PRIVATE CONDITIONAL QUANTILE FUNCTIONS THROUGH SHARED LOCATION LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Conditional quantile functions support prediction at uncertainty levels chosen after training. A direct approach under differential privacy fits separate models at several levels. Training these models on the same records divides a fixed privacy budget among them, increasing the noise required for each fit. In this study, we propose DP-ShareQ, which combines a shared input-dependent location with a learned shift for each quantile level. We develop two learners for this representation. Joint learns the location and these offsets together, using quantile ordering to calibrate gradient noise. CDF first learns a private location predictor, then privately estimates its residual distribution and uses its quantiles as offsets. For both learners, we derive prediction bounds that quantify the approximation cost of shared quantile spacing and the error of private learning. Across six datasets with public linear and additive features, Joint and CDF reduce average normalized integrated pinball loss relative to output-perturbed separate fits by 25.9–30.3% and 31.0–31.5%, respectively, across privacy budgets $\varepsilon \in \{1, 2, 4\}$.

1 INTRODUCTION

A prediction model is often used at several uncertainty levels. Inventory policies depend on shortage costs, service commitments depend on tolerable delays, and prediction intervals depend on the desired coverage. These choices may change after training. A conditional quantile function supports them with a single predictor indexed by level (Koenker & Bassett, 1978; Tagasovska & Lopez-Paz, 2019).

Differential privacy (DP) limits how much a published predictor can reveal about any training example (Dwork & Roth, 2014). A direct approach fits private quantile regressions at several levels and interpolates between them. Because these fits use the same records, they share a total privacy budget, and each fit requires more noise than it would under the full budget. A joint predictor can instead take both the input and the quantile level, answering later requests from a single private fit. However, coefficients describing differences in input effects across levels must also be estimated in the presence of privacy noise. When a common input effect is a good approximation, sharing it can reduce private estimation error.

In this study, we propose DP-ShareQ, which privately learns a shared input-dependent location and level-dependent shifts, or offsets (Zou & Yuan, 2008). The location moves all quantiles together, while the offsets determine their spacing. We develop two learners for this representation (Figure 1). DP-ShareQ-Joint (Joint) fits the location and offsets together using losses across quantile levels. It uses quantile ordering to bound record gradients and allocates noise across the location and offset coefficients to reduce prediction error. DP-ShareQ-CDF (CDF) first fits a private location predictor with Huber regression. It then privately estimates the residual cumulative distribution function and uses its quantiles as offsets. Both learners support later quantile requests without additional privacy cost.

We derive prediction bounds for Joint and CDF that separate the approximation cost of shared quantile spacing from private learning error. Joint's bound quantifies the trade-off between reducing error from privacy noise by sharing coefficients and approximating input-dependent quantile spacing with common offsets. For CDF, it connects accuracy in Huber regression to quantile prediction error under conditions on residual distributions, justifying learning the input effect separately from the offsets.

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.

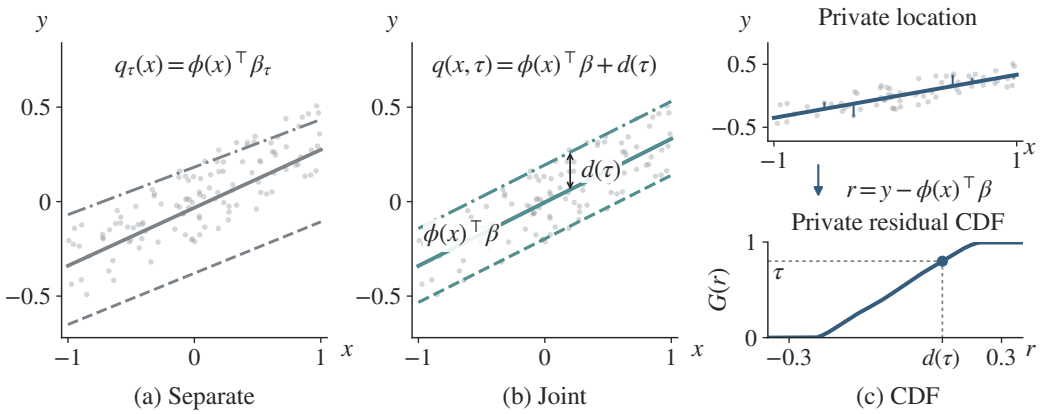


Figure 1: Private quantile learning on the same synthetic data ($n = 1024, \varepsilon = 2, \delta = n^{-2}$). (a) Separate estimates the input effect independently at each fitted level. (b) Joint learns the common effect and offsets together with directional gradient noise. (c) CDF splits its budget between a Huber location fit and a private Fourier residual CDF. Vertical segments show residuals; the guide lines show how CDF inversion gives a quantile offset. Panels (a,b) show levels $0.05, 0.5, 0.95$.

We evaluate both learners on six datasets with public linear and additive features at privacy budgets $\varepsilon \in \{1, 2, 4\}$. We compare with separate fits using output or objective perturbation and a private joint model with level-dependent input effects. At each budget, both learners achieve lower normalized integrated pinball loss than all three alternatives, averaged across datasets and feature maps. Ablations identify ordered-gradient noise calibration as an important source of Joint's gains.

2 A Shared Model for Conditional Quantiles

Let $D = \{(x_i, y_i)\}_{i=1}^n$ be an i.i.d. sample of an input X and a response Y. We estimate the conditional quantile of Y at an input x and a level $\tau \in (0, 1)$.

DP-ShareQ represents this function as a shared location plus a level-dependent offset:

$$q(x, \tau) = \phi(x)^\top \beta + d(\tau), \tag{1}$$

where $\phi(x) \in \mathbb{R}^p$ is a public feature map containing a constant, β is the coefficient vector determining the input effect, and d is nondecreasing. In this representation, the prediction difference between two inputs is the same at every level, and the gap between two levels does not depend on the input.

3 Private Learning of the Shared Model

We protect each input–response record under replacement differential privacy: neighboring datasets have the same size and differ in one record. We calibrate Gaussian noise using Gaussian differential privacy (GDP) (Dong et al., 2022), choosing the total parameter μ to meet the requested (ε, δ) guarantee. Appendix A.1 gives this conversion.

The response y lies in a known bounded interval, rescaled to $[-1, 1]$. On the public input domain $\mathcal{X}$, the feature map satisfies $\|\phi(x)\|_2 \leq 1$ for all $x \in \mathcal{X}$. The feature map is fixed independently of the training data, for example a representation pretrained on public data. Learning it from the private records adds its own privacy cost.

Joint and CDF's location stage use noisy gradient updates and return the average of the last half of the iterates (Varshney et al., 2022; Shejwalkar et al., 2025).

3.1 JOINT: FITTING LOCATION AND OFFSETS TOGETHER

On a public interval $\mathcal{T} \subset (0,1)$, we represent d by a nondecreasing piecewise-linear function on r public knots. Its constant component is absorbed into the location intercept, so the shared model has $p + r - 1$ free coefficients.

Learning objective. We combine the pinball loss $\rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\})$ over public training quantile levels $\tau_1 < \cdots < \tau_m$, with positive weights w_j summing to one, and smooth it with bandwidth $h_J > 0$ for gradient learning (Zou & Yuan, 2008; Moon & Zhou, 2023). Let θ collect the coefficients encoding β and d. The objective is

$$L_D(\theta) = \frac{1}{n} \sum_i \sum_j w_j \mathbb{E}_{V \sim \mathrm{Unif}[-h_J, h_J]} \rho_{\tau_j}\{y_i - \phi(x_i)^\top \beta - d(\tau_j) + V\}. \tag{2}$$

Let $\mathcal{K}$ be the closed convex set of coefficient vectors θ whose offset values at successive knots are nondecreasing. With a fixed nonnegative quadratic regularizer $R(\theta)$, we minimize $L_D(\theta) + R(\theta)$ over $\theta \in \mathcal{K}$. Since the offset is linear between knots, ordering its knot values ensures monotonicity throughout $\mathcal{T}$.

Noise from ordered gradients. Monotone offsets restrict the gradients a record can produce. At a fixed input, the sign of $q(x,\tau) - y$ switches at most once as the level increases. Smoothing averages these threshold patterns, so every record gradient lies in the public convex hull

$$\mathcal{G} = \mathrm{conv}\{g_l(x) : x \in \mathcal{X},\ 0 \le l \le m\}, \qquad g_l(x) = \sum_j w_j \{\mathbf{1}(j > l) - \tau_j\} \nabla_\theta q(x, \tau_j). \tag{3}$$

Here l indexes the point at which the sign pattern switches. Since the prediction is linear in θ, $\nabla_\theta q$ depends only on the input and level. Replacing one record changes the average gradient by $(g - g')/n$, with $g, g' \in \mathcal{G}$.

Private updates and output. Let $\Omega \succ 0$ be a public matrix specifying the noise across coefficient directions. Starting from a public feasible point, we take a fixed even number T of steps with learning rate $\eta > 0$:

$$\theta_{t+1} = \mathrm{prox}_{\eta(R + I_\mathcal{K})}\{\theta_t - \eta(\nabla L_D(\theta_t) + \xi_t)\}, \qquad \xi_t \sim N\left(0, \frac{T\Omega}{n^2\mu^2}\right). \tag{4}$$

For both learners, the learning rate, iteration count, and other tuning parameters are fixed independently of the training data. The noises are independent. The proximal operator applies the quadratic regularizer and the monotonicity constraint after each noisy gradient step. Here $I_\mathcal{K}$ is zero on $\mathcal{K}$ and infinite outside it. We return the coefficients $\widehat{\theta} = (2/T) \sum_{t=T/2+1}^{T} \theta_t$. Averaging both β and d preserves the ordered offsets.

Proposition 1 (Private joint learning)**.** *If* $\sup_{g,g' \in \mathcal{G}} (g - g')^\top \Omega^{-1} (g - g') \le 1$*, the* T *iterates in Equation* (4) *have a* μ*-GDP joint distribution. The averaged coefficients* $\widehat{\theta}$ *therefore define a* μ*-GDP quantile function on* $\mathcal{T}$*.*

We choose Ω independently of the training data, allocating relative variances to the intercept, input coefficients, and offset shape to reduce prediction perturbation (Li et al., 2015; Asi et al., 2021; Nikolov & Tang, 2024). For our feature maps, sensitivity reduces to a comparison of pairs of threshold patterns. Appendices A.5 and B.4 give the construction and proof of Proposition 1.

3.2 CDF: FITTING A LOCATION AND ITS RESIDUAL DISTRIBUTION

For a fixed location predictor, the quantiles of the pooled residuals give the offsets in Equation (1). For a fixed fraction $f \in (0,1)$, the location and residual stages receive budgets $\mu_A = \sqrt{1-f}\,\mu$ and $\mu_C = \sqrt{f}\,\mu$, respectively.

Private location learning. Huber regression estimates the location while bounding each record's gradient (Huber, 1964; Avella-Medina et al., 2023; Chang et al., 2026). A public threshold $h > 0$ separates the quadratic and linear parts of the scaled loss:

$$\ell_h(u) = \begin{cases} u^2/(4h), & |u| \le h, \\ |u|/2 - h/4, & |u| > h, \end{cases} \qquad H_D(\beta) = \frac{1}{n}\sum_i \ell_h(y_i - \phi(x_i)^\top \beta). \tag{5}$$

This loss equals the uniformly smoothed pinball loss at $\tau = 1/2$ with bandwidth h, up to an additive constant. Each record gradient has norm at most $1/2$, so replacing one record changes the average by at most $1/n$. With a quadratic penalty of strength $\lambda > 0$, we start from $\beta_0 = 0$ and update

$$\beta_{t+1} = \frac{\beta_t - \eta(\nabla H_D(\beta_t) + \sigma_A Z_t)}{1 + \eta\lambda}, \qquad Z_t \sim N(0, I_p), \quad 0 < \eta \le 4h. \tag{6}$$

The Gaussian noises are independent. We set

$$\sigma_A = \frac{\sqrt{T}}{n\mu_A}, \qquad \widehat{\beta} = \frac{2}{T}\sum_{t=T/2+1}^{T} \beta_t. \tag{7}$$

The resulting location estimate $\widehat{\beta}$ is μ_A-GDP.

Private residual distribution. For each observation used in the location fit, we compute the residual $y_i - \phi(x_i)^\top \widehat{\beta}$. The estimated coefficients $\widehat{\beta}$ and the public input and response domains determine an interval containing every possible residual. We normalize this interval to $[0, 1]$, obtaining residuals u_i. At a normalized threshold $t \in [0, 1]$, we estimate their CDF using M Fourier terms (Fawaz et al., 2013; Tao & Sarwate, 2026):

$$c_k(u) = \frac{\sqrt{2}\cos(\pi k u)}{\pi k}, \qquad \widetilde{F}(t) = t + \sum_{k=1}^{M} \left\{ \frac{1}{n}\sum_i c_k(u_i) + Z_k \right\} \sqrt{2}\sin(\pi k t). \tag{8}$$

A single residual determines all $c_k(u_i)$, so we calibrate noise from changes in the coefficient vector as a whole. The resulting sensitivity constant U_M depends only on M and the Fourier basis. Appendix A.7 defines this constant and proves the privacy guarantees for both CDF stages. For M between 16 and 64, it lowers every noise standard deviation by 22–24% relative to separate coordinate bounds. We add mutually independent Gaussian noises, independent of the location noise:

$$Z_k \sim N(0, \sigma_k^2), \qquad \sigma_k^2 = \frac{2U_M}{\pi^2 n^2 \mu_C^2 k}. \tag{9}$$

Proposition 2 (Residual release and composition)**.** *For every fixed location output, the coefficient release in Equation* (8) *with the noise in Equation* (9) *is μ_C-GDP. Composing it with the location estimate in Equations* (6)–(7) *on the same training data gives $\sqrt{\mu_A^2 + \mu_C^2}$-GDP.*

The noisy series need not be nondecreasing. We project $\widetilde{F}$ onto nondecreasing values on a uniform public grid, with endpoints zero and one (Tao & Sarwate, 2026). Linear interpolation and rescaling give a valid residual CDF $\widehat{G}$, whose quantiles define the offsets $\widehat{d}(\tau) = \widehat{G}^{-1}(\tau)$.

For both learners, we clip the fitted shared curve $\phi(x)^\top \widehat{\beta} + \widehat{d}(\tau)$ to the known response range $[-1, 1]$, denoting the final predictions by $\widehat{q}_J$ and $\widehat{q}_C$. Subsequent quantile requests are postprocessing and incur no additional privacy cost. Joint supports levels in $\mathcal{T}$, and CDF supports all $0 < \tau < 1$.

4 Statistical Guarantees for Shared Quantile Learning

Sharing reduces the coefficients learned privately, but fixes the quantile gaps across inputs. We first quantify this approximation cost, then bound the additional learning error of Joint and CDF.

Let P denote the population distribution and q^* its conditional quantile function. We evaluate predictions by population pinball loss averaged over a symmetric band $B = [a, 1-a]$, with $0 < a < 1/2$ and width $w = 1 - 2a$:

$$\mathcal{R}_B(q) = \frac{1}{w}\int_B \mathbb{E}\rho_\tau\{Y - q(X, \tau)\}\, d\tau. \tag{10}$$

Risk over all levels is denoted by $\mathcal{R}_{(0,1)}$.

4.1 Optimal shared quantile prediction

Consider a population $Y = \phi(X)^\top \beta_* + E$, where $E \mid X = x$ is symmetric about zero and $\mathbb{E}|E| < \infty$. Let G_* be the CDF of the residuals pooled over inputs, with quantiles $d_*(\tau) = G_*^{-1}(\tau)$. Using these quantiles as shared offsets defines the population curve $q_{\text{pool}}(x, \tau) = \phi(x)^\top \beta_* + d_*(\tau)$.

Proposition 3 (Optimal shared curve)**.** *The curve q_{pool} minimizes $\mathcal{R}_B$ over shared curves q of the form* (1)*, before final clipping. The same conclusion holds over all levels* $(0, 1)$.

Proposition 3 identifies the best predictor under shared quantile gaps, even when the true residual distributions differ across inputs. Its approximation cost is $\mathcal{A}_B = \mathcal{R}_B(q_{\text{pool}}) - \mathcal{R}_B(q^*)$.

To quantify this cost, let $F_x(r) = \mathbb{P}(E \le r \mid X = x)$ be the residual CDF at input x, so $G_*(r) = \mathbb{E}_X F_X(r)$. Over all levels, the cost is exactly

$$\mathcal{A}_{(0,1)} = \frac{1}{2} \int_{\mathbb{R}} \mathbb{E}_X \left[(F_X(r) - G_*(r))^2 \right] \, dr. \tag{11}$$

When the residual distributions are identical across inputs, $q_{\text{pool}} = q^*$ and $\mathcal{A}_B = 0$. Proofs of Proposition 3 and Equation (11) appear in Appendices A.3 and A.4.

4.2 Joint: private learning with shared coefficients

Joint learns the shared curve using the piecewise-linear offset in Section 3.1. Let $\mathcal{E}_J$ combine the cost of approximating q_{pool} with a piecewise-linear offset and the costs of regularization, loss smoothing, and numerical integration over the training quantile levels.

Theorem 4 (Joint prediction error)**.** *Under the population assumptions of Section 4.1, the bounded features and responses of Section 3, and the learning conditions in Appendix A.6, Joint with T updates, learning rate η, and training quantile levels in $B \subseteq \mathcal{T}$ satisfies*

$$\mathbb{E}\mathcal{R}_B(\widehat{q}_J) - \mathcal{R}_B(q^*) \le \mathcal{A}_B + \mathcal{E}_J + O\left(\frac{1}{T} + \frac{\eta T}{n} + \frac{\eta T \operatorname{tr} \Omega}{n^2 \mu^2} \right). \tag{12}$$

Here the outer expectation is over training data and privacy noise, with the learning settings fixed. The terms inside $O(\cdot)$ describe optimization, generalization, and privacy noise, respectively. Appendix A.6 defines $\mathcal{E}_J$ precisely and gives the full bound.

The privacy benefit of sharing. For fixed n, T, μ, both the total variance of the gradient noise and the privacy term in the bound are proportional to $\operatorname{tr} \Omega = \sum_j \Omega_{jj}$. Under matched sensitivity calibration and equal noise variance per coefficient in an orthonormal quantile basis, this sum is proportional to the coefficient count. Sharing reduces this count from pr, when input effects vary across levels, to $p + r - 1$. Appendix A.6 derives the corresponding trace comparison.

4.3 CDF: from Huber loss to quantile risk

We analyze how errors in CDF's two stages affect the same band risk used for Joint. For coefficients β, let $G_\beta(r) = \mathbb{P}\{Y - \phi(X)^\top \beta \le r\}$ be the population residual CDF, so $G_{\beta_*} = G_*$.

Transfer from Huber loss. Under the population model of Section 4.1, assume each conditional residual density is nonincreasing away from zero and the pooled residual density is bounded by $1/(2h)$. Here h is the Huber threshold in Section 3.2. The population Huber risk is $H_P(\beta) = \mathbb{E}\ell_h(Y - \phi(X)^\top \beta)$.

