LEARNING CONDITIONAL DISTRIBUTIONS FROM REPLICATED AGGREGATES

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning an individual response distribution from group totals becomes difficult when each measurement carries unknown offsets and the underlying distribution is heavy-tailed. To tackle this, we introduce Pair-Triple Spectral Recovery (PTSR) where pair differences recover the conditional law up to reflection and triple contrasts use spectral phase to resolve the remaining orientation. PTSR provides moment-free recovery while also eliminating the shared nuisance offsets. We provide finite-sample guarantees for reflection selection and connect spectral signal-to-noise to intrinsic reflection information. Across two heavy-tailed benchmarks, PTSR improves conditional-density accuracy and held-out likelihood over winsorized method-of-moments and aggregate-likelihood estimator baselines.

1 INTRODUCTION

A single measurement can combine many unit-level responses. Household electricity meters record appliance demand (Hart, 1992), transcriptomic assays aggregate molecular counts across cells (Fishman et al., 2026), and portfolio profit and loss sums position-level value changes weighted by holdings (Danielsson, 2011). Recovering individual behaviour from these observations is an inverse problem. With unit covariates, the target is a shared conditional law $x \mapsto Q^\star(\cdot \mid x)$ describing response variability, asymmetry, and tail behaviour.

We study a repeated-measurement version of this problem. Group membership and weights are known, and an instrument reports several totals under the same conditional state. An unknown baseline is shared within each replicate group but can change between groups. A conditional likelihood can connect individual responses to aggregate supervision (Zhang et al., 2020; Law et al., 2018); here, estimating the response law also requires separating its shape from the group baselines. Heavy tails make moment-based approaches delicate because response variance and skewness may be undefined. We target shape and orientation relative to a fixed location convention.

Replication exposes a useful structure where subtracting two totals cancels their common baseline, but the resulting distribution is unchanged when every individual response law is reflected about its reference location. For a family identified up to reflection by pair spectra, this leaves two candidate orientations. A zero-sum contrast of three totals also removes the baseline and retains phase information that can distinguish the candidates (Figure 1). This connects replicate contrasts to the power-spectrum and bispectral invariants used in signal reconstruction (Bendory et al., 2018).

We introduce *Pair–Triple Spectral Recovery* (PTSR), which uses this structure to separate shape estimation from orientation recovery. A shared conditional model fits pair spectra and supplies the two reflected candidates at each covariate location. The model then predicts their opposite triple-phase patterns. PTSR compares these predictions with independent triple observations and pools the signed evidence across nearby covariates. Its empirical quantities are averages of bounded Fourier features (Feuerverger & McDunnough, 1981), so the estimating statistics remain defined even when moments of the underlying responses do not exist.

Contributions. We develop a conditional-distribution estimator that eliminates shared additive offsets and combines pair-based orbit estimation with local triple-based orientation recovery (Section 3). We give a moment-free analysis of local reflection selection, including a finite-sample guarantee and a spectral information interpretation of the implemented score (Section 6). We eval-

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.

uate the complete estimator through matched heavy-tail comparisons and controls that separate the roles of moment availability and spatial coverage (Section 5).

2 LEARNING FROM REPLICATED AGGREGATES

2.1 OBSERVATIONS AND TARGET

We observe repeated totals from groups of units whose covariates are known. The target is one shared conditional rule $x \mapsto Q_x^\star$, where $Q_x^\star = Q^\star(\cdot \mid x)$ is a probability distribution on $\mathbb{R}$ and $x \in \mathcal{X} \subseteq \mathbb{R}^d$. Units with different covariates may therefore have different response distributions. As in aggregate-observation learning (Zhang et al., 2020), the individual responses enter a known observation operator. Let b index a group design $\mathcal{D}_b = \{(x_{bi}, a_{bi})\}_{i=1}^{k_b}$, which is observed. For independent repetition $n \in \{1, \dots, L_b\}$ of that design, we observe a vector Z_{bn} of R_b totals:

$$Z_{bn}^{(r)} = M_{bn} + S_{bn}^{(r)}, \qquad \text{where} \quad S_{bn}^{(r)} = \sum_{i=1}^{k_b} a_{bi} U_{bni}^{(r)} + \varepsilon_{bn}^{(r)}, \quad r = 1, \dots, R_b. \tag{1}$$

Here $U_{bni}^{(r)} \mid x_{bi} \sim Q_{x_{bi}}^\star$ is hidden. The known design fixes membership (b), covariates (x_b), and weights (a_b) across replicates. Conditional on the design, responses are independent across units and replicates, and each replicate uses fresh draws from the same conditional laws. Different repetitions are independent. The offset M_{bn} is shared within the replicate vector and can vary arbitrarily between repetitions. Additional noise $\varepsilon_{bn}^{(r)}$ is independent of the responses and across repetitions and replicates, with the same known or separately calibrated symmetric law within a design. Setting it to zero gives the main recovery setting.

Since in general the mean of $Q_x^\star$ may not exist (e.g. Cauchy distribution), we estimate shape and orientation with respect to a fixed coordinate function $c : \mathcal{X} \to \mathbb{R}$. This fixes a reference location for each $Q_x^\star$ such that taking $U \mapsto U + c(x)$ and $M_{bn} \mapsto M_{bn} - \sum_i a_{bi} c(x_{bi})$ leaves every observed total, Z_{bn}, unchanged.

We write the characteristic functions[1]: $\Phi_x^{\text{unit}}(t) = \mathbb{E}[e^{itU} \mid x]$ for an individual response, $\Phi_b^\varepsilon(t)$ for measurement noise, and $\Phi_b^{\text{agg}}(t)$ for the unshifted aggregate $S_{bn}^{(r)}$. Conditional independence gives

$$\Phi_b^{\text{agg}}(t) = \Phi_b^\varepsilon(t) \prod_{i=1}^{k_b} \Phi_{x_{bi}}^{\text{unit}}(a_{bi} t). \tag{2}$$

In practice, we approximate the empirical characteristic functions, which are simply the average of bounded Fourier features (Feuerverger & McDunnough, 1981). Also, characteristic functions exist without requiring existence of moments of Q_x since $|e^{itU}| = 1$.

2.2 REFLECTION STRUCTURE OF REPLICATE CONTRASTS

For $U \sim Q$, let $\mathcal{R}Q$ be the law of $-U$ in the fixed location coordinate. Choose a reference orientation $Q_{x,+}$ and define

$$Q_{x,-} = \mathcal{R}Q_{x,+}, \qquad [Q_x] = \{Q_{x,+}, Q_{x,-}\}.$$

The set $[Q_x]$ is the *reflection orbit*. The labels $+$ and $-$ distinguish its representatives; the data determine which is correct. Their unit characteristic functions satisfy $\Phi_{x,-}^{\text{unit}}(t) = \overline{\Phi_{x,+}^{\text{unit}}(t)}$. For a symmetric law, the two representatives coincide.

Proposition 1 (Pair invariance and triple separation)**.** *For three independent equal-design replicates, define the normalized contrasts*

$$D = \frac{Z^{(1)} - Z^{(2)}}{\sqrt{2}}, \qquad C = \frac{Z^{(1)} + Z^{(2)} - 2Z^{(3)}}{\sqrt{6}}.$$

[1]We use here and across the paper these superscripts to label the transformed variable so it is easy to follow; these are not to be confused with powers, it should be quite clear to differentiate these

Both eliminate the shared offset. The factors $\sqrt{2}$ and $\sqrt{6}$ make the coefficient vectors unit length. If an unshifted replicate has finite variance V_b, each normalized contrast also has variance V_b. The normalization itself requires no finite moments.

Their characteristic functions are

$$\Phi_b^{\mathrm{pair}}(t) = \left|\Phi_b^{\mathrm{agg}}\left(\frac{t}{\sqrt{2}}\right)\right|^2, \qquad \Phi_b^{\mathrm{triple}}(t) = \left[\Phi_b^{\mathrm{agg}}\left(\frac{t}{\sqrt{6}}\right)\right]^2 \Phi_b^{\mathrm{agg}}\left(\frac{-2t}{\sqrt{6}}\right). \tag{3}$$

Reflecting all unit laws preserves Φ_b^{pair} and conjugates Φ_b^{triple}. A frequency with $\operatorname{Im}\Phi_b^{\mathrm{triple}}(t) \neq 0$ therefore separates the two reflected triple-contrast laws.

Proof. The contrast coefficients sum to zero, cancelling the offset. Independence gives the products in Equation (3). Symmetry makes Φ_b^{ε} real, so reflecting the unit laws conjugates Φ_b^{agg}. Its squared magnitude is unchanged, while its triple product is conjugated. A nonzero imaginary component distinguishes the characteristic functions and hence the two distributions. □

Figure 1 follows the same two reflection hypotheses through the individual, pair, and triple distributions.

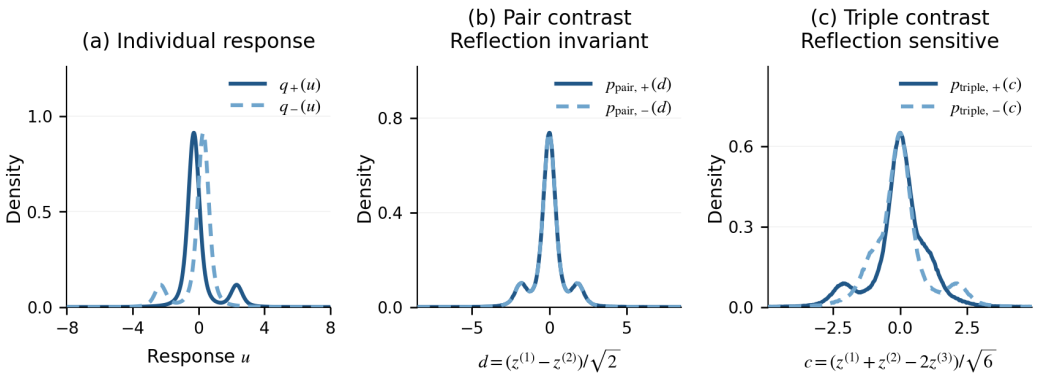

Figure 1: **Pair and triple contrasts under reflected response laws.** Dark solid curves correspond to $Q_{x,+}$; lighter dashed curves correspond to $Q_{x,-} = \mathcal{R}Q_{x,+}$. Panel (a) shows the individual response densities. Panels (b) and (c) show the induced density estimates of $D = (Z^{(1)} - Z^{(2)})/\sqrt{2}$ and $C = (Z^{(1)} + Z^{(2)} - 2Z^{(3)})/\sqrt{6}$ respctively. Their axes and density heights include this change of units. Pair densities coincide under reflection; triple densities retain orientation information. The shared group offset cancels in both contrasts.

Recovering a conditional orbit additionally requires an identifiable response family and sufficient pair designs. We assume that equality of the population pair spectra over the available designs determines $[Q_x]$ on the covariate domain of interest. Pair fitting then estimates a reference representative $Q_{x,+}$ and its reflection $Q_{x,-} = \mathcal{R}Q_{x,+}$; local triple evidence addresses the remaining choice. The triple product in Equation (3) is a diagonal bispectral quantity, connecting the construction to phase recovery from higher-order invariants (Bendory et al., 2018).

3 Pair-Triple Spectral Recovery

PTSR estimates a conditional reflection orbit from pairs, constructs local orientation evidence from triples, and pools that evidence across covariates. We describe these operations for a conditional model with an evaluable characteristic function. Pair fitting uses the weighted aggregate map in

Equation (2). For local orientation recovery, we use groups whose units share an anchor context x_j, as in the recovery benchmark. Each anchor has a known group design, so its two candidate aggregate laws are reflections of one another.

3.1 Estimating shape from pair spectra

Let $Q_{\theta,x,+}$ be a canonical representative of the model's reflection orbit and let $Q_{\theta,x,-} = \mathcal{R}Q_{\theta,x,+}$. The parameter vector θ is shared across covariates. For pair-training design b, average the cosine features of the normalized differences $D_{bn}^{rs} = (Z_{bn}^{(r)} - Z_{bn}^{(s)})/\sqrt{2}$:

$$\widehat{P}_b(t) = \frac{1}{L_b\binom{R_b}{2}} \sum_{n=1}^{L_b} \sum_{1\le r<s\le R_b} \cos(tD_{bn}^{rs}). \tag{4}$$

Here L_b is the number of independent replicate vectors for that design, and the target has expectation $\Phi_b^{\mathrm{pair}}(t)$. Differences from the same vector are dependent; averaging them forms one pair-spectrum target.

With fixed frequency sets $\mathcal{T}_b^{pair} = \{t_{b1}, \dots, t_{bm}\}$, fit the model by

$$\widehat{\theta} \in \arg\min_{\theta} \frac{1}{B_{\mathrm{p}}} \sum_{b=1}^{B_{\mathrm{p}}} \frac{1}{|\mathcal{T}_b^{\mathrm{p}}|} \sum_{t\in\mathcal{T}_b^{\mathrm{p}}} \left(\Phi_{\theta,b}^{\mathrm{pair}}(t) - \widehat{P}_b(t)\right)^2. \tag{5}$$

The model prediction is real because it is a squared modulus of the aggregate characteristic function. For an equal-weight local group without additional noise, it reduces to $|\Phi_{\theta,x_b}^{\mathrm{unit}}(t/\sqrt{2})|^{2k_b}$. A separate pair-validation set selects the checkpoint that has lowest validation loss under the same fixed-frequency objective. We select ('freeze') the resulting model as our estimator and write its reflected candidates as $\widehat{Q}_{x,+}$ and $\widehat{Q}_{x,-} = \mathcal{R}\widehat{Q}_{x,+}$.

3.2 Constructing bounded triple evidence

At anchor x_j, collect N_j fresh independent triple groups and form their normalized contrasts C_{jn}. For fixed frequencies $\mathcal{T}_j^{triple} = \{t_{j1}, \dots, t_{jm}\}$, define

$$h_j(c) = \big(\sin(t_{j1}c), \dots, \sin(t_{jm}c)\big)^\top, \qquad \overline{h}_j = \frac{1}{N_j}\sum_{n=1}^{N_j} h_j(C_{jn}). \tag{6}$$

This empirical mean is the average of bounded features and remains well defined for heavy-tailed contrasts. Using the aggregate design at anchor j, the frozen pair model predicts

$$\begin{aligned} \widehat{\Phi}_{j,+}^{\mathrm{triple}}(t) &= \left[\widehat{\Phi}_{j,+}^{\mathrm{agg}}\left(\frac{t}{\sqrt{6}}\right)\right]^2 \widehat{\Phi}_{j,+}^{\mathrm{agg}}\left(\frac{-2t}{\sqrt{6}}\right), \\ \widehat{\mu}_j &= \big(\operatorname{Im}\widehat{\Phi}_{j,+}^{\mathrm{triple}}(t_{jq})\big)_{q=1}^{m}. \end{aligned} \tag{7}$$

Its reflected candidate predicts $-\widehat{\mu}_j$. PTSR measures how strongly the observed phase aligns with this predicted direction:

$$\widehat{v}_j = \frac{\widehat{\mu}_j}{\|\widehat{\mu}_j\|_2}, \qquad z_j = \sqrt{N_j}\,\widehat{v}_j^\top \overline{h}_j, \tag{8}$$

where $\widehat{v}_j = 0$ if the predicted phase vanishes. Positive evidence favours $\widehat{Q}_{x_j,+}$ and negative evidence favours its reflection. The factor $\sqrt{N_j}$ accounts for the number of independent triple observations. This is a Euclidean normalization of the phase direction. The implementation regularizes it near vanishing phase; Appendix A gives the exact expression and the frequency conversion used by our code implementation.

3.3 POOLING EVIDENCE AND RECONSTRUCTING THE LAW

Nearby covariate locations can share orientation information. PTSR pools the local scores using nonnegative weights that sum to one:

$$\widehat{f}_\ell(x) = \sum_j \omega_{\ell,j}(x) z_j, \qquad \omega_{\ell,j}(x) \geq 0, \qquad \sum_j \omega_{\ell,j}(x) = 1. \tag{9}$$

We use normalized Gaussian-kernel weights, so the bandwidth ℓ determines how far evidence is shared across the covariate domain. We select ℓ by agreement between the pooled training scores and independent triple evidence at validation anchors. Both pair-checkpoint selection and bandwidth selection use aggregate observations only. The exact kernel, numerical safeguards, and validation criterion are collected in Appendix A.

The pooled sign selects the conditional law:

$$\widehat{s}(x) = \operatorname{sign}\big(\widehat{f}_{\widehat{\ell}}(x)\big), \qquad \widehat{Q}_x = \widehat{Q}_{x,\widehat{s}(x)}. \tag{10}$$

A zero score selects the positive candidate. For a symmetric candidate, either sign gives the same distribution. The pooling stage selects orientation while retaining the pair-estimated shape.

Algorithm 1 Pair–Triple Spectral Recovery

Input: pair-training and pair-validation observations; separate local triple-training and triple-validation observations; a conditional reflection family; fixed frequencies and candidate bandwidths.

1. Form pair-spectrum targets and fit the shared orbit model by Equation (5).
2. Select the checkpoint on pair validation and freeze its two reflected candidates.
3. Compute local triple-phase evidence at training and validation anchors using Equation (18) in Appendix A.
4. Pool the evidence and select the bandwidth using Equations (19) and (20) in Appendix A.
5. Return the conditional law selected by the pooled sign in Equation (10).

4 EXPERIMENTAL PROTOCOL

We evaluate conditional-density recovery from shifted totals using a heavy-tailed location mixture with a closed-form characteristic function. This instantiates the general construction in Section 3. Aggregate likelihood and spectral estimators use the same α-stable response family.

4.1 RESPONSE FAMILY AND OBSERVATION DESIGN

For a covariate $x = (x_1, x_2) \in [-1, 1]^2$, the unshifted individual response is

$$U \mid x = A_{\eta(x),d(x)} + \sigma(x) S_\alpha, \qquad \mathbb{E}[e^{itS_\alpha}] = \exp(-|t|^\alpha), \qquad \alpha = 1.5. \tag{11}$$

The bounded two-point component $A_{\eta,d}$ has zero mean, separation d, and orientation parameter η; S_α is independent symmetric stable noise. Reflection about the fixed location changes η to $-\eta$ while preserving σ and d. The response has infinite variance and no finite third moment (Samoradnitsky & Taqqu, 2017), providing a setting in which to evaluate bounded spectral recovery.

