	-0.021 ± 0.012
	network route	**0.033**	**0.512**	0.892	11.39	0.0	–	–
	RLCP, selected c	0.055	0.464	0.888	11.80	0.1	520	-0.022 ± 0.011
bio ($d = 9$)	raw score, global	0.043	0.807	0.901	7.35	0.0	–	-0.011 ± 0.003
	network route	**0.032**	**0.833**	0.901	6.63	0.0	–	–
	RLCP, selected c	0.037	0.823	0.902	7.29	0.4	8143	-0.005 ± 0.002
bike ($d = 18$)	raw score, global	0.049	0.690	0.900	0.74	0.0	–	-0.015 ± 0.005
	network route	**0.035**	**0.798**	0.902	0.72	0.0	–	–
	RLCP, selected c	0.043	0.726	0.900	0.74	0.3	3355	-0.008 ± 0.005
star ($d = 39$)	raw score, global	0.013	**0.683**	0.902	512.96	0.0	–	-0.007 ± 0.009
	network route	**0.006**	0.670	0.895	511.48	0.0	–	–
	RLCP, selected c	0.011	0.681	0.901	511.24	0.2	1084	-0.006 ± 0.007
facebook ($d = 53$)	raw score, global	0.051	0.732	0.903	1.01	0.0	–	-0.028 ± 0.002
	network route	**0.023**	**0.836**	0.901	0.87	0.0	–	–
	RLCP, selected c	0.043	0.776	0.903	0.99	0.7	11878	-0.021 ± 0.003
community ($d = 100$)	raw score, global	0.032	0.565	0.898	0.27	0.0	–	-0.025 ± 0.009
	network route	**0.008**	**0.666**	0.903	0.25	0.0	–	–
	RLCP, selected c	0.026	0.579	0.900	0.27	0.1	1034	-0.018 ± 0.011
MEPS ($d = 139$)	raw score, global	0.025	0.726	0.902	2.05	0.0	–	-0.014 ± 0.004
	network route	**0.011**	**0.815**	0.902	1.96	0.0	–	–
	RLCP, selected c	0.021	0.755	0.903	2.03	0.7	5531	-0.010 ± 0.004
blog ($d = 280$)	raw score, global	0.073	0.714	0.901	1.48	0.0	–	-0.040 ± 0.003
	network route	**0.033**	**0.818**	0.900	1.30	0.0	–	–
	RLCP, selected c	0.065	0.756	0.902	1.45	1.0	13416	-0.032 ± 0.003

The bandwidth trap. Figure 4 sweeps c over the selection grid on the calibration fold and records the share of infinite test sets and the local mass m_{eff}/m. Call a multiple *local* if the median effective size is at most 10% of the fold and *usable* if at most 5% of the test sets are infinite. On this base-two grid a multiple that is both exists on bio only ($c = 4$); elsewhere the smallest usable multiple exceeds the largest local one by a factor 2 (bike, star, facebook, MEPS) or 4 (concrete, community, blog). A bandwidth between grid points, or other thresholds, could open a narrow window, and a local mass above 10% is not by itself global calibration; on this grid, however, no multiple is at once local and usable on seven datasets. Along the sweep the group error is smallest at c between 4 (concrete) and 128 (blog), where 0.7–32% of the test sets are infinite and the neighbourhood holds 2–26% of the fold; there RLCP reaches the network's level within 0.002 on bike, star, MEPS and blog and stays 0.004–0.012 above it elsewhere, part of these minima being abstention since infinite sets count as covered. The restricted selector settles to the right of the minimum on every dataset (mean multiple 35–474), 0.003 (star) to 0.046 (blog) of group error above it, and within 0.006 only on bio.