Theorem 5 (Huber-to-quantile risk bound)**.** *Under these conditions, β_* minimizes H_P. For every coefficient vector β and residual CDF G with a finite first moment, the shared predictor $q = \phi^\top \beta + G^{-1}$ satisfies*

$$\mathcal{R}_B(q) - \mathcal{R}_B(q^*) \le \mathcal{A}_B + \frac{1}{w} \left\{ H_P(\beta) - H_P(\beta_*) + \tfrac{1}{2} \left\| G - G_\beta \right\|_{L^2(\mathbb{R})}^2 \right\}. \tag{13}$$

The Huber excess risk $H_P(\beta) - H_P(\beta_*)$ measures the error in learning the shared input effect. The squared distance $\|G - G_\beta\|_{L^2(\mathbb{R})}^2$ measures how well the CDF G used for prediction approximates the population residual distribution G_β at the same β. Appendix A.8 proves the bound and the optimality of its constants over $(0, 1)$.

Combined prediction error. We apply this transfer to the estimates $(\widehat{\beta}, \widehat{G})$ from Section 3.2 and bound the errors of the two private stages. Let $\mathcal{E}_C$ collect the costs of regularizing the input coefficients and approximating the residual CDF with a finite Fourier series and a grid.

Theorem 6 (CDF prediction error)**.** *Under the distributional assumptions of Theorem 5 and the learning conditions in Appendix A.12, CDF with T input-effect updates, learning rate $0 < \eta \leq 4h$, and M Fourier terms satisfies*

$$\mathbb{E}\mathcal{R}_B(\widehat{q}_C) - \mathcal{R}_B(q^*) \leq \mathcal{A}_B + \mathcal{E}_C + O\left(\frac{1}{T} + \frac{1}{n} + \frac{\eta T p}{n^2 \mu_A^2} + \frac{U_M \log(eM)}{n^2 \mu_C^2}\right). \tag{14}$$

The terms inside $O(\cdot)$ describe optimization, generalization, and privacy noise in the location and residual stages, respectively. Appendix A.12 defines $\mathcal{E}_C$ and gives the full bound.

Learning all offsets from one distribution. Increasing M reduces the CDF approximation term in $\mathcal{E}_C$ but raises the residual noise term. All requested quantiles use this same distribution estimate, so their number does not enter the bound. Allocating more privacy budget to the residual stage lowers its noise but leaves less for learning the input effect.

5 Experiments

Setting. We use the regression datasets California Housing, Concrete, Energy Efficiency, Airfoil Self-Noise, Power Plant, and Real Estate, each with public linear and additive maps. The linear map gives each continuous input a linear effect. The additive map uses 12 fixed, equal-width bins per continuous input and sums the resulting piecewise-constant effects.

We split each dataset once into 60% training, 15% validation, and 25% testing data. Each method selects hyperparameters, such as regularization strength λ, by the mean validation normalized integrated pinball loss (nIPL) over ten independent noise repetitions. Keeping these hyperparameters fixed, we retrain on the combined training and validation data (75%) and evaluate on the held-out test data (25%). We repeat the private learning procedure 100 times with independent noise.

We use total privacy budgets $\varepsilon \in \{1, 2, 4\}$ and $\delta = n^{-2}$, where n is the number of training observations. For every method, the reported privacy budget covers one fit with fixed hyperparameters and excludes hyperparameter selection. Appendix B.6 charges this step to the same budget through private selection (Liu & Talwar, 2019; Papernot & Steinke, 2022; Koskela & Kulkarni, 2023). All methods are evaluated at 513 equally spaced quantile levels on $[0.05, 0.95]$. Appendix B gives the preprocessing, prediction postprocessing, hyperparameter search ranges, and validation losses over the residual budget fraction and Fourier order.

Methods and reporting. We compare separate quantile regression using output perturbation (Sep-Out) or objective perturbation (Sep-Obj) (Chen & Chua, 2023, Algorithm 2). Both methods fit ridge-regularized smoothed quantile regressions independently at selected levels and interpolate between them. Full jointly learns a function $q(x, \tau)$ whose input effects can vary with the quantile level. It uses the same quantile basis and smoothed loss as Joint, with private gradient descent (Bassily et al., 2014). At the same total (ε, δ), Sep-Obj uses its objective-perturbation guarantee, while all other methods calibrate Gaussian noise using GDP.

We report nIPL for average accuracy across the quantile band and pinball loss differences (Joint minus CDF) for the comparison at individual levels. We also report the normalized score for central 90% prediction intervals (nIS_{90}), penalizing width and the distance of observations outside the interval. All losses use the normalized response and are multiplied by 100.

Tables report means and sample standard deviations over 100 repetitions, and figures show means. When aggregating across datasets or feature maps, we average their losses within each repetition and compute the reported statistics from the resulting 100 averages.

Table 1: nIPL $\times 100$ at $\varepsilon = 2$, for both input maps. Entries are mean $\pm$ standard deviation over 100 noise repetitions. Bold marks the lowest displayed mean in each row.

		Reference methods			DP-ShareQ (ours)	
Dataset	Map	Sep-Out	Sep-Obj	Full	Joint	CDF
California	Linear	10.783 ± 1.217	9.577 ± 0.137	7.718 ± 0.058	7.737 ± 0.037	$\mathbf{7.646} \pm 0.026$
	Additive	9.573 ± 0.046	8.829 ± 0.103	8.057 ± 0.049	7.901 ± 0.028	$\mathbf{7.899} \pm 0.030$
Concrete	Linear	8.969 ± 0.221	8.807 ± 0.387	8.033 ± 0.351	6.652 ± 0.386	$\mathbf{6.406} \pm 0.328$
	Additive	8.018 ± 0.176	8.695 ± 0.429	7.713 ± 0.254	6.403 ± 0.254	$\mathbf{6.123} \pm 0.212$
Energy	Linear	8.292 ± 0.407	8.608 ± 0.914	6.310 ± 0.524	3.697 ± 0.256	$\mathbf{3.176} \pm 0.134$
	Additive	3.492 ± 0.401	5.725 ± 0.627	4.458 ± 0.306	1.602 ± 0.202	$\mathbf{1.420} \pm 0.167$
Airfoil	Linear	5.122 ± 0.071	5.159 ± 0.089	4.757 ± 0.130	3.702 ± 0.132	$\mathbf{3.611} \pm 0.132$
	Additive	4.610 ± 0.071	4.935 ± 0.149	4.654 ± 0.080	3.939 ± 0.108	$\mathbf{3.704} \pm 0.118$
Power Plant	Linear	2.692 ± 0.171	2.539 ± 0.075	2.335 ± 0.011	$\mathbf{2.310} \pm 0.005$	2.318 ± 0.005
	Additive	3.136 ± 0.172	2.955 ± 0.096	2.557 ± 0.027	2.331 ± 0.011	$\mathbf{2.329} \pm 0.013$
Real Estate	Linear	5.854 ± 0.821	6.683 ± 0.984	5.491 ± 0.616	3.878 ± 0.367	$\mathbf{3.763} \pm 0.471$
	Additive	5.000 ± 0.607	8.357 ± 1.265	6.082 ± 0.398	3.711 ± 0.235	$\mathbf{3.526} \pm 0.300$

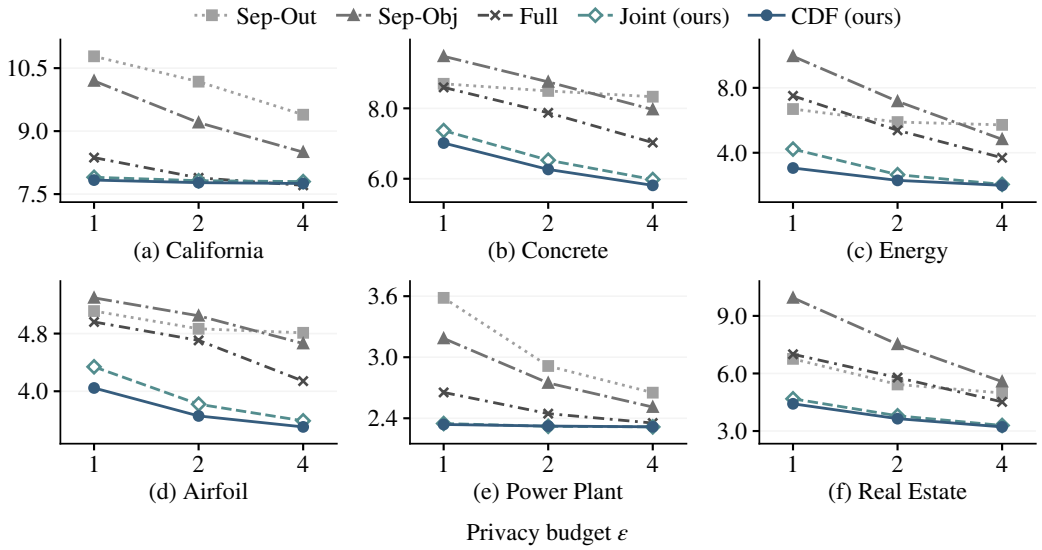


Figure 2: Mean nIPL $\times 100$ by dataset across privacy budgets over 100 noise repetitions, averaging the two input maps within each repetition.

5.1 Quantile Prediction

Table 1 reports nIPL for each dataset and map at $\varepsilon = 2$. Joint and CDF both improve on Sep-Out and Sep-Obj in every dataset–map setting. Their gains over Full are substantial on Energy, Concrete, and Real Estate. Full is much closer on California and Power Plant.

Across the budgets in Figure 2, Full has lower nIPL than both Sep-Out and Sep-Obj on average over datasets and maps. Joint and CDF further reduce nIPL relative to Full by 15.0–21.0% and 16.5–26.5%, respectively. Their reductions relative to Sep-Out are larger, at 25.9–30.3% and 31.0–31.5%. On these benchmarks, sharing the input effect together with the learners' noise calibration yields lower error than fitting level-dependent effects.

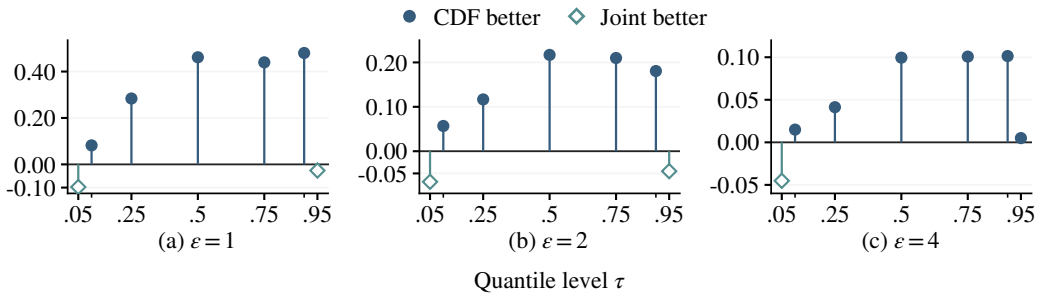


Figure 3: Mean pinball loss difference $\times 100$, Joint minus CDF, averaged over datasets, maps, and 100 noise repetitions. Positive values favor CDF, and negative values favor Joint.

Table 2: Ablations of Joint (upper block) and CDF (lower block). Losses are multiplied by 100. Entries average datasets and maps within each repetition and report mean $\pm$ standard deviation over 100 repetitions. Bold marks the lowest displayed mean within each block and column.

Method	nIPL			nIS_{90}		
	$\varepsilon = 1$	$\varepsilon = 2$	$\varepsilon = 4$	$\varepsilon = 1$	$\varepsilon = 2$	$\varepsilon = 4$
Joint	**5.144** $\pm$ 0.104	**4.489** $\pm$ 0.059	**4.171** $\pm$ 0.028	**78.589** $\pm$ 2.413	**66.355** $\pm$ 1.267	**61.054** $\pm$ 0.492
w/o ord. bound	5.973 $\pm$ 0.207	5.126 $\pm$ 0.149	4.403 $\pm$ 0.064	91.741 $\pm$ 3.938	77.391 $\pm$ 2.438	66.378 $\pm$ 1.164
w/o ordering	5.963 $\pm$ 0.208	5.120 $\pm$ 0.149	4.399 $\pm$ 0.064	91.277 $\pm$ 4.307	76.883 $\pm$ 2.546	66.044 $\pm$ 1.222
$\Omega \propto I$	5.498 $\pm$ 0.131	4.711 $\pm$ 0.078	4.271 $\pm$ 0.038	96.955 $\pm$ 4.825	82.106 $\pm$ 4.165	70.162 $\pm$ 2.631
CDF	**4.785** $\pm$ 0.091	**4.327** $\pm$ 0.061	**4.099** $\pm$ 0.033	**81.069** $\pm$ 6.922	**68.632** $\pm$ 4.639	**61.858** $\pm$ 2.164
Quantile loc.	5.091 $\pm$ 0.134	4.511 $\pm$ 0.085	4.183 $\pm$ 0.039	83.695 $\pm$ 6.473	70.520 $\pm$ 4.829	61.947 $\pm$ 2.076
Coord. bound	4.838 $\pm$ 0.101	4.350 $\pm$ 0.067	4.105 $\pm$ 0.035	86.821 $\pm$ 7.143	72.795 $\pm$ 6.474	63.607 $\pm$ 3.154

CDF's average gain over Joint decreases from 7.0% at $\varepsilon = 1$ to 1.7% at $\varepsilon = 4$. Full has the lowest mean nIPL on California at $\varepsilon = 4$. Appendix C gives the remaining dataset–map comparisons.

Accuracy across quantile levels. Averaged over datasets and maps (Figure 3), CDF has lower pinball loss at the five interior levels from 0.1 to 0.9 at every budget, while Joint has lower loss at 0.05. Joint also has lower loss at 0.95 for $\varepsilon = 1, 2$; the two are nearly tied there at $\varepsilon = 4$. These differences favor CDF for average accuracy across the band and Joint for the pair of extreme quantiles defining a 90% prediction interval. Appendix D reports interval scores, widths, and coverage.

5.2 Contributions of the Learning Designs

Table 2 reports the ablations. Results by dataset appear in Appendices E.1 and E.2. Each variant retains the selected settings of Joint or CDF except for the component being studied. Appendix E.4 isolates coefficient sharing within output perturbation, as in composite quantile regression.

Joint: ordering. The variant w/o ordered bound increases the scale of Ω using a sensitivity bound that allows arbitrary patterns of gradient signs across quantile levels, instead of the ordered patterns in $\mathcal{G}$. Joint reduces nIPL relative to this variant by 13.9%, 12.4%, and 5.3% at $\varepsilon = 1, 2, 4$. The variant w/o ordering additionally removes $\theta \in \mathcal{K}$ from Equation (4), allowing $d(\tau)$ to decrease between knots. After the common monotone rearrangement of predictions, its nIPL is slightly lower than that of w/o ordered bound, with differences below 0.2%. Under this postprocessing, ordered-gradient noise calibration gives the larger gain.

Joint: directional noise. Isotropic noise sets Ω proportional to the identity matrix. Joint reduces nIPL by 6.4%, 4.7%, and 2.3%, and the 90% interval score by 18.9%, 19.2%, and 13.0% at $\varepsilon = 1, 2, 4$. The larger gain in interval score shows that noise allocation matters particularly for the extreme quantiles.

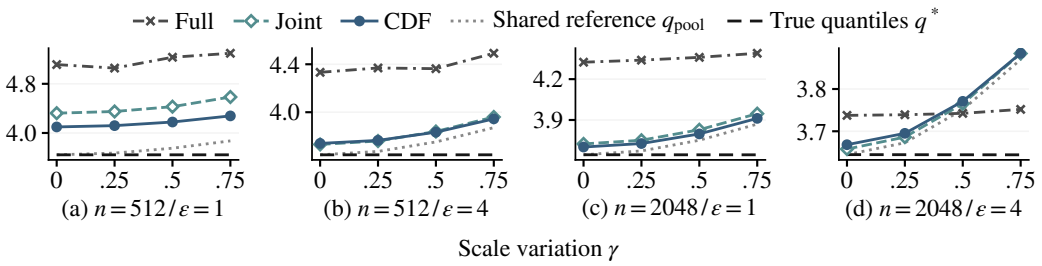


Figure 4: Sharing under input-dependent scale. Curves show mean nIPL $\times 100$ over 100 independent repetitions with newly generated training data and noise.

CDF: location learning. Quantile location fits the shared input effect and offsets with the pinball objective over multiple levels in Equation (2) and uses its median curve in place of the Huber location predictor. CDF reduces nIPL by 6.0%, 4.1%, and 2.0% relative to this variant at $\varepsilon = 1, 2, 4$.

CDF: residual noise. Coordinate bound replaces U_M in Equation (9) by $4 \sum_{k=1}^{M} k^{-1}$, bounding each coefficient change separately. CDF's joint-change bound reduces nIPL by 1.08%, 0.54%, and 0.14%, and the 90% interval score by 6.6%, 5.7%, and 2.7% at $\varepsilon = 1, 2, 4$.

5.3 Simulation under Input-Dependent Scale

We test whether learning fewer coefficients still pays when quantile spacing varies with the input. We generate $Y = 0.25X_1 + 0.2(1 + \gamma X_2)U$, with $X_1, \ldots, X_8, U$ independently uniform on $[-1, 1]$. Larger γ makes quantile spacing depend more strongly on the input. All learners use the linear map $\phi(x) = (1, x)^\top / 3$. We compare training sample sizes $n \in \{512, 2048\}$ and privacy budgets $\varepsilon \in \{1, 4\}$.

At $n = 512$, Joint and CDF outperform Full at both budgets throughout the tested range of γ (Figure 4). With $n = 2048, \varepsilon = 4$, the comparison reverses as scale variation increases. In this setting, the shared learners approach q_{pool}, but its gap from q^* grows with γ. Larger samples and privacy budgets reduce the privacy-noise terms in the bounds, but not the approximation cost of sharing.

6 Related Work

Composite quantile regression learns a common input effect across levels (Zou & Yuan, 2008). Private quantile regression (Chen & Chua, 2023; Tran et al., 2024) builds on private convex estimation (Chaudhuri et al., 2011; Kifer et al., 2012). Recent work studies locally private online quantile regression (Liu & Hu, 2026) and high-dimensional distributed inference (Shen et al., 2026).

Private conformal methods calibrate the coverage of prediction sets using held-out data (Angelopoulos et al., 2022) or training data (Wu et al., 2026). DP-ShareQ learns a conditional quantile curve and controls its integrated prediction risk.

7 Limitations and Conclusion

DP-ShareQ uses the same quantile gaps at every input. When the true gaps vary with the input, the approximation cost of sharing remains even as privacy noise decreases. For CDF, the bound connecting Huber loss to quantile prediction error assumes symmetric, unimodal conditional residual densities and a bound on the pooled density. Appendix F.2 evaluates both learners with skewed and bimodal residuals. Our experiments use tabular datasets with fixed feature maps.

Joint and CDF share input effects across quantile levels to reduce privacy noise. A natural extension is to let quantile gaps vary with the input while retaining the noise reduction from sharing.

AI USE STATEMENT

Generative AI tools assisted with research ideation and execution, drafting, and polishing manuscript text. The authors take responsibility for the final text, methods, claims, and results.

REFERENCES

Anastasios Nikolas Angelopoulos, Stephen Bates, Tijana Zrnic, and Michael I. Jordan. Private prediction sets. *Harvard Data Science Review*, 2022.

Hilal Asi, John Duchi, Alireza Fallah, Omid Javidbakht, and Kunal Talwar. Private adaptive gradient methods for convex optimization. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 383–392, 2021.