The orientations have the form $\eta(x) = m(x) \tanh(g(x)/0.12)$, with two distinct nonlinear choices of g defining Recovery Tasks A and B, specified in Appendix B.1. Tasks A and B share smooth scale, separation, and magnitude functions. Thus the response law varies continuously across each orientation boundary. A local group contains k independent units at the same x, with $a_{bi} = 1$. Every replicate vector receives a fresh contaminated-normal offset shared by all its totals; the target is the unshifted unit law.

Pair data use $k \in \{2, 4\}$ and eight repeated totals per context. The triple stage uses $k = 4$, 72 training anchors, and 36 validation anchors. We vary $N \in \{512, 2048, 8192\}$ independent triple groups per training anchor while keeping the pair data and anchor locations fixed. Smaller budgets use prefixes of the largest stream. Covariates are scrambled Sobol points generated independently of the true orientation boundary. Pair training, pair validation, triple training, triple validation, and individual evaluation use separate random streams. We conduct twelve independent runs per task; the exact sample counts and training settings are in Appendix B.2.

4.2 COMPARATORS AND EVALUATION

Winsorized GMM shares PTSR's frozen pair estimate, observed triples, validation anchors, and Gaussian-kernel smoother. It replaces sine-phase evidence with a clipped cubic statistic; the clipping threshold and bandwidth are selected using aggregate validation observations. This matched comparison isolates the choice of triple statistic.

LIO/LAO MLE instantiates the aggregate-likelihood framework of Zhang et al. (2020) using a stable-mixture likelihood adapter. **Nuisance-profiled MLE** additionally optimizes one offset for each replicate vector. **Raw ECF** fits the signed model to empirical characteristic functions of the unadjusted totals.[2] As information-rich references, **Shift-known MLE** receives the simulated offsets, **individual supervision** receives unit-level outcomes, and the **true law** gives the generating distribution. Appendix B.3 defines the estimators and profiling audit.

All methods are evaluated on a common held-out panel of 12,000 covariates and unit responses per run. The primary metric is mean conditional total variation,

$$\widehat{R}_{\mathrm{TV}} = \frac{1}{n_{\mathrm{test}}} \sum_{i=1}^{n_{\mathrm{test}}} \frac{1}{2} \int_{\mathbb{R}} |\widehat{q}(u \mid x_i) - q^{\star}(u \mid x_i)| \, du, \tag{12}$$

computed by common-grid numerical quadrature. Individual negative log likelihood evaluates predictive density at the held-out response. Orientation-region F1 measures positive-versus-negative orientation on points with $|\eta^{\star}(x)| > 0.1$, where the reflected candidates are appreciably different. These metrics distinguish density accuracy from orientation recovery. Aggregate learners use aggregate validation; the individual test panel scores the frozen models. Uncertainty and paired comparisons are calculated across independent runs (Appendix B.5).

5 RESULTS

5.1 CONDITIONAL-DENSITY RECOVERY

Table 1 reports recovery at the largest triple budget. PTSR attains conditional TV of 0.0674 on Task A and 0.0694 on Task B, reducing Winsorized GMM's errors by 45.0% and 37.6% respectively. The paired intervals favour PTSR on both tasks (Appendix C.1). Since the estimators share their pair model, observations, and pooling family, this comparison isolates triple-stage estimation, including aggregate-selected smoothing.

Table 1: **Recovery at $N = 8192$ triple groups per anchor.** Means over twelve independent runs; uncertainty is reported in Appendix C.1. The top block contains generating or privileged references. Daggered likelihood rows use the reduced pair pool in Appendix B.4. PTSR and GMM share both pair sizes and the same triples.

	Recovery Task A			Recovery Task B		
Method	TV ↓	NLL ↓	F1 ↑	TV ↓	NLL ↓	F1 ↑
True law	0.0000	1.5525	1.0000	0.0000	1.5509	1.0000
Individual supervision	0.0235	1.5543	0.9951	0.0263	1.5532	0.9981
Shift-known MLE†	0.0634	1.5657	0.9699	0.0638	1.5649	0.9644
Nuisance-profiled MLE†	0.2679	1.8257	0.8820	0.3028	1.8942	0.7820
LIO/LAO MLE†	0.3556	1.9019	0.5508	0.3546	1.9012	0.4024
Winsorized GMM	0.1225	1.6004	0.4325	0.1112	1.5940	0.5662
PTSR (ours)	**0.0674**	**1.5679**	**0.9601**	**0.0694**	**1.5676**	**0.9694**

The improvement also appears in predictive likelihood. PTSR attains individual NLL of 1.5679 and 1.5676, whereas Winsorized GMM obtains 1.6004 and 1.5940. The corresponding individual-supervision values are 1.5543 and 1.5532. Thus the spectral estimator approaches the predictive

[2]The frozen likelihood and raw-ECF pools retain the pair-$k = 2$ and triple-$k = 4$ blocks; PTSR and Winsorized GMM also use the pair-$k = 4$ block. We report this coverage difference in Appendix B.4.

accuracy of a model trained with unit-level outcomes while recovering the conditional density from repeated totals.

Figure 2 connects these scores to the reconstructed fields. Its upper row shows orientation, while its lower row evaluates the complete density.

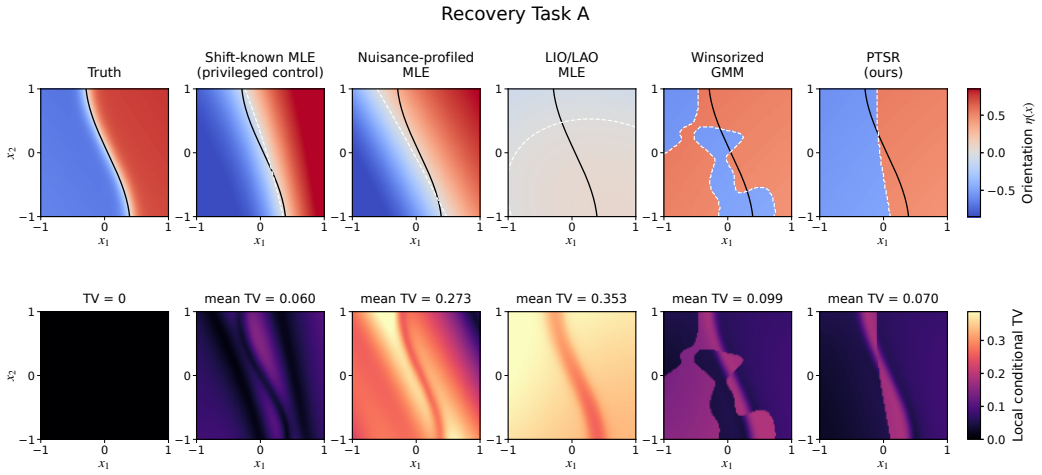

Figure 2: **Recovery Task A.** Upper panels show orientation; black and white dashed contours mark true and fitted zero sets. Lower panels show pointwise TV. Coordinates (x_1, x_2) and ticks are shared, with x_2 labelled only in the Truth column. Run 903 is shown; information access and visualization details appear in Table 1 and Appendix C.2.

PTSR recovers the large orientation regions in both tasks, whereas the clipped-moment fields introduce additional sign changes (Figures 2–3). Their shared shape model makes this difference particularly informative. Across all runs, PTSR attains orientation-region F1 of 0.9601 and 0.9694. The likelihood panels also show why a plausible contour is insufficient: errors in scale, separation, and mixture weights can leave a much larger density error.

PTSR's remaining error includes the magnitude inherited from pair fitting. Its orbit TV is 0.0640 and 0.0668, accounting for most of the full-TV values in Table 1. Thus a paler but correctly signed region can still have shape error. Appendix E examines the corresponding response densities at fixed covariates.

5.2 Evidence budget and supporting controls

Appendix C.3 examines the triple-stage budget with the pair data and anchor locations fixed. At $N = 512$, PTSR attains mean TV of 0.0842 and 0.0871, below Winsorized GMM at $N = 8192$ in both tasks. Figure 5 reports all three measured budgets. This sixteen-fold comparison concerns triple groups per anchor, rather than the complete training dataset.

The Gaussian control and nuisance audit in Appendices C.5 and C.4 examine the roles of moment evidence and shared offsets. With Gaussian noise and supplied quotient parameters, third-order GMM and PTSR achieve nearly equal orientation F1 (Figure 7). The accompanying fields show both estimators recovering the same boundary (Figure 8). Refining the nuisance-profile search does not improve the recorded density recovery, while shift-known MLE remains close to PTSR (Figure 6). The likelihood comparisons retain the data-coverage qualification in Section 4.2.

Appendix D tests spatial pooling on three disconnected positive regions. PTSR recovers every broad component across the eight runs, while smaller components are sensitive to anchor coverage and bandwidth selection. The appendix gives the generating field, reports all component scales, and displays the all-negative realization alongside successful recovery. These studies distinguish the roles of pair estimation, local triple evidence, and spatial coverage.

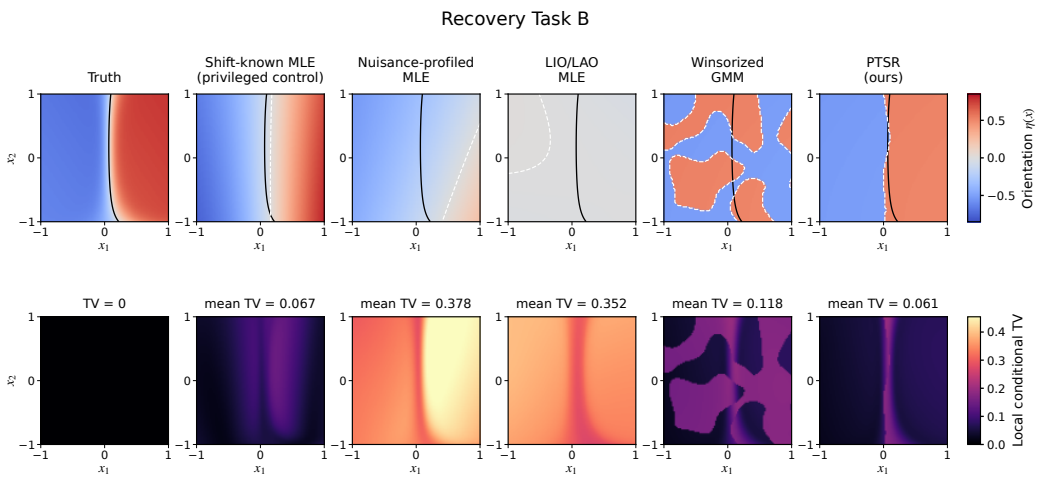

Figure 3: **Recovery Task B.** Layout and information access match Figure 2; the displayed run is 907. Colour saturation measures orientation magnitude, not confidence. Each task uses a common TV scale across its methods. Table 1 reports the twelve-run test-panel means rather than the local grid means shown here.

6 Reflection information and local guarantees

PTSR combines bounded triple features into local orientation evidence. We now relate the reliability of this evidence to its signal and variability. Fix an anchor and condition on the independently fitted pair model and frequency set. Let P_+ and $P_- = \mathcal{R}P_+$ be the two true contrast laws. For the feature vector h in Equation (6), define

$$\mu = \mathbb{E}_+[h(C)], \qquad \Sigma = \mathrm{Cov}_+[h(C)], \qquad j(w) = \frac{(w^\top \mu)^2}{w^\top \Sigma w}. \tag{13}$$

These are moments of bounded sine features, so they exist even when the response has no finite mean or variance. PTSR uses the predicted phase direction $w = \widehat{v}_j$ from Equation (8); positive rescaling of w preserves its local decision and $j(w)$.

Proposition 2 (Local reflection selection)**.** *Let $C_1, \ldots, C_N$ be independent contrasts from one candidate law, and fix w before observing them. If $a = w^\top \mu > 0$, selecting the sign of $\sum_n w^\top h(C_n)$ gives*

$$\max_{s \in \{+,-\}} \Pr_s(\widehat{s} \neq s) \leq \exp\left(-\frac{Na^2}{2\|w\|_1^2}\right). \tag{14}$$

The bound follows from Hoeffding's inequality because $|w^\top h(C)| \leq \|w\|_1$ (Hoeffding, 1963). Appendix F.1 gives the proof and a variance-sensitive bound. For positive score variance, a normal approximation further yields

$$\Pr_s(\widehat{s} = s) \approx \Phi_{\mathrm{N}}\left(\sqrt{Nj(w)}\right), \tag{15}$$

where Φ_{N} is the standard normal CDF. The approximation error is controlled by bounded-score moments in Appendix F.2. The positive-mean condition expresses alignment between the learned phase direction and the true candidate.

Figure 4 compares predicted and observed local reflection accuracy using supplied candidate laws for Tasks A and B. For each horizontal coordinate q, the grey dashed curve is $\Phi_{\mathrm{N}}(c\sqrt{q})$, with one coefficient c fitted on separate calibration scenes and shared across tasks. In panel (c), the green solid curve is the normal-approximation prediction in Equation (15), which fixes $c = 1$. The curves nearly coincide: fitting gives $c = 1.0025$, and both yield MSEs of approximately 0.00112 and 0.00101. Calibrated squared phase has almost identical error (Appendix G.3).

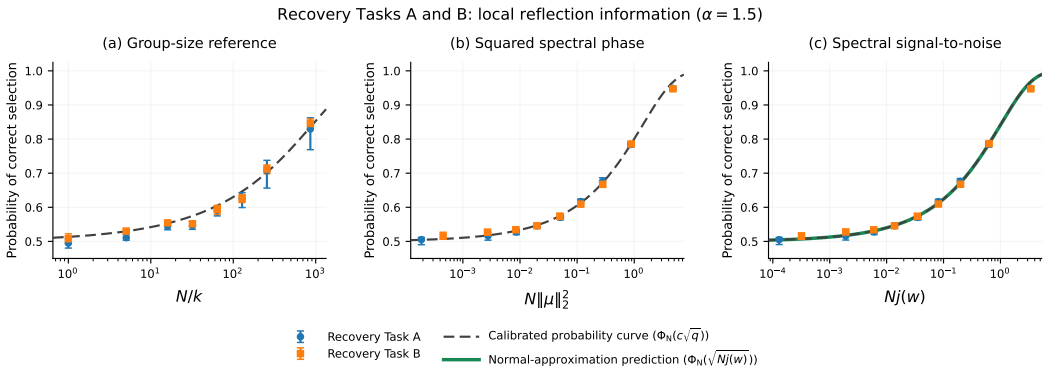

Figure 4: **Local reflection accuracy at $\alpha = 1.5$.** Every panel describes the same held-out decisions. Markers show binned Task A/B accuracies with scene-bootstrap 95% intervals. Grey dashed curves show $\Phi_N(c\sqrt{q})$, with c estimated from separate calibration scenes and shared across tasks. The green solid curve in (c), drawn beneath the calibrated curve, shows $\Phi_N(\sqrt{Nj(w)})$. Appendix G gives the calibration protocol and unbinned prediction errors.

The same feature bank also admits a population interpretation: its mean and second-moment matrix describe a projection of the likelihood contrast between reflected laws. Appendix F.4 proves this identity and connects it to intrinsic testing information. Appendix G.4 evaluates Gaussian, Cauchy, and $\alpha = 0.75$ controls, while Appendix F.8 relates local branch errors to conditional-density error.

7 Related work

Learning from aggregate observations. Zhang et al. (2020) formulate set-level supervision through an induced likelihood and analyse consistency up to observational equivalence. Aggregate-output Gaussian processes infer fine-scale latent functions (Law et al., 2018), while frequency-domain models address spatio-temporal aggregation (Bhowmik et al., 2017). Count Bridges combines aggregate-conditioned sampling with generalized EM to learn unit-level distributions (Fishman et al., 2026). PTSR handles shared unknown offsets through replicate contrasts that estimate conditional shape and orientation.

Invariant reconstruction and phase recovery. Power spectra and bispectra are established tools for reconstruction up to transformations (Kakarala, 2012). In multireference alignment, Bendory et al. (2018) recover a signal from noisy circularly shifted copies through translation-invariant features, and Perry et al. (2019) study the associated sample complexity. PTSR applies the magnitude–phase distinction to independent repeated totals: pairs constrain conditional shape, and triples supply orientation evidence pooled across covariates.

Characteristic-function estimation. Feuerverger & McDunnough (1981) establish efficiency results for empirical-characteristic-function estimators, while Carrasco & Kotchoni (2017) study regularized characteristic-function moment estimation. PTSR uses phase direction predicted by the pair model and analysis connects that score to reflection testing and interprets the sine-feature span through a likelihood-contrast projection (Appendix F.4).

8 Conclusion

PTSR uses pair spectra to estimate conditional shape and bounded triple phase to recover orientation, eliminating shared group offsets before estimation. The heavy-tail experiments show improved recovery over matched winsorized moments, with competitive moment-based recovery in the Gaussian control. Local guarantees explain spectral reliability, and the disconnected-region study identifies spatial coverage as a practical constraint. Together, these results support replicated spectral estimation for location-referenced conditional laws under heavy-tailed aggregate supervision.

AI USE STATEMENT

In this work, we used generative AI tools for tasks requiring disclosure, including generation of synthetic-experiment code, development and refinement of research methodology and experiments, assistance in formatting and checking proofs,theoretical formulations and mathematical claims, parts of implementation of methods in code, and for suggestions on how to present interpretation of experimental results. We also used generative AI tools for recommended-disclosure tasks including suggesting more informative experimental parameters, editing and refactoring software code for readability and distribution, creating reproducibility artifacts, modifying scientific figures to standardize presentation throughout the manuscript, searching for and summarizing related literature to ensure proper attribution, identifying potentially relevant prior work, editing parts of the manuscript text for improving readability and structure. We reviewed all AI-assisted work. Suggested literature and citations were independently checked against the underlying sources. AI-assisted mathematical arguments suggestions were reviewed by the authors, including verification of assumptions, calculations, and proof steps. AI-assisted code was inspected, tested, and executed by the authors, and reported experimental results were checked against the stored outputs. We take responsibility for the final content of this work, including text, claims, code, figures, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide the information needed to reproduce both the theoretical and empirical results. The complete PTSR estimator, including contrast conventions, frequency scaling, numerical stabilization, kernel pooling, and validation procedures, is specified in Appendix A. Appendix B gives the full synthetic data-generating process, model parameterization, sample sizes, optimization settings, comparator definitions, information access, evaluation metrics, and independent run identifiers. Additional controls and uncertainty calculations are documented in Appendices C–E. Complete proofs and assumptions for the theoretical results are provided in Appendix F, with the information-validation protocols and numerical reference calculations in Appendix G. We additionally provide anonymous supplementary source code, fixed experiment configurations, plotting scripts, and the frozen numerical outputs used to regenerate the reported tables and figures. All experiments use explicitly specified synthetic distributions and recorded random seeds/run identifiers, so no external dataset preprocessing is required.