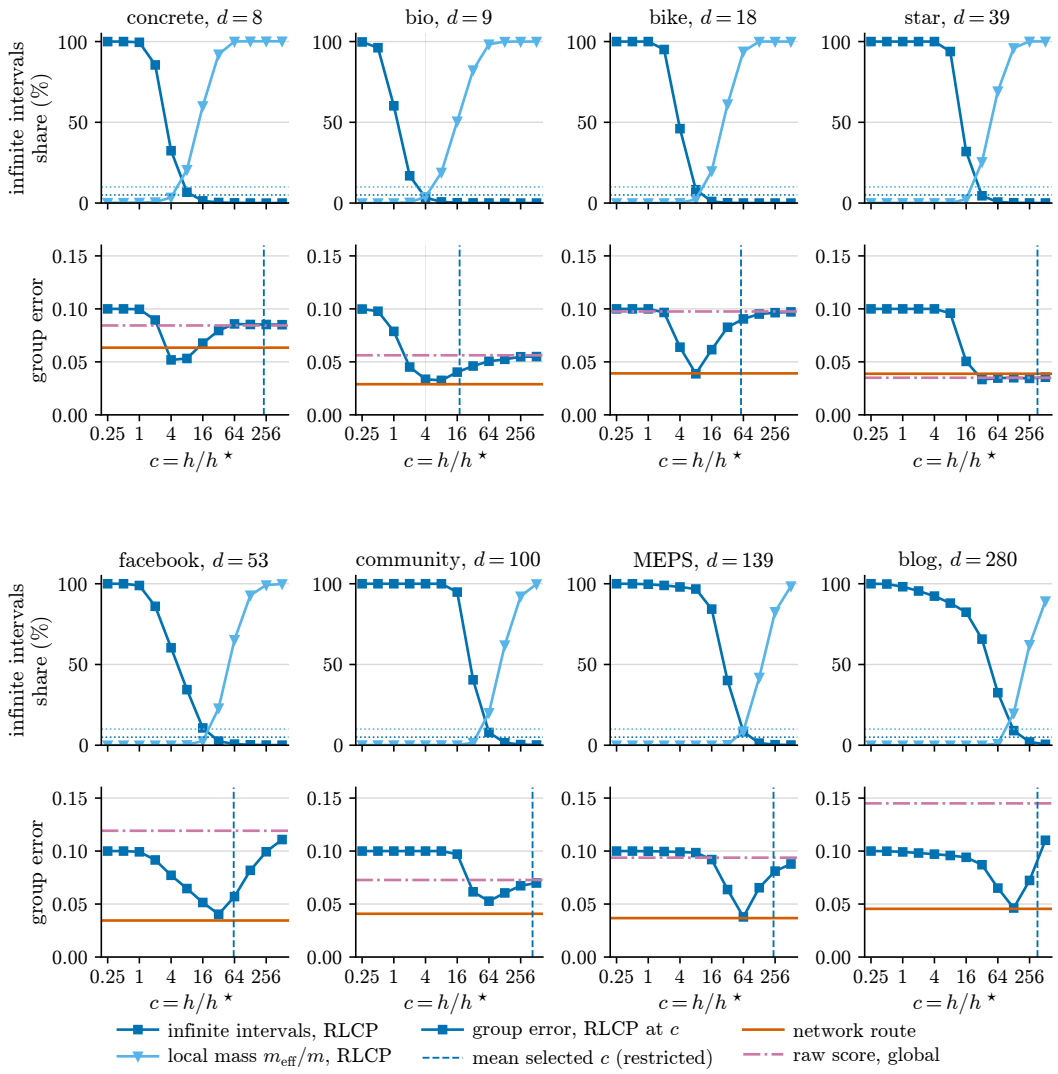

Figure 4: Ambient RLCP on the eight datasets against $c = h/h_\star$ (base-two axis), means over twenty splits. Upper panels: share of infinite intervals and local mass m_{eff}/m; the dotted lines are the two thresholds of the text, 5% infinite intervals and 10% of the fold. Lower panels: the group error of Figure 2 along the sweep, the network route and the raw score as horizontal marks, the mean selected multiple as a dashed tick.

F.4 Dimension on a simple synthetic law

One experiment isolates the ambient dimension on a law with closed-form coverage and a one-index quantile map. Let $X \sim \mathrm{Unif}([0,1]^d)$, $\varepsilon \sim N(0,1)$ independent, $z_\alpha = \Phi^{-1}(1-\alpha/2)$ and

$$\sigma(x) = 1 + 0.8\,(2u(x) - 1), \qquad Y = \frac{\sigma(X)\,\varepsilon}{z_\alpha}, \qquad S(x,y) = |y|,$$

so that $q_S^{1-\alpha}(x) = \sigma(x) \in [0.2, 1.8]$ exactly, with the sparse index $u(x) = x_1$ (the other coordinates inert) or the dense index $u(x) = \Phi(\sqrt{12d}\,(\bar{x} - 1/2))$ (every coordinate active). The sparse index is uniform; the dense one is only approximately so, by the central limit theorem, with support $(\Phi(-\sqrt{3d}), \Phi(\sqrt{3d}))$, so the two variants do not share the law of $\sigma(X)$ exactly (at $d = 2$, σ ranges over $[0.211, 1.789]$); the identity $q_S^{1-\alpha}(x) = \sigma(x)$ and the closed-form coverage hold in both. Throughout, $N = 8000$, $d \in \{2, 5, 10, 20, 50\}$, five replications, 1024 evaluation covariates: an exploratory design in which only the large effects are read.