Marco Avella-Medina, Casey Bradshaw, and Po-Ling Loh. Differentially private inference via noisy optimization. *The Annals of Statistics*, 51(5), 2023.

Raef Bassily, Adam Smith, and Abhradeep Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pp. 464–473, 2014.

Olivier Bousquet and André Elisseeff. Stability and generalization. *Journal of machine learning research*, 2(Mar):499–526, 2002.

Thomas F Brooks, D Stuart Pope, and Michael A Marcolini. Airfoil self-noise and prediction. Technical Report NASA-RP-1218, National Aeronautics and Space Administration, 1989.

Jinyuan Chang, Lin Yang, Mengyue Zha, and Wen-Xin Zhou. Adapting to noise tails in private linear regression. *Journal of the American Statistical Association*, pp. 1–14, 2026.

Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12(29):1069–1109, 2011.

Du Chen and Geoffrey A. Chua. Differentially private stochastic convex optimization under a quantile loss function. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 4435–4461, 2023.

Jinshuo Dong, Aaron Roth, and Weijie J. Su. Gaussian differential privacy. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(1):3–37, 2022.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9:211–407, 2014.

Nadia Fawaz, S. Muthukrishnan, and Aleksandar Nikolov. Nearly optimal private convolution. In *Algorithms – ESA 2013*, pp. 445–456, 2013.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

Somesh Das Gupta. On Anderson's probability inequality. In *Studies in Econometrics, Time Series, and Multivariate Statistics*, pp. 407–417. Elsevier, 1983.

Moritz Hardt, Ben Recht, and Yoram Singer. Train faster, generalize better: Stability of stochastic gradient descent. In *Proceedings of The 33rd International Conference on Machine Learning*, pp. 1225–1234, 2016.

Xuming He, Xiaoou Pan, Kean Ming Tan, and Wen-Xin Zhou. Smoothed quantile regression with large-scale inference. *Journal of Econometrics*, 232(2):367–388, 2023.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35 (1):73–101, 1964.

Daniel Kifer, Adam Smith, and Abhradeep Thakurta. Private convex empirical risk minimization and high-dimensional regression. In *Proceedings of the 25th Annual Conference on Learning Theory*, pp. 25.1–25.40, 2012.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33, 1978.

Antti Koskela and Tejas D Kulkarni. Practical differentially private hyperparameter tuning with subsampling. In *Advances in Neural Information Processing Systems*, volume 36, pp. 28201–28225, 2023.

Chao Li, Gerome Miklau, Michael Hay, Andrew McGregor, and Vibhor Rastogi. The matrix mechanism: optimizing linear counting queries under differential privacy. *The VLDB Journal*, 24 (6):757–781, 2015.

Jingcheng Liu and Kunal Talwar. Private selection from private candidates. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 298–309, 2019.

Yi Liu and Qirui Hu. Locally private online quantile regression: Estimation and inference. *arXiv preprint arXiv:2607.05312*, 2026.

Haeseong Moon and Wen-Xin Zhou. High-dimensional composite quantile regression: Optimal statistical guarantees and fast algorithms. *Electronic Journal of Statistics*, 17(2), 2023.

Aleksandar Nikolov and Haohua Tang. General Gaussian Noise Mechanisms and Their Optimality for Unbiased Mean Estimation. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, pp. 85:1–85:23, 2024.

R Kelley Pace and Ronald Barry. Sparse spatial autoregressions. *Statistics & Probability Letters*, 33 (3):291–297, 1997.

Nicolas Papernot and Thomas Steinke. Hyperparameter tuning with renyi differential privacy. In *International Conference on Learning Representations*, 2022.

Virat Shejwalkar, Arun Ganesh, Rajiv Mathews, Yarong Mu, Shuang Song, Om Thakkar, Abhradeep Thakurta, and Xinyi Zheng. Recycling scraps: Improving private learning by leveraging checkpoints. *Proceedings on Privacy Enhancing Technologies*, 2025(2):607–628, 2025.

Ziliang Shen, Caixing Wang, Shaoli Wang, and Yibo Yan. High-dimensional differentially private quantile regression: Distributed estimation and statistical inference. *arXiv preprint arXiv:2508.05212*, 2026.

Lukas Steinberger and Hannes Leeb. Conditional predictive inference for stable algorithms. *The Annals of Statistics*, 51(1), 2023.

Natasa Tagasovska and David Lopez-Paz. Single-model uncertainties for deep learning. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Ye Tao and Anand D. Sarwate. Functional approximation methods for differentially private distribution estimation. *arXiv preprint arXiv:2501.06620*, 2026.

Tran Tran, Matthew Reimherr, and Aleksandra Slavkovic. Differentially private quantile regression. In *Privacy in Statistical Databases*, pp. 18–34, 2024.

Athanasios Tsanas and Angeliki Xifara. Energy Efficiency. UCI Machine Learning Repository, 2012. DOI: https://doi.org/10.24432/C51307.

Pınar Tüfekci. Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. *International Journal of Electrical Power & Energy Systems*, 60:126–140, 2014.

Prateek Varshney, Abhradeep Thakurta, and Prateek Jain. (nearly) optimal private linear regression for sub-gaussian data via adaptive clipping. In *Proceedings of Thirty Fifth Conference on Learning Theory*, pp. 1126–1166, 2022.

Jiamei Wu, Ce Zhang, Zhipeng Cai, Jingsen Kong, Bei Jiang, Linglong Kong, and Lingchen Kong. Differentially private conformal prediction. *arXiv preprint arXiv:2604.14621*, 2026.

I-Cheng Yeh. Concrete Compressive Strength. UCI Machine Learning Repository, 1998. DOI: https://doi.org/10.24432/C5PK67.

I-Cheng Yeh. Real Estate Valuation. UCI Machine Learning Repository, 2018. DOI: https://doi.org/10.24432/C5J30W.

Hui Zou and Ming Yuan. Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, 36(3), 2008.

A PROOFS OF THE MAIN GUARANTEES

We use a risk decomposition to establish the shared target and approximation cost, then prove privacy and prediction bounds for Joint and CDF. The two learners use the same Gaussian privacy accounting. For the averaged outputs, write $\mathcal{I}_T = \{T/2+1, \dots, T\}$, with T even.

A.1 GAUSSIAN PRIVACY ACCOUNTING

Under replacement adjacency, neighboring datasets have the same size and differ in one record. A mechanism $\mathcal{M}$ is (ε, δ)-DP if, for every neighboring pair and measurable set S,

$$\mathbb{P}\{\mathcal{M}(D) \in S\} \le e^{\varepsilon}\mathbb{P}\{\mathcal{M}(D') \in S\} + \delta.$$

Let Φ be the standard normal CDF. The Gaussian trade-off function is $G_\mu(\alpha) = \Phi(\Phi^{-1}(1-\alpha) - \mu)$. A mechanism is μ-GDP when the hypothesis-testing trade-off between its outputs on neighboring datasets is bounded below by G_μ. We use three standard properties (Dong et al., 2022). A Gaussian mean release with Mahalanobis sensitivity at most μ is μ-GDP. Adaptive GDP mechanisms compose by the Euclidean norm of their parameters. A μ-GDP mechanism is $(\varepsilon, \delta_\mu(\varepsilon))$-DP, where

$$\delta_\mu(\varepsilon) = \Phi\left(-\frac{\varepsilon}{\mu} + \frac{\mu}{2}\right) - e^{\varepsilon}\Phi\left(-\frac{\varepsilon}{\mu} - \frac{\mu}{2}\right). \tag{15}$$

We solve $\delta_\mu(\varepsilon) = \delta$ to obtain the total parameter.

A.2 THE LOCATION–DISTRIBUTION DECOMPOSITION

For coefficients β, write $R_\beta = Y - \phi(X)^\top\beta$. Using its population CDF G_β to form the offsets gives the risk

$$J(\beta) = \mathcal{R}_{(0,1)}\left(\phi^\top\beta + G_\beta^{-1}\right). \tag{16}$$

Proposition 7 (Errors in the input effect and residual distribution)**.** *If R_β and the distribution with CDF G have finite first moments, the shared curve $q = \phi^\top\beta + G^{-1}$ satisfies*

$$\mathcal{R}_{(0,1)}(q) = J(\beta) + \tfrac{1}{2}\left\|G - G_\beta\right\|^2_{L^2(\mathbb{R})}. \tag{17}$$

Proof. For a CDF G and real r, the continuous ranked probability score is

$$\mathrm{CRPS}(G, r) = \int_{\mathbb{R}} (G(t) - \mathbf{1}\{r \le t\})^2\, dt = 2\int_0^1 \rho_\tau(r - G^{-1}(\tau))\, d\tau.$$

The equality holds for distributions with finite first moments (Gneiting & Raftery, 2007). Taking expectation over R_β gives

$$\mathcal{R}_{(0,1)}(q) = \tfrac{1}{2}\left\|G - G_\beta\right\|^2_{L^2} + \tfrac{1}{2}\int_{\mathbb{R}} G_\beta(t)(1 - G_\beta(t))\, dt.$$

For independent identically distributed R_β, R'_β, $\mathbb{E}|R_\beta - R'_\beta| = 2\int G_\beta(1-G_\beta)$. Setting $G = G_\beta$ identifies the last term with $J(\beta)$, so

$$J(\beta) = \tfrac{1}{4}\mathbb{E}|R_\beta - R'_\beta|.$$

This identity proves Proposition 7. Minimizing over G shows that $J(\beta)$ is the risk over all levels with optimal residual offsets.

□

A.3 THE OPTIMAL SHARED CURVE

We prove Proposition 3 for the shared curve in Equation (1), before final clipping, by symmetrizing its location and offsets. This argument also applies to symmetric discrete learning measures and does not require unique quantiles. Let κ be a probability measure on $(0,1)$ invariant under $\tau \mapsto 1-\tau$,

and let $\mathcal{R}_\kappa$ be population pinball risk integrated against κ. Assume these risks are finite. For an offset function d, define its reflection by $(\mathsf{T}d)(\tau) = -d(1-\tau)$. Conditional symmetry and $\rho_\tau(-r) = \rho_{1-\tau}(r)$ give

$$\mathcal{R}_\kappa(\phi^\top\beta + d) = \mathcal{R}_\kappa\{\phi^\top(2\beta_* - \beta) + \mathsf{T}d\}.$$

By convexity, averaging these two curves cannot increase risk:

$$\mathcal{R}_\kappa\left(\phi^\top\beta_* + \frac{d + \mathsf{T}d}{2}\right) \le \mathcal{R}_\kappa(\phi^\top\beta + d). \tag{18}$$

If d is nondecreasing, so are $\mathsf{T}d$ and their average. With coefficients β_*, expected pinball loss at each level is minimized by a marginal quantile of E. Thus $q_{\mathrm{pool}}(x,\tau) = \phi(x)^\top\beta_* + d_*(\tau)$ minimizes risk over this shared class. Taking κ uniform on B, or on $(0,1)$, proves the proposition.

From all quantile levels to a symmetric band. Apply the same argument to the measure assigning mass $1/2$ to each of τ and $1-\tau$. Every shared curve q has nonnegative excess risk over q_{pool} on this pair. Both the band and its complement consist of such pairs. Integrating gives

$$0 \le \mathcal{R}_B(q) - \mathcal{R}_B(q_{\mathrm{pool}}) \le \frac{\mathcal{R}_{(0,1)}(q) - \mathcal{R}_{(0,1)}(q_{\mathrm{pool}})}{w}. \tag{19}$$

For $|Y| \le 1$, clipping a prediction to $[-1,1]$ cannot increase pinball loss. Thus an upper bound for a shared curve also bounds its clipped output.

A.4 Approximation cost of a shared residual law

Let F_x be the CDF of $Y - \phi(x)^\top\beta_*$ conditional on $X = x$, so $G_* = \mathbb{E}_X F_X$. Write $\mathcal{A}_{(0,1)}$ for the excess risk of q_{pool} over q^* across all levels. The conditional version of Proposition 7 gives

$$\mathcal{A}_{(0,1)} = \frac{1}{2}\int_{\mathbb{R}} \mathrm{Var}_X\{F_X(r)\}\,dr = \frac{1}{4}\mathbb{E}_{X,X'}\|F_X - F_{X'}\|^2_{L^2(\mathbb{R})}, \tag{20}$$

where X' is an independent copy. Since $G_* = \mathbb{E}_X F_X$, the variance equals $\mathbb{E}_X[(F_X(r) - G_*(r))^2]$, as used in Equation (11). At each level the conditional quantile minimizes expected pinball loss, so for a band of width w, $0 \le \mathcal{A}_B \le \mathcal{A}_{(0,1)}/w$.

Proposition 8 (Input-dependent scale)**.** *Suppose $Y = \phi(X)^\top\beta_* + s(X)Z$, with Z independent of X, $S = s(X) \ge s_{\min} > 0$, and $\mathbb{E}S^2 < \infty$. Let Z have a finite first moment and density f_0 with $\kappa_0 = \int_{\mathbb{R}} u^2 f_0(u)^2\,du < \infty$. Then*

$$\mathcal{A}_B \le \frac{2\kappa_0}{w}\,\mathrm{Var}(\sqrt{S}) \le \frac{\kappa_0}{2w s_{\min}}\,\mathrm{Var}(S). \tag{21}$$

Proof. For the CDF F_0 of Z, write $F_s(r) = F_0(r/s)$. Differentiation in s and a change of variables give $\|\partial_s F_s\|^2_{L^2} = \kappa_0/s$. Integrating this derivative and using Minkowski's inequality yields

$$\|F_s - F_t\|_{L^2} \le \sqrt{\kappa_0}\left|\int_t^s v^{-1/2}\,dv\right| = 2\sqrt{\kappa_0}|\sqrt{s} - \sqrt{t}|.$$

The pairwise identity therefore bounds $\mathcal{A}_{(0,1)}$ by $2\kappa_0\,\mathrm{Var}(\sqrt{S})$. Since $|\sqrt{s}-\sqrt{t}| \le |s-t|/(2\sqrt{s_{\min}})$, the variance bound and division by w prove Equation (21).

□

In particular, for fixed $s_0 > 0$ and bounded V, $S = s_0(1+\gamma V)$ gives $\mathcal{A}_B = O(\gamma^2)$ as $\gamma \to 0$.

A.5 Privacy and gradient geometry of Joint

Coefficient coordinates. Write $\phi(x) = (\phi_0, \phi_-(x))$, where $\phi_0 > 0$ is constant, and let β_- denote the coefficients of the nonconstant features. Let ψ_r be the piecewise-linear basis constructed in Appendix B.4, orthonormal under the learning weights and containing a constant. We use

$$\theta = (c, \beta_-), \qquad z(x,\tau) = (\phi_0\psi_r(\tau), \phi_-(x)), \qquad q_\theta(x,\tau) = z(x,\tau)^\top\theta.$$

For a fixed reference level $\tau_0 \in \mathcal{T}$, the location and offset in Equation (1) can be written as

$$\beta = (\psi_r(\tau_0)^\top c, \beta_-), \qquad d(\tau) = \phi_0\{\psi_r(\tau) - \psi_r(\tau_0)\}^\top c.$$

Thus $d(\tau_0) = 0$ fixes the common intercept and $q_\theta = \phi^\top \beta + d$. We use $\tau_0 = 1/2$ in our setting. Both components are linear in θ, with $\nabla_\theta q = z$.

Threshold decomposition. For any $0 \le u_1 \le \cdots \le u_m \le 1$, define

$$\alpha_0 = u_1, \qquad \alpha_l = u_{l+1} - u_l \ (1 \le l < m), \qquad \alpha_m = 1 - u_m.$$

These coefficients are nonnegative and sum to one, and $u_j = \sum_{l<j} \alpha_l$. Consequently

$$\sum_j w_j(u_j - \tau_j) z(x, \tau_j) = \sum_{l=0}^{m} \alpha_l g_l(x) \in \mathcal{G}.$$

Smoothed quantile gradients have this form because the same smoothing CDF is evaluated along a nondecreasing prediction curve. For fixed threshold l, $g_l(x)$ is affine in the feature vector. When the public feature domain is a polytope, its vertices suffice to construct the convex hull.

Proof of Proposition 1. At a fixed current iterate, replacing one record changes the empirical gradient by $(g - g')/n$, with $g, g' \in \mathcal{G}$. With the covariance in Equation (4), its squared Mahalanobis change is at most μ^2/T. The noisy gradient is therefore $\mu/\sqrt{T}$-GDP, conditional on the previous outputs. The proximal map is deterministic postprocessing. Adaptive Gaussian composition over T iterations gives total parameter μ. Taking the fixed last-half average and evaluating its function at any further levels are postprocessing.

Intercept, input, and shape directions. Let P_I, P_B, P_S be orthogonal projectors onto the constant offset, input slopes, and nonconstant offset shape, respectively. They sum to the identity. For data-independent ratios $r_I, r_S > 0$, define

$$R^2(r_I, r_S) = \sup_{g, g' \in \mathcal{G}} \left\{ \frac{\|P_I(g - g')\|^2}{r_I} + \|P_B(g - g')\|^2 + \frac{\|P_S(g - g')\|^2}{r_S} \right\},$$
$$\Omega = R^2(r_I, r_S)(r_I P_I + P_B + r_S P_S).$$

Orthogonality gives $(g - g')^\top \Omega^{-1} (g - g') \le 1$. For fixed ratios, R^2 is the smallest multiplier covering the public set of gradient changes. For a fixed, data-independent reference measure on $\mathcal{X}$, define $W_z = \mathbb{E}_X \sum_j w_j z(X, \tau_j) z(X, \tau_j)^\top$. The average squared prediction perturbation is

$$\mathbb{E}_{X, \xi_t} \sum_j w_j \{z(X, \tau_j)^\top \xi_t\}^2 = \frac{T}{n^2 \mu^2} \operatorname{tr}(W_z \Omega). \tag{22}$$

In terms of the three coefficient groups,

$$\operatorname{tr}(W_z \Omega) = R^2(r_I, r_S)\{r_I \operatorname{tr}(W_z P_I) + \operatorname{tr}(W_z P_B) + r_S \operatorname{tr}(W_z P_S)\}.$$

We minimize this expression over a finite set of ratios specified independently of the training data. It balances the cost of perturbing one intercept, the common input effects, and the offset shape. The construction specializes Gaussian covariance design (Li et al., 2015; Asi et al., 2021; Nikolov & Tang, 2024) to the shared ordered quantile model.

A.6 Prediction risk of joint learning

We first bound prediction error at the training quantile levels, then transfer the bound to the quantile band.

Let $\nu = \sum_j w_j \delta_{\tau_j}$ be Joint's public learning measure, and let $\mathcal{R}_\nu$ denote unsmoothed population pinball risk averaged over this measure. Write $\mathcal{F}_D = L_D + R + I_{\mathcal{K}}$, and choose a measurable empirical minimizer θ_D^*. Denote Joint's smoothing bandwidth by h_J. Let L_J bound the smoothness of L_D, $G_J = \sup_{g \in \mathcal{G}} \|g\|$, and $\Delta_J = \sup_{g, g' \in \mathcal{G}} \|g - g'\|$. For $0 < \eta L_J < 2$, define the initialization term

$$\mathcal{B}_J = \mathbb{E}_D \left[\frac{\|\theta_0 - \theta_D^*\|^2}{\eta} + \frac{2\eta L_J}{2 - \eta L_J} \{\mathcal{F}_D(\theta_0) - \mathcal{F}_D(\theta_D^*)\} \right].$$

For coefficients $u \in \mathcal{K}$, let q_u be the corresponding shared curve before clipping. We use u, fixed independently of the training data and noise, as a reference for the learned predictor.