REFERENCES

Tamir Bendory, Nicolas Boumal, Chao Ma, Zhizhen Zhao, and Amit Singer. Bispectrum inversion with application to multireference alignment. *Trans. Sig. Proc.*, 66(4):1037–1050, February 2018. ISSN 1053-587X. doi: 10.1109/TSP.2017.2775591. URL https://doi.org/10.1109/TSP.2017.2775591.

Rabi N Bhattacharya and R Ranga Rao. *Normal approximation and asymptotic expansions*. SIAM, 2010.

Avradeep Bhowmik, Joydeep Ghosh, and Oluwasanmi Koyejo. Frequency domain predictive modelling with aggregated data. In *Artificial Intelligence and Statistics*, pp. 971–980. PMLR, 2017.

Marine Carrasco and Rachidi Kotchoni. Efficient estimation using the characteristic function. *Econometric Theory*, 33(2):479–526, 2017.

Jon Danielsson. *Financial risk forecasting: The theory and practice of forecasting market risk with implementation in R and Matlab*. John Wiley & Sons, 2011.

Andrey Feuerverger and Philip McDunnough. On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 20–27, 1981.

Nic Fishman, Gokul Gowri, Tanush Kumar, Jiaqi Lu, Valentin De Bortoli, Jonathan Gootenberg, and Omar Abudayyeh. Count bridges enable modeling and deconvolving transcriptomic data. In *International Conference on Learning Representations*, volume 2026, pp. 129437–129477, 2026.

Adityanand Guntuboyina, Sujayam Saha, and Geoffrey Schiebinger. Sharp inequalities for f -divergences. *IEEE Trans. Inf. Theor.*, 60(1):104–121, January 2014. ISSN 0018-9448. doi: 10.1109/TIT.2013.2288674. URL https://doi.org/10.1109/TIT.2013.2288674.

George W Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

Ramakrishna Kakarala. The bispectrum as a source of phase-sensitive invariants for fourier descriptors: a group-theoretic approach. *Journal of Mathematical Imaging and Vision*, 44(3):341–353, 2012.

Ho Chung Law, Dino Sejdinovic, Ewan Cameron, Tim Lucas, Seth Flaxman, Katherine Battle, and Kenji Fukumizu. Variational learning on aggregate outputs with gaussian processes. *Advances in neural information processing systems*, 31, 2018.

Frank Nielsen. Revisiting chernoff information with likelihood ratio exponential families. *Entropy*, 24(10):1400, 2022.

Amelia Perry, Jonathan Weed, Afonso S Bandeira, Philippe Rigollet, and Amit Singer. The sample complexity of multireference alignment. *SIAM Journal on Mathematics of Data Science*, 1(3): 497–517, 2019.

Gennady Samoradnitsky and Murad S Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*. Routledge, 2017.

Yivan Zhang, Nontawat Charoenphakdee, Zhenguo Wu, and Masashi Sugiyama. Learning from aggregate observations. *Advances in Neural Information Processing Systems*, 33:7993–8005, 2020.

A NUMERICAL IMPLEMENTATION OF SPECTRAL SCORING AND POOLING

This section records the numerical forms used for the pair targets, triple evidence, and spatial pooling in Section 3. The normalized contrasts in Section 2.2 and the raw contrasts stored by the implementation give the same Fourier features after the frequency conversion below.

A.1 CONTRAST UNITS AND FREQUENCY CONVERSION

Our code forms

$$D^{\mathrm{raw}} = Z^{(1)} - Z^{(2)} = \sqrt{2}\, D, \qquad C^{\mathrm{raw}} = Z^{(1)} + Z^{(2)} - 2Z^{(3)} = \sqrt{6}\, C.$$

If ν is a frequency used with a raw contrast, the corresponding normalized-contrast frequencies are

$$t_{\mathrm{pair}} = \sqrt{2}\,\nu, \qquad t_{\mathrm{triple}} = \sqrt{6}\,\nu. \tag{16}$$

Consequently,

$$\cos(\nu D^{\mathrm{raw}}) = \cos(t_{\mathrm{pair}} D), \qquad \sin(\nu C^{\mathrm{raw}}) = \sin(t_{\mathrm{triple}} C).$$

The model predictions transform in exactly the same way: evaluating $\Phi^{\mathrm{pair}}(\sqrt{2}\nu)$ gives $|\Phi^{\mathrm{agg}}(\nu)|^2$, and evaluating $\Phi^{\mathrm{triple}}(\sqrt{6}\nu)$ gives $[\Phi^{\mathrm{agg}}(\nu)]^2\Phi^{\mathrm{agg}}(-2\nu)$. Thus changing the contrast convention changes neither the fitted pair loss nor the triple features.

In the stable-mixture recovery implementation, the raw frequency bank is

$$\nu_q = u_q k^{-1/\alpha}, \qquad (u_q)_{q=1}^{7} = (0.20, 0.32, 0.48, 0.68, 0.95, 1.30, 1.75), \qquad \alpha = 1.5.$$

Equation (16) gives the two normalized banks used in the exposition. These frequencies are fixed from the group design before the corresponding validation outcomes are evaluated. The square-root factors normalize contrast coefficients; they use no raw-response variance.

For a displayed raw-contrast density, the same change of units includes the density Jacobian:

$$p_D(d) = \sqrt{2}\, p_{D^{\mathrm{raw}}}(\sqrt{2}d), \qquad p_C(c) = \sqrt{6}\, p_{C^{\mathrm{raw}}}(\sqrt{6}c). \tag{17}$$

Figure 1 uses these normalized axes and density values.

A.2 PAIR TARGETS AND STABILIZED PHASE EVIDENCE

For each observed pair-training vector, the code averages the cosine feature over all $\binom{R_b}{2}$ differences. This is Equation (4) expressed in raw units. In the recovery benchmark, each training context supplies one such replicate vector, so $L_b = 1$. Shared replicates induce dependence among its pair differences; the vector remains the independent sampling unit.

The implemented triple score is

$$z_j^{\mathrm{impl}} = \frac{\widehat{\mu}_j^{\top}\overline{h}_j}{\sqrt{\|\widehat{\mu}_j\|_2^2/N_j + \epsilon_{\mathrm{num}}}}, \qquad \epsilon_{\mathrm{num}} = 10^{-8}. \tag{18}$$

For nonzero $\widehat{\mu}_j$, its relationship to Equation (8) is

$$z_j^{\mathrm{impl}} = z_j \left(1 + \frac{N_j \epsilon_{\mathrm{num}}}{\|\widehat{\mu}_j\|_2^2}\right)^{-1/2}.$$

The factor damps evidence when the predicted phase is small and gives zero when the predicted phase vanishes. It preserves the sign of every nonzero local score. Because pooling combines scores from different anchors, this damping is part of the numerical estimator. The normalization uses the predicted phase norm.

A.3 KERNEL POOLING AND VALIDATION

The Gaussian kernel and implemented pooled score are

$$K_\ell(x, x_j) = \exp\left(-\frac{\|x - x_j\|_2^2}{2\ell^2}\right),$$
$$\widehat{f}_\ell^{\,\mathrm{impl}}(x) = \frac{\sum_j K_\ell(x, x_j) z_j^{\mathrm{impl}}}{\max\{\sum_j K_\ell(x, x_j), \epsilon_{\mathrm{num}}\}}. \tag{19}$$

When the kernel sum exceeds the numerical floor, the kernel weights sum to one, as in Equation (9). The floor stabilizes evaluation when all anchors are far from a query point.

Let $\mathcal{V}$ index validation anchors with independently computed triple scores z_v^{impl}. Write $\langle a_v \rangle_{\mathcal{V}} = |\mathcal{V}|^{-1} \sum_{v \in \mathcal{V}} a_v$. The bandwidth maximizes normalized agreement with the validation evidence:

$$\widehat{\ell} \in \arg\max_{\ell \in \mathcal{L}} \frac{\langle \widehat{f}_\ell^{\,\mathrm{impl}}(x_v) z_v^{\mathrm{impl}} \rangle_{\mathcal{V}}}{\sqrt{\langle [\widehat{f}_\ell^{\,\mathrm{impl}}(x_v)]^2 \rangle_{\mathcal{V}} \langle [z_v^{\mathrm{impl}}]^2 \rangle_{\mathcal{V}} + \epsilon_{\mathrm{num}}}}. \tag{20}$$

For the main recovery runs, $\mathcal{L} = (0.16, 0.22, 0.30, 0.42, 0.58)$. The first maximizing bandwidth in this ordered list is retained in an exact tie. The final branch is positive when $\widehat{f}_{\widehat{\ell}}^{\,\mathrm{impl}}(x) \geq 0$ and negative otherwise. Pair checkpoints and pooling bandwidths are selected from aggregate validation observations; individual outcomes and true branch labels are reserved for evaluation.

B RECOVERY BENCHMARK SPECIFICATION

This appendix specifies the implementation used in Section 4. The general pair and triple objectives are given in Section 3; their contrast-unit conversions, stabilization, and pooling rules are recorded as provided in Appendix A.

B.1 RESPONSE DISTRIBUTION AND TASK FUNCTIONS

Write $\pi = (1 + \eta)/2$. The bounded component in Equation (11) has law

$$A_{\eta,d} = \begin{cases} -(1-\pi)d, & \text{with probability } \pi, \\ \pi d, & \text{with probability } 1 - \pi. \end{cases} \tag{21}$$

Its mean is zero and its two support points are separated by d. With independent symmetric stable noise, the unit characteristic function is

$$\Phi_x^{\mathrm{unit}}(t) = e^{-|\sigma(x)t|^\alpha} \left[\pi(x) e^{-i(1-\pi(x))d(x)t} + (1 - \pi(x)) e^{i\pi(x)d(x)t}\right]. \tag{22}$$

Replacing η with $-\eta$ conjugates this expression. The main recovery experiment fixes $\alpha = 1.5$ and uses

$$\begin{aligned} \sigma(x) &= 0.42 + 0.12\ \mathrm{sigmoid}(1.1x_1 - 0.7x_2), \\ d(x) &= 1.65 + 0.30 \sin(1.3x_1)\cos(0.8x_2), \\ m(x) &= 0.58 + 0.22\ \mathrm{sigmoid}(0.9x_1 + 0.6x_2), \\ \eta(x) &= m(x) \tanh(g(x)/0.12). \end{aligned} \tag{23}$$

The two tasks differ only through

$$\begin{aligned} g_A(x) &= x_1 + 0.35 \sin(1.35x_2) - 0.05, \\ g_B(x) &= \sin(1.2x_1)\cos(1.05x_2) + 0.32 \sin(0.72x_1x_2) - 0.08. \end{aligned} \tag{24}$$

Both tasks are evaluated on $[-1, 1]^2$. The names Task A and Task B identify the two fields without imposing a complexity ordering.

Each independent replicate vector receives offset $M = 3G + V$, where $G \sim \mathcal{N}(0, 1)$ and V independently takes values $-6, 0, 6$ with probabilities $0.1, 0.8, 0.1$. The same offset is added to every total within the vector. There is no additional ε noise in the main recovery experiment. Within a local group all units share the same context and have equal weights; their draws are independent. The stable part of their sum has scale $k^{1/\alpha}\sigma(x)$, which allows aggregate simulation without generating each stable unit draw separately.

B.2 Sampling, model class, and optimization

Table 2 gives the fixed recovery protocol. Scrambled Sobol contexts and separate random streams distinguish training, validation, and evaluation. The smaller triple budgets are nested within the largest stream, so their differences are paired rather than independent repetitions. At each pair context, the eight totals yield 28 pair differences. Their cosine features are averaged into one spectrum target. The replicate vector is the independent unit; its 28 differences are dependent.

Table 2: Recovery-benchmark sampling and optimization settings. Counts for pair data apply separately to each pair size. Exact run identifiers are recorded here for reproducibility.

Quantity	Setting
Pair group sizes	$2, 4$
Pair training / validation contexts per size	$1024/512$
Replicates at each pair context	8
Triple group size	4
Triple training / validation anchors	$72/36$
Triple training groups per anchor	$512, 2048, 8192$
Triple validation groups per anchor	768
Common individual test contexts and outcomes	$12{,}000$
Individual-reference training / validation examples	$24{,}000/6{,}000$
Independent-run identifiers	$900, \ldots, 911$
Hidden layers / width / activation	$3/72/\mathrm{SiLU}$
Maximum pair / direct-model updates	$1500/1600$
Optimizer / learning rate / weight decay	AdamW / 0.002 / 10^{-5}
Nominal batch size / gradient-norm cap	$256/10$

The pair and direct models have the same hidden multilayer-perceptron architecture and three output coordinates r_1, r_2, r_3. They use

$$\sigma_\theta = 0.20 + 0.80\ \mathrm{sigmoid}(r_1), \qquad d_\theta = 0.35 + 2.15\ \mathrm{sigmoid}(r_2).$$

The quotient model outputs $|\eta_\theta| = 0.01 + 0.94\ \mathrm{sigmoid}(r_3)$, whereas direct signed models use $\eta_\theta = 0.96 \tanh(r_3)$. Pair training alternates between the two group sizes and restores the checkpoint with the lowest fixed-frequency pair-validation loss. Direct models restore the checkpoint selected by their own aggregate-validation objective. The original profile learner uses batches of 32 replicate vectors; the refined audit uses 21, reflecting its larger grid cost. Other learners use the nominal batch size, capped by the number of available contexts.

PTSR and Winsorized GMM reuse one frozen quotient checkpoint across all three budgets in a given run and task. Their frequency, stabilizer, and bandwidth settings appear in Appendix A. Validation-based early stopping can reduce the update count.

B.3 Comparator definitions and information access

Matched moment comparator. Winsorized GMM uses the same pair checkpoint and triple observations as PTSR. Let C_{jn} be the normalized triple contrast of Section 2.2. For a clipping level λ, it averages

$$g_\lambda(C_{jn}) = \mathrm{clip}\left(\left[\frac{\sqrt{6}\, C_{jn}}{k^{1/\alpha}}\right]^3 ; -\lambda, \lambda\right), \qquad \lambda \in \{2, 4, 8, 16, 32\}. \tag{25}$$

The factor $\sqrt{6}$ converts the normalized contrast back to the raw contrast used by the code. For each anchor and clipping level, 1,024 independent model simulations from the learned positive candidate estimate the feature mean $\widehat{a}_{j,\lambda}$ and variance $\widehat{v}_{j,\lambda}$. Its local evidence is

$$z^{\mathrm{GMM}}_{j,\lambda} = \frac{\widehat{a}_{j,\lambda}\, \overline{g}_{j,\lambda}}{\sqrt{\max\{\widehat{v}_{j,\lambda}, 10^{-6}\}/N_j + 10^{-8}}}. \tag{26}$$

The variance estimate uses the average squared deviation with divisor 1,024. Separate simulated streams orient training and validation scores. Clipping and bandwidth jointly maximize the aggregate-validation alignment in Equation (20). The simulated direction uses the learned candidate. "Winsorized GMM" denotes this clipped-cubic implementation.

Aggregate likelihood. For a homogeneous group, let $L \sim \mathrm{Binomial}(k, \pi(x))$ count units taking the first atom in Equation (21). Conditional on $L = \ell$, the unshifted sum has location $(k\pi(x) - \ell)d(x)$ and stable scale $k^{1/\alpha}\sigma(x)$. Writing f_α for the standard symmetric stable density,

$$p_{\theta,k}(z \mid x) = \sum_{\ell=0}^{k} \binom{k}{\ell} \pi_\theta(x)^\ell [1 - \pi_\theta(x)]^{k-\ell} \frac{f_\alpha\big([z - (k\pi_\theta(x) - \ell)d_\theta(x)]/[k^{1/\alpha}\sigma_\theta(x)]\big)}{k^{1/\alpha}\sigma_\theta(x)}. \quad (27)$$

LIO/LAO MLE applies the aggregate-likelihood principle of Zhang et al. (2020) using this task-specific density adapter. The released LIO Gaussian regression losses are retained for native implementation checks. The stable-mixture adapter is our implementation. The finite mixture in Equation (27) is exact for the model; its stable log density is evaluated numerically by interpolation of a stored table and a tail approximation.

Nuisance-aware likelihood and raw spectral fitting. Nuisance-profiled MLE maximizes the sum of replicate log likelihoods over a shared scalar shift for each vector. The original implementation uses 21 training and 31 validation grid points around the replicate median. The refined audit uses 81 coarse plus 41 local points during training and 121 plus 61 during validation, with further evaluation at finer grids. It is a numerical profile search rather than a guarantee of global maximization. Raw ECF instead minimizes the squared complex discrepancy between empirical raw-total CFs and $[\Phi^{\mathrm{unit}}_{\theta,x}(t)]^k$. Both unadjusted likelihood and raw ECF fit totals without an explicit model for the shared offset.

Privileged references. Shift-known MLE subtracts the simulated shift before fitting Equation (27). Individual supervision fits the signed family to separate unit-level training and validation samples. The true-law row evaluates the generating distribution. Shift-known and individually supervised models are learned references and may retain finite-sample and optimization error.

B.4 Data coverage in the frozen comparator runs

The raw likelihood, raw ECF, and profiled-likelihood implementations store data by group size. Assigning triple data at $k = 4$ replaces the pair-$k = 4$ entry, leaving pair-$k = 2$ and triple-$k = 4$ data in these pools. Refined profiling and shift-known MLE inherit this pattern. PTSR and Winsorized GMM retain both pair sizes and share the pair checkpoint, triples, validation protocol, and test panel.

PTSR versus Winsorized GMM is therefore the matched branch-statistic comparison. The likelihood and raw-ECF rows retain a different pair-data budget. Their validation also uses fixed prefixes: raw MLE scores up to 96 totals per context, and profiling scores up to 128 replicate vectors per group-size block. The reported results use these frozen choices.