No bandwidth localizes above a few dimensions (Figure 5a). Ambient RLCP on the whole budget runs at thirteen multiples $ch_\star(N)$, $c = 2^{k/2}$, $-3 \le k \le 9$, one centre per point. Against the local mass, R has a clear minimum at $d = 2$ (0.044 at $c = 1/2$, the neighbourhood holding 2% of the fold), a shallow one at $d = 5$ (0.080–0.083 at $c = 1$), and stays within 0.005 of the raw-score level 0.101 at every multiple from $d = 10$: the procedure returns infinite sets or calibrates globally. The uniform cube degenerates gently (the largest local and smallest usable multiples still meet at $d = 5$ and miss each other by one grid step, a factor $\sqrt{2}$, from $d = 10$), so the error curve, not the infinite share, is the evidence.

The fit is the binding constraint of the network route (Figure 5b). On the same replications, the network of Section F.2 trained for a fixed 200 epochs explains a share of the variance of σ falling from 0.98 ($d = 2$) to 0.28 ($d = 50$; 0.25 sparse), against 0.99 to 0.83 (0.85 sparse) with early stopping. The fixed schedule overfits the inert coordinates, and a larger fitting sample does little for it: in a separate check at $d = 50$, the fixed schedule explains 0.32 of the variance with $n_{\mathrm{fit}} = 4000$ and 0.40 with 16000; every network in this paper is early-stopped.

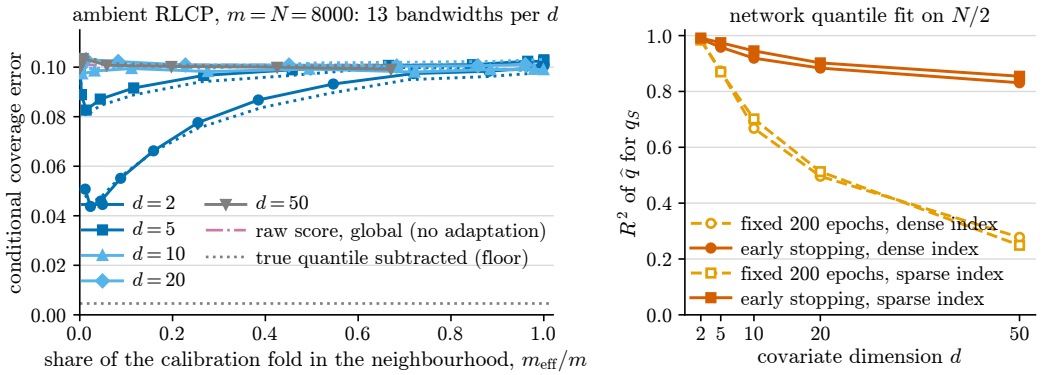

Figure 5: (a) R for ambient RLCP, $m = N = 8000$, against the share of the calibration fold in the neighbourhood, one curve per d (solid: dense, dotted: sparse), thirteen multiples of $h_\star$ per curve, one centre per point, means over five replications; raw-score level and exact floor horizontal. (b) R^2 of the fitted quantile map for σ on $N/2$ observations, fixed 200 epochs against early stopping.

F.5 Low dimension with a roughness dial

The procedures of the body run in $d \in \{1, 2\}$ on quantile maps of controlled regularity. Let $X \sim \mathrm{Unif}([0,1]^d)$, $t(x) = d^{-1/2} \sum_j x_j$, $\varepsilon \sim N(0,1)$ and

$$q^\star(x) = 1 + 0.4\,g(t(x)), \qquad Y = \frac{q^\star(X)\,\varepsilon}{z_\alpha}, \qquad S(x,y) = |y|,$$

so that $q_S^{1-\alpha}(x) = q^\star(x) \in [0.59, 1.41]$ and $\{|y| \le h\}$ has conditional coverage $2\Phi(hz_\alpha/q^\star(x)) - 1$. The maps are the lacunary sums

$$w_\gamma(t) = c_\gamma^{-1} \sum_{j=0}^{40} (1 - 2^{-\gamma}) 2^{-j\gamma} \cos(2\pi 2^j t + \phi_j), \qquad \phi_j = 2\pi\{\tfrac{1}{2} + \vartheta j\}, \quad \vartheta = 0.6180339887, \tag{52}$$

at $\gamma = \beta_q \in \{0.5, 2.5\}$, braces denoting the fractional part and c_γ a grid normalizer with $|w_\gamma| \le 1.0135$, and $g(t) = \cos(2\pi t)$. The truncated sum is smooth: β_q is the nominal index of the series, not a Hölder exponent of the simulated map.