Lemma 9 (Risk at the training quantile levels)**.** *Assume $|Y| \le 1$, $\mathcal{B}_J < \infty$, and public fixed learning settings. For every fixed $u \in \mathcal{K}$, the averaged output of Equation* (4) *satisfies*

$$\mathbb{E}\mathcal{R}_\nu(\operatorname{clip} q_{\widehat{\theta}}) - \mathcal{R}_\nu(q_u) \le R(u) + \frac{h_J}{4} + \frac{\mathcal{B}_J}{T} + \frac{4\eta T \operatorname{tr}(\Omega)}{(2-\eta L_J)^2 n^2 \mu^2} + \frac{\eta G_J \Delta_J (3T+2)}{4n}. \tag{23}$$

Here clipping is to $[-1, 1]$. The terms account for regularization, smoothing, optimization, privacy noise, and generalization.

Proof. Noisy proximal learning. Condition on D, and put $\delta_t = \theta_{t+1} - \theta_t$. The proximal optimality inequality, convexity of L_D, and L_J-smoothness give

$$\begin{aligned}\mathcal{F}_D(\theta_{t+1}) - \mathcal{F}_D(\theta_D^*) &\le \frac{\|\theta_t - \theta_D^*\|^2 - \|\theta_{t+1} - \theta_D^*\|^2}{2\eta} + \frac{\eta L_J - 1}{2\eta}\|\delta_t\|^2 \\ &\quad - \langle \xi_t, \theta_t - \theta_D^* + \delta_t \rangle, \\ \mathcal{F}_D(\theta_{t+1}) - \mathcal{F}_D(\theta_t) &\le -\frac{2-\eta L_J}{2\eta}\|\delta_t\|^2 - \langle \xi_t, \delta_t \rangle.\end{aligned}$$

Multiply the second inequality by $\eta L_J/(2-\eta L_J)$ and add it to the first. The coefficient of $\|\delta_t\|^2$ becomes $-1/(2\eta)$. Young's inequality bounds this term together with $-2\langle \xi_t, \delta_t \rangle/(2-\eta L_J)$ by $2\eta\|\xi_t\|^2/(2-\eta L_J)^2$. The remaining noise inner product has conditional mean zero. Summing over all T updates telescopes both the squared distance and the objective difference. Since $\mathbb{E}\|\xi_t\|^2 = T \operatorname{tr}(\Omega)/(n^2\mu^2)$, this gives

$$\begin{aligned}\sum_{t=1}^{T} \mathbb{E}_{\text{noise}}[\mathcal{F}_D(\theta_t) - \mathcal{F}_D(\theta_D^*)] &\le \frac{\|\theta_0 - \theta_D^*\|^2}{2\eta} \\ &\quad + \frac{\eta L_J}{2-\eta L_J}\{\mathcal{F}_D(\theta_0) - \mathcal{F}_D(\theta_D^*)\} \\ &\quad + \frac{2\eta T^2 \operatorname{tr}(\Omega)}{(2-\eta L_J)^2 n^2 \mu^2}.\end{aligned}$$

Every summand is nonnegative. Restricting to the last half, dividing by $T/2$, and using convexity at the averaged output bounds its empirical suboptimality by the optimization and privacy terms in Equation (23). This argument applies throughout $0 < \eta < 2/L_J$.

Generalization and pinball risk. Couple the Gaussian increments on neighboring datasets. The gradient step is nonexpansive for $\eta \le 2/L_J$ (Hardt et al., 2016, Lemma 3.6), as is the proximal map. Changing one observation therefore gives $\|\theta_t - \theta_t'\| \le \eta t \Delta_J/n$. The average of t over $\mathcal{I}_T$ is $(3T+2)/4$. Since each smoothed record loss is G_J-Lipschitz on $\mathcal{K}$, replacement stability (Bousquet & Elisseeff, 2002) bounds the expected generalization error by $\eta G_J \Delta_J (3T+2)/(4n)$. Comparing the empirical minimizer with the fixed u adds at most $R(u)$. Uniform smoothing satisfies

$$0 \le \mathbb{E}_{V \sim \text{Unif}[-h_J, h_J]} \rho_\tau(r+V) - \rho_\tau(r) \le h_J/4.$$

Finally, clipping cannot increase pinball loss for $|Y| \le 1$. These observations give the stated population bound. □

Adding $\mathcal{R}_\nu(q_u) - \mathcal{R}_\nu(q^*)$ separates the piecewise-linear shared model's approximation error from learning error. The noise term uses total coefficient variance $\operatorname{tr}(\Omega)$. Equation (22) evaluates the same perturbation in the public prediction metric.

The initialization term is finite for the experimental model, including $\lambda_a = 0$. The inequality $\mathcal{F}_D(\theta_D^*) \le \mathcal{F}_D(\theta_0)$ and $\lambda > 0$ bound the input slopes uniformly. The levels lie in $[a, 1-a]$ for $a > 0$, so $\rho_\tau(r) \ge a|r|$ bounds the weighted absolute offset values. The nonzero constant feature and full-rank basis Gram matrix then bound the offset coefficients. Bounded responses and the public initial curve bound $\mathcal{F}_D(\theta_0)$ uniformly in D.

From training quantile levels to a quantile band. Let $B = [a, 1-a] \subseteq \mathcal{T}$, $w = 1 - 2a$, and suppose all training quantile levels lie in B. Define the discrepancy of their weighted distribution from the uniform band measure by

$$d_m = \sup_{t \in B} \left| \sum_{\tau_j \le t} w_j - \frac{t-a}{w} \right|, \qquad V_B(M) = 2M + (1+M)w.$$

For a nondecreasing curve $q(x, \cdot)$ bounded in absolute value by M, and $|y| \le 1$, the function $f_{x,y}(\tau) = \rho_\tau\{y - q(x,\tau)\}$ has total variation at most $V_B(M)$. For consecutive partition points $s < t$,

$$|f_{x,y}(t) - f_{x,y}(s)| \le |q(x,t) - q(x,s)| + (1+M)(t-s).$$

Summing and using monotonicity proves the bound. Integration by parts against the difference of the two probability measures gives

$$|\mathcal{R}_B(q) - \mathcal{R}_\nu(q)| \le V_B(M) d_m.$$

Joint's clipped output has $M = 1$. For a fixed feasible comparator u, let $M_u = \sup_{x,\tau \in B} |q_u(x,\tau)| < \infty$ and put

$$\mathcal{Q}_m(u) = \{V_B(1) + V_B(M_u)\} d_m.$$

Applying the variation bound to the output and comparator yields

$$\mathbb{E}\mathcal{R}_B(\widehat{q}_J) - \mathcal{R}_B(q_u) \le \mathbb{E}\mathcal{R}_\nu(\widehat{q}_J) - \mathcal{R}_\nu(q_u) + \mathcal{Q}_m(u).$$

Define the cost of the finite offset representation by $\Delta_{r,B}(u) = \mathcal{R}_B(q_u) - \mathcal{R}_B(q_{\mathrm{pool}})$. The decomposition

$$\mathcal{R}_B(q_u) - \mathcal{R}_B(q^*) = \mathcal{A}_B + \Delta_{r,B}(u)$$

separates this representation cost from the cost of sharing.

Theorem 4 (Detailed form)**.** *Assume the population model of Proposition 3, the conditions of Lemma 9, and training quantile levels in B. For every fixed $u \in \mathcal{K}$,*

$$\mathbb{E}\mathcal{R}_B(\widehat{q}_J) - \mathcal{R}_B(q^*) \le \mathcal{A}_B + \Delta_{r,B}(u) + R(u) + \frac{h_J}{4} + \frac{\mathcal{B}_J}{T} + \mathcal{Q}_m(u) + \frac{4\eta T \operatorname{tr}(\Omega)}{(2 - \eta L_J)^2 n^2 \mu^2} + \frac{\eta G_J \Delta_J (3T+2)}{4n}. \tag{24}$$

Proof. Combine Lemma 9 with the variation bound above and add $\mathcal{R}_B(q_u) - \mathcal{R}_B(q^*)$. □

For the chosen u, collect the approximation and regularization terms as

$$\mathcal{E}_J = \Delta_{r,B}(u) + R(u) + \frac{h_J}{4} + \mathcal{Q}_m(u).$$

These are the costs of the finite offset representation, regularization, loss smoothing, and numerical integration, respectively. For fixed geometry, comparator, initialization, and learning rate, with $2 - \eta L_J$ bounded below by a positive constant, the remaining terms give Equation (12). Conditional symmetry makes $\Delta_{r,B}(u) \ge 0$ by Proposition 3. The bound at the training quantile levels requires bounded responses but not conditional symmetry. The finite basis and learning measure determine $\Delta_{r,B}(u)$ and $\mathcal{Q}_m(u)$. These approximation terms remain as sample size increases.

Coefficient sharing under matched calibration. Use the orthonormal basis ψ_r of Appendix B.4 and put $C^2 = \sum_j w_j \max(\tau_j, 1-\tau_j)^2$. For any coefficient matrix A,

$$\sum_j w_j \{\phi(x)^\top A \psi_r(\tau_j)\}^2 = \|A^\top \phi(x)\|^2 \le \|A\|_F^2.$$

Cauchy–Schwarz therefore bounds each record gradient norm by C for Full and its shared restriction, Joint. Both have gradient diameter at most $2C$ and loss smoothness at most $1/(2h_J)$. The matched isotropic choices $\Omega_V = 4C^2 I_{pr}$ and $\Omega_J = 4C^2 I_{p+r-1}$ are valid for the two learners. Substitution in the noise term of Equation (23) gives dimension factors pr and $p + r - 1$, respectively. Each learner retains its own representation and regularization costs.

A.7 PRIVACY OF CDF

Location stage. Let $\chi = (1 + \eta\lambda)^{-1}$ denote the shrinkage factor in Equation (6). The record gradient is $-\phi(x)\,\mathrm{clip}\{(y-\phi(x)^\top\beta)/h, -1, 1\}/2$. Its norm is at most $1/2$, and replacing one record changes the average gradient by at most $1/n$. Conditional on the preceding iterates, neighboring update means differ by at most $\chi\eta/n$, while the Gaussian standard deviation is $\chi\eta\sigma_A$. Each update is thus $1/(n\sigma_A) = \mu_A/\sqrt{T}$-GDP. Adaptive Gaussian composition makes the entire trajectory μ_A-GDP. The fixed average in Equation (7) is postprocessing of that trajectory.

Residual stage. Fix a possible coefficient output β. Its residuals lie in the interval

$$I_\beta = [-1 - \sup_{x\in\mathcal{X}} \phi(x)^\top\beta, 1 - \inf_{x\in\mathcal{X}} \phi(x)^\top\beta]. \tag{25}$$

With lower endpoint a_β and length B_β, the normalized residuals are $u_i = (y_i - \phi(x_i)^\top\beta - a_\beta)/B_\beta$. The projection uses a uniform public grid of $K+1$ points on $[0,1]$, with $K > M$, followed by linear interpolation.

For the Fourier release, define the public quantities

$$H_M = \sum_{k=1}^{M} \frac{1}{k}, \qquad C_M(t) = \sum_{k=1}^{M} \frac{\cos(kt)}{k}, \qquad U_M = \min\{4H_M,\, 2H_M - 2\min_{t\in[0,2\pi]} C_M(t)\}. \tag{26}$$

The same residual determines every coordinate of the released vector.

The interval in Equation (25) contains the residual of every record in the public domain, on either neighboring dataset. Put

$$S_M(u,v) = \sum_{k=1}^{M} \frac{[\cos(\pi ku) - \cos(\pi kv)]^2}{k}.$$

Since the difference of real parts is bounded by the complex difference,

$$[\cos(\pi ku) - \cos(\pi kv)]^2 \le |e^{i\pi ku} - e^{i\pi kv}|^2 = 2 - 2\cos\{\pi k(u-v)\}.$$

Consequently $S_M(u,v) \le 2H_M - 2C_M(\pi(u-v))$. The independent bound $|\cos s - \cos t| \le 2$ also gives $S_M(u,v) \le 4H_M$. Combining the two proves $S_M(u,v) \le U_M$. For the covariance in Equation (9), the squared Mahalanobis change of the empirical coefficient mean is

$$\sum_{k=1}^{M} \frac{\{[c_k(u) - c_k(v)]/n\}^2}{\sigma_k^2} = \frac{\mu_C^2}{U_M} S_M(u,v) \le \mu_C^2.$$

The Gaussian mechanism gives μ_C-GDP conditional on every first-stage output. Adaptive composition with the location stage proves Proposition 2. Projection, interpolation, rescaling, and any subsequent quantile evaluation are postprocessing.

Computing the public constant. Differentiating C_M yields

$$C_M'(t) = -\sum_{k=1}^{M} \sin(kt) = -\frac{\sin(Mt/2)\sin((M+1)t/2)}{\sin(t/2)}.$$

Thus its minimum on $[0, 2\pi]$ is attained among the endpoints and the points $2\pi j/M$ and $2\pi j/(M+1)$ within this interval. The coordinate-bound mechanism uses $\sigma_{k,\square}^2 = 8H_M/(\pi^2 n^2 \mu_C^2 k)$, so the ratio of standard deviations is $\sqrt{U_M/(4H_M)}$ in every coordinate. It equals 0.78, 0.77, and 0.76 for $M = 16, 32, 64$.

A.8 PROOF OF THE HUBER-TO-QUANTILE TRANSFER

Write f_x for the density of $E \mid X = x$ and $\bar{f} = \mathbb{E}_X f_X$ for the pooled density. The theorem assumes $\|\bar{f}\|_\infty \le 1/(2h)$.

Examples of the conditions. Conditional symmetry makes the center independent of the admissible bandwidth h. Conditional uniform noise on $[-s(x), s(x)]$, with $s(x) \geq h$, satisfies the theorem. Conditional Gaussian noise with standard deviation at least $h\sqrt{2/\pi}$ does as well. A common bound on the conditional densities is sufficient for the marginal density bound, but is not necessary.

Population minimizers. Let (X', E') be an independent copy of (X, E), and write $v(x) = \phi(x)^\top(\beta_* - \beta)$. Then the residuals in Appendix A.2 satisfy $R_\beta = E + v(X)$ and $R'_\beta = E' + v(X')$. For fixed $X = x, X' = x'$, the difference $E - E'$ has a symmetric conditional law. Its expected absolute loss is an even convex function of the shift $v(x) - v(x')$, minimized at zero. Averaging shows that β_* minimizes J. The same argument with the even convex Huber loss shows that it minimizes H_P. Neither argument requires independence between a residual and its covariates.

Conditional smoothing and central probabilities. We compare probabilities of shifted intervals using symmetry and unimodality (Gupta, 1983). For an independent $U \sim \mathrm{Unif}[-h, h]$,

$$\ell_h(r) = \tfrac{1}{2}\mathbb{E}_U|r + U| - h/4.$$

This is the uniform convolution representation of the Huber loss (He et al., 2023), with the normalization in Equation (5). The marginal density bound gives

$$\mathbb{P}(|E'| \leq t) \leq \min\{2t\|\bar{f}\|_\infty, 1\} \leq \min\{t/h, 1\} = \mathbb{P}(|U| \leq t).$$

For fixed x, t, put $\Psi_{x,t}(z) = \mathbb{P}(|E+z| \leq t \mid X = x)$. The probability $\Psi_{x,t}$ is even and nonincreasing in $|z|$. It is locally absolutely continuous, and for almost every $z \geq 0$,

$$\Psi'_{x,t}(z) = f_x(t + z) - f_x(|t - z|) \leq 0.$$

The stochastic ordering of $|U|$ and $|E'|$ therefore implies

$$\mathbb{P}(|E + U| \leq t \mid X = x) = \mathbb{E}_U\Psi_{x,t}(U) \geq \mathbb{E}_{E'}\Psi_{x,t}(-E') = \mathbb{P}(|E - E'| \leq t \mid X = x).$$

Here E' remains a marginal copy independent of the current (X, E); the convolution kernel is the conditional density f_x.

For a symmetric integrable continuous random variable Z, $\mathbb{E}|Z + z| - \mathbb{E}|Z| = \int_0^{|z|} \mathbb{P}(|Z| \leq t)\, dt$. Together with the Huber representation, this gives

$$\mathbb{E}[\ell_h(E + z) - \ell_h(E) \mid X = x] \geq \tfrac{1}{2}\mathbb{E}[|E - E' + z| - |E - E'| \mid X = x].$$

Applying this inequality at the fixed conditional shift $z = v(x)$ and averaging over X gives

$$\tfrac{1}{2}\{\mathbb{E}|R_\beta - E'| - \mathbb{E}|E - E'|\} \leq H_P(\beta) - H_P(\beta_*).$$

From a shift to a residual distribution. The one-dimensional CDF identity is

$$2\mathbb{E}|R_\beta - E'| - \mathbb{E}|R_\beta - R'_\beta| - \mathbb{E}|E - E'| = 2\,\|G_\beta - G_*\|^2_{L^2(\mathbb{R})}\,.$$

Consequently,

$$\begin{aligned} J(\beta) - J(\beta_*) + \tfrac{1}{2}\,\|G_\beta - G_*\|^2_{L^2(\mathbb{R})} &= \tfrac{1}{2}\{\mathbb{E}|R_\beta - E'| - \mathbb{E}|E - E'|\} \\ &\leq H_P(\beta) - H_P(\beta_*). \end{aligned} \tag{27}$$

Dropping the nonnegative CDF term in Equation (27) and applying Proposition 7, with $J(\beta_*) = \mathcal{R}_{(0,1)}(q_{\mathrm{pool}})$, gives the bound over all levels

$$\mathcal{R}_{(0,1)}(q) - \mathcal{R}_{(0,1)}(q_{\mathrm{pool}}) \leq H_P(\beta) - H_P(\beta_*) + \tfrac{1}{2}\,\|G - G_\beta\|^2_{L^2(\mathbb{R})}\,. \tag{28}$$

Apply Equation (19) and add $\mathcal{A}_B$ to obtain Equation (13). The first-moment assumptions make these expectations finite. Also, $\int (G_\beta - G_*)^2 \leq \int |G_\beta - G_*| \leq \mathbb{E}|v(X)|$. The result applies pointwise to fitted coefficients before taking expectation over the training data and noise.

Sharpness. Take $E \sim \mathrm{Unif}[-s, s]$, $s > h$, and an independent $v(X) = \pm a$ with equal probabilities, where $0 < a \leq s - h$. Direct integration gives

$$H_P(\beta) - H_P(\beta_*) = \frac{a^2}{4s}, \qquad J(\beta) - J(\beta_*) = \frac{a^2}{4s} - \frac{a^3}{12s^2}.$$

Their ratio tends to one as $a \downarrow 0$. A bounded scalar covariate $X = \pm 1$ realizes this displacement. Taking $G = G_\beta$ proves sharpness of the Huber coefficient in the bound over all levels (28). At $\beta = \beta_*$, Proposition 7 gives equality in that bound for any G. Choosing $G \neq G_*$ proves sharpness of the coefficient $1/2$ on the CDF error.