B.5 Density, orientation, and uncertainty metrics

Equation (12) is approximated by the trapezoidal rule on 401 equally spaced response values in $[-60, 60]$, using the same stable-density evaluator for truth and predictions. This is a finite-grid approximation to full-line TV. Test NLL is

$$-\frac{1}{n_{\mathrm{test}}} \sum_i \log \widehat{q}(u_i \mid x_i),$$

with the same held-out u_i for all methods in a task and run. Orbit TV takes the smaller pointwise TV error between the prediction and its reflection before averaging over contexts.

Orientation-region F1 classifies $\eta(x) > 0$ as positive and is evaluated only on the held-out points satisfying $|\eta^\star(x)| > 0.1$. It equals $2\mathrm{TP}/(2\mathrm{TP} + \mathrm{FP} + \mathrm{FN})$, with zero assigned when the denominator vanishes. The source CSV calls this quantity `boundary_f1`; it measures region classification, not contour distance. The same mask is used for every estimator.

Means include all twelve runs. Original 95% intervals use 1.96 times the run-level standard error; paired intervals are computed from within-run method differences. Nested budgets from the same run are dependent. Boxplots retain every run, and representative surfaces use the frozen method-neutral selection rule. Quantitative comparisons use the common held-out panel; surface grids are visual diagnostics.

C Recovery uncertainty and supporting controls

C.1 Run-level uncertainty and matched comparisons

Table 3 expands the main recovery table with 95% intervals calculated from the independent-run standard errors. Each estimate includes all twelve runs. The true-law NLL also varies across runs because it is evaluated on different sampled test responses; its conditional TV and orientation F1 are fixed by definition. Likelihood rows use the frozen comparator pools specified in Appendix B.4.

PTSR and Winsorized GMM share each run's pair checkpoint, triple observations, and test panel. Table 4 therefore calculates uncertainty from within-run differences, rather than from two independent method averages. PTSR has lower TV and individual NLL in every paired run on each task. The mean orbit TV is identical for the two methods, 0.0640 on Task A and 0.0668 on Task B. Their full-TV differences measure the effect of the fitted orientation fields on a common learned orbit.

Table 3: **Run-level uncertainty at the main budget.** Entries are means $\pm$ the 95% interval half-width (1.96 times the run-level standard error). Every row includes twelve runs. Daggered rows inherit the comparator-pool qualification in Table 1.

Method	Conditional TV	Individual NLL	Region F1
Recovery Task A			
True law	0.0000 ± 0.0000	1.5525 ± 0.0052	1.0000 ± 0.0000
Individual supervision	0.0235 ± 0.0019	1.5543 ± 0.0051	0.9951 ± 0.0013
Shift-known MLE†	0.0634 ± 0.0036	1.5657 ± 0.0055	0.9699 ± 0.0110
Nuisance-profiled MLE†	0.2679 ± 0.0673	1.8257 ± 0.1199	0.8820 ± 0.0587
LIO/LAO MLE†	0.3556 ± 0.0005	1.9019 ± 0.0034	0.5508 ± 0.1568
Winsorized GMM	0.1225 ± 0.0085	1.6004 ± 0.0093	0.4325 ± 0.1231
PTSR (ours)	0.0674 ± 0.0025	1.5679 ± 0.0053	0.9601 ± 0.0101
Recovery Task B			
True law	0.0000 ± 0.0000	1.5509 ± 0.0037	1.0000 ± 0.0000
Individual supervision	0.0263 ± 0.0016	1.5532 ± 0.0036	0.9981 ± 0.0008
Shift-known MLE†	0.0638 ± 0.0022	1.5649 ± 0.0039	0.9644 ± 0.0105
Nuisance-profiled MLE†	0.3028 ± 0.0715	1.8942 ± 0.1133	0.7820 ± 0.1071
LIO/LAO MLE†	0.3546 ± 0.0003	1.9012 ± 0.0019	0.4024 ± 0.1792
Winsorized GMM	0.1112 ± 0.0150	1.5940 ± 0.0110	0.5662 ± 0.1396
PTSR (ours)	0.0694 ± 0.0050	1.5676 ± 0.0045	0.9694 ± 0.0049

Table 4: **Matched triple-estimator comparison.** Differences are PTSR minus Winsorized GMM at $N = 8192$. Negative values favour PTSR for TV and NLL; positive values favour PTSR for F1. Intervals use the twelve paired run differences.

Task	Metric	Mean difference	95% interval
A	Conditional TV	-0.0551	$[-0.0642, -0.0461]$
A	Individual NLL	-0.0325	$[-0.0390, -0.0260]$
A	Orientation-region F1	$+0.5277$	$[+0.4060, +0.6493]$
B	Conditional TV	-0.0418	$[-0.0546, -0.0289]$
B	Individual NLL	-0.0263	$[-0.0354, -0.0173]$
B	Orientation-region F1	$+0.4033$	$[+0.2651, +0.5414]$

C.2 Representative surfaces and their interpretation

The frozen Task A and Task B figures show runs 903 and 907. The method-neutral selection standardizes conditional-TV vectors across feasible methods and chooses a run close to the across-run median vector. All columns within a task use that same run. Each panel uses the same (x_1, x_2) coordinates and ticks; the vertical label x_2 appears only in the Truth column of each row. Quantitative claims include every run.

Displayed mean TV averages a 100×100 surface grid; numerical tables average the separate held-out panels. A panel value can therefore differ from a table mean. The top-row saturation measures orientation magnitude. The lower row also measures errors in scale, separation, and mixture weight. The shift-known model remains a learned privileged reference, so it may be outperformed on an individual realization.

C.3 Increasing the triple-stage budget

Figure 5 holds the pair estimate and anchor locations fixed while varying the number of independent triples per training anchor. Between $N = 512$ and $N = 8192$, PTSR's mean TV decreases from 0.0842 to 0.0674 on Task A and from 0.0871 to 0.0694 on Task B. Its smallest-budget result is already below the largest-budget Winsorized GMM result in both tasks.

The budgets are nested within a run, so they are dependent experimental conditions. Boxes display the variation across independent runs at each measured budget; they do not summarize additional measurements between budgets. The likelihood curve in this budget figure is the original profiling implementation. Table 1 instead uses the refined-profile audit at the main budget.

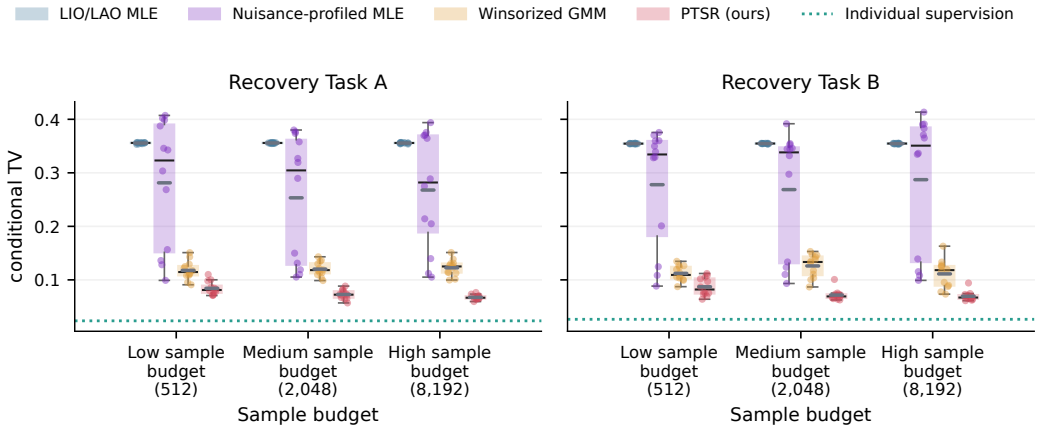

Figure 5: **Triple-stage sample budget.** Low, medium, and high budgets contain 512, 2,048, and 8,192 independent triple groups per training anchor. Boxes show the IQR and median; grey dashes show arithmetic means; translucent points show the twelve independent runs. Whiskers extend to the most extreme observations within 1.5 IQR of the box. The dotted line is the learned individual-supervision reference. Pair data and anchors remain fixed. The likelihood rows retain the data-pooling difference in Appendix B.4.

C.4 Numerical resolution of nuisance profiling

The audit replaces the original 21-point training search by 81 coarse plus 41 local shifts. Validation uses 121 plus 61 candidates, with further post-training checks through 161 plus 81. Details are in Appendix B.3.

Mean TV remains 0.2679 and 0.3028 after refinement, compared with 0.2678 and 0.2872 originally. Both implementations vary substantially across runs (Figure 6). Revealing the shifts instead gives TV 0.0634 and 0.0638, close to PTSR. This pattern is consistent with the value of nuisance elim-

ination within the tested implementations. The audit refines their numerical search; it retains their recorded data pools and does not characterize the best possible nuisance-aware likelihood estimator.

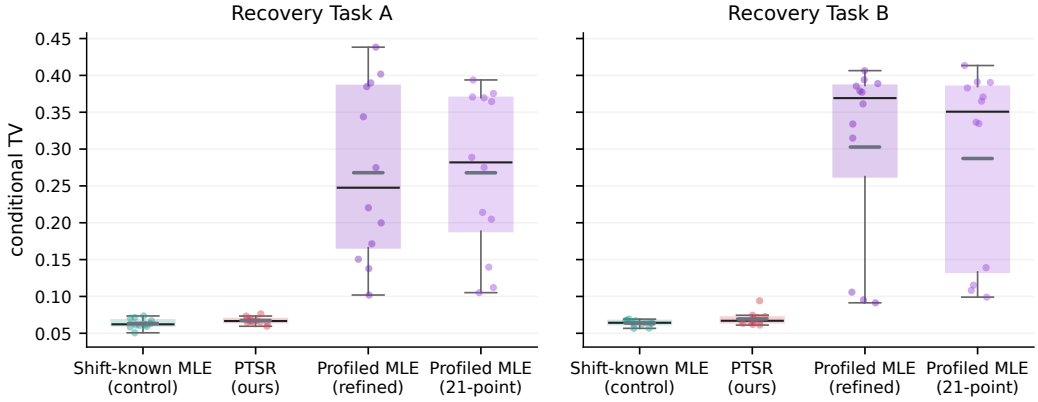

Figure 6: **Nuisance-profile likelihood audit.** Original and refined profiling are compared with a learned shift-known control and PTSR. Boxes show IQR and median, grey dashes show arithmetic means, and points show all twelve runs. Shift-known MLE receives the true simulated offsets. Finer profiling does not improve the recorded reconstruction accuracy. The same data-coverage qualification applies to both profile implementations and to the shift-known control.

C.5 Third-order evidence in the finite-moment control

The Gaussian control uses $U = A_{\eta(x),d(x)} + \sigma(x)E$, where $E \sim \mathcal{N}(0,1)$ and the bounded component and parameter fields are given in Appendix B.1. Each triple receives an independent shared offset $M = 2.5G$, with $G \sim \mathcal{N}(0,1)$. Both estimators receive the true quotient parameters. The control retains the Task A orientation field, triple size $k = 4$, the three evidence budgets, and the aggregate-validation pooling procedure. PTSR uses bounded sine features; the moment comparator uses third-order Hermite evidence. This separates local orientation estimation from error in the learned pair model.

More precisely, writing C for the normalized contrast of Section 2.2, the comparator uses

$$H_3(\widetilde{C}) = \widetilde{C}^3 - 3\widetilde{C}, \qquad \widetilde{C} = \frac{C}{\sqrt{k(\sigma^2 + d^2/4)}}.$$

The denominator is the implemented scale, rather than the exact mixture standard deviation. Positive-candidate simulations calibrate the statistic's direction and variance before the same aggregate-validation bandwidth search. This untruncated third-order control is distinct from the clipped-cubic comparator used with stable noise.

Figure 7 shows similar recovery for both scores. At budgets 512, 2,048, and 8,192, mean F1 is 0.9663, 0.9750, and 0.9816 for GMM, and 0.9623, 0.9746, and 0.9829 for PTSR. Their run-level uncertainty intervals overlap. This establishes a regime in which third-order evidence is effective and places the heavy-tail comparison in context.

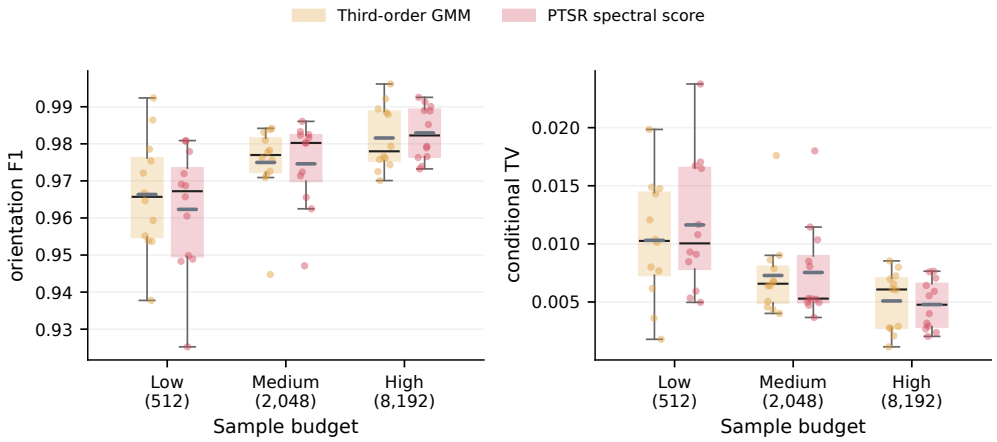

Figure 7: **Gaussian finite-moment control.** Both estimators receive the true quotient parameters and differ in their triple evidence. The three sample budgets count independent triple groups per training anchor. Boxes show IQR and median, grey dashes show arithmetic means, and translucent points show the twelve runs. GMM and PTSR have closely matched orientation recovery.

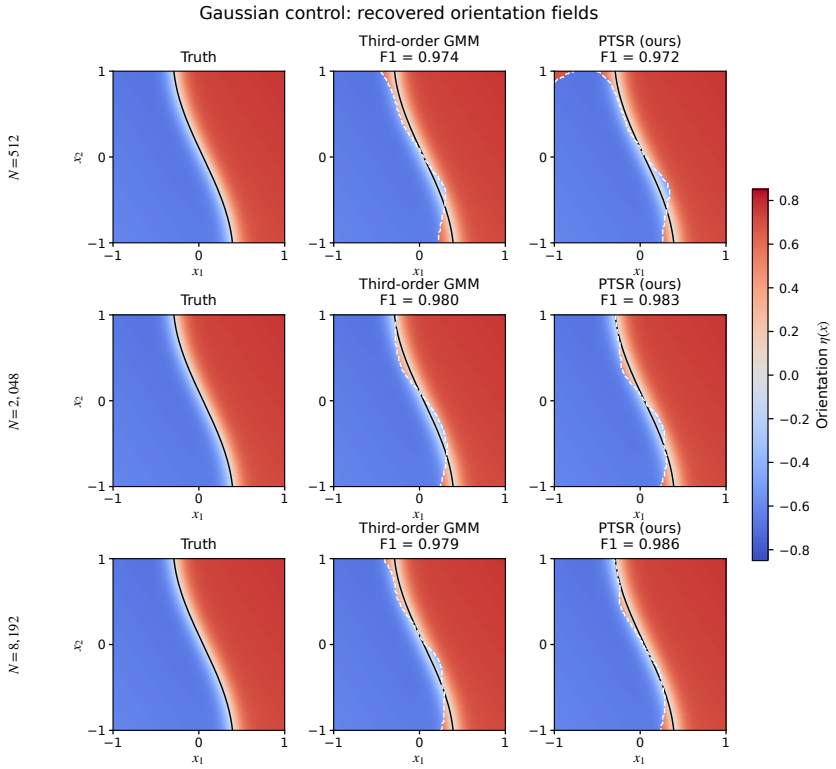

Figure 8: **Gaussian-control orientation recovery.** Rows show $N = 512$, 2,048, and 8,192 triple groups per training anchor. Columns show truth, third-order GMM, and PTSR; all use the same true orientation magnitude. Black contours mark the true zero set and white dashed contours mark the fitted zero set. Panel F1 values are measured on the common held-out context panel using the mask $|\eta^\star(x)| > 0.1$. The displayed run is selected jointly across estimators and budgets from the separate CPU visual replication. Figure 7 retains the original twelve-run boxplots.

Recovered Gaussian orientation fields. Figure 8 complements the run-level boxplots with the estimated fields at each budget. Both estimators follow the true transition at each budget, with small local boundary deviations in the displayed run. Since the supplied quotient fixes $|\eta(x)|$, differences in the coloured regions arise from reflection selection. This control shows why third-order GMM is an informative comparator: when its moment exists, it recovers the orientation effectively.

The field figure uses a separate visual replication of the same Gaussian protocol with twelve independent runs. The boxplots and surface study therefore use separate realizations. Sampling, scoring, and aggregate-validation rules are the same in both studies. One run is chosen by proximity to the across-run median TV vector over both estimators and all three budgets. This method-neutral choice is used for every panel. Both methods receive 72 training and 36 validation anchors, 768 validation triples per anchor, and a common 12,000-context test panel. The candidate bandwidths are $0.16, 0.22, 0.30, 0.42, 0.58$.

D SPATIAL RESOLUTION OF DISCONNECTED COMPONENTS

The main recovery tasks have connected zero boundaries. This experiment asks whether local triple evidence and spatial pooling can recover three separated positive regions. The moderate geometry is the predeclared primary condition; broader and more compact geometries vary component scale.

D.1 DATA-GENERATING PROCESS AND SPATIAL DESIGN

For $x \in [-1, 1]^2$, define the orientation field by

$$g_\zeta(x) = \sum_{\ell=1}^{3} \beta_\ell \exp\left\{ -\frac{(x_1 - c_{\ell 1})^2}{\zeta v_{\ell 1}} - \frac{(x_2 - c_{\ell 2})^2}{\zeta v_{\ell 2}} \right\} - 0.58, \qquad (28)$$
$$\eta_\zeta(x) = m(x) \tanh(g_\zeta(x)/0.12) .$$

Here $m(x)$ is the magnitude function in Equation (23), and Table 5 specifies the three localized bumps. The geometric scale is $\zeta = 1.60$ for broad, 1.40 for moderate, and 1.00 for compact components. It multiplies the squared-width parameters, so the characteristic widths grow as $\sqrt{\zeta}$. The positive set $\{x : g_\zeta(x) > 0\}$ has three connected components in each case. The corresponding grid-area fractions are approximately 20%, 17.5%, and 12%.