Protocol. $N \in \{2000, 8000, 32000\}$, five replications, 2048 evaluation covariates, one centre per point; network route, raw score and floor as in Section F.2. A second score-route arm replaces the network by the estimator of the paper, the local-polynomial fit of Equation (17) at the single level $1 - \alpha$ (exact solver, degree the largest integer below β, bandwidth $(\log(8n/0.1)/n)^{1/(2\beta+d)}$ on $n = N/2$, with $\beta = \beta_q$ on the lacunary maps and $\beta = 4$ on the cosine map), the recentred score being calibrated on the other half. Ambient RLCP runs on the whole budget at $h_\star = (\log 40/N)^{1/(2\min(\beta_q,1)+d)}$ ($\beta_q = 1$ on the cosine map) and at its multiples $c \in \{1/4, \dots, 8\}$; the selected multiple follows the restricted selector of Section F.2 at the same exponent.

Readings (Figure 6). On the rough map at $d = 1$, RLCP at the rule beats the network at every budget (0.017 against 0.025 at $N = 32000$) and the local polynomial matches the rule within 0.001: the network's deficit there is a deficit of the learner, not of the score route. At $d = 2$ the three are within 0.003 of each other. On the smooth maps the two score-route arms agree within 0.004 and lead RLCP from $N = 8000$, with an R at $N = 32000$ that is 1.7 to 2.4 times smaller than the rule's. The selected multiple is large at $N = 2000$ and settles between 0.5 and 2.4 times $h_\star$ from $N = 8000$. Minimum coverage separates the routes on the smooth maps only, where the network reaches 0.82–0.89 at $N = 32000$ against 0.75–0.87 for RLCP; the local polynomial has the lowest minimum at $d = 2$, $N = 2000$, its degree-2 or 3 fit being poorly determined near the corners of the square. Infinite sets stay below 1%.

Geometry of the sets (Figure 7). The score route returns one smooth curve shifted by a single calibrated scalar: at degree 0 the local-polynomial fit tracks the rough map at the scale of its bandwidth ($b = 0.06$), the network smooths it ($R^2 = 0.76$), and on the smooth maps both reproduce $q^\star$ up to a bias of the order of the shift. RLCP returns one noisy set per centre draw, whose error changes with the draw: the thick draw has rms 0.028 on the rough map and 0.016 to 0.019 on the smooth ones, the thin draws up to 0.049 and 0.025, since a displaced centre reads the quantile at $x + hV$. The selected multiple is 4 on the rough map ($h = 0.092$, about 1275 points per neighbourhood) and 1/4 on the smooth ones ($h = 0.020$, about 280 points) at this budget, so the sets are smoother on the former and rougher on the latter.

F.6 Budget allocation with a fixed score

This experiment allocates $N = 1000$ labels between learning a rectification of the fixed score $S(x, y) = |y|$ and calibrating it, against RLCP on the whole budget. Let $X \sim \mathrm{Unif}([0,1]^d)$ and

$$r(x) = \frac{\exp\{A\, d^{-1/2} \sum_{j=1}^d |\sin(2^k \pi x_j)|\}}{\mathbb{E}\exp\{A\, d^{-1/2} \sum_{j=1}^d |\sin(2^k \pi X_j)|\}}, \qquad Y = \frac{r(X)\,\varepsilon}{\Phi^{-1}(0.95)}, \quad \varepsilon \sim N(0,1),$$

so that $\mathbb{E}[r(X)] = 1$ and $q_{|Y||X}(0.9) = r(X)$: the common oracle is $[-r(x), r(x)]$, of mean length 2, and k changes the spatial frequency of r but not its law, since $u \mapsto 1 - |2u - 1|$ preserves Lebesgue measure. The panels of Figure 8 are $(d, k) = (1, 0)$ with $A = 0$ (homogeneous control), then $(1, 0)$, $(1, 3)$ and $(2, 0)$ with A chosen so that the raw score calibrated globally has population RMS coverage error 0.08.

Procedures. Each of 200 paired replications supplies $N = 1000$ labels; learning shares 1/4, 1/2, 3/4 are nested prefixes of the same draw. The rectifier is a local polynomial fit of the conditional 0.9-quantile of $|Y|$ with uniform weights on a fraction f of nearest neighbours, the degree $\ell \in \{0, 1, 2\}$ and $f \in \{0.2, 0.4, 0.7, 1\}$ chosen by three-fold pinball cross-validation