Role of unimodality. Let $X \sim \mathrm{Unif}[-1,1]$, and independently let E be an equal mixture of uniforms on $[-a-w,-a]$ and $[a,a+w]$, where $a > h$, $w \geq h$, and $a + w < 1$. This noise is symmetric and has density bounded by $1/(2h)$. Choose a location whose residual is $E + dX$, with $0 < |d| \leq \min\{a-h, w/2\}$, and denote its coefficient by β_d. The residual stays on the same linear branch of the Huber loss, so

$$H_P(\beta_d) - H_P(\beta_*) = 0, \qquad J(\beta_d) - J(\beta_*) = \frac{d^2}{12w} - \frac{|d|^3}{30w^2} > 0.$$

The second identity follows from the triangular law of the residual differences within a mixture component and $J(\beta_d) = \mathbb{E}|R_{\beta_d} - R'_{\beta_d}|/4$. Thus symmetry and a marginal density bound alone do not give the excess-risk transfer.

A.9 Finite-sample location risk

We separate optimization error, Gaussian noise, regularization bias, and the generalization cost of using the training data. Assume $|Y| \leq 1$. Let $\mathcal{F}_D(\beta) = H_D(\beta) + \lambda \|\beta\|_2^2/2$, let $\beta_{\lambda,D}$ minimize $\mathcal{F}_D$, and suppose $\|\beta_*\|_2 \leq B_0$. The gradient bound and comparison with zero give

$$\|\beta_{\lambda,D}\|_2 \leq B_\lambda = \min\left\{\frac{1}{2\lambda}, \sqrt{\frac{2\ell_h(1)}{\lambda}}\right\}.$$

For $h \leq 1$, $\ell_h(1) = 1/2 - h/4$. For each $1 \leq t \leq T$, let v_t bound accumulated noise, r_t bound mean squared distance to the empirical minimizer, and s_t bound sensitivity to one record:

$$v_t = p(\eta\chi\sigma_A)^2 \frac{1-\chi^{2t}}{1-\chi^2}, \qquad r_t = \chi^{2t} B_\lambda^2 + v_t, \qquad s_t = \frac{1-\chi^t}{\lambda n}. \tag{29}$$

With $L = 1/(2h)$, the regularized objective has smoothness $L + \lambda$. Put $S_T = 2T^{-1}\sum_{t\in\mathcal{I}_T} s_t$.

Lemma 10 (Location learning error)**.** *For $0 < \eta \leq 2/L$, the averaged estimate in Equation (7) satisfies*

$$\mathbb{E}[H_P(\widehat{\beta}) - H_P(\beta_*)] \leq \frac{L+\lambda}{T}\sum_{t\in\mathcal{I}_T} r_t + \frac{\lambda B_0^2}{2} + \frac{S_T}{2}. \tag{30}$$

Proof. Empirical error. For $\eta \leq 2/L$, the gradient step on H_D is nonexpansive (Hardt et al., 2016, Lemma 3.6). Scaling by χ makes the noiseless update χ-contractive, with fixed point $\beta_{\lambda,D}$. The independent centered Gaussian increment then gives

$$\mathbb{E}[\|\beta_{t+1} - \beta_{\lambda,D}\|_2^2 \mid \beta_t, D] \leq \chi^2 \|\beta_t - \beta_{\lambda,D}\|_2^2 + p(\eta\chi\sigma_A)^2.$$

Iterating from zero bounds the mean squared distance by r_t. Smoothness bounds the expected objective gap at β_t by $(L+\lambda)r_t/2$. Convexity then bounds the gap at the suffix average by $(L+\lambda)T^{-1}\sum_{t\in\mathcal{I}_T} r_t$. Comparing the empirical minimum with β_* adds at most $\lambda B_0^2/2$ after dropping the penalty on $\widehat{\beta}$.

Stability. Couple the Gaussian increments identically on neighboring datasets. The parameter difference satisfies $d_{t+1} \leq \chi d_t + \chi\eta/n$, hence $d_t \leq s_t$. Averaging gives parameter difference at most S_T and loss difference on a new record at most $S_T/2$. Replacement-sample stability (Bousquet & Elisseeff, 2002) bounds the expected generalization error by this quantity. The fixed comparator β_* has equal expected empirical and population risks. Combining these terms proves Equation (30).

□

Let β_t^0 be the noiseless iterates from the same data and initial point, and put $\widehat{\beta}^0 = 2T^{-1}\sum_{t\in\mathcal{I}_T}\beta_t^0$. Coupling the noisy and noiseless trajectories with the contraction argument of Lemma 10 gives

$$\mathbb{E}_{\mathrm{noise}}[\left\|\widehat{\beta} - \widehat{\beta}^0\right\|_2^2 \mid D] \leq \frac{2}{T}\sum_{t\in\mathcal{I}_T} v_t \leq v_T. \tag{31}$$

This bound holds for every dataset D.

Expected residual range. Let $D_\phi = \sup_{x,x'} \|\phi(x) - \phi(x')\|_2 \le 2$. The noiseless trajectory decreases $\mathcal{F}_D$, since $\eta\chi < 2/(L+\lambda)$, so its norm is at most $\sqrt{2\ell_h(1)/\lambda}$. The bounded-gradient recursion also bounds every noiseless iterate by $1/(2\lambda)$, so $\|\widehat{\beta}^0\|_2 \le B_\lambda$. Equation (31) and Minkowski's inequality give

$$\left\|B_{\widehat{\beta}}\right\|_{L^2} \le \overline{B}_T := 2 + D_\phi(B_\lambda + \sqrt{v_T}). \tag{32}$$

Thus the fitted residual range has finite first and second moments.

A.10 Proof of residual CDF stability

Prediction stability controls the effect of fitting the coefficients and estimating their residual distribution on the same training data (Bousquet & Elisseeff, 2002; Steinberger & Leeb, 2023).

Lemma 11 (Residual CDF stability)**.** *Let f_D be trained on an i.i.d. sample, with randomness independent of the data. Suppose that coupling this randomness on neighboring datasets gives $\sup_x |f_D(x) - f_{D'}(x)| \le S$ almost surely. Let F_D be the empirical CDF of $y_i - f_D(x_i)$, and let F_P^D be the conditional CDF of $Y - f_D(X)$ for a fresh observation. Suppose a measurable interval I_D contains both the training residuals and the support of the residual distribution for a fresh observation almost surely. For $B_D = |I_D|$ with finite expectation,*

$$\mathbb{E}\left\|F_D - F_P^D\right\|^2_{L^2(\mathbb{R})} \le \frac{\mathbb{E}B_D}{4n} + 4S. \tag{33}$$

Proof. Condition on the data-independent randomness used by the predictor. Replace observation i with an independent copy, obtaining $D^{(i)}$. For each original observation j, define

$$g_j^D(r) = \mathbf{1}\{y_j - f_D(x_j) \le r\} - F_P^D(r).$$

We evaluate $g_i^{D^{(i)}}$ on the original observation i. For two real thresholds, the integral of the absolute difference of their indicators equals the distance between them. Prediction stability therefore implies

$$\left\|g_j^D - g_j^{D^{(i)}}\right\|_{L^1(\mathbb{R})} \le 2S.$$

One term accounts for the empirical indicator and the other for the CDF for a fresh observation. Since $|g_j| \le 1$, changing both factors in $g_i g_j$ changes its integral by at most $4S$.

For $j \ne i$, condition on the $n-1$ other observations, the independent replacement, and the predictor randomness. Under this conditioning, $f_{D^{(i)}}$ and $g_j^{D^{(i)}}$ are fixed, whereas $\mathbb{E}[g_i^{D^{(i)}}(r)] = 0$. Thus the replaced cross term has zero expectation. For $j = i$, conditional independence gives

$$\mathbb{E}\int (g_i^{D^{(i)}}(r))^2\, dr = \mathbb{E}\int F_P^{D^{(i)}}(r)(1 - F_P^{D^{(i)}}(r))\, dr \le \frac{\mathbb{E}B_{D^{(i)}}}{4}.$$

The modified dataset $D^{(i)}$ has the same distribution as D. Expanding the squared norm of $n^{-1}\sum_i g_i^D$ gives n diagonal terms and $n(n-1)$ cross terms. Their sum is at most $\mathbb{E}B_D/(4n) + 4S$, proving Lemma 11. The finite expected range and the integrated stability bounds justify the exchanges of expectation and integration. □

A.11 Finite approximation error of the residual CDF

The CDF error has four sources: discretizing the coefficients, truncating the series, adding privacy noise, and interpolating the projected values. On the normalized interval $[0,1]$, use knots $t_j = j/K$, $j = 0, \dots, K$, with $M < K$. For $z_k = \pi k/(2K)$, define

$$b_k = \frac{\sqrt{2}}{2K} + \frac{\sqrt{2}}{\pi k}\frac{z_k^2}{6 - z_k^2}, \qquad \mathcal{A}_{M,K} = \sum_{k=1}^{M} b_k^2,$$

$$\mathcal{T}_{M,K} = \sum_{k=M+1}^{K-1} \frac{1}{2K^2 \sin^2(\pi k/(2K))}.$$

Lemma 12 (Projected Fourier CDF)**.** *For any fixed dataset and first-stage output,*

$$\mathbb{E}_{\mathrm{CDF}} \left\| \widehat{G} - G_{D,\beta} \right\|_{L^2(\mathbb{R})}^2 \le B_\beta \mathcal{C}_{n,M,K},$$

$$\mathcal{C}_{n,M,K} = \left[\sqrt{\mathcal{A}_{M,K} + \mathcal{T}_{M,K} + \frac{2U_M H_M}{\pi^2 n^2 \mu_C^2}} + \frac{1}{\sqrt{3K}} \right]^2. \tag{34}$$

Here $G_{D,\beta}$ is the empirical residual CDF and the expectation is over the CDF noise.

Proof. Coefficient error. Let F be any empirical CDF, including one with atoms at either endpoint. Put $v_j = F(t_j)$ for interior knots and fix $v_0 = 0, v_K = 1$. The sine vectors $e_k(t_j) = \sqrt{2}\sin(\pi k t_j)$, $1 \le k < K$, are orthonormal under $K^{-1}\sum_{j=1}^{K-1}$. For a point mass at u, put $l = \max\{1, \min\{K, \lceil Ku \rceil\}\}$ and $w_u = (l - 1/2)/K$. Summing the sine progression from l to $K-1$ and subtracting the coefficient of the baseline t_j gives the exact discrete coefficient

$$a_k(u) = \frac{\sqrt{2}\cos(\pi k w_u)}{2K \sin(\pi k/(2K))}.$$

The inequalities $|u - w_u| \le 1/(2K)$ and $0 \le z/\sin z - 1 \le z^2/(6 - z^2)$ imply $|c_k(u) - a_k(u)| \le b_k$. Let $a_k(F)$ be the average of $a_k(u)$ over the empirical residuals. Using $|\cos(\pi k w_u)| \le 1$ and Jensen's inequality gives $\sum_{k=M+1}^{K-1} a_k(F)^2 \le \mathcal{T}_{M,K}$. The bound $\mathcal{T}_{M,K} \le 1/(2M)$ follows from $\sin(\pi k/(2K)) \ge k/K$. Orthonormality and centered coefficient noise therefore bound the expected squared knot error of the raw Fourier reconstruction by $V = \mathcal{A}_{M,K} + \mathcal{T}_{M,K} + \sum_{k=1}^{M} \sigma_k^2$.

Projection and interpolation. The true knot vector belongs to the closed convex set of bounded nondecreasing sequences. Projecting the Fourier reconstruction onto this set cannot increase its distance to the true vector (Tao & Sarwate, 2026). If F_{lin} interpolates the true knot values and $\widehat{F}$ interpolates the projected values, integrating the linear error on each interval gives $\mathbb{E}\left\| \widehat{F} - F_{\mathrm{lin}} \right\|_{L^2}^2 \le V$. For a monotone CDF on one interval, the squared error of endpoint interpolation is bounded by one third of the interval length times the squared CDF increment. The increments are nonnegative and sum to one, including endpoint atoms. Consequently $\|F - F_{\mathrm{lin}}\|_{L^2}^2 \le 1/(3K)$. Minkowski's inequality and rescaling to I_β give Equation (34). □

A.12 Prediction error of CDF

We give the detailed form of Theorem 6 by combining Huber-to-quantile transfer with CDF estimation and residual stability. Write $\mathcal{E}_{\mathrm{loc}}(T) = \mathbb{E}[H_P(\widehat{\beta}) - H_P(\beta_*)]$, and use $S_T = 2T^{-1}\sum_{t \in \mathcal{I}_T} s_t$ and $\overline{B}_T \ge \mathbb{E}B_{\widehat{\beta}}$ from Appendix A.9.

Theorem 6 (Detailed form)**.** *Under the distributional assumptions of Theorem 5, suppose $|Y| \le 1$ and use the updates in Equation* (6) *with $0 < \eta \le 4h$. CDF's final prediction satisfies*

$$\mathbb{E}\mathcal{R}_B(\widehat{q}_C) - \mathcal{R}_B(q^*) \le \mathcal{A}_B + \frac{1}{w}\left[\mathcal{E}_{\mathrm{loc}}(T) + \tfrac{1}{2}\mathbb{E}\left\| \widehat{G} - G_{\widehat{\beta}} \right\|_{L^2(\mathbb{R})}^2 \right]$$

$$\le \mathcal{A}_B + \frac{1}{w}\left[\mathcal{E}_{\mathrm{loc}}(T) + \overline{B}_T \left(\mathcal{C}_{n,M,K} + \frac{1}{4n} \right) + 4S_T \right]. \tag{35}$$

Here $\mathcal{C}_{n,M,K}$ is defined in Lemma 12, with $M < K$. For the main-text summary, fix $h, \lambda > 0$, $0 < \eta \le 4h$, $w > 0$, and a bound $\|\beta_*\|_2 \le B_0$, and take $K \ge 2M$. Then Equation (14) *holds with*

$$B_C = 2 + 2\left(B_\lambda + \frac{1}{\sqrt{\lambda(2 + \eta\lambda)}} \right), \qquad \mathcal{E}_C = \frac{\lambda B_0^2}{2w} + \frac{2B_C}{wM}.$$

The constants in $O(\cdot)$ depend only on the fixed h, λ, η, B_0, w, not on $n, p, T, M, K, \mu_A, \mu_C$.

Proof. Apply Theorem 5 at the fitted coefficients and CDF, then take expectations. Final clipping cannot increase pinball risk for $|Y| \le 1$, so the first inequality holds for $\widehat{q}_C$. Here $G_{\widehat{\beta}}$ is the population residual CDF for a fresh observation, conditional on the fitted coefficients.

Lemma 12 and Lemma 11 give

$$\mathbb{E}\left\|\widehat{G} - G_{D,\widehat{\beta}}\right\|_{L^2}^2 \le \overline{B}_T \mathcal{C}_{n,M,K}, \qquad \mathbb{E}\left\|G_{D,\widehat{\beta}} - G_{\widehat{\beta}}\right\|_{L^2}^2 \le \frac{\overline{B}_T}{4n} + 4S_T.$$

Minkowski's inequality on the joint probability space and residual axis yields

$$\mathbb{E}\left\|\widehat{G} - G_{\widehat{\beta}}\right\|_{L^2}^2 \le \left[\sqrt{\overline{B}_T \mathcal{C}_{n,M,K}} + \sqrt{\overline{B}_T/(4n) + 4S_T}\right]^2. \tag{36}$$

The stability term accounts for dependence between the fitted location and its training residuals.

Using $(\sqrt{x} + \sqrt{y})^2 \le 2x + 2y$ in Equation (36) gives the second inequality in Equation (35). Equation (30) bounds $\mathcal{E}_{\mathrm{loc}}(T)$ for $0 < \eta \le 4h$. The shared approximation cost $\mathcal{A}_B$ is already normalized by the band width.

Main-text summary. Put $v = \eta T p/(n^2 \mu_A^2)$. Equation (29) gives $v_t \le v/\{\lambda(2+\eta\lambda)\}$, $S_T \le 1/(\lambda n)$, and

$$\frac{1}{T}\sum_{t \in \mathcal{I}_T} r_t \le \frac{B_\lambda^2 \chi^T}{T(1-\chi^2)} + \frac{v}{2\lambda(2+\eta\lambda)}.$$

For fixed h, λ, η, Equation (30) therefore yields

$$\mathcal{E}_{\mathrm{loc}}(T) \le \frac{\lambda B_0^2}{2} + O(T^{-1} + n^{-1} + v).$$

If $v \le 1$, Equation (32) gives $\overline{B}_T \le B_C$. For $K \ge 2M$, $b_k \le 1/K$, hence $\mathcal{A}_{M,K} \le 1/(4M)$ and $\mathcal{T}_{M,K} \le 1/(2M)$. It follows that

$$\mathcal{C}_{n,M,K} \le \frac{2}{M} + \frac{U_M H_M}{n^2 \mu_C^2} \le \frac{2}{M} + \frac{U_M \log(eM)}{n^2 \mu_C^2}.$$

Substitution in Equation (35) proves the summary for $v \le 1$. If $v > 1$, bounded responses and final predictions give excess risk at most two, which is absorbed into $O(v)$. □

B Experimental Setting

B.1 Datasets and public feature maps

The six regression tasks are California Housing, Concrete, Energy Efficiency, Airfoil Self-Noise, Power Plant, and Real Estate (Pace & Barry, 1997; Yeh, 1998; Tsanas & Xifara, 2012; Brooks et al., 1989; Tüfekci, 2014; Yeh, 2018). Energy uses heating load. The main comparison uses one fixed 60/15/25 training–validation–test partition per dataset. Hyperparameters are selected using training fits and validation losses. The final fit uses the combined training and validation data (75%). Table 3 lists this final training size. The test partition is shared by all methods.

Table 3: Dataset sizes and dimensions of the public feature maps. Dimensions include the constant feature.

Dataset	Training	Test	Linear	Additive
California	15,480	5,160	9	89
Concrete	772	258	9	89
Energy	576	192	15	75
Airfoil	1,127	376	6	56
Power Plant	7,176	2,392	5	45
Real Estate	310	104	7	67

All preprocessing uses fixed ranges and categories, given in Table 4. A continuous variable in the linear map is centered at its public midpoint, scaled, and clipped to $[-1, 1]$. In the additive map, each continuous variable is divided into 12 equal-width public bins, represented by orthonormal contrasts. Values outside the range enter the first or last bin. Categories use the same contrast construction in

both maps. We concatenate these representations and a constant, then divide by their public maximum norm. This guarantees $\|\phi(x)\|_2 \le 1$ throughout the domain. The additive representation is a sum of binned univariate effects.

For a response range $[L_Y, U_Y]$, we use the normalization $y = 2(y_{\text{raw}} - L_Y)/(U_Y - L_Y) - 1$ in both training and evaluation. The specified ranges contain every response in the six datasets. Predictions from every method are clipped to $[-1, 1]$, evaluated at 513 equally spaced levels in $[0.05, 0.95]$, and rearranged in increasing order. Fixed-level regressions are linearly interpolated to this grid. For Joint and CDF, monotonicity already follows from the fitted offset constraint and CDF inversion, respectively.

Table 4: Public domains in the original units. Input intervals follow the dataset's feature order, with Energy's continuous and categorical variables listed separately.