Table 5: Parameters of the disconnected orientation field in Equation (28). All three tasks share these coefficients and vary only ζ.

ℓ	β_ℓ	$c_{\ell 1}$	$c_{\ell 2}$	$v_{\ell 1}$	$v_{\ell 2}$
1	1.20	-0.50	-0.32	0.10	0.08
2	1.08	0.50	-0.08	0.08	0.10
3	1.00	0.00	0.60	0.07	0.06

Responses follow $U = A_{\eta_\zeta(x), d(x)} + \sigma(x) S_{1.5}$, with the same $\sigma(x)$, $d(x)$, two-point law, and shared-offset distribution as Appendix B.1. Within each group, independently generated units share the anchor covariates and have equal weights. Only the orientation field changes. The pair sizes are 2 and 4, the triple size is 4, and each training anchor supplies 8,192 independent triples.

The primary comparison uses 192 training and 96 validation anchors, with 1,024 validation triples per anchor. Eight independent runs use identifiers 1200–1207. Scrambled Sobol anchors are sampled independently of the component boundaries. The bandwidth candidates are $0.08, 0.11, 0.15, 0.20, 0.28, 0.38$. PTSR and Winsorized GMM share their pair checkpoint, observed triples, and pooling family. Component metrics use the 192×192 covariate grid shown in the source experiment.

D.2 COMPONENT SCALE AND COMPLETE-DENSITY ERROR

Figure 9 displays all three geometries, and Table 6 summarizes every run. PTSR recovers all broad components in all runs, with mean positive-region F1 0.8319 and component recall 1.000. The

moderate and compact means are less stable, with F1 0.7117 and 0.5984. PTSR has lower TV and higher mean F1 than Winsorized GMM at each component scale.

Unlike the masked orientation F1 in the main benchmark, positive-region F1 here is calculated on the complete surface grid. Component recall matches predicted and true positive components at an IoU threshold of 0.15; predicted speckles smaller than 0.25% of the grid are excluded from component counting. A control that combines the learned quotient with the true sign attains F1 and recall one. Since the sign is supplied, this control checks the representability of the regions and the metric implementation, rather than learned orientation recovery. Its remaining density error measures the learned quotient.

The primary moderate-task F1 falls below its predeclared threshold of 0.75. Its recall exceeds 0.75, and its conditional TV is below 0.10. Retaining all three geometries makes this distinction visible: PTSR improves over the matched clipped-moment estimator throughout the ladder, while reliable recovery of every small component requires stronger spatial evidence.

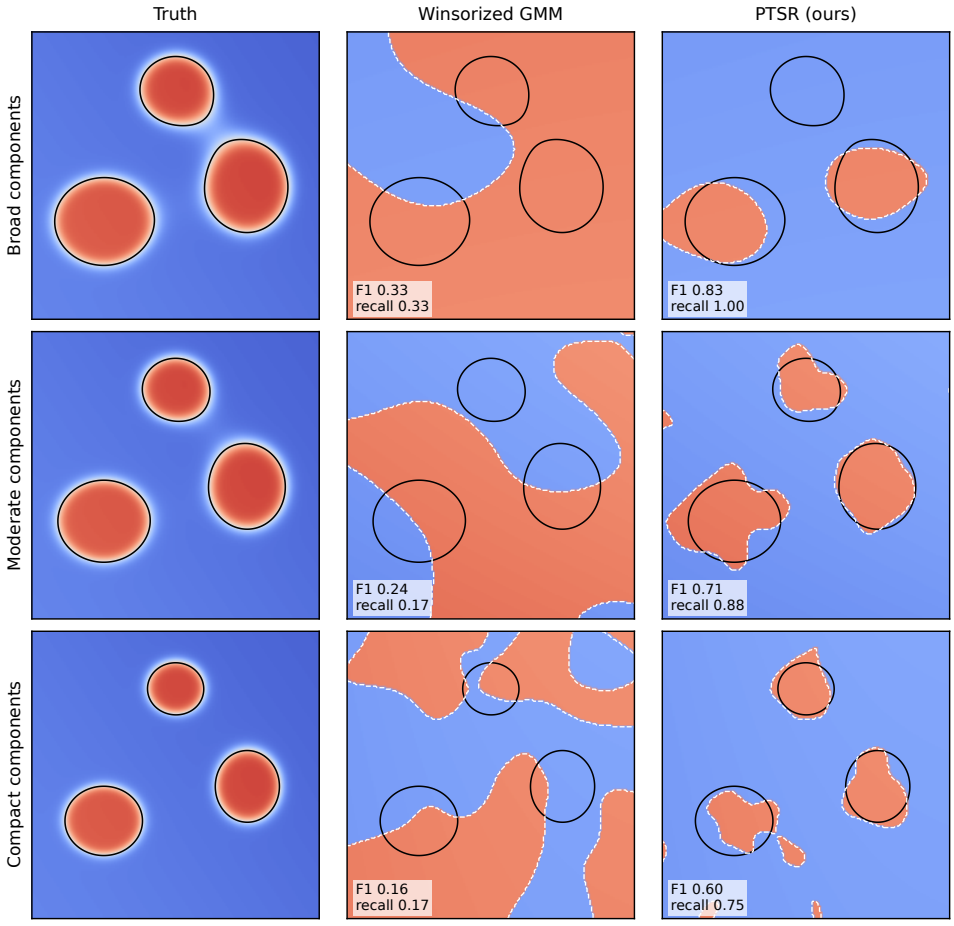

Figure 9: **Disconnected-component difficulty ladder.** Rows show the broad, moderate, and compact fields of Equation (28); columns show the truth, Winsorized GMM, and PTSR. The covariate domain is $[-1, 1]^2$. Colours indicate signed orientation. Black contours mark the true regions and white dashed contours mark fitted regions. These are the frozen representative runs; all eight runs contribute to Table 6 and the stability plot below.

Table 6: **Disconnected components at 192 training anchors.** Means include all eight runs. Positive-region F1 uses the complete surface grid; recall uses component IoU at least 0.15. The primary moderate condition is reported alongside both predeclared scale variants.

Geometry	Method	Conditional TV ↓	Positive F1 ↑	Recall ↑
Broad	Winsorized GMM	0.1187	0.3310	0.333
Broad	**PTSR (ours)**	0.0708	0.8319	1.000
Moderate (primary)	Winsorized GMM	0.1183	0.2363	0.167
Moderate (primary)	**PTSR (ours)**	0.0686	0.7117	0.875
Compact	Winsorized GMM	0.1093	0.1638	0.167
Compact	**PTSR (ours)**	0.0652	0.5983	0.750

D.3 Anchor coverage and the oversmoothed realization

The left panel of Figure 10 retains the full run-to-run variation on the moderate task. Seven of eight PTSR runs recover all three components. The remaining run selects bandwidth 0.38 and predicts the negative branch throughout the domain, producing F1 zero. This realization remains included in every reported mean.

The right panel increases spatial coverage through nested sets of 72, 128, and 192 training anchors, with the per-anchor triple budget fixed. Mean PTSR F1 increases from 0.537 to 0.654 to 0.712, and mean TV decreases from 0.0744 to 0.0708 to 0.0686. This comparison increases both coverage and the total number of triples; it does not hold the total acquisition budget constant.

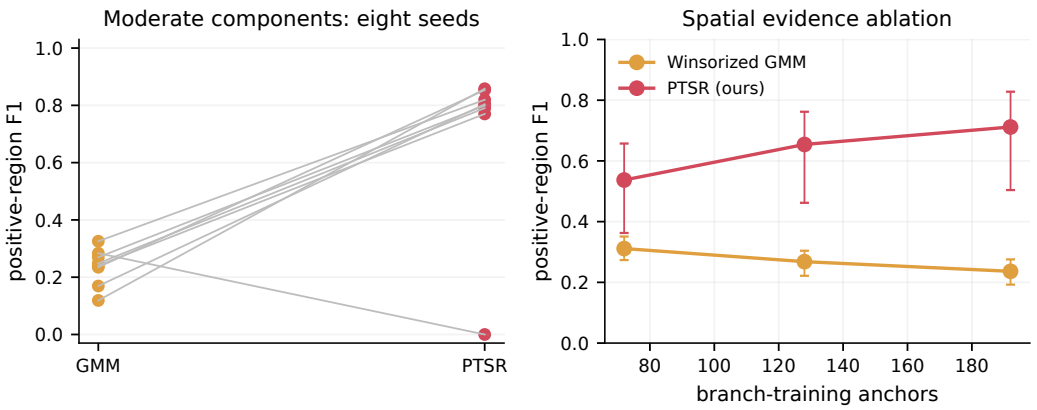

Figure 10: **Moderate-task stability and anchor coverage.** Left: all eight paired runs, including the PTSR zero-F1 realization. Right: nested sets of 72, 128, and 192 anchors, each with 8,192 triples. Intervals use run-level standard errors. Increasing anchor count improves spatial coverage and increases the total triple budget.

Figure 11 compares the frozen representative successful run with the zero-F1 realization. The first uses bandwidth 0.08 and preserves the three positive islands. The second uses 0.38 and pools their evidence into the surrounding negative region. Its error map localizes the loss to the missed components. The coexistence of low average TV and poor component recovery is explained by the small positive area and the broad response densities: integrated density error and small-region topology are complementary measurements.

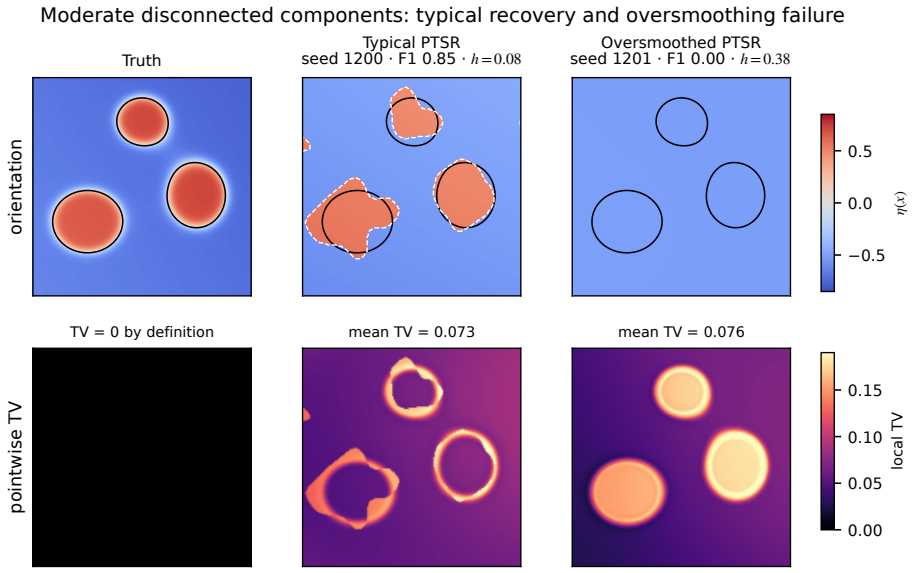

Figure 11: **Successful and oversmoothed PTSR reconstructions.** The moderate truth, representative run 1200, and failed run 1201 are shown with their pointwise-TV maps. The all-negative field is reconstructed from the frozen checkpoint and recorded branch. Both runs remain in Table 6; no model was retrained for the figure.

E CONDITIONAL DENSITIES AT FIXED COVARIATES

Figures 12 and 13 complement the orientation and error maps in Section 5.1. They evaluate the fitted unshifted response densities at four fixed covariate locations. The horizontal variable is the unshifted individual response, labelled y in these frozen plots, rather than either covariate. Each panel title identifies the covariate pair whose conditional law is shown.

The slices distinguish errors in orientation from errors in scale, separation, and mixture weight. In several panels, the raw aggregate-likelihood fit is too diffuse, whereas the profiled fit concentrates excessive mass around narrow peaks. PTSR follows the broader stable-mixture shape more closely in these displayed examples. The numerical table remains the domain-wide comparison: a finite set of slices illustrates local behaviour rather than measuring average accuracy.

PTSR and Winsorized GMM use the same learned pair orbit. Their densities are therefore identical at a point whenever both methods select the same reflection. Where the signs differ, the two curves are reflected representatives of that shared orbit. Overlapping curves consequently show agreement in local branch selection, rather than two independent estimates of all shape parameters.

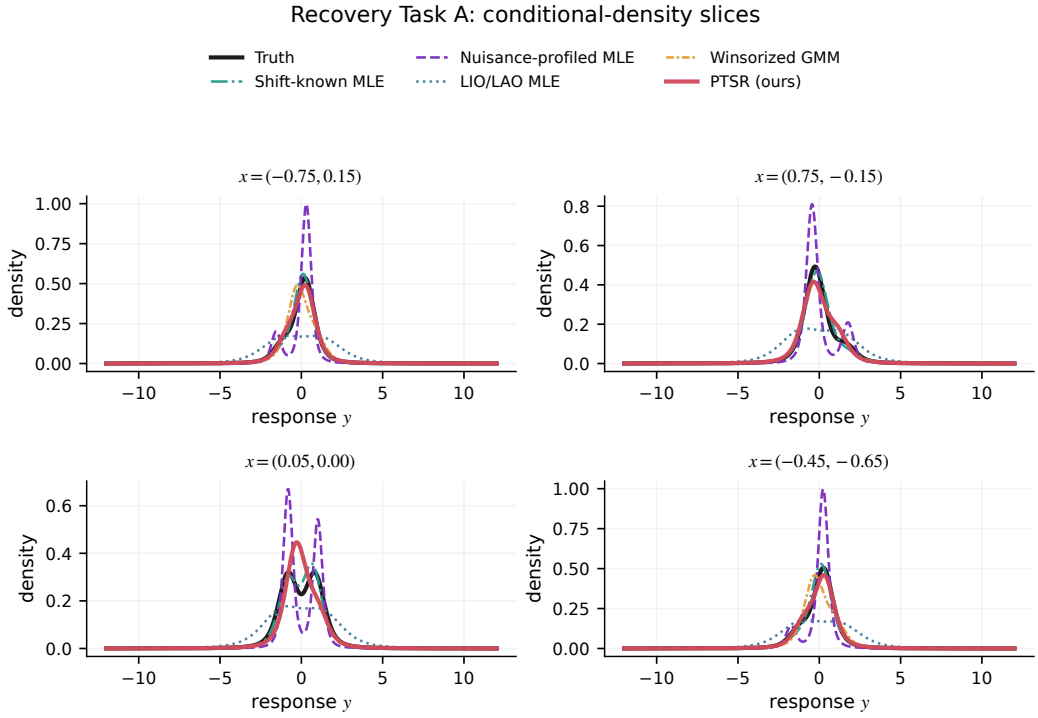

Figure 12: **Conditional-density slices for Task A.** Each panel fixes a covariate pair and plots the density of the unshifted individual response. PTSR and GMM share the pair model; shift-known MLE is a privileged control. The displayed fits use the same representative run as Figure 2. The finite response window emphasizes the mixture shape; the predictive NLL is evaluated at the full held-out response samples.

Task B exhibits the same distinction (Figure 13). A method can place an orientation boundary plausibly while allocating the wrong mass within the conditional density. Reading these slices alongside the local-TV maps shows how that calibration error contributes to the integrated result. Likelihood curves retain the comparator-pool qualification in Appendix B.4.

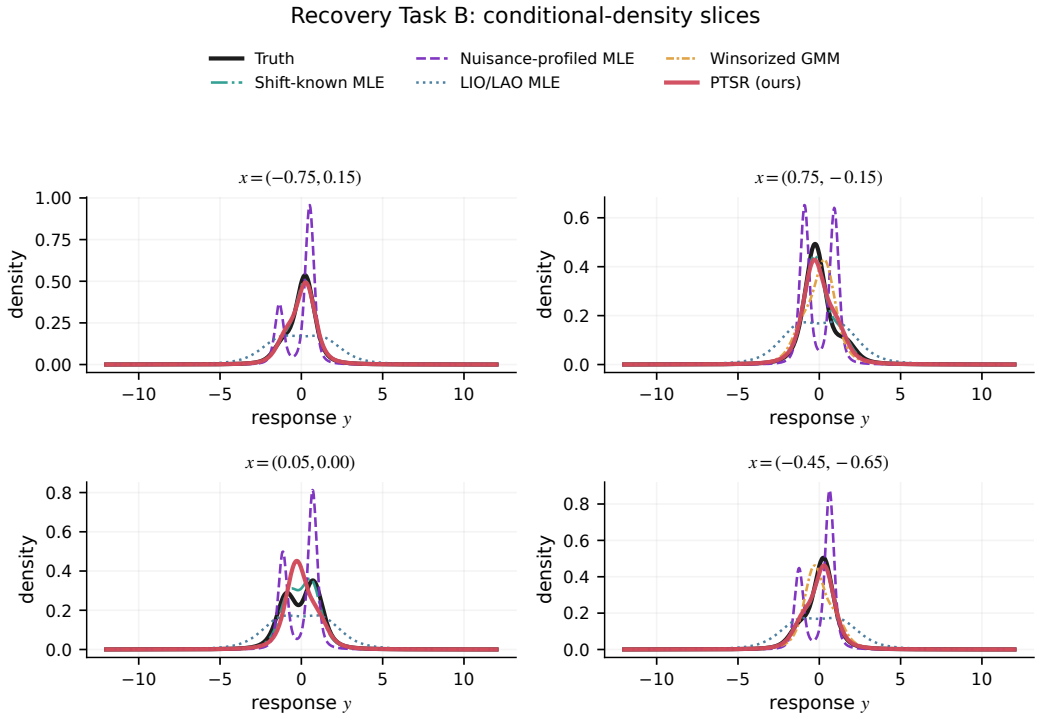

Figure 13: **Conditional-density slices for Task B.** The covariate locations, method order, and density convention match Figure 12; the fitted models correspond to the representative run in Figure 3. Line colour and style are consistent across panels.

F Reflection information and proofs for local selection

The analysis follows the local triple score from its bounded observations to the information in the underlying binary experiment. Throughout, the pair fit, frequency set, and score direction are fixed before the contrasts being tested. Statements for an estimated direction are conditional on this independent fit.