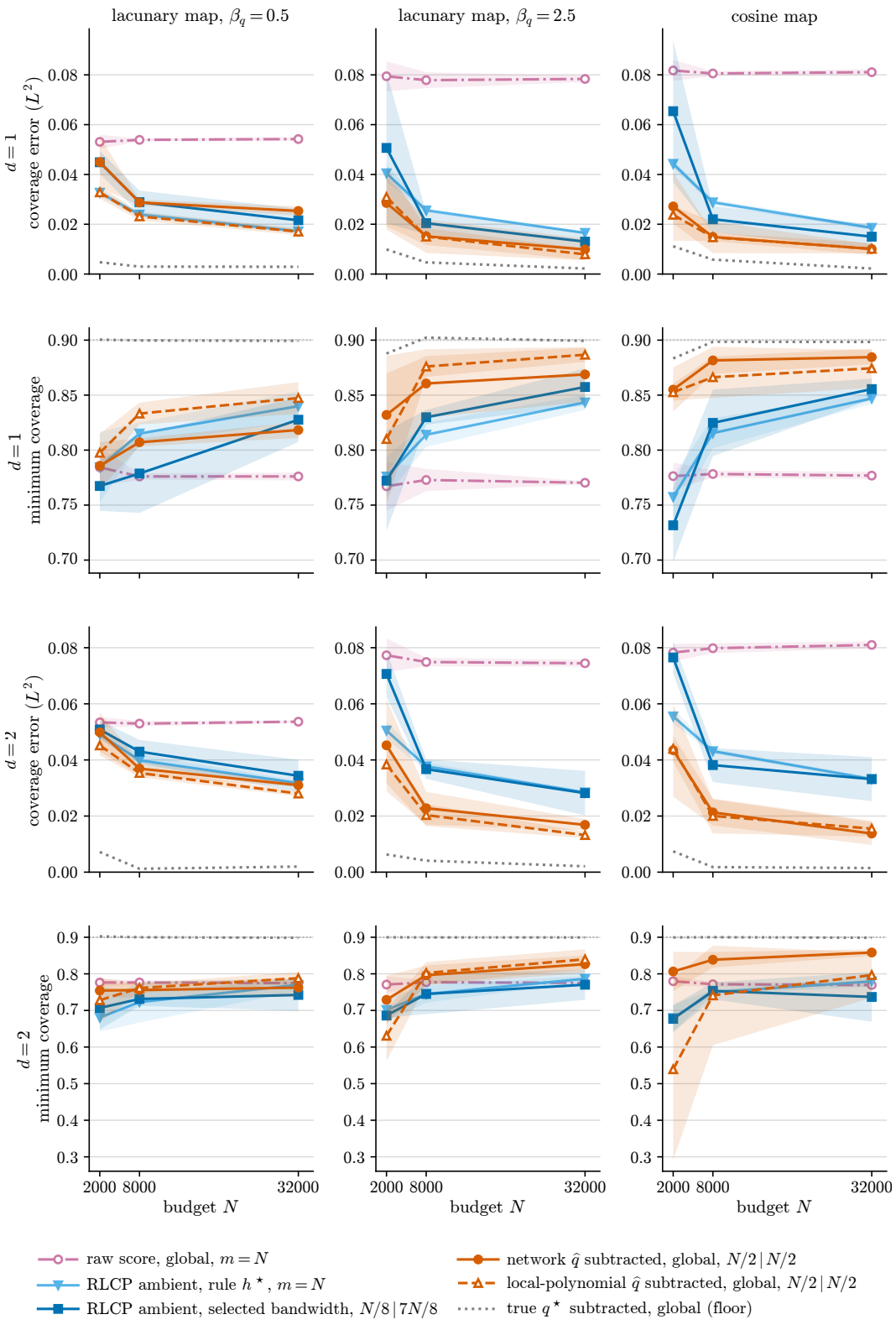

Figure 6: One column per map, two rows per d: R and the minimum evaluated coverage against N (linear axis); means and 95% t-intervals over five replications. The score route appears twice, with the network and with the local-polynomial estimator.