Dataset	Input domains	Response
California	$[0, 20]$, $[0, 100]$, $[0, 20]$, $[0, 5]$, $[0, 10000]$, $[0, 15]$, $[30, 45]$, $[-130, -110]$	$[0, 5.1]$
Concrete	$[0, 600]$, $[0, 400]$, $[0, 400]$, $[0, 300]$, $[0, 50]$, $[0, 1200]$, $[0, 1200]$, $[0, 365]$	$[0, 100]$
Energy	Continuous $X_1, X_2, X_3, X_4, X_5, X_7$: $[.5, 1]$, $[400, 1000]$, $[100, 500]$, $[50, 300]$, $[0, 10]$, $[0, .5]$; $X_6 \in \{2, 3, 4, 5\}$, $X_8 \in \{0, \dots, 5\}$	$[0, 60]$
Airfoil	$[0, 25000]$, $[0, 30]$, $[0, .5]$, $[0, 100]$, $[0, .1]$	$[80, 160]$
Power Plant	$[0, 40]$, $[20, 90]$, $[980, 1050]$, $[0, 110]$	$[400, 520]$
Real Estate	$[2012, 2014]$, $[0, 50]$, $[0, 10000]$, $[0, 15]$, $[24.8, 25.2]$, $[121.3, 121.7]$	$[0, 150]$

B.2 Baseline settings and privacy budgets

All methods use replacement adjacency and the same total (ε, δ), with $\delta = n^{-2}$. Gaussian mechanisms use GDP calibration. Sep-Out divides the squared budget μ^2 equally among its fitted levels. Full composes its gradient queries over the full horizon. Joint uses its full GDP budget μ for joint learning. Sep-Obj composes the published objective-perturbation guarantees of its fixed-level fits. CDF uses $\mu_A^2 + \mu_C^2 = \mu^2$ for its two stages on the same training data. The conversion from μ to (ε, δ) is given in Appendix A.1.

CDF. The location loss and noisy updates are Equations (5) and (6). The CDF has $K = 1024$ intervals and Fourier order $M \in \{16, 32, 64\}$. Its range is Equation (25), computed over the public feature domain conditional on the private location.

Joint. Joint uses the shared model in Section 3.1, with $r = 8$ equally spaced knots, including the endpoints of $\mathcal{T} = [0.05, 0.95]$, for its nondecreasing piecewise-linear offset. It minimizes smoothed pinball loss at 63 training quantile levels on $[0.05, 0.95]$, chosen by Gauss–Legendre quadrature with weights summing to one. Validation selects the offset-to-input penalty ratio $\lambda_a/\lambda \in \{0, 0.1, 1\}$. Appendix B.4 specifies the coefficient penalties and gradient-noise covariance.

Sep-Out. This method fits ridge-regularized smoothed quantile regressions at $m \in \{5, 9, 31\}$ equally spaced levels. It adds Gaussian output noise to each fitted coefficient vector and linearly interpolates between fitted levels. Let D_τ be the diameter of the record gradients over the public feature domain. The coefficient sensitivity is bounded by $(D_\tau/n + 2\zeta)/\lambda$, including an optimization gradient error bound $\zeta = 5 \times 10^{-6}$. The standard deviation of the output noise is $\sqrt{m}(D_\tau/n + 2\zeta)/(\lambda\mu)$. The m Gaussian releases compose to μ-GDP.

Sep-Obj. Using the same smoothed loss and candidate level counts as Sep-Out, we apply the published fixed-level objective-perturbation procedure (Chen & Chua, 2023, Algorithm 2) independently at each level. Its ridge coefficient is raised when necessary to satisfy the lower bound required by the privacy guarantee. Each level receives equal privacy parameters. The total budget and level

count determine whether basic or advanced composition requires less noise. The guarantee under replacement adjacency uses the same total (ε, δ) as the other methods. Both fixed-level methods use L-BFGS-B, with at most 3000 iterations.

Full. This joint model uses Joint's quantile basis, integrated loss, and candidate penalty ratios, but lets input coefficients vary with τ. Isotropic Gaussian noise perturbs each full gradient, followed by a proximal penalty step. After selecting the learning settings, validation nIPL chooses the final iterate or the last-half average. Appendix B.4 gives the joint-learning objectives, updates, and noise calibration.

Validation selection. All methods use $\lambda \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ and smoothing scale $a \in \{0.025, 0.05, 0.1\}$. CDF, Joint, and Full use uniform smoothing with bandwidth a, denoted h for CDF and h_J for Joint. The other methods use logistic smoothing with bandwidth a/π, matching the smoothing variance. The three gradient-based methods select $T \in \{128, 512, 2048, 8192\}$ and $\eta \in \{1, 1.8\}/L$, where L bounds the smoothness of the unregularized loss in the model coordinates. For CDF, $L = 1/(2h)$. Fixed-level methods select m instead of T.

The three gradient-based methods use three repetitions to shortlist two learning settings per iteration count, followed by ten repetitions for selection. Joint and Full apply this rule within each penalty ratio. CDF initially uses $f = 0.3, M = 32$. Its two best settings are compared at $f \in \{0.1, 0.3, 0.5\}$ and $M \in \{16, 32, 64\}$, again using ten repetitions. Sep-Out and Sep-Obj use ten repetitions for all candidate settings. Final selection minimizes validation nIPL separately for each dataset, feature map, and privacy budget. The fixed-partition comparisons use 100 noise repetitions at the selected settings.

Appendix E specifies the learning-stage comparisons.

Comparing residual distributions. Polynomial CDF (Tao & Sarwate, 2026) perturbs raw residual moments and transforms them to Legendre coefficients, following Algorithm 1 of that work. For M coefficients, its replacement sensitivity is $\sqrt{4\lceil M/2 \rceil + \lfloor M/2 \rfloor}/n$. Histogram perturbs a vector of bin proportions with sensitivity $\sqrt{2}/n$. It projects the noisy proportions onto the affine set with total mass one, then forms a cumulative sum and linearly interpolates between bin boundaries. Both use Gaussian noise with the same CDF budget μ_C as CDF.

All three estimators use the same private location, residual range, and bounded isotonic CDF projection. Their representation sizes are selected separately by mean validation nIPL over ten noise repetitions: $M \in \{16, 32, 64\}$ for Fourier, $M \in \{2, 4, 6, 8, 10\}$ for Polynomial CDF, and $16, 32, 64, 128,$ or 256 bins for Histogram. The location fit and budget split are fixed across these comparisons.

B.3 Losses, intervals, and averaging

All losses use the normalized response. For test pairs (x_i, y_i) and levels τ_j, we report

$$\mathrm{nIPL} = \frac{1}{n_{\mathrm{test}}\, 513} \sum_i \sum_{j=1}^{513} \rho_{\tau_j}\{y_i - \widehat{q}(x_i, \tau_j)\}.$$

For nominal coverage $1 - \alpha$, the central interval has endpoints $l_i = \widehat{q}(x_i, \alpha/2)$ and $u_i = \widehat{q}(x_i, 1 - \alpha/2)$. The normalized interval score is

$$\mathrm{nIS}_{1-\alpha} = \frac{1}{n_{\mathrm{test}}} \sum_i \left[(u_i - l_i) + \frac{2}{\alpha}(l_i - y_i)_+ + \frac{2}{\alpha}(y_i - u_i)_+ \right].$$

Width is the average $u_i - l_i$, and coverage is the fraction with $l_i \le y_i \le u_i$.

An overall result first averages the six datasets and two maps within each noise repetition. A dataset result averages its two maps. Means and sample standard deviations are then computed across the 100 repetitions. Privacy budgets are kept separate. Losses and widths are multiplied by 100 in tables and figures; coverage is reported in percent.

B.4 QUANTILE BASES AND GEOMETRY FOR JOINT LEARNING

Let $(\tau_j, w_j)_{j=1}^m$, with $m = 63$, be the Gauss–Legendre rule on $[0.05, 0.95]$, normalized so that $\sum_j w_j = 1$. For r equally spaced knots, let $b_r(\tau)$ be the vector of piecewise-linear hat functions. Define

$$Q_r = \sum_j w_j b_r(\tau_j) b_r(\tau_j)^\top, \qquad \psi_r(\tau) = Q_r^{-1/2} b_r(\tau).$$

Thus $\sum_j w_j \psi_r(\tau_j)\psi_r(\tau_j)^\top = I_r$. Joint and Full use the uniformly smoothed composite loss in Equation (2).

Joint. Joint uses $r = 8$ in Section 3.1. Use the coordinates $\theta = (c, \beta_-)$ and their mapping to β, d from Appendix A.5. The feasible set $\mathcal{K}$ requires the knot values $Q_8^{-1/2} c$ to be nondecreasing, equivalently making d nondecreasing. The input coefficients β_- are unconstrained. The objective is $L_D(\theta) + \lambda_a \|c\|^2/2 + \lambda \|\beta_-\|^2/2$. The proximal step in Equation (4) divides the offset and input blocks by $1 + \eta\lambda_a$ and $1 + \eta\lambda$, respectively, then projects onto $\mathcal{K}$ in the Euclidean coefficient norm. The initial curve is $2\tau - 1$, independent of x.

Let $e = \sum_j w_j \psi_8(\tau_j)$. Then $\|e\| = 1$ and $e^\top \psi_8(\tau) = 1$. In (c, β_-) coordinates, the three projectors in Appendix A.5 are

$$P_I = \mathrm{diag}(ee^\top, 0), \quad P_B = \mathrm{diag}(0, I_{p-1}), \quad P_S = \mathrm{diag}(I_8 - ee^\top, 0).$$

We choose both covariance ratios from $\{2^{-6+j/8} : j = 0, \ldots, 96\}$ to minimize $\mathrm{tr}(W_z\Omega)$. For W_z, we use a fixed product reference measure: uniform over the continuous domain and uniform over categorical levels or additive bins. The ratios, reference measure, and covariance are all determined without using the training data.

Public-domain calculation. Every vertex of the normalized feature domain has norm one and constant coordinate ϕ_0. Let $s_* = \min_{x,x'} \phi(x)^\top \phi(x')$ over these vertices. This minimum separates across the public feature blocks. For threshold l, write

$$a_l = \sum_j w_j \{\mathbf{1}(j > l) - \tau_j\}, \qquad v_l = \sum_j w_j \{\mathbf{1}(j > l) - \tau_j\} \psi_8(\tau_j).$$

The covariance radius in Appendix A.5 becomes

$$R^2(r_I, r_S) = \max_{\substack{0 \le l, l' \le m \\ s \in \{s_*, 1\}}} \left\{ \frac{\phi_0^2 (a_l - a_{l'})^2}{r_I} + \frac{\phi_0^2 \{\|v_l - v_{l'}\|^2 - (a_l - a_{l'})^2\}}{r_S} + (1 - \phi_0^2)(a_l^2 + a_{l'}^2) - 2(s - \phi_0^2) a_l a_{l'} \right\}.$$

The first two terms are the intercept and shape contributions. The input contribution is affine in the feature inner product, so one of its two attainable endpoints maximizes it. Each covariance requires only comparisons between threshold patterns, without enumerating the full product domain.

Selected ratios. Table 5 lists the selected ratios for the twelve feature maps, the resulting split of $\mathrm{tr}(W_z\Omega)$ among the intercept, input, and shape directions, and its value relative to the isotropic covariance $\Omega \propto I$ under the same sensitivity calibration. The shape ratio r_S lies between 0.04 and 0.09, so each of the seven shape coefficients receives far less variance than one input coefficient, while the input block carries 53–84% of the average squared prediction perturbation. The directional covariance reduces this perturbation to 30–57% of its isotropic value, most strongly for the linear maps, whose input blocks are small.

Full. The general joint model is $q_V(x, \tau) = \phi(x)^\top V \psi_8(\tau)$, with the same basis as Joint. Write $V_0 = c^\top$ for the constant-input row and V_- for the remaining rows. The penalty is $\lambda_a \|c\|^2/2 + \lambda \|V_-\|_F^2/2$. Restricting $V_- = \beta_- e^\top$, with $e^\top \psi_8(\tau) = 1$ and $\|e\| = 1$, gives Joint's shared predictor and penalty. Full leaves these rows unrestricted and imposes no ordering constraint during fitting. Each noisy gradient step is followed by dividing the constant-input row by $1 + \eta\lambda_a$ and

Table 5: Selected covariance ratios and the resulting allocation of the average squared prediction perturbation $\mathrm{tr}(W_z\Omega)$ for each public feature map. The last column divides $\mathrm{tr}(W_z\Omega)$ by its value under the isotropic covariance with the same sensitivity calibration.

					Share of $\mathrm{tr}(W_z\Omega)$ (%)			
Dataset	Map	p	r_I	r_S	Intercept	Input	Shape	Ratio to $\Omega \propto I$
California	Linear	9	0.595	0.044	16.7	74.7	8.7	0.36
California	Additive	89	1.000	0.053	11.5	84.3	4.2	0.57
Concrete	Linear	9	0.595	0.044	16.7	74.7	8.7	0.36
Concrete	Additive	89	1.000	0.053	11.5	84.3	4.2	0.57
Energy	Linear	15	0.707	0.048	15.3	77.4	7.3	0.42
Energy	Additive	75	1.000	0.053	11.8	83.8	4.4	0.56
Airfoil	Linear	6	0.595	0.062	22.0	61.8	16.2	0.31
Airfoil	Additive	56	1.000	0.074	16.4	75.1	8.5	0.49
Power Plant	Linear	5	0.648	0.074	25.9	53.3	20.8	0.30
Power Plant	Additive	45	1.000	0.088	18.9	69.4	11.7	0.45
Real Estate	Linear	7	0.545	0.053	18.7	68.7	12.6	0.33
Real Estate	Additive	67	0.917	0.062	13.4	80.2	6.4	0.52

Table 6: CDF validation nIPL $\times 100$, averaged over the six datasets, two maps, and ten noise repetitions. Hyperparameters for location learning are held fixed while f and M vary.

	$\varepsilon = 1$			$\varepsilon = 2$			$\varepsilon = 4$		
f	$M = 16$	$M = 32$	$M = 64$	$M = 16$	$M = 32$	$M = 64$	$M = 16$	$M = 32$	$M = 64$
0.1	5.211	5.208	5.256	4.640	4.587	4.599	4.385	4.296	4.296
0.3	5.231	5.197	5.215	4.682	4.618	4.620	4.412	4.319	4.317
0.5	5.483	5.445	5.455	4.818	4.754	4.755	4.473	4.378	4.375

the other rows by $1 + \eta\lambda$. The initial curve is $2\tau - 1$. Its per-record gradient norm is bounded by $C = [\sum_j w_j \max\{\tau_j, 1 - \tau_j\}^2]^{1/2}$, so the average gradient has replacement sensitivity at most $2C/n$. GDP composition over all updates uses isotropic Gaussian noise with standard deviation $2C\sqrt{T}/(n\mu)$.

B.5 Residual budget and Fourier order

Table 6 reports validation losses for the CDF budget fraction f and Fourier order M, with each task's selected hyperparameters for location learning held fixed. Larger M improves resolution while adding noisy coefficients. Larger f reduces residual noise while leaving less budget for location learning. At $\varepsilon = 2, 4$, orders 32 and 64 give similar average losses, whereas order 16 is less accurate. Across the tested orders and privacy budgets, $f = 0.5$ gives higher average loss than $f = 0.1$ or 0.3.

B.6 Charging hyperparameter selection to the privacy budget

We evaluate private hyperparameter selection within the same total (ε, δ) budgets as the main comparison.

Procedure. Candidates are drawn uniformly from each method's full search grid in Appendix B.2 and fitted privately on the 60% training split. Their nIPL on the disjoint 15% validation split is released with Gaussian noise. The candidate with the lowest noisy score is returned and evaluated on the test split without retraining. Each validation loss lies in $[0, 1]$, so the released mean has replacement sensitivity $1/n_{\mathrm{val}}$. We calibrate its noise to the privacy parameter of the fit. Because a record lies in only one split, the fit and its score together have the same privacy guarantee as the fit. For the four GDP-calibrated methods, K is Poisson with mean 10, each candidate is μ_0-GDP, and

the search guarantee follows by applying the Poisson selection bound to the candidate's Gaussian privacy curve (Papernot & Steinke, 2022, Theorem 6). The largest μ_0 meeting the (ε, δ) of the main comparison satisfies $0.51 \leq \mu_0/\mu \leq 0.60$. Sep-Obj's published guarantee is stated per level in (ε, δ) form, so it uses the hard-stopping search (Liu & Talwar, 2019, Theorem 3.4), which stops with probability 0.1 after each candidate and at the latest after $T = \lceil 10 \log(2/\delta) \rceil$, with $(\varepsilon_1, \delta_1)$-DP candidates satisfying $3\varepsilon_1 + 3\sqrt{2\delta_1} = \varepsilon$ and $\sqrt{2\delta_1}\, T = \delta/2$, so that $\varepsilon_1 \approx \varepsilon/3$. The reference fits the settings selected in the main comparison on the same 60% split at the full budget. The procedures differ in candidate search, validation-score perturbation, and per-fit privacy budget. Both use 100 independent noise repetitions, with candidates redrawn in each repetition of private selection.

Results. Private selection has higher mean nIPL than the reference (Table 7): 30–43% for the GDP-calibrated methods and 53% for Sep-Obj, averaged over datasets, maps, and budgets. CDF and Joint remain the two methods with the lowest mean nIPL at every budget. CDF has lower nIPL than Full and Sep-Out in all 36 dataset–map–budget conditions, and Joint in 34 and 36 of them. Within the private search, the selected candidate's unperturbed validation nIPL exceeds that of the best drawn candidate by 1.9–3.5 points ($\times 100$) across methods on Real Estate, with 62 validation records, averaging over maps and budgets. On California, with 3,096 validation records, this gap is below 0.2.

Table 7: nIPL $\times 100$ with hyperparameter selection charged to the privacy budget (Private selection) and with the settings of the main comparison at the full budget (Reference), both fitted on the 60% training split. Entries average the six datasets and two maps within each repetition and report mean $\pm$ standard deviation over 100 repetitions. Bold marks the lowest mean under private selection.

	$\varepsilon = 1$		$\varepsilon = 2$		$\varepsilon = 4$	
Method	Private selection	Reference	Private selection	Reference	Private selection	Reference
Sep-Out	10.512 ± 1.337	7.194 ± 0.211	8.229 ± 1.097	6.528 ± 0.233	7.005 ± 0.712	6.078 ± 0.094
Sep-Obj	12.227 ± 0.609	8.525 ± 0.359	11.156 ± 0.598	7.121 ± 0.217	9.733 ± 0.536	5.978 ± 0.114
Full	10.758 ± 1.031	6.816 ± 0.150	8.122 ± 0.644	5.956 ± 0.118	6.668 ± 0.314	5.154 ± 0.067
Joint	7.979 ± 0.731	5.363 ± 0.146	6.336 ± 0.391	4.638 ± 0.080	5.251 ± 0.235	4.235 ± 0.039
CDF	$\mathbf{7.382} \pm 0.664$	5.061 ± 0.140	$\mathbf{5.822} \pm 0.444$	4.428 ± 0.075	$\mathbf{4.990} \pm 0.248$	4.141 ± 0.040

B.7 Computational cost

Joint's gradient costs $O\{n(p+m) + mr\}$ per update. Its monotone projection acts on the r offset coefficients. CDF's location gradient costs $O(np)$. Its residual stage computes the Fourier expansion in $O(np + nM + KM)$, followed by isotonic projection and inversion. Both methods reuse the fitted predictor at every requested level.