F.1 Feature moments and a fixed spectral score

At a fixed design, let P_+ be the law of the normalized triple contrast C and let $P_- = \mathcal{R}P_+$, where reflection acts by $c \mapsto -c$. Symmetric additional measurement noise preserves this relation. Write

$$h(c) = (\sin(t_1 c), \dots, \sin(t_m c))^\top, \quad \mu = \mathbb{E}_+ h(C), \quad K = \mathbb{E}_+[h(C)h(C)^\top], \quad \Sigma = K - \mu\mu^\top.$$

The product-to-sum identity gives the feature moments directly from the contrast CF:

$$\mu_q = \operatorname{Im} \Phi_+^{\text{triple}}(t_q), \qquad K_{q\ell} = \frac{1}{2}\left\{\operatorname{Re} \Phi_+^{\text{triple}}(t_q - t_\ell) - \operatorname{Re} \Phi_+^{\text{triple}}(t_q + t_\ell)\right\}. \tag{29}$$

The means under P_+ and P_- are opposites. Their second-moment matrices and covariances agree because products of two odd functions are even. All these quantities are finite, regardless of moments of C.

Proof of Proposition 2. Put $X_n = w^\top h(C_n)$ and $L = \|w\|_1$. Each X_n lies in $[-L, L]$. Under P_+ its mean is $a > 0$, so an error is contained in the event $N^{-1}\sum_n (X_n - a) \le -a$. Hoeffding's inequality (Hoeffding, 1963) gives

$$\Pr_+\left(\frac{1}{N}\sum_n X_n \le 0\right) \le \exp\left(-\frac{2N^2a^2}{N(2L)^2}\right) = \exp\left(-\frac{Na^2}{2L^2}\right).$$

Under P_-, the score has the law of $-X_n$ under P_+, so the upper-tail argument gives the same bound. Bounding the non-strict tail also covers ties under the fixed tie convention of Section 3.3. □

The argument uses independence between the tested triple groups. Several overlapping triples extracted from one replicate vector remain dependent and cannot be counted as additional independent observations in this bound.

Proposition 3 (Variance-sensitive selection bound)**.** *For independent contrasts indexed by (b, n), let $X_{bn} = w_b^\top h_b(C_{bn})$ have mean a_b and variance v_b under the positive hypothesis. Suppose*

$$A = \sum_b N_b a_b > 0, \qquad V = \sum_b N_b v_b, \qquad L = \max_b \|w_b\|_1.$$

The test selecting the sign of $\sum_{b,n} X_{bn}$ has each conditional error at most

$$\exp\left\{-\frac{A^2}{2(V + 2LA/3)}\right\}. \tag{30}$$

Proof. Since $|X_{bn}| \le L$, the centered score is bounded in absolute value by $2L$. Apply the one-sided Bernstein inequality to the independent centered sum, with threshold A and total variance V. Reflection reverses every score and preserves its variance, giving the second conditional bound. □

Here all designs test the same binary reflection hypothesis. Evidence about different unknown local signs is instead combined through the spatial procedure in Section 3.3.

F.2 Normal approximation and score normalization

For one fixed design, let $X = w^\top h(C)$, $a = \mathbb{E}_+ X > 0$, $v = \mathrm{Var}_+(X) > 0$, and $\beta_3 = \mathbb{E}_+|X - a|^3$. Boundedness of X implies $\beta_3 < \infty$, even when C has no raw moments.

Proposition 4 (Normal approximation to reflection accuracy)**.** *For N independent copies of X, the sign test obeys*

$$\left|\Pr_s(\widehat{s} = s) - \Phi_{\mathrm{N}}\left(\frac{\sqrt{N}\,a}{\sqrt{v}}\right)\right| \le \frac{C_{\mathrm{BE}}\beta_3}{v^{3/2}\sqrt{N}}, \qquad s \in \{+, -\}, \tag{31}$$

where C_{BE} is a universal Berry–Esseen constant.

Proof. Apply the Berry–Esseen inequality to $N^{-1/2}\sum_n (X_n - a)/\sqrt{v}$ (Bhattacharya & Rao, 2010). The sign threshold corresponds to $-\sqrt{N}a/\sqrt{v}$. Taking the CDF or its left limit at that threshold gives the bound for either tie convention, since the limiting normal CDF is continuous. Reflection gives the statement under the negative hypothesis. Substituting $j(w) = a^2/v$ yields Equation (15). □

The approximation is pointwise for a fixed law and score. Across changing designs, its quality depends on the standardized feature-score moment $\beta_3/v^{3/2}$ as well as on N.

The implemented evidence in Appendix A is a positive multiple of $\widehat{\mu}^\top \overline{h}$. This scalar normalization preserves the local sign and $j(w)$, although it affects the relative scale of evidence passed to the spatial smoother. Variance normalization of a fixed direction would instead divide by $\sqrt{w^\top \Sigma w/N}$. Both use only moments of bounded features.

The information simulations store the raw contrast $C_{\mathrm{raw}} = Z_1 + Z_2 - 2Z_3$. With the convention $C = C_{\mathrm{raw}}/\sqrt{6}$ used in Sections 2–3, set $t_q = \sqrt{6}\,t_{q,\mathrm{raw}}$. Then $\sin(t_q C) = \sin(t_{q,\mathrm{raw}} C_{\mathrm{raw}})$ exactly. Hence μ, K, Σ, the score, and its information are unchanged by this conversion.

F.3 Intrinsic information of a reflection experiment

To compare spectral evidence with the information in the whole contrast observation, define

$$\overline{P} = (P_+ + P_-)/2, \qquad \frac{dP_+}{d\overline{P}} = 1 + \mathfrak{r}, \qquad \frac{dP_-}{d\overline{P}} = 1 - \mathfrak{r}.$$

The symmetrized law $\overline{P}$ is reflection invariant. Its likelihood contrast $\mathfrak{r}$ is odd and satisfies $|\mathfrak{r}| \le 1$ almost surely. These definitions apply to continuous, discrete, and mixed laws. Set

$$\rho(P_+, P_-) = \int \sqrt{dP_+\, dP_-} = \mathbb{E}_{\overline{P}}\sqrt{1 - \mathfrak{r}(C)^2}, \qquad \mathcal{I}(P_+, P_-) = -\log \rho(P_+, P_-). \tag{32}$$

We take $-\log 0 = +\infty$ and use the Hellinger convention $H^2 = 1 - \rho$. This is the Bhattacharyya information; reflection also makes it equal to Chernoff information (Nielsen, 2022).

Theorem 5 (Reflection symmetry and independent evidence)**.** *For reflected laws, the Chernoff coefficient is minimized at exponent $1/2$. For independent observations testing one common reflection hypothesis, with N_b observations of design b,*

$$\mathcal{I}_{\mathrm{tot}} = \sum_b N_b \mathcal{I}(P_{b,+}, P_{b,-}).$$

The equal-prior optimal error R^* satisfies

$$\frac{1 - \sqrt{1 - e^{-2\mathcal{I}_{\mathrm{tot}}}}}{2} \le R^* \le \frac{1}{2} e^{-\mathcal{I}_{\mathrm{tot}}}. \tag{33}$$

Proof. **Midpoint optimality.** With $p_\pm = dP_\pm/d\overline{P}$, define $Z(u) = \int p_+^u p_-^{1-u}\, d\overline{P}$ for $0 < u < 1$. Reflection gives $Z(u) = Z(1-u)$, and Hölder's inequality makes $\log Z$ convex. Therefore

$$\log Z(1/2) \le \tfrac{1}{2}\{\log Z(u) + \log Z(1-u)\} = \log Z(u).$$

The minimum coefficient is ρ. If the affinity vanishes, both informations are infinite.

Independent products. Tonelli's theorem gives multiplication of affinities under product laws. Negative logarithms then give the sum defining $\mathcal{I}_{\mathrm{tot}}$.

Testing bounds. For the product distributions, write their common-measure densities as p, q. The equal-prior Bayes error is $R^* = (1 - \mathrm{TV}(P, Q))/2$. Since $\min(p, q) \le \sqrt{pq}$, we obtain $R^* \le \rho/2$. Also,

$$\mathrm{TV}(P, Q) \le \frac{1}{2}\sqrt{\int (\sqrt{p} - \sqrt{q})^2 \int (\sqrt{p} + \sqrt{q})^2} = \sqrt{1 - \rho^2}$$

by Cauchy–Schwarz. Substituting $\rho = e^{-\mathcal{I}_{\mathrm{tot}}}$ proves the lower bound. □

Thus vanishing total information implies chance-level optimal resolution, whereas total information tending to infinity is equivalent to vanishing optimal error. These conclusions concern the specified contrast experiment. A scalar triple contrast is one projection of the two-dimensional translation-invariant three-replicate observation.

F.4 Information captured by the sine-feature span

The spectral feature moments have a useful geometric meaning. Let $\Pi_{\mathcal{T}}$ be orthogonal projection in $L^2(\overline{P})$ onto the span of $\{\sin(t_q c)\}_{q=1}^m$, and define

$$\mathcal{A}_{\mathcal{T}} = \mu^\top K^\dagger \mu, \qquad \mathcal{A}_* = \mathbb{E}_{\overline{P}}[\mathfrak{r}(C)^2].$$

Here $K^\dagger$ is the Moore–Penrose inverse. The complete energy $2\mathcal{A}_*$ is the classical triangular discrimination, an f-divergence (Guntuboyina et al., 2014).

Theorem 6 (Spectral projection and information certificate)**.** *The spectral energy satisfies*

$$\mathcal{A}_{\mathcal{T}} = \|\Pi_{\mathcal{T}}\mathfrak{r}\|^2_{L^2(\overline{P})} = \sup_{w:\, w^\top K w>0} \frac{(w^\top \mu)^2}{w^\top K w}, \tag{34}$$

$$\mathcal{A}_* - \mathcal{A}_{\mathcal{T}} = \|\mathfrak{r} - \Pi_{\mathcal{T}}\mathfrak{r}\|^2_{L^2(\overline{P})}. \tag{35}$$

In particular, $0 \le \mathcal{A}_{\mathcal{T}} \le \mathcal{A}_* \le 1$*. For* $\mathcal{A}_* < 1$*,*

$$-\tfrac{1}{2}\log(1-\mathcal{A}_*) \le \mathcal{I}(P_+, P_-) \le -\log(1-\mathcal{A}_*). \tag{36}$$

Every finite sine bank therefore certifies

$$\mathcal{I}(P_+, P_-) \ge -\tfrac{1}{2}\log(1-\mathcal{A}_{\mathcal{T}}). \tag{37}$$

If nested frequency sets have a union dense in $(0,\infty)$*, then* $\mathcal{A}_{\mathcal{T}} \uparrow \mathcal{A}_*$*.*

Proof. **Projection.** Oddness of h_q gives

$$\langle \mathfrak{r}, h_q\rangle_{\overline{P}} = \tfrac{1}{2}(\mathbb{E}_+ h_q - \mathbb{E}_- h_q) = \mu_q, \qquad K = \mathbb{E}_{\overline{P}}[hh^\top].$$

The projection normal equations are $Kw = \mu$. If $u \in \ker K$, then $u^\top h = 0$ almost surely under $\overline{P}$, so $u^\top \mu = 0$. The equations are compatible and the minimum-norm solution is $K^\dagger \mu$. Its projected squared norm is $\mu^\top K^\dagger \mu$. Cauchy–Schwarz gives the variational expression, and Pythagoras gives Equation (35).

Information certificate. For $0 \le u \le 1$, $\sqrt{1-u} \ge 1-u$. Jensen's inequality in the other direction gives

$$1 - \mathcal{A}_* \le \mathbb{E}_{\overline{P}}\sqrt{1-\mathfrak{r}^2} \le \sqrt{1-\mathcal{A}_*}.$$

Taking negative logarithms proves Equation (36). Since $\mathcal{A}_{\mathcal{T}} \le \mathcal{A}_*$, the finite-bank certificate follows. The endpoint with energy one follows by continuity.

Complete-frequency limit. Nested spans give monotonicity. To prove completeness, suppose an odd $g \in L^2(\overline{P})$ is orthogonal to the sines at a dense set of frequencies. The signed measure $d\nu = g\, d\overline{P}$ is finite by Cauchy–Schwarz. Its cosine transform vanishes by oddness, and its sine transform vanishes everywhere by continuity. Uniqueness of Fourier transforms of finite measures gives $\nu = 0$. Thus the sine span is dense in the odd subspace, which contains $\mathfrak{r}$, and the projections converge in L^2. □

The gap in Equation (35) is exactly the distinguishing signal missed by the chosen feature span. The certificate can be computed from the CF identities in Equation (29).

F.5 Covariance weighting and estimated candidate laws

Covariance weighting is a standard tool in empirical characteristic-function estimation (Feuerverger & McDunnough, 1981; Carrasco & Kotchoni, 2017). In the reflected experiment it has an exact relation to the projection energy.

Proposition 7 (Optimal linear sine score)**.** *If* $\mathcal{A}_{\mathcal{T}} < 1$*, then*

$$\mathcal{J}_{\mathcal{T}} := \sup_{w:\, w^\top \Sigma w>0} j(w) = \frac{\mathcal{A}_{\mathcal{T}}}{1-\mathcal{A}_{\mathcal{T}}}, \qquad \mathcal{I}(P_+, P_-) \ge \tfrac{1}{2}\log(1+\mathcal{J}_{\mathcal{T}}). \tag{38}$$

On the effective feature span, an optimal direction is proportional to $\Sigma^{-1}\mu$*.*

Proof. For $b = w^\top K w > 0$ and $a = w^\top \mu$, the variance is $b - a^2$. Hence $j(w) = u/(1-u)$, with $u = a^2/b$. This function is increasing for $0 \le u < 1$, and Theorem 6 gives $\sup u = \mathcal{A}_{\mathcal{T}}$. Cauchy–Schwarz in the Σ inner product gives the optimizing direction. Substitution into Equation (37) proves the information bound. □

Here the effective span is the complement of $\ker K$. If all features vanish, set both spectral informations to zero. If $\mathcal{A}_{\mathcal{T}} = 1$, a zero-variance separating score exists and its signal-to-noise is infinite. PTSR instead fixes the phase direction learned from pairs. Its local information is $j(\widehat{v}_j)$, and the optimum above measures the potential of the feature bank.

Proposition 8 (Perturbation of a regularized spectral certificate)**.** *For positive-semidefinite $K, \widehat{K}$ and $\lambda > 0$, define $\mathcal{A}_\lambda = \mu^\top (K + \lambda I)^{-1}\mu$ and $\widehat{\mathcal{A}}_\lambda = \widehat{\mu}^\top (\widehat{K} + \lambda I)^{-1}\widehat{\mu}$. If $\|\widehat{\mu} - \mu\|_2 \le d$ and $\|\widehat{K} - K\|_{\mathrm{op}} \le e$, then*

$$|\widehat{\mathcal{A}}_\lambda - \mathcal{A}_\lambda| \le \frac{2\|\mu\|_2 d + d^2}{\lambda} + \frac{\|\mu\|_2^2 e}{\lambda^2}, \qquad \mathcal{A}_\lambda \le \mathcal{A}_{\mathcal{T}}. \tag{39}$$

Proof. Insert and subtract $\mu^\top(\widehat{K} + \lambda I)^{-1}\mu$. The vector-error terms are bounded using $\|(\widehat{K} + \lambda I)^{-1}\|_{\mathrm{op}} \le 1/\lambda$. The resolvent identity bounds the remaining inverse difference by e/λ^2. Diagonalizing K proves $\mathcal{A}_\lambda \le \mathcal{A}_{\mathcal{T}}$, since μ is orthogonal to its null space. □

With n independent calibration contrasts, entrywise Hoeffding bounds and a union bound give, jointly with probability at least $1 - \delta$,

$$d \le \sqrt{\frac{2m\log(4m/\delta)}{n}}, \qquad e \le m\sqrt{\frac{2\log(4m^2/\delta)}{n}}.$$

These bounds describe simulation error relative to the law generating the calibration contrasts. If a learned contrast law differs from the true law by TV at most ϵ, boundedness adds mean error at most $2\sqrt{m}\epsilon$ and Gram-matrix error at most $2m\epsilon$. This separates numerical estimation of feature moments from error in the pair-estimated model.

F.6 Finite-variance local limit

The moment-free information recovers standardized skewness in a local Gaussian regime. Consider a sequence of aggregates with finite third absolute moments, including their symmetric independent noise. Write

$$V = \sum_i a_i^2 \kappa_2(U_i) + \mathrm{Var}(\varepsilon), \qquad T = \sum_i a_i^3 \kappa_3(U_i), \qquad \Gamma = T/V^{3/2}.$$

For this calculation, normalize the raw triple contrast by $\sqrt{6V}$. This invertible rescaling preserves intrinsic information.

Assumption 1 (Gaussian reflection tangent)**.** As $\Gamma \to 0$ through nonzero values, the symmetrized contrast law and likelihood contrast satisfy

$$\mathfrak{r}(c) = \pm\frac{\Gamma}{6\sqrt{6}} H_3(c) + o_{L^2(\overline{P})}(\Gamma), \quad \mathbb{E}_{\overline{P}} H_3(C)^2 \to 6, \quad \mathbb{E}_{\overline{P}}\mathfrak{r}(C)^4 = o(\Gamma^2),$$

where $H_3(c) = c^3 - 3c$.

This assumption specifies the local tangent and remainder control used by the expansion. Normal and Edgeworth approximation theory provides the corresponding framework for particular sum arrays (Bhattacharya & Rao, 2010).

Corollary 9 (Standardized skewness as local reflection information)**.** *Under Assumption 1,*

$$\mathcal{A}_* = \frac{\Gamma^2}{36} + o(\Gamma^2), \qquad \mathcal{I} = \frac{\Gamma^2}{72} + o(\Gamma^2), \qquad \mathcal{J}_* = \frac{\Gamma^2}{36} + o(\Gamma^2), \tag{40}$$

where $\mathcal{J}_* = \mathcal{A}_*/(1 - \mathcal{A}_*)$ denotes the complete-frequency limit.