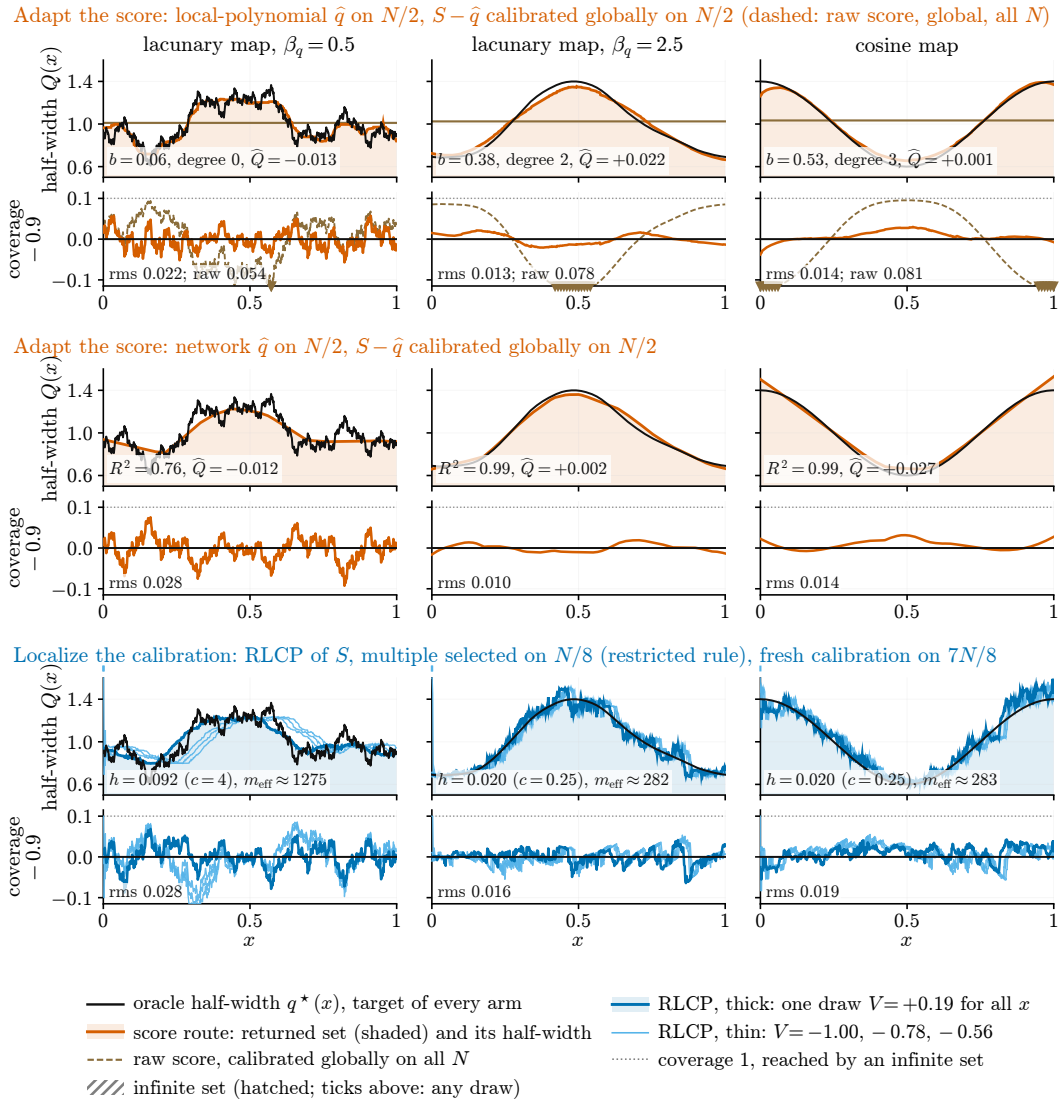

Figure 7: The sets returned by three procedures on one realization at $d = 1$, $N = 8000$ (replication 0, fixed before plotting), one column per map. Every set is $[-Q(x), Q(x)]$; each block shows its half-width $Q(x)$ (shaded) against $q^\star(x)$ (black) and, below, the exact conditional coverage minus 0.9. Top two blocks, the score route: $\widehat{q}$ by local polynomial, then by the network, fitted on $N/2$, the recentred score calibrated globally on $N/2$; dashed, the raw score calibrated globally on all N. Bottom block, RLCP of the same score at the multiple selected on $N/8$, calibrated on the remaining $7N/8$. RLCP draws a centre $x + hV$ at every x; for display one V is shared by all x (thick), three others are thin, the same four in every panel; hatching and ticks mark infinite sets. The rms printed in each panel is the L^2 error of the drawn curve; one realization illustrates the geometry, the replicated risks are in Figure 6.

($f = 1$ is a global polynomial), refitted on the whole fitting sample and projected onto $[0, \infty)$; additive rectification subtracts it from S, multiplicative rectification divides S by it, floored away from zero, each variant selected separately, and the remaining labels calibrate the transformed score globally. RLCP uses all N labels at the smallest bandwidth whose expected local count, including the test mass, reaches 60, computed on an independent bank of uniform covariates. Each replication evaluates on a fresh scrambled Sobol design of 4096 covariates with one centre per covariate.