Table 8 measures a single fit on California with 15,480 training observations at $\varepsilon = 2$. Both methods select $T = 8192$. Measurements use one NVIDIA RTX PRO 6000 Blackwell GPU, double precision, and one CPU thread. We average five executions after one warm-up. Public setup includes construction of the noise geometry. Fit time includes all private stages and, for Joint, monotone projection. Prediction evaluates all 513 levels on 5,160 test inputs, including the common clipping and rearrangement.

C Quantile Prediction Across Datasets

Tables 9 and 10 complement the $\varepsilon = 2$ comparison in Table 1. Together, they report every dataset, input map, and privacy budget for the five methods. Values are nIPL multiplied by 100, reported as mean $\pm$ sample standard deviation over 100 noise repetitions. Bold marks the lowest mean in each row.

Table 8: Single-model computational costs, excluding data loading and hyperparameter selection. Public setup is reusable across fits. Memory is the peak increase over the GPU allocation before fitting.

Map	Learner	Setup (s)	Fit (s)	Predict (ms)	Memory (MiB)
Linear	Joint	0.150	1.586	10.23	30.58
Linear	CDF	0.001	0.374	10.85	23.48
Additive	Joint	0.128	1.664	10.42	30.67
Additive	CDF	0.002	0.614	10.61	23.48

Table 9: nIPL $\times 100$ at $\varepsilon = 1$, for both input maps. Entries are mean $\pm$ standard deviation over 100 noise repetitions.

		Reference methods			DP-ShareQ (ours)	
Dataset	Map	Sep-Out	Sep-Obj	Full	Joint	CDF
California	Linear	11.779 ± 0.097	10.732 ± 0.221	8.074 ± 0.107	7.768 ± 0.078	$\mathbf{7.663} \pm 0.055$
	Additive	9.785 ± 0.112	9.659 ± 0.169	8.664 ± 0.062	8.033 ± 0.045	$\mathbf{8.005} \pm 0.047$
Concrete	Linear	9.137 ± 0.386	9.347 ± 0.516	8.693 ± 0.281	7.497 ± 0.545	$\mathbf{7.181} \pm 0.541$
	Additive	8.257 ± 0.300	9.609 ± 0.653	8.496 ± 0.249	7.240 ± 0.360	$\mathbf{6.847} \pm 0.322$
Energy	Linear	8.998 ± 1.035	11.222 ± 1.377	8.989 ± 0.676	4.883 ± 0.698	$\mathbf{4.128} \pm 0.517$
	Additive	4.389 ± 0.709	8.667 ± 1.132	6.027 ± 0.517	3.568 ± 0.320	$\mathbf{1.993} \pm 0.356$
Airfoil	Linear	5.358 ± 0.209	5.288 ± 0.181	5.066 ± 0.055	4.246 ± 0.373	$\mathbf{3.960} \pm 0.301$
	Additive	4.866 ± 0.191	5.302 ± 0.168	4.856 ± 0.111	4.432 ± 0.188	$\mathbf{4.131} \pm 0.180$
Power Plant	Linear	3.197 ± 0.533	2.816 ± 0.196	2.416 ± 0.038	2.318 ± 0.011	$\mathbf{2.315} \pm 0.007$
	Additive	3.970 ± 0.390	3.554 ± 0.142	2.892 ± 0.056	2.378 ± 0.028	$\mathbf{2.363} \pm 0.022$
Real Estate	Linear	6.741 ± 1.493	9.049 ± 1.657	6.395 ± 0.679	4.572 ± 0.458	$\mathbf{4.325} \pm 0.613$
	Additive	6.781 ± 1.345	10.831 ± 1.695	7.607 ± 0.674	4.787 ± 0.461	$\mathbf{4.512} \pm 0.405$

Table 10: nIPL $\times 100$ at $\varepsilon = 4$, for both input maps. Entries are mean $\pm$ standard deviation over 100 noise repetitions.

		Reference methods			DP-ShareQ (ours)	
Dataset	Map	Sep-Out	Sep-Obj	Full	Joint	CDF
California	Linear	9.311 ± 0.637	8.727 ± 0.073	$\mathbf{7.608} \pm 0.035$	7.731 ± 0.018	7.643 ± 0.019
	Additive	9.471 ± 0.026	8.272 ± 0.064	$\mathbf{7.810} \pm 0.024$	7.866 ± 0.014	7.862 ± 0.016
Concrete	Linear	8.834 ± 0.095	8.195 ± 0.309	7.245 ± 0.240	6.007 ± 0.192	$\mathbf{5.831} \pm 0.178$
	Additive	7.827 ± 0.087	7.737 ± 0.285	6.808 ± 0.164	5.949 ± 0.120	$\mathbf{5.794} \pm 0.116$
Energy	Linear	8.343 ± 0.273	6.115 ± 0.692	4.728 ± 0.287	3.036 ± 0.072	$\mathbf{2.980} \pm 0.087$
	Additive	3.096 ± 0.244	3.541 ± 0.406	2.657 ± 0.199	1.068 ± 0.041	$\mathbf{1.004} \pm 0.039$
Airfoil	Linear	5.075 ± 0.045	4.747 ± 0.267	4.146 ± 0.118	3.541 ± 0.041	$\mathbf{3.515} \pm 0.031$
	Additive	4.546 ± 0.047	4.577 ± 0.123	4.136 ± 0.071	3.639 ± 0.077	$\mathbf{3.496} \pm 0.059$
Power Plant	Linear	2.544 ± 0.063	2.421 ± 0.030	2.315 ± 0.005	$\mathbf{2.309} \pm 0.002$	2.317 ± 0.002
	Additive	2.759 ± 0.059	2.599 ± 0.053	2.391 ± 0.013	2.321 ± 0.008	$\mathbf{2.315} \pm 0.012$
Real Estate	Linear	5.459 ± 0.377	5.281 ± 0.498	4.499 ± 0.224	3.348 ± 0.171	$\mathbf{3.261} \pm 0.246$
	Additive	4.506 ± 0.347	5.869 ± 0.582	4.538 ± 0.203	3.235 ± 0.109	$\mathbf{3.169} \pm 0.168$

D PREDICTION INTERVALS

We compare central 90% prediction intervals. Tables 11–13 report interval scores for all five methods. For each dataset, we average the two input maps within each repetition, then compute means and standard deviations over the 100 noise repetitions. Scores are multiplied by 100, and bold marks the lowest mean.

Table 11: Central 90% interval scores $\times 100$ at $\varepsilon = 1$.

Dataset	Sep-Out	Sep-Obj	Full	Joint	CDF
California	165.841 ± 5.275	164.075 ± 6.110	136.903 ± 4.450	120.231 ± 0.632	**119.317** ± 0.503
Concrete	135.367 ± 11.043	160.176 ± 13.953	136.427 ± 10.316	**100.500** ± 5.224	109.950 ± 17.434
Energy	120.873 ± 23.231	171.147 ± 28.605	130.924 ± 11.865	**63.359** ± 5.820	75.143 ± 20.990
Airfoil	79.305 ± 10.789	90.459 ± 5.940	72.412 ± 3.941	**59.943** ± 2.964	60.707 ± 9.380
Power Plant	58.092 ± 5.894	65.707 ± 7.870	52.535 ± 2.937	**33.340** ± 0.402	33.419 ± 0.297
Real Estate	109.628 ± 18.736	174.052 ± 35.833	122.891 ± 14.118	94.164 ± 12.757	**87.878** ± 26.654

Table 12: Central 90% interval scores $\times 100$ at $\varepsilon = 2$.

Dataset	Sep-Out	Sep-Obj	Full	Joint	CDF
California	159.971 ± 7.735	146.233 ± 3.478	126.445 ± 3.438	119.604 ± 0.343	**118.538** ± 0.346
Concrete	136.248 ± 8.865	144.993 ± 11.473	137.015 ± 9.650	88.066 ± 3.092	**86.645** ± 5.736
Energy	86.718 ± 5.098	122.637 ± 14.142	109.841 ± 8.631	**42.926** ± 2.922	46.817 ± 11.403
Airfoil	73.844 ± 3.828	81.089 ± 4.747	69.600 ± 3.554	**50.860** ± 1.169	52.031 ± 3.428
Power Plant	49.001 ± 4.419	51.027 ± 3.670	41.566 ± 1.585	**32.854** ± 0.172	33.048 ± 0.175
Real Estate	88.951 ± 13.958	130.290 ± 20.959	113.837 ± 18.544	**63.817** ± 6.615	74.715 ± 23.705

Table 13: Central 90% interval scores $\times 100$ at $\varepsilon = 4$.

Dataset	Sep-Out	Sep-Obj	Full	Joint	CDF
California	150.220 ± 8.501	132.841 ± 2.268	119.242 ± 2.358	119.406 ± 0.173	**118.283** ± 0.166
Concrete	123.422 ± 4.351	128.346 ± 8.480	120.266 ± 7.547	81.514 ± 1.432	**81.118** ± 2.136
Energy	82.693 ± 5.837	93.361 ± 9.870	107.699 ± 10.693	**33.093** ± 1.659	33.642 ± 2.936
Airfoil	68.895 ± 2.767	75.608 ± 6.554	69.230 ± 3.699	**49.716** ± 0.777	50.530 ± 0.804
Power Plant	46.917 ± 3.221	41.835 ± 1.856	36.000 ± 0.737	**32.741** ± 0.135	32.848 ± 0.203
Real Estate	83.779 ± 6.379	98.334 ± 13.314	97.901 ± 14.289	**49.854** ± 1.536	54.728 ± 12.981

Table 14 separates interval width, miss penalties, and coverage at $\varepsilon = 2$. The average widths are almost equal, 53.910 for Joint and 53.877 for CDF, while the miss penalties are 12.445 and 14.756. CDF's higher 90% interval score comes from the larger penalties for observations outside the interval.

E LOCATION LEARNING AND NOISE GEOMETRY

We give results by dataset for the Joint and CDF ablations in Section 5.2. The final subsections compare residual representations and isolate sharing within output perturbation.

Table 14: Width, total miss penalty, and coverage of central 90% intervals at $\varepsilon = 2$, averaging the two maps. Width and miss penalty sum to the interval score; both are multiplied by 100. Coverage is in percent. Entries are mean $\pm$ standard deviation over 100 noise repetitions.

	Width		Miss penalty		Coverage (%)	
Dataset	Joint	CDF	Joint	CDF	Joint	CDF
California	80.820 ± 0.334	85.902 ± 0.675	38.784 ± 0.410	32.636 ± 0.700	88.849 ± 0.098	89.881 ± 0.168
Concrete	79.361 ± 2.759	73.738 ± 6.800	8.705 ± 2.019	12.907 ± 3.263	94.043 ± 1.074	92.134 ± 1.666
Energy	38.179 ± 3.076	36.533 ± 12.592	4.747 ± 3.016	10.284 ± 7.752	95.583 ± 1.502	91.003 ± 2.585
Airfoil	41.465 ± 1.200	41.207 ± 4.089	9.396 ± 1.158	10.824 ± 2.093	90.762 ± 0.800	90.644 ± 1.571
Power Plant	23.751 ± 0.092	23.842 ± 0.183	9.103 ± 0.167	9.206 ± 0.235	89.953 ± 0.215	89.985 ± 0.257
Real Estate	59.884 ± 7.096	62.038 ± 26.152	3.933 ± 1.770	12.677 ± 5.594	96.514 ± 1.404	90.875 ± 3.725
Overall	53.910 ± 1.340	53.877 ± 5.155	12.445 ± 0.638	14.756 ± 1.713	92.617 ± 0.368	90.753 ± 0.852

E.1 Joint: ordering and noise allocation

Table 15 gives results by dataset for the Joint ablations in Section 5.2. Table 5 reports the corresponding covariance ratios and the reduction in $\mathrm{tr}(W_z\Omega)$ relative to isotropic noise.

Number of knots. Table 16 varies the number of offset knots $r \in \{4, 8, 16\}$ with every other setting fixed at the values selected for $r = 8$, recomputing the public geometry for each r. Average nIPL changes by less than 0.6% across r, whereas the 90% interval score rises by 6–9% at $r = 16$ and falls by 2–5% at $r = 4$ for $\varepsilon \le 2$. A larger r adds shape coefficients that receive privacy noise. The resulting $\mathrm{tr}(W_z\Omega)$ is 1.05–1.26 times its $r = 8$ value at $r = 16$ and 0.87–0.97 times at $r = 4$. The 90% interval score is more sensitive to r than average nIPL.

Unordered-score bound. Use the minimum feature inner product s_* and covariance ratios $r_I, r_S > 0$ from Appendix B.4. The variants w/o ordered bound and w/o ordering use the radius bound

$$R^2_{\mathrm{box}} = \max_{s \in \{s_*, 1\}} \left[\max\left\{ \frac{\phi_0^2}{r_I} + \frac{1 + s - 2\phi_0^2}{2}, \frac{\phi_0^2}{r_S} \right\} + \frac{1 - s}{2} \right].$$

To see this, let $a = \sum_j w_j(u_j - \tau_j)$ and $v = \sum_j w_j(u_j - \tau_j)\psi_r(\tau_j)$, and use primes for a second record. Bessel's inequality gives $\|v - v'\|^2 \le 1$, while $|a + a'| \le 1$. At feature inner product s, the squared slope change is $(1 - s)(a + a')^2/2 + (1 + s - 2\phi_0^2)(a - a')^2/2$. The constant-curve component of $v - v'$ is $a - a'$, so the shape component has squared norm at most $1 - (a - a')^2$. Maximizing over $(a - a')^2, (a + a')^2 \le 1$, then over the endpoints $s_*, 1$, gives the displayed radius.

E.2 CDF: location and residual noise

Location learning. Quantile location replaces $H_D(\beta)$ with the pinball loss over multiple levels in Equation (2), adding the ordered offset d to the first-stage fit. Its median prediction $\phi(x)^\top \widehat{\beta} + \widehat{d}(1/2)$ is used to form the residuals. The added offset coefficients receive a quadratic penalty of strength λ. The Gaussian scale for this objective follows Proposition 1, with an isotropic covariance and location budget μ_A.

Residual noise. Coordinate bound replaces U_M by $4\sum_{k=1}^{M} k^{-1}$ in Equation (9), obtained by bounding $|c_k(u) - c_k(u')| \le 2\sqrt{2}/(\pi k)$ separately for each k.

Table 17 reports both variants alongside CDF.

E.3 Residual representations

We compare Fourier, Histogram, and Polynomial CDF using the estimators and selection ranges in Appendix B.2. Two additional references retain CDF's private location. Fourier without CDF noise keeps the selected order, grid, projection, and inversion, but removes coefficient perturbations. The

Table 15: Ordering, score calibration, and noise allocation in Joint. Results average the two input maps. Losses are multiplied by 100; entries are mean ± standard deviation over 100 noise repetitions. Bold marks the lowest displayed mean for each dataset and column.

Method	nIPL			nIS_{90}		
	$\varepsilon = 1$	$\varepsilon = 2$	$\varepsilon = 4$	$\varepsilon = 1$	$\varepsilon = 2$	$\varepsilon = 4$
California						
Joint	**7.901** ± 0.044	**7.819** ± 0.023	**7.798** ± 0.011	**120.231** ± 0.632	119.604 ± 0.343	119.406 ± 0.173
w/o ord. bound	8.025 ± 0.084	7.869 ± 0.041	7.811 ± 0.020	121.747 ± 1.222	120.148 ± 0.637	119.514 ± 0.305
w/o ordering	8.025 ± 0.084	7.869 ± 0.041	7.811 ± 0.020	121.747 ± 1.222	120.148 ± 0.637	119.514 ± 0.305
$\Omega \propto I$	7.906 ± 0.044	7.820 ± 0.022	**7.798** ± 0.011	120.265 ± 0.781	**119.546** ± 0.401	**119.358** ± 0.209
Concrete						
Joint	**7.368** ± 0.345	**6.527** ± 0.225	**5.978** ± 0.118	**100.500** ± 5.224	**88.066** ± 3.092	**81.514** ± 1.432
w/o ord. bound	8.577 ± 0.721	7.435 ± 0.501	6.370 ± 0.255	119.337 ± 10.093	103.315 ± 6.416	88.538 ± 3.160
w/o ordering	8.575 ± 0.721	7.431 ± 0.501	6.369 ± 0.255	119.287 ± 10.427	103.087 ± 6.519	88.423 ± 3.188
$\Omega \propto I$	7.685 ± 0.402	6.739 ± 0.254	6.060 ± 0.126	123.758 ± 12.847	106.527 ± 9.429	90.387 ± 4.058
Energy						
Joint	**4.225** ± 0.367	**2.650** ± 0.171	**2.052** ± 0.042	**63.359** ± 5.820	**42.926** ± 2.922	**33.093** ± 1.659
w/o ord. bound	5.908 ± 0.748	4.019 ± 0.481	2.321 ± 0.108	84.839 ± 11.210	62.267 ± 6.606	37.616 ± 2.571
w/o ordering	5.910 ± 0.750	4.012 ± 0.481	2.320 ± 0.108	86.778 ± 12.302	61.900 ± 6.735	37.578 ± 2.582
$\Omega \propto I$	4.581 ± 0.366	2.963 ± 0.210	2.108 ± 0.055	77.467 ± 11.855	62.990 ± 11.430	38.413 ± 4.918
Airfoil						
Joint	**4.339** ± 0.216	**3.821** ± 0.094	**3.590** ± 0.042	**59.943** ± 2.964	**50.860** ± 1.169	**49.716** ± 0.777
w/o ord. bound	4.947 ± 0.443	4.080 ± 0.219	3.738 ± 0.085	72.539 ± 5.811	56.438 ± 3.092	52.667 ± 1.548
w/o ordering	4.943 ± 0.443	4.076 ± 0.218	3.737 ± 0.085	72.086 ± 6.016	56.127 ± 3.111	52.616 ± 1.552
$\Omega \propto I$	4.502 ± 0.238	3.892 ± 0.110	3.615 ± 0.042	78.020 ± 9.912	58.531 ± 4.274	52.537 ± 1.298
Power Plant						
Joint	**2.348** ± 0.016	**2.321** ± 0.006	**2.315** ± 0.004	**33.340** ± 0.402	**32.854** ± 0.172	**32.741** ± 0.135
w/o ord. bound	2.416 ± 0.034	2.344 ± 0.014	2.327 ± 0.008	34.828 ± 0.801	33.302 ± 0.396	33.011 ± 0.246
w/o ordering	2.416 ± 0.034	2.344 ± 0.014	2.327 ± 0.008	34.806 ± 0.803	33.302 ± 0.396	33.011 ± 0.246
$\Omega \propto I$	2.355 ± 0.016	2.322 ± 0.006	2.316 ± 0.004	34.146 ± 0.451	32.917 ± 0.173	32.778 ± 0.136
Real Estate						
Joint	**4.679** ± 0.319	**3.794** ± 0.215	**3.292** ± 0.101	**94.164** ± 12.757	**63.817** ± 6.615	**49.854** ± 1.536
w/o ord. bound	5.963 ± 0.601	5.012 ± 0.527	3.851 ± 0.245	117.156 ± 18.561	88.874 ± 11.559	66.921 ± 5.157
w/o ordering	5.911 ± 0.604	4.988 ± 0.530	3.831 ± 0.245	112.958 ± 20.727	86.735 ± 12.454	65.122 ± 5.472
$\Omega \propto I$	5.957 ± 0.540	4.529 ± 0.298	3.727 ± 0.179	148.073 ± 22.675	112.126 ± 19.775	87.501 ± 14.245

Table 16: Joint with r offset knots and all other settings fixed at those selected for $r = 8$. nIPL and 90% interval score $\times 100$, averaged over the six datasets and two maps within each repetition; entries are mean ± standard deviation over 100 noise repetitions. The $r = 8$ row reproduces Table 2.

	nIPL			nIS_{90}		
r	$\varepsilon = 1$	$\varepsilon = 2$	$\varepsilon = 4$	$\varepsilon = 1$	$\varepsilon = 2$	$\varepsilon = 4$
4	5.139 ± 0.108	4.463 ± 0.060	4.177 ± 0.029	74.45 ± 2.07	64.82 ± 0.92	61.82 ± 0.40
8	5.144 ± 0.104	4.489 ± 0.059	4.171 ± 0.028	78.59 ± 2.41	66.35 ± 1.27	61.05 ± 0.49
16	5.162 ± 0.102	4.486 ± 0.061	4.182 ± 0.025	85.33 ± 3.81	72.27 ± 2.57	64.75 ± 1.15

empirical residual CDF also removes the finite Fourier approximation. Both serve as nonprivate references for residual estimation.