Proof. The normalized triple has third cumulant $(1+1-8)T/(6V)^{3/2} = -\Gamma/\sqrt{6}$. The assumption gives $\mathbb{E}_{\overline{P}}\mathfrak{r}^2 = \Gamma^2\mathbb{E}_{\overline{P}}H_3^2/216 + o(\Gamma^2) = \Gamma^2/36 + o(\Gamma^2)$. For $u \in [0,1]$,

$$1 - \sqrt{1-u} - \frac{u}{2} = \frac{u^2}{2(1+\sqrt{1-u})^2}.$$

Substitute $u = \mathfrak{r}^2$ and take expectations to obtain $1 - \rho = \mathcal{A}_*/2 + o(\Gamma^2)$. Expanding $-\log\rho$ gives the intrinsic-information formula. Expanding $\mathcal{A}_*/(1-\mathcal{A}_*)$ gives the last formula. □

For equal-weight homogeneous summands without additional noise, $\Gamma^2 = \gamma_1^2/k$, where γ_1 is unit skewness. Unequal positive weights introduce the factor $(\sum_i a_i^3)^2/(\sum_i a_i^2)^3$. Thus the familiar count and effective-size coordinates arise from a finite-moment local specialization.

F.7 Bounded asymmetry with symmetric stable noise

In the family of Appendix B.1, write $U = A_{\eta,d} + \sigma S_\alpha$, where the bounded two-point variable $A_{\eta,d}$ has mean zero and $\mathbb{E}e^{itS_\alpha} = e^{-|t|^\alpha}$. Its third cumulant is

$$\kappa_3(A_{\eta,d}) = \frac{\eta(1-\eta^2)d^3}{4}.$$

This is a moment of the bounded component. For $\alpha < 2$, the full response has neither a finite variance nor a third absolute moment.

Proposition 10 (Stable-scale spectral sensitivity)**.** *For k independent identically distributed units, $0 < \alpha < 2$, $\sigma > 0$, and $\kappa_3(A_{\eta,d}) \neq 0$, use the normalized triple contrast of Section 2.2. For fixed nonzero u and $t_k = \sqrt{6}\, uk^{-1/\alpha}$,*

$$\operatorname{Im} \Phi^{\text{triple}}_{k,+}(t_k) = e^{-(2+2^\alpha)\sigma^\alpha |u|^\alpha} \kappa_3(A_{\eta,d}) u^3 k^{1-3/\alpha}\{1 + o(1)\}. \tag{41}$$

The associated one-frequency squared signal-to-noise is of order $k^{2-6/\alpha}$.

Proof. Let $m(t)$ be the bounded component's CF. It has a continuous logarithm near zero, and

$$2\log m(t) + \log m(-2t) = -3\kappa_2(A)t^2 + i\kappa_3(A)t^3 + O(t^4),$$

with the next imaginary term of order t^5. Symmetric stable noise contributes the real exponent $-k(2+2^\alpha)\sigma^\alpha|t|^\alpha$ to the raw triple CF. Substituting $t = uk^{-1/\alpha}$ makes that exponent constant. The bounded even term tends to zero because $1 - 2/\alpha < 0$; its leading odd term is $\kappa_3(A)u^3k^{1-3/\alpha}$. Expanding the sine of that phase proves Equation (41) after the contrast-scale conversion.

Finally, the sine variance tends to

$$\frac{1 - e^{-(2+2^\alpha)\sigma^\alpha |2u|^\alpha}}{2} > 0.$$

Dividing the squared mean by this variance gives the stated order. □

At $\alpha = 1.5$ this order is k^{-2}. The result applies to bounded asymmetric structure convolved with symmetric stable noise, including $\alpha \leq 1$ through the fixed location coordinate. All finite-frequency quantities in Equations (29) and (13) remain defined.

The same analysis can detect higher odd structure when the cubic coefficient vanishes. For example, let A take values $(-2, -1, 0, 1, 3)$ with probabilities $(.08, .20, .40, .30, .02)$. Direct substitution gives $\mathbb{E}A = \mathbb{E}A^3 = 0$ and $\mathbb{E}A^5 = 2.4$. For the raw contrast $A_1 + A_2 - 2A_3$, the fifth cumulant is -72, yielding nonzero fifth-order imaginary phase near zero. Independent symmetric stable noise multiplies that phase by a positive real CF factor. Spectral reflection information therefore remains positive even though cubic information is absent.

F.8 From local selection to conditional-density error

Fix learned candidates $\widehat{Q}_{x,+}$ and $\widehat{Q}_{x,-} = \mathcal{R}\widehat{Q}_{x,+}$. Let s°_x minimize their TV distance to the true conditional law and put $d_{\text{orb}}(x) = \text{TV}(\widehat{Q}_{x,s^\circ_x}, Q^\star_x)$. Then any selected field satisfies

$$\text{TV}(\widehat{Q}_{x,\widehat{s}(x)}, Q^\star_x) \leq d_{\text{orb}}(x) + \mathbf{1}\{\widehat{s}(x) \neq s^\circ_x\}\, \text{TV}(\widehat{Q}_{x,+}, \widehat{Q}_{x,-}). \tag{42}$$

If the better learned representative is selected, the first term is the full error. Otherwise, the triangle inequality through that representative gives the display. Averaging over evaluation covariates yields the orbit-plus-branch risk decomposition. This connects the local testing analysis to the density results: remaining pair-shape error and spatial selection error contribute separately.

G Information validation across tail regimes

This experiment compares predicted and observed local reflection accuracy for the response functions of Recovery Tasks A and B in Appendix B.1. The supplied candidate contrast laws determine the spectral feature means and covariances through their analytic CFs. Independent simulated contrasts provide the empirical selection frequencies. All coordinate comparisons describe the same fixed PTSR-form decision rule.

G.1 Design and observation protocol

The response is $U = A_{\eta(x),d(x)} + \sigma(x)S_\alpha$, with the same transition width 0.12 and task functions as the recovery benchmark. We evaluate $\alpha \in \{2, 1.5, 1, 0.75\}$. The Gaussian endpoint uses the convention $\mathbb{E}e^{itS_2} = e^{-t^2}$, so its stable-noise variance is $2\sigma^2$. The Cauchy and sub-unit-index cases use the same specified location coordinate; their raw means are undefined.

The original experiment comprises 3,584 designs. Eight independent scenes form the calibration split and eight form the held-out split. Each scene generates two anchors per task. Designs cross $k \in \{4, 16, 64\}$, uniform or dense unequal positive weights, and symmetric measurement-noise ratios 0 and 0.25. The primary layout repeats the same covariate within each group. An additional layout at $k = 16$ uses nearby heterogeneous covariates. All variants of one scene remain together on the same side of the split.

For each design, 192 independent binary trials are balanced between the two reflection hypotheses. Each trial supplies nested prefixes at $N \in \{64, 256, 512, 1024, 2048, 4096, 8192\}$. Thus comparisons across budgets are paired. The seven base frequencies are $u \in \{0.20, 0.32, 0.48, 0.68, 0.95, 1.30, 1.75\}$, with the frozen aggregate-scale adjustment. Frequencies applied to the stored raw triple contrast convert to the normalized convention by multiplication by $\sqrt{6}$, as in Appendix F.2.

The primary score follows the PTSR phase direction evaluated at the supplied candidate law. The archive also evaluates a numerically regularized covariance-weighted sine-bank score. We report the fixed PTSR-form score here, so all coordinate comparisons predict the same decisions.

G.2 Coordinates, calibration, and uncertainty

Table 7 defines the six displayed coordinates. The count and effective-size quantities describe the measurement design. The spectral quantities use the bounded-feature distribution. The optimal-bank quantity describes the best population linear direction in that bank, which can differ from the fixed PTSR direction.

Table 7: **Information coordinates.** Raw cumulant geometry is used only where its population moments exist. Undefined entries remain explicitly labelled in every regime plot.

Coordinate	Definition and interpretation
N/k	Group-count reference.
$N/k_{\text{eff}}^{(3)}$	Positive-weight reference, with $k_{\text{eff}}^{(3)} = (\sum_i a_i^2)^3/(\sum_i a_i^3)^2$.
$N\Gamma^2$	Squared standardized aggregate third cumulant; defined here only for $\alpha = 2$.
$N\lVert\mu\rVert_2^2$	Squared signed phase in the fixed sine bank.
$Nj(w)$	Squared signal-to-noise of the implemented phase direction.
$N\mathcal{J}_{\mathcal{T}}$	Optimal linear sine-bank information from Equation (38).

For a horizontal coordinate q, the *calibrated probability curve* is $\Phi_{\mathrm{N}}(c\sqrt{q})$. One coefficient c is estimated from the calibration scenes and then held fixed when predicting the held-out scenes. It is shared across Tasks A and B within each tail-regime, scope, and score setting. The grey dashed curves in Figure 4 use this calibration.

For $q = Nj(w)$, Equation (15) instead fixes $c = 1$: the candidate laws supply the score mean and variance, giving the *normal-approximation prediction* $\Phi_{\mathrm{N}}(\sqrt{Nj(w)})$. This is the green solid curve drawn beneath the grey calibrated curve in the information plots. Its comparison with the calibrated curve assesses the probability scale supplied directly by the score's signal-to-noise. A separate phase-only prediction uses $\Phi_{\mathrm{N}}(\sqrt{2N\lVert\mu\rVert_2^2})$; it additionally approximates the sine-feature covariance by $\Sigma \approx I/2$. Both are probability approximations, distinct from the concentration bound in Proposition 2.

MSE is computed from the unbinned design–budget results. Intervals resample whole held-out scenes, keeping together their nested budgets, weights, and group sizes. The binned points in Fig-

ure 4 summarize the same held-out decisions; their horizontal bin edges are fixed using the calibration split. An intrinsic-information reference is computed separately by Fourier inversion. It describes optimal testing of the contrast law rather than a specific linear score.

G.3 Primary heavy-tail experiment

The main figure fixes $\alpha = 1.5$, equal-context groups, and no extra measurement noise. Table 8 reports its unbinned errors. Spectral phase and signal-to-noise both describe reflection resolution much more closely than the two count-based references. After fitting the probability coefficient, the phase and signal-to-noise errors differ by only about 3×10^{-6} in each task; the paired scene-bootstrap intervals for those differences contain zero.

Table 8: **Primary local-information experiment.** Held-out MSE for the fixed PTSR-form score at $\alpha = 1.5$, with no additional measurement noise. The bottom row uses the normal approximation without fitting its probability coefficient.

Coordinate	Probability mapping	Task A MSE	Task B MSE
N/k	Calibrated	0.007624	0.006228
$N/k_{\mathrm{eff}}^{(3)}$	Calibrated	0.005665	0.003800
$N\|\mu\|_2^2$	Calibrated	0.001119	0.001012
$Nj(w)$	Calibrated	0.001116	0.001009
$Nj(w)$	$c = 1$	0.001116	0.001008

The signal-to-noise coefficient fitted across both tasks is 1.002495. Fixing it to one gives almost unchanged MSE, so the green normal-approximation curve nearly coincides with the grey calibrated curve. Figure 14 compares the normal-approximation prediction with the phase-only covariance approximation. The first uses the exact bounded-feature covariance; the second replaces it by $I/2$. Both use supplied candidate laws. Section 5 evaluates the additional effects of pair estimation and spatial pooling in the complete learner.

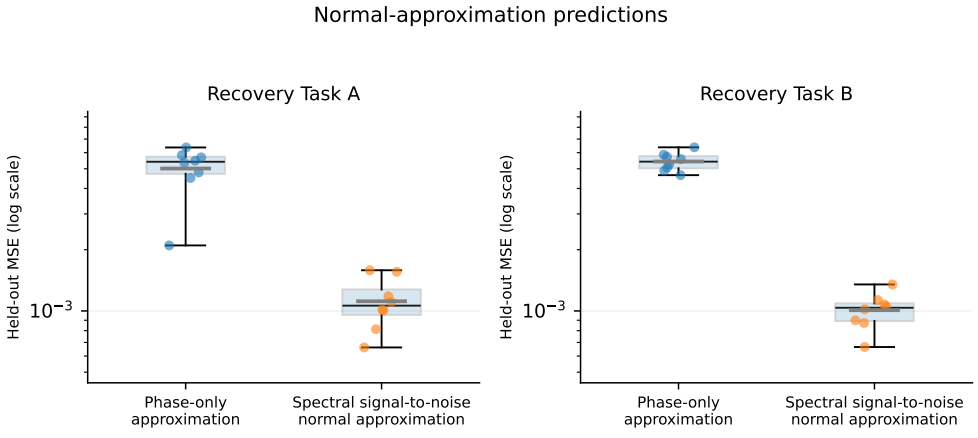

Figure 14: **Normal-approximation predictions for local reflection accuracy.** Spectral signal-to-noise uses the exact bounded-feature covariance; the phase-only prediction approximates it by $I/2$. Both probability mappings are specified without estimating a calibration coefficient and predict the same held-out decisions. Boxes show IQRs and medians across eight scenes; translucent points are scenes and grey dashes are arithmetic means. All values are retained on logarithmic axes.

G.4 Gaussian, Cauchy, and sub-unit-index controls

Table 9 extends the comparison across tail regimes, pooling the two measurement-noise settings within each regime. It reports both observed local accuracy and prediction MSE. These answer dif-

ferent questions: accuracy measures recovery, while MSE measures whether a coordinate correctly describes the available resolution.

Table 9: **Local information across noise-tail regimes.** The same-context control layout includes both symmetric measurement-noise settings. Correct selection is averaged over the measured designs and budgets. The last three columns are prediction MSEs: the count and phase references use calibration-scene fits; SNR fixes its coefficient to one.

α	Task	Correct selection	N/k	Phase	SNR, $c = 1$
2	A	0.706	0.008774	0.000901	0.000895
2	B	0.716	0.004469	0.000930	0.000930
1.5	A	0.620	0.007377	0.001149	0.001146
1.5	B	0.629	0.005810	0.001047	0.001049
1	A	0.548	0.005728	0.001251	0.001251
1	B	0.550	0.005736	0.001248	0.001249
0.75	A	0.521	0.002648	0.001310	0.001309
0.75	B	0.520	0.003038	0.001267	0.001258

Gaussian control. At $\alpha = 2$, all six coordinates are defined. Figure 15 compares their scene-level errors. The standardized third-cumulant coordinate has the local interpretation in Corollary 9; the exact spectral coordinates retain the complete chosen-frequency signal. The normal-approximation SNR MSE is approximately 0.00090 for Task A and 0.00093 for Task B across the control designs.

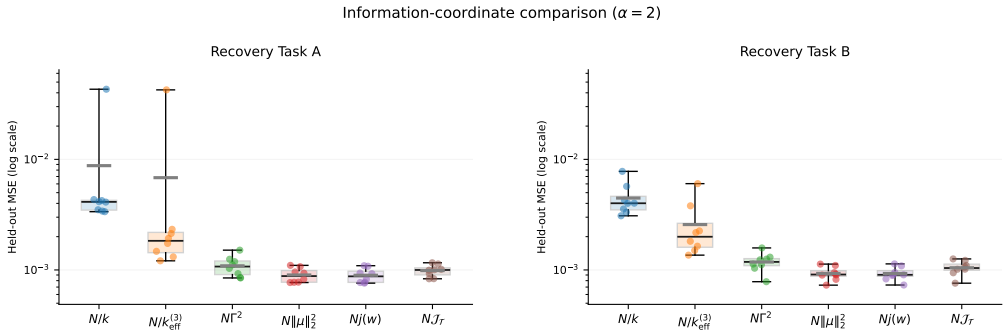

Figure 15: **Gaussian control, $\alpha = 2$.** All six population coordinates are available. Boxes show held-out scene MSE, with medians, translucent scene points, and grey arithmetic-mean dashes. Logarithmic axes preserve extreme values. Each coordinate predicts the fixed PTSR-form score.

Stable noise used by the recovery benchmark. At $\alpha = 1.5$, the full-response variance and third absolute moment do not exist. Figure 16 therefore leaves the $N\Gamma^2$ position labelled undefined. The phase and covariance-standardized score coordinates remain well defined and have similar calibrated error. The optimal-bank coordinate concerns a different direction in feature space, so it need not best predict the fixed score.

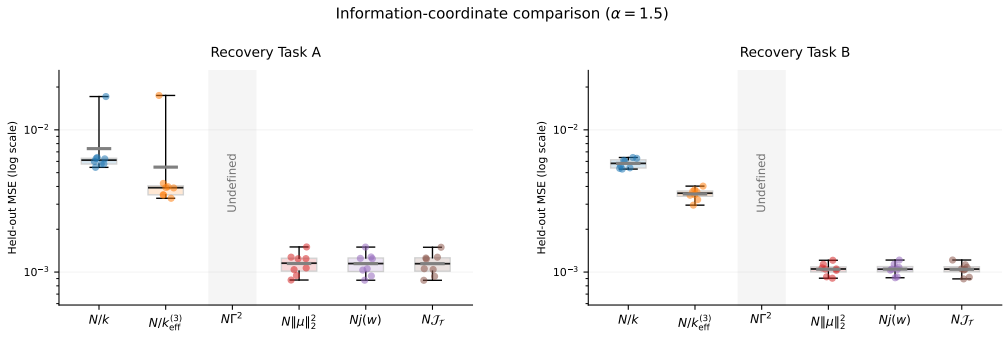

Figure 16: **Stable control, $\alpha = 1.5$.** The empty third-cumulant position marks an undefined population quantity. The remaining coordinates are evaluated on the same held-out decisions. This control pools both measurement-noise settings, whereas Table 8 reports the noiseless main setting.

Cauchy control. At $\alpha = 1$, even the raw mean is undefined. The feature means and covariances remain bounded expectations. Figure 17 shows SNR prediction MSE near 0.00125 on both tasks. Mean correct selection over this design range is about 0.55, indicating that many groups provide weak local orientation evidence. The information calculation correctly reflects this difficult regime.

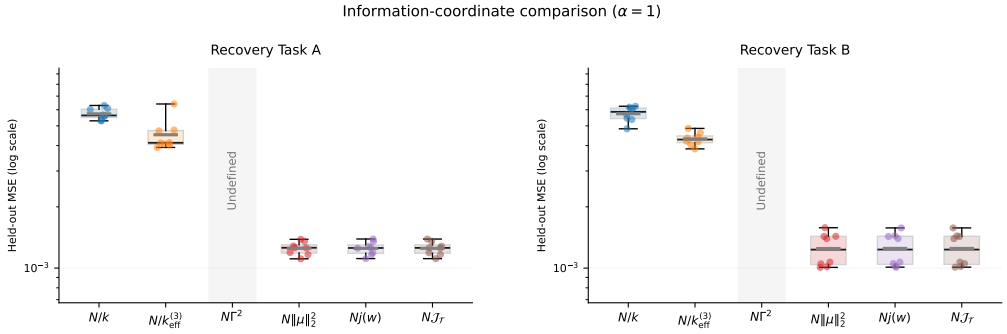

Figure 17: **Cauchy control, $\alpha = 1$.** The spectral quantities remain defined despite the absence of a raw mean or variance. Lower probability-prediction MSE describes the local score more accurately; observed recovery probability is reported separately in Table 9.

Index below one. At $\alpha = 0.75$, mean accuracy is about 0.52 over the tested designs, while the normal-approximation SNR MSE is about 0.00131 and 0.00126. Figure 18 thus validates prediction of weak local evidence in a regime without a finite mean. It evaluates the same binary contrast question as the other information controls, using supplied candidate laws.

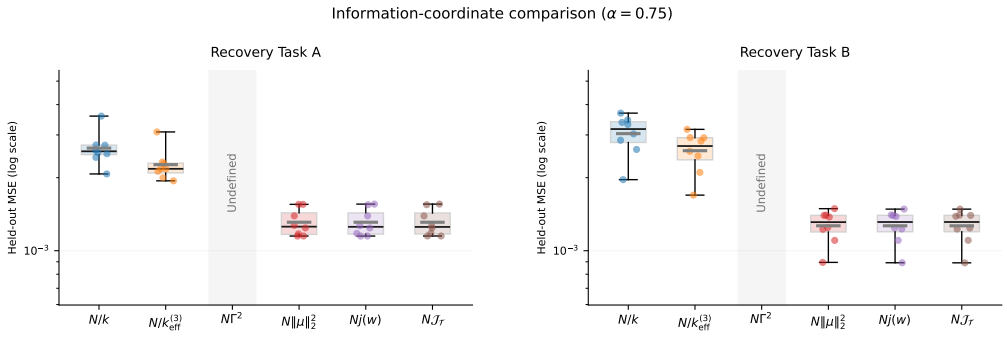

Figure 18: **Sub-unit-index control, $\alpha = 0.75$.** All feature-based quantities remain defined. Mean correct selection is close to chance over much of this design range, as recorded in Table 9. The MSE distributions measure how well each coordinate predicts that limited resolution.

G.5 Transfer across aggregation sizes

A subsequent design-range extension adds 2,560 cases at $k \in \{8, 12, 24, 32, 48\}$ on the same held-out scenes. It preserves the original probability coefficients and the original $k \in \{4, 16, 64\}$ observations. The resulting eight-size grid is therefore a test of transfer to additional aggregate designs with fixed calibration. Figures 19–22 report every measured size and retain all scene values.

Figures 19 and 20 show the Gaussian and main stable comparisons, respectively. The Gaussian third-cumulant coordinate is available only in the first.

The response on each trial is a binary correct/incorrect decision. Its empirical accuracy over 192 trials has a Monte Carlo variance contribution $p(1-p)/192$, with maximum approximately 0.00130. As aggregation weakens the score and success probability approaches $1/2$, this contribution grows. A rise in prediction MSE toward that scale can therefore reflect noisier empirical accuracies even when the probability approximation remains accurate. The boxplots make scene-to-scene variability visible without assuming measurements between the eight tested sizes.

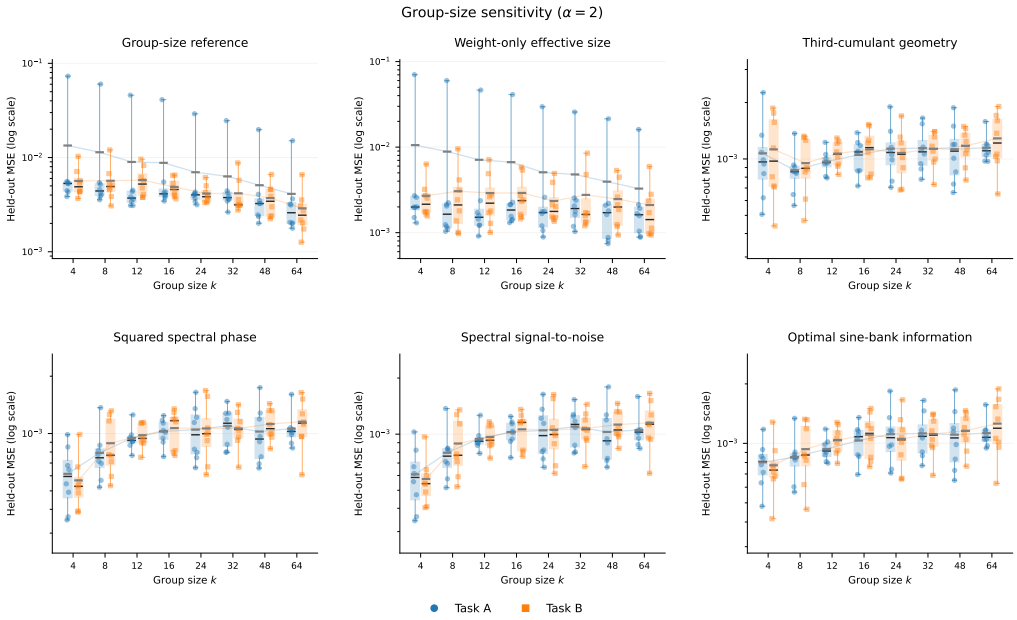

Figure 19: **Aggregation-size transfer under Gaussian noise.** Each box summarizes the eight held-out scene MSEs at a measured k; faint connectors join arithmetic means. Calibration is frozen. The vertical scale is logarithmic and shared across tasks within a coordinate panel, but differs between coordinate panels.

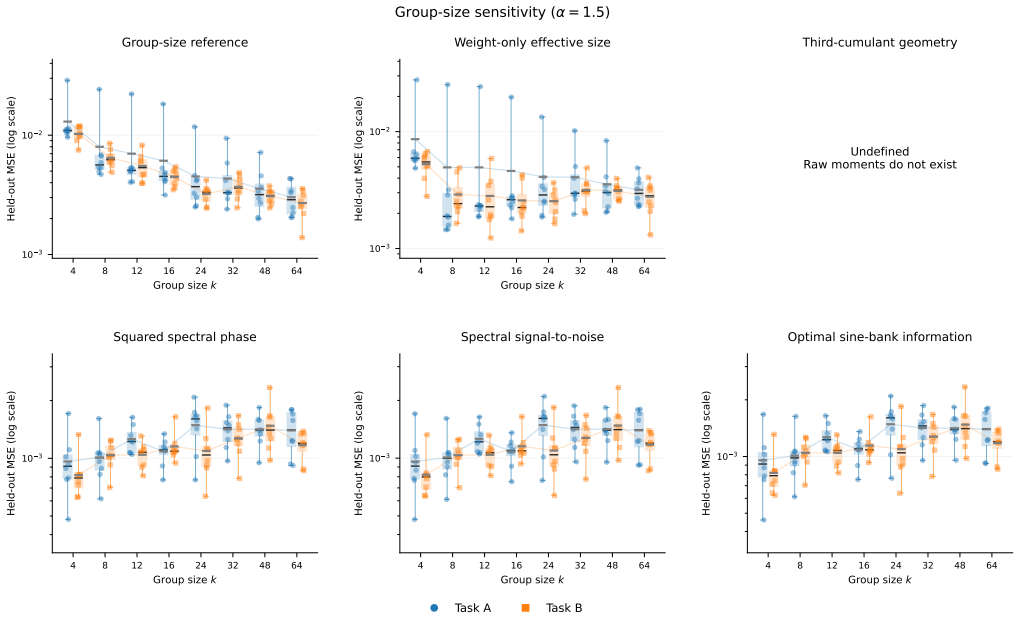

Figure 20: **Aggregation-size transfer at $\alpha = 1.5$.** The third-cumulant coordinate is undefined. Count-based references and bounded spectral quantities use the same scene-level decisions, with the original calibration applied to every added group size.

Under Cauchy and sub-unit-index noise, the bounded asymmetric component becomes increasingly weak relative to the stable sum at large k. Proposition 10 gives its small-frequency scaling. Fig-

ures 21 and 22 show how the probability coordinates behave across the corresponding measured design range. The recovery difficulty and the accuracy of its prediction remain separate quantities.

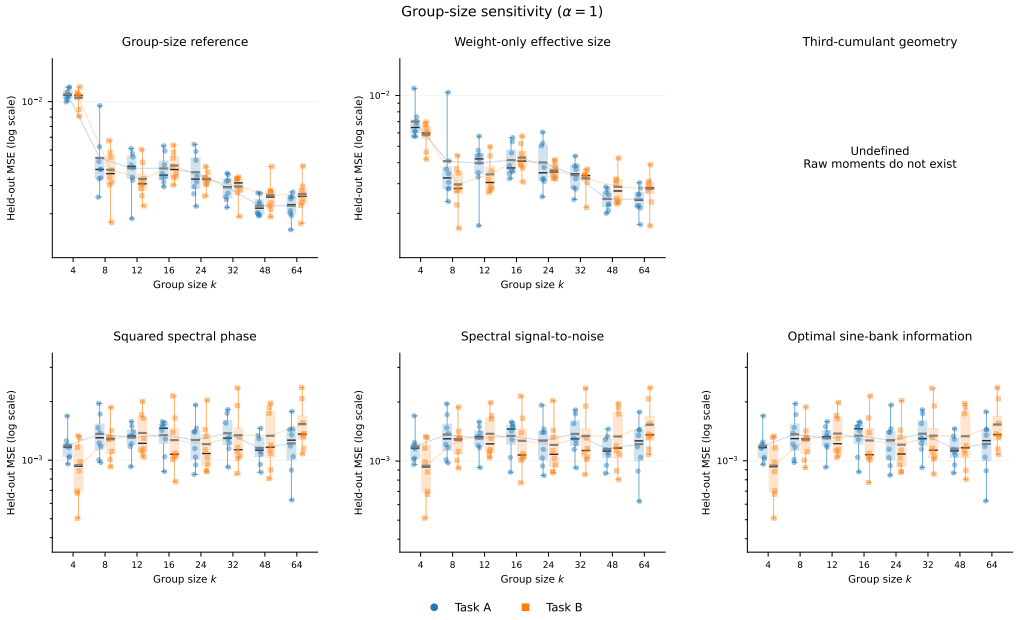

Figure 21: **Aggregation-size transfer under Cauchy noise.** The design-range extension uses the original held-out scenes and coefficients. Boxes summarize scene-level probability-prediction errors; all values remain included on the logarithmic axes.

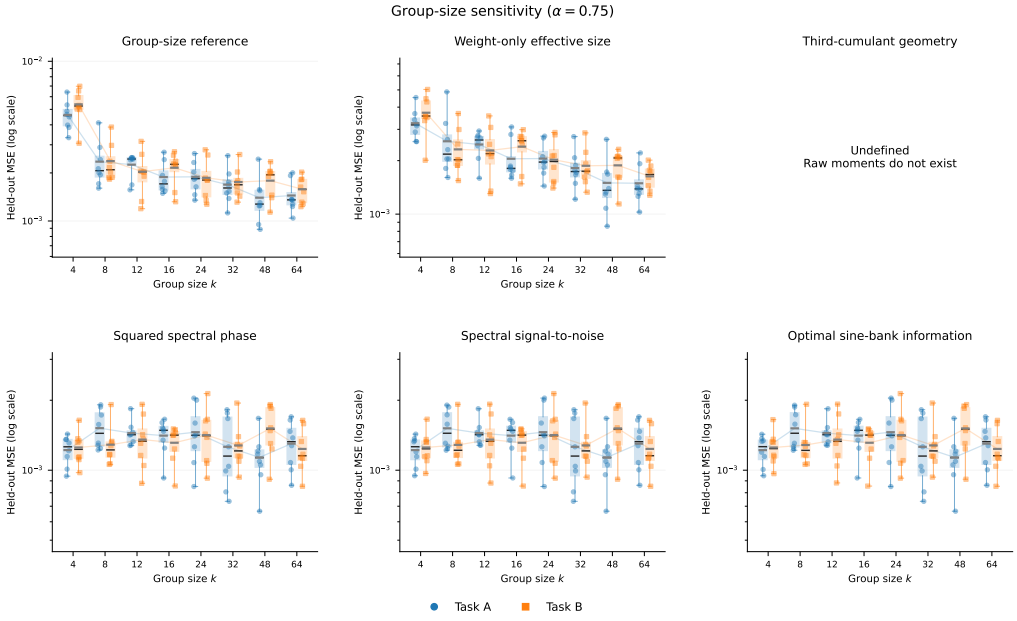

Figure 22: **Aggregation-size transfer at $\alpha = 0.75$.** The spectral coordinates describe local reflection evidence without raw moments. As in the other size panels, the connector indicates the ordering of tested designs and is not a fitted continuous response curve.

G.6 NUMERICAL REFERENCES AND REPRODUCIBILITY

For each candidate law, the feature means and Gram matrices are computed using Equation (29). Intrinsic affinity is evaluated by Fourier inversion on extent 128 with 2^{15} grid points; doubling extent and resolution supplies convergence checks with relative tolerance 0.03 and absolute tolerance 10^{-8}. This numerical reference is kept separate from analytic feature moments and empirical trial accuracy.

The released data contain 50,176 score–design–budget rows from the original 3,584 designs, covering the fixed phase score and the sine-bank reference. The group-size extension is stored separately. Correct-selection counts, coordinate availability, fitted coefficients, and plotted values allow the reported MSEs and figures to be reproduced.

H NOTATION

Symbol	Meaning
b, i, n, r	Group design, unit, independent repetition, and replicate indices.
$x_{bi}, a_{bi}, \mathcal{D}_b$	Unit covariates, known weights, and the complete aggregate design.
k_b, R_b, L_b	Units in an aggregate, replicates in a vector, and independent replicate vectors of a pair design.
$U_{bni}^{(r)}, S_{bn}^{(r)}, Z_{bn}^{(r)}$	Hidden unit response, unshifted aggregate, and observed shifted total.
$M_{bn}, \varepsilon_{bn}^{(r)}$	Arbitrary shared group offset and additional independent measurement noise.
$Q_x^\star, \mathcal{R}Q_x^\star$	Conditional unit law in a fixed location coordinate and its reflection.
$[Q_x], Q_{x,+}, Q_{x,-}$	Reflection orbit and its two representatives, with $Q_{x,-} = \mathcal{R}Q_{x,+}$.
$\Phi_x^{\mathrm{unit}}, \Phi_b^{\varepsilon}, \Phi_b^{\mathrm{agg}}$	Unit-response, measurement-noise, and unshifted-aggregate characteristic functions.
D, C	Normalized contrasts $(Z_1 - Z_2)/\sqrt{2}$ and $(Z_1 + Z_2 - 2Z_3)/\sqrt{6}$.
$\Phi_b^{\mathrm{pair}}, \Phi_b^{\mathrm{triple}}$	Characteristic functions of the normalized pair and triple contrasts. Descriptive superscripts identify the transformed variable.
$\widehat{P}_b(t)$	Empirical cosine average used as a pair-spectrum target.
$\theta, \widehat{Q}_{x,\pm}$	Shared model parameters and fitted reflected candidates.
j, x_j, N_j	Local anchor index, its covariate, and its number of independent triple groups.
$\mathcal{T}_j^{\mathrm{p}}, \mathcal{T}_j^{\mathrm{t}}, t_{jq}$	Pair/triple frequency sets and an individual triple frequency.
$h_j, \overline{h}_j$	Bounded sine-feature vector and its empirical average over independent triples.
$\widehat{\mu}_j, \widehat{v}_j, z_j$	Predicted phase mean, its normalized direction, and local signed evidence.
ϵ_{num}	Numerical stabilizer in the implemented evidence score; distinct from measurement noise.
$\omega_{\ell,j}(x), \ell$	Normalized Gaussian-kernel weight and bandwidth.
$\widehat{f}_\ell, \widehat{s}(x), \widehat{Q}_x$	Pooled evidence, selected reflection sign, and reconstructed conditional law.
$P_+, P_-, \overline{P}$	Reflected laws of a local triple contrast and their symmetrization $(P_+ + P_-)/2$.
$\mathfrak{r}(c)$	Bounded odd likelihood contrast, defined by $dP_\pm/d\overline{P} = 1 \pm \mathfrak{r}$.
$\rho, \mathcal{I}, H^2$	Hellinger affinity, intrinsic information $-\log\rho$, and squared Hellinger distance using $H^2 = 1 - \rho$.
$\mathcal{I}_{\mathrm{tot}}, R^*$	Information accumulated for one common binary hypothesis and its equal-prior optimal testing error.
μ, K, Σ	True bounded-feature mean, uncentered second-moment matrix, and covariance.
w, a, v	Frozen score direction, score mean $w^\top\mu$, and score variance $w^\top\Sigma w$.
$j(w)$	Squared score signal-to-noise; distinct from the anchor index j.
Φ_{N}	Standard normal CDF, distinct from the characteristic functions with descriptive superscripts.
$\Pi_{\mathcal{T}}, K^\dagger$	Projection onto a sine-feature span in $L^2(\overline{P})$ and the Moore–Penrose inverse.
$\mathcal{A}_{\mathcal{T}}, \mathcal{A}_*$	Projected likelihood-contrast energy and complete energy $\mathbb{E}_{\overline{P}}\mathfrak{r}^2$.
$\mathcal{J}_{\mathcal{T}}$	Optimal linear sine-bank information, $\mathcal{A}_{\mathcal{T}}/(1 - \mathcal{A}_{\mathcal{T}})$, when $\mathcal{A}_{\mathcal{T}} < 1$.
$\lambda, \mathcal{A}_\lambda$	Ridge parameter and regularized spectral energy.

Symbol	Meaning (continued)
$\pi, \sigma, d, \eta, m(x)$	Stable-mixture component probability, noise scale, component separation, signed orientation, and orientation envelope.
α, S_α	Stable index and standard symmetric stable variable with CF $e^{-\lvert t\rvert^\alpha}$.
Γ, V, T	Standardized aggregate third cumulant, variance, and third cumulant in the finite-moment specialization.
$k_{\mathrm{eff}}^{(3)}$	Weight-only reference $(\sum_i a_i^2)^3/(\sum_i a_i^3)^2$.
ζ, g_ζ	Component-scale parameter and disconnected-region generating field.
TV, NLL	Conditional total variation and individual negative log likelihood.
MSE	Error in predicting empirical local reflection accuracy, distinct from conditional-density TV.
Orientation-region F1	Positive-region F1 on the specified evaluation subset; stored as `boundary_f1`.
Scene / run	Independent information-calibration cluster / independently trained recovery realization.