Readings. In the homogeneous control both rectifications lead RLCP at every share; in the two smooth heterogeneous panels they overtake it as the learning share grows; in the

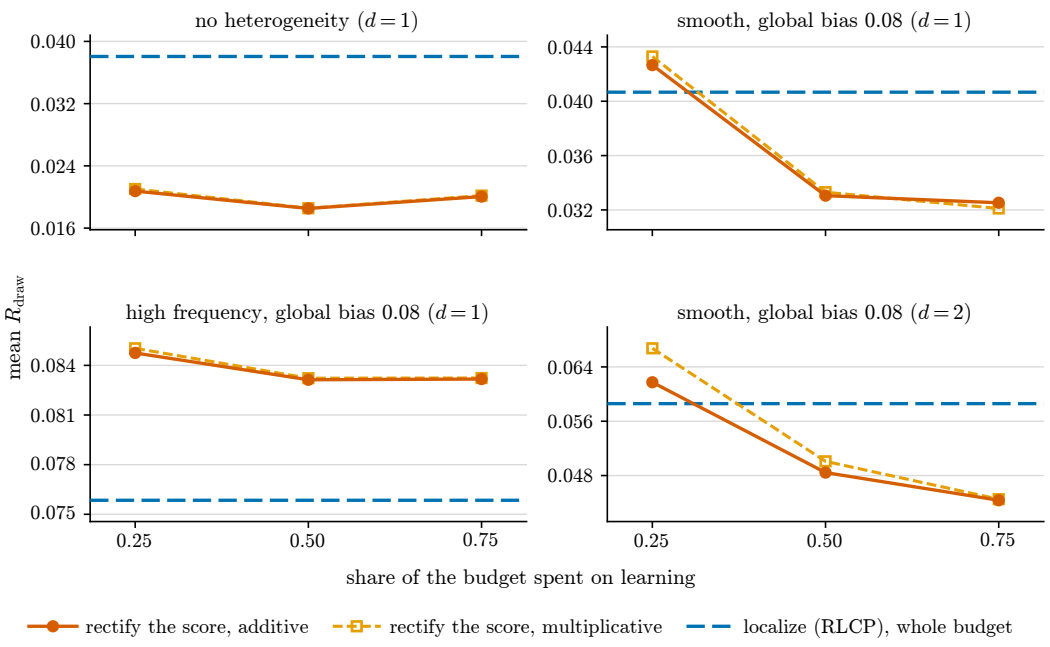

Figure 8: Mean R over 200 paired replications against the share of the budget spent on learning, four design-fixed panels, $N = 1000$. Solid: additive rectification then global calibration; dashed: multiplicative; horizontal: RLCP on the whole budget. Multiplicative rectification is outside the additive guarantees of Proposition 7.

high-frequency panel RLCP leads at every share, its neighbourhood of width about 0.06 resolving the period 1/8 of $|\sin(8\pi x)|$ that the smallest neighbour fraction allowed, $f = 0.2$, averages over. RLCP sits at 0.038 to 0.076 across the panels with at most 1.6% infinite intervals; the raw score calibrated globally sits at the designed 0.08.

F.7 Population mechanisms

Two calculations use population thresholds, no fitted model and no calibration sample, to isolate centre randomization and score-error propagation (Figure 9); both use CDF inversion and deterministic quadrature over centres. Define the centre-averaged risk

$$R^2_{\text{avg}} = \mathbb{E}_X\left[\left(\mathbb{E}_{\tilde{X}|X}\left[e(X,\tilde{X})\right]\right)^2\right], \qquad R^2 - R^2_{\text{avg}} = \mathbb{E}_X\left[\text{Var}_{\tilde{X}|X}\left(e(X,\tilde{X})\right)\right] \geq 0, \tag{53}$$

the risk of the signed error averaged over the draw before squaring; it needs several centres per point and is used here only.

Centre displacement without interior smoothing bias. Let $X \sim \text{Unif}[0,1]$, $a_0 = 0.2$ and, for $0 \leq t \leq 1$,

$$f_{Y|X=x}(y) = \tfrac{1}{2}\{1+a_0x(3y^2-1)\}\mathbf{1}_{[-1,1]}(y), \qquad H_x(t) = \mathbb{P}\{|Y| \leq t \mid X = x\} = t+a_0x(t^3-t),$$

strictly increasing, so $q_S^{1-\alpha}(x)$ solves $H_x(q) = 0.9$. H_x is affine in x and the design uniform, so a box neighbourhood centred at $\tilde{x}$ inside $[0,1]$ has population score CDF $H_{\tilde{x}}$ exactly: no smoothing bias relative to the centre. The interval is nevertheless thresholded at $q_S^{1-\alpha}(\tilde{x})$, so with $\tilde{x} = x + hV$, $V \sim \text{Unif}[-1,1]$,

$$e(x,\tilde{x}) = -a_0hV\{q_S^{1-\alpha}(\tilde{x})^3 - q_S^{1-\alpha}(\tilde{x})\}.$$

The error is linear in hV to first order, hence $R \asymp h$, and that term is odd in V, so it cancels under the signed average and $R_{\text{avg}} \asymp h^2$; the slopes of Figure 9(a) are fitted at $x = 0.5$ over six bandwidths from 0.005 to 0.12, all interior. For a neighbourhood clipped by the boundary

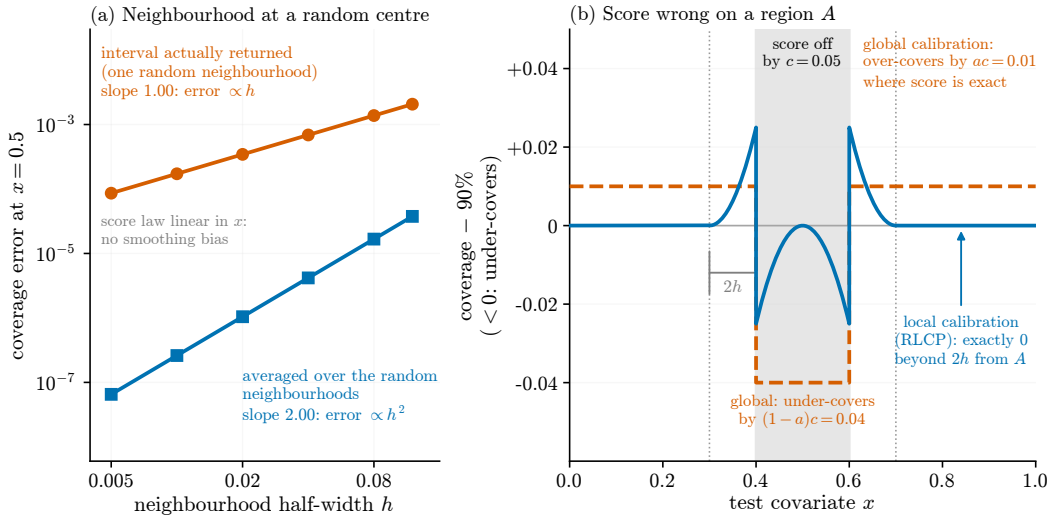

Figure 9: Population thresholds, no fitting and no calibration sample. (a) At $x = 0.5$, where every neighbourhood drawn is interior, the error of the interval actually returned decays like h (fitted log–log slope 1.0002) while the signed error averaged over the centre decays like h^2 (slope 2.0001). (b) A score error $c = 0.05$ on $A = [0.4, 0.6)$ shifts the global threshold by $ac = 0.01$ everywhere; the local curve is the signed error averaged over centres at $h = 0.05$, zero beyond distance $2h$ from A for every centre.

the population CDF is $H_{\bar{u}}$ with $\bar{u}$ the midpoint of $[\tilde{x} - h, \tilde{x} + h] \cap [0, 1]$, so the threshold is read at $\bar{u} \neq \tilde{x}$ and R_{avg} leaves the h^2 line near the edges.

Global propagation and local confinement. Let $X \sim \text{Unif}[0, 1]$, $Y \sim \text{Unif}[-1, 1]$ independent, $p = 1 - \alpha$, and

$$S(x, y) = |y| - p, \qquad \widehat{S}(x, y) = S(x, y) + c\mathbf{1}_A(x), \qquad a = |A|, \quad 0 < c < \min(\alpha, p),$$

a fixed perturbation with $\Delta_S(x) = c\mathbf{1}_A(x)$ and $\|\Delta_S\|_{L^1(P_X)} = ac$. With $[z]_0^1 = \min(1, \max(0, z))$, the marginal CDF of $\widehat{S}$ is $G(t) = (1 - a)[p + t]_0^1 + a[p + t - c]_0^1$, so the population global threshold is ac and $e_{\text{glob}}(x) = ac - c\mathbf{1}_A(x)$: calibration changes coverage outside A, where the score is exact, with signed length error $2e_{\text{glob}}(x)$. Local calibration replaces a by the neighbourhood fraction $\lambda_{\tilde{x},h}(A) = |A \cap [\tilde{x} - h, \tilde{x} + h] \cap [0, 1]|/|[\tilde{x} - h, \tilde{x} + h] \cap [0, 1]|$, giving threshold $c\lambda_{\tilde{x},h}(A)$ and error $c\{\lambda_{\tilde{x},h}(A) - \mathbf{1}_A(x)\}$; since $|\tilde{x} - x| \leq h$, $\text{dist}(x, A) > 2h$ excludes A from every neighbourhood and the local error is zero for every centre. The figure uses $A = [0.4, 0.6)$, $c = 0.05$, $h = 0.05$, neighbourhoods intersected with $[0, 1]$ and renormalized, centres not clipped.