Table 17: CDF ablations by dataset, averaging the two input maps. Quantile location replaces the location learner. Coordinate bound calibrates residual noise by bounding each coefficient change separately. Losses are multiplied by 100; entries are mean ± standard deviation over 100 noise repetitions. Bold marks the lowest displayed mean for each dataset and column.

Method	nIPL			nIS_{90}		
	$\varepsilon = 1$	$\varepsilon = 2$	$\varepsilon = 4$	$\varepsilon = 1$	$\varepsilon = 2$	$\varepsilon = 4$
California						
CDF	**7.834** ± 0.037	**7.773** ± 0.021	**7.753** ± 0.012	**119.317** ± 0.503	118.538 ± 0.346	118.283 ± 0.166
Quantile loc.	7.907 ± 0.046	7.820 ± 0.023	7.798 ± 0.014	119.329 ± 0.632	**118.389** ± 0.342	**118.103** ± 0.189
Coord. bound	**7.834** ± 0.037	**7.773** ± 0.021	**7.753** ± 0.012	119.381 ± 0.522	118.557 ± 0.351	118.285 ± 0.167
Concrete						
CDF	**7.014** ± 0.312	**6.265** ± 0.183	**5.813** ± 0.103	**109.950** ± 17.434	**86.645** ± 5.736	**81.118** ± 2.136
Quantile loc.	7.415 ± 0.408	6.532 ± 0.239	5.949 ± 0.127	112.234 ± 14.560	90.022 ± 7.217	81.477 ± 1.917
Coord. bound	7.089 ± 0.345	6.284 ± 0.186	5.819 ± 0.104	117.197 ± 18.830	89.373 ± 7.954	81.897 ± 2.828
Energy						
CDF	**3.061** ± 0.310	**2.298** ± 0.096	**1.992** ± 0.046	**75.143** ± 20.990	**46.817** ± 11.403	**33.642** ± 2.936
Quantile loc.	3.858 ± 0.421	2.660 ± 0.190	2.062 ± 0.054	82.383 ± 20.404	50.905 ± 11.902	33.803 ± 2.683
Coord. bound	3.168 ± 0.351	2.324 ± 0.105	1.996 ± 0.048	86.626 ± 24.057	55.589 ± 18.478	35.767 ± 5.782
Airfoil						
CDF	**4.045** ± 0.185	**3.658** ± 0.082	**3.506** ± 0.032	**60.707** ± 9.380	**52.031** ± 3.428	50.530 ± 0.804
Quantile loc.	4.257 ± 0.215	3.798 ± 0.114	3.589 ± 0.057	62.750 ± 9.758	52.270 ± 3.141	**49.290** ± 0.946
Coord. bound	4.072 ± 0.191	3.664 ± 0.082	3.508 ± 0.032	67.506 ± 13.699	53.662 ± 5.175	50.922 ± 1.085
Power Plant						
CDF	**2.339** ± 0.011	**2.323** ± 0.007	2.316 ± 0.006	33.419 ± 0.297	33.048 ± 0.175	32.848 ± 0.203
Quantile loc.	2.346 ± 0.015	**2.323** ± 0.007	**2.315** ± 0.007	**33.296** ± 0.408	**32.884** ± 0.186	**32.692** ± 0.220
Coord. bound	2.340 ± 0.011	2.324 ± 0.007	2.316 ± 0.006	33.463 ± 0.308	33.059 ± 0.174	32.852 ± 0.203
Real Estate						
CDF	**4.419** ± 0.353	**3.645** ± 0.298	**3.215** ± 0.154	**87.878** ± 26.654	**74.715** ± 23.705	**54.728** ± 12.981
Quantile loc.	4.760 ± 0.463	3.933 ± 0.332	3.386 ± 0.193	92.177 ± 26.476	78.652 ± 24.240	56.317 ± 11.951
Coord. bound	4.523 ± 0.386	3.732 ± 0.336	3.237 ± 0.162	96.751 ± 29.243	86.531 ± 30.989	61.918 ± 18.081

Among the three private residual estimators, Fourier has the lowest mean nIPL and 90% interval score at all three budgets (Table 18). Its nIPL reductions over Histogram are 2.49%, 1.02%, and 0.74%. The Fourier coefficient change decays as $1/k$, so the resulting allocation uses less variance at higher frequencies.

Removing only the Fourier coefficient noise reduces nIPL by 1.37%, 0.64%, and 0.15%, but the 90% interval score by 16.5%, 9.4%, and 2.9%. The noise-free Fourier and empirical CDF scores are close, indicating that residual noise accounts for most of their gap from the private Fourier estimator.

E.4 Sharing within output perturbation

Table 19 compares independent and shared input coefficients under Gaussian output perturbation. Each model selects settings from the same tuning grid. Sharing reduces mean nIPL by 4.6%, 1.9%, and 7.1% at $\varepsilon = 1, 2, 4$.

Let $\phi(x) = (\phi_0, \phi_-(x))$, where ϕ_0 is constant. Sep-Out minimizes the average of its m regularized quantile losses, with an independent coefficient vector at each fitted level. Tying these vectors to $\beta_j = (a_j, \beta_-)$ gives

$$q_j(x) = \phi_0 a_j + \phi_-(x)^\top \beta_-, \qquad R(a, \beta_-) = \frac{\lambda}{2m} \sum_j a_j^2 + \frac{\lambda}{2} \|\beta_-\|^2.$$

Table 18: Residual representations with the same private location and CDF budget. nIPL and 90% interval score are multiplied by 100. Entries are mean ± standard deviation over 100 repetitions. The two reference columns combine the private location with nonprivate residual estimation. Bold compares the private estimators.

		Private residual CDFs			Residual references	
ε	Metric	Polynomial	Histogram	Fourier	Fourier no noise	Empirical
1	nIPL	8.428 ± 0.607	4.907 ± 0.103	**4.785** ± 0.091	4.720 ± 0.082	4.713 ± 0.084
	nIS_{90}	133.801 ± 7.597	84.610 ± 7.020	**81.069** ± 6.922	67.666 ± 1.111	66.967 ± 1.145
2	nIPL	8.026 ± 0.607	4.371 ± 0.066	**4.327** ± 0.061	4.299 ± 0.056	4.293 ± 0.057
	nIS_{90}	128.986 ± 7.606	74.115 ± 6.492	**68.632** ± 4.639	62.210 ± 0.870	62.216 ± 0.826
4	nIPL	7.180 ± 0.389	4.130 ± 0.037	**4.099** ± 0.033	4.093 ± 0.031	4.091 ± 0.031
	nIS_{90}	120.253 ± 6.344	66.038 ± 4.380	**61.858** ± 2.164	60.089 ± 0.480	60.061 ± 0.460

Table 19: Sharing input slopes within output perturbation. Both models select settings from the same grid. nIPL is multiplied by 100, averaging the two input maps; entries are mean ± standard deviation over 100 noise repetitions.

	$\varepsilon = 1$		$\varepsilon = 2$		$\varepsilon = 4$	
Dataset	Sep-Out	Shared	Sep-Out	Shared	Sep-Out	Shared
California	**10.782** ± 0.080	10.828 ± 0.055	10.178 ± 0.613	**9.803** ± 0.586	9.391 ± 0.320	**8.936** ± 0.295
Concrete	8.697 ± 0.255	**8.583** ± 0.194	**8.494** ± 0.147	8.552 ± 0.121	8.330 ± 0.070	**7.911** ± 0.663
Energy	6.694 ± 0.642	**6.192** ± 0.552	**5.892** ± 0.283	5.908 ± 0.244	5.719 ± 0.184	**4.229** ± 0.454
Airfoil	5.112 ± 0.139	**4.900** ± 0.098	4.866 ± 0.049	**4.855** ± 0.059	**4.811** ± 0.031	4.839 ± 0.029
Power Plant	3.584 ± 0.315	**3.256** ± 0.277	2.914 ± 0.118	**2.790** ± 0.119	2.652 ± 0.044	**2.587** ± 0.034
Real Estate	6.761 ± 0.997	**5.964** ± 0.916	5.427 ± 0.513	**5.128** ± 0.489	4.983 ± 0.249	**4.847** ± 0.227

In coordinates $\theta = (a/\sqrt{m}, \beta_-)$, this penalty is $\lambda\|\theta\|^2/2$. The smoothing, candidate levels, and selection rule are those of Sep-Out in Appendix B.2.

At symmetric fitted levels, scores $s_j \in [-\tau_j, 1-\tau_j]$ satisfy $|\bar{s}| \le 1/2$. The gradient is $(\phi_0 s/\sqrt{m}, \phi_- \bar{s})$. Between two records, its offset block changes by at most $|\phi_0|$, and its slope block by at most $\sqrt{1-\phi_0^2}$, so its diameter is at most one. Strong convexity gives coefficient sensitivity $(1/n + 2\zeta)/\lambda$, with the same optimization tolerance as Sep-Out. Isotropic Gaussian output perturbation uses this sensitivity and the total GDP budget μ.

F Residual Variation and Approximation by Shared Models

We vary residual scale and shape to examine approximation by shared models and private learning. Noise-free fits and population references separate these effects from privacy noise.

F.1 Sharing under input-dependent scale

For the scale model in Section 5.3, with $\gamma \in \{0, 0.25, 0.5, 0.75\}$, the conditional quantile is $q^*(x,\tau) = 0.25x_1 + 0.2(1+\gamma x_2)(2\tau - 1)$. Full's functional class contains this curve. The shared population curve uses the true location and pooled residual quantiles, $q_{\text{pool}}(x,\tau) = 0.25x_1 + d_*(\tau)$. Conditional symmetry makes it optimal among curves of the form (1) on a symmetric band (Proposition 3).

For $\gamma > 0$, the marginal residual density has maximum

$$\|\bar{f}\|_\infty = \frac{\log\{(1+\gamma)/(1-\gamma)\}}{0.8\gamma},$$

with limit 2.5 at zero. It is at most 3.244 for the chosen values of γ, so every candidate $h \le 0.1$ satisfies $\|\bar{f}\|_\infty \le 1/(2h)$. The transfer conditions hold throughout this family, and Proposition 8 gives $\mathcal{A}_B = O(\gamma^2)$. At zero scale variation, the residual is uniform on $[-0.2, 0.2]$.

Private learning. Methods use their settings and selection rules from Appendix B.2, with $\delta = n^{-2}$. Validation averages results over ten independently generated pairs of training and validation datasets, of sizes n and 512, respectively. The final comparison uses 100 repetitions with newly generated training data and noise. Evaluation averages 1024 fresh inputs per repetition and integrates the conditional uniform law at the 513 quantile levels. Methods share the data and evaluation inputs.

Figure 4 gives private nIPL. Table 20 removes Gaussian perturbations at the same selected settings in both stages where applicable. At $n = 512, \varepsilon = 1, \gamma = 0$, Joint has lower noise-free loss than CDF, 3.711 versus 3.806. Their private losses in Figure 4 are higher by 0.612 and 0.293, which reverses the order. At $n = 2048, \varepsilon = 4, \gamma = 0.75$, Full instead has lower loss both with and without noise.

Table 20: Noise-free nIPL at the endpoints of the scale sweep. Gaussian perturbations are removed at the settings selected for each privacy budget. Entries are mean ± standard deviation over 100 independently generated training datasets; losses are multiplied by 100.

n	ε	γ	Full	Joint	CDF
512	1	0	4.609 ± 0.055	**3.711** ± 0.023	3.806 ± 0.041
		0.75	4.747 ± 0.072	4.073 ± 0.054	**3.982** ± 0.058
512	4	0	3.767 ± 0.024	**3.683** ± 0.016	3.703 ± 0.023
		0.75	**3.762** ± 0.058	3.913 ± 0.054	3.917 ± 0.055
2048	1	0	3.735 ± 0.012	3.686 ± 0.011	**3.665** ± 0.008
		0.75	**3.835** ± 0.047	3.909 ± 0.048	3.883 ± 0.048
2048	4	0	3.684 ± 0.010	**3.655** ± 0.004	3.665 ± 0.008
		0.75	**3.658** ± 0.048	3.881 ± 0.048	3.882 ± 0.048

F.2 Residual shape at fixed scale

We keep the input distribution and location from Appendix F.1, but vary residual shape at a fixed conditional mean and variance:

$$Y = 0.25X_1 + \frac{0.2}{\sqrt{3}} \frac{Z - \mathbb{E}[Z \mid X]}{\sqrt{\mathrm{Var}(Z \mid X)}}.$$

The reference uses $Z \sim \mathrm{Unif}[-1, 1]$. The skewed law has conditional density $(1 + 0.9X_2 z)/2$ on $[-1, 1]$. The bimodal law mixes $\mathrm{Unif}[-1, -0.8]$ and $\mathrm{Unif}[0.8, 1]$ with equal probabilities, independently of X. Every residual has conditional mean zero and variance $0.2^2/3$. The skewed and bimodal laws depart from conditional symmetry and unimodality, respectively.

Each learner retains its settings selected on the uniform reference, including Joint's smoothing bandwidth and CDF's Huber threshold. We use 100 repetitions with newly generated training data and noise. Sample sizes, privacy budgets, and the linear map follow Section 5.3; $\delta = n^{-2}$, with the evaluation levels in Appendix F.1. We integrate the pinball loss exactly under each conditional law and average over 1024 fresh inputs per repetition. Noise-free fits remove perturbations at the same settings.

Table 21 shows similar relative performance under uniform and skewed residuals. Under bimodality, Joint has lower loss as the sample size or privacy budget increases, while CDF remains better at $n = 512, \varepsilon = 1$. The gap under bimodal residuals persists even without privacy noise.

Table 21: Residual shape at fixed conditional variance: nIPL $\times 100$, with settings inherited from the uniform reference. Entries are mean $\pm$ standard deviation over 100 independently generated training datasets, with or without privacy noise. Bold marks the lower mean within each private or noise-free pair.

			Private		Noise-free	
n	ε	Residual law	Joint	CDF	Joint	CDF
512	1	Uniform	4.383 ± 0.362	**4.096** ± 0.215	**3.713** ± 0.027	3.811 ± 0.050
		Skewed	4.367 ± 0.362	**4.063** ± 0.212	**3.694** ± 0.026	3.782 ± 0.045
		Bimodal	4.438 ± 0.360	**4.254** ± 0.221	**3.466** ± 0.032	4.049 ± 0.065
512	4	Uniform	**3.727** ± 0.038	3.735 ± 0.044	**3.682** ± 0.017	3.703 ± 0.026
		Skewed	**3.709** ± 0.036	3.713 ± 0.040	**3.664** ± 0.016	3.683 ± 0.023
		Bimodal	**3.489** ± 0.047	3.686 ± 0.098	**3.333** ± 0.014	3.639 ± 0.077
2048	1	Uniform	3.718 ± 0.032	**3.699** ± 0.023	3.688 ± 0.010	**3.665** ± 0.008
		Skewed	3.700 ± 0.031	**3.684** ± 0.021	3.669 ± 0.010	**3.653** ± 0.009
		Bimodal	**3.512** ± 0.044	3.823 ± 0.095	**3.439** ± 0.010	3.828 ± 0.070
2048	4	Uniform	**3.658** ± 0.007	3.666 ± 0.009	**3.654** ± 0.005	3.664 ± 0.009
		Skewed	**3.641** ± 0.007	3.658 ± 0.011	**3.637** ± 0.005	3.656 ± 0.010
		Bimodal	**3.331** ± 0.004	3.855 ± 0.117	**3.320** ± 0.002	3.867 ± 0.132

F.3 Marginal and conditional coverage

For a predictor's central 80% interval $I(x)$, define conditional coverage error as $\mathbb{E}_X |\Pr\{Y \in I(X) \mid X\} - 0.8|$. Computed from the population law, the pooled oracle has exact 80% marginal coverage at every γ but 18.1 percentage points of conditional error at $\gamma = 0.75$. Table 22 compares the fitted intervals at this scale variation. Full has lower conditional error at $n = 2048, \varepsilon = 4$. At $n = 512, \varepsilon = 1$, CDF has the lowest interval score but higher conditional error than Joint.

Table 22: Central 80% intervals at $\gamma = 0.75$. Scores and normalized widths are multiplied by 100; coverage is in percent and conditional error in percentage points. Entries are mean $\pm$ standard deviation over 100 independent repetitions with newly generated training data and noise.

n	ε	Method	nIS_{80}	Width	Coverage	Cond. error
512	1	Full	66.195 ± 4.651	52.964 ± 6.682	83.062 ± 5.065	16.229 ± 1.565
		Joint	59.696 ± 4.887	55.694 ± 5.715	92.837 ± 2.619	14.862 ± 1.249
		CDF	49.883 ± 2.680	35.724 ± 4.464	78.583 ± 4.650	15.894 ± 2.099
512	4	Full	72.963 ± 8.757	70.147 ± 9.015	94.779 ± 2.089	16.654 ± 1.007
		Joint	45.805 ± 0.538	35.297 ± 1.601	81.714 ± 2.077	16.606 ± 0.672
		CDF	45.783 ± 0.802	32.256 ± 1.968	77.974 ± 2.678	18.079 ± 1.252
2048	1	Full	67.446 ± 7.466	63.733 ± 8.369	93.061 ± 3.008	15.689 ± 1.306
		Joint	45.563 ± 0.552	34.209 ± 0.811	80.537 ± 1.169	17.023 ± 0.540
		CDF	45.381 ± 0.545	32.324 ± 1.465	78.377 ± 1.910	18.177 ± 0.919
2048	4	Full	42.542 ± 2.430	39.402 ± 2.663	90.060 ± 2.019	11.932 ± 1.353
		Joint	44.907 ± 0.445	34.042 ± 0.598	80.872 ± 0.916	17.580 ± 0.327
		CDF	44.913 ± 0.467	33.064 ± 0.726	79.655 ± 1.065	18.042 ± 0